

Daniel Van Olmen, Marta Andersson, Jonathan Culpeper and Riccardo Giomi (Eds.)
The Grammar of Impoliteness

Trends in Linguistics Studies and Monographs

Editors

Chiara Gianollo

Daniel Van Olmen

Editorial Board

Walter Bisang

Tine Breban

Volker Gast

Hans Henrich Hock

Karen Lahousse

Natalia Levshina

Caterina Mauri

Heiko Narrog

Salvador Pons

Niina Ning Zhang

Amir Zeldes

Editor responsible for this volume

Chiara Gianollo

Volume 392

The Grammar of Impoliteness

Edited by
Daniel Van Olmen, Marta Andersson,
Jonathan Culpeper and Riccardo Giomi

DE GRUYTER
MOUTON

ISBN 978-3-11-147527-1
e-ISBN (PDF) 978-3-11-147708-4
e-ISBN (EPUB) 978-3-11-147717-6
ISSN 1861-4302
DOI <https://doi.org/10.1515/9783111477084>



This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. For details go to <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

This book is an OA title and might need an additional note regarding the use of figures, photos, etc. Please add the note if necessary:

Creative Commons license terms for re-use do not apply to any content (such as graphs, figures, photos, excerpts, etc.) not original to the Open Access publication and further permission may be required from the rights holder. The obligation to research and clear permission lies solely with the party re-using the material.

Library of Congress Control Number: 2025941730

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the internet at <http://dnb.dnb.de>.

© 2025 the author(s), editing © 2025 Daniel Van Olmen, Marta Andersson,
Jonathan Culpeper and Riccardo Giomi, published by Walter de Gruyter GmbH, Berlin/Boston,
Genthiner Straße 13, 10785 Berlin
The book is published open access at www.degruyterbrill.com.

Typesetting: Integra Software Services Pvt. Ltd.
Printing and binding: CPI books GmbH, Leck

www.degruyterbrill.com
Questions about General Product Safety Regulation:
productsafety@degruyterbrill.com

Content

Introduction

Daniel Van Olmen, Jonathan Culpeper, Riccardo Giomi and Marta Andersson

1 The grammar of impoliteness — 3

From morphemes to periphrastic structures

Elisa Mattiello

2 Impolite suffixoids in English slang — 31

Colin Davis & Hyewon Jang

3 A theoretical and experimental investigation of the morpho-syntax of an anti-honorific prefix in Korean — 61

Nina Dobrushina

4 Blessings and curses are structurally different: Data from Daghestan — 95

Annick Paternoster

5 *Che ti venga* NP, a conventionalised impoliteness formula for Italian disease curses (14th–20th century) — 129

Rita Finkbeiner

6 *Wehe* ‘woe’ + verb-second conditional clause in German: A conventionalized threat construction? — 179

Gerhard B. van Huyssteen, Adri Breed and Suléne Pilon

7 ‘What the hell?!’ vs. ‘Wat de hel?!’: Contrasting the intensifying *whx* construction in English and Afrikaans — 215

Angela Queisser & Monika Pleyer

8 Such an impoliteness: Evidence for the ‘evaluative *such* construction’ — 253

Daniel Van Olmen and Marta Andersson

9 Conventionalized impoliteness in English and Polish: The case of ‘you idiot!’ — 283

Jonathan Culpeper, Isolde van Dorst and Mathew Gillings

10 A corpus-based exploration of British English impoliteness formulae — 307

Conclusion

Marina Terkourafi

11 *What's in a word?* Reflections about impoliteness and future directions — 339

Index — 349

Introduction

Daniel Van Olmen, Jonathan Culpeper, Riccardo Giomi
and Marta Andersson

1 The grammar of impoliteness

Abstract: This chapter introduces the grammar of impoliteness as a field (worthy) of study. It argues that more attention should be paid in the literature to linguistic forms that are specialized for expressing impoliteness, in particular to not purely lexical ones. To frame this type of research, we first discuss how the concept of impoliteness is understood in the field at large and how it is interpreted in the present volume and in its contributions. The chapter then moves on to the notion of grammar, examining how it is viewed in different theoretical frameworks and how those views relate to this volume and the studies that it brings together. We also consider challenges for research into the grammar of impoliteness and outline avenues for future inquiry. The focus here is on issues of a methodological, typological, diachronic and theoretical nature. The chapter concludes with a discussion of the contribution that this volume as a whole makes to the study of the grammar of impoliteness.

Key words: diachrony, grammar, impoliteness, methodology, theory, typology

1 Introduction

(Im)politeness has been studied in fields as diverse as psychology, sociology and neuroscience. The dominant view in linguistics, especially since the “discursive” and “poststructuralist” turns of the research on the topic (e.g. Mills 2003; Locher 2006; Van der Bom and Mills 2015), is that (im)politeness is an essentially socio-pragmatic phenomenon related to the negotiation of societal norms. It is seen as not intrinsic to language but as arising from a situational assessment by the speech

Daniel Van Olmen, Lancaster University, Bailrigg, LA1 4YW, Lancaster, United Kingdom,
e-mail: d.vanolgen@lancaster.ac.uk

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

Riccardo Giomi, University of Amsterdam, Postbus 1642, 1000 BP Amsterdam, the Netherlands,
e-mail: r.giomi@uva.nl

Marta Andersson, Uppsala University, Thunbergsvägen 3 L, 753 10 Uppsala, Sweden,
e-mail: marta.andersson@engelska.uu.se

participants. As a result, issues of linguistic form have not received much attention in the field so far. It would obviously be absurd to claim that context plays no part in (im)polite linguistic behavior (e.g. in banter, *you bastard!* may serve to strengthen rather than challenge the rapport between friends). Still, scholars like Terkourafi (e.g. 2005a) and Culpeper (2011) have argued that no account of (im)politeness can be complete without a thorough understanding of the role of actual linguistic form in it and that there do exist words as well as more complex structures that are, to varying degrees, conventionally associated with (im)politeness. In other words, (im)politeness is not merely of a socio-pragmatic nature in their view: it also has a linguistic component and perhaps even its own grammar. This position has been generally overlooked in the literature. Even Knoblock's (2022) *The Grammar of Hate* volume, for instance, deals primarily with purely pragmatic uses of specific morphosyntactic features for impolite purposes (see Giomi 2023).

The present volume seeks to help redress this neglect of form, by bringing together studies dealing with the grammatical expression of impoliteness in particular. This aim assumes an understanding of impoliteness as well as grammar, of course. Impoliteness could be characterized as involving negatively evaluated (linguistic) behaviors that have (often intentional) offensive effects (see Culpeper 2011: 23) and as encompassing phenomena such as insults, threats, curses, condescensions and reproaches. It is important to acknowledge, though, that impoliteness is a complex notion. Section 2 will therefore discuss it in more detail, with reference to the ways in which it is understood in the contributions to this edited collection. Grammar is not easy to define either. The volume's focus is certainly not on discursive aspects of impoliteness or on individual words like Dutch *eikel* 'dickhead' and ready-made multi-word lexemes like English *son of a bitch*. A linguistic form such as French *espèce de NP!* (lit. 'species of NP!'), by contrast, falls within the present scope, as the structure itself appears to have the potential to create novel insults (e.g. *espèce de linguiste!* 'you linguist!'; Van Olmen and Grass 2023). The exact sense(s) in which something can be regarded as grammar should still be spelled out, however. For that reason, Section 3 will examine the concept of grammar in more depth, in relation to the range of forms studied in this volume.

The grammatical expression of impoliteness merits more attention but is not entirely uncharted territory. The existing literature consists mostly of isolated studies of specific structures in individual languages – especially European and East Asian ones (e.g. Mel'čuk and Milićević 2011 on Bosnian-Croatian-Serbian; Hudson 2018 on Japanese; Mattiello 2022 on English). But few attempts have been made thus far to draw together the research for a more comprehensive picture of grammatical impoliteness and bring it to bear on issues of wider/theoretical significance. Giomi and Van Oers (2022) is a recent exception, with their cross-linguistic survey of structures expressly reserved for direct insults and with their conclusion

that several languages across the world distinguish insults as a sentence type in its own right. Through its different contributions, the present volume too wishes to weigh in on a number of broader issues – of a methodological, typological and diachronic nature, among other things. These challenges for the field will be presented in Section 4. Section 5, finally, will discuss how the various chapters in this collection help address them.

2 Impoliteness

People who research grammar probably assume that the concept of impoliteness is relatively straightforward; and people who research impoliteness probably assume that the concept of grammar is relatively straightforward. Neither assumption, of course, is true. Impoliteness, as a concept, has had a particularly tortuous history. We should note immediately that the label *impoliteness* for this concept is not the only possible one, other candidates being, for example, *rudeness*, *verbal aggression*, *verbal abuse* and *incivility*. Different labels have different nuances of meaning (see Culpeper 2011: Chapter 3) and different disciplines have gravitated toward different labels (*verbal aggression* is important in psychology, for instance). Of course, those are but some of the labels in English. In other languages, we see *scortesia* (Italian), *unhöflichkeit* (German), *kukosa adabu* (Swahili), 失禮 (Mandarin), to name but a few (and needless to say, within each language, there are multiple terms for the notion of impoliteness). An upshot of all this is that we cannot rely on a notion of impoliteness determined by the English lexical item *impoliteness*. Instead, we need a definition of the concept itself. In other words, we need a second-order notion of politeness (a theoretical construct), not first-order (the layperson's commonsense notion), though the latter may shape the former.

It is not the place of this section to attempt to review all second-order definitions of impoliteness. Despite the apparent confusion in the field, we can say of politeness studies that something of a consensus is emerging. Haugh and Watanabe (2017: 67) remark that in politeness studies:

the focus has shifted squarely to politeness as involving 'subjective judgements about the social appropriateness of verbal and non-verbal behaviour' (Spencer-Oatey 2005, 97), and (im)politeness itself is broadly conceptualised as a type of interpersonal attitude or attitudinal evaluation.

Attitudes and evaluations are key. So far so good. The tricky bit, however, is to spell out the factors motivating the attitudes and evaluations that connect with politeness or impoliteness.

Note that one possible key factor is flagged in the quotation above: “social appropriateness”. Schneider (2012) argues that it is key for both politeness and impoliteness, and it also looms large for both in the relational approach espoused by Richard Watts and Miriam Locher (Locher 2004: 51; see also Watts 2003: 19; Locher and Watts 2005: 11; Watts 2005: xliii). However, Culpeper (2011) points out that the term *inappropriate* has a particularly weak link with impoliteness-related terms of the kind mentioned at the beginning of this section. Even for politeness, although appropriateness encompasses much, that very factor makes it vague. Arndt and Janney (1987: 376) argue that “appropriacy-based approaches to politeness” are “too vague”. They suggest that, rather than social situations and their norms of appropriacy, people should be the focus of politeness and we should “focus on cross-modal emotive behaviour as a means by which politeness is negotiated” (Arndt and Janney 1987: 377).

Emotion, in fact, is key in making impoliteness what it is, and thus is one way of making a definition less vague. As Garcés-Conejos Blitvich (2010: 69) points out, with reference to Kienpointner (2008: 41): “we tend to associate impoliteness, but not necessarily politeness, with true emotions”. Most of the impoliteness definitions in this volume allude to emotion but, with Van Huyssteen, Breed and Pilon’s present contribution, it is foregrounded: “(Im)politeness is typically associated with emotive psychological states of mind, emotively motivated human behaviour, perceptions and expectations of what is appropriate or not, (dis)agreeable social interactions and relationships, cultural identity, etc.”. One issue here is: which emotions are we talking about? Knowing this would help us be more precise. Culpeper (2011: Section 2.3) considered this issue for British culture. Culpeper (2011: Section 2.3) characterized and quantified the emotion labels people reported in describing impoliteness events where they have been offended. The vast majority of the emotion labels, 70%, fell into the category which emotion scholars call “sadness”, a category that can be made more transparent by considering the emotion labels it included from the data: *embarrassed*, *humiliated*, *hurt* and *upset*. Henceforth, we will refer to this category as “hurt”. “Anger” was the next most important category, accounting for 14.3% of labels used, included the labels *angry*, *irritation* and *annoyed*. Culpeper et al. (2014) showed that these two emotion groups, hurt and anger, also accounted for the vast majority of emotion labels reported by informants experiencing impoliteness events in Germany, Finland, Turkey and China, though there was slight variation between the weightings of those two emotion groups.

The hurt emotion group brings us into contact with another way of conceiving of impoliteness, i.e. via the concept of face. This is the approach taken by Matiello and Finkbeiner in this volume. When it comes to the academic concept of face, most scholars – including the most cited work on politeness, namely, Brown and Levinson (1987) – connect with Goffman’s (1967: 5) definition of face:

the positive social value a person effectively claims for himself by the line others assume he has taken during a particular contact. Face is an image of self delineated in terms of approved social attributes.

Losing face means that one's public image is damaged, and that often results in emotional consequences. Goffman notes the emotional consequences of face loss at various points: "If events establish a face for him [and] if his ordinary expectations are not filled, one expects that he will 'feel bad' or 'feel hurt'" (Goffman 1967: 6); "He may become embarrassed and chagrined; he may become shamefaced" (Goffman 1967: 8). These clearly involve the hurt emotion group. Impoliteness is a matter of facework that attacks or aggravates face, and indeed the first generation of works on impoliteness all took it this way (see, for example, Lachenicht 1980; Austin 1990; Culpeper 1996, 2005; Kienpointner 1997; Bousfield 2008). Face is closely linked to identity, and so violations of identity are part of facework (see, for example, Garcés-Conejos Blitvich 2013 and references therein). However, face, although a rich and useful notion, does not easily accommodate all impoliteness-relevant negative emotions or the beliefs that give rise to them. Note that the anger emotion group is not directly accommodated by face. People have strong beliefs about social organization and behaviors within social organizations, about how people should be treated, about what is fair and what is not, and so forth. For example, in British culture, the rude act of jumping the queue is not so much a matter of face but perceived to be a violation of the fair and "right" practice of awaiting your turn, and it is something that is likely to provoke the emotional response of anger. Beliefs about rights are underpinned by morality and constitute part of society's "moral order". The concept of moral order is essentially "a culture-specific ideology about what counts as right or wrong" (Culpeper and Tantucci 2021: 148; see also Garfinkel 1964: 225) and is often referred to in recent (im)politeness research (e.g. Parvaresh and Tayebi 2018; Xie 2020).

In more recent years, and keying into the notions of attitude and evaluation, approaches to politeness and impoliteness have generally been more inclusive. Spencer-Oatey's (e.g. 2008; Spencer-Oatey and Kádár 2021) rapport management framework accommodates a range of evaluative beliefs, including some based on types of face and some based on what she terms "sociality rights". In a similar vein but focusing specifically on impoliteness, Culpeper (2011: 23) pulls together the factors that shape impoliteness:

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. Various factors can exacerbate how offensive an impolite behaviour is taken to be, including for example whether one understands a behaviour to be strongly intentional or not.

The papers in this volume not mentioned in this section thus far all orientate to this definition. Of course, there is no claim here that this is the one and only way of defining impoliteness. Indeed, interestingly, some papers in this volume highlight specific parts of it, perhaps as a way of compensating for the fact that the definition's very broadness makes it lose precision. As with most definitions, if not all, there is the problem of infinite regress: the concepts that defined the concept themselves need definition. In the above definition, the notion of "offence" would be a case in point (see Culpeper and Haugh's 2021 attempt to pin it down). Furthermore, impoliteness will always have controversial boundaries and boundaries that are difficult to navigate. For example, Matiello's morphopragmatic analyses in this volume are focused on items that belong to slang. Obviously, much slang is oriented toward in-group membership and positive emotions, not the stuff of impoliteness. Thus, Matiello's analyses required an extra step to identify the items that are "generally perceived as impolite, offensive, and face-threatening, both to the speaker's and to the hearer's face".

3 Grammar

Quite problematically for a research agenda centered around the notion of *grammar of impoliteness*, it is not only the boundaries of the concept of impoliteness that are fuzzy and potentially controversial: decades, if not centuries of debate in the (narrowly or broadly defined) field of language studies go to show that exactly the same is true of the notion of grammar. And after all, if this was not the case, there probably would not be so many different linguistic theories around. What most of these theories have in common is that, in one way or another, they describe the grammar as a structured set of constraints and operations that has some sort of psychological reality in the mind of language users. What is far from consensual, on the other hand, are the workings of these constraints and operations and the nature of the building blocks to which they apply.

On a restrictive (and usually prescriptive) approach, grammar tends to be equated with morphological and syntactic structure only. This has been referred to as "the traditional sense [of the word "grammar"] in linguistics, and the usual popular interpretation of the term" (Crystal 2008: 217). Crystal is probably right in submitting that this is also the concept of grammar (however vague and uncon-

scious) that one may tentatively ascribe to most laypeople. And it also does not seem too far-fetched to say that this pre-theoretical conception of grammar is precisely what informs the vast majority of the grammars (intended as grammar *books* for individual languages) that are used in language teaching, of both L1 in basic schooling and L2, for whatever age range. As soon as the question “what is grammar?” is taken to the next, more theoretically-oriented level of reflection, however, the picture becomes more complicated.

To start with, at least some implicit recognition that grammar is more than just syntax and morphology is to be found in most contemporary frameworks. This is also the case for Generative Grammar, the theory that *par excellence* tends to offer the most restrictive definition of grammar “proper”, essentially equating it (at least in some of its versions) with syntax, or at best *morphosyntax*. As a matter of fact, the practice of sticking semantic labels such as *tense*, *aspect*, *modality* etc. onto the various syntactic nodes has been ubiquitous in generative grammarians’ famous tree diagrams since the early days of the framework. This is, in itself, already quite meaningful. And when this practice has been criticized, as in Ray Jackendoff’s Parallel Architecture (Jackendoff and Audring 2019, 2020), this was done in the spirit of arguing that syntax and semantics constitute separate modules of the grammar (not that syntax is *the only* component of the grammar). In fact, the status of the so-called *Logical Form* with respect to the grammatical system has been a topic of debate between proponents of different versions of Generative Grammar (see Hornstein 1995: 3–4).

At face value, one may be tempted to take as a starting point the traditional divide of the linguistics world into formalists and functionalists, and automatically ascribe to the former the assumption that grammar is essentially concerned with the formal properties of language, and to the latter the competing assumption that grammar encompasses at least a certain amount of meaning representation. In fact, this would be an utter oversimplification of a much more complex and diverse landscape. Not only have various formally-oriented linguists proposed that what they call Logical Form is a level of grammatical analysis in its own right, but the other way round, there also are linguistic frameworks such as Halliday’s Systemic Functional Grammar (Halliday and Matthiessen 2014) that typically focus much more on the meaning and function of linguistic utterances rather than their form, but which all the same regard semantics and *lexicogrammar* as two sharply separate systems. On this approach, the lexicogrammar *realizes* a semantic structure, but the latter is not *part* of the former. What Crystal (2008) refers to as the “traditional” view of grammar, in sum, is still very much alive and kicking, even in some of the otherwise most radically functionalist framework.

By contrast with the restrictive approach, in other frameworks the grammar is explicitly argued to include not only the strictly semantic (i.e. denotational,

truth-conditional) properties of linguistic expressions but even (some of) their pragmatic properties. In Functional Discourse Grammar (Hengeveld and Mackenzie 2008), for instance, the grammatical component is comprised of an Interpersonal and a Representational Level (dealing with pragmatics and semantics, respectively) which hierarchically govern the Morphosyntactic and the Phonological Level; precisely for this reason, the first two levels do not encompass any possible facet of an utterance's meaning but are restricted to those aspects of pragmatic and semantic content that receive overt encoding in linguistic form, whether morphosyntactically, phonologically, or both (e.g. the different syntactic templates for declarative and interrogative sentences in English and the falling versus rising prosodic contours associated with these two sentence types in so many languages). This point is particularly important here, because much (though certainly not all) of what is generally understood to belong to the realm of impoliteness is pragmatic, rather than semantic in nature: not, or not necessarily in the sense that it is strictly a matter of discourse (and as such not relevant to the grammar) but in the sense that it concerns what the speaker *does* with their utterance rather than what they *describe*. After all, the notion of 'doing' is precisely what the word *pragmatics* is literally about: and there is nothing in this notion that is inherently in contrast with the possibility of being conventionally associated with a given linguistic form. One fundamental assumption of what we have referred to as the *grammar of impoliteness* research agenda is precisely that some aspects of grammatically encoded meaning are not denotational but rather interpersonal in nature (i.e. pragmatic in the sense just described). For instance, the descriptive, truth-conditional content of a curse or threat may not differ at all from that of a mere statement about the future: what distinguishes curses, threats and predictions from each other is that they each realize a different type of communicative action, i.e. a different speech act. To the extent that such speech acts are explicitly indicated by dedicated formal means, such as the morphological markers or syntactic constructions discussed in the chapters by Dobrushina, Finkbeiner and Paternoster in this volume, there is no reason why the illocutionary distinctions in question should not be regarded as *bona fide* grammatical features of the language at stake.

It should be stressed that, once again, this perspective is not necessarily restricted to functionally-oriented approaches. Clearly pragmatic notions such as topic, focus and illocutionary force are nowadays an integral part of the hierarchy of functional projections assumed in generative syntax, and some generative accounts have proposed further interpersonal concepts as part of the grammar, whether as syntactic nodes in their own right or as features associated with certain items or positions within a syntactic tree. This is precisely what is suggested in the one chapter of the present volume that adopts a generative approach, authored by Davis and Jang: in Korean, the features [+honorific] and [–honorific], which

encode the speaker's subjective evaluation of a referent rather than its objective, truth-conditional properties, are specified in the head position of a noun phrase (and trigger the use of the prefixes *si-* and *che-* on a verb agreeing with that noun phrase), so they are also understood as being part of the grammar to all intents and purposes. Even though their meaning is non-truth-conditional, it is still not defeasible, i.e. it is not an inference but an inherent specification of the linguistic forms expressing them: in other words, such meanings correspond to what Grice (1975) called conventional implicatures, and which would later come to be known as use-conditional meanings (Recanati 2004; Gutzmann 2015). Yet other models, such as Discourse Grammar (Kaltenböck, Heine and Kuteva 2011) regard such meaning components as belonging to a module of the linguistic system (Thetical Grammar) which is distinct and separate from that dealing with the semantic "proper", i.e. denotational aspects of meaning (Sentence Grammar), but which nonetheless can be shown to be systematically associated with certain linguistic forms and hence deserves to be called a *grammar*.

The linguistic frameworks mentioned so far in this section may be divided (admittedly somewhat roughly) into those that regard meaning and grammar as separate, though of course interconnected entities and those that include some aspects of meaning within their respective notions of grammar. Within the latter group, one may draw further lines depending on exactly how much, and what kind of meaning is taken to be grammatical in nature (especially, semantic/denotational only, or pragmatic/interpersonal as well?). For instance, another theoretical approach that would presumably include interpersonal meaning in its conception of grammar is the Morphopragmatics framework (Dressler and Merlini Barbaresi 1991, 1994; see Mattiello's chapter in this volume). While it is beyond the scope of this chapter to pursue a detailed classification of linguistic frameworks with respect to this criterion, we could not conclude this section without mentioning the constructionist approach, which plays an important role in several of the following chapters.

Typically, scholars working with one or another version of Construction Grammar,¹ or who anyway make reference to the basic principles of this family of approaches, endorse a rather encompassing perspective when it comes to the boundaries of grammar. A key assumption of this perspective is that each specific construction of a language (understood as a conventionalized pairing of form and meaning) is specified for a variety of properties of different types. Namely, these subsets of properties correspond to what in other models would be regarded as

¹ "Construction Grammar" is of course a bit of a misnomer, but it has become rather customary in the literature to use this term to refer to a variety of more specific frameworks. For an overview, see Hoffmann (2017).

different levels of grammatical analysis, such as pragmatics, semantics, syntax, morphology, phonology. It should be noted, in this connection, that the type of pragmatic properties included in this set of specifications is usually more encompassing than in other frameworks. For instance, the fact that a construction belongs to a certain register, or even to a certain variety of a language, would be part of this set in a good many constructionist accounts (see for instance the diasystematic approach adopted in Van Huyssteen, Breed and Pilon's chapter) but not in, say, Discourse Grammar, Functional Discourse Grammar or Generative Grammar.

There are two last aspects of the term *grammar* that must be mentioned here. First, so far we have discussed the notion of grammar as (more or less) equivalent to "language competence" (Saussure's *langue*) and hence opposed to all that is extra-grammatical in the sense of pertaining to linguistic performance and/or the sheer articulation of sounds (*parole*). There is however a more restrictive, but equally important interpretation of the term, namely, one in which "grammar" essentially denotes the procedural, abstract knowledge relevant to linguistic competence (the "know-how" of the language faculty); in this sense, grammar contrasts with *lexicon*, understood as the declarative, propositional component of linguistic competence (the "know-that": see Ullman 2001 for an overview). Simplifying somewhat, for most theories of grammar the former type of knowledge is observable in the form of rules (e.g. mainstream Generative Grammar) or mapping constraints (as in Optimality Theory: Prince and Smolensky 1993), whereas lexical knowledge provides the building blocks with which procedural knowledge operates. In Construction Grammar, on the other hand, linguistic knowledge basically consists of a network of constructions, interconnected with each other by various types of taxonomical relations of inheritance. In this perspective, the procedural/declarative distinction is a matter of degree and not an ontological divide: the types of linguistic competence that other frameworks regard as grammatical or lexical in nature do not stand in a dichotomic opposition to each other but correspond to a continuum of constructions. While all constructions consist of a number of slots and are specified for the same types of properties (see above), they vary as regards their degree of *schematicity*, that is, as to how many slots they include and how many of these are pre-instantiated or are left open, and if the latter, what and how many types of units they can host. The more open slots a construction has, and the more different types of fillers can go into these slots, the more schematic that construction will be (i.e. the more procedural knowledge is involved in using that construction). In this vein, some of the contributions included in this volume more or less explicitly discuss the expressions examined in terms of open or fixed slots: for instance, Culpeper, Van Dorst and Gillings compare the productivity of more abstract and schematic impoliteness-related constructions with that of more fixed, lexical-like ones, whereas both Van Huyssteen, Breed and Pilon for Afrikaans and Van Olmen and

Andersson for English and Polish address the interaction between the impoliteness meaning of the construction as a whole and that of the individual items that may fill the open slots.

Finally, one prominent topic in constructionist research (and the chapters mentioned above are no exception) is the issue of conventionalization. This too is obviously very relevant to the notion of grammar – at a pre-theoretical level (and going back once again to good-old Saussure), because language is by definition conventional; in synchrony, because most definitions of “construction” make overt reference to conventionalized pairings of form and meaning (e.g. Goldberg 2006); and in diachrony, in that much contemporary research on grammaticalization regards the context- and frequency-induced conventionalization of an inference as the mechanism responsible for the emergence of new meanings (König and Traugott 1988; Traugott and König 1991; Heine 2002). Indeed, frequency, productivity and context feature prominently in discussions of conventionalization, see e.g. Terkourafi’s (2005b: 247) definition of linguistic norms as “regularities of co-occurrence between linguistic expressions and their extra-linguistic contexts of use”. As we will see in Section 4, the question of determining to what extent a construction is conventionalized for the impoliteness-related meaning it expresses is a recurrent theme in this volume, and is often explicitly addressed in connection with both frequency and context – for instance in Queisser and Pleyer’s discussion of the conventionalized insulting meaning of ‘such’-constructions in German and English.

4 Challenges and directions

As pointed out in Section 1, the grammar of impoliteness has, of course, not gone completely unexplored in the literature. In our view, however, the existing body of research raises a number of questions that warrant further investigation. There are also areas that it has paid little attention to but that we deem relevant for an in-depth understanding of the topic. The present section will briefly discuss these theoretical, methodological, typological and diachronic issues. They are, as we will see in Section 5, taken up to varying degrees in the contributions to this volume. Our primary aim here, though, is to set an agenda for future research.

4.1 Methodology

A first issue that merits more consideration is how we can establish that particular grammatical structures in a language are specialized for impolite purposes.

If its conventionalization for impoliteness is “a correlate of the (statistical) frequency with which” it “is used in one’s experience of” impolite “contexts” (Terkourafi 2005a: 213, originally about politeness), corpus linguistics looks like a fruitful approach. It can give us a good idea of whether and how often a certain structure is meant and/or taken as offensive, through careful examination of the context (see, for instance, Van Olmen, Andersson and Culpeper 2023: 29–30 for the details of such an analysis). It will not come as a surprise therefore that this method has been adopted quite regularly, perhaps most notably by Culpeper (2011) for the identification of impoliteness formulae in British English (see also, for example, Kleinke and Bös 2015 on German; Lai 2019 on Chinese; Andersson 2022 on Swedish).

The corpus-based approach still has its challenges, however. A rather self-evident but not unimportant one, in light of some of our questions below, is its non-applicability to the many languages for which we have no adequate corpora yet. Determining impoliteness based on corpus data is also susceptible to (unconscious) analytical bias, though this problem can be solved with, say, interrater reliability testing (cf. Landone 2022: 221). More significantly, impoliteness is, all in all, a relatively rare phenomenon in language (Culpeper 2011: 9). As a result, to find sufficient data for grammatical structures of potential interest, one may be obliged to resort to extremely large corpora, like web-crawled ones, or highly specific corpora, like discussions on contentious topics. The former run the risk of being unmanageable and the latter that of being skewed. Heated debates about politics, for example, are unlikely to feature sincere compliments taking the form of *you genius!* and would fail to give us all the necessary information about the structure in question. It thus seems desirable to reflect more on which (combinations of) corpora are most appropriate for the present aims.

Conventionalization as correlative to frequency furthermore prompts the question of how often some structure actually has to fulfill an impolite function to make up an expression dedicated or even just partly dedicated to impoliteness. Culpeper (2011: 134) sets the bar at half of all hits for his formulae. For Dobrushina (2024: 615), by contrast, the exceptional usage of an optative marker for blessings instead of curses in two Turkic languages is enough to write that it is not specialized for curses and that positive versus negative “evaluation is still the job of the communicative context and the lexical meaning of the words”. The discrepancy between these scholars reflects different takes on what counts as conventionalization and, from the perspective of pragmatics (see Terkourafi 2005b: 251 on generalized, i.e. default but still cancellable, implicatures), one could probably make a case for this optative being partially dedicated. The disagreement nonetheless suggests that other or supplementary ways to establish conventionalized impoliteness may be needed.

One possible approach centers around linguistic coercion, a process whereby “the meaning of [a] lexical item conforms to the meaning of the structure in which it is embedded” (Michaelis 2004: 25). It has been employed implicitly as a criterion/measure in various publications (e.g. Jones 1996: 223; Finkbeiner, Meibauer and Wiese 2016: 4; Jain 2022: 389) and Giomi, Van Olmen and Van Oers (2025) propose it as one of their conditions for a grammatical structure to constitute an insulting sentence type: it can solely contain negatively evaluative expressions or, if other types of expressions are tolerated, a negatively evaluative reading is imposed on them. A noun like ‘linguist’ becoming an insult when it occurs in a particular structure (see Section 1 on *espèce de NP!* in French) is a good indication of that structure’s conventionalized impoliteness. Coercion even appears to be helpful in candidates for partial conventionalization, considering how *you linguist!* – despite the acceptability of genuine *you genius!* – would usually be interpreted (without context). For that reason, it is worth exploring, in our view, whether the use of such effects can be extended beyond insults in some way. Coercion has the further advantage that it can be tested fairly easily, for example, by asking speakers to assess the well-formedness and (im)politeness of suitable stimuli in a questionnaire.

This line of experimental research more generally avoids the corpus-based approach’s potential for analytical bias and also has – its own complications notwithstanding (for reasons of space, we refer to Landone 2022: 151–167 for an evaluation) – a long and productive history in the field (e.g. Hill et al. 1986; Nadeu and Prieto 2011; Terkourafi, Weissman and Roy 2020). A more intensive application of such methods to the topic of conventionalized impoliteness in grammatical structures would undoubtedly prove useful, providing us with (quantitative) data on how they are interpreted, perceived and the like by speakers themselves. Yet, not all these methods will be equally effective for the present goals. The popular instruments of discourse completion tasks and production questionnaires (see Landone 2022: 125–139) expect participants to supply language themselves and they have generated interesting insights into conventionalized structures for, among other things, apologies, requests and compliments (see Blum-Kulka, House and Kasper’s 1989 ground-breaking work and its many follow-up studies). For research on impoliteness, however, there are obviously ethical issues around asking people to produce the kind of language required. Participants are also likely to feel uncomfortable doing so and to moderate their answers accordingly.

There is room for more online experimental approaches as well. As Raizen, Vergis and Christianson (2015: 213) correctly point out, methods such as questionnaires can only tell us about forms “after they have been processed”, while the authors’ own eye-tracking study of taboo words reveals that speakers’ assessments of them as impolite may in part happen pre-consciously. Linguistic research into (im)politeness that adopts such experimental approaches is, as a whole, still in its

comparative infancy. It is not difficult to imagine, though, how measuring reaction times, event-related potentials, skin conductance responses or heart rates (e.g. Jiang and Zhou 2015; Ruytenbeek, Allaert and Vanderhasselt 2024; Zlov and Zlatev 2024) could contribute to investigations into the grammar of impoliteness. One could hypothesize, for instance, that relevant structures would trigger heightened arousal, as indicated by psychophysiological responses (cf. Fox et al. 2018), even when they are lexically nonsensical or incomplete or when they are used jokingly.

The preceding paragraphs have discussed a range of methods that may help establish that some grammatical structure is specialized for impolite purposes. They do all require access to a substantial amount of data and/or speakers, which cannot be assumed for most of the world's languages. As such, these methods are ill-suited for any study that wants to examine the grammar of impoliteness from a typological point of view. That type of research will almost inevitably have to draw on the limited resources available for each language – perhaps, the intuitions of a few of its speakers and, more likely, whatever information is included in its description by a field linguist. Reliability is an obvious concern here. However, the difficulties for cross-linguistic research in this area are more fundamental. (Im)politeness and its formal side especially have, understandably so, not been on the minds of most people documenting languages and are therefore only occasionally mentioned in their grammars, if at all. The typologist's initial task is thus simply to find (data on) structures of potential interest. An approach that could be fruitful in this respect is querying parallel corpora with numerous languages (e.g. Open Subtitles) for equivalents of known impolite structures. Another one is to search vast collections of grammars with corpus tools for words that would occur in descriptions of relevant structures (e.g. *threat**) and to check the concordances and then the primary sources.

4.2 Areas of research

There are clearly serious methodological challenges to overcome. At the same time, they offer opportunities for further research into more languages, which would enable us to answer questions that have largely remained open. We know little, for instance, about whether (different types of) grammatical expressions of impoliteness have formal features in common across languages and, if so, which ones and why. Early indications of a positive answer come from a pilot study by Giomi and Van Oers (2022) on structures that are specialized for direct insults. What may be called “insultives” are found to display marking usually associated with possession in a variety of languages. *Din* ‘your’ in Norwegian *din idiot!* ‘you idiot!’ (Julien 2016: 88) can serve as an example and so can the second person singular possessive

suffix *-u* in *Tukang Besi pai'i-'u la!* 'you stupid!' (Donohue 1999: 455) and *se-n-kin* 'its' with genitive *-n* in Finnish *senkin pässi!* 'you oaf!' (Hakulinen et al. 2004: §1726) (see Oda 2019 on Japanese too). Another feature that (partially) conventionalized insult structures appear to share and that can again be illustrated with Norwegian and *Tukang Besi* is the presence of a second person form (see also Corver 2008 on Dutch; Hu and Van Olmen 2024 on Chinese).

More research is needed, though, to determine how widespread these phenomena really are and, of course, to see if other types of expressions (e.g. silencers, ill-wishes) exhibit any cross-linguistic similarities in form. Only then can we properly assess the validity of explanations like Van Olmen, Andersson and Culpeper's (2023: 37) appeal to "pragmatic explicitness" (Culpeper and Haugh 2014: 170) for the occurrence of 'you' in insultives. That is to say: is it true for all languages/cultures at issue that spelling out the target with a second person is a manifestation of directness that is not very compatible with politeness (cf. Brown and Levinson 1987: 131) and instead allows speakers to overtly "associate the other with a negative aspect" (Culpeper 2005: 41)?

A question that is closely related to the one about shared features is whether there are any recurrent grammatical and/or lexical sources for (different types of) expressions of impoliteness and, if so, which ones and why. A (for now) tentative observation in this regard is that the imperative, for one, seems to be the basis for a range of structures in various languages. English *don't you dare* VI, for instance, clearly originates from a negative imperative. However, unlike regular negative imperatives (e.g. *don't (you) worry!*), this expression of a threat can no longer omit the subject. In the same vein, Aikhenvald (2020: 53–55) notes – for languages as diverse as Thai, Russian and Amharic – that curses often take an imperative-like form but do not have all the syntactic characteristics typical of conventionally directive imperatives. The imperative is, as Van Olmen (2018) argues, also a source for structures in a number of European languages dedicated to conveying a reproach, i.e. 'you should have Ved!' (e.g. the Dutch so-called "reproachative" *had gebeld!* lit. 'had called!'; see Mori 2024: 34 on Japanese too). These remarks make it tempting to assume that the imperative's apparent versatility can at least to some extent be accounted for by its ostensible potential, as an imposition on the addressee, for impoliteness. Support for this idea, in a way, comes from Aikhenvald's (2020: 55) assertion that imperatives in Manambu are actually "judged too strong to be used in curses and maledictions". Still, in many languages, the imperative is, in fact, among the more polite strategies to issue a directive (e.g. Kasanga 2006: 70 on Northern Sotho). For that reason, any sweeping statements about its role here are probably somewhat premature.

The same holds for any claims about more lexical sources. Guillaume (2018: 118), for instance, contends that Tacana's depreciative suffixes derive from lexemes

meaning ‘bad’ and ‘be wrong’ but such negatively evaluative items are clearly not the only possible lexical sources for the grammatical expression of impoliteness. The French insultive mentioned in Section 1 features a noun meaning ‘species, type’ (cf. Italian *razza* ‘race, breed’ in *razza di scerno!* ‘you fool!’), its Hebrew equivalent one meaning ‘piece’ (Fishman 2018; cf. English *piece of shit*). A more comprehensive picture of the origins of structures of impoliteness is needed, however. It will enable us to identify potential tendencies and provide us with a stronger cross-linguistic foundation for our attempts at explaining findings.

In this endeavor, there is a vital role for diachronic research too. To our knowledge, little attention has been paid so far to how (different types of) grammatical forms of impoliteness emerge and evolve over time. The Spanish insultive *so NP!* (Giomi and Van Oers 2022) already raises interesting questions, though, about how frequently expressions of impoliteness develop out of ones of politeness or vice versa. Real Academia Española (2023: s.v. *so*, our translation), a reference dictionary of Spanish, writes that *so* serves “to enhance the meaning of the adjective or noun it precedes, generally with a derogatory meaning” (e.g. *so cabrón!* ‘you bastard!’) and traces its etymology back to *señor* ‘sir, mister’. In other words, the structure appears to originate in some kind of politeness strategy and it is not implausible that the former is the result of the ironic/sarcastic usage of the latter. Pragmatic reversal (e.g. Mazzon 2017; Fedriani 2019) may therefore be one of the mechanisms of change that gives rise to grammar of impoliteness. Van Olmen’s (2018: 141–149) account of the Dutch reproachative adds analogy and insubordination (see Evans 2007) to the mix of relevant processes but it still requires checking against actual historical language data. In short, it very much remains to be seen, for instance, which mechanisms are most significant for the development of grammatical expressions of impoliteness (e.g. reanalysis?), whether it involves any typical bridging contexts or how conventionalization really unfolds diachronically.

Another issue that more research in general will be able to shed light on is which types of impoliteness (do not) get conventionalized regularly in the grammar of languages. Preliminary results by Aikhenvald (2020) and Dobrushina (2024) indicate, for instance, that morphologically marked curses are (even more) infrequent cross-linguistically (than morphologically marked blessings). More periphrastic structures deserve to be taken into account too, of course. Similarly, a cautious comparison of Van Olmen’s (2018) findings for reproachatives in European languages with those for insultives in Europe by Giomi and Van Oers (2022) and others referred to above suggests that the latter occur considerably more often than the former, which seem to be limited to just six languages. An attractive explanation for this observation comes from Culpeper, Iganski and Sweiry (2017: 15). They note that, in the Crown Prosecution Service records for England and Wales on religiously aggravated hate crime, insults are by far the most common type of impo-

liteness. The comparative frequency of insultives in the languages of Europe could thus be argued to exemplify Du Bois's (1985: 363) famous dictum that "grammars code best what speakers do most". This claim does presume that Culpeper, Iganski and Sweiry's (2017) findings can be extended to English and (European) language(s) at large, an assumption that merits further scrutiny itself.

A final challenge for the study of the grammar of impoliteness is how to account for the phenomenon and how to capture it in theoretical models of language. This issue has received some attention in the generative paradigm and formal semantics (e.g. Corver 2008; Gutzmann 2019; Jain 2022) but it is typically examined through the wider lens of evaluative or expressive language. We would advocate for an approach that is more focused on impoliteness in particular, also within other frameworks (see, for instance, Giomi, Van Olmen and Van Oers 2025 on insultives as a sentence type from a Functional Discourse Grammar perspective).

5 Contribution of the volume

The present volume addresses the issue that most of the literature so far is made up of separate studies of specific structures in individual European or East Asian languages (see Section 1) in different ways. First, many chapters here explicitly compare impoliteness structures across languages. Mattiello's, for instance, looks at English and Italian, Queisser and Pleyer's at English and German and Van Olmen and Andersson's at English and Polish (Italian and German are also the subject of Paternoster's and Finkbeiner's studies respectively). Second, several contributions explore languages for which impoliteness remains under-researched. Davis and Jang's focuses on Korean while Dobrushina's contrasts a number of Nakh-Daghestanian languages and Van Huyssteen, Breed and Pilon's compares English to Afrikaans. Third, Culpeper, Van Dorst and Gillings's study takes a whole set of impoliteness structures or formulae that were originally identified for British English, though subsequent research has revealed that they are also relevant to a number of other languages, and examines exactly how robust these structures are in British English.

Regarding the question of conventionalized impoliteness from a methodological perspective (see Section 4.1), this volume showcases the potential of a range of different approaches. For example, Davis and Jang report on an online questionnaire testing the relative order and acceptability of the anti-honorific prefix alongside other Korean verbal prefixes in constructed sentences. This method yields insights into how these prefixes are understood in natural language usage and their syntactic domains. The chapters by Queisser and Pleyer on 'you are such

a N' and by Van Olmen and Andersson on 'you NP!' deploy experimental methods too, relying on questionnaire data assessing judgments of well-formedness and (im)politeness. A key focus in these studies is the idea of testing for coercion effects, i.e. whether the structure in question forces an impolite reading onto lexical content that is not inherently negative, as an indication of conventionalization. This idea also emerges in Paternoster's contribution. Whilst investigating the Italian formula *che ti venga* NP! 'may NP come to you!' as a conventionalized linguistic expression of impoliteness, particularly a disease curse, she examines the sarcastic use of seemingly benedictive cases of the structure.

Paternoster's approach is explicitly corpus-based, however, and an excellent example of the necessary in-depth analysis of the co-text of authentic attestations to determine their function. The study also brings metapragmatics into play, looking at how contemporaries conceived of the expression under examination and thereby gaining insight into whether, for example, it was viewed as expressing positive or negative attitude. Corpus-based methods are also central to Mattiello's quantitative and qualitative investigation of the pragmatic functions of English suffixoids like *-ass* and *-head* and their corresponding expressions in Italian, as well as to Van Huyssteen, Breed and Pilon's collexeme analysis for Afrikaans *wat de . . .!* 'what the . . .!'. The latter also addresses possible objections to the use of certain corpus data, by contrasting a corpus of comments removed by moderators, likely to contain offensive language, to a more general corpus of unedited online comments. Culpeper, Van Dorst and Gillings further innovate in the area of corpus studies, combining sophisticated large-scale queries with a meticulous analysis of representative samples of the retrieved data *and* interrater reliability tests to establish a consensus that a particular use of an expression really does have impoliteness effects. Another noteworthy approach found in some chapters is the use of information from dictionaries alongside that from corpus data. In Finkbeiner's contribution, this method is applied in an investigation of a threat structure in German. Mattiello uses it both in her study of English suffixoids and in her examination of their translational equivalents in Italian. Dobrushina, finally, is faced with the problem that, for her Nakh-Daghestanian languages, there simply exist no extensive corpora and solves it by employing dictionaries of Avar, Lak and Rutul as a source for examples of curses (as well as blessings).

The present volume also provides new data relevant for questions about shared features of and sources for grammatical expressions of impoliteness (see Section 4.2). Across various contributions, it highlights the potential for impoliteness within morphology. Davis and Jang, for instance, examine a unique case of impolite verbal morphology in Korean (drawing parallels with Japanese) while Mattiello explores how morphemes and their pragmatic effects are translated between the morphological (and syntactic) systems of English and Italian. The partial con-

ventionalization of optative suffixes for curses is the topic of Dobrushina's contribution, which also notes the unexpected presence and marked position of second person pronouns in negatively oriented wishes in particular. This observation ties in nicely with the structures that are the focus of Queisser and Pleyer's and Van Olmen and Andersson's chapters, i.e. 'you are such a N' and 'you NP!'. Together, they draw attention to explicit second person pronouns as an important feature of conventionalized grammatical impoliteness. The volume also addresses the role of (presumably insubordinate) 'that'-clauses as a stable source for curse structures, as discussed in Paternoster's study, and examines the German interjection *wehe* 'woe' combined with conditional clauses as a foundation for threat structures, as explored in Finkbeiner's contribution. The latter structure is, moreover, argued to involve a second person pronoun in the embedded clause, with overt reference to the addressee – a pattern that is, of course, reminiscent of the curses and the insult structures just mentioned. Finally, Culpeper, Van Dorst and Gillings's corpus-based analysis of British English invites comparable large-scale investigations into the form and function of impoliteness across other languages, for which the authors suggest their method could be replicated.

As to the diachrony of grammatical expressions of impoliteness (see, again, Section 4.2), the volume offers new insights too. For instance, Paternoster's study of Italian *che ti venga* NP! as a disease curse from the 14th to the 20th century highlights its generally stable and conventionalized use over the years but it also notes some possible shifts toward greater conventionalization. The analysis reveals that the structure, which combines verbs in the subjunctive mood with direct address and disease nouns, functions as an expression of impoliteness in both cultural and legal contexts. Furthermore, even some studies whose primary focus is not on tracing their structures' historical trajectory still engage with their development over time. One example is the evolution of English compound constituents into bound morphemes with specialized meanings in Mattiello's chapter. Van Huyssteen, Breed and Pilon's contribution is also mainly synchronic in nature, but they present interesting reflections on Afrikaans *wat de . . . !* as a structural borrowing from English that has been extended considerably in the target language. Likewise, Finkbeiner's chapter on a threat structure in Present-day German offers insight into its development from an interjection with lamentation as its primary meaning and a structure used to predict a calamity. Her findings align rather straightforwardly with approaches to grammaticalization that discuss the conventionalization of conversational implicatures and propose shifts from the representational to the interpersonal domain. Lastly, Queisser and Pleyer touch upon the historical development of 'such/so' as an intensifier in English and German and suggest a grammaticalization-like process as a result of which non-evaluative nouns in the nominal slot of the structures under examination tend to be coerced into an evaluative reading.

Another, final strength of this volume is that the grammar of impoliteness is studied from a variety of theoretical perspectives, showing that research into the phenomenon does not and need not depend on one's theoretical framework. Davis and Jang's chapter, for instance, investigates anti-honorific marking in Korean from a generative point of view, thus challenging the common assumption that impoliteness is primarily a pragmatic/discursive phenomenon rather than a structural one. Mattiello's contribution adopts a morphopragmatic perspective, indicating that impoliteness can be studied within word-formation processes too. An approach that underlies many of the other chapters is the usage-based constructionist one. It is, for example, assumed in Queisser and Pleyer's and Van Olmen and Andersson's experimental investigations into the (partial) conventionalization for impoliteness of the structures that they are interested in. Paternoster's diachronic study aligns with this framework as well, highlighting how recurring pragmatic inferences contribute to the conventionalization of impoliteness within grammatical structures over time. The usage-based constructionist approach is also present in Finkbeiner's chapter, whose findings – as stressed above – can furthermore be related to the same principles of grammaticalization theory that are relevant to Paternoster's work. Van Huyssteen, Breed and Pilon, then, are perhaps the most explicit in their acknowledgment of this theoretical perspective: they situate their study within a diasystematic construction grammar model, which aims to explain emergent bilingual or multilingual phenomena. Finally, Culpeper, Van Dorst and Gillings's contribution challenges the mainstream view that impoliteness is entirely context-driven by demonstrating that impoliteness operates on a spectrum, with some structures showing high conventionalization (e.g. *fuck off*) while others are more contextually dependent (e.g. *get lost*) – and also with the more productive, more abstract formulae (more dependent on grammar as opposed to specific words) attracting slightly lower impoliteness scores. Consequently, the study can be seen as arguing for a middle ground between grammaticalization and pragmatics, further strengthening the volume's focus on theoretical diversity.

Note, to conclude, that the volume is structured in the following way. The first part contains the present introduction. In the third and last part, Marina Terkourafi offers her reflections, as a leading scholar in the field of (im)politeness studies, on the topic of this collection and on the various contributions. The middle part starts off with studies that focus on morphology, i.e. Mattiello on English suffixoids and Davis and Jang on an anti-honorific prefix in Korean. Dobrushina's chapter is next, as it looks at optative suffixes in Nakh-Daghestanian languages but also at some syntactic peculiarities of curses in particular. Part two continues with contributions dealing with periphrastic structures that feature specific lexical content. The first one is Paternoster's study of an Italian disease curse, a topic that is closely related to that of the preceding chapter. The second one is Finkbeiner's investiga-

tion of a threat structure in German. The middle part then moves on to studies that examine individual periphrastic structures with no specific lexical content, i.e. Van Huyssteen, Breed and Pilon on Afrikaans *wat de . . .!*, Queisser and Pleyer on English and German ‘you are such a N!’ and Van Olmen and Andersson on ‘you NP!’. The latter two have the topic of insults in common too. Part two ends with Culpeper, Van Dorst and Gillings’s contribution, which looks at a whole range of different structures in British English and can be said to have the widest focus of all chapters.

Acknowledgments: This volume emanates from a workshop on the grammar of impoliteness held at the 56th Annual Meeting of the Societas Linguistica Europaea in Athens in August 2023. We would like to thank all participants for their presentations and the stimulating discussions, without which the present collection would not be what it is now. We are most grateful to Marina Terkourafi in particular. Her feedback as a discussant was invaluable and she also kindly agreed to share her views on this volume’s topic and contributions in the concluding chapter. The final choice of papers presented here is ours, of course, and includes work by a number of scholars that we contacted after the workshop. The evaluations of the following referees have been a great help in the selection process: Sungdai Cho, Chiara Ghezzi, Andreas Jucker, Manfred Kienpointner, Natalia Knoblock, Katharina Korecky-Kröll, Maarten Mous, Na Song, Vittorio Tantucci, An Van linden and Nathalie Verelst. Thanks are also due to our authors for their internal reviews and for going on this – hopefully, not too impolitely long – journey with us. Finally, we want to acknowledge the support for this project from Chiara Gianollo as the book series’ editor-in-chief, from Barbara Karlson and Birgit Sievert as the publisher and from Lancaster University, whose generous financial contribution has made open access publication possible.

References

- Aikhenvald, Alexandra Y. 2020. “Damn your eyes!” (Not really): Imperative imprecatives, and curses as commands. In Nico Nassenstein & Anne Storch (eds.), *Swearing and Cursing: Contexts and Practices in a Critical Linguistic Perspective*, 53–78. Berlin: De Gruyter.
- 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.
- Arndt, Horst & Richard Wayne Janney. 1987. *Intergrammar: Toward an Integrative Model of Verbal, Prosodic and Kinesic Choices in Speech*. Berlin: Mouton de Gruyter.
- Austin, J. Paddy M. 1990. Politeness revisited – the dark side. In Alan Bell & Janet Holmes (eds.) *New Zealand Ways of Speaking English*, 277–293. Clevedon & Philadelphia: Multilingual Matters.

- Blum-Kulka, Shoshana, Juliane House & Gabriele Kasper. 1989. *Cross-cultural Pragmatics: Requests and Apologies*. Norwood: Ablex.
- Bousfield, Derek. 2008. *Impoliteness in Interaction*. Amsterdam: Benjamins.
- Brown, Penelope & Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Corver, Norbert. 2008. Uniformity and diversity in the syntax of evaluative vocatives. *Journal of Comparative Germanic Linguistics* 11. 43–93.
- Crystal, David. 2008. *A Dictionary of Linguistics and Phonetics*. Oxford: Blackwell.
- Culpeper, Jonathan. 1996. Towards an anatomy of impoliteness. *Journal of Pragmatics* 25(3). 349–367.
- Culpeper, Jonathan. 2005. Impoliteness and entertainment in the television quiz show *The Weakest Link*. *Journal of Politeness Research* 1(1). 35–72.
- Culpeper, Jonathan. 2011. *Impoliteness: Using Language to Cause Offence*. Cambridge: Cambridge University Press.
- Culpeper, Jonathan & Michael Haugh. 2014. *Pragmatics and the English Language*. Basingstoke: Palgrave.
- Culpeper, Jonathan & Michael Haugh. 2021. The metalinguistics of offence in (British) English: A corpus-based metapragmatic approach. *Journal of Language, Aggression and Conflict* 9(2). 185–214.
- Culpeper, Jonathan, Paul Iganski & Abe Sweiry. 2017. Linguistic impoliteness and religiously aggravated hate crime in England and Wales. *Journal of Language Aggression and Conflict* 5(1). 1–29.
- Culpeper, Jonathan, Gila Shauer, Leyla Marti, Meilian Mei & Minna Nevala. 2014. Impoliteness and emotions in a cross-cultural perspective. *SPELL: Swiss Papers in English Language and Literature* 30. 67–88.
- Culpeper, Jonathan & Vittorio Tantucci. 2021. The principle of (im)politeness reciprocity. *Journal of Pragmatics* 175. 146–164.
- Dobrushina, Nina. 2024. Evaluation between grammar and context: The case of blessings and curses. *Diachronica* 41(5). 605–634.
- Donohue, Mark. 1999. *A Grammar of Tukang Besi*. Berlin: Mouton de Gruyter.
- Dressler, Wolfgang U. & Lavinia Merlini Barbaresi. 1991. Elements of morphopragmatics. In Jef Verschuren (ed.), *Levels of Linguistic Adaptation*, 33–51. Amsterdam: Benjamins.
- Dressler, Wolfgang U. & Lavinia Merlini Barbaresi. 1994. *Morphopragmatics: Diminutives and Intensifiers in Italian, German, and Other Languages*. Berlin: Mouton de Gruyter.
- Du Bois, John W. 1985. Competing motivations. In John Haiman (ed.), *Iconicity in Syntax*, 343–365. Amsterdam: Benjamins.
- Evans, Nicholas. 2007. Insubordination and its uses. In Irina Nikolaeva (ed.), *Finiteness: Theoretical and Empirical Foundations*, 366–431. Oxford: Oxford University Press.
- Fedriani, Chiara. 2019. A pragmatic reversal: Italian *per favore* ‘please’ and its variants between politeness and impoliteness. *Journal of Pragmatics* 142. 233–244.
- 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: Benjamins.
- Fishman, Alon. 2018. Presupposing expressive meaning: The case of Hebrew “xatixat”. Paper presented at the 40th Annual Conference of the German Linguistics Society, Stuttgart, 7–9 August, 2018.

- Fox, Alexa K., George D. Deitz, Marla B. Royne & Joseph D. Fox. 2018. The face of contagion: Consumer response to service failure depiction in online reviews. *European Journal of Marketing* 52(1/2). 39–65.
- Garcés-Conejos Blitvich, Pilar. 2010. A genre approach to the study of im-politeness. *International Review of Pragmatics* 2(1). 46–94.
- Garcés-Conejos Blitvich, Pilar. 2013. Introduction: Face, identity and im/politeness. Looking backward, moving forward: From Goffman to practice theory. *Journal of Politeness Research*. 9(1). 1–33.
- Garfinkel, Harold. 1964. Studies of the routine grounds of everyday activities. *Social Problems* 11(3). 225–250.
- Giomi, Riccardo and Denise van Oers. 2022. Insultive constructions: A crosslinguistic perspective. Paper presented at 55th Annual Meeting of the Societas Linguistica Europaea, Bucharest, 24–27 August, 2022.
- Giomi, Riccardo, Daniel Van Olmen & Denise van Oers. 2025. Insults as a sentence type: A cross-linguistic perspective. Paper presented at the 10th International Workshop on Functional Discourse Grammar, Schoorl, 2–5 July, 2025.
- Giomi, Riccardo. 2023. Review of Knoblock (ed.) (2022). *Journal of Linguistics* 59(4). 919–923.
- Goffman, Erving. 1967. *Interactional Ritual: Essays on Face-to-Face Behavior*. Garden City: Anchor Books.
- Goldberg, Adele E. 2006. *Constructions at Work*. Oxford: Oxford University Press.
- Grice, H. Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and Semantics III: Speech Acts*, 183–198. New York: Academic Press.
- Guillaume, Antoine. 2018. The grammatical expression of emotions in Tacana and other Takanan languages. *Studies in Language* 42(1). 114–145.
- Gutzmann, Daniel. 2015. *Use-Conditional Meaning: Studies in Multidimensional Semantics*. Oxford: Oxford University Press.
- Gutzmann, Daniel. 2019. *The Grammar of Expressivity*. Oxford: Oxford University Press.
- Hakulinen, Auli, Maria Viikuna, Riitta Korhonen, Vesa Koivisto, Tarja-Riitta Heinonen & Irja Alho. 2004. *Iso Suomen Kielioppi*. Helsinki: Finnish Literature Society.
- Halliday, M.A.K. & Christian M.I.M. Matthiessen. 2014. *Halliday's Introduction to Functional Grammar*. Abingdon: Routledge.
- Haugh, Michael & Yasuhisa Watanabe. 2017. (Im)politeness theory. In Bernadette Vine (ed.) *The Routledge Handbook of Language in the Workplace*, 65–76. New York: Routledge.
- Heine, Bernd. 2002. On the role of context in grammaticalization. In Ilse Wischer & Gabriele Diewald (eds.), *New Reflections on Grammaticalization*, 83–101. Amsterdam: Benjamins.
- Hengeveld, Kees & J. Lachlan Mackenzie. 2008. *Functional Discourse Grammar: A Typologically-based Theory of Language Structure*. Oxford: Oxford University Press.
- Hill, Beverly, Sachiko Ide, Shoko Ikuta, Akiko Kawasaki & Tsunao Ogino. 1986. Universals of linguistic politeness: Quantitative evidence from Japanese and American English. *Journal of Pragmatics* 10(3). 347–371.
- Hoffmann, Thomas. 2017. Construction Grammars. In Barbara Dancygier (ed.), *The Cambridge Handbook of Cognitive Linguistics*, 310–329. Cambridge: Cambridge University Press.
- Hornstein, Norbert. 1995. *Logical Form: From GB to Minimalism*. Cambridge, MA: Blackwell.
- Hu, Yue & Daniel Van Olmen. 2024. A corpus study of Mandarin Chinese impolite constructions: The case of *ni zhe(ge)/ge shazhi* 'you idiot!'. Paper presented at the 13th International Conference on Construction Grammar, Gothenburg, 26–28 August, 2024.

- Hudson, Mutsuko Endo. 2018. *Ne* as an “impoliteness” (“detachment”) marker? In Mutsuko Endo Hudson, Yoshiko Matsumoto & Junko Mori (eds.), *Pragmatics of Japanese: Perspectives on grammar, interaction and culture*, 197–216. Amsterdam: Benjamins.
- Jackendoff, Ray & Jenny Audring. 2019. The parallel architecture. In András Kertész, Edith Moravcsik & Csilla Rákosi (eds.), *Current Approaches to Syntax: A Comparative Handbook*, 215–239. Berlin: De Gruyter Mouton.
- Jackendoff, Ray & Jenny Audring. 2020. *The Texture of the Lexicon: Relational Morphology and the Parallel Architecture*. Oxford: Oxford University Press.
- Jain, Kate H. 2022. You Hoboken! Semantics of an expressive label maker. *Linguistics and Philosophy* 45(2). 365–391.
- Jiang, Xiaoming & Xiaolin Zhou. 2015. Impoliteness electrified: ERPs reveal the real time processing of disrespectful reference in Mandarin utterance comprehension. In Marina Terkourafi & Staci Defibaugh (eds.), *Interdisciplinary Perspectives on Im/politeness*, 239–266. Amsterdam: Benjamins.
- Jones, Michael A. 1996. *Foundations of French Syntax*. Cambridge: Cambridge University Press.
- Julien, Marit. 2016. Possessive predicational vocatives in Scandinavian. *Journal of Comparative Germanic Linguistics* 19. 75–108.
- Kaltenböck, Gunther, Bernd Heine & Tania Kuteva. 2011. On thetical grammar. *Studies in Language* 35(4). 848–893.
- Kasanga, Luanga A. 2006. Requests in a South African variety of English. *World Englishes* 25(1). 65–89.
- Kienpointner, Manfred. 1997. Varieties of rudeness: Types and functions of impolite utterances. *Functions of Language* 4(2). 251–287.
- Kienpointner, Manfred. 2008. Impoliteness and emotional arguments. *Journal of Politeness Research* 4(2). 243–265.
- Klinke, Sonja & Birte Bös. 2015. Intergroup rudeness and the metapragmatics of its negotiation in online discussion fora. *Pragmatics* 25(1). 47–71.
- Knoblock, Natalia (ed.). 2022. *The Grammar of Hate: Morphosyntactic Features of Hateful, Aggressive and Dehumanizing Discourse*. Cambridge: Cambridge University Press.
- König, Ekkehard & Elizabeth C. Traugott. 1988. Pragmatic strengthening and semantic change: The conventionalizing of conversational implicature. In Werner Hülsen & Rainer Schulze (eds.), *Understanding the Lexicon: Meaning, Sense and World Knowledge in Lexical Semantics*, 110–124. Tübingen: Max Niemeyer.
- Lachenicht, Lance G. 1980. Aggravating language: A study of abusive and insulting language. *International Journal of Human Communication* 13(4). 607–688.
- Lai, Xiaoyu. 2019. Impoliteness in English and Chinese online diners’ reviews. *Journal of Politeness Research* 15(2). 293–322.
- Landone, Elena. 2022. *Methodology in Politeness Research*. Cham: Springer.
- Locher, Miriam A. 2004. *Power and Politeness in Action: Disagreements in Oral Communication*. Berlin: Mouton de Gruyter.
- Locher, Miriam A. 2006. Polite behavior within relational work: The discursive approach to politeness. *Multilingua* 25(3). 249–267.
- Locher, Miriam A. & Richard J. Watts. 2005. Politeness theory and relational work. *Journal of Politeness Research* 1(1). 9–33.
- Mattiello, Elisa. 2022. Language aggression in English slang: The case of the -o suffix. In Natalia Knoblock (ed.), *The Grammar of Hate: Morphosyntactic Features of Hateful, Aggressive and Dehumanizing Discourse*, 34–58. Cambridge: Cambridge University Press.

- Mazzon, Gabriella. 2017. Paths of development of English DMs: (Inter)subjectification, deontic reversal and other stories. In Chiara Fedriani & Andrea Sansó (eds.), *Pragmatic Markers, Discourse Markers and Modal Particles: New Perspectives*, 289–304. Amsterdam: Benjamins.
- Mel'čuk, Igor & Jasmina Miličević. 2011. The “*Budalo jedna!*”-type construction in Contemporary Serbian. *Journal of Slavic Linguistics* 19(1). 85–118.
- Michaelis, Laura A. 2004. Type shifting in construction grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics* 15(1). 1–67.
- Mills, Sara. 2003. *Gender and Politeness*. Cambridge: Cambridge University Press.
- Mori, Hideki. 2024. Characteristics of the *X ie* imperative expression: Three criteria for the classification of imperatives. *Journal of Japanese Linguistics* 40(1). 31–57.
- Nadeu, Marianne & Pilar Prieto. 2011. Pitch range, gestural information, and perceived politeness in Catalan. *Journal of Pragmatics* 43(3). 841–854.
- Oda, Kenji. 2019. Towards the non-predicate modification analysis of the expressive small clause in Japanese. *Proceedings of the Pacific Asia Conference on Language, Information and Computation* 33. 196–202.
- Parvaresh, Vahid & Tahmineh Tayebi. 2018. Impoliteness, aggression and the moral order. *Journal of Pragmatics*. 132. 91–107.
- Prince, Alan & Paul Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. New Brunswick: Rutgers University Center for Cognitive Science.
- Raizen, Adina, Nikos Vergis & Kiel Christianson. 2015. Using eye-tracking to examine the reading of texts containing taboo words. In Marina Terkourafi & Staci Defibaugh (eds.), *Interdisciplinary Perspectives on Im/politeness*, 213–238. Amsterdam: Benjamins.
- Real Academia Española. 2023. *Diccionario de la lengua española*. Madrid: Real Academia Española.
- Recanati, François. 2004. *Literal Meaning*. Cambridge: Cambridge University Press.
- Ruytenbeek, Nicolas, Jens Allaert & Marie-Anne Vanderhasselt. Psychophysiological effects of evaluative language use on Twitter complaints and compliments. *Internet Pragmatics* 7(2). 193–218.
- Schneider, Klaus .P. 2012. Appropriate behaviour across varieties of English. *Journal of Pragmatics* 44(9). 1022–1037.
- Spencer-Oatey, Helen D.M. 2005. (Im)politeness, face and perceptions of rapport: Unpackaging their bases and interrelationships. *Journal of Politeness Research* 1(1). 95–119.
- Spencer-Oatey, Helen D.M. 2008. *Culturally Speaking: Managing Rapport through Talk across Cultures*. London & New York: Continuum.
- Spencer-Oatey, Helen D.M. & Dániel Z. Kádár. 2021. *Intercultural Politeness: Managing Relations across Cultures*. Cambridge: Cambridge University Press.
- 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.
- Terkourafi, Marina, Benjamin Weissman & Joseph Roy. 2020. Different scalar terms are affected by face differently. *International Review of Pragmatics* 12. 1–43.
- Traugott, Elizabeth C. & Ekkehard König. 1991. The semantics-pragmatics of grammaticalization revisited. In Elizabeth C. Traugott & Bernd Heine (eds.), *Approaches to Grammaticalization I: Focus on Theoretical and Methodological Issues*, 189–218. Amsterdam: Benjamins.

- Ullman, Michael T. 2001. The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research* 30(1). 37–69.
- 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 Olmen, Daniel & Delphine Grass. 2023. *Espèce de linguiste!* – An impoliteness construction in French? Paper presented at the 18th International Pragmatics Conference, Brussels, 9–14 July, 2023.
- 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.
- Van Olmen, Daniel. 2018. Reproachatives and imperatives. *Linguistics* 56(1). 115–162.
- Watts, Richard J. 2003. *Politeness*. Cambridge: Cambridge University Press.
- Watts, Richard J. 2005. Linguistic politeness research: Quo vadis? In Richard J. Watts, Sachiko Ide & Konrad Ehlich (eds.), *Politeness in Language: Studies in its History, Theory and Practice*, xi–xlvii. Berlin: De Gruyter.
- Xie, Chaoqun (ed.). 2020. *(Im)politeness and Moral Order in Online Interactions*. Amsterdam: Benjamins.
- Zlov, Vladislav & Jordan Zlatev. 2024. A cognitive-semiotic approach to impoliteness: Effects of conventionality and semiotic system on judgements of impoliteness by Russian and Swedish speakers. *Journal of Politeness Research* 20(2). 249–296.

From morphemes to periphrastic structures

Elisa Mattiello

2 Impolite suffixoids in English slang

Abstract: This study conducts a corpus-assisted analysis to investigate the role of some slang morphological means in contexts of impoliteness and offensiveness. In particular, the study adopts a morphopragmatic approach to explore the pragmatic functions/effects associated with the slang suffixoids *-ass* (e.g. *fat ass*), *-brain* (e.g. *birdbrain*), *-face* (e.g. *shitface*), and *-head* (e.g. *airhead*) used in verbal aggression. The productivity of *-head* and similar elements positions them on the borderline between compounding and derivation, as part of transitional morphology, i.e. transitional between sub-components of word-formation. The combination of quantitative and qualitative analyses of data drawn from the *Corpus of Contemporary American English* (COCA, Davies 2008) demonstrates the frequency of the morphological processes, their privileged co-texts, and their effects in context. The contrastive English-Italian analysis using *Open Parallel Corpus – English* (OPUS2) illustrates how the two languages express impoliteness and offensiveness through different morphological and syntactic means.

Keywords: impoliteness, English slang, morphopragmatics, affixoid, transitional morphology

1 Introduction

While lexical and discourse strategies of impoliteness and offensiveness have been widely studied thus far (Culpeper 2011), there is limited research devoted to the contribution of grammatical and morphological aspects to impolite language (see Van Olmen et al., this volume; see also Davis and Jang, this volume, on impolite morphology in Korean). A language variety which provides ground for an investigation of impoliteness is English slang.

Slang has for long been associated with informal, vulgar, or bad language (Dumas and Lighter 1978; Andersson and Trudgill 1990; Allen 1998; Stenström 2000) and defined as a special vocabulary used by any set of persons of a low or disreputable character (OED). According to Mattiello (2008: 31), the word ‘slang’ may be defined with at least two senses: “First, slang is the restricted speech of marginal

Elisa Mattiello, Department of Philology, Literature and Linguistics, University of Pisa, Via Santa Maria 67, 56126, Pisa (Italy), elisa.mattiello@unipi.it

or distinct subgroups in society and, second, it is a quite temporary, unconventional vocabulary characterized primarily by connotations of informality and novelty". While, on the one hand, slang can be viewed as a means of sociability and in-group aggregation, especially among teenagers, college students, and young adults (Eble 1996; Munro 2001; Stenström, Andersen and Hasund 2002), it can, on the other hand, act as a means of social exclusion and even of impoliteness and offensiveness.

This paper provides a corpus-assisted analysis of morphological means used with offensive effects in slang, including phenomena such as insults, criticisms, contempt, and verbal aggression. The focus is on four suffixoids (*-ass*, *-brain*, *-face*, *-head*), which are often used in English slang to form complex words referring to "specific groups, such as homosexuals, fools, or ineffectual people" (Mattiello 2023: 101).

The study builds on previous research (Mattiello 2023) that examined the compound families *X-ass*, *X-brain*, *X-face*, and *X-head* in the context of verbal aggression. This research focuses on how the elements *-ass*, *-brain*, *-face*, and *-head* evolve from compound constituents to bound morphemes (suffixoids). In particular, the paper aims to demonstrate that, because of their frequency and productivity in the formation of forms of impoliteness, *-head* and similar elements have developed from compound constituents to affixoids, i.e. bound morphemes originated from free morphemes. Affixoids are here described as transitional elements between affixes and compound constituents: like affixes, they have a bound status (cf. *writ(e)* + *-er* and *air/bone/cabbage/egg/meat* + *-head*), yet, like compound constituents, they correspond to autonomous words (cf. *-head* vs. *head*; see 'transitional phenomena' in Mattiello 2022a). However, affixoids have acquired a more distinct meaning from that of their parent morphemes: e.g., in *Green's Dictionary of Slang* (Green 2023), the suffixoid *-head* is said to create terms 'with a negative personal meaning', often referring to 'foolish people' or 'idiots'. In particular, 1) suffixoids have acquired a figurative, namely metonymic and metaphorical meaning (e.g., in *cabbagehead*, *-head* stands for 'a person's brain' made of/full of what is specified by the first element), and 2) they are used in English slang with a specific pragmatic function, i.e. to express rudeness or even offend others.

The paper adopts a morphopragmatic approach (Dressler and Merlini Barbaresi 1994) to investigate the pragmatic meanings obtained by morphological elements. It explores the pragmatic functions/effects associated with slang words in situations of impoliteness: e.g., *birdbrain* (*Ha, even a birdbrain can get that one*), *airhead* (*You do realise she is a total airhead*), or *fat ass* (*Get in the goddamn car, you fat ass*)¹ display a derisive, critical, and offensive function in their respective

¹ All contextualized examples in English are drawn from the *Corpus of Contemporary American English* (COCA, Davies 2008).

contexts. Derision, criticism, and offense will be differentiated as follows: derision involves ridicule and mockery, criticism entails condemnation and disapproval, and offense reflects attack and aggression.

The combination of quantitative and qualitative analyses of data drawn from the *Corpus of Contemporary American English* (COCA, Davies 2008) will show the frequency of the morphological processes, their privileged co-texts and collocates, as well as their negative potential and face-threatening power. A contrastive English-Italian analysis in the *Open Parallel Corpus – English* (OPUS2), available on Sketch Engine (Kilgarriff et al. 2004), will show how the two languages differ in terms of morphological richness and how they exploit their different morphological systems to express impoliteness. Since English and Italian belong to different language families (Germanic vs. Romance), differences in the use of morphological or syntactic means are expected in the two languages (see also Paternoster, this volume, on a curse construction in Italian).

The paper is organized as follows. After this introductory section, Section 2 introduces the background to the study, with special focus on the theoretical frameworks of morphopragmatics and transitional morphology, and on studies on impoliteness. Section 3 clarifies the distinction between affixoid, combining form, and compound constituent, either strict or loose compound. Section 4 explains the methodology used for data selection and presents the morphological means under investigation. The analysis conducted in Sections 5.1–5.2 is both quantitative and qualitative and aims to investigate the different distributions of slang morphological means and their pragmatic features and contexts of occurrence. Finally, Section 5.3 provides a contrastive analysis of the different morphological means that the two languages use to render comparable concepts and produce similar effects. A general discussion and some conclusive remarks are made in the final section.

2 Theoretical background

The Theory of Morphopragmatics (TMP) adopted in this study was pioneered by Wolfgang U. Dressler and Lavinia Merlini Barbaresi in several studies (Dressler and Merlini Barbaresi 1986, 1987), further expanded as a model in Dressler and Merlini Barbaresi (1994).

According to Merlini Barbaresi (2006: 332), morphopragmatics is a subdiscipline integrating morphology and pragmatics which can be defined as “the set of general pragmatic meanings/effects obtained by morphological rules”. It focusses on various phenomena, from Italian and German evaluative affixes (Dressler and Merlini Barbaresi 1994) to German and Hungarian excessives (Dressler and Kiefer

1990), from Italian ethnic adjectives (Crocco Galeas 1992) to French and English suffixes (the French *-o* suffix in Dressler and Kilani-Schoch 1994; the English *-y/ie* suffix in Merlini Barbaresi 2001; the English *-o* suffix in Mattiello 2022b), and even to Spanish phenomena (Cantero 2003) and, recently, German expressive compounds (Korecky-Kröll and Dressler 2022).

TMP relies on the tenet that morphological rules do not only convey semantic meanings, but also possess some autonomous pragmatic features which are not derivable from semantics. For instance, Italian diminutives semantically express the meaning [small] or [young], but they also possess a pragmatic feature [nonserious] which is not derived from the denotational meaning. Thus, in the Italian utterance *Mangi come un maial-ino!* (lit. you eat like a pig-DIM! ‘you eat like a piggy!’), the suffix *-ino* is used to downgrade the illocutionary force of the evaluative assertion via an added playful character (Bazzanella, Caffi and Sbisà 1991). In the diminutive formation *maial-ino* ‘pig-DIM’, the pragmatic feature [nonserious/playful] can constrain the type of speech situation and reduce the strength of the illocutionary force by decreasing the speaker’s responsibility in uttering the sentence, thus obtaining mitigation of the critical remark (Caffi 2001).

Morphopragmatic effects created via morphological means range from positivity (e.g. pleasantness) to negativity (e.g. meanness), and often express a variety of emotions: from endearment, sympathy, or empathy to irony, sarcasm, or aggressiveness. As Hamilton (2012: 6) remarks, verbal aggression is “the act of using aggressive language on a target” in order to insult, tease, or provoke them.

Impoliteness and hate speech have recently attracted the attention of many scholars addressing the issue of hostile or antagonistic communication, especially in digital environments (Erjavec and Kovačič 2012; Culpeper, Iganski and Sweiry 2017; Miro-Llinares, Moneva and Esteve 2018) and social media (Oksanen et al. 2014; Knoblock 2017; Zhang, Robinson and Tepper 2018; Demata 2021). However, until now, attention has primarily been focused on lexical aspects and discursive strategies, with Knoblock (2022) being an exception, as her edited volume focuses on the grammar of hate, specifically on the morphological and syntactic means used to express aggressiveness.

In this study, the focus is on morphological aspects that contribute to impoliteness in English slang. The study is based on the assumption that English slang has developed some specific morphological means that are generally perceived as impolite, offensive, and face-threatening, both to the speaker’s and to the hearer’s face. Culpeper’s (2005: 38) definition of impoliteness accommodates both the speaker’s and the hearer’s perceptions of intentionality: “Impoliteness comes about when: (1) the speaker communicates face-attack intentionally, or (2) the hearer perceives behaviour as intentionally face-attacking, or a combination of (1) and (2)”. Here I use impolite and offensive discourse as a cover term for Culpeper’s (2011)

face-aggravation and face attack against an individual or group of individuals, deliberately involving aggressive and offensive content.

The study builds on previous research (Mattiello 2023), where the author had already explored the emergence of the compound families *X-ass*, *X-brain*, *X-face*, and *X-head* in the context of verbal aggression. The present study, however, emphasizes the transition of the elements *-ass*, *-brain*, *-face*, and *-head* from compound constituents frequently found in word families to bound morphemes (suffixoids). These suffixoids, like compound constituents, correspond to autonomous words, but acquire a more specific meaning or function when used productively across various examples/types. The productivity of these suffixoids will be verified through corpus-based quantitative analyses, showing their type/token frequencies (both raw and normalized). Their specific meanings will be investigated through collocational analysis, while their functions will be explored through a morphopragmatic analysis.

The following research questions will be addressed in this paper:

1. What is the morphological productivity of the four slang suffixoids under investigation?
2. What specific connotative meanings do these suffixoids convey, based on a comprehensive collocation analysis?
3. Does Italian use the same morphological means to convey impoliteness or offensiveness, and do these serve the same functions as English slang elements?

A corpus-assisted analysis will help clarify the productivity of the four slang suffixoids and their connotative meanings and functions. A systematic comparison with Italian, using parallel corpora, will reveal the different morphological means employed in a Romance language to convey the specific connotations of English slang suffixoids in impolite contexts. The following section outlines the distinction between affixoid, combining form, and compound constituent.

3 Affixoids, combining forms, and compound constituents

Before we move on to the data and methods, a preliminary distinction between ‘affixoid’, ‘combining form’, and ‘compound constituent’ is in order. The term ‘affixoid’ is used within Construction Morphology by Booij and Hüning (2014), who note how the element *free* has developed the more general meaning ‘without what is denoted by the base word’ when it is used as the right constituent of compounds,

as in *sugar-free* and *fat-free*. Moreover, unlike the suffix *-less* having the same denotative meaning in *sugar-less* and *fat-less*, the affixoid *-free* also presupposes a slightly negative evaluation (e.g., that the presence of sugar is not good for one's health). Therefore, affixoids exist alongside formally identical and usually free 'parent' morphs; yet, they acquire a more generalized meaning (Stevens 2005: 73) or 'desemanticization' (Amiot and Dugas 2020), as *-free* 'clear of something which or someone who is regarded as problematic' in *trouble-free*, *pollution-free*, *risk-free*, etc. Moreover, Amiot and Dugas (2020) specify that affixoids are "productive", i.e., are part of schematic constructions with high type frequency (e.g., [N-*free*]), and are bound to their bases (e.g., in *sugar-free*, the suffixoid and the base cannot be split up as **sugar very free*).

A comparable example is reported by Jespersen (1942: 457–458), who lists *-proof* among suffixes (NB: without using the label 'suffixoid'). Originating from the homonymous adjective, *-proof* has become more specifically used to form words meaning 'impenetrable by the thing denoted by the first component' (e.g. *rain-proof*, *sound-proof*, *water-proof*). These examples reflect the idea that affixoids have acquired a more generalized or specialized meaning than their lexemic counterparts, and that they can be used productively with this meaning. A very recent comparable example is *-exit* (in *Brexit*, after *Grexit*), which is becoming productively used with the meaning 'withdrawal (of a country) from the European Union', as in *Spexit*, *Frexit*, *Italexit*, etc. According to the above analysis, *-exit* can be regarded as a suffixoid because it is more generalized than the formally identical free morph (cf. Mattiello 2019).

A different approach is found in Meibauer (2013), where he examines the German prefixoids *ratten-*, *sau-*, *Hammer-*, and *Arsch-* in complex words such as *rattenscharf* ('rat sharp'), *sauschlecht* ('sow bad'), *Hammerauftritt* ('hammer performance'), and *Arschgesicht* ('arse face') as evaluative elements in adjectival and nominal compounds. Meibauer (2013) argues that these elements are not semi-prefixes because they remain lexemes, having only undergone metaphorical extension. Since their meanings are systematically connected with expressive meaning, I argue that they are German prefixoids used to add negative connotations to the base words.

In my model (Mattiello 2022a: 34), I classify affixoids as an instance of transitional morphology. Transitional morphology is that part of morphology that lies at the boundaries of morphological grammar or straddles the demarcation line between two (sub)components. It includes (a) phenomena that are transitional between morphology and other components (i.e. phonology, syntax, lexicon, semantics, pragmatics), (b) phenomena situated at the interface between different subcomponents of morphology (i.e. inflection, derivation, compounding), and (c) borderline phenomena that are non-prototypical representatives of one of the

above subcomponents. Therefore, if we focus on the internal boundaries of morphology, affixoids do not constitute a prototypical category of either compounding or derivation, but may be considered transitional between these two morphological subcomponents. Like compound constituents, affixoids correspond to autonomous words; yet, like affixes, they have a bound status, that is, they attach to words (e.g. *sugar+free*), and sometimes also to word parts (*Br(itish)+exit*). In this latter aspect, affixoids can be compared to combining forms.

Like affixoids, combining forms are either initial or final bound morphemes. However, unlike pre- and suffixoids, combining forms are either allomorphic variants of classical Latin or Greek words (e.g., *bio-* and *-logy* in *biology*), or shortenings of (native or non-native) English words (e.g., *e-* from *electronic* in *e-book*), often with the intervention of a secretion process (e.g., *-aholic* ‘person addicted to’ in *workaholic*) (Warren 1990).² Sánchez Fajardo (2022) has widely studied the pejorative connotation of *-head* (e.g. *egghead*) and *-pants* (e.g. *smartypants*), which he considers to be native combining forms. In particular, he considers them “pejorative formatives on account of their metonymic and synecdochic etymology” (Sánchez Fajardo 2022: 165), as they actually refer to human beings through some of their body parts by adding a depreciative connotation. However, while *bio-*, *e-*, *-aholic*, and *-logy* are bound morphemes which do not correspond to formally identical parent morphs, affixoids do correspond to free morphemes (cf. freestanding *head* vs. bound *-head*). Hence, in my approach, *-aholic* is a combining form while *-head* is an affixoid.

Furthermore, affixoids also differ from compound constituents. According to Bauer et al. (2013: 441), “[t]he most important criterion in distinguishing a compound element from a suffix is its relatedness to a free form”. Thus, if the bound form such as *-man* consistently differs in meaning from the free form *man*, one should assume the existence of an affix. This is also relevant to suffixoids. Indeed, the semantic differentiation between the suffixoid *-head* in *airhead* and the compound constituent *head* in *spearhead* suggests that the two elements should be treated as distinct.

Compound constituents are bases which constitute compounds and may occupy either the right or the left position, as in *birdcage*, which is a ‘cage’ (head) for ‘birds’ (modifier). As Bauer (2017: 1) observes, compounds “are made up of two elements each of which is used elsewhere in the language as a word in its own right”. However, unlike affixoids which correspond to free morphemes but

² According to Warren (1990: 119), “secretion” is “a process in which certain semantic elements in a linguistic unit are kept and others discarded”. Thus, for instance, in *workaholic* ‘a person addicted to working’, the semantic element ‘alcohol’ which is present in the meaning of *alcoholic* is discarded.

undergo desemanticization, compound constituents do not undergo reinterpretation such as generalization or specification, but acquire meaning depending on the base with which they are combined. Thus, *birdcage* is ‘a cage in which birds are kept’, *birdbath* is ‘a bowl filled with water for birds to drink and bathe in’, *seabird* is ‘a bird that lives near the sea and gets its food from it’, and a *songbird* is ‘any of many different types of bird that make musical sounds’. In the former two compounds *bird* acts as modifier, whereas in the latter two compounds it acts as head, the relationship between the compound constituents changes according to the other constituent combined with *bird*. By contrast, in *birdbrain* ‘a fool’ and *dog-ass* ‘an objectionable, unpleasant person’, the bases *bird* and *dog* lose their original meaning because they are combined with two suffixoids (*-brain* and *-ass*) whose pragmatic meaning/function is to convey impoliteness.

Another crucial distinction to be introduced here, especially because it is related to the Italian renditions of suffixoids, is that between strict and loose compounds. According to Scalise (1992: 181), “the presence of a strong boundary is indication of a loose compound while a weak boundary indicates a strict compound”. In line with this claim, in Italian, *sala stampa* ‘press room’ could be viewed as a loose compound because the two words *sala* and *stampa* are closely knit in form and meaning, as commonly happens in compounds, but still carry two-word accents and are spelt as two separate words, rather than being solid. By contrast, *quintessenza* ‘quintessence’ is a strict compound because of the weak boundary between *quinta* ‘fifth’ and *essenza* ‘essence’. A comparable distinction in English would be between the loose compound *head office*, with two-word accents, and the strict compound *spearhead*, stressed on the first constituent. Actually, in Italian we also find syntactic constructions characterized by prepositions. For instance, the [N+Adj]_N compound *testa vuota* ‘lit. head empty’ for ‘airhead’ is a strict compound, i.e. a morphological construction, whereas the [N+Prep+N]_N formation *testa di cavolo* ‘lit. head of cabbage’ for ‘pinhead’ is a phrase closer to Scalise’s (1992) loose compound, i.e. a syntactic construction.

For the latter construction, various labels have been used: namely, *composto sintagmatico* ‘syntagmatic compound’ (Scalise 1994), *lessema complesso* ‘complex lexeme’ (Voghera 1996), and (*unità*) *polirematica* ‘multi-word (lexical) unit’ (Voghera 2004). The problem remains as to whether they should be included in morphology or in the lexicon: some of these constructions are included in dictionaries as separate entries, like compounds, others are not recognized and rather fall under the category of lexical units or syntactic constructions. In this study, I will consider these loose compounds as far as they are attested in dictionaries and as prepositional phrases when they are not recognized. The distinctions between strict and loose compounds and between morphological and syntactic constructions will be essential in the contrastive analysis conducted in Section 5.3.

In the present study, I will use the term ‘suffixoid’ for such recurrent elements as slang *-ass*, *-brain*, *-face*, and *-head*, because like compound constituents they correspond to free words, but unlike them they acquire specific pragmatic features connected with impoliteness (cf. Mattiello and Dressler 2018). In particular, I will focus on the pragmatic contribution that such slang constituents can make to offensive discourse and face-threatening acts.

4 Data and methodology

The data analyzed in this study were collected from the *Corpus of Contemporary American English* (COCA, Davies 2008). It is the most widely used corpus of American English created by Mark Davies, freely accessible at <https://www.english-corpora.org/>. It contains more than one billion words of text (20 million words each year) from eight genres: i.e., spoken, fiction, popular magazines, newspapers, academic texts, and (recently updated) TV and movies subtitles, blogs, and other web pages.

For the selection of relevant data, I used a twofold process. First, I automatically searched for complex words ending in the selected slang suffixoids (**ass*, **brain*, **face*, **head*). The asterisk allowed me to obtain as results all words ending in the selected morphemes. However, sometimes these were word chunks (e.g. *compass* or *embarrass* do not include the suffixoid *-ass*). Hence, manual selection was necessary to clean the data from irrelevant examples.

Second, the selected data were searched in the digital edition of *Green’s Dictionary of Slang* (henceforth, GDS) for meaning and relevance. The digital version of GDS is available in updated online form at <https://greensdictofslang.com/>, including advanced search tools and the words’ etymologies. In GDS, the suffixoids under investigation were all labelled ‘suffixes’, with the exception of *-brain*. The senses of the complex words ending in a suffixoid are all from GDS.

The lexicographic inspection allowed me to restrict the number of formations to recognized slang nouns. For instance, words such as *busy-ass*, *bony-ass*, *lame-ass*, *long-ass*, *stupid-ass*, *sweet-ass*, or *ugly-ass* attested in COCA but not attested in GDS have been excluded from the analysis. However, they show the productivity of the suffixoids in terms of availability and profitability in the formation of nonce words or even of hapax legomena (once-only attestations) (Bauer 2001).³ In addition, *cheap-ass*, *fancy-ass*, *fine-ass*, *nasty-ass*, *weak-ass*, and *weird-ass* have been

³ By ‘availability’ I mean the potential a process offers the language user to produce new words, while ‘profitability’ is the actual use of the process to coin new words.

rejected because they are only attested as slang adjectives, whereas *sorry-ass* has been included because attested both as a noun and as an adjective, even if the raw frequency in COCA does not distinguish between the two parts-of-speech.

When two spellings (i.e. solid and hyphenated) were attested in the corpus, they have both been included in the quantitative analysis in order to identify the most frequent spelling: e.g., solid *badass* is more frequent than hyphenated *bad-ass* in COCA (see Table 1). Spelling as two separate words was not taken into account because affixoids are attached to bases.

The selection process yielded 28 *-ass* words, 15 *-brain* words, 31 *-face* words, and 51 *-head* words (different spellings are considered different words here because COCA gives them as separate entries). Most of the senses associated with these words have to do with foolishness, stupidity, incompetence, or are terms of abuse associated with vulgarity and derogatoriness. For *-head* words, those having meaning related to drugs have been excluded (e.g. *cokehead* ‘a regular cocaine user’, *dopehead* ‘a drug user’, *gearhead* ‘a narcotics addict’, *horsehead* ‘a heroin addict’) because their use is generally not connected with either impoliteness or offensiveness.

For the data analysis, I used the quantitative data obtained from COCA (both raw and normalized frequencies). Some of the words were only attested once in the corpus, as nonce words, and this confirms the productivity of the affixoids not only in terms of frequency but also in terms of profitability in the formation of new (less well-known) words (Bauer 2001).

In a previous study, Mattiello (2023) investigated a set of relevant examples of such formations referring to people with negative or undesirable characteristics, such as foolishness or stupidity, and presented their raw frequencies in COCA, along with normalized frequencies across different genres. In this study, I consider both high-frequency examples and hapaxes to demonstrate the productivity of the suffixoids and their evolution toward bound morphemes used with different meanings. These meanings are identified through a comprehensive analysis of the collocates of complex words ending in a suffixoid.

For the morphopragmatic analysis, I first selected the most frequent words from the quantitative analysis and examined their contexts and co-texts. The qualitative analysis supports the findings of previous research with additional examples and contexts, thereby confirming the functions previously discussed by Mattiello (2023). Some tokens of the most frequent types in COCA were excluded from the qualitative analysis because they were personal names (e.g. *Butt-head* is also a character’s personal name) or irrelevant to my analysis (e.g. *Bubblebrain* ‘a music label’ and *Fatbrain* ‘a toy company’ are brand names of business companies, while *boldface* is a technical term of printing referring to ‘a thick font type’). A close reading of the words in context (mainly spoken discourse or blog and web

genres) was necessary in order to select relevant examples showing the pragmatic functions/effects associated with *-ass*, *-brain*, *-face*, and *-head* words. The main pragmatic functions/effects investigated were the derisive function (§ 5.2.1), the critical function (§ 5.2.2), and the offensive function (§ 5.2.3).

Finally, for the English-Italian contrastive analysis, I used the *WordReference* dictionary and the parallel English-Italian subcorpus OPUS2 to explore the possible Italian translations of such English formations. I also compared English and Italian in terms of their different morphological systems, diverse morphological means and morphological richness.

5 Analysis

5.1 Quantitative analysis

The quantitative analysis of the four suffixoids under investigation (*-ass*, *-brain*, *-face*, *-head*) includes an exploration of their most common types, which are also classified into semantic categories based on the GDS meanings. Additionally, the quantitative analysis presents the raw and normalized frequencies of the tokens and the immediate collocates of a representative type (*badass*, *fatbrain*, *fuckface*, *butt-head*), to investigate whether the collocates confirm the impolite contexts in which these slang suffixoids are used.

5.1.1 The suffixoid *-ass*

Table 1 reports the quantitative data (raw and normalized (per million words) frequency in COCA) of formations exhibiting the suffixoid *-ass*. This suffixoid shows 22 different types (excluding different spellings), with 5,616 tokens overall.

Table 1: Raw and normalized frequency of *-ass* words in COCA.

-ASS WORD	RAW FREQUENCY IN COCA	NORMALIZED FREQUENCY (PMW) IN COCA
badass	1,807	1.82
bad-ass	322	0.32
dumbass	825	0.83
dumb-ass	416	0.42

Table 1 (continued)

-ASS WORD	RAW FREQUENCY IN COCA	NORMALIZED FREQUENCY (PMW) IN COCA
big-ass	420	0.42
bigass	19	0.02
smart-ass	397	0.40
hard-ass	204	0.21
hardass	46	0.05
punk-ass	150	0.15
fat-ass	142	0.14
fatass	63	0.06
crazy-ass	136	0.14
wiseass	120	0.12
wise-ass	89	0.09
sorry-ass	101	0.10
half-ass	84	0.08
tight-ass	75	0.08
candy-ass	57	0.06
shit-ass	23	0.02
skinny-ass	21	0.02
dead-ass	19	0.02
jive-ass	16	0.02
pansy-ass	16	0.02
lard-ass	13	0.01
funky-ass	13	0.01
mean-ass	12	0.01
cracker-ass	10	0.01

According to GDS, *-ass* is “used to form generally negative (but increasingly positive too) adj[ective]s and occas[ionally] nouns”. Hence, its function is mainly intensifying in the case of adjectives. Since my analysis focuses on nouns, *-ass* nouns can be classified as:

1. referring to arrogant, stubborn or presumptuous people: *big-ass/bigass*, *smart-ass*, *hard-ass/hardass*, *wiseass/wise-ass*, *jive-ass*;
2. referring to foolish or aggressive people: *badass/bad-ass*, *dumbass/dumb-ass*, *crazy-ass*;
3. referring to unpleasant people (as terms of abuse): *punk-ass*, *tight-ass*, *shit-ass*, *funky-ass*, *mean-ass*;

4. referring to overweight or skinny people: *fat-ass/fatass, skinny-ass, lard-ass, cracker-ass*;
5. referring to stupid or incompetent people: *sorry-ass, half-ass*;
6. referring to cowards or effeminate people: *candy-ass, pansy-ass*;
7. referring to idle, lazy people: *dead-ass*.

These seven semantic categories, listed in order of type frequency, confirm the negative character of the suffixoid, with a potential use in contexts of impoliteness, offensiveness and aggressiveness.

When considering the most frequent type – i.e. *badass/bad-ass* – its most relevant collocates in COCA are noteworthy. These collocates, arranged by their Mutual Information (MI) score (indicated in brackets), reflect the strength of association between the word and its immediate collocates.

- *badass* + NOUN: *motherfucker* (6.36), *chick* (6.34), *dude* (5.50), *hunter* (5.10), *bitch* (4.53), *monster* (3.85), *killer* (3.53), *character* (3.28), *boss* (2.42), *shit* (2.02);
- *badass* + ADJ: *pretty* (6.60), *ultimate* (5.19), *total* (4.48), *inner* (4.43), *fucking* (4.19), *super* (3.93), *favorite* (3.58), *cool* (3.01), *complete* (2.68), *real* (2.55);
- *badass* + VERB: *fool* (3.48), *name* (2.44), *embrace* (2.38), *swear* (2.36), *sound* (2.30), *remind* (2.30), *feature* (2.19), *kick* (2.10), *slip* (2.03), *mix* (2.01);
- *badass* + ADV: *miserably* (5.38), *fatally* (5.21), *vaguely* (4.85), *boldly* (4.92), *alright* (3.52), *totally* (2.66), *definitely* (2.58), *fucking* (2.54), *constantly* (2.36), *pretty* (2.35).

The collocates for *badass* (e.g., *motherfucker, bitch, shit, fool, miserably*) and the adjectives/adverbs used to emphasize this noun (*total, fucking, complete, definitely*) highlight its negative connotation in some contexts. However, *badass* is a polysemous word that can also carry a positive meaning in slang when used as an adjective, i.e., ‘formidable, admirable, first-rate,’ as suggested by some of its collocates (e.g., *pretty, cool, real*). Therefore, it will be excluded from the qualitative analysis in Section 5.2.

5.1.2 The suffixoid *-brain*

Table 2 reports the frequency in COCA of formations exhibiting the suffixoid *-brain*, amounting to 14 different types (excluding different spellings) and 39 tokens.

Table 2: Raw and normalized frequency of *-brain* words in COCA.

-BRAIN WORD	RAW FREQUENCY IN COCA	NORMALIZED FREQUENCY (PMW) IN COCA
birdbrain	52	0.05
bird-brain	5	0.01
fatbrain	8	0.01
lame-brain	7	0.01
bubblebrain	3	0.00
pea-brain	3	0.00
mush-brain	2	0.00
mouse-brain	2	0.00
wet-brain	2	0.00
shit-brain	1	0.00
beanbrain	1	0.00
bean-brain	1	0.00
butterbrain	1	0.00
fuckbrain	1	0.00
dickbrain	1	0.00
meatbrain	1	0.00

In GDS, *brain* is not considered a suffix(oid) but rather an independent word displaying several meanings, one of which is “a stupid person”. However, words ending in *-brain* are compared to *-head* words in the dictionary (e.g. *fat-brain* and *fat-head*). All *-brain* words reported in Table 2 convey a derogatory sense of ‘a fool, a simpleton’. Some of them are marked as derogatory (e.g. *fatbrain*, *dickbrain*) and others clearly activate a metaphorical sense: e.g., *bubblebrain* ‘a foolish, careless person, with a brain like a bubble’, *bird-brain* ‘person having a brain comparable to that of a bird’, *pea-/bean-brain* ‘person having a brain of the dimensions of a pea or bean’.

These meanings confirm the negative connotation of the words containing the suffixoid *-brain*. As for the most frequent type, *fatbrain*, given the small number of contexts provided by the corpus, no collocates could be identified.

5.1.3 The suffixoid *-face*

Table 3 shows the frequency in COCA of formations exhibiting the suffixoid *-face*, amounting to 22 different types (excluding different spellings) and 675 tokens.

Table 3: Raw and normalized frequency of *-face* words in COCA.

-FACE WORD	RAW FREQUENCY IN COCA	NORMALIZED FREQUENCY (PMW) IN COCA
boldface	283	0.28
bold-face	9	0.01
fuckface	71	0.07
fuck-face	40	0.04
paleface	51	0.05
leatherface	50	0.05
pigface	29	0.03
shitface	27	0.03
shit-face	2	0.00
dogface	26	0.00
dog-face	2	
yellow-face	15	0.02
yellowface	5	0.01
assface	8	0.01
ass-face	3	0.00
buttface	7	0.01
butt-face	2	0.00
fish-face	6	0.01
fishface	3	0.00
moon-face	6	0.01
puke-face	6	0.01
dickface	5	0.01
stoneface	4	0.00
pie-face	3	0.00
cuntface	3	0.00
arseface	2	0.00
rat-face	2	0.00
ratface	2	0.00
monkey-face	1	0.00
apectace	1	0.00
cow-face	1	0.00

According to GDS, *-face* is “a s[uf]f[i]x used in comb[ination] with an abusive epithet to form a derog[atory] term of address”. Hence, its function is mainly derogatory, offensive and abusive. In particular, *-face* nouns fall into the following semantic categories, with some words belonging to more than one in GDS:

1. referring to unpleasant, contemptible or stupid people: *boldface/bold-face, fuck-face/fuck-face, shitface/shit-face, dogface/dog-face, assface/ass-face, fish-face/fishface, puke-face, dickface, pie-face, arseface, rat-face/ratface, monkey-face*;

2. referring to unattractive people: *pigface*, *buttface/butt-face*, *fish-face/fishface*, *cuntface*, *monkey-face*, *apeface*;
3. forming racist terms: *paleface*, *yellow-face/yellowface*, *moon-face*, *pie-face*;
4. referring to unemotional people: *leatherface*, *stoneface*;
5. referring to despicable women: *cow-face*.

These five semantic categories, again listed in order of type frequency, suggest a negative feature of the suffixoid.

Given the ambiguity of *boldface*, also used as font name, *fuckface* has been taken into consideration for collocations. The term *fuckface* shows collocates (mainly nouns and adjectives) with negative, offensive or vulgar meaning, arranged in MI order: e.g., piece-of-shit (13.35), cocksucker (11.84), asshole (8.82), fascist (8.69), motherfucker (8.50), dick (5.49), stupid (5.45), fuck (5.03), ass (4.94). This and related formations, therefore, encourage a morphopragmatic analysis to verify the specific use and functions of the suffixoid.

5.1.4 The suffixoid *-head*

Formations exhibiting the suffixoid *-head* are collected in Table 4, with information about their frequency in COCA. This suffixoid shows 37 different types (excluding different spellings), with 7,488 tokens overall.

Table 4: Raw and normalized frequency of *-head* words in COCA.

-HEAD WORD	RAW FREQUENCY IN COCA	NORMALIZED FREQUENCY (PMW) IN COCA
buckhead	1,219	1.23
butt-head	669	0.67
butthead	226	0.23
pinhead	669	0.67
pin-head	8	0.01
dickhead	661	0.67
dick-head	22	0.02
shithead	495	0.50
shit-head	6	0.01
flathead	445	0.45
flat-head	26	0.03
bonehead	427	0.43
bonehead	8	0.01

Table 4 (continued)

-HEAD WORD	RAW FREQUENCY IN COCA	NORMALIZED FREQUENCY (PMW) IN COCA
hammerhead	338	0.34
egghead	335	0.34
knucklehead	333	0.34
marblehead	288	0.29
airhead	183	0.18
sheepshead	182	0.18
meathead	174	0.18
fuckhead	173	0.17
fuck-head	9	0.01
fathead	131	0.13
fat-head	6	0.01
motorhead	64	0.06
motor-head	3	0.00
conehead	60	0.06
cheesehead	58	0.06
bubblehead	35	0.04
leatherhead	24	0.02
chickenhead	23	0.02
potatohead	22	0.02
potato-head	6	0.01
ironhead	22	0.02
butterhead	22	0.02
knothead	18	0.02
knot-head	5	0.01
buckethead	17	0.02
bucket-head	8	0.01
hardhead	12	0.01
stupidhead	11	0.01
beanhead	11	0.01
bean-head	3	0.00
dumbhead	9	0.01
bullet-head	5	0.01
asshead	4	0.00
lemonhead	4	0.00
melon-head	4	0.00
melonhead	3	0.00
wethead	1	0.00
lardhead	1	0.00

In GDS, *-head* is included as a main entry to indicate a suffix “used in a variety of comb[ination]s in which *-head* is linked to a n[oun] to create a term with a neg[ative] personal meaning”. Several examples of *-head* formations are provided by the dictionary whose meanings are related to foolishness: e.g., *airhead*, *blockhead*, *bonehead*, *cabbage-head*, *dickhead*, *egghead*, *meathead*, *pinhead*, *pointy-head*, and *shithead*. They mostly have metaphorical meaning, implying that the head in question is shaped like or otherwise resembles the noun it accompanies. All the formations included in Table 4 refer to ‘fools or stupid people’, with the exception of *buckhead* ‘a person with buck teeth’.

If we consider the most frequent type meaning ‘a stupid or obnoxious person’ – i.e. *butt-head* – its most relevant collocates in COCA include:

- *butt-head* + NOUN: cartoon (6.52), chick (6.12), angle (5.44), aisle (5.24), pause (5.08), creator (4.82), butt (4.17), plane (3.98), cow (3.81), idiot (3.54);
- *butt-head* + ADJ: nerdy (7.14), upsetting (6.74), muddy (5.81), unaware (5.25), dumb (5.15), cool (4.65), overweight (4.37), asleep (3.68), excited (3.44), stupid (2.24);
- *butt-head* + VERB: butt (7.67), suck (4.45), stare (3.85), kick (3.75), shut (3.70), wake (3.58), check (3.49), laugh (3.10), close (2.22), watch (2.17);
- *butt-head* + ADV: blankly (7.04), dead (4.78), wide (4.31), randomly (4.30), tight (3.93), suddenly (3.04), aside (2.99), barely (2.91), sure (2.17).

Not all the collocates for *butt-head* display a negative connotation; yet, some collocates are negative, especially vulgar (*butt*) or offensive (*cow*, *dumb*, *idiot*, *overweight*, *stupid*), or imply an aggressive tone (*furiously*, *shut (up)*, *suck*, *kick*, etc.), thus encouraging a morphopragmatic approach for the qualitative analysis.

5.2 Qualitative analysis

For the qualitative analysis, the suffixoids explored so far from a quantitative viewpoint will be contextualized to ascertain their morphopragmatic functions. The specific pragmatic functions that dominate the language of impoliteness and offensiveness can be grouped into three conflict functions; namely, derision, criticism, and offence (see “conflict situations” in Hamilton 2012). Hence, each of these functions – namely, derisive, critical and offensive – will be dealt with in Subsections 5.2.1–5.2.3 in order to investigate in more detail how they are served by the morphological means under investigation.

5.2.1 Derisive function

Speakers often opt for slang suffixoids as an indication of scorn, mockery, disdain, ridicule, and consequently disrespect and impoliteness, with the intent to laugh at someone perceived as stupid or of no value. The feature [derisive] deriving from a character [playful] or [non-serious] (Dressler and Merlini Barbaresi 1994) may produce different morphopragmatic effects. For instance, speakers may use slang *-brain* words with a mocking intent, as in examples (1)–(2):

- (1) Of course, Mitten's never did that, did he? # Instead, it was revealed this past Monday, that it was Mormon Bishop Mittens who lied for POLITICAL GAIN. # You are a **bird-brain**! We are electing a president, not an Imam or priest. (BLOG, *US intel believes some Benghazi attackers tied to al Qaeda in Iraq*, 2012)
- (2) Hey! Paul, Staten Island, what do you got? Hey, Dogg, how you doing? Um, listen, I'm just calling in response to that **pea-brain** Philadelphia Phil. I'm listening here. It's unbelievable! I mean, you ever hear of innocent until proved guilty? It's a little thing called the American Constitution. (MOVIE, *Big Fan*, 2009)

The general tone in (1) is hilarious, humorous, or even sarcastic. Indeed, the accusation *You are a bird-brain!* is followed by the statement *We are electing a president, not an Imam or priest*, which indicates a mocking intent towards the addressee. By contrast, in (2) the word *pea-brain* has an external referent (*I'm just calling in response to that pea-brain Philadelphia Phil*), who is laughed at by the following utterances *It's unbelievable! I mean, you ever hear of innocent until proved guilty? It's a little thing called the American Constitution*. The speaker is clearly sarcastic when (s)he defines *the American Constitution* by using an ironic understatement (*a little thing*), which confirms his/her derisive attitude.

The function of *-head* words is similarly derisive in (3)–(5):

- (3) Show me proof of what you are saying is more than what you've gathered from your own opinions or watching Fox News. Give me concrete evidence so that I can believe you. Prove it to me. Otherwise, all you are is just a **pinhead** with an inflated sense of ego. And that's unfortunate. # HAHA, you people are so ignorant and so blind, you wouldn't see a door if it splits your heads wide open. . . your comments keep me entertained. . . (WEB, *Romney staff refusing to let frostbitten voters leave PA rally*, 2012)
- (4) Ash, You are a simpering, childish fool whose intellect is dwarfed by your stupidity. You come in here and talk like a little **dickhead** in front of the

grownups and you get pissy when we ignore you. I crap out better men than you every morning. You are a vapid, ignorant, small minded fool whose last happy day will be the one before the election. (WEB, *We Will Not Be Silenced*, 2012)

- (5) Yes, I realize it's a catfish. What the fuck do you know about fish, you're just a stupid beetle. Shut up, asshole. # What's so fucking surprising, **shithead**? Like you weren't just sitting around waiting to be discovered by humans. (WEB, *Fuck You, Penguin*, 2012)

Again, the general tone is aggressive and even sarcastic and *-head* words contribute to helping speakers in their mocking intent. Many collocates confirm the impolite context: e.g., nouns (*asshole, beetle, fool, stupidity*), adjectives (*blind, ignorant, pissy, small minded, stupid, unfortunate, vapid*), verbs (*crap out, shut up*), adverbs (*fucking*), and longer vulgar expressions (*what the fuck*).

Another vulgar context that also shows the use of an *-ass* word with a derisive function is illustrated in (6):

- (6) Bullshit. Translucent. What the f. . . Well, come on, you lime-sucking **smart-ass**. How the hell did you do it? (TV, *The Boys*, 2019)

In this example, the word *smart-ass* is derisory for the arrogant, know-it-all behaviour of the addressee (*Well, come on, you lime-sucking smart-ass*), as confirmed by the various negative (esp. aggressive and vulgar) collocates (*bullshit, what the f. . ., how the hell. . .*).

The derisive function also partly overlaps with the critical function, as most of the above contexts also suggest a critical attitude, more closely investigated in Section 5.2.2.

5.2.2 Critical function

Some of the same morphological means can take a disparaging meaning/function when used in a critical context, to express negative opinions or judgments about the bad qualities of someone or to condemn their behaviour. For instance, *-ass* words can be used to express critical remarks such as:

- (7) He's a dick it seems for the sake of being a dick; he's just a **dumbass** on nearly everything. . . a real hatchet mouth. . . (BLOG, *Do You Want To Watch Devin Faraci Get - Ain't It Cool News*, 2012)

- (8) My dad is such a **hard-ass**. He's breaking my balls 'cause I'm flunking history. Who needs to learn about shit that already happened? If they had a class called "future," I'd be all over it. (TV, *F Is for Family*, 2015)
- (9) Babe, are you still upset because of what **fat-ass** said? 'Cause I'll kick his ass if you want me to. What, I will, for you. I'm not gonna let anyone disrespect my girl. (MOVIE, *Night of Something Strange*, 2017)

All these words express criticism towards a third person, an external referent (*he's just a dumbass on nearly everything*, *My dad is such a hard-ass*, *Babe, are you still upset because of what fat-ass said?*). The speakers' disparaging attitudes are confirmed by the presence of offensive words (*dick*) and other vulgar expressions (*breaking my balls*, *shit*, *kick his ass*), showing the negative connotations of *dumbass*, *hard-ass*, and *fat-ass*.

In blogs and on the Web, *-brain* and *-head* words display a similar condemning function:

- (10) I'm not say it's not true just I having see the correct data. # Is there no subject of which he has no understanding that the **pea-brain** Michael Moore will not offer up an ignorant, emotional, and wrong opinion about? (BLOG, *Michael Moore on climate change and Superstorm Sandy*, 2012)
- (11) Jim Braselman # Here is the deal, and it is beyond dispute and self-evident. The GOP, got SPANKED in the election - badly spanked. And they are desparate (sic) to get back at Obama any way they can. This **pinhead** is elected in California, where GOP registration has fallen BELOW 30% statewide!! (BLOG, *Anderson Cooper Gets Heated With GOP Rep. Over Benghazi*, 2012)
- (12) You cannot impeach Obama because he is a **dickhead**. Being stupid is not against the law, you can not legislate to make people be smart. Obama is part of the price everybody pays for living in a free, democractic (sic) country. (WEB, *The American Spectator: Building 'Frankenstein' in the Middle East?*, 2012)

These three contexts show that the Internet is often an attractor of criticism, because people tend to use online posts to comment on political or other public people's behaviour (*the pea-brain Michael Moore*. . . , *This pinhead is elected in California*, *You cannot impeach Obama because he is a dickhead*). Collocates such as *ignorant*, *wrong*, or *stupid* confirm the negative attitude of the bloggers or web users. Yet, again, it is difficult to provide a clear distinction among the three functions of the slang suffixoids analyzed. The effect produced by *pea-brain*, *pinhead*, and *dickhead* is indeed offensive.

5.2.3 Offensive function

The third function that slang morphological means can perform is the offensive one, i.e. slang suffixoids are often deliberately used to attack and offend someone. Words frequently associated with offence are *fuckface* and *paleface*:

- (13) Hey, what the fuck, two out of three ain't bad. You expect everyone who runs a fucking football team to be a fucking Mother Theresa, like the fucking piece of shit who wrote that fucking article, right assholes? # Get fucking cancer, **fuckface**, and hey, if you've already got it, maybe you'd like to buy a nice comfy mattress to rest on during your convalescence. (WEB, *The Cranky Redskins Fan's Guide to Dan Snyder - Washington City*, 2012)
- (14) Now, is the salad bar still free, though? Why would you care? Doesn't look like you've had a salad in your whole life. I'm not telling you again. Now, get in here! - Die, **paleface**! - You little bastards! Hey! Hey, now! Hey, now! Hey! Hey! Knock that shit off! Those are my good golf clubs! (MOVIE, *Delta Farce*, 2007)

Both contexts are aggressive and extremely offensive, as the speakers are directly attacking their addressees by wishing them to get sick (*Get fucking cancer, fuckface*) or to die (*Die, paleface!*). Co-textual material confirms the speakers' attacking tone: in (13), the adverb *fucking* is used five times, accompanied by other abusive expressions (*piece of shit, assholes, what the fuck*), in (14), *paleface* is followed by the epithet *bastards* and the colourful expression *Knock that shit off*.

A similar offensive function is displayed by *-head* and *-ass* words in (15)–(18):

- (15) Would Amazon be better if there weren't any fake reviews at all? # Go read them now, before some **pinhead** complains to Amazon and they get removed. # After reading my reviews, take some time to read other reviews of those same products. Look at the pics customers have uploaded as well. You'll probably enjoy it as much as I did. (BLOG, *A Newbie's Guide to Publishing*, 2012)
- (16) I get that sometimes the other side is that people are ignorant idiots. Unfortunately, being a **dumbass** does not preclude you from reproducing. (BLOG, *Motherhood Uncensored: The other side*, 2012)
- (17) You never heard about him running around on any of his wives with other women. So where is he getting his sex? He's either a porn addict, or he is

GAY!!!! # H # This is really offensive in several ways. A little tip from me to you: try not being such a fucking **dickhead** next time, maybe you'll actually get a chuckle out of someone! (WEB, *Tom Cruise And Katie Holmes Divorce And I Think It's Hilarious*, 2012)

- (18) Hey, Juan, slow. . . Come on, guys. Slow it down. Juan. Hey, hey. Hey, Alvey! Alvey! The fucking fight is tonight, motherfuckers. You wanna lose that fucking million dollars, you **shithead**? You fucking eat yet, by the way? Hey. You fucking eat? I'll make you some fucking oatmeal. (TV, *Kingdom*, 2017)

In (15)–(16) the offence has no specific addressee (*before some pinhead complains. . . , being a dumbass does not preclude you from reproducing*), whereas in (17)–(18) the speakers are more direct and address their addressee with an offensive word (*try not being such a fucking dickhead next time, you shithead*). The contexts and collocates confirm that the offensive function overlaps with derision (in (16)) and with criticism (in (17)).

5.3 Contrastive analysis

For the contrastive analysis, the words resulting from the quantitative analysis reported in Tables 1–4 will be reconsidered from the translation perspective. The *WordReference* lexicographic tool and the parallel English-Italian OPUS2 corpus will be used to inspect how words ending in an English slang suffixoid can be translated in a dictionary and are rendered in the corpus. For *-ass* words, only seven are found in *WordReference* and six in OPUS2. Table 5 shows their possible Italian translations.

Table 5: Meaning and translation of *-ass* words in GDS, *WordReference*, and OPUS2.

-ASS WORD	MEANING IN GDS	TRANSLATION IN WORDREFERENCE	TRANSLATION IN OPUS2
badass	an unpleasant, aggressive individual	tosto, tipo tosto, duro, cazzuto	bastardo
dumbass dumb-ass	a fool	fesso, cretino, idiota, imbecille	stronzo, idiota, imbecille
smart-ass	one who sees themselves as cleverer than they really are	saputello, saccente, spocchioso, signor sottutoio, sputasentenze	sapientone, genietto, furbacchione

Table 5 (continued)

-ASS WORD	MEANING IN GDS	TRANSLATION IN WORDREFERENCE	TRANSLATION IN OPUS2
hard-ass hardass	a tough person, a thug	duro	osso duro, stronzo
fat-ass fatass	a very fat person; also as a term of address	ciccione, panzone, palla di lardo	ciccione, grassone
wiseass wise-ass	one who sees themselves as cleverer than they really are	sapientone, sotuttoio	furbone, furbino, sapientone, sbruffone
tight-ass	a mean person, a skinflint	tirchio	—

As Table 5 shows, the most common morphological strategy used to translate *-ass* words into Italian is the use of evaluative suffixes, especially augmentatives with a suffix *-one*: e.g., *smart-ass* → It. *sapient-one* ‘wise-AUG’, *furb-acchi-one* ‘sly-INFIX-AUG’; *fat-ass* → It. *cicci-one*, *grass-one* ‘fat-AUG’. Diminutives are less frequently used in translation: e.g., *smart-ass* → It. *saput-ello* ‘known-DIM’, *geni-etto* ‘genius-DIM’; *wiseass* → It. *furb-ino* ‘sly-DIM’. Moreover, most of these words (*sapientone*, *sbruffone*, *saputello*) are now lexicalized in Italian, meaning that they have lost their morphological boundaries. A derived word as Italian translation is in *badass* → It. *cazz-uto* ‘dick-SUFF’, where the suffix *-uto* expresses approbation (Rossi 2011).

As an alternative strategy, Italian may recur to compounding: e.g., V+N strict compounds (*smart-ass* → It. *sputa+sentenze* ‘spill+sentences’), loose compounds (*badass* → It. *tipo+tosto* ‘guy+determined’), or compounds deriving from phrases (*smart-ass* → It. (*signor*) *so+tutto+io* ‘Mr+know+all+I’ ‘know-it-all’; *hard-ass* → It. *osso+duro* ‘bone+hard’) (see Section 3). Prepositional phrases occur in the translations of *fat-ass* → It. *palla di lardo* ‘ball+of+lard’.

As for *-brain* words, only two types are found in both *WordReference* and *OPUS2*. Table 6 shows their possible Italian translations.

Table 6: Meaning and translation of *-brain* words in GDS, *WordReference*, and *OPUS2*.

-BRAIN WORD	MEANING IN GDS	TRANSLATION IN WORDREFERENCE	TRANSLATION IN OPUS2
lame-brain	a fool, a simpleton	stupido, scervellato	testa di caprone, idiota
bird-brain	a fool; also as a term of address	cervello di gallina	sciocco, coglione, cervello di gallina

As the table shows, for the rendering of both *-brain* words prepositional phrases are suggested: e.g., *bird-brain* → It. *cervello di gallina* ‘brain+of+hen’, *lame-brain* → It. *testa di caprone* ‘head+of+goat-AUG’, the latter also exhibiting an augmentative suffix. A derived word as translation is in *lame-brain* → It. *s-cervell-ato* ‘PREF-brain-SUFF’, with a privative prefix *s-* (‘without (brain)’) and a past participle suffix *-ato* (the origin of the term is from the verb *scervellare*).

Similarly, only two *-face* words are found in both *WordReference* and OPUS2. Table 7 shows their possible Italian translations.

Table 7: Meaning and translation of *-face* words in GDS, *WordReference*, and OPUS2.

-FACE WORD MEANING IN GDS		TRANSLATION IN WORDREFERENCE	TRANSLATION IN OPUS2
fuckface	a fool, an idiot, a generally contemptible person	faccia di merda, faccia da culo, faccia da cazzo	faccia di merda, faccia da cazzo
paleface	a white person	viso pallido	viso pallido

For *paleface*, both the dictionary and the corpus provide the compound *viso pallido* ‘face+pale’, which is a literal translation of the English word. For *fuckface*, instead, some prepositional phrases are suggested: *fuckface* → It. *faccia di merda* ‘face+of+shit’, *faccia da/di culo* ‘face+of+ass’, *faccia da cazzo* ‘face+of+ dick’.

The last group of *-head* words is more numerous and for thirteen cases both *WordReference* and OPUS2 provide at least one possible translation. All translations are reported in Table 8.

Table 8: Meaning and translation of *-head* words in GDS, *WordReference*, and OPUS2.

-HEAD WORD	MEANING IN GDS	TRANSLATION IN WORDREFERENCE	TRANSLATION IN OPUS2
butt-head butthead	a stupid or obnoxious person	coglione, cazzone, minchione	stupidotto, idiota
pinhead pin-head	a stupid person	testa di cavolo, zucca vuota	testa vuota, cervello di formica, cervello di gallina
dickhead dick-head	a fool, an incompetent	testa di cazzo	testa di cazzo, testa di minchia, coglione, deficiente
shithead shit-head	a derog[atory] term of general abuse, occas[ionally] nickname	stronzo, pezzo di merda, testa di cazzo	stronzo, faccia di merda, testa di cazzo

Table 8 (continued)

-HEAD WORD	MEANING IN GDS	TRANSLATION IN WORDREFERENCE	TRANSLATION IN OPUS2
bonehead bonehead	a fool, a dullard, an idiot; thus as a term of address	tonto, zuccone, fesso	testone, testa vuota, testa di rapa
hammerhead	anyone stupid and obstinate	testa di legno, zuccone	testa di legno
egghead	a fool; a bald person	testa d'uovo	cervellone
knucklehead	a term of abuse, a description for any foolish, stupid, slow person	testone, zuccone; testa di legno	testa dura, cretino, deficiente
airhead	an idiot, a fool, someone who has nothing but air, and no brains, in their head	testa vuota, testa di legno; zuccone, somaro	testa vuota, svampito
meathead	a stupid person; thus as a term of address	testa di rapa	testa di rapa, testa vuota, polpettone
fuckhead fuck-head	a fool, a complete idiot; esp[ecially] as a term of address	testa di cazzo; stronzo, coglione	testa di cazzo, pezzo di merda, stronzo, coglione
fathead fat-head	a fool, an idiot; often used affectionately as well as derog[atory]	stupido, scemo	testa di rapa, testa di cazzo, capoccione, imbecille
hardhead	a fool	testa dura	testa dura

As Table 8 shows, *-head* words are often translated into Italian by means of an augmentative suffix *-one*: e.g., *butt-head* → It. *cazz-one*, *minchi-one* ‘dick-AUG’ (*minchia* is the popular South Italian word for ‘penis’); *bonehead*, *airhead* → It. *zucc-one* ‘pumpkin-AUG’. Only *butt-head* → *stupid-otto* ‘stupid-DIM’ exhibits a diminutive suffix.

Compounds are provided, for instance, for *pinhead* → *zucca vuota* ‘pumpkin+empty’; *bonehead* → *testa vuota* ‘head+empty’; *knucklehead* → *testa dura* ‘head+hard’. The only translation which involves derivation is *airhead* → *s-vampito* ‘PREF+flame+SUFF’, with a privative prefix *s-* used to refer to someone who is ‘absent-minded’.

Most translations involve a prepositional phrase: e.g., *pinhead* → It. *testa di cavolo* ‘head+of+cabbage’, *cervello di gallina* ‘brain+of+hen’; *dickhead* → It. *testa di cazzo*; *shithead* → It. *pezzo di merda* ‘piece+of+shit’, *faccia di merda* ‘face+of+shit’; *bonehead*, *meathead* → It. *testa di rapa* ‘head+of+turnip’; *hammerhead* → It. *testa di legno* ‘head+of+wood’.

In summary, a contrastive analysis shows that English slang suffixoids are rendered into Italian using a variety of different morphological means: namely, augmentatives (*fat-ass* → It. *grassone* ‘fat-AUG’), diminutives (*wiseass* → It. *furbin* ‘sly-DIM’), derived words (*lame-brain* → It. *scervellato* ‘PREF-brain-SUFF’), strict compounds (*smart-ass* → It. *sputa+sentenze* ‘spill+sentences’), loose compounds (*badass* → It. *tipo tosto* ‘guy+determined’), and prepositional phrases (*bird-brain* → It. *cervello di gallina* ‘brain+of+hen’). The most productive class of *-head* words is generally translated using *testa di X* ‘head of X’ (*pinhead* → It. *testa di cavolo* ‘head+of+cabbage’).

Thus, the predominance of compounding over derivation in English has facilitated the formation and use of suffixoids, which are rare in Italian, neither in slang nor in standard language. Italian, therefore, resorts to other morphological means to compensate for this morphological gap.

6 Conclusions

In Culpeper’s (2005: 38) definition of impoliteness, either the speaker deliberately communicates face-attack or the hearer perceives the speaker’s behaviour as deliberately face-attacking. This study has shown how slang suffixoids *-ass*, *-brain*, *-face*, and *-head* are used in face-aggravation and face attack against an individual or group of individuals, deliberately involving aggressive and offensive content. Unlike compound constituents, suffixoids are productively used to create series of offensive words, most of which are attested in slang dictionaries and in corpora, with some occurring only once, as confirmation of their availability and profitability in the creation of new words.

The quantitative analysis in COCA has shown that the suffixoid *-ass* has produced 22 different types and 5,616 tokens overall, *-brain* has 14 different types and 91 tokens, *-face* has 22 different types and 675 tokens, and *-head* has 37 different types and 7,488 tokens, the latter being the most productive.

The qualitative analysis has shown that these slang suffixoids serve three specific pragmatic functions that dominate the language of impoliteness and offensiveness: namely, derision, criticism, and offense. Many of the suffixoids are added to vulgar or derogatory bases: e.g., *dickbrain*, *fuckbrain*, *shit-brain*; *arseface*, *assface*, *cuntface*, *fuckface*, *shitface*; *butt-head*, *dickhead*, *fuckhead*, *shithead*, *stupidhead*. Hence, they add derogatory and offensive pragmatic meanings to the negatively denoted words. Other suffixoids convey an overall metaphoric meaning: e.g., *bean-brain* or *airhead* suggest that the addressee has a brain of the size of a bean or a head filled with air and no brain at all. Others display metonymic meanings: e.g.,

badass, *big-ass*, and *fat-ass* respectively refer to a person having ‘a bad, big, or fat ass’, i.e. ‘unpleasant, aggressive’; ‘superior’; ‘very fat’.

An exploration of the collocates of *-ass*, *-brain*, *-face*, and *-head* words confirms their impolite and offensive character: they range from other offensive epithets (*asshole*, *motherfucker*, *piece of shit*) to insulting and aggressive adjectives (*small minded*, *stupid*), to imperative forms (*crap out*, *shut up*), intensifiers (*fucking*), and even to colourful expressions (*break my balls*, *kick his ass*, *what the fuck*).

A contrastive analysis reveals that English slang suffixoids are translated into Italian using a variety of morphological strategies, highlighting the differences between the two languages’ morphological systems. This also confirms that English suffixoids occupy a transitional space between derivation and compounding, which explains their rendering in Italian through either derivational or compounding processes, or even through syntactic patterns.

References

- Allen, Irwing L. 1998. Slang: Sociology. In Jacob L. Mey & R. E. Asher (eds.), *Concise Encyclopedia of Pragmatics*, 878–883. Amsterdam: Elsevier.
- Amiot, Dany & Edwige Dugas. 2020. Combining forms and affixoids in morphology. In Rochelle Lieber (ed.), *The Oxford Encyclopedia of Morphology*. Oxford: Oxford University Press. <https://www.oxfordreference.com/view/10.1093/acref/9780190682361.001.0001/acref-9780190682361-e-569>.
- Andersson, Lars G. & Peter Trudgill. 1990. *Bad Language*. Oxford: Blackwell.
- Bauer, Laurie. 2001. *Morphological Productivity*. Cambridge: Cambridge University Press.
- Bauer, Laurie. 2017. *Compounds and Compounding*. Cambridge: Cambridge University Press.
- Bazzanella, Carla, Claudia Caffi & Marina Sbisà. 1991. Scalar dimensions in illocutionary force. In Igor Ž. Žagar (ed.), *Speech Acts: Fiction or Reality?*, 63–76. Ljubljana: Institute for Social Sciences.
- Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.
- Booij, Geert E. & Matthias Hüning. 2014. Affixoids and constructional idioms. In Ronny Boogaart, Timothy Coleman & Gijsbert Rutten (eds.), *Extending the Scope of Construction-based Grammar*, 77–106. Berlin & Boston: De Gruyter Mouton.
- Caffi, Claudia. 2001. *La mitigazione: un approccio pragmatico alla comunicazione nei contesti terapeutici*. Munster: Lit.
- Cantero, Mónica. 2003. *La morfopragmática del español*. München: LINCOM Europa.
- Crocco Galéas, Grazia. 1992. *Gli etnici italiani: Studio di morfologia naturale*. Padova: Unipress.
- Culpeper, Jonathan. 2005. Impoliteness and entertainment in the television quiz show. *The Weakest Link. Journal of Politeness Research: Language, Behaviour, Culture* 1(1). 35–72.
- Culpeper, Jonathan. 2011. *Impoliteness: Using Language to Cause Offence*. Cambridge: Cambridge University Press.
- Culpeper, Jonathan, Paul Iganski & Abe Sweiry. 2017. Linguistic impoliteness and religiously aggravated hate crime in England and Wales. *Journal of Language Aggression and Conflict* 5(1). 1–29.

- Davies, Mark. 2008–. *Corpus of Contemporary American English* (COCA). <https://corpus.byu.edu/coca/> (accessed 18 October 2024).
- Demata, Massimiliano. 2021. Keeping the threat out: Trump's discourse, the wall and the 'other'. *Altre Modernità* 25(5). 270–286.
- Dressler Wolfgang U. & Ferenc Kiefer. 1990. Austro-Hungarian morphopragmatics. In Wolfgang U. Dressler, Hans C. Luschützky, Oskar E. Pfeiffer & John R. Rennison (eds.), *Contemporary Morphology*, 69–77. Berlin: Mouton de Gruyter.
- Dressler, Wolfgang U. & Marianne Kilani-Schoch. 1994. Morphopragmatique interactionnelle: les formations en -o du français branché. In Livia Tonelli & Wolfgang U. Dressler (eds.), *Natural Morphology – Perspectives for the Nineties*, 31–52. Padua: Unipress.
- Dressler Wolfgang U. & Lavinia Merlini Barbaresi. 1986. How to fix interfixes? *Acta Linguistica Hungarica* 36. 53–68.
- Dressler Wolfgang U. & Lavinia Merlini Barbaresi. 1987. Elements of Morphopragmatics. In Jef Verschueren (ed.), *Levels of Linguistic Adaptation*, 33–51. Amsterdam: Benjamins.
- Dressler, Wolfgang U. & Lavinia Merlini Barbaresi. 1994. *Morphopragmatics: Diminutives and Intensifiers in Italian, German, and Other Languages*. Berlin & New York: Mouton de Gruyter.
- Dumas, Bethany K. & Jonathan Lighter. 1978. Is *slang* a word for linguistics? *American Speech* 53. 5–17.
- Eble, Connie. 1996. *Slang and Sociability: In-Group Language among College Students*. Chapel Hill-London: The University of North Carolina Press.
- Erjavec, Karmen & Melita P. Kovačič. 2012. "You don't understand, this is a new war!" Analysis of hate speech in news web sites' comments. *Mass Communication and Society* 15(6). 899–920.
- Green, Jonathon. 2023. *Green's Dictionary of Slang*. Available at <https://greensdictofslang.com/>.
- Hamilton, Mark A. 2012. Verbal aggression: Understanding the psychological antecedents and social consequences. *Journal of Language and Social Psychology* 31(1). 5–12.
- Jespersen, Otto. 1942. *A Modern English Grammar on Historical Principles*, Part VI: *Morphology*. Copenhagen: Ejnar Munksgaard.
- Kilgarrieff, Adam, Pavel Rychlý, Pavel Smrz & David Tugwell. 2004. The Sketch Engine, in *Proceedings of the Eleventh EURALEX International Congress*. Lorient, France.
- Knoblock, Natalia. 2017. Xenophobic Trumpeters: A corpus-assisted discourse study of Donald Trump's Facebook conversations. *Journal of Language Aggression and Conflict* 5(2). 295–322.
- Knoblock, Natalia (ed.). 2022. *Grammar of Hate: Morphosyntactic Features of Hateful, Aggressive, and Dehumanizing Discourse*. Cambridge: Cambridge University Press.
- Korecky-Kröll, Katharina & Wolfgang U. Dressler. 2022. Expressive German adjective and noun compounds in aggressive discourse. In Natalia Knoblock (ed.), *Grammar of Hate: Morphosyntactic Features of Hateful, Aggressive, and Dehumanizing Discourse*, 197–221. Cambridge: Cambridge University Press.
- Mattiello, Elisa. 2008. *An Introduction to English Slang: A Description of its Morphology, Semantics and Sociology*. Monza: Polimetrica.
- Mattiello, Elisa. 2019. "Brexit, Grexit, with the possibility of Spexit": Blend splinters and secreted affixes as creative word-formation mechanisms. In Veronica Bonsignori, Gloria Cappelli & Elisa Mattiello (eds.), *Worlds of Words: Complexity, Creativity, and Conventionality in English Language, Literature and Culture*, 525–538. Pisa: Pisa University Press.
- Mattiello, Elisa. 2022a. *Transitional Morphology: Combining Forms in Modern English*. Cambridge: Cambridge University Press.
- Mattiello, Elisa. 2022b. Language aggression in English slang: The case of the -o suffix. In Natalia Knoblock (ed.), *Grammar of Hate: Morphosyntactic Features of Hateful, Aggressive, and Dehumanizing Discourse*, 34–58. Cambridge: Cambridge University Press.

- Mattiello, Elisa. 2023. Slang and verbal aggression: A morphopragmatic analysis of the compound families *X-ass*, *X-brain*, *X-face*, and *X-head*, *Journal of Language Aggression and Conflict* 11(1). 101–120.
- Mattiello, Elisa & Wolfgang U. Dressler. 2018. The morphosemantic transparency/opacity of novel English analogical compounds and compound families. *Studia Anglica Posnaniensia* 53. 67–114.
- Meibauer, Jörg. 2013. Expressive compounds in German. *Word Structure* 6(1). 21–42.
- Merlini Barbaresi, Lavinia. 2001. The pragmatics of the ‘diminutive’ -y/ie suffix in English. In Chris Schaner-Wolles, John Rennison & Friedrich Neubarth (eds.), *Naturally! Linguistic Studies in Honour of Wolfgang Ulrich Dressler on the Occasion of his 60th Birthday*, 315–326. Torino: Rosenberg and Sellier.
- Merlini Barbaresi, Lavinia. 2006. Morphopragmatics. In Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, Volume 8, 332–335. 2nd edn. Oxford: Elsevier.
- Miró-Llinares, Fernando, Asier Moneva & Miriam Esteve. 2018. Hate is in the air! But where? Introducing an algorithm to detect hate speech in digital microenvironments. *Crime Science* 7(1). 1–12.
- Munro, Pamela (ed.). 2001. *U.C.L.A. Slang 4*. Los Angeles: Department of Linguistics, University of California.
- Oksanen, Atte, James Hawdon, Emma Holkeri, Matti Näsi & Pekka Räsänen. 2014. Exposure to online hate among young social media users. In M. Nicole Warehime (ed.), *Soul of Society: A Focus on the Lives of Children and Youth*, 253–273. Bingley: Emerald.
- Open Parallel Corpus – English* (OPUS2). <https://app.sketchengine.eu/> (accessed 18 October 2024).
- Rossi, Fabio. 2011. Parole oscene. In *Enciclopedia dell’Italiano. Treccani online*. <https://www.treccani.it/enciclopedia/parole-oscene>.
- Sánchez Fajardo, José A. 2022. *Pejorative Suffixes and Combining Forms in English*. Amsterdam/Philadelphia: John Benjamins.
- Scalise, Sergio. 1992. Compounding in Italian. *Italian Journal of Linguistics* 4(1). 175–199.
- Stenström, Anna-Brita. 2000. From Slang to Slanguage: A Description Based on Teenage Talk. In Tamas Kis (ed.), *Mi a szleng?*, 89–108. Debrecen: Kossuth Lajos University Press.
- Stenström, Anna-Brita, Gisle Andersen & Ingrid K. Hasund. 2002. *Trends in Teenage Talk: Corpus Compilation, Analysis and Findings*. Amsterdam & Philadelphia: Benjamins.
- Stevens, Christopher M. 2005. Revisiting the affixoid debate. On the grammaticalization of the word. In Torsten Leuschner, Tanja Mortelmans & Sara de Groodt (eds.), *Grammatikalisierung im Deutschen*, 71–83. Berlin & New York: de Gruyter.
- Voghera, Miriam. 1994. Lessemi complessi: Percorsi di lessicalizzazione a confronto. *Lingua e Stile* 29. 185–214.
- Voghera, Miriam. 2004. Polirematiche. In Maria Grossmann & Franz Rainer (eds.), *La formazione delle parole in italiano*, 56–69. Tübingen: Niemeyer.
- Warren, Beatrice. 1990. The importance of combining forms. In Wolfgang U. Dressler, Hans C. Luschützky, Oskar E. Pfeiffer & John R. Rennison (eds.), *Contemporary Morphology*, 111–132. Belin & New York: Mouton de Gruyter.
- WordReference*. <https://www.wordreference.com/> (accessed 18 October 2024).
- Zhang, Ziqi, David Robinson & Jonathan Tepper. 2018. Detecting hate speech on Twitter using a Convolution-LSTM Based Deep Neural Network. In Aldo Gangemi, Roberto Navigli, Maria-Esther Vidal, Pascal Hitzler, Raphaël Troncy, Laura Hollink, Anna Tordai & Mehwish Alams (eds.), *The Semantic Web*, 745–760. Lecture Notes in Computer Science, 10843. Springer Verlag.

Colin Davis & Hyewon Jang

3 A theoretical and experimental investigation of the morpho-syntax of an anti-honorific prefix in Korean

Abstract: We examine the morphology and syntax of the anti-honorific prefix *che-* in Korean. This morpheme is of interest for research on grammatical politeness, given the commonality of politeness-sensitive morphology, and the rarity of impolite morphology. This element is also of interest for research on Korean, which has many suffixes, but extremely few prefixes. We argue that this prefix *che-* is an anti-honorific, which expresses an irreverent attitude towards the subject of the verb. This is the opposite of the function of the Korean honorific suffix *-si*, which is well-known in linguistic research. Based on an online experiment with native Korean speakers which tested the possible orderings of *che-* relative to the few other Korean prefixes, we argue that *che-* occupies a syntactic position close to the verb. This is similar to what previous work on Korean argues about the honorific *-si*. Some such works argue that *-si* should be understood as the reflex of a morphological agreement relation between the verb and an honorific subject. Analogously, we argue that *che-* can be analyzed as manifesting morphological agreement with a non-honorific subject. We go on to discuss a few remaining details and puzzles about this under-studied morpheme.

Keywords: Korean, honorificity, prefixes, syntax, morphology

1 Introduction

In this paper, we examine an element in Korean verbal morphology which has an impolite or vulgar interpretation.¹ This is the prefix *che-*, which attaches to verbs, as exemplified in (1) below. This element conveys the speaker's lack of reverence or

¹ Unless otherwise specified, the Korean examples in this paper were prepared by Korean native speaker Hyewon Jang. We transliterate Korean using the Yale romanization system.

Colin Davis, Faculty of Education and Arts, Nord University, Universitetsalléen 11, 8026 Bodø, Norway, e-mail: colin.davis@nord.no

Hyewon Jang, Department of Linguistics, University of Konstanz, Universitätsstrasse 10, 78457, Konstanz, Germany, e-mail: hye-won.jang@uni-konstanz.de

politeness towards the subject of the clause that contains it, as we discuss in more detail below. While many languages, including Korean, have honorific markers which encode a respectful attitude towards some referent, the meaning of *che-* appears to be essentially the opposite of an honorific. Therefore we consider this prefix an “anti-honorific”, and gloss it as “AHON”:

(1) Korean anti-honorific prefix

Kyay emcheng che-mek-ess-e
 (s)he much AHON-eat-PST-DECL
 ‘(S)he ate a lot.’

As far as we know, there is no previous work in linguistics which examines this morpheme. This is not surprising, since the colloquial and potentially offensive nature of it likely excludes it from consideration as a topic suitable for formal research. This paper is an initial attempt to fill this gap, and in particular, to situate this element within theoretical work on the syntax and morphology of politeness. A number of questions will remain open for future research, however.

The existence of *che-* stands in contrast to the well-known fact that Korean has a variety of morphological strategies for expressing politeness/honorificity. These include, for example, the hearer-oriented honorific suffix *-yo*, which appears sentence-finally (2a), and the subject-oriented honorific suffix *-si* which is found close to the verb root (2b):

(2) Some Korean polite/honorific suffixes

a. Hearer-oriented honorific

Na-nun hakkyo-ey ka-yo
 I-TOP school- DAT go-HON
 ‘I go to school.’
 (Cho 2022: 113)

b. Subject-oriented honorific

Halape-nim-kkayse cinci-ul capswu-si-n-ta
 Grandfather-HON-NOM cooked.rice-ACC eat-HON-PRES-DECL
 ‘Grandfather eats cooked rice.’
 (Cho 2022: 106)

The subject anti-honorific *che-* is similar to *-si* in that it also attaches directly to the verb in the basic case, though their linear positions happen to be opposites: *-si* is a suffix, while *che-* is a prefix. They also both make a semantic contribution about the speaker’s attitude towards the subject, though in opposite ways. We do not claim

that there is any significance in the fact that the two are opposites in both their meaning and linear position, though we will argue for an analysis in which *che-* and *-si* arise from similar mechanisms. We describe the nature of the meaning of *che-* in more detail in Section 2 below.

1.1 Preview of methodology and analysis

A growing body of work has taken an interest in the syntax and morphology of elements expressing politeness/honorificity (see Miyagawa 2017, Choi and Harley 2019, Alok and Haccidan 2022, Wang 2023, and references therein). However, importantly, there is a lack of work on grammatical expressions of impoliteness (see Van Olmen et al., this volume, and Culpeper, Van Dorst and Gillings, this volume). As far as we know, the marking of impoliteness by dedicated verbal morphology is essentially undocumented in cross-linguistic research.² This impolite morpheme *che-* is worth investigating for this reason alone, but is also of interest to the study of Korean more generally, because of the fact that Korean has very few prefixes. Like its neighbor Japanese, Korean is a typical language of the “Subject-Object-Verb” variety. These are well-known to have an overwhelming preference for suffixation in their morphology.³ As we will discuss, in Korean verbal morphology there are only two previously known verb-selecting prefixes, *an-*, a plain negator, and *mos-*,

2 One case that may also represent impoliteness in verbal morphology occurs in Japanese. Japanese has a (hearer-oriented) polite auxiliary *-mas*, as is well-known (Miyagawa 2017, Yamada 2019), which appears between the verb and tense morphology (i). The same position can contain the auxiliary verb *-yagar(u)* (Kaiser et al. 2001), which encodes a sense of vulgarity and intensification (ii):

i. Japanese polite auxiliary

Sensei-ga *ki-mas-ita*
 teacher-NOM come-HON-PST
 ‘The teacher came’

ii. Japanese vulgar auxiliary

Hee - ano yaroo kekkoo shibutoku nari-yagat-ta na
 Well that guy quite tough become-AHON-PST PRT
 ‘Well, that guy’s gotten quite tough!’
 (adapted from Kaiser et al. 2001: 442)

Kaiser et al. (2001) describe such forms as “pejorative” and disrespectful. While we cannot say to what extent this element is analogous to the Korean *che-*, it is the closest parallel in another language that we are aware of.

3 Another common characteristic of such languages is *pro*-drop, which also occurs in Korean. Therefore in many of our Korean examples, there is no overt subject. We argue that the *che-* prefix

a negative modal (Chung 2007a, Chung 2007b). This third prefix *che-* thus provides a new angle from which to explore Korean morpho-syntax, and simultaneously, to investigate the grammatical expression of impoliteness.

To investigate these topics, we conducted an online experiment in which native Korean speakers judged constructed examples with different relative orders of the three prefixes *che-*, *an-*, and *mos-*. This experiment and its results are discussed in detail in Section 3 below. This experiment revealed a strong tendency for *che-* to appear closer to the verb than the other two prefixes. We use this result as a basis for analyzing the syntax and morphology of *che-*.

Following the analysis of Korean negative prefixes in (Chung 2007a/b), we argue that this experimental result indicates that *che-* occupies a position in the syntax of the clause that is near the verb, structurally intervening between it and functional material such as negation and tense. This proposal serves as a foundation for our theoretical analysis of *che-*. In this analysis, we adopt a theory of syntactic structure that is common in research on what is often termed “generative” or “minimalist” syntax. While this is not the only way one might approach the study of syntax, this will provide a concrete foundation for developing an explicit grammatical analysis of *che-*. We describe in detail the necessary components of the theory in Section 4 below. In summary, the central hypothesis of this syntactic theory is that words instantiate structural units, termed “phrases”, which are arranged in nested hierarchical structures that bind words together into complete sentences. For approachable introductions to this theory of syntax, see for instance Adger (2003) and Carnie (2021), and for a deeper foundational introduction, Chomsky (1995).

We use these syntactic concepts to develop a preliminary structural analysis of *che-*, based on which we then refine our proposal about how exactly this morpheme arises within the structure of the sentence. Many works on grammatical politeness, and Korean politeness in particular such as Choi and Harley (2019) and references therein, assume that politeness morphemes instantiate a grammatical feature [+HON(ORIFIC)]. We propose that there are also instances of this feature that are negatively valued, as [-HON]. We hypothesize that *che-* is the manifestation of such a grammatical feature, though for clarity we gloss this morpheme as [AHON] in our Korean example sentences.⁴ We propose an analysis of the arising of *che-* that

encodes a speaker's non-honorific attitude towards the subject of a clause, but due to *pro-drop*, that subject is not always expressed.

⁴ Wang (2023) argues that honorificity should not be analyzed as stemming from a specific grammatical feature like [HON], because honorificity has a cross-linguistic tendency to be expressed via semantically less-specified morphology, recruited from elsewhere in the grammar. While this is often true, for a language like Korean with clear honorific marking that is not related to any other part of the grammar, it is difficult to avoid positing a concrete grammatical feature like [HON], as

extends the analysis of the Korean subject honorific *-si* in Choi and Harley (2019). These authors argue that this honorific suffix, which we previewed in (2b) above, is the reflex of a morphological agreement relation between the verb, and a subject specified as [+HON]. We will argue that similarly, *che-* can be analyzed as the reflex of morphological agreement with a subject marked as [-HON]. However, we will see that there are differences in the distribution of *che-* and *-si*, which will require us to carefully define the morphological rule that triggers its appearance in the sentence.

1.2 Contents of the paper

In Section 2, we overview the behavior and meaning of *che-*. In Section 3, we describe our online experiment about the position of *che-* relative to the other prefixes in Korean. In Section 4, we introduce the basics of the syntactic theory we assume, and provide a preliminary analysis of the syntactic location of *che-*. In Section 5, we refine this analysis by considering in more detail the mechanisms that cause *che-* to arise in the sentence, extending the analysis of the Korean honorific *-si* in Choi and Harley (2019). In Section 6 we conclude, and mention two remaining puzzles about the distribution of *che-*.

2 Basic facts about *che-* and its meaning

We are aware of no previous linguistic work on *che-*, and very little descriptive documentation of it. The national institute of Korean language has an online dictionary containing an entry for the verb for ‘eat’ combined with *che-*, which it describes as a vulgar form.⁵ Since this morpheme is colloquial, and likely to cause offense, it is unsurprising that it has either been unnoticed or ignored by formal academic work. Therefore there is an opportunity to investigate this topic more closely.

As mentioned above, Korean grammar has a variety of ways of marking politeness/honorificity (Cho 2022), such as the subject-oriented honorific *-si* and hearer-directed politeness marker *-yo*. The proper use of such linguistic forms is essential in Korean, because Korean culture places great significance on displaying appropriate respect to individuals being spoken to or about, depending on their

we do here. See Choi and Harley (2019) for other work that explicitly depends on the presence of such a feature.

⁵ See https://stdict.korean.go.kr/search/searchView.do?word_no=533628&searchKeywordTo=3 (in Korean).

relative position within the culturally-determined social hierarchy. Not using the expected grammatical strategies for expressing politeness/honorificity is therefore interpreted as rude in Korean. While it is possible to communicate a lack of honorificity simply by not using honorific morphology that would otherwise be expected, we propose that *che-* constitutes an overt morphological expression of a lack of honorificity. For this reason, we have termed it “anti-honorific”. Specifically, this morpheme encodes a lack of honorificity with regard to the grammatical subject of a verb that carries it, which thus yields a sense of rudeness, crudeness, or vulgarity.⁶ This prefix can also be used with reference to the speaker, giving rise to a sense of self-deprecation. As a reviewer notes, this prefix can also convey a sense of intensity and/or sarcasm, though these properties do not seem distinct enough to classify them as separate functions, so we set them aside in this paper. Several additional examples demonstrating natural uses of anti-honorific *che-* are provided in (3):

(3) More examples of *che-*

- a. *Ya mwe hay, yaca sikan-ey kongpwu an-ha-ko*
 hey what do self.study time-LOC study NEG-do-and
che-ca-nya?
AHON-sleep-Q?
 ‘Hey, what are you doing? Are you not studying, but sleeping during self-study time?’
- b. *Taiethu ha-nta-mye mwe-l tto che-mek-e?*
 diet do-FUT-said what-ACC again **AHON**-eat-DECL
 ‘You said you were going to be on a diet, but what are you eating?’
- c. *Kyay-nun hakchangsicel-ey yenay-man che-ha-yss-nunteyto cikum*
 3SG-TOP studenthood-LOC dating-only **AHON**-do-PST-despite now
cal sal-a.
 well live-DECL
 ‘All s/he did when s/he was a student was date, but even so, s/he is doing well now.’

Example (2a) above showed that Korean also has an addressee-oriented honorific *-yo*. This can co-occur with *che-*. In these situations, the subject of the verb that carries *che-* is interpreted anti-honorifically, while at the same time, the presence of *-yo* indicates that the utterance is respectful with reference to the listener, as

⁶ Though the morpheme may not be interpreted as rude in casual conversations with socially close individuals, in which case considerations of social hierarchy are not relevant. This is the same context in which honorific/polite morphology is not used anyway, in languages like Korean and Japanese.

exemplified in (4) below. These examples are all appropriate in situations where one is speaking irreverently of the subject in question, while nevertheless being respectful towards a higher-status conversation partner:

(4) Co-occurrence of *che-* and *-yo*

- a. *Kyay-ka emcheng che-wul-myense na-hanthey*
 he/she-NOM lots AHON-cry-while me-DAT
sakwahay-ss-e-yo.
 apologize-PST-DECL-HON.
 ‘(S)he cried a lot while apologizing to me.’
- b. *Ton-ul kulehkey che-ssuteni mang-hay-ss-napwa-yo.*
 money-ACC like.that AHON-use get.screwed-do-PST-seem-HON.
 ‘(S)he got screwed due to using money like that.’

These examples are bi-clausal sentences in which *che-* and *-yo* occur on separate verbs, though the final *-yo* attached to the main verb takes scope over the entire utterance, adding hearer-oriented politeness to the entire sentence. Since the subject of the verb marked with *che-* and the hearer of the utterance are different individuals in these examples, there is no conflict between the politeness of *-yo* and the anti-honorificity of *che-*. This fact is important, because it makes clear that *che-* specifically encodes an anti-honorific attitude towards the subject, rather than adding a general sense of disrespectfulness to the entire utterance. Given that the irreverence conveyed by *che-* is specifically directed towards the subject, it is descriptively convenient to view it as the opposite of the subject-oriented honorific *-si* that we have already previewed above, shown once more in (5):

(5) Subject honorific suffix *-si*

- Pwucang-nim ecey swul manh-i tu-si-ess-e*
 boss/manager-HON yesterday alcohol much-ADV drink-HON-PST-DECL
 ‘My boss drank a lot yesterday.’

In order to be more precise about the meaning of *che-*, we turn to Culpeper (2011), who provides extensive discussion about what impoliteness is, and how it is manifested (see also Van Olmen et al., this volume). As Culpeper (2011: 20) overviews, there are numerous proposals about what exactly constitutes impoliteness. As he describes, a significant number of definitions involve one’s “face”, that is one’s self-image, reputation, or social identity, being aggravated or attacked. When an individual perceives their “face” to be intentionally under attack by a given communicative act, it is regarded as impolite. Culpeper (2011: 23) further qualifies that

impoliteness most fundamentally arises from the perception of a negative attitude, the details and intensity of which are mediated by contextual factors and social expectations. It is clear that the meaning of *che-* falls within these general descriptions. Under our proposal, *che-* encodes that the speaker is intentionally communicating a lack of honorificity towards the subject in question. In the context of Korean culture, for which one's place in the social hierarchy is an important consideration, this clearly constitutes an instance of face-attacking linguistic behavior. Unsurprisingly, the use of *che-* is highly likely to be interpreted as involving a negative attitude towards the subject, since if the subject were viewed positively, there is no reason why the expected degree of linguistic politeness would not have been used instead.

The meaning of *che-* thus encodes information about the speaker's irreverent attitude towards the subject. However, this meaning does not obviously encode at-issue descriptive information about the situation in question. As Kroeger (2022: 192) overviews, this is the case for honorifics in general, as well as a variety of other linguistic phenomena which appear to fundamentally be supplemental to, and thus logically separate from, the truth-conditional meaning of a sentence. Kroeger (2022) follows Gutzmann (2015) in referring to such meanings as *use-conditional*. As Kroeger (2022: 195) summarizes, use-conditional meanings have properties in common with presuppositions, but are distinct from them to an extent that deserves a separate classification (Potts 2005, Gutzmann 2015). Not being presuppositional, use-conditional meanings must presumably be learned as inherent to particular words/constructions (Kroeger 2022: 195; see also Van Olmen et al, this volume, and Mattiello, this volume). Thus, in summary, we can regard *che-* as a morpheme whose inherent semantics contains the use-conditional meaning that the speaker has an irreverent, “face-attacking” attitude towards the subject of the verb. In the remainder of this paper, we will consider in detail where exactly in the sentence structure *che-* arises, and why.

3 An experimental investigation of the distribution of *che-*

To investigate the possible positions for *che-*, we collected data using an online experiment through the participant recruitment platform Prolific (<https://www.prolific.co/>). This service allows researchers to set up experiments and solicit online participation of registered users, depending on the experiment's specified target audience. Our experiment was constructed for native Korean speakers, 16 of which completed it. The mean age of the participants was 34, and 13 of the 16 were female,

with the remaining 3 being male. The majority of the participants, 14 of 16, were born in Korea, while one was born in the United States, and one in Canada.

This experiment tested the possible positions for *che-* when combined with the other two verb-selecting prefixes in Korean: the negator *an-* and the negative ability modal *mos-*, each exemplified in isolation below.⁷

(6) Plain negative prefix⁸

Na-nun amuto an(i)-manna-ss-ta

I-TOP anyone NEG-meet-PST-DECL

‘I didn’t meet anyone.’

(7) Negative modal prefix

Na-nun amuto mos-manna-ss-ta

I-TOP anyone NEG.MOD-meet-PST-DECL

‘I couldn’t meet anyone.’

(adapted from Chung 2007b: 106)

The experiment gathered data on this by showing speakers context-setting examples with potential response examples that use some combination of prefixes, and asking participants to judge how acceptable a response would be, given the specified context. Judgments were solicited using a four point scale (1=worst, 4=best), which we later simplified by counting scores of 1 or 2 as negative judgments, and 3 to 4 as positive judgments. An example of a context-response pair that participants were asked to judge is provided in (8) below. In the original experiment, both the context and response were provided purely in Korean orthography, though we show transcribed glosses with translations here for the convenience of the reader.

7 Korean also has several other negative prefixes, historically borrowed from Chinese:

(i) a. **pul**-kanungha

‘impossible’

b. **pi**-kyoywukcek-i

‘uneducational’

(Chung 2007: 99)

Chung (2007b) argues that these prefixes are lexical phenomena that lack the productivity of the negators *an-* and *mos-*, which in contrast Chung argues represent actual syntactic positions in the clausal structure of Korean, as we assume in this paper.

8 In this example we see that the negative prefix can be realized as *an-* or *ani-*. As Chung (2007b: 1, footnote 1) describes, the former version is colloquial, while the latter is formal, and archaic for some speakers.

- (8) Context *Ne nac-tongan naynay phe ca-ss-ci?*
 you afternoon-during throughout PRT sleep-PST-Q?
 You slept through the whole afternoon, didn't you?
- Response *An-che-ca-ss-ketun!*
 NEG-AHON-sleep-PST-PRT
 No I did not!

The experiment tested all logical possibilities for combinations of the three prefixes, which are the following twelve: *an+mos*, *mos+an*, *an+che*, *che+an*, *mos+che*, *che+mos*, *an+mos+che*, *an+che+mos*, *mos+an+che*, *mos+che+an*, *che+an+mos*, *che+mos+an*. Each morpheme combination was presented in two sentences using intransitive verbs (“sleep”, “run”), and two using transitive verbs (“eat”, “use”). The 16 participants were evenly split into four groups, and the four participants of each group judged 16 sentences each, which differed for each group. The sentences were presented in randomized order. We focused our experiment on the combinations *an+che*, *che+an*, *mos+che*, and *che+mos*, which are the most important for our analysis. Each participant judged two sentences for each of these four combinations, and one sentence for each of the remaining 8 combinations. Those remaining 8 consist of examples that either use just *an-* and *mos-*, which are less relevant for our study due to not containing *che-*, or combinations of all three prefixes. Examples with three prefixes are difficult to judge, and receive low acceptability ratings in any case, so we did not emphasize these in the experiment. The result of this experimental design is that for the most important four combinations, we have a total of 32 judgments for each, whereas for the remaining combinations we have a total of 16 judgments. Ultimately, every participant did judge examples of all 12 types.⁹

The results are as follows. The majority of speakers, specifically 75%, accepted the order *an-che-*, while 59% accepted *mos-che-*, exemplified respectively in (9) and (10) below. Many speakers accepted both, since these are not complementary percentages, given that all participants judged all example types.

- (9) Negative prefix preceding *che-*
 Ttekpokki an-che-mek-ess-tako
 rice.cake NEG-AHON-eat-PST-EMPH
 ‘(I) couldn’t eat the rice cakes.’

⁹ The full set of survey examples and the resulting judgments are documented here: https://docs.google.com/document/d/1zSqfsl6BRSSStQYwPN_PVXCHolaN4xBjhS-eXus9syjs/edit?tab=t.0

(10) Negative modal prefix preceding *che-*

Ne onul elkwul-i way ilay? Ecey cam
 you today face-NOM why be.like? yesterday sleep
mos-che-ca-ss-nya?

NEG.MOD-AHON-sleep-PST-Q?

‘What’s wrong with your face today? Could you not sleep last night?’

Two orders that were not widely accepted were *che-an-*, which scored 31%, and *che-mos-*, which received a mere 13% acceptance. While the 13% for *che-mos-* is small enough to be potentially negligible, the 31% for *che-an-* appears more substantial. We comment on this in Section 6.1 below, after the main analysis is complete.

The judgments for combinations of negative prefixes are as follows. The order *mos-an-* received minimal acceptance (6%), while *an-mos-* is much more acceptable (38%), though clearly still degraded, likely due to the processing difficulty of double negation.

(11) Combining negative prefixes requires the order *an-mos-*

Ttekpokki an-mos-mek-ess-tako

Rice.cakes **NEG-NEG.MOD**-eat-PST-DECL-EMPH

‘I wasn’t unable to eat the rice cakes’

The only ordering of all three prefixes with any notable acceptability was *an-mos-che-* (13%), which is what we would expect given the preferability of the pairs *an-che-*, *mos-che-*, and *an-mos-*. However, a triple prefix verb may be too unusual or complex to receive high acceptability.

(12) Both negative prefixes combined with *che-*

Ttekpokki an-mos-che-mek-ess-tako

Rice.cakes **NEG-NEG.MOD-AHON**-eat-PST-EMPH

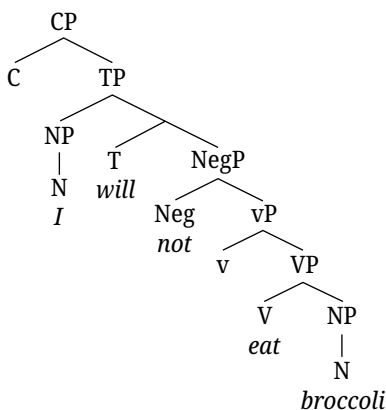
‘I wasn’t unable to eat the rice cakes’

In what follows, we focus on the most-accepted orders *an-che-* and *mos-che-*, which we use as the basis for our syntactic and morphological analysis of *che-*. We argue that the high acceptability of these prefix orders indicates that *che-* occupies a low syntactic position close to the verb. We make this argument with a preliminary syntactic analysis in Section 4, which we refine in Section 5, extending proposals from work on the morphology of Korean honorifics in Choi and Harley (2019).

4 A preliminary syntactic analysis of the position of *che-*

Here we provide a preliminary analysis of the syntactic position of *che-*. In order to do this, it is now necessary to be specific about the syntactic assumptions that we adopt. As mentioned above, we use a theory common in research on what is often termed “generative” or “minimalist” syntax, which will provide a concrete foundation for developing an explicit analysis of *che-*. A central hypothesis of this syntactic theory is that every word, and even most morphemes, instantiate a structural unit. These units are termed “phrases”. Each phrase has a word or morpheme that determines its grammatical category, which is termed the “head” of the phrase. Phrases are combined in a hierarchical structure, which binds them together into a complete sentence. We see an exemplifying syntactic diagram for an English sentence in (13) below. This example contains all phrasal types that will be required for the analysis. Notice that the word that acts as the head of the phrase bears an appropriate category label, and that category is passed up to the label that marks the top of the phrase. Thus a word which is a noun is marked “N”, short for “noun”, and this category is inherited by the label of its phrase, which in this case is NP for “noun phrase”. The core phrases here are the noun phrase (NP), verb phrase (VP), and tense phrase (TP). We also take negative sentences to contain a negation phrase (NegP), which is included in this diagram as well. These phrases are hierarchically arranged in such a way that the words that they contain have the correct left-to-right order.

(13) A structure for an English clause



Note that in this diagram, the subject NP is attached in the TP, and the object NP sits in the VP. This is a typical way of diagramming these constituents, which we will carry over to Korean, though this is not vital for the analysis of *che-*.

Two phrases in this diagram do not contain any word. Right above the VP, there is a phrase commonly termed the “little verb phrase”, or “vP”. This is phrase headed by a functional element, “little v”, which is widely argued to sit above the VP and perform an array of important grammatical functions that relate to the verb. Though this element is often not expressed by a dedicated word/morpheme, the existence of such a syntactic position is widely argued for, though the way it is labeled in syntactic analyses varies (Larson 1988, Hale and Keyser 1993, Chomsky 1995, Kratzer 1996, Harley and Noyer 2000). This position in the sentence structure will be important in Section 5 below. The exemplifying diagram in (13) above also contains a phrase “CP” which stands for “complementizer phrase”. This phrase contains no word in basic main clauses, but would be filled by the complementizer word “that” if this clause were a subordinate one. This phrase serves many syntactic functions, and will become relevant as our analysis of Korean progresses.

We now return to the analysis of Korean, building on the syntactic concepts just discussed. Based on the distributional data gathered in our experiment, here we provide a preliminary analysis of the syntactic position of *che-*, framing it in context with the rest of the Korean verbal complex. We base our syntactic analysis on Chung (2007a,b), who provides an analysis of Korean clause structure, including the negative prefixes we have seen in the previous section. Chung (2007a,b) argues that these negative prefixes in Korean head a Neg(ation)P which is structurally directly below TP, analogous to the position of negation in English, as the diagram in (13) above showed. Chung (2007b: 114) adopts a clause structure for Korean that is fundamentally very similar to the English clause structure shown above, though with several modifications. First, while heads of phrases in a language like English lean leftward, in languages like Korean, heads lean rightward. For this reason, languages like English are often referred to as “head-initial”, and languages like Korean as “head-final”. It is also necessary to account for the fact that in Korean, the verb and other functional elements in the clause are combined into one word, as demonstrated once more in example (14):

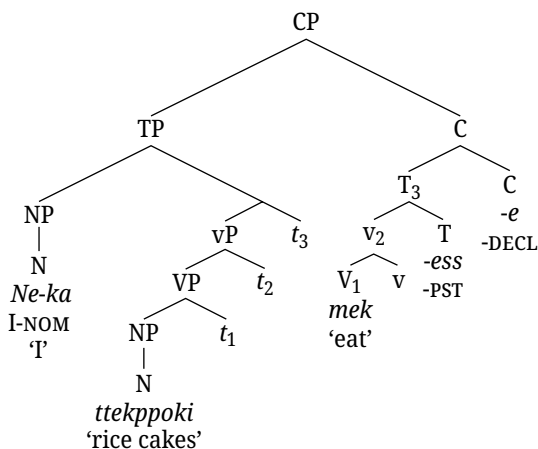
(14) A basic Korean sentence

Ne-ka ttekpokki mek-ess-e
 I-NOM rice.cakes eat-PST-DECL
 ‘I ate the rice cakes’

Chung (2007a–b) posits that all functional heads in the Korean clause are united into one word by movement. The movement of an element from one place to another is

a common theme in syntactic research, including movement of both phrases, and in certain situations, only the heads of phrases. Furthermore, much work argues that when two or more heads move together to form a single unit, the morphemes that express them form a single word (see for instance Embick 2010, Arregi and Nevins 2012, Bobaljik 2012). For Korean, Chung (2007a–b) argues that all functional heads in the clause move together to the C head, forming a single word in that position. This is demonstrated in the diagram in (15) below,¹⁰ which represents the Korean example just shown in (14) above. Since head movement occurs in a maximally local fashion (Travis 1984), this must be an incremental process, whereby V moves to little v, little v moves to T, and T moves to C. This process carries V and all intervening functional heads into C, uniting them into a single constituent, as (15) shows:

- (15) Diagram for example (14) above: Heads united into one word by movement¹¹



Since Korean is a head-final language (with the exception of the prefixes, which we address below), this movement process shifts the main verb rightward, along with all other functional heads between it and C. Note that when one head moves to

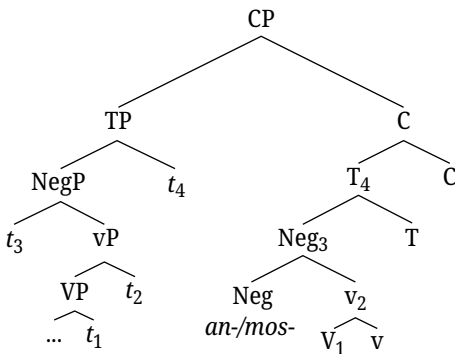
¹⁰ Here “t” is short for “trace”, which is a standard way of marking a position where a moved element used to be. The moved element, and the trace filling its previous position, are labeled with matching numerical indices so that the path of movement is unambiguously encoded.

¹¹ Korean has many sentence-final morphemes/particles that encode information about clause type, as well as various pragmatic effects. Here we have the informal declarative suffix *-e*, and many other examples will include the more formal (“dictionary”) declarative suffix *-ta*. Following the analysis of such morphemes in Choi and Harley (2019), we diagram them as occupying the C head.

another to form a complex head-unit, the category label of the head that was moved into is inherited by the new unit formed by that movement. This is why, in the diagram above, the labels for v, T, and C are duplicated. This is common in syntactic diagrams for a variety of reasons, but in this paper, it is necessary only to focus on the non-branching positions in the head complex, where words/morphemes attach.

This basic analysis of Korean clause structure makes it possible to discuss more precisely the nature of Korean prefixes. In Korean orthography, the negative prefixes *an-* and *mos-* are separated from the verb by a space (though *che-* is not). Contrary to what this orthographic convention implies, Chung (2007a–b) argues using facts about semantic scope and morphology that these prefixes are in fact structurally incorporated with the verb. Specifically, he argues that the head of NegP, where these prefixes are attached, also participates in the head movement process that derives the agglutinative Korean verbal complex, combining these heads into the single constituent which represents one complex morphological word. Chung (2007a,b) argues that these negative elements are prefixes rather than suffixes, unlike most morphemes in Korean, because the Neg head that represents them leans leftward, rather than rightward. Thus, following Chung (2007a,b), we consider *an-* and *mos-* to be morphemes attached at the head of NegP, which exceptionally leans to the left, as diagrammed in the schema in (16) below:¹²

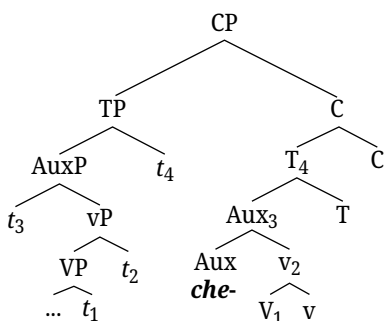
(16) Schema for a Korean verb with a negative prefix



¹² Considering *an-* and *mos-* to be instantiations of the same phrase NegP is likely an oversimplification, since as mentioned in Section 3, the two can co-occur. This is not surprising, since they are semantically different: *an-* is plain negation, while *mos-* expresses negative (ability) modality. Therefore it is possible that *an-* heads a NegP, while *mos-* occupies a higher functional position relating to modality. This possibility does not have any important relation to our main analysis, so we will proceed by simply considering *an-* and *mos-* to represent different versions of a single phrase NegP.

This form of analysis can also be extended to *che-*. Preliminarily, we propose that *che-* is the head of a phrase that is above vP, which leans left, thus making *che-* a prefix in the surface order. We thus temporarily diagram *che-* as the head of an Aux(iliary)P, but we will refine this analysis and eliminate AuxP in the next section. Tentatively using this AuxP for convenience, and following the analysis of Chung (2007a–b) for Korean negative prefixes, we assume that the Aux which *che-* expresses is carried along into C by head movement of V, and that this head leans leftward just as negative prefixes do. This preliminary analysis is shown in the schema in (17) below:

(17) Structure for a verb with *che-*



In the previous section, we saw that *che-* most productively appears between the verb and any instance of *an-* or *mos-*, when present, as demonstrated again in (18)–(19):

(18) Negative prefix preceding *che-*

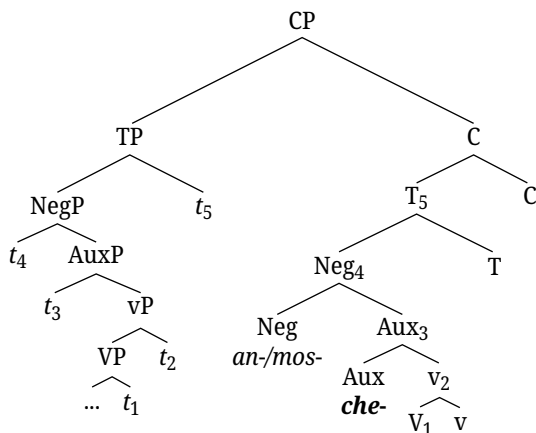
Ttekpokki an-che-mek-ess-tako
 rice.cake NEG-AHON-eat-PST-EMPH
 ‘(I) didn’t eat the rice cakes.’

(19) Negative modal prefix preceding *che-*

Ne onul elkwul-i way ilay? Ecey cam
 you today face-NOM why be.like? yesterday sleep
mos-che-ca-ss-nya?
 NEG.MOD-AHON-sleep-PST-Q?
 ‘What’s wrong with your face today? Could you not sleep last night?’

If *an-* and *mos-* represent NegP, the fact that these prefixes must precede *che-* is an automatic prediction of our hypothesis that *che-* sits in a syntactic position directly above vP. This is demonstrated in the schema in (20), which combines the negative prefixes with *che-*:

(20) Korean verbal complex including negative prefix and *che-*



This structural analysis matches the facts. We refine this analysis in the next section, in which we eliminate AuxP in order to account for the presence of *che-* in a way that better fits with previous research on Korean morpho-syntax.

Before turning to the next section, we address an alternative hypothesis that *che-* might be an adverb, rather than the head of a syntactic phrase. The default position for adverbs in Korean is preceding the verb, either preceding or following an object if present:

(21) Some adverb positions in Korean

a. Adverb Object Verb

Kuphakey susi-lul mek-ess-e
quickly sushi-ACC eat-PST-DECL
 ‘(I/(s)he. . .) quickly ate sushi’

b. Object Adverb Verb

Susi-lul **kuphakey** mek-ess-e
 sushi-ACC **quickly** eat-PST-DECL
 ‘(I/(s)he. . .) quickly ate sushi’

Therefore adverbs and *che-* are superficially similar in that they precede the verb. However, *che-* and adverbs do not have the exact same distribution. While it is possible for an adverb to precede the object as in (21a), this order is impossible for *che-*, as (22a) below shows. The only possibility is for *che-* to be adjacent to the verb, as in (22b):

(22) *Che-* does not pattern like an adverb

- a. **Che susi-lul mek-ess-e*
 AHON sushi-ACC eat-PST-DECL
 ‘(I/(s)he. . .) ate sushi’
- b. *Susi-lul che-mek-ess-e*
 sushi-ACC **AHON**-eat-PST-DECL
 ‘(I/(s)he. . .) ate sushi’

These facts are consistent with the hypothesis that *che-* is indeed a prefix of the verb, rather than an adverbial phrase.

5 Refining the analysis: On *che-* versus *-si* and (anti)honorificity as agreement

Above, we have proposed that *che-* heads a phrase directly above the vP and below NegP. The subject honorific *-si* appears to have a similar hierarchical position, though this morpheme is a suffix rather than a prefix. This becomes evident by using another strategy for negation in Korean, which involves post-verbal negation of the form *V-ci an-hay*. Here there is the converbial marker *-ci* suffixed to the verb, followed by a supporting verb *hay* (‘do’), to which the negative prefix attaches.¹³ (Note that the Korean ‘do’ alternates between *hay* and *ha* for morpho-phonological reasons.) Importantly, when *-si* is present in such a structure, it is common for it to attach directly to the verb, preceding any of the morphemes involved in this negation strategy. We discuss the properties of this construction further later in this section, but as a preliminary analysis, it is consistent with the facts to posit that *-si* is hosted by a phrase right above the vP. For convenience we can label this AuxP as we did for *che-*, as illustrated in (23) below, which shows such an example sentence along with a corresponding syntactic diagram:

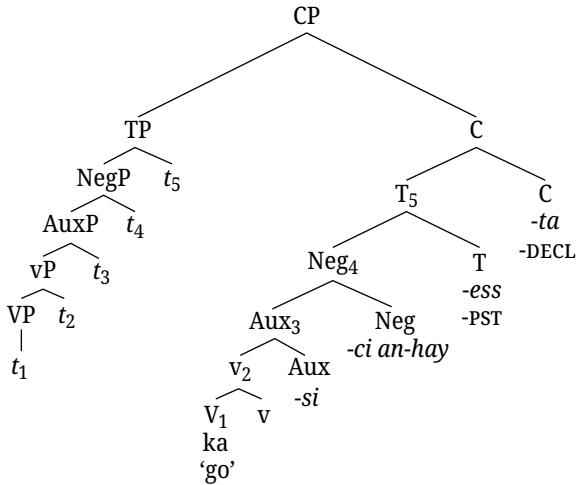
¹³ In colloquial speech, *V-ci an-hay* is contracted to *V-ci anh-*. See further Chung (2007b: 97, footnote 1).

(23) Honorific suffix structurally beneath negation

- a. *Halapeci-kkeyse ka-si-ci an(i)-hay-ess-ta.*
 grandfather-NOM.HON go-HON-CVB NEG-do-PST-DECL
 ‘Grandfather didn’t go.’

(Choi and Harley 2019: 1333)

b.



Thus both *che-* and *-si* appear to be structurally below negation. However, there is a complication in the distribution of *-si* which shows that these morphemes cannot be considered totally structurally analogous.

This becomes evident by further investigating the negative construction we just introduced. While in this configuration *-si* can indeed appear directly next to the verb as in the above example, there are two other options (Yi 1994). The honorific suffix can instead attach to the support verb *hay* ‘do’ as in (24a) below, or to both the main verb and the extra supporting ‘do’ verb at the same time, as in (24b) below:

(24) More options for the honorific suffix with negation

- a. Honorific on supporting verb

Halapeci-kkeyse ka-ci an(i)-ha-si-ess-ta.
 grandfather-NOM.HON go-CVB NEG-do-HON-PST-DECL
 ‘Grandfather didn’t go.’

(Choi and Harley 2019: 1333)

- b. Honorific on main verb and support verb

Halapeci-kkeyse ka-si-ci an(i)-ha-si-ess-ta.
 Grandfather-NOM.HON go-HON-CVB NEG-do-HON-PST-DECL
 ‘Grandfather didn’t go.’

(Choi and Harley 2019: 1333)

We will show that the distribution of *che-* is more restricted than this. Before discussing *che-* in further detail, though, next we summarize the analysis from Choi and Harley (2019) for the facts about *-si*, which we will extend to account for *che-*.

In brief, Choi and Harley (2019) argue that *-si* is not hosted by a dedicated syntactic position in the clause (such as the AuxP we invoked for convenience above), but rather is inserted by the morphology through what is essentially an agreement relation, between an honorifically-interpreted subject and little *v*. We will argue for a similar proposal for *che-*, thus eliminating the AuxP from the analysis, and instead appealing to a morphological mechanism.

5.1 The morphological process that yields *-si*

Choi and Harley (2019) build on previous works arguing that *-si* is syntactically-sensitive. In particular, they take it to fundamentally be an agreement marker.¹⁴ Instead of being a morpheme that encodes the presence of a subject with certain case/number/gender features, which are frequently found in Indo-European languages, they take this to be a morpheme that registers the presence of a subject bearing the feature [+HON]. Choi and Harley (2019) frame this analysis within the Distributed Morphology theory (Halle and Marantz 1993), which argues that morphology is only determined after the syntactic component of grammar has finished completing the structure of the sentence. The details of this theory are not vital for the present analysis of *che-*. However, the reason why Distributed Morphology is relevant for Choi and Harley (2019) is because this theory allows the morphological part of the grammar to augment a syntactic structure, in certain limited ways, when required for morphological well-formedness (see Harley and Noyer 1999 for an overview). Importantly, Choi and Harley (2019) hypothesize a mechanism of this sort for Korean which they term “HON-sprouting”, defined in (25) below, which they argue is responsible for the arising of *-si*:

- (25) [HON] sprouting rule for *-si*¹⁵
 $v \rightarrow [v \text{ Hon}] / [\text{NP}_{[+HON]} \dots [\dots __ \dots]]$
 (Choi and Harley 2019: 1336)

The format of this rule is that on the left side of the slash (“/”), it defines the operation that is triggered, while on the right side of the slash, it states the grammatical context

¹⁴ For arguments about politeness marking as agreement, see Miyagawa (2017).

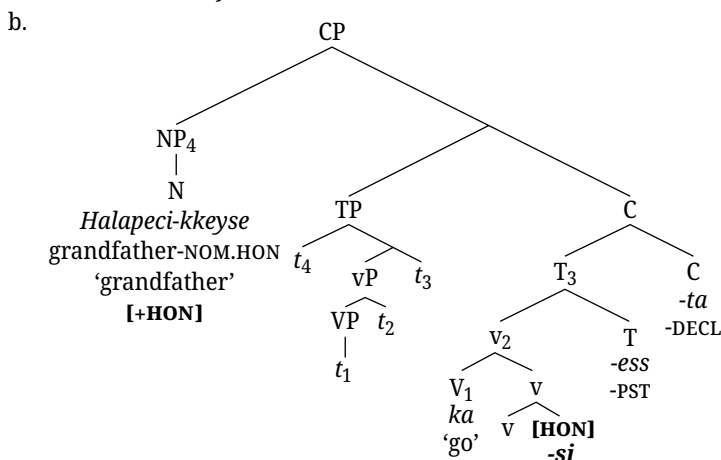
¹⁵ Choi and Harley (2019: 1334, footnote 15) argue that this rule must indeed apply to *v* and not *V*, since the honorific marker is separated from *V* by material that overtly expresses *v*, such as causative morphology, when present.

in which that operation occurs, with the blank space (“_”) marking the position that the affected element must be in. Specifically, then, what this rule defines is that a [HON] feature must be attached to the right of the little *v* head ($v \rightarrow [v \text{ Hon}]$) when it is in a syntactic position below an NP that itself bears a [+HON] feature (/ [NP_[+HON] . . . [. . . _ . . .]]). The [+HON] feature on the NP is the only meaningful one, since it is the NP in question that is understood to be honorific, while the [HON] inserted in little *v* is essentially agreement, with no semantic contribution. This extra [HON] does have a morphological effect, however, in that it is expressed as *-si*, as Choi and Harley (2019) argue.

In basic positive sentences, this morphological rule applies once, after all heads move together to C in the syntax, as exemplified in (26) below. For theory-internal reasons Choi and Harley (2019) must assume that the subject moves into the CP from the usual subject position in TP, as shown here. We assume this as well for consistency with Choi and Harley (2019), though it is not vital for the present analysis. Note that in this diagram, Choi and Harley (2019) must add another layer within the little *v* head, in order to explicitly diagram the insertion of [HON] in it:

(26) Morphological sprouting of HON in a basic positive sentence¹⁶

- a. *Halapeci-kkeyse* *cip-ey* *ka-si-ess-ta*.
 grandfather-NOM.HON home-DAT go-HON-PST-DECL
 ‘Grandfather went home.’
 (Choi and Harley ex. 6a)



(Adapted from Choi and Harley 2019: 1337)

¹⁶ Choi and Harley (2019) adopt a hypothesis common in work using Distributed Morphology: that lexical heads like V, N, and so on, in fact originate as un-categorized roots, notated as “*v*”, which gain their category when merged with a separate category-defining head. In this context, a given

This process also applies in negative sentences. We have seen two strategies for verbal negation in Korean: the prefixation of *an-* to the main verb as in (6) above, and the use of a more complex construction containing a support verb meaning ‘do’ to which *an-* prefixes instead, as in (23) above. In examples of the latter type, importantly, *-si* can in fact suffix to either the main verb, the supporting “do”, or both, as (23)–(24) above demonstrated. Choi and Harley (2019) argue that this is possible because the supporting *hay* (‘do’) in such constructions is an extra instance of little *v*, to which the morphological rule in (25) above can also apply.

Before demonstrating the implementation of this analysis, we must discuss why this extra “do” appears. In English, there are a variety of situations, including the use of negation, that motivate the insertion of a supporting “do” in certain tenses:

(27) English ‘do’-support

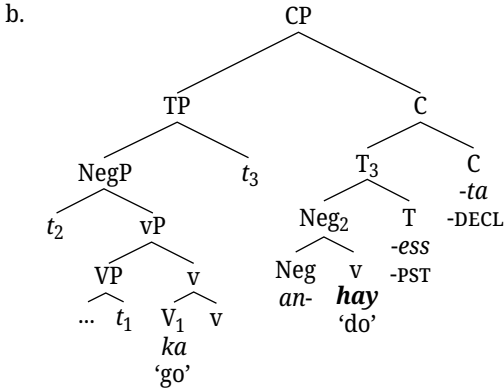
- a. *The cat likes meat.*
- b. *The cat **does not** like meat.*

Choi and Harley (2019) follow the analysis of this phenomenon in Bobaljik (1994), who argues that when *T* is not sufficiently local to a *V* that can realize its morphology, it is supported by an extra instance of *v*, which is realized as “do”. One of the elements that can disrupt locality between *T* and *V* in English is negation, motivating “do”-support as in (27) above. Choi and Harley (2019) adopt the same general idea for Korean. Specifically, they argue that in Korean “do”-support constructions with negation, *V* moves to little *v*, but little *v* cannot move onward to *Neg*. Rather, *Neg* moves alone to *T* and *C*. In this case, *V* and *v* do not reach *C*. However, the negative prefix *an-*, which expresses the *Neg* head, nevertheless requires a verbal element to support it. For this reason, an additional instance of *hay* (“do”) is inserted by the morphological component of the grammar. This “do” is taken to be an instance of *v*, as mentioned. This fits with the fact that it does not carry the semantics of a fully fledged lexical verb, though it is nevertheless grammatically verbal. Such an example is illustrated below in (28a), and the corresponding structure for its verbal morphological complex is shown in (28b). Choi and Harley (2019) remain neutral about the exact identity of the converbial marker *-ci* that appears in these constructions, and therefore omit it from their diagrams. We will do the same.

root functions as a *V* when selected by *v*. The hypothesis of category-neutral lexical roots is not relevant to the topics we discuss in this paper, so for simplicity, we replace all instances of “√” from Choi and Harley (2019) with *V*.

(28) Do-support with negation in Korean

- a. *Inho-ka ka-ci an(i)-hay-ess-ta.*
 Inho-NOM go-CVB NEG-do-PST-DECL
 'Inho didn't go'



(Adapted from Choi and Harley 2019: 1332)

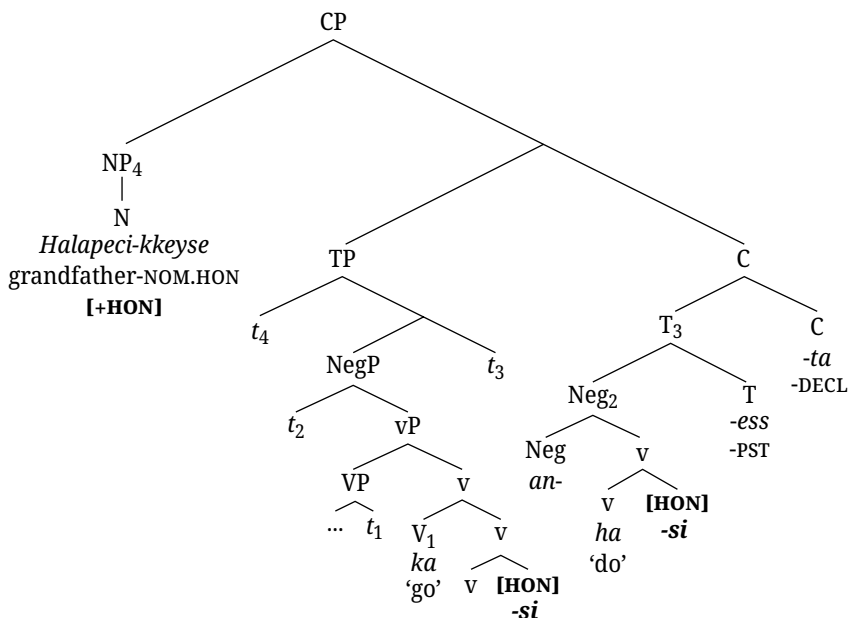
Notice that under this analysis, these negation constructions contain two instances of *v*: one that originates directly above the actual main verb, and another introduced by “do”-support at Neg. Therefore there are two positions where the [HON]-sprouting rule defined in (25) above can apply. In (29) below we diagram situations where [HON]-sprouting applies at both instances of *v*, as example (24b) above demonstrated. It is also permitted for this rule to apply at only one of them, as in examples like (23a) and (24a) above.¹⁷ Importantly, Choi and Harley (2019) argue that the possibility of having multiple simultaneous instances of *-si* indicates that this morpheme should not be analyzed as expressing a specific syntactic head in Korean, but is better understood as the result of a morphological rule. Such a rule can indeed be triggered more than once if a structure contains multiple positions in which the rule is applicable. This analysis is illustrated for an example with two instances of *-si* in (29) below:

(29) Honorific -si on main verb and support verb

- a. *Halapeci-kkeyse ka-si-ci an(i)-ha-si-ess-ta.*
 Grandfather-NOM.HON go-HON-CVB NEG-do-HON-PST-DECL
 'Grandfather didn't go.'
 (Choi and Harley 2019: 1333)

¹⁷ See Choi and Harley (2019: 1340) for discussion about the optionality in this pattern.

b.



5.2 An analogous analysis for *che-*

We argue that *che-* can be understood as arising from a mechanism very similar to what Choi and Harley (2019) posit for *-si*. We thus dispense with the preliminary convenience of diagramming *che-* as the head of an AuxP that sits above vP, and instead argue that this morpheme can be understood as arising from, in essence, an agreement operation achieved in morphology.

There is an additional relevant fact about *che-* which we have not mentioned. We have seen that in negative constructions in Korean that use “do”-support, *-si* can be inserted at either the v above the main verb, or at the v that instantiates the supporting “do”. If *-si* and *che-* were fully analogous, we would expect *che-* to also be able to be inserted at either of these positions. This is not the case. It appears that *che-* can only prefix to the main verb, as (30a) below shows, and not to a negation-supporting “do” (30b), or to both of them simultaneously (30c):

(30) *Che-* only appears on the main verb¹⁸a. Prefixation to main verb¹⁹

Manhi che-mek-ci an(i)-hay-ss-e

a.lot AHON-eat-CVB NEG-do-PST-DECL

‘(I/(s)he/etc.) did not eat a lot’ [Same translation for the following two examples]

b. No prefixation to the supporting “do”²⁰

**Manhi mek-ci an-che-hay-ss-e*

a.lot eat-CVB NEG-AHON-do-PST-DECL

c. No prefixation to both

**Manhi che-mek-ci an-che-hay-ss-e.*

a.lot AHON-eat-CVB NEG-AHON-do-PST-DECL

To capture the more restricted distribution of *che-*, we argue that it arises from a morphological rule slightly different from that for *-si-*, defined in (31) below. While the rule in (25) above posited by Choi and Harley (2019) inserts a feature [HON] which is expressed by *-si-*, the rule in (31) below instead inserts a negatively valued honorific feature, namely [-HON], which we argue is expressed by *che-*. Furthermore, to capture the restricted distribution of *che-*, we hypothesize that this rule only applies to the main V, not little *v*. Since *che-* is a prefix rather than a suffix, we also define the added feature as being attached to the left of V, as (31) shows. In summary, this rule has the same format as the one in (25) above, but this additional rule for *che-* is defined to add a [-HON] feature on the left side of V ($V \rightarrow [-\text{Hon } V]$) when V is structurally beneath an anti-honorifically construed NP ($[\text{NP}_{[-\text{HON}]} \dots [\dots _ \dots]]$).

(31) [-HON] sprouting rule for *che-*

$V \rightarrow [-\text{Hon } V] / [\text{NP}_{[-\text{HON}]} \dots [\dots _ \dots]]$

Since this rule only applies to V, it does not predict the possibility of inserting multiple instances of *che-* in negative constructions with “do”-support. Following Choi

¹⁸ These judgments are provided by native speaker Hyewon Jang.

¹⁹ This example is most natural if it undergoes the phonological simplification mentioned in footnote 13 above, which would result in the string *Manh-i che-mek-ci anh-ass-e*.

²⁰ In Section 2, we mentioned that the prefix order *che-an-* had marginal acceptability. For exhaustiveness, we show that using this order does not improve an example like (30b):

i. **Manhi mek-ci che-an-hay-ss-e*

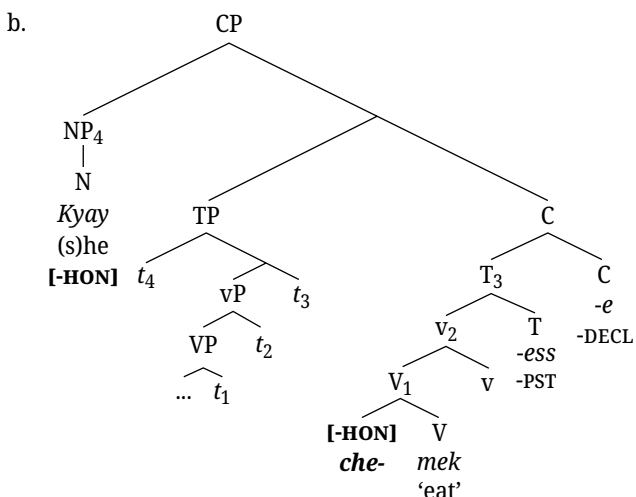
a.lot eat-CVB AHON-NEG-do-PST-DECL

‘(I/(s)he/etc.) didn’t eat a lot’

and Harley (2019), such constructions have multiple instances of little *v* at which the insertion of *-si* can occur, but nevertheless such constructions contain only one main verb *V*. Thus given the rule in (31) above, we correctly predict the possibility of only one *che-* per clause. The successful application of this rule in a basic positive sentence is demonstrated in (32) below:

(32) Korean clause with *che-* under the sprouting/agreement analysis

- a. *Kyay emcheng che-mek-ess-e*
 (s)he much AHON-eat-PST-DECL
 '(S)he ate a lot.' (= ex. 1 above)



We assume that the [-HON] feature on the subject NP is the semantically meaningful one, since this NP is what is construed anti-honorifically, whereas the additional instance of [-HON] expressed as *che-* is a meaningless reflex inserted in the morphology. Since *che-* does not have the positional variation in negative constructions with “do”-support that *-si* displays, we do not show diagrams for *che-* in such constructions. Whether or not “do”-support occurs in the higher part of a clause, *che-* insertion only ever applies at *V*.

This analysis avoids positing a dedicated phrase in the clausal spine that is specifically used to host anti-honorific morphology. This would be ad hoc and difficult to verify on any independent grounds, especially given the fact that we are aware of little to no cross-linguistic parallel with this Korean phenomenon. Analyzing this morpheme as the reflex of a Korean-specific morphological rule, in contrast, avoids these potential concerns. Much research since Chomsky (1995), including a great deal of research in Distributed Morphology, has made productive use of the hypoth-

esis that while syntactic structures are relatively stable cross-linguistically, variation is primarily found in the morpho-phonological components of grammar. Since anti-honorific morphology is indeed far from universal, it is theoretically appealing to understand *che-* as the result of a Korean-specific morphological rule, as Choi and Harley (2019) do for *-si*. Thus we have adopted a similar rule for *che-*, though as stated above, the rule must be defined differently to capture this morpheme's more restricted distribution.

5.2.1 The co-occurrence of *che-* and *-si*

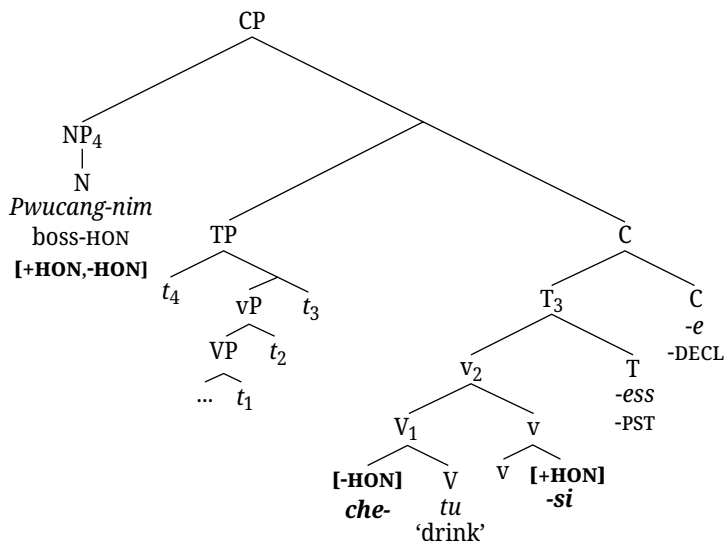
Since *che-* and *-si* have opposite meanings, it would be unsurprising if they could not co-occur. However, they can coexist on the same verb. In this case, there is no contradiction, but rather this combination is interpreted as sarcastic, insincere honorification, as (33) demonstrates.²¹

(33) Combination of *che-* and *-si* on one verb

- a. *Pwucang-nim ecey swul manh-i*
 boss/manager-HON yesterday alcohol much-ADV
che-tu-si-ess-e
 AHON-drink-HON-PST-DECL
 'My boss drank a lot yesterday.'
- b. *Che-cwumwu-si-ko nacwungey yayki-hay.*
 AHON-sleep-HON-and later talk-do
 'Go to sleep, and lets talk later'

Since *che-* and *-si* are attached in different positions, V versus little v, from the perspective of morpho-syntax it is expected that they should be able to co-occur. Given what we have said above, we must assume that the subject in such situations bears both the features [+HON] and [-HON], which each separately trigger the sprouting rules defined in (25) and (31) above.

21 These judgments are provided by native speaker Hyewon Jang.

(34) Co-occurrence of *che-* and *-si* (based on example 33a above)

Since these features are clearly semantically conflicting, it is unexpected that the presence of both of them on a single subject would actually be interpretable. We suggest that pragmatics is able to resolve this state of affairs as sarcastic honorification rather than a total contradiction, but we must leave the details of such an analysis to other work. In summary, while the differing attachment points of these morphemes explain why this morpho-syntactic structure is grammatical in the first place, this does not clarify the pragmatic mechanisms that allow such sentences to be interpretable in actual use. This represents a topic for future research.

6 Conclusion and further questions

Here we have analyzed several aspects of the Korean prefix *che-*, which as far as we know has been unnoticed by linguistic research so far. We have argued that this is an anti-honorific morpheme, which expresses an irreverent attitude towards the subject of the verb, interestingly in contrast to the many other politeness-encoding morphemes in Korean.²² Extending the analysis of Choi and Harley (2019) for the

²² We are aware of another potential anti-honorific morpheme in Korean, *-ssa*, which is mainly used by older speakers (i). We have not had the opportunity to examine this morpheme in detail,

subject-oriented honorific *-si*, we hypothesized that *che-* arises from a morphological agreement process, which attaches this prefix to V when structurally beneath a subject specified as [-HON]. This analysis avoids positing a specific position in the underlying sentence structure that is responsible for hosting anti-honorific morphology, which would be questionable on at least theoretical and methodological grounds. If this morphologically-oriented analysis of *che-* is correct, this morpheme constitutes another possible case of (im)politeness marking as grammatical agreement (Miyagawa 2017).²³ Below we mention two remaining questions for future research.

6.1 Another potential position for *che-*

As mentioned in Section 2, our experiment found marginal acceptance (31%) of the prefix order *che-an-*. Though such examples are evidently degraded, they pose a challenge to the above analysis. Since in examples of this variety *che-* does not prefix directly to the verb, this configuration does not fit the definition of the rule we hypothesized in (31) above. We could describe such examples by positing a second version of the rule, which sprouts [-HON] off Neg rather than V, as defined in (35) below. This rule has exactly the same format as those we've seen so far in (25) and (31), but differs from (31) simply in sprouting [-HON] to the left of the Neg head, rather than V:

- (35) [-HON] sprouting rule, alternative version
 Neg → [-Hon Neg] / [NP_[-HON] . . . [. . . __ . . .]]

The application of this rule would correctly derive examples like the following:

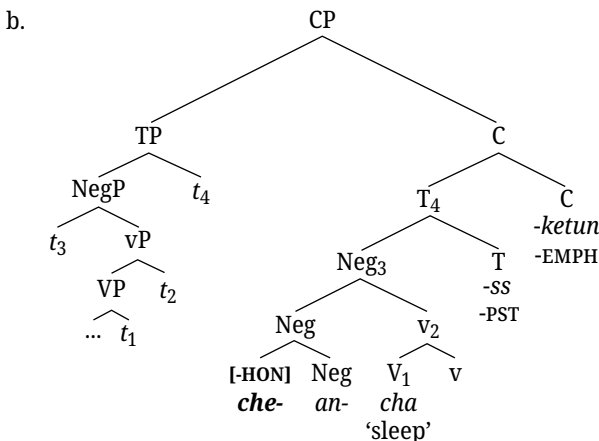
- (36) Marginally acceptable *che-an-*

but we mention it as a topic for future research:

- i. Mwel kulehkey wul-e-ssa?
 What like.that cry-DECL-**AHON**
 'What are you crying about?'

²³ This analysis of *che-*, in combination with the proposals from Choi and Harley (2019) about *-si*, also represents further evidence for the distinction between V and little v. This is because it is necessary to have at least two different verbal heads which can be separately targeted by the morphological rules that give rise to these morphemes, in order to account for their distributional differences.

- a. (na) **che-an-ca-ss-ketun**
 (I) **AHON-NEG-sleep-PST-EMPH**
 ‘I did not sleep’



If *che-* is a new development in Korean, it is possible that the definition of its corresponding rule is to an extent under-determined at this point in time.²⁴ Furthermore, it is conceivable that there might be other differences between more typical *an-che-* examples, and those using the marginal *che-an-*, which would provide a basis for analyzing this state of affairs. However, we are not currently aware of any such informative differences. We leave this puzzle for future research.

6.2 A puzzle about adjectives

In Korean, property-denoting predicates that in English would generally be translated as adjectives are morpho-syntactically verb-like, since they conjugate/infect in ways that are very similar and often even identical to verbs. In many cases, though not always, such predicates even contain a semantically empty light verb “do”, as we see in (37b) below. As (37) shows, such property-denoting predicates can host the honorific *-si*, just as verbs do:

²⁴ Gyu-Hwan Lee (p.c.) suggests that this morpheme is likely a recent innovation, though it is beyond the scope of this paper to verify this. Tracing the history of the morpheme represents an opportunity for future work.

(37) Subject honorific with property-denoting predicates

- a. *Ku pan sensayng-nim-i yeppu-si-e*
 that class teacher-HON-NOM pretty-**HON**-DECL
 ‘The teacher of that class is pretty.’
- b. *Ku pwun-un swuncin-ha-si-e.*
 that person-**TOP** naive-do-**HON**-DECL
 ‘That person is naive.’

However, these predicates do not naturally combine with *che-*, as (38) below shows, using exactly the same predicates that (37) above contains:²⁵

(38) No or marginal use of anti-honorific with property-denoting predicates

- a. **Ku yeca che-yepp-e*
 that woman **AHON**-pretty-DECL
 ‘That woman is pretty.’
- b. *Ku salam ??che-swuncin-(*che)-hay.*
 that person **AHON**-naive-(**AHON**)-do
 ‘That person is naive.’

While these predicates in Korean do not use any obvious copular verb in the way that English does, example (37b) does contain an instance of “do”, at least morphologically. Nevertheless use of *-si* is permitted in both (37b), as well as (37a), which lacks anything obviously verbal. If these predicates are both verb-like enough to allow use of *-si*, it is not clear why use of *-che* should not also be permitted for them, as (38) shows. We leave this puzzle for future consideration.

Abbreviations

ACC	accusative
AHON	anti-honorific
CVB	converb
DAT	dative
DECL	declarative
EMPH	emphasis
FUT	future
HON	honorific
LOC	locative

²⁵ The judgments in (37) and (38) are provided by native speaker Hyewon Jang.

MOD	modal
NEG	negation
NOM	nominative
PRES	present
PRT	particle
PST	past
Q	question
TOP	topic

Acknowledgments: For feedback on various instantiations of this work, thanks to Rajesh Bhatt, Pavel Caha, Eunsun Jou, Gyu-Hwan Lee, Jeffrey Parrot, Yasu Sudo, and the participants of the 2023 International Circle of Korean Linguistics, the 25th Seoul International Conference on Generative Grammar, the 16th Conference on Syntax, Phonology and Language Analysis, as well as the organizers of the Grammar of Impoliteness workshop at the 56th Annual Meeting of the Societas Linguistica Europaea, especially Riccardo Giomi.

References

- Adger, David. 2003. *Core Syntax: A Minimalist Approach*. Oxford: Oxford University Press.
- Alok, Deepak & Bill Haddican. 2022. The formal heterogeneity of allocutivity. *Glossa: a Journal of General Linguistics* 7. 1–41.
- Arregi, Karlos & Andrew Nevins. 2012. *Morphotactics: Basque Auxiliaries and the Structure of Spellout*. Dordrecht: Springer.
- Baker, Mark. 1985. The Mirror Principle and morphosyntactic explanation. *Linguistic Inquiry* 16. 373–415.
- Bobaljik, Jonathan. 1994. What does adjacency do? *The Morphology-Syntax Connection* 32. 1–32.
- Bobaljik, Jonathan. 2012. *Universals in Comparative Morphology: Suppletion, Superlatives, and the Structure of Words*. Cambridge, MA: MIT Press.
- Carnie, Andrew. 2021. *Syntax: A Generative Introduction*. 4th edn. Chichester: Wiley-Blackwell.
- Cho, Sungdai. 2022. Politeness strategies in Korean. In Sungdai Cho & John Whitman (eds.), *The Cambridge Handbook of Korean Linguistics*, 105–132. Cambridge, UK: Cambridge University Press.
- Choi, Jaehoon & Heidi Harley. 2019. Locality domains and morphological rules: Phases, heads, node-sprouting and suppletion in Korean honorification. *Natural Language and Linguistic Theory* 37. 1319–1365.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Culpeper, Jonathan. 2011. *Impoliteness*. Cambridge, UK: Cambridge University Press.
- Chung, Inkie. 2007a. *Ecology of PF: A Study of Korean Phonology and Morphology in a Derivational Approach*. University of Connecticut doctoral dissertation.
- Chung, Inkie. 2007b. Suppletive negation in Korean and distributed morphology. *Lingua* 117. 95–148.
- Embick, David. 2010. *Localism versus Globalism in Morphology and Phonology*. Cambridge, MA: MIT Press.

- Gutzmann, Daniel. 2015. *Use-conditional Meaning: Studies in Multidimensional Semantics*. Oxford: Oxford University Press.
- Hale, Ken & Samuel Jay Keyser. 1993. On argument structure and the lexical representation of syntactic relations. In Ken Hale & Samuel Jay Keyser (eds.), *The View from Building 20*, 53–109. Cambridge, MA: MIT Press.
- Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Ken Hale & Samuel Jay Keyser (eds.), *The View from Building 20*, 1–52. Cambridge, MA: MIT Press.
- Harley, Heidi & Rolf Noyer. 1999. State-of-the-article: Distributed morphology. *GLOT International* 4(4). 3–9.
- Harley, Heidi & Rolf Noyer. 2000. Formal versus encyclopedic properties of vocabulary: evidence from nominalizations. In Bert Peeters (ed.), *The Lexicon-Encyclopedia Interface*, 349–374. Amsterdam: Elsevier.
- Kaiser, Stefan, Yasuko Ichikawa, Noriko Kobayashi & Hirofumi Yamamoto. 2001. *Japanese - A Comprehensive Grammar*. London, UK: Routledge.
- Kratzer, Angelika. 1996. Severing the external argument from its verb. In Johan Rooryck & Laurie Zaring (eds.), *Phrase Structure and the Lexicon*, 109–137. Dordrecht: Kluwer.
- Kroeger, Paul. 2022. *Analyzing Meaning: An Introduction to Semantics and Pragmatics*. 3rd edn. Berlin: Language Science Press.
- Larson, Richard. 1988. On the double object construction. *Linguistic Inquiry* 19. 335–391.
- Miyagawa, Shigeru. 2017. *Agreement beyond Phi*. Cambridge, MA: MIT Press.
- Potts, Christopher. 2005. *The Logic of Conventional Implicatures*. Oxford: Oxford University Press.
- Travis, Lisa. 1984. *Parameters and Effects of Word Order Variation*. MIT doctoral dissertation.
- Wang, Ruonan. 2023. Honorifics without [hon]. *Natural Language and Linguistic Theory* 41. 1287–1347.
- Yamada, Akita. 2019. *The Syntax, Semantics and Pragmatics of Japanese Addressee-honorific Markers*. Georgetown University doctoral dissertation.
- Yi, Eun-Young. 1994. NegP in Korean. In Howard Grabois, David Parkinson & Deborah Yeager (eds.), *Papers on negation*, 193–208. Ithaca: CLC Publications.

Nina Dobrushina

4 Blessings and curses are structurally different: Data from Daghestan

Abstract: The paper examines the expression of curses – one of the main ways of conveying impoliteness in Daghestan (North Caucasus). Blessings and curses, an integral part of communication in Daghestan, provide a case to look for grammatical and constructional distinctions between wishes with positive value (politeness) and wishes with negative value (impoliteness). The comparison of the morphological, syntactic and semantic properties of blessings and curses in Nakh-Daghestanian languages contributes to the discussion of whether impoliteness can be inherently associated with linguistic structures or is primarily determined by context. Data from three Nakh-Daghestanian languages are considered – Avar (Avar-Ando-Tsezic), Lak (Lak) and Rutul (Lezgian). The study reveals several structural features of curses. First, there are morphological forms in all three languages that are primarily associated with curses. Secondly, three constructional features were found to be particularly present in curses: the absence of words for ‘God’ and ‘Allah’, the presence of second person pronouns and their final position in the utterance, the last feature being the most consistent. Finally, an important finding of this study is that both the morphological and constructional features associated with curses manifest themselves only as tendencies.

Keywords: blessings, curses, impoliteness, Nakh-Daghestanian languages, optative, wishes

1 Introduction

Depending on the culture, wishes can be limited to a small number of idiomatic expressions used in specific situations such as English *stay well* or *go to hell*, or be an elaborate field with hundreds of formulae catering to the needs of (im)politeness. Formulaic language is highly characteristic of politeness, deeply rooted in social and cultural conventions (Terkourafi 2002). The conventionalized formulae ensure that “the speaker gets what he/she wants and is perceived as an individual within the group” (Wray and Perkins 2000: 18), since a hearer is more likely to

Nina Dobrushina, Laboratoire Dynamique Du Langage, CNRS, and Université Lyon 2, DDL - MSH-LSE, 14 avenue Berthelot, 69363 Lyon Cedex 07, France, e-mail: nina.dobrushina@cnrs.fr

interpret a message correctly if it is in a form he/she has heard before. While the use of formulae is thought to be a prominent feature of polite discourse in any culture (Terkourafi 2002), the role of the conventionalization in impoliteness – negatively evaluated (linguistic) behaviors that have offensive effects (Culpeper 2011: 117; Van Olmen, Andersson and Culpeper 2023) – is much less studied (see also Van Olmen et al., this volume) Are there any linguistic structures that are directly associated with impoliteness?

A pilot cross-linguistic study of insulting expressions showed that there is strong evidence for the existence of grammaticalized ‘insultive’ constructions in the languages of the world (Giomi and Van Oers 2022). In rural Daghestan, curses (ill wishes) are among the main means of expressing offense (cf. curses and ill wishes in the list of the conventionalized impoliteness formulae attested in the *Oxford English Corpus*, Culpeper 2011: 135–136; see also Culpeper, Van Dorst and Gillings, this volume). This paper seeks to find evidence for grammatical structures characteristic of curses in the languages of Daghestan (North Caucasus) (see Pater-noster, this volume, for a curse construction in Italian).

In Daghestan, wishes are an integral part of communication. Blessings are common speech rituals, almost obligatory in everyday village life as one of the main means of expressing politeness. Curses are used to express bad feelings directed at a specific person. For example, in the highland village of Archib where the one-village language Archi (Lezgif, Nakh-Daghestanian) is spoken, if someone comes from one of the downhill villages to bring an Archi person some goods, he can be greeted with the following wish:

- (1) *danna-k uq'a-šaw χ:wari-ši zaba šummar la:χa b-a*
 where-LAT 1.go.PFV-CONC be.glad-CVB come.IMP life be.long 3-do.IMP
 ‘Wherever you go, be happy when you come back, long live!’
 (Karim Musaev pc)

If the neighbors are fighting, one of them might use the following curse:

- (2) *ik^w o'rc-a*
 heart 4.freeze-IMP
 ‘Let (your) heart freeze (= May you die quickly).’
 (Karim Musaev pc)

The importance of wishes in the culture of Daghestanian people has been noticed by the specialists on Nakh-Daghestanian languages (Khaidakov 1961; Musaev 2002: 111; Mallaeva 2007: 69–60, Dobrushina 2011), but their structural properties have yet to be studied by linguists.

The expressions of wishes are usually considered by linguists as a whole, indiscriminately as to the value encoded by the wish. Meanwhile, as shown in Dobrushina (2024), curses can be expressed by dedicated morphological means:

- (3) Kazakh (Turkic)
aram qat-qyr
 filthy freeze-MAL.OPT
 ‘Drop dead.’ (‘Die filthy = pagan’)

Since both blessings and curses are very widely used in Daghestan, and are often expressed by the same suffixes, they present an excellent case to look for grammatical and constructional distinctions between wishes with positive value (politeness) and wishes with negative value (impoliteness). The comparison of the morphological, syntactic and semantic properties of blessings and curses in the languages of the Nakh-Daghestanian family is the aim of this paper. The study of conventionalized expression of curses will contribute to the discussion whether impoliteness can be inherently associated with linguistic structures or is primarily determined by context.

I will start with a short introduction to the region and the family. I will then briefly describe the main grammatical means that are used for wishes in the languages of the family. In the main part of the paper I will look at the data from three Nakh-Daghestanian languages (Avar, Rutul, Lak) to understand how they encode blessings and curses.

1.1 Daghestan and Nakh-Daghestanian languages

The republic of Daghestan is an area of high language density and linguistic diversity. Most of the languages spoken in Daghestan belong to the Nakh-Daghestanian (East Caucasian) language family, which counts up to 50 languages (depending on the researcher; see Figure 1). There are also three Turkic (Kumyk, Nogai, Azerbaijani) and two Indo-European (Tat and Russian) languages spoken in the area (for an overview of languages see Wixman 1980; Tuite 1999; Koryakov 2002; Berg 2005; Chirikba 2008). The estimated number of languages ranges from 30 to 50. The size of the language population varies significantly, starting from 500 people (Hinug) to half a million (Avar).

With so many languages packed in a relatively small area, it is not infrequent that in two adjacent villages located at a walking distance from one another, different languages are spoken. Economic and social ties between adjacent villages have always been strong. Interethnic communication requires knowledge of a shared

language. There was never a lingua franca common to all of Daghestan (Chirikba 2008: 30). Until the mid 19th century, the lowlands were dominated by speakers of Turkic languages, which led to the use of Kumyk and Azerbaijani as lingua francas in adjacent territories (Wixman 1980: 108–119). The local language Avar was used as a lingua franca in some parts of highland North Daghestan. Besides their native language, most Daghestanian highlanders also spoke the language of the closest village(s), and many of them additionally acquired some distant language due to the seasonal works or trade. A significant change occurred in the middle of the 20th century, when, due to the spread of secular schools, the Russian language came to Daghestan and became the all-Daghestanian lingua franca.

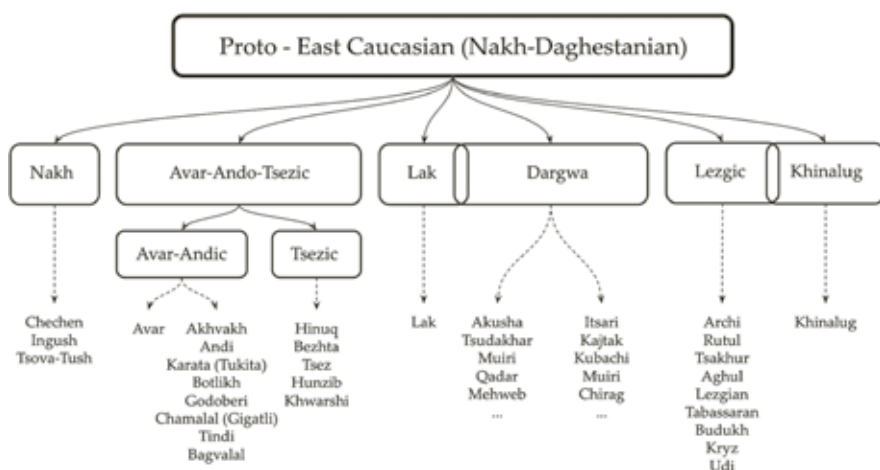


Figure 1: East Caucasian language family (courtesy of Michael Daniel).

In contrast to the extreme diversity in terms of languages and situations of language contact, Daghestan is rather homogeneous in terms of culture and history (Wixman 1980: 107). The vast majority of Daghestanians are muslims, and a command of classical Arabic was typical for some part of the population. Arabic was sporadically used for writing in local languages, but indigenous languages remained largely unwritten until the 20th century.

Nakh-Daghestanian languages are predominantly agglutinative, with most inflectional categories being expressed by suffixation. They are especially famous for their rich nominal paradigms with large systems of spatial cases. Most languages have the category of gender distinguishing from two to six values. The systems of moods are usually rich in volitional forms, but lack forms with general, non-specified irrealis meaning. Most languages have an inflectional prohibitive (negative

imperative) which is usually not derived from imperative. The basic word order in Nakh-Daghestanian languages is SOV, with nominal dependents situated to the left from the head. As all families endemic for the Caucasus, Nakh-Daghestanian languages are consistently ergative, which means that case marking distinguishes S/P vs A (where S is the sole argument of intransitive verb, P is a patient, A is an agent) and the agreement is largely controlled by nominative arguments.

1.2 Morphological means of expressing wishes in the languages of Daghestan

In the languages of the Nakh-Daghestanian family, wishes are most often expressed by synthetic optatives – morphological forms dedicated to expressing a wish or hope of the speaker that something would happen (Bybee, Perkins and Pagliuca 1994: 321; Dobrushina, van der Auwera and Goussev 2013; Nikolaeva 2016: 77).

Dedicated morphological optatives are cross-linguistically rather infrequent, as was shown in the World Atlas of Language Structures (Dobrushina, van der Auwera and Goussev 2013). Only 48 of 319 languages surveyed in WALS have it, and many of them are found in northern India and Nepal and in the Caucasus. Most languages of the Caucasus have special morphological forms to express wishes; some of them have several such forms.

The forms which express optative meanings can have different functional scopes in Nakh-Daghestanian languages. In some languages, the optative expresses only blessings or curses, while another form is used to express indirect commands (the latter form is also referred to as 3rd person imperative or jussive); this is the case of Mehweb (Dargwa group of Nakh-Daghestanian family) in (4) and (5):

(4) Mehweb optative

ja-allah ħušab taliħ g-a-b!

PTCL-Allah 2PL.DAT luck give.PFV-IRR-OPT

‘May [Allah] give [you] luck!’

(Dobrushina 2019:143)

(5) Mehweb jussive

musa uz-e bet’a

Musa 1.work.IPFV-IMP say.PFV-IMP

‘Let Musa work.’

(Dobrushina 2019:143)

Many other languages combine the two functions in one form. For example, Rutul (Lezgic group of Nakh-Daghestanian family) has the same form to express blessings and curses addressed to 1st, 2nd or 3rd persons (6), and to express 3rd person imperatives (7). This form cannot be used for commands addressed to 1st or 2nd persons (8):

- (6) *zi / wi / had saʁ r-iš-ij*
 I / you / that healthy 2-become-OPT
 ‘May I / you / she be healthy.’
 (own fieldwork)

- (7) *said-a uqʼ sej-ij*
 Said-ERG grass 4.mow-OPT
 ‘Said should mow the grass.’
 (own fieldwork)

- (8) **wa-d uqʼ sej-ij*
 2SG-ERG grass 4.mow-OPT
 ‘You should mow the grass.’
 (own fieldwork)

In this paper, the forms that pattern similarly to those of Rutul will be classified as *optatives*. I reserve the term *jussive* for the forms which are restricted to third person only and solely express indirect commands as in Mehweb (Dobrushina 2019: 133–138), or maximally combine it with the expression of a wish with a 3rd person subject, as in Kumyk (Gadziakhmedov 2000).

Most languages of the family have either several dedicated morphological optatives or at least one dedicated optative and some additional means of expressing the wish of the speaker. These forms can be synonymous or can show some semantic distinctions. Semantic distinctions relevant for this paper (between blessings and curses) will be discussed in Section 2. As for the morphosyntactic distinctions, the most striking is that some optatives which will be discussed in this paper are not verbs in the proper sense. Nominal optative forms were mentioned, for example, in Abdullaev (1954) for standard Dargwa and thoroughly discussed in Sumbatova and Lander (2014) for Tanty Dargwa, where one of the optative forms takes nominal plural endings (9b), can occur in argument position and can be marked for case (9c):

- (9) a. *ʃuʳ allah-li w-at*
 2SG Allah-ERG 1-leave.PFV
 ‘May Allah let you be (i.e. May Allah bless you)!’
- b. *ʃuʳx:a allah-li d-at-are*
 2PL Allah-ERG 1/2PL-leave.PFV-PL
 ‘May Allah let you (plural) be!’
- c. *sa-j allah-li w-at-la kulpat b-erq-ab*
 self-M Allah-ERG 1-leave.PFV-GEN family HPL-blossom.PFV-OPT
 ‘May the family of (this person who was) let be by Allah multiply!’
 (Sumbatova and Lander 2014: 132)

In this paper, I will mainly focus on dedicated optative forms, but will also consider other means of expressing wishes if they are frequent enough in my data.

1.3 Blessings and curses in Daghestan

As was mentioned above, blessings and curses are very typical of the communication in Daghestanian villages. Blessings are abundantly used for greetings. The choice of blessing depends on the specific situation such as gender of the person who greets and is greeted, or the current activity in which one or the other party is involved (for example a woman who is milking the cow will be greeted differently from a woman who is mowing). Blessings will be used when someone leaves the village or comes back from a trip, bought a new cloth or recovered from a disease, moved into a new house or came back from the garden with the harvest. Curses are used when people fight or when an inferior misbehaves. There could be specific curses for different occasions, for example addressed to someone who eats too much, is stubborn or greedy, disrespectful or lazy. Unlike blessings, curses can be frequently addressed to animals.

Wishes usually have a clear value – they call for something bad or good to happen. There are however very rare cases when the wish is neutral; in my data, these are wishes which are not addressed to anyone particular, see example (10):

- (10) Lak
q:a-b-u<w>ǰʳ-ajnan k:a<w>k:u-n kʰul b-an-naw
 NEG-3-<3>know-CVB.CNT <3>see-INF know 3-do-OPT
 ‘May those who do not understand see for themselves.’ (i.e., it is difficult to understand something unless one is convinced in practice)
 (Ramazanova 2005: 119)

There is some cross-linguistic evidence for forms dedicated to curses (maybe referred to as maledictives; cf. also malefactive imperative in Aikhenvald 2020). For example, such a form is reported for Tariana (Arawak, Aikhenvald 2020) and Kalamang (Trans-New Guinea, Visser 2022: 479–481). In Dobrushina (2024), I show that Kazakh (Turkic language of Central Asia) has a form which is used exclusively for curses. I did not however find forms dedicated to the expression of curses in Daghestan (neither in the Nakh-Daghestanian family, nor in the other languages of the region). Usually blessings and curses are expressed by the same morphological form, but in some languages certain forms tend to specialize towards blessings and curses. More than that, as will be shown in this paper, there are certain constructional properties associated with curses.

For this study I chose three Nakh-Daghestanian languages which belong to different branches and are not or almost not in contact: Avar, Rutul and Lak.

1.4 Data

My goal in this paper is twofold. On the one hand, I look at the optative forms in order to understand whether they tend to specialize towards blessings and curses or not. On the other hand, I want to analyze wishes irrespective of their form with the aim of finding all possible morphological and syntactic means of expressing wishes in each language. For this, besides the grammars which account for the main forms, I need a source of examples of wishes. Corpora can hardly help to collect a sufficient number of wishes. Wishes are mainly used in dialogues. Since most available corpora of Daghestanian languages consist of narratives, examples of wishes are not many. For instance, in the corpus of the Kina dialect of Rutul (about 20,000 words) there are 28 examples of wishes, from which only five are curses, while all other examples instantiate blessings. These data are not sufficient to study the distinctions between blessings and curses. This is why I used another type of data – dictionaries.

Dictionaries of Nakh-Daghestanian languages proved to be a good source of examples of wishes. Due to the local lexicographic tradition, dictionaries contain plenty of examples illustrating the usage of some words, wishes among them. For example, in the Rutul-Russian dictionary the verb *qiq'as* ‘to return’ is illustrated by the following wish:

- (11) *za ki-wi-d xiw wi-di ul-ab-a: qi-q'-ij*
 1SG.ERG PV-3.give.PFV-ATTR bread 2SG-ATTR eye-OBL.PL-IN.EL PV-4.come-OPT
 Curse: ‘May the good which I have done for you not benefit you.’ (lit. ‘Let the bread with which I have fed you go out of your eyes.’)
 (Alisultanov and Suleimanova 2019: 397)

Some dictionaries contain hundreds of wish expressions, thus providing sufficient data for the current research. For some languages, there are also dictionaries of idiomatic expressions or collections of folklore genres which often include wishes.

Since all these dictionaries and collections are bilingual in Russian, I used several Russian queries in order to identify wishes: *pust'* (jussive particle), *čtob* (conjunction used in insubordinate constructions denoting curses), *(ne) dai Bog* ('God (don't) give'), *da* (optative particle). Examples were manually extracted from the dictionaries, accumulated in an Excel table and coded according to their value (positive or negative) and to their constructional properties. Table 1 shows the sources which were used for this study and the number of examples collected for each language. In addition to these sources, I also cite examples from various grammars and examples provided by language consultants, but the quantitative study is based only on the sources listed in Table 1.

Table 1: The sources of the examples of wish expressions.

	Bilingual Russian dictionary	Dictionary of idiomatic expressions or collection of folklore genres	Number of examples
Avar	Gimbatov (2006)	Magomedkhanov (1993)	191
Rutul	Alisultanov and Sulejmanova (2019)	Makhmudova (2016)	96
Lak	Abdullaev (2015)	Ramazanova (2005)	238

Relative frequency of blessings and curses in the dictionaries is not meaningful; their actual frequency depends on the situation. For example, in the above mentioned corpus of Kina Rutul most of the blessings were used in one text, which was a toast. The distribution of formal means in relation to the value of the wish however can be meaningful. In this paper, the numbers are used in order to compare blessings and curses and to find out the formal means that are more typical of curses.

2 Blessings and curses in Avar, Rutul and Lak

In what follows I will consider the data from three Nakh-Daghestanian languages: Avar (Avar-Ando-Tsezic), Lak (Lak) and Rutul (Lezgic). My first question will be whether some of the morphological forms used for wishes show a preference for a positive or negative value. Second, I will consider whether curses have some special constructional properties which make them distinct from blessings.

2.1 Avar

Avar is the largest language in Daghestan with about 700,000 speakers. Avar was chosen for this study for two reasons. First, there are two sources from which wishes can be extracted: a large Avar-Russian dictionary and a dictionary of Avar idiomatic expressions. Second, there are several optative forms in Avar, which allows me to test whether there is a specialization of one of the forms towards the expression of curses. The following three forms were considered: *-gi*, *-ad*, and *-ja-G*, where G is a gender marker.

By far the most common optative form (according to the examples in the dictionary by Gimbatov (2006)) is the one with the suffix *-gi*.

- (12) *une-s-e* *nux* *b-it'a-gi*, *č'ole-s-e* *ruq'* *b-it'a-gi*
 leave-OBL-DAT road 3-straighten-OPT stay-OBL-DAT house 3-straighten-OPT
 Blessing: 'May the one who leaves have a good trip, may the one who stays
 stay happily at home'
 (Alekseev et al. 2012: 2024)

- (13) *ara-tu-sa* *w-us:un-ge-gi* *mun*
 go.PTCP-IN-EL 1-return-PROH-OPT you.sg
 Curse: 'May you not return from where you go (may you perish)'
 (Gimbatov 2006: 230)

The other two forms are morphologically nominal. The form with the suffix *-ad* can be classified as a kind of masdar (verbal noun). The optative in *-ad* can have plural ending *-al*, which is also found with nouns:

- (14) *gull-i-ca* *w-orł-ad*
 bullet-OBL-ERG 1-pierce-OPT
 Curse: 'May a bullet pierce you'.
 (Gimbatov 2006: 207)

- (15) *xur-ad-al!*
 exterminate-OPT-PL
 Curse: 'May you all be exterminated!'
 (Gimbatov 2006: 325)

The optative in *-ad* can inflect for case: ergative in example (17a), dative in (17b), genitive in (17c) (the last three examples provided by Zulaikhat Mallaeva).

- (16) *raλ'-aλ' k^w-ad*
 ground-ERG eat-OPT
 Curse: 'May you be swallowed up by the ground'
 (Gimbatov 2006: 929)
- (17) a. *rak' b-oχ-ad-i-c:a b-ic-ana he-b χabar*
 heart 3-rejoice-OPT-OBL-ERG 3-tell-AOR DEM-3 story
 Blessing: 'He brought this message, may his heart rejoice'
- b. *rak' b-oχ-ad-i-je b-ic-e he-b χabar*
 heart 3-rejoice-OPT-OBL-DAT 3-tell-IMP DEM-3 story
 Blessing: 'Tell him this message, may his heart rejoice'.
- c. *rak' b-oχ-ad-i-l χarb-i-l bercin-λi*
 heart 3-rejoice-OPT-OBL-GEN story-OBL-GEN beautiful-NMLS
 Blessing: 'How beautiful is his speech, may his heart rejoice'.

The third form, the optative in *-ja-G* is a participle with a position for gender marker that depends on the gender of the addressee, even if the addressee is not an argument of the verb:

- (18) a. *dur k^wer b-aq'^wa-ja-j*
 you.sg.GEN hand 3-become.dry.IMP-OPT-2
 Curse: 'May your hand become withered' (addressed to a woman)
- b. *dur k^wer b-aq'^wa-ja-w*
 you.sg.GEN hand 3-become.dry.IMP-OPT-1
 Curse: 'May your hand become withered' (addressed to a man)
 (Testelels n.d.)
- (19) *waj dur dada č'aqa-ja-j*
 INTJ you. SG.GEN dad be.healthy.IMP-OPT-2
 Blessing: 'May your dad be healthy' (addressed to a woman)
 (Gimbatov 2006: 453)
- (20) *insul ruq' b-aq'^w-ara-b roł-u-l c'e-ja-w*
 father.GEN house 3-become.dry.IMP-PST.PTCP-3 wheat-OBL-GEN fill.up-OPT-1
 Blessing: 'May your father's house be filled with dried wheat' (addressed to a man)
 (Gimbatov 2006: 526)

2.1.1 Distribution of the forms according to the value of the wish

There is no visible difference in the meanings of optatives in *-gi* and *-ja-G*. For example, in the dictionary by Gimbatov (2006) there are two synonymous examples with different optative forms:

- (21) *ʃumru λ'-un w-oχa-gi*
 life give-CVB 1-rejoice-OPT
 Blessing: 'May you be happy to live a long life'
 (Gimbatov 2006: 442)

- (22) *ʃumru λ'-un w-oχa-ja-w*
 life give-CVB 1-rejoice-OPT-1
 Blessing: 'May you be happy to live a long life'
 (Gimbatov 2006: 442)

Both forms are used to express both blessings and curses. The value can be deduced only from the sentential context. For example, the two following examples have an identical structure with the same transitive verb 'to wash'. The Agent participant is absent in both constructions. The first wish is positive; it is typically used in reference to a deceased person. The second wish is negative; it hints at the fact that dead bodies are washed before being buried.

- (23) *munah-al čura-ja-w*
 sin-PL wash-OPT-1
 Blessing: 'May his sins be forgiven!'
 (Magomedkhanov 1993: 223)

- (24) *čex čura-ja-w*
 stomach wash-OPT-1
 Curse: 'May you die!' (lit. 'May your belly be washed!')
 (Magomedkhanov 1993: 388)

As for the optative in *-ad*, all examples in my data (though not very numerous) are curses. Table 2 presents the distribution of the optative forms according to the value of the wish:

Table 2: Distribution of blessings and curses in Avar optatives

	Blessings	Curses	total
<i>-gi</i>	63	88	151
<i>-ad</i>	0	13	13
<i>-ja-G</i>	14	13	27

The preference for expressing curses by the optative in *-ad* is only a tendency; the expression of blessings is not ruled out. Although they were not found in dictionary data, the Avar consultant (Zulajkhat Mallaeva) came up with the following examples:

- (25) *rak'* *b-oχ-ad*
heart 3-rejoice-PTCP
Blessing: 'May your heart rejoice!'

- (26) *bet'er* *č'aq-ad*
head prosper-OPT
Blessing: 'May your head prosper'

- (27) *ħal* *č'aq-ad*
well.being prosper-OPT
Blessing: 'May you be well'

To sum up, the majority of wish expressions contain a verb form which is neutral (not associated with blessings or curses). I will now consider whether curses have some special constructional properties which make them distinct from blessings.

2.1.2 Distribution of the constructions according to the value of wish

There are several constructional peculiarities of blessings versus curses that manifest themselves at the quantitative level: the mention of a supernatural power, the explicit mention of the second person, and the final position of the second person pronoun.

Blessings and curses are peculiar in that, like imperatives, they imply performing a wish, but, unlike imperatives, the intended performer is not the addressee but a supernatural power. The implicit performer of most wishes in Daghestan is generally understood to be Allah; native speakers often translate optative utterances into Russian using the words ‘Allah’ or ‘God’, even when the reference to the god was not present in the original sentence.

Most often, the performer of the wish is omitted. This is why transitive wish constructions usually lack the ergative argument, as in the following example with the transitive verb ‘give’ and omitted subject:

- (28) *du-je* *ʃumru* *λ'e-ja-w*
 you(sg)-DAT life give-OPT-1
 Blessing: ‘May you live long’ (lit. ‘to you life may give’)
 (Gimbatov 2006: 442)

Wish expressions with an ergative are infrequent in my data, and are less typical for blessings than for curses. In blessings, the ergative position is always filled by the word *Allah*:

- (29) *allah-as* *c'una-gi* *hedina-b* *balah-al-da-s*
 Allah-ERG save-OPT such-3 trouble-OBL-LOC-ABL
 Blessing: ‘May Allah protect you from such a misfortune’
 (Gimbatov 2006: 98)

Besides ergative, the word *Allah* can be present in the form of the nominative (in intransitive clauses) or the genitive in a possessive construction:

- (30) *allah* *gurħa-gi*
 Allah have.mercy-OPT
 Blessing: ‘May Allah have mercy’
 (Gimbatov 2006: 44)
- (31) *allah-as-ul* *barkat* *daim-ʔa-gi*
 Allah-OBL-GEN grace be.eternal-become-OPT
 Blessing: ‘May God’s grace be everlasting’
 (Gimbatov 2006: 454)

Conversely, the ergative argument in curses is usually not the name of the supernatural force but rather a real entity which is meant to harm the addressee of the wish, such as ‘bullet’ (14), ‘wolf’ (32), or ‘raven’ (33):

- (32) *bac'-i-ca* *č^wa-gi* *mun*
 wolf-OBL-ERG kill-OPT you.SG
 Curse: 'May you be killed by a wolf'
 (Gimbatov 2006: 126)
- (33) *nuq-i-ca* *r-ik'a-gi* *dur* *ber-al*
 raven-OBL-ERG NPL-dig.out-OPT you.sg.GEN eye-PL
 Curse: 'May the raven pluck out your eyes'
 (Gimbatov 2006: 177)

I conclude that the presence of the word 'Allah' is typical almost exclusively of blessings.

Another feature that distinguishes blessings and curses is the explicit mention of the second person. In Avar, verbs do not inflect for person, so the only way for a personal reference to be expressed is the usage of a pronoun. Most wishes are addressed to the hearer, but the second person is often not present. Nevertheless, speakers may use second person pronoun when they translate wishes into Russian:

- (34) *k^wer-al* *q'ota-gi*
 hand-PL cut-OPT
 Curse: 'May your hands be cut off!'
 (Gimbatov 2006: 562)

Second person pronouns can be present in the nominative (as an object in transitive clauses) (35), genitive form (as a possessor) (36), or dative form (as an addressee) (37).

- (35) *mun* *č'efer* *raλ'-al-d-e* *w-ača-ja-w*
 you.sg black soil-OBL-SUP-LAT 1-carry-OPT-1
 Curse: 'I wish you were dead!' (lit. 'May you be carried away onto black soil!')
 (Magomedkhanov 1993: 222)
- (36) *dur* *ruq'* *b-uħa-ja-w*
 you.sg.GEN house 3-burn-OPT-1
 Curse: 'May your house burn down!'
 (Magomedkhanov 1993: 127)
- (37) *du-je* *ʃumru* *λ'e-ja-w*
 you.sg-DAT life give-OPT-1
 Blessing: 'May you live long!'
 (Gimbatov 2006: 442)

Whatever the form, second person pronouns occur much more frequently in curses than blessings.

More than that, the second person pronouns in curses often occupy the final position in the clause (38–39), but are not excluded in blessings (40). Nakh-Daghestanian languages typically are verb-final (Forker and Belyaev 2016), so having a personal pronoun at the end of the clause is a rather marked word order:

- (38) *χab-a-t'a mič' b-iža-gi dur*
grave-OBL-SUP nettle 3-grow-OPT you.sg.GEN
Curse: 'May your grave grow nettles'
(Gimbatov 2006: 172)

- (39) *dandamaj-gi mun!*
swell-OPT you.sg
Curse: 'May you swell up!'
(Gimbatov 2006: 456)

- (40) *alžan λ'e-gi du-je*
paradise give-OPT you.sg-DAT
Blessing: 'May you go to heaven'
(Gimbatov 2006: 43)

Blessings feature such a word order much less frequently (Table 3). The counting shows that only 11% of blessings expressed by the form in *-gi* contain a second person pronoun, as compared to more than 60% of curses. As for the pronoun-final position, 51% of curses expressed by the form in *-gi* and only 6% of blessings have it. In the wishes expressed by *-ja* and *-ad* this tendency is not present.

Table 3: Second person pronouns in blessings and curses in Avar.

		absent	final position	non-final position	total
-gi	blessings	56	4	3	63
	curses	32	45	11	88
-ja	blessings	10	0	3	13
	curses	10	0	4	14
-ad	blessings	0	0	0	0
	curses	13	0	0	13

This observation complies with the evidence from some non-Nakh-Daghestanian languages. As shown in Van Olmen, Andersson and Culpeper (2023), insults tend to increase the pragmatic explicitness of the addressee, spelling out its second person target openly (see also Van Olmen and Andersson, this volume). This was noticed based on the data of several European languages which all use the construction *YOU+NP* with explicit second person pronouns (*you idiot*). In the further sections, I will test whether this tendency holds for other Nakh-Daghestanian languages.

To summarize, there are three optatives in Avar. None of them is used exclusively for curses, but one (the form in *-ad*) seems to prefer such contexts. There are several constructional distinctions between blessings and curses:

- curses usually do not contain the mention of Allah,
- curses more often contain second person pronoun,
- curses frequently feature the final position of the second person pronoun.

2.2 Rutul

Rutul belongs to the Lezgian group of the Nakh-Daghestanian family. Unlike Avar, Rutul is a minority language (about 10,000 speakers in Southern Daghestan). There was no bilingualism between Avar and Rutul as far as it can be traced; they are separated geographically and culturally. Rutul people were bilingual in Azerbaijani and used it for communication with adjacent language groups, Tsakhur and Lezgian. For this study, I collected about one hundred examples of wishes in the dictionary of standard Rutul (dialect of the village of Rutul, also called Mukhad Rutul) supplemented by two dozen examples in the dictionary of idiomatic expressions by Makhmudova (2016).

There is only one dedicated optative form in Rutul, a form with the suffix *-xʲ* or *-j* (variant of the same suffix) after a vowel.

- (41) *zi wi-di ul-i-s r-iq'-ij*
 I you.sg-ATTR eye-OBL-DAT 2-die-OPT
 Curse: 'May I be sacrificed for your eye!'
 (Makhmudova 2016: 178)

A plausible hypothesis expressed in Makhmudova (2002) is that the Rutul optative in *-xʲ/-j* originates from the imperative of the verb 'say' *jixʲ*. But as was discovered by the current study, apart from the form in *-xʲ/-j*, wishes can also be expressed by a form in *-di* (eleven examples in Alisultanov and Suleimanova 2019):

- (42) *kan l-ešu-di*
 core PV-1.take.PFV-DI
 Curse: ‘May you die!’
 (Alisultanov and Suleimanova 2019: 178)

The existence and origin of the form in *-di*, to my knowledge, has never been discussed.

In Mukhad Rutul, the suffix *-di* is found in attributives (including participles). This suffix has two realizations depending on the preceding phoneme: *-di* after a consonant and *-d* after a vowel (see Alekseeva 2024):

- (43) *tʰi<wi>rka-d tila*
 <3>lame-ATTR dog
 ‘lame dog’
- (44) *husur-di kantʰ*
 sharp-ATTR knife
 ‘sharp knife’

The participial origin of the form used in wishes is very likely, as we find many parallels in other Nakh-Daghestanian languages. For example, the participle as the main predicate of wishes was found in another dialect of Rutul, that of the village of Kina. The problem with the wishes found in the Mukhad Rutul dictionary is that they have suffix *-di* after the vowels, which is not expected in participles. As I will show in Section 2.2.2, the empirical findings of this paper help to suggest a solution to this problem and to explain the origin of the *-di* form.

2.2.1 Distribution of the forms according to the value of the wish

While the optative in *-xʰ/-j* is equally used in blessings (45) and curses (46), the forms in *-di* (47) are found only in curses (see Table 4).

- (45) *duχ-re saχ d-iš-ij*
 son-PL healthy HPL-become.PFV-OPT
 Blessing: ‘May your sons be well!’
 (Alisultanov and Suleimanova 2019: 301)

- (46) *wi-di gadam midi qu[˘]ma[˘]bu[˘]x^j*
 you.sg-ATTR step here RE-PROH-4.come
 Curse: 'May you never set foot here again!'
 (Alisultanov and Suleimanova 2019: 203)

- (47) *k'a[˘]šen jiši-di*
 coal 1.become.PFV-Dİ
 Curse: 'May you burn!'
 (Alisultanov and Suleimanova 2019: 230)

Although there are no examples of the *-di* form used in positive wishes in the dictionaries, my language consultants think that blessings can be also expressed by this form. The following example was suggested by Mariza Ibragimova (pc):

- (48) *baχt-li jiši-di*
 happiness-ADJ 1.become.PFV-Dİ
 Blessing: 'May you be happy!'

Table 4: Distribution of blessings and curses in Rutul optatives.

	Blessings	Curses	total
<i>-x/-j</i>	51	34	85
<i>-di</i>	0	11	11

2.2.2 Distribution of the constructions according to the value of wish

As mentioned above, the optative in *-x/-j* is used both for blessings and curses. The constructional tendencies observed for Avar show up in Rutul wishes with the optative in *-x/-j* as well. Only blessings contain mentions of Allah or God (*jiniš* in Mukhad Rutul):

- (49) *jinc-i-re u[˘]χ-u[˘]x^j wi*
 god-OBL-ERG 1.save-OPT you.sg
 Blessing: 'May God save you!'
 (Alisultanov and Suleimanova 2019: 168)

- (50) *allah-a-ra wa-s nusrat wij-ixʲ*
 Allah-OBL-ERG you.sg-DAT help 4.give-OPT
 Blessing: ‘May Allah help you!’
 (Alisultanov and Suleimanova 2019: 279)

Conversely, curses expressed by the optative in *-xʲ/-j* more often contain the mention of the second person, and especially in the final position (Table 5).

- (51) *ul xed w-iš-ij wi-di*
 eye water 3-become.PFV-OPT you.sg-ATTR
 Curse: ‘May your eye become water’ ≈ ‘May you go blind!’
 (Alisultanov and Suleimanova 2019: 408)

- (52) *naʰhnet jik-ij wa-s*
 curse 4.become-OPT you.sg-DAT
 Curse: ‘Damn you!’
 (Alisultanov and Suleimanova 2019: 277)

Table 5: Second person pronouns in Rutul blessings and curses expressed by optative in *-xʲ/-j*.

	absent	final position	non-final position	total
Blessings	32	4	14	50
Curses	8	18	8	34

Wishes with the forms in *-di*, by contrast, never contain a second person subject in our data, although, according to the language consultant Alisultan Alisultanov, they are always used with reference to second person.

The examples which were elicited with the help of the language consultant showed that the form is identical to a participle in all respects apart from the fact that it has the suffix *-di* instead of the expected *-d*. It is negated with the prefix *že-*, which is only possible in non-finite forms (unlike the optative in *-xʲ/-j*, negated as a prohibitive, the dedicated form for the negative imperative). If the form is addressed to several addressees, the nominal plural suffix *-bir* will be used:

- (53) *haj ke-že-d-gi-d-bir*
 INTJ PV-NEG-HPL-be.delayed.PFV-ATTR-PL
 Curse: ‘I wish you (plural) were gone!’
 (pc with Alisultan Alisultanov)

It can take a case form, such as the ergative case in the following example, and has an attributive type of declension:

- (54) *ha ke-že-r-gi-d-nij-e* *midi? ul*
 that PV-NEG-1-be.delayed.PFV-ATTR-OBL.H-ERG here eye
li-qi-wi-r diš
 PV-PV-3.give.PFV-CVB NEG
 Curse: 'He never once looked this way/never showed up here (I wish he were lost!)'
 (pc with Alisultan Alisultanov)

Notably, in the cases of further inflection added to the participle-like form, the unexpected *-i* in *-di* disappears, and the form aligns with the regular participle. The uninflected form can be used with the variant *-d* only if the construction contains the name of the addressee of the wish (there are no such cases in the dictionary). Compare (55a,b) and (56a,b) (examples courtesy Timur Maisak and Mariza Ibragimova):

- (55) a. *dünja:-di:* *ke-že-r-gi-di*
 world-OBL.SUP(ESS) PV-NEG-1-be.delayed.PFV-DI
 Curse: 'May you be gone from the world!'
 b. *dünja:-di:* *ke-že-r-gi-d* *maḡač*
 world-OBL.SUP(ESS) PV-NEG-1-be.delayed.PFV-ATTR Maḡač
 Curse: 'May you be gone from the world, Makhach!'
- (56) a. *haj hu<r>xu-di*
 INTJ <2>burn.PFV-OPT
 Curse: 'May you burn!'
 b. *haj hu<r>xu-d* *riši*
 INTJ <2>burn.PFV-ATTR sister
 Curse: 'May you burn, sister!'

Based on this distribution, I suggest that the form in *-di* grammaticalized from the combination of the participle in *-d* with the second person pronoun *wi* (nominative):

kežergid wi -> kežergidi

The hypothesis explains why second person pronoun is never found in the wishes with *-di* form, why *-di* form can refer exclusively to second person and is incompatible with the further nominal inflection (54) or with reference to the addressee

of wish by a personal name (55b). On the other hand, it is also in line with the tendency found for Avar optative in *-gi* and Rutul optative in *-x/-j* that curses often have second person pronouns in the final position.

Interestingly, in another dialect of Rutul, that of the village of Kina, the contraction of second person pronoun and participle suffix did not happen. In this dialect, participles are used for wishes with the suffix *-d* and can be combined with the second person pronoun:

- (57) *wi-di* *mo'ho'bad* *düz* *ž-i<w>ši-d*
 you.sg-ATTR wish straight NEG-3.become.PFV-ATTR
 Curse: 'May your wishes not come true!'
 (personal fieldwork)
- (58) *hej* *kowχa* *hej* *sax* *hiši-d* *wi* *kowχa*
 hey elder hey healthy 1.become.PFV-ATTR you.sg elder
 Blessing: 'Hey, foreman - hey, be well, foreman.'
 (Kina Rutul corpus)

To sum up, Rutul has an optative which is equally used for blessings and curses, and a form in *-di(-d)* (originally participle) used predominantly for curses. Wish-constructions with the optative share the same properties as Avar wishes: the mention of Allah (God) is more typical of blessings, while the presence of clause-final second person pronoun is found mainly in curses. I have also suggested a diachronic scenario of how a special form in *-di* has appeared through the contraction of the participle and second person pronoun.

2.3 Lak

Lak is another large language of Daghestan (an estimated 150,000 speakers). It is taught at schools and has a relatively long writing tradition. Lak constitutes a separate branch within the family.

The examples of wishes for this study were taken from the dictionary by Abdulaev (2015) and the collection of idiomatic expressions by Ramazanova (2005). Among 262 examples of wish expressions there are two dedicated forms (*-naw* and *-iwu*) and several non-dedicated forms; from the latter, I will only discuss the most frequent form.

Overall, the most common form is that with the suffix *-naw* (Desiderative I in Friedman 2021). Unlike most other Lak forms, the optative in *-naw* does not have personal endings.

First person

- (59) *harx:unu-s:a x:u d-an q:a-itan-naw*
 tonight-ADJ night 4-do-INF NEG-leave-OPT
 Curse: 'May I not survive tonight.'
 (Abdullaev 2015: 559)

Second person

- (60) *ja wi-x busu-l b-usan-naw ja ina*
 or you.sg-AD teller-ERG 3-tell-OPT or you.sg
lasu-l lasun-naw
 taker-ERG take-OPT
 Curse: 'May you either learn a lesson (lit. told by someone) or be taken away.'
 (Lit. 'Either let he who tells tell you, or let he who takes take you')
 (Abdullaev 2015: 393)

Third person

- (61) *c^w-an d-iwa-n allaha-na-l b-an-naw*
 REFL.I.PL-DAT 4-punish-INF Allah-OBL-ERG 3-do-OPT
 Curse: 'May God condemn (punish) them!'
 (Abdullaev 2015: 151)

Another form which comes up frequently in examples has the suffix *-iwu* (Desiderative II in Friedman 2021):

- (62) *ina a^rš:-ara-l o^q'-iwu*
 you.sg ground-OBL-ERG 1.swallow-OPT
 Curse: 'May you be swallowed up by the earth!'
 (Abdullaev 2015: 242)

It is often used with an additional element *-j* also found in vocatives (79 cases out of 96):

- (63) *š:al-li-l o^q'-iwuj*
 soil-OBL-ERG 1.swallow-OPT
 Curse: 'Go through the earth!' (literally 'May the earth swallow you up!')
 (Abdullaev 2015: 218)

The optative in *-iwu* can attach a nominal plural ending *-qul*, compare:

- (64) *ina c'ullu iwuj*
 you.sg healthy 1.be.OPT
 Blessing: 'May you be well!'
 (Rosa Eldarova pc)
- (65) *zu c'ullu b-iwu-qul*
 you.pl healthy HPL-be.OPT-PL
 Blessing: 'May you (plural) be well!'
 (Rosa Eldarova pc)

According to Eldarova (1999: 58), the form in *-iwu(j)* is a masdar with an optative meaning.

Finally, wishes can be expressed by second person imperatives. In Lak, as well as in some other Nakh-Daghestanian languages, along with their regular usage in commands (66), second person imperatives can be used with third person subjects to express wishes (67–68):

- (66) *ca balaj uča žu-n*
 one song say.IMP we-DAT
 'Sing us a song'
 (Kazenin 2013: 286)
- (67) *ka d-ahi wi-l*
 hand 4-fall.IMP you.SG-GEN
 Curse: 'May your hand fall!'
 (Ramazanov 2005: 106)
- (68) *čahar b-uč'a wi-l*
 letter 3-arrive.IMP you.SG-GEN
 Curse: 'May news of you come!' (i.e., death notification)
 (Ramazanov 2005: 205)

There is a tendency for such imperatives to take the suffix which typically marks intransitive imperatives even if the construction is transitive (Eldarova 1999: 56; Friedman 2013). There is no strict correspondence between the imperative suffix and transitivity in Lak, but the tendency is that intransitive imperatives have *-u*,

while transitive imperatives have *-a* or *-i* (Friedman 2013). In the following sentence, the verb *öq'an* is used. This verb can be intransitive, meaning 'drown', or transitive, meaning 'swallow' (Abdullaev 2015). The wish below is a transitive construction with ergative and nominative arguments. However, in example (69) the verb is used with the suffix *-u* as if it were intransitive:

- (69) *š:al-li-l o'q'u ina*
 soil-OBL-ERG 1.swallow.IMP you.sg
 Curse: 'Go through the earth!' (literally 'May the earth swallow you up!')
 (Abdullaev 2015: 218)

The same is true for the following example with the verb *biš:un* 'strike'. This verb usually has an imperative in *-a*: *r-iš:-a* 'strike!' (Eldarova 1999: 57). In the following wish, however, the verb is used with the suffix *-u*:

- (70) *t:u-j-χ-gu i<w>š:u-n š:a-χ iš:-u ina*
 I-SUP-TRANS-ADD <1>1.strike-CVB ground-AD 1.strike-IMP you.sg
 Curse: 'May you first hit me, then the ground' (cursing at the address of the offender)
 (Ramazanova 2005: 176)

The suffix *-a*, typical for transitive imperatives, can also occur in wishes:

- (71) *ka-ru burk'-unt:u-lu d-iš-a (wi-l)*
 hand-PL gravestone-OBL-SUB NPL-put-IMP (you.SG-GEN)
 Curse: 'May (your) hands be put under the tombstone!'
 (Ramazanova 2005: 109)

2.3.1 Distribution of the forms according to the value of the wish

Optatives in *-naw* and *-iwu(j)* can be used in the same context:

- (72) *žan c'ullu d-an-naw*
 soul healthy 4-do-OPT
 Blessing: 'May (you) be healthy!'
 (Ramazanova 2005: 54)

- (73) *wi-l žan c'ullu d-iwuj*
 you.sg-GEN soul healthy 4-be.OPT
 Blessing: 'May your soul be healthy!'
 (Ramazanov 2005: 55)

- (74) *ina c'ullu iwuj*
 you.sg healthy 1.be.OPT
 Blessing: 'May (you) be healthy!'
 (Rosa Eldarova pc)

There is however a distinction between them in terms of their value, visible in the collected data. The optative in *-naw* is much more often used for blessings, while the optative in *-iwu(j)* occurs significantly more frequently with negative value (curses) (the latter fact was also noticed by Murkelinskij 1971: 216) (see Table 6).

Wish-expressions with imperatives seem synonymous with other optative forms (compare (75) and (76), (77) and (78)):

- (75) *š:al-li-l o'q'u ina*
 soil-OBL-ERG 1.swallow.IMP you.sg
 Curse: 'Go through the earth!' (literally 'May the earth swallow you up!')
 (Abdullaev 2015: 218)

- (76) *š:al-li-l o'q'-iwuj*
 soil-OBL-ERG 1.swallow-OPT
 Curse: 'Go through the earth!' (literally 'May the earth swallow you up!')
 (Abdullaev 2015: 218)

- (77) *š:ajt'an-nu-l lasi wi-l o'rč-ru*
 shaitan-OBL-ERG take.IMP you.sg-GEN child-PL
 Curse: 'May shaitan take your children!'
 (Rosa Eldarova pc)

- (78) *wi-l o'rč-ru š:ajt'an-nu-l las-iwuj*
 you.sg-GEN child-PL shaitan-OBL-ERG take-OPT
 Curse: 'May shaitan take your children!'
 (Rosa Eldarova pc)

However, all examples with imperatives in my database denote a curse (see Table 6).

Table 6: Distribution of blessings and curses in Lak optatives

	blessings	curses	total
naw	96	25	121
wu, wuj	21	74	95
imperative	0	22	22

Therefore, all wish-forms in Lak tend to specialize in either blessings or curses.

2.3.2 Distribution of the constructions according to the value of wish

The two constructional properties which were found in Avar and Rutul – the presence of the words *Allah* or *God* and the presence of second person pronouns – are less pronounced in Lak. Wishes with the optatives in *-naw* and *-wu(j)* do not show value-induced preferences with respect to either of those constructional properties. Conversely, imperative constructions, which always express a curse, contain the second person pronoun in the final position in 14 cases out of 22; and, vice versa, all cases with the second person pronoun in the final position which exist in my database are imperatives (see Table 7).

Table 7: Second person pronouns in Lak blessings and curses.

		absent	final position	non-final position	total
-naw	blessings	81	0	15	96
	curses	20	0	5	25
-wu	blessings	5	0	16	21
	curses	47	0	27	74
imperative	blessings	0	0	0	0
	curses	8	14	0	22

To summarize, Lak manifests a tendency to specialize on blessings or curses on the level of morphology – all three forms which are typical for wishes show a preference for a certain value. Constructional tendencies are pronounced weaker in Lak than in Avar and Rutul. The second person pronoun non-final positions and the presence of the words ‘Allah’ and ‘God’ are found both in blessings and curses, while the final position of the second person pronoun is attested only in curses.

2.4 Second person pronoun in Karata, Botlikh and Godoberi wishes

Since the tendency to have explicit second person pronouns was less prominent in Lak than in Avar and Rutul, I decided to undertake a brief count in three other languages. Karata, Botlikh and Godoberi were chosen for practical reasons – there exist digitized versions of the dictionaries (Alekseev and Azayev 2019, Magomedova and Khalidova 2001, Saidova 2006, Saidova and Abusov 2012; Chuprinko, Moroz and Finkelberg 2023), and the number of examples of wishes is more than fifty for each language. All three languages belong to the same Andic branch of Avar-Ando-Tsezic languages.

We can see from Table 8 that a feature most strongly associated with curses is the final position of the second person pronoun. As for the presence of the words for ‘Allah’ and ‘God’, the number of examples found in the dictionaries of these languages was not representative.

Table 8: Second person pronouns in Karata, Botlikh and Godoberi blessings and curses.

languages		absent	final position	non-final position	total
Karata	blessings	65	3	3	71
	curses	12	11	4	17
Botlikh	blessings	30	7	8	45
	curses	19	18	9	46
Godoberi	blessings	37	7	5	49
	curses	3	7	0	10

3 Discussion and conclusion

Before summarizing the results and drawing conclusions, I want to discuss certain limitations of this study. As was explained in the beginning of the paper, I used the dictionary examples to examine wish expressions. These data might be skewed for several reasons. First, the examples in the dictionaries may reflect the personal preferences of the authors. Second, dictionary examples serve specific aims (demonstrating the usage of words), and the words that are not necessary for these aims can be omitted by the authors (including second person pronouns or words for ‘God’ and ‘Allah’). For example, in the small corpus of dialectal Rutul the word ‘Allah’ was found in 20% of the examples of blessings (5 out of 23), while in the dictionary of standard Rutul the proportion is 10%. However, corpora of a size which is sufficient for the quantitative study of wishes do not yet exist for any of Nakh-Daghestanian languages.

I consider curses to be a type of linguistic / interactional impoliteness. The aim of this study was to test whether there are structures that are associated with impoliteness realized as curses. As I showed in this paper, the answer is positive: there are structural properties of wishes which exhibit a bias towards impoliteness in every studied language.

First, in all three languages – Avar, Rutul and Lak – there are forms that are associated primarily with curses: optatives in *-ad* in Avar, *-di* in Mukhad Rutul, *-wu(j)* in Lak and second person imperatives in Lak. It is noteworthy that three forms out of four are morphologically nominal. The Avar form in *-ad* and the Lak form in *-iwu(j)* both inflect for case and number and thus may be qualified as masdars (verbal nouns) dedicated to the expression of wishes. The Mukhad Rutul form in *-di* originates from a participle and keeps the ability to inflect for case and number. This tendency needs to be tested in other languages of the family. If the association between verbal nominal forms and negative value is confirmed, this might be another structural feature related to impoliteness.

Second, three constructional features were found mainly in curses: the absence of the words for ‘God’ and ‘Allah’, the presence of second person pronouns and their final position in the utterance. Only the last feature was confirmed for all three languages.

With regard to the first feature, the usage of words for ‘God’ and ‘Allah’, I did not come across similar evidence in languages outside Daghestan. Although the presence of the word ‘god’ in wishes has been documented in several West African languages (Creissels 2022, Dombrowsky-Hahn and Francesco Zappa 2024), I have found no evidence suggesting that this phenomenon is more prevalent in blessings than in curses. Residents of Daghestanian villages often claim that cursing is not godly; although curses are widely used in certain situations, cursing is considered as committing sin. This may be the reason why Allah is not usually mentioned in curses.

The second feature, the presence of the second person pronouns, complies with the observations made for some European languages. As discussed in Van Olmen, Andersson and Culpeper (2023), the second person pronouns are superfluous in an address. Even without the presence of ‘you’, wishes are understood as addressed to the listener; it is the third rather than the second person that has to be specifically mentioned. The explicit expression of the second person increases the effect of the curse, in the same line as in the construction *YOU+NP* as an insult (e.g. *you (stupid) idiot*), studied in Corver (2008) and Van Olmen, Andersson and Culpeper (2023).

The most robust feature of the curses turned out to be the final position of the second person pronoun. It was found in all three languages in the study. Moreover, in Mukhad Rutul the final second person pronoun was most likely the source of the development of the optative affix *-di*. The post-verbal position is highly atypical for subject pronouns in Nakh-Daghestanian languages. In the corpus of Rutul, the imper-

ative followed by subject pronoun occurs only once; apparently, the word order has a contrastive value ('you go, I stay'). Probably, the post-verbal position of second person pronouns in curses is similarly motivated by emphatic reinforcement.

Finally, an important finding of this study is that both morphological and constructional features attested in curses manifest themselves only as tendencies. There are no structures that are associated with curses on the grammatical level: even when the data show a hundred per cent association with impoliteness, language consultants readily produce examples of blessings with the same structure, which means that evaluation is still partly a matter of the communicative context and the lexical meaning of the words. This may be one reason why forms and constructions that tend to express curses are usually overlooked by the grammarians.

The expression of curses studied in this paper illustrates the process of conventionalization, which, as was shown in Terkourafi (2005), is a matter of degree, and may well vary for different speakers, as well as for the same speaker over time. A similar conclusion is drawn by Van Olmen, Andersson and Culpeper (2023): in their study of *YOU*+NP in Dutch, English and Polish they show that impoliteness can be strongly conventionalized in a similar way across languages, but it is never completely conventional. The study of curse expressions in Nakh-Daghestanian languages, carried out in this paper, is consistent with this finding. The data revealed numerous forms and constructions expressing curses almost exclusively, but not entirely exclusively.

This observation supports the results in Dobrushina (2024): the cases of Kumyk and Nogai inflectional optatives and Russian optative constructions showed that grammatical items that have evaluative usages often exhibit a "fluid" value whose positive or negative character needs to be specified by their contexts. Even if the majority of examples is associated with negative value, there could be infrequent examples with positive value. Qualitative evaluation (positive or negative) emerges as a contextually defined speaker's perspective on a given situation. It comes as a satellite to some other meaning, and even if it specializes towards negative or positive attitude, it can still maintain its unstable character and display the opposite meaning in a certain context, as it happens in all cases considered in this paper.

Abbreviations

- 1 first gender (masculine)
- 2 second gender (feminine)
- 3 third gender (non-human)
- 4 fourth gender (non-human)

ABL	ablative
AD	adessive
ADD	additive
ADJ	adjectivizer
AOR	aorist
ATTR	attributivizer
CONC	concessive
CNT	continuative
CVB	converb
DAT	dative
DEM	demonstrative
EL	motion from a spatial domain
ERG	ergative
ESS	essive
G	gender
GEN	genitive
HPL	human plural
I	first person
IMP	imperative
INF	infinitive
INTJ	interjection
IPFV	imperfective
IRR	irrealis
LAT	motion into a spatial domain
MAL	maledictive
NEG	negative
NMLZ	nominalization
NPL	non-human plural
OBL	oblique stem
OPT	optative
PFV	perfective
PL	plural
PROH	prohibitive
PST	past
PTCL	particle
PTCP	participle
PV	preverb
RE	refactive
REFL	reflexive
SG	singular
SUB	subessive
SUP	spatial domain on the horizontal surface of the landmark
TRANS	translative

Acknowledgements: I am deeply grateful to Karim Musaev, Zulaikhat Mallaeva, Alisultan Alisultanov, Mariza Ibragimova and Rosa Eldarova for their consultations in Archi, Avar, Rutul and Lak. I am also grateful to Michael Daniel and Timur Maisak for discussing some of the preliminary findings of the study with me. Finally, I thank the editors and two anonymous reviewers for their helpful comments on this paper.

References

- Abdullaev, Saigid. 1954. *Grammatika darginskogo jazyka* [Grammar of Dargwa]. Makhachkala.
- Abdullaev, isa X. 2015. *Laksko-russkij slovar'* [Lak-Russian dictionary]. Makhachkala: IJALI.
- Aikhenvald, Alexandra Y. 2020. "Damn your eyes!" (Not really): Imperative imprecatives, and curses as commands. In Nico Nassenstein & Anne Storch (eds.), *Swearing and cursing*, 53–78. Berlin: De Gruyter Mouton.
- Alekseev, Mikhail E., Boris M. Ataev, Magomed A. Magomedov, Magomed I. Magomedov, Gulzhagan I. Madiyeva, Patimat A. Saidova & Dzhalil S. Samedov. 2012. *Sovremennyy avarskij jazyk* [Contemporary Avar language]. Makhachkala: Aleph.
- Alekseev, Mikhail E. & Khalil G. Azayev. 2019. *Botlixsko-russkij slovar'* [Botlikh-Russian dictionary] / Moskva: Academia.
- Alekseeva, Anastasija. 2024. Attributive markers. In Anastasija Alekseeva, Nikita Beklemishev, Michael Daniel, Nina Dobrushina, Konstantin Filatov, Anastasija Ivanova, Timur Maisak, Maxim Melenchenko, George Moroz, Ivan Netkachev & Ilya Sadakov. *Atlas of Rutul dialects*. Moscow: Linguistic Convergence Laboratory, HSE University. https://lingconlab.github.io/rutul_dialectology/index.html
- Alisultanov, Alisultan S. & Taisija A. Sulejmanova. 2019. *Russko-rutul'skij slovar'* [Russian-Rutul dictionary]. Makhachkala: Alef.
- Berg, Helma van den. 2005. The East Caucasian language family. *Lingua* 115(1). 147–190.
- Bybee, Joan L., Revere Dale Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.
- Chirikba, Viacheslav A. 2008. The problem of the Caucasian Sprachbund. In Pieter Muysken (ed.), *From linguistic areas to areal linguistics*, 25–94. Amsterdam: Benjamins.
- Chuprinko, Kirill, Georgij Moroz & Leja Finkelberg. 2023. Comparative Andic dictionary database, v. 0.0.2 Moscow: Linguistic Convergence Laboratory, HSE University. [cited 14/06/2024]. https://github.com/LingConLab/andic_dictionary_examples_database.
- Corver, Norbert. 2008. Uniformity and diversity in the syntax of evaluative vocatives. *Journal of Comparative German Linguistics* 11. 43–93.
- Creissels, Denis. 2022. Directive and optative clauses in Manding languages. *Mandenkan* 69. 3–29.
- Culpeper, Jonathan, 2011. *Impoliteness: Using Language to Cause Offence*. Cambridge: Cambridge University Press.
- Dobrushina, Nina. 2011. The optative domain in East Caucasian languages. In Gilles Authier & Timur Maisak (eds.), *Tense, aspect, modality and finiteness in East Caucasian languages*, 95–130. Bochum: Brockmeyer.

- Dobrushina, Nina. 2019. Moods of Mehweb. In Michael Daniel, Nina Dobrushina & Dmitry Ganenkov (eds.), *The Mehweb language: Essays on phonology, morphology and syntax*, 117–165. Berlin: Language Science Press.
- Dobrushina, Nina. 2024. Evaluation between grammar and context: The case of blessings and curses. *Diachronica*. 41(4).
- Dobrushina, Nina, Johan van der Auwera & Valentin Goussev. 2013. The Optative. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/73>, accessed on 2019-09-26.)
- Dombrowsky-Hahn, Klaudia & Francesco Zappa. 2024. “May God preserve us from evil sicknesses”: Grammar, Arabic origin, and pragmatics of request prayers in Bambara (Mande). *Linguistique et langues africaines* 10(1). 1–28.
- Eldarova, Rosa G. 1999. *Lakskij glagol (sistema glagol'nogo slovoizmenenija)*. [Lak verb. The system of verbal forms]. Makhachkala: IPC DGU.
- Forker, Diana, & Oleg Belyaev. 2016. Word order and focus particles in Nakh-Daghestanian languages. In Jocelyne M.M. Fernandez-Vest & Robert D. Van Valin (eds.), *Information structuring of spoken language from a cross-linguistic perspective*, 239–262. Berlin: De Gruyter Mouton.
- Friedman, Victor. 2013. The morphology of imperatives in Lak: Stem vowels in the second singular simplex transitive affirmative. In Balthasar Bickel, Lenore A. Grenoble, David A. Peterson & Alan Timberlake (eds.), *Language typology and historical contingency: In honor of Johanna Nichols*, 445–462. Amsterdam: Benjamins.
- Friedman, Victor. 2021. Lak. In Maria Polinsky (ed.), *The Oxford handbook of languages of the Caucasus*, 201–241. Oxford: Oxford University Press.
- Gadzhiahmedov, Nurmagomed E. 2000. Slovoizmenitel'nye kategorii imeni i glagola v kumykskom yazyke [Inflectional categories of noun and verb in the Kumyk language]. Makhachkala: Jupiter.
- Gimbatov, Magomed-Kamil M. 2006. *Avarsko-russkij slovar'* [Avar–Russian dictionary]. Unpublished manuscript. <http://avar.me/>
- Giomi, Riccardo & Denise van Oers. 2022. Insultive constructions: A crosslinguistic perspective. Paper presented at Societas Linguistica Europaea 55, University of Bucharest, 24–27 August.
- Kazenin, Konstantin I. 2013. *Sintaksis sovremennogo lakskogo jazyka* [Syntax of contemporary Lak]. Makhachkala: IJaLI.
- Khaidakov, Said. M. 1961. *Očerki po leksike lakskogo jazyka* [Essays on the lexicon of the Lak language]. Moscow: Izdatel'stvo akademii nauk SSSR.
- Koryakov, Yuri B. 2002. *Atlas of the Caucasian languages*. Moscow: Institut jazykoznanija RAN.
- Magomedova, Patimat & Rashidat Khalidova. 2001. *Karatinsko-russkij slovar'* [Karata-Russian dictionary]. Makhachkala: IJaLI.
- Magomedkhanov, Magomedkhan M. 1993. *Avarsko-russkij frazeologičeskij slovar'* [Avar-Russian phraseological dictionary]. Moskva: Russkij jazyk.
- Mallaeva, Zulaikhat M. 2007. *Glagol avarskogo jazyka: Struktura, semantika, funkcii* [The verb in Avar: Structure, semantics, functions]. Makhachkala: IJaLI.
- Makhmudova, Svetlana M. 2002. *Grammatičeskie klassy slov i grammatičeskie kategorii rutul'skogo jazyka* [Grammatical word classes and grammatical categories of Rutul]. Makhachkala: Dagestan State University PhD dissertation.
- Makhmudova, Svetlana M. 2016. *Frazeologičeskij slovar' rutul'skogo jazyka* [Rutul phraseological dictionary]. Makhachkala: Epokha.
- Murkelinskij Gadzhi B. 1971. *Grammatika lakskogo jazyka*. [Grammar of Lak] Makhachkala: Dagestanskoe učebno-pedagogičeskoe izdatel'stvo.

- Musaev, Magomed-Said M. 2002. *Darginskij jazyk* [The Dargwa language]. Moscow: DGU, Akademija.
- Nikolaeva, Irina. 2016. Analyses of the semantics of mood. In Jan Nuyts & Johan van der Auwera (eds.), *The Oxford handbook of modality and mood*, 68–88. Oxford: Oxford University Press.
- Ramazanova, Mijasat R. 2005. *Laksko-russkij frazeologičeskij slovar'* [Lak-Russian phraseological dictionary]. Makhachkala.
- Saidova, Patimat A. 2006. *Godoberinsko-russkij slovar'* [Godoberi-Russian dictionary]. Makhachkala: DNC RAN Institut jazyka, literatury i iskusstva im. G. Cadasy i Institut évoljucionnoj antropologii im. M. Planka.
- Saidova, Patimat A. & Magomed G. Abusov. 2012. *Botlixsko-russkij slovar'* [Botlikh-Russian dictionary]. Makhachkala: DNC RAN Institut jazyka, literatury i iskusstva im. G. Cadasy i Institut évoljucionnoj antropologii im. M. Planka.
- Sumbatova, Nina R. & Jury A. Lander. 2014. *Darginskij govor selenija Tanty: Grammatičeskij očerk, voprosy sintaksisa* [The Dargwa Variety of the Tanti Village: A grammatical sketch. Aspects of syntax]. Moscow: Jazyki slavjanskoj kul'tury.
- Terkourafi, Marina. 2002. Politeness and formulaicity: Evidence from Cypriot Greek. *Journal of Greek Linguistics* 3. 179–201.
- Terkourafi, Marina. 2005. 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.
- Testelefs, Jakov. n.d. Avar. Unpublished manuscript.
- Tuite, Kevin. 1999. The myth of the Caucasian Sprachbund: The case of ergativity. *Lingua* 108(1). 1–29.
- 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.
- Visser, Eline. 2022. *A grammar of Kalamang*. Berlin: Language Science Press.
- Wixman, Ronald. 1980. *Language aspects of ethnic patterns and processes in the North Caucasus*. Chicago: University of Chicago.
- Wray, Alison & Michael R. Perkins. 2000. The functions of formulaic language: An integrated model. *Language & Communication* 20(1). 1–28.

Annick Paternoster

5 *Che ti venga* NP, a conventionalised impoliteness formula for Italian disease curses (14th–20th century)

Abstract: This essay argues that the Italian formula *che ti venga* NP is a conventionalised linguistic expression of impoliteness, particularly as a disease curse, from the fourteenth to the twentieth century. Still used in contemporary Italian, examples such as *che ti venga il gavocciolo* ('may a plague sore take you') appear in medieval legislative texts and judicial records. Theologically, these curses were deemed sinful and blasphemous.

Drawing on two historical corpora – the *CORpus Diacronico dell'Italiano* (CODIT) and an eighteenth-century corpus of Carlo Goldoni's dramatic oeuvre – the study finds 132 instances of the curse. Quantitative analysis examines subjunctive and pronoun use, word order, intensifiers, and disease nouns, revealing a preference for severe illnesses like the plague and rabies. The qualitative analysis focuses on rare benedictive uses, reinforcing the hypothesis that *che ti venga* NP is predominantly impolite. Seventeenth- and eighteenth-century pedagogical texts provide further metapragmatic evidence of its conventionalised status. This interdisciplinary approach fills a research gap by highlighting the syntactic stability of the *Che ti venga* NP phrase over centuries, paving the way for cross-cultural comparisons of similar expressions.

Keywords: *Che ti venga* NP, disease curses, cursing, Italian, impoliteness, conventionalised expressions, maledictions, swearing

1 Introduction

This paper was prompted by an analysis of *impoliteness* in the theatre play *La Lena*, written by Ludovico Ariosto and performed in 1528 at the Este court in Ferrara (Paternoster 2015: 131–165). Renaissance comedies portray conflict, disorder, corruption, etc. outside the court, with happy endings celebrating ducal rule. Unsurprisingly, the genre is rich in impolite exchanges. *Lena*, an aging prostitute

Annick Paternoster, University of Lugano, Istituto di Studi italiani, Palazzo nero, Via Buffi 13, 6900 Lugano, Switzerland, e-mail: annick.paternoster@usi.ch

exploited both by her husband pimp and her lover, directs verbal abuse at anybody thwarting her attempts at making some extra money. The play includes eight occurrences of a *disease curse*, four pronounced by Lena. Examples (1) to (8) come from Ariosto (2007 [1528]):¹

- (1) *Deh, ti venga il malanno!*
oh 2SG.DAT come.SBJV.3SG the calamity
Oh, a calamity strike you! [lit. 'Oh, that to you may come the calamity!']
- (2) *Oh, che ti venga il mal di Santo Antonio!*
oh COMP 2SG.DAT come.SBJV.3SG the disease of Saint Anthony
Oh, Saint Anthony's fire get you! [lit. 'Oh, that to you may come the disease of Saint Anthony!']
- (3) *Cancar ti venga*
canker 2SG.DAT come.SBJV.3SG
'May you rot'² [lit. 'Canker to you may come!']
- (4) *Brutto impiccato, che ti venghi il cancaro!*
nasty hanged.man.VOC COMP 2SG.DAT come.SBJV.3SG the canker
'Nasty man, you deserve to be hanged, may you rot! [lit. 'Nasty hanged man, that to you may come the canker!']
- (5) *Che la fistola Ti venga!*
COMP the fistula 2SG.DAT come.SBJV.3SG
'A running fistula take you! [lit. 'That to you may come the fistula!']
- (6) *Deh, manigoldo, ti venga la fistula!*
oh hangman.VOC 2SG.DAT come.SBJV.3SG the fistula
'Oh, hangman, a running fistula take you! [lit. 'Oh, hangman, to you may come the fistula!']
- (7) *Doh, che ti venga il morbo!*
oh COMP 2SG.DAT come.SBJV.3SG the plague
'Oh, the plague take you!' [lit. 'Oh, that to you may come the plague!']

¹ In a handful of longer examples, only the part of the example relevant to the curse will be glossed. Respectful uses of the pronouns *voi/vi/vostro* addressing a single recipient are glossed with POL. I provide a free and a literal translation; all translations are mine. Rather than translating with a medical term, I use the historical disease term to maintain the negative valence associated with it. The most helpful bilingual dictionary was Florio (1598).

² *Cancar* indicates a festering or ulcerative sore, a canker. This translation occurs in Palermo (1755: 342).

- (8) *Che venir possa il morbo a mastro Lazzaro*
 COMP come.INF may.SBJV.3SG the plague to master Lazzaro
 ‘May the plague take master Lazzaro’ [lit. ‘That to master Lazzaro may come the plague’]
Che m’ Arrecò alle man questa casipula!
 COMP 1SG.DAT bring.PST.3SG to the hands this hovel
 ‘who got me this hovel’ [lit. ‘who brought me into the hands this hovel!’]

Angry characters wish for diseases – including the plague – or severe misfortune to happen to someone else. Note the recurrent use of the verb *venire* ‘to come’, in the present subjunctive, to express a wish, with or without the conjunction *che* ‘that’. The disease is the subject of the verb, translating to ‘may this disease come to you!’. Less literal translations are ‘this disease get/take you!’. The indirect object is expressed with the personal pronoun of the 2nd person singular *ti* ‘you’, but example (8) uses a Noun Phrase indicating a third person *mastro Lazzaro* ‘master Lazzaro’. Also in (8), *venire* ‘to come’ is used with the modal verb *potere* in the present subjunctive: *possa venire* ‘may come’. These examples suggest *che ti venga* NP is a *conventionalised impoliteness formula* for disease curses (Culpeper 2011: 120–139; see also Culpeper, Van Dorst and Gillings, this volume). I test this hypothesis in two historical corpora.

Research into impoliteness in Old Italian utilises judicial records from central Italian communes (e.g., Marcheschi 1983; Breschi 1994; Fantappiè 2000; Larson 2004). When victims reported verbal abuse to a notary, the offence was noted in Latin, but from the thirteenth century the offensive words were increasingly quoted in the vernacular. *Verba iniuriosa* (literally, ‘unlawful words’, from Latin *in-ius* ‘not-law’) cover insults, strong criticism, curses, and defamation. Historical insults are well researched. Although historians are interested in how insults interact with law, hierarchy and gender, they focus on semantics to uncover the values of Italian medieval society, i.e. purity, loyalty and, mainly, honour (Burke 1987; Lesnick 1991; Dean 2007; Vise 2015, 2025; Lett 2018; Raveggi 2018).³ Dardano et al. (1992) and Alfonzetti and Spampinato Beretta (2012) pioneered studies on the syntactic structure of insults, listing:

- derogatory nouns, which may be accompanied by adjectives: *Sossa puttana marcia* [dirty rotten whore] (Dardano et al. 1992: 13);
- a declarative *tu sei* ‘you are’ with emphasis on the personal pronoun: *Tu se’ uno mentechatto* [you are mad] (Dardano et al. 1992: 15)

³ For a semantic approach on historical insults outside Italy, for French see Delumeau (1989); Gonthier (2007); for Old Frisian Bremmer (1998); for Spanish Madero (1992). Lagorgette (1994, 2003) on French and Jucker and Taavitsainen (2000) on English use a pragmatic approach.

- multifunctional connective *che* ‘that’, for a relative clause or complement clause (or both): *Bastardo mulo che tu se’* [A bastard mule that you are] (Dardano et al. 1992: 14)
- rhetorical questions: *No’ avete voi vergogna d’accompagnare le puctane?* [Are you not ashamed of accompanying the whores?] (Dardano et al. 1992: 16).⁴

Since neither Dardano et al. (1992) nor Alfonzetti and Spampinato Beretta (2012) systematically explore curses (imprecations of ill-fortune, *maledictions*), this study fills an empirical research gap.

Section 2 starts by placing curses within current impoliteness research to differentiate them from self-curses, oaths, promises, insults, swearing, and threats. However, whereas contemporary studies examine secular cursing, historical curses invoked supernatural powers and therefore they constituted a different speech act. This section zooms in on the link with theology and the legal history of *verba iniuriosa* in medieval Italy. Section 3 investigates *conventionalisation* through quantitative analysis using two historical corpora, where *che ti venga* NP is mostly maledictive. The few benedictive uses are qualitatively analysed in Section 4. Section 5 finds metapragmatic evidence of conventionalisation in two textbooks for language learning. Section 6 offers concluding remarks and pointers for future research directions.

2 Cursing

2.1 Present-day cursing

The legal perspective highlights the difference between past and present. Historically, curses were punishable by law (see Section 2.3), and their legal consequences today remain a topic of discussion on Italian legal advice websites and forums.⁵ The key point is that the law is not superstitious (Greco 2021); for Italian lawmakers, a curse cannot cause the wished-for disease and the curser has no agency. While cursing is not a criminal offence per se, it can be interpreted as defamation or a threat, both punishable by law. How to unravel these aspects of cursing? What is the link with superstition and how can curses be distinguished from other speech acts?

⁴ On insults in present-day Italian see Alfonzetti (2009, 2017) and Domaneschi (2020).

⁵ <https://forum.finanzaonline.com/threads/e-reato-dire-ti-maledico-a-qualcuno.1920926/>; <https://it.diritto.narkive.com/EZnc0uXl/insultare-e-reato-ma-maledire>; <https://it.quora.com/È-reato-augurare-la-morte-o-una-brutta-malattia-una-scongiura-ecc-a-qualcuno> (accessed 8 January 2025).

Firstly, how do curses fit into impoliteness research? Culpeper (2011: 136) considers present-day curses – such as *damn you!* – conventionalised impoliteness expressions or formulae, routinely evaluated as impolite (see also Dobrushina, this volume, who links forms, impoliteness and curses in Nakh-Daghestanian languages). Impoliteness formulae suggest that “some words and structures are more regularly perceived as impolite than others”.⁶ Terkourafi’s frame-based politeness approach (e.g. 2001, 2005) first theorised the role of conventionalisation, “understood as a three-way relationship between an expression, a context and a speaker”, that is, “an expression is conventionalised for some use relative to a context for a speaker if it is used frequently enough in that context to achieve a particular illocutionary goal to that speaker’s experience” (Terkourafi and Kadar 2017: 182). Thus, frequent use of a polite expression in a particular context makes politeness its default meaning. However, impoliteness may be conventionalised in different ways. Although less frequent, it is more noticeable, it is typically challenged, it is also shaped by exposure beyond direct interactions: “people acquire a knowledge of impoliteness formulae that far exceeds *their own direct experience of usage* of formulae associated with impolite effects”, i.e., by “indirect experience, and in particular metadiscourse” (Culpeper 2011: 130–132, italics in original). Examples of conventionalised curses include “[go] [to hell/hang yourself/fuck yourself]; [damn/fuck] [you]” (Culpeper 2011: 136). Curses wish for something bad to happen to someone, and that is why they are offensive, but in a secular Western European context they are not believed to cause actual harm. With Culpeper (2011: 23) I see impoliteness as a negative evaluation of a behaviour, verbal or not, that is perceived to be offensive or inappropriate, given the expectations of how someone would like to be treated in a certain context. Whether these expectations are more individual or social in nature, offensive behaviour tends to cause strong emotions (Culpeper and Haugh 2021).

Investigating the recurrent phraseology of late-modern French and Spanish votive formulae, López-Simo (2023) distinguishes curses from neighbouring speech acts. Curses are classified as interpersonal routine formulae (López-Simo 2023: 22). *Swearing*, like *Diable!* [What the devil!], is therefore not included in her study, because it does not typically address someone (López-Simo 2023: 23; Lagorgette 2003).⁷ Blessings or benedictions wish for something positive to happen to someone (e.g., *Que Dieu vous garde* [may God protect you]); curses or maledictions wish for something negative (e.g., *Dieu te maudisse* [God damn you], López-Simo 2023: 24). Curses differ from insults, which directly address the recipient with a negative voc-

⁶ <https://www.lancaster.ac.uk/fass/projects/impoliteness/forms.htm> (accessed 8 January 2025)

⁷ Swearing can be offensive if perceived as disrespectful or contextually inappropriate. Culpeper (2018) notes that taboo words primarily intensify impoliteness.

ative (López-Simo 2023: 25)⁸. Curses and threats differ in agency (Duranti 2004). Both want harm to befall the interlocutor, but a threat involves the speaker committing to cause the harm, whilst a curse implies no personal agency. Timing also differs: threats must seem urgent to be maximally efficacious; maledictions may take effect in an indeterminate future (on urgency in threats see Limberg 2009).

Secondly, outside present-day urban society, curses may entail a supernatural or cosmological aspect. Kádár and Szalai (2020) study contemporary ritual curses in the Gabor Roma communities (Transylvania, Romania). The Roma use curses in a supernatural, i.e., an “archetypal” way, attributing a “harmful effect to them” (Kádár and Szalai 2020: 16–17). Examples are “May your mother die! May your grandmother die!” (Kádár and Szalai 2020: 28).⁹ These curses are not to be confused with self-curses. Like López-Simo (2023: 26), Szalai (2023) does not see self-curses as a rite of aggression; rather, they are relationship-forging: the Roma wish harm unto themselves if the recipient does not accept an offer of food, as in the example “May my father die, if you don’t drink that coffee!” (Kádár and Szalai 2020: 16). Self-curses are oaths or promises. In the Old Testament “an oath, or a solemn promise, required an act of conditional self-cursing, or calling upon divine power to inflict dire punishment should the promise be broken” (Ramos 2015: 2). When cursing, the Roma invoke supernatural or cosmological powers to wish harm upon others (and occasionally upon themselves), likely reflecting an old tradition rooted in widely held beliefs that certain words, and specifically curses, possessed the power to cause real-world effects. In short, curses were believed to have magic power.

2.2 Historical cursing

Indeed, historical and contemporary curses fundamentally differ in terms of the speech act involved. Nowadays, curses may aim to intimidate and act as directives, but primarily they fall into the category of expressives, serving as expressions of anger or irritation (Searle 1979). The sincerity condition for an expressive curse is that the speaker genuinely feels the emotion or psychological state conveyed. In contrast, historical curses were believed to bring harm simply by being spoken. While sincere emotions might have accompanied the curse, these feelings were, so to speak, irrelevant to its *efficacy*. Early-Modern witchcraft trials in England

⁸ For Alfonzetti and Spampinato Beretta (2012: 2) insults directly target the interlocutor. For Culpeper (2011: 135) formulaic insults also include “personalized third-person negative references (in the hearing of the target)”.

⁹ Cursing a mother and grandmother in front of their (grand)child seems cruel, but the curses are used teasingly, in a safe space where Roma children learn about cursing (Kádár and Szalai 2020: 19–20).

offer a telling example. Culpeper and Semino (2000: 102) note how accusations of witchcraft were often based on the presence of a curse: when an older, poor woman expressed anger by wishing harm on those who had wronged her, “her words could be interpreted as a witch’s curse and be subsequently used against her in court”. In such cases, the interpretation of the speech act ‘curse’ relies on the preparatory condition that the alleged witch is believed to have a pact with a supernatural power such as the devil (Culpeper and Semino 2000: 107–109). Witches’ curses and historical curses more broadly (Danet and Bogoch 1992: 136) align with Searle’s category of declarations (1979: 16–17), where the mere act of speaking the curse was believed to bring harm. This belief is so deeply rooted that, given the appropriate context – such as when spoken by a marginalised elderly woman – a simple outburst of anger could be perceived as a declaration of supernatural hostility (see also Arnovick 1999).

Cursing dates to the ancient world. Kitz (2007: 615, 2014) notes that in the Ancient Near East deities played a central role in executing maledictions since they were viewed “as the agents who would, in one way or another, realize the requested punishment should they judge in the speaker’s favor”. In the Greco-Roman world curses were often written on lead tablets, known as *katadesmos* in Greek or *tabella defixionis* in Latin, invoking Gods to bring the cursed under the curser’s power or inflict harm or disease (Gager 1992: 21); tomb epitaphs contained curses to deter looters. The Bible contains many curses, and in the early Middle Ages Benedictine monks of northern France pronounced liturgical curses against Vikings or unruly feudal lords (Little 1993). Danet and Bogoch (1992: 132) analyse ‘whoever’ curses commonly found in Anglo-Saxon legal documents, “wills, grants or leases of land, and some royal writs”, which were included to deter anyone from violating the writer’s wishes. ‘Whoever’ curses were frequently included in medieval manuscripts to warn against tampering with the document (Baker 2023).

Vecchio (2014) offers a valuable theological distinction between vertical and horizontal dimensions of a curse, affecting either God or other humans. The Church Fathers were primarily concerned with the vertical dimension: when is a curse legitimate and thus efficacious? For sixth-century Gregory the Great, curses were only legitimate when pronounced by a serene judge acting with the righteousness of God (Vecchio 2014: 353). This view is closely tied to the concept of excommunication, where a divine pact grants the Church authority to exclude the damned from eternal salvation (Vecchio 2014: 357). However, private curses driven by anger or vengeance were considered sinful. From the thirteenth century onward, the focus shifted to the horizontal or social dimension of curses, condemning them for their lack of regard for fraternal love (Vecchio 2014: 360). Cursing God’s creations was seen as a violation of the Second Commandment (Fisher 1908). Thomas Aquinas in the *Summa Theologiae* (1920–1942: II-II, q. 76 a. 1) explains that if a man “desires another’s evil, as evil, being intent on the evil itself, then evil speaking will be

unlawful [. . .], and this is what is meant by cursing” (see Casagrande and Vecchio [1987] on the sins of the tongue). Wishing evil on someone was seen as a mortal sin.

Both swearing and cursing were considered sinful, but they also risked being prosecuted as acts of *blasphemy* by the Inquisition, which was particularly active in the latter half of the sixteenth century and throughout the seventeenth century. While the Church showed a certain tolerance for common swearing, the boundary between casual and heretical swearing – which carried severe punishments – was never clearly defined (Biasiori 2024). When cursing someone, the curser implicitly invoked God to carry out the negative wishes. This act was indirectly blasphemous because it usurped divine authority, challenging God’s exclusive right to decide whether to punish someone. The belief in the vertical (blasphemous) implications of cursing is still echoed in the eighteenth-century Italian manual *L’Uomo apostolico* (Maria 1732), a guide for confessors.¹⁰ It includes an *Esortazione X* [Exhortation X] addressing mothers prone to cursing their children, citing examples such as “Che ti venga la Peste; che ti possa rompere il Collo; Che il Diavolo ti porti, ec.” [The plague take you; may you break your neck; may the Devil take you] (Maria 1732: 290):

[. . .] essendo Egli, che per i suoi Giusti Giudizj manda la Peste, la Morte, e gli altri Mali di pena, pare, che si voglia obbligarlo ad eseguire il male iniquamente imprecato. [since it is He, who by His Righteous Judgments sends the Plague, Death, and the other punishing Evils, it appears, that one wants to force Him to execute the evil unjustly imprecated.] (Maria 1732: 290)

The curser is sinfully usurping the role of God. The belief that curses offended both God and the cursed persisted, in Italy, at least into the eighteenth century, but note that *L’Uomo apostolico* enjoyed numerous editions up till the nineteenth century.

2.3 Cursing in Italian communal statutes and trial records

Recognising the potential of curses to disrupt social harmony, secular authorities also sought to regulate their use. Most statute books of fourteenth-century Italian communes covered *verba iniuriosa* (Tardivel: 2020: 303–304). Statute books (or *statuti*) were official collections of laws and regulations that governed city life and they often included specific prohibitions on language, detailing words and phrases that were considered offensive. In a fourteenth-century statute from Le Marche d’Ancona, a province within the Papal State, Tardivel (2019: 92–93) identifies four types of verbal offences: insults (*iniuria*); reproaches (*reimproperatio*), maledictions (*blasfemia*) wishing future harm. The statute lists the following curses:

¹⁰ There exist earlier 1726 and 1727 editions.

- (9) *che si impeccato* [may you be hanged]
- (10) *che te venga languenaglia* [tongue paralysis take you]
- (11) *che te venga lo carbone* [the carbuncle take you]
- (12) *che vencha lo male de la cadia* [falling-sickness take you]
- (13) *che te venga l'antrace* [a carbuncle take you]¹¹
- (14) *che venga la lepra o vero la malsania* [leprosy or an illness take you] (Tardivel 2019: 93)

Five curses use *che ti venga* NP. Tardivel (2019: 94) notes these curses invoke the worst diseases, highlighting their blasphemous nature as they question the universe created by God. The fourth type of *verba iniuriosa* is calumny.

How are the communal lawtexts applied in the courts of justice? Along with insults, “cursing was certainly the privileged form of verbal abuse in medieval and early modern times” and “the common form of cursing was a disease curse” (Tardivel forthcoming). Studying Bolognese trial documents, Tardivel finds 21 disease curses. A few examples:

- (15) *Te nasca el vermo chano tu e perzerto cativo ribaldo* [the staggers take you, you are a depraved nasty rogue]
- (16) *putana marça quod naschavit tibi mille vermi canes, tu es bastarda et fuisti filia unus meretricis, tu es excommunicata ab ore domini pape, soça putana* [rotten whore may you get a thousand times the staggers, you are a bastard and the daughter of a whore, you are excommunicated through the Pope's mouth, filthy whore]
- (17) *O te nascha la postema in le gola tu vo* [Oh, may you get a pus-filled abscess in your throat]
- (18) *Che te vengina le posteme de santo Antonio in la gola* [May you get St. Anthony's pus-filled abscesses in the throat]

¹¹ Anthrax causes carbuncles, seen in (11).

- (19) *Che dio te dia le mala Pasqua* [God give you misfortune] (Tardivel forthcoming)¹²

Note the use of *nascere* ‘to be born’ as an alternative for *venire* ‘to come’, along with intensifiers like interjections, insults, hyperbole, and body parts (*in la gola*). The presence of these curses in trial records (as mediated by court clerks) and statutes is significant. While my findings from historical corpora appear in literary dialogues, Tardivel’s examples show that the disease curse *che ti venga* NP has currency in speech-based (Culpeper and Kytö 2010: 17), non-fictional documents.

3 Quantitative analysis

3.1 Corpora and method

To investigate the conventionalisation of *che ti venga* NP, this study conducts a quantitative analysis using two historical corpora. First, the CORPUS Diacronico dell’Italiano (CODIT ‘Diachronic Corpus of Italian’, Micheli 2022), a balanced corpus of 33 million tokens from the thirteenth century until 1947. Secondly, a corpus of theatre plays by eighteenth-century playwright Carlo Goldoni: ItalianLISCortIta XVIII secolo – commedie Goldoni with 3,7 million tokens. Compiled by Chiara Ghezzi, it comprises 177 documents amounting to the entire dramatic oeuvre of Goldoni, in a (1955) edition by Giuseppe Ortolani for the collection *Classici Mondadori*.¹³

In both corpora I used the concordancer with initial search strings *ti venga* and *ti vengano* ‘come to you’, 3rd person singular and plural of the present subjunctive of *venire* ‘to come’ with the second personal pronoun singular *ti* ‘to you’ in the oblique case for the indirect object. These searches were complemented by a reverse search using the diseases found, revealing variation in verb forms. The infinitive *venire* or *venir* can be accompanied by *possa*, the 3rd person present subjunctive of the modal verb *potere* ‘may’, also as the inversion *venir possa*. Other personal pronouns were found, leading to further searches for new diseases. This resulted in 132 disease curses: 61 in CODIT and 71 in ItalianLISCortIta XVIII secolo – commedie Goldoni.

The quantitative analysis will comprise the following aspects:

- the connective *che* ‘that’;

¹² Literally, a bad Easter. The translations are loosely based on Tardivel (forthcoming).

¹³ Available from <https://liberliber.it/autori/autori-g/carlo-goldoni/> (last accessed 8 January 2025). Goldoni’s comedies are currently being reedited in critical editions for the *Edizione nazionale* ‘national edition’ by the publisher Marsilio, Venice.

- the augurative subjunctive *venga* ‘come’;
- the personal pronouns in a t/v system;
- intensifications;
- the disease noun;
- the maledictive context.¹⁴

3.2 *Che ti venga* NP in CODIT

The CODIT, *Corpus Diacronico dell’Italiano*, compiled by Maria Silvia Micheli, is hosted on the Czech National Corpus website and can be queried through the KonText interface.¹⁵ CODIT is a balanced diachronic corpus of written Italian, containing approximately 33 million tokens. It follows the structure and periodization of the MIDIA corpus,¹⁶ but includes entire texts rather than 8000-token samples. CODIT has five subcorpora: 1) thirteenth century – 1375; 2) 1376–1532; 3) 1533–1691; 4) 1692–1840; 5) 1841–1947. Each subcorpus includes six text genres: essays, literary prose, poetry, letters, scientific texts, and drama. The first subcorpus lacks scientific texts and has fewer tokens, while the others have approximately 6 million each (Micheli 2022). Due to its “speech-related” genres (Culpeper and Kytö 2010: 17), i.e. literary prose, letters and drama, CODIT is useful for analysing interpersonal interactions. However, its relatively small context panel makes context interpretation challenging, although not impossible.

3.2.1 Connective *che* ‘that’

The curses tend to be introduced with the connective *che* ‘that’. Salvi and Renzi (2010: 1211–1218) distinguish between optative and augurative clauses. Optative clauses express a general wish, using the imperfect subjunctive (*Avess’io mille marche d’oro* ‘if only I had a thousand gold marks’); augurative clauses specify the person for whom the wish is intended, whether for his/her benefit or harm, and employ the present subjunctive (*Che tu possa guarire presto* ‘may you recover quickly’). The augurative clause behaves like an elliptic subordinate, depending from a performative verb such as *augurare* ‘to wish’ (Dardano et al. 1992: 19; Salvi & Renzi 2010: 1215). Evans (2007) studies how clauses originally functioning as subordinates

¹⁴ The dataset containing all examples is available from LaRS - Language Repository of Switzerland (Paternoster 2025).

¹⁵ <https://www.korpus.cz/kontext/query?corpname=codit> (accessed 8 January 2025)

¹⁶ <https://www.corpusmidia.unito.it/> (accessed 8 January 2025)

developed into main clauses in their own right in a process called insubordination, whereby the clauses received specialised usages. Evans (2007: 387) lists politeness as an important function of insubordination since “insubordinating ellipsis has the effect of putting the face-threatening act ‘off the record’”. However, when impoliteness is the aim, the presence of the performative verb makes the utterance laborious and formal, undermining its pragmatic effectiveness (Dardano et al. 1992: 16). The insubordinate clause sounds more peremptory. Another difference between optative and augurative insubordinate clauses regards the use of the connective *che*, frequent with the augurative clause, but infrequent with the optative clause (Salvi and Renzi 2010: 2016).

Exceptionally in CODIT, with the disease appearing before the verb, *che* separates the disease (the subject) from the verb, as in (20):

- (20) *Il cancro che vi venga,*
 the canker COMP/REL 2PL.POL.DAT come.SBJV.3SG
 ‘May you rot,’ [lit. ‘The canker that to you may come,’]
messer lo compare di Puglia!
 mister the compeer.VOC from Puglia
 ‘mister compeer from Puglia!’ [lit. ‘mister the compeer from Puglia’]
 (3_PROSA_BANDELLO_NOVELLE)

Besides (20), CODIT only contains one other case of mid-sentence *che*. This mid-sentence *che* has a syncretic function, being both a relative pronoun and a conjunctive (indicated by the gloss COMP/REL; see Dardano et al. 1992: 32 specifically on mid-sentence *che* in disease curses in sixteenth-century comedy). Dardano et al. (1992: 17–20) discuss the hypothesis that this generic *che* is a hyperconnective, a pragmatic indicator of close textual cohesion, which has thematic and information-organizing purposes. It aims to achieve emphasis, particularly in less formal contexts.

The sentence-initial connective *che*, on the other hand, is present in 69% of the examples in CODIT: 42 out of 61. This is in line with Salvi and Renzi’s analysis (2010: 2016), as in (21):

- (21) *Che ti venga l'anticuore, vecchio cucco!*
 COMP 2SG.DAT come.SBJV.3SG the chest abscess old fool.VOC
 ‘A chest abscess take you, old fool!’ [lit. ‘That to you may come the chest abscess, old fool!’]
 (3_TEATRO_ANDREINI_LO SCHIAVETTO)

All use *che* except for one occurrence of *così* ‘so’, as in (22):

- (22) *Così vi venga il canchero e a voi*
 thus 2PL.POL.DAT come.SBJV.3SG the canker and to 2PL.POL
 ‘So may you rot, you’ [lit. ‘So to you may come the canker and to you’]
e a tutti quegli che non credono
 and to all those REL NEG believe.IND.PRS.3PL
 ‘and all those who believe’ [lit. ‘and to all those who do not believe’]
che io nolla finisca.
 COMP 1SG.NOM NEG_3SG.F.ACC finish.SBJV.1SG
 ‘I won’t finish it.’ [lit. ‘that I won’t finish it’]
 (3_PERS_CELLINI_VITA)

Finally, *che* is absent from 17 curses, as in (23):

- (23) *Oh, ti venga il cancro ne i denti.*
 oh 2SG.DAT come.SBJV.3SG the canker in the teeth
 ‘Oh, may you rot in your teeth’ [lit. ‘Oh, to you may come the canker in the teeth.’]
 (3_TEATRO_CROCE_LA FARINELLA)

Because *che* is uncommon in optative sentences (which are not directed at a specific individual), its presence, especially when delivered in a threatening tone, would alert the hearer to the possibility of a curse. Its pragmatic function is to command attention, and its modern equivalent would be “mark my words”. This usage aligns with the numerous forms of intensification commonly found in curses (see § 3.2.4).

3.2.2 The augurative subjunctive *venga* ‘come’

Section 2.2 highlights that historical curses rely on a higher power to enact the curse, thereby reducing the agency of the speaker. The semantics of *venire* ‘to come’ aligns perfectly with this notion.¹⁷ For *venire* ‘to come’ the *Grande Dizionario della Lingua Italiana* (s.v., henceforward GDLI)¹⁸ lists a meaning related to disease, mentioning curses: “Prodursi in un organismo (una malattia, una sensazione dolorosa, ecc.). – Anche in imprecazioni.” [To occur in an organism (a disease, a painful sensation, etc.). – Also in curses].

¹⁷ It is not a coincidence that Italian can use *venire* for the passive voice.

¹⁸ Consulted at <https://www.gdli.it/> (accessed 8 January 2025)

As seen above, the augurative sentence uses the present subjunctive. In CODIT 72% of occurrences – 44 out of 61 – use the present subjunctive *venga* or its plural *vengano*. There are 9 instances of *possa venir* ‘may come’, as in (24), which is the oldest curse:

- (24) *che venir possa fuoco da cielo*
 COMP come.INF may.SBJV.3SG fire from heaven
 ‘May a fire come from heaven’ [lit. ‘That nothing may come fire from heaven’]
che tutte v’ arda,
 REL all.F.PL 2PL.ACC burn.SBJV.3SG
 ‘to burn you all’ [lit. ‘that may burn you all,’]
generazion pessima che voi siete!
 generation.VOC worst REL 2PL.NOM be.IND.PRS.2PL
 ‘worst generation that you are!’
 (1_PROSA_BOCCACCIO_DECAMERON)

6 examples use the regional form *vegna*. For occurrences in periods 1 and 2, “before the normalization of literary Italian proposed by Pietro Bembo” (Micheli 2022), CODIT metadata mention the regional variety: all examples are from fourteenth-century Tuscan author Franco Sacchetti. The regional form *venghi* appears twice: once in Ludovico Ariosto’s *La Lena* (Tuscan) and once in seventeenth-century Giovanni Della Porta from Naples.

As regards the word order of the augurative sentence, the Subject can either precede the Verb, as in (19), or follow it (Salvi and Renzi 2010: 1216). In CODIT, the subject NP usually follows the main verb and this inversion is present in 50 out of 61 cases, as in (21), (22), (23), (24) and (25):

- (25) *Che vi venga il colèra!*
 COMP 2PL.DAT come.SBJV.3SG the cholera
 ‘Cholera take you!’ [lit. ‘That to you may come the cholera!’]
 (5_PROSA_VERGA_MALAVOGLIA)

Instances like (20) and (26), where the disease – the Subject – precedes the Verb are far less common, with 11 occurrences. These cases are marked and put into focus the disease (for intensification see Section 3.2.4):

- (26) *Che duol ti venga,*
 COMP suffering 2SG.DAT come.SBJV.3SG
 ‘Sorrow strike you’ [lit. ‘That suffering to you may come,’]

o canti tu o no!
 or sing.SBJV.2SG 2SG.NOM or NEG
 ‘whether you sing or not!’
 (1_POES_SACCHETTI_RIME)

While this marked word order could imply a stronger illocutionary force, it can also result from the constraints of meter and rhyme.¹⁹

So far, the most common form is *che venga* followed by a disease. What happens with the Indirect Object indicating the cursed?

3.2.3 The personal pronoun

Old Italian has a t/v system with *tu* for unmarked address and *voi* for deference towards higher rank. Pronoun use is linked to power and social rank: lower and middle classes use a reciprocal *tu* while aristocrats use *voi*, unless close. In the sixteenth century, the pronouns *Lei*, *Ella* start to express deference towards the high-power recipient, usually addressed with a title. *Voi* expresses respect, but it also continues to index social superiority. *Tu* is used amongst lower and middle classes and in top-down exchanges, e.g. between master and servant (Molinelli 2018; Ghezzi 2021). In 31 out of 61 cases, the cursed is addressed with *ti* (or *te* or *t*). *Ti* is used in (26) and in:

(27) *O che ti venga il gavàcciolo, pinchellone!*
 oh COMP 2SG.DAT come.SBJV.3SG the plague sore idiot.VOC
 ‘Oh a plague sore take you, you noddly!’ [lit. ‘Oh, that to you may come the plague sore, idiot!’]
 (3_TEATRO_ANDREINI_LO SCHIAVETTO)

Vi can be a plural *ti*. Plural *vi*, used in (24), (25) and (28), occurs in 8 cases:

(28) *Che venga fuoco del cielo*
 COMP come.SBJV.3SG fire of the heaven
 ‘May a fire come from heaven’ [lit. ‘That may come fire from the heaven’]
che tutte v’ arda!
 REL all.F.PL 2PL.ACC burn.SBJV.3SG
 ‘to burn you all!’ [lit. ‘that may burn you all!’]
 (3_PROSA_BANDELLO_NOVELLE)

19 Of the 11 cases of marked word order – Subject before Verb – 7 appear in verse texts.

In 5 cases *vi* concerns a single recipient. Example (29) occurs in Old Italian when *voi* expresses deference:

- (29) *venir vi possa fame grande e sete,*
 come.INF 2PL.POL.DAT may.SBJV.3SG hunger great and thirst
 ‘May great hunger and thirst come to you’
che d’ogni vostro danno lieto so’.
 since of every 2PL.POL.POSS harm glad be.IND.PRS.1SG
 ‘for of all your harm I am glad.’
 (1_POES_SACCHETTI_RIME)

It follows (26) which is the top-down pronoun use between the same characters in a poem by Franco Sacchetti. The 4 remaining cases of *vi* – as in (20) – all regard the sixteenth century, where *vi* expresses respect. It is, hence, possible to curse someone who is addressed with deference or respect:

- (30) *Che vi vengano tremila cacasangui!*²⁰
 COMP 2PL.POL.DAT come.SBJV.3PL three-thousand bloody fluxes
 ‘The bloody flux take you three thousand times!’ [lit. ‘That to you
 may come three thousand bloody fluxes!’]
 (3_PROSA_BANDELLO_NOVELLE)

Example (30) is a bottom-up curse against a priest, accompanied by physical violence. No curses use deferential *Le*.

The use of second-person pronouns indicates that the curser directly addresses the cursed, who is within earshot and aware of the curse. As far as the context panel allows us to see, one curse using the second person pronoun occurs in a monologue, and another, (31), represents the character’s thoughts:

- (31) *Che ti possa venir la rognà.*²¹
 COMP 2SG.DAT may.SBJV.3SG come.INF the scabies
 ‘The itch get you’ [lit. ‘That to you may come the itch:’]
(dicevo io tra me e me)
 say.IND.IPFV.1SG 1SG.NOM between me and me
 ‘(I said to myself)’

²⁰ *Cacasangue* is a historical term for dysentery.

²¹ *Rogna* is a historical term for scabies.

pezzo d'asino imbastardito tra le sacrestie!
 piece of donkey.voc bastardised among the sacristies
 'piece of bastardised jackass amongst the sacristies!'
 (5_PERS_TANINI_LA VITA DI GIULIO PANE)

Example (31) dates from 1922; it is the most recent within CODIT.

The pronoun for the cursed usually appears as an indirect object of *venga*, but in (28) it appears in a relative clause. In the remaining 17 cases, the cursed is referred to by a proper (32) or a common noun (33), or by a third-person pronoun (34):

(32) *O venga il canchero a Cain!*²²
 oh come.SBJV.3SG the canker to Cain
 'Oh, may Cain rot!' [lit. 'Oh, may the canker come to Cain!']
 (3_TEATRO_ANDREINI_LO SCHIAVETTO)

(33) *Che venga il canchero ai peccati!*²³
 COMP come.SBJV.3SG the canker to the sins
 'May the sins rot!' [lit. 'That the canker may come to the sins!']
 (3_PERS_CELLINI_VITA)

(34) *Che li venga il gavocciolo,*
 COMP 3SG.DAT.M come.SBJV.3SG the plague sore
 'A plague sore take him,' [lit. 'That to him may come the plague sore,']
ignorante che egli è!
 ignorant man REL 3SG.NOM.M be.IND.PRS.3SG
 'the fool that he is!' [lit. 'ignorant man that he is!']
 (3_PROSA_BANDELLO_NOVELLE)

In sum, over two thirds of curses directly address a character in the fictional dialogue. Unmarked *ti* and its plural *vi* (39 cases) are far more common than respectful and deferential pronouns.²⁴ Some curses refer to third parties.

²² Cain is not the biblical figure, but a Jewish character in the play.

²³ In (33) exceptionally the target is not a human being, but a personified entity. The speaker is angry: he sees his current misfortune – he has been wounded in a fight – as divine punishment for sins he committed earlier. The sins themselves are held responsible and cursed as though they were living beings.

²⁴ For Goldoni – see § 3.3.3 – I studied eighteenth-century cases of pronoun switches, where the curse switches to a non-deferential form. This strategy is very common in earlier Italian, however, the limited size of the context panel did not allow to detect cases of pronoun switches in CODIT.

3.2.4 Intensification

Section 3.2.2 discussed a first type of intensification, the marked word order where the disease precedes the verb. There are many other forms of intensification. Culpeper notes that intensification exacerbates offensiveness, making the impolite intent more obvious (2011: 139).

Hyperbole

The hyperbole exaggerates numbers, wishing for multiple episodes of the disease, even three thousand, as in (30). 5 curses include hyperbole:

- (35) *Ribaldo! Che ti vengano Cento cancri!*
 scoundrel.VOC COMP 2SG.DAT come.SBJV.3PL a hundred cankers
 ‘Scoundrel! May you rot a hundred times over!’ [lit. ‘Scoundrel! That to you may come a hundred cankers!’]
 (2_TEATRO_GABRIELE ARIOSTO_STUDENTI)

Body part

Another intensifier is the addition of a body part. CODIT contains three cases:

- (36) *Empio signor, che della robba altrui*
 ‘Impious lord, who with other people’s possessions’
lieto ti vai godendo e del sudore,
 ‘and sweat happily goes enjoying yourself,’
venir ti possa un cancaro nel cuore,
 come.INF 2SG.DAT may.SBJV.3SG a canker into the heart
 ‘may a canker come to your heart’ [lit. ‘may to you come a canker into the heart,’]
che ti porti di peso ai regni bui.
 ‘that carries you at once to the dark realms.’ [lit. ‘which may carry you entirely to the dark realms’]
 (2_POES_BERNI_RIME)

In (36) the reference to the heart makes the curse especially menacing. Incidentally, this example is a striking case of the grammar of impoliteness overriding the deference system (see Section 3.2.3): the second person singular is used to curse (and insult with *empio signor*) a high-power aristocrat, the lord of Rimini Sigismondo Malatesta. Curses (23) and (37) target the mouth:

- (37) *Che li venghi la peste alla lingua!*
 COMP 3SG.DAT.M come.SBJV.3SG the plague to the tongue
 ‘May he get the plague in his tongue!’ [lit. ‘That to him may come the plague in the tongue!’]
 (3_TEATRO_DELLA PORTA_LA SORELLA)

Body orifices make the disease particularly debilitating and humiliating, intensifying the level of taboo.

Multiple diseases

Curse (38) uses two diseases:

- (38) *quello sguaiato tristo facimale, quel disgraziato, quel sciaguratello,*
 ‘that rude, sad rascal, that wretch, that little scoundrel,
che gli venga un gavocciolo, un cassale,
 COMP 3SG.DAT.M come.SBJV.3SG a plague sore a (mortal) fever
 ‘a plague sore take him, a mortal fever,’ [lit. ‘that to him may come a plague sore, a mortal fever;’]
s’è tolto quel pensiero del cervello?
 ‘has he taken that thought out of his mind?’ [lit. ‘has he taken that thought from the brain?’]
 (4_POES_PARINI_ALCUNE POESIE)

Another curse counts three diseases. However, instead of a usage, it is a “mention” (Jucker 2020: 19), where each disease is an example of a different curse. Sixteenth-century short-story writer Matteo Bandello, who is no stranger to disease curses – CODIT contains no fewer than 17 curses from his *Novelle* – writes this metapragmatic comment in (39):

- (39) *I nostri vicini bergamaschi quando sentono alcuno che maledicendo il compagno gli dice:*
 ‘Our neighbours in Bergamo, when they hear someone cursing his companion by saying:’
“Ti venga il cacasangue, la febre, il cancaro”
 2SG.DAT come.SBJV.3SG the bloody flux the fever the canker
 “‘The bloody flux, the fever, the canker take you’” [lit. “To you may come the bloody flux, the fever, the canker”]

e simili imprecazioni, sogliono dire: “Io non so dir tante cose, ma io vorrei che tu fussi morto”.

‘and similar imprecations, they tend to say: “I don’t know how to say so many things, but I wish you were dead”.’

(3_PROSA_BANDELLO_NOVELLE)

For Bandello (or his narrator) the Bergamasque, instead of using a formulaic disease curse, simply wish someone dead.

Exclamation marks

The exclamation mark suggesting an intensified prosody is present from the fourteenth century, in (24) and (26), to the most recent example from 1922, in (31). In all, 38 curses out of 61 have an exclamation mark, over half.²⁵

Emotive interjections

The emotive interjections *o*, *oh*, *deh*, *doh*, *olà* precede the curse. In the GDLI (s.v.) *O* and *Oh*, treated as one lemma, can, besides positive emotions, also express “sdegno, indignazione, rimprovero, ammonizione o, anche, ironia e sarcasmo” [outrage, indignation, reproach, admonition or, also, irony and sarcasm]. *Oh* and *O* introduce 8 curses, as seen, resp. in examples (2) and (27). However, in (40) *oh* expresses surprise:

- (40) *Oh! Che ti venga il bene.*
 oh COMP 2SG.DAT come.SBJV.3SG the good
 ‘Oh! Bless you.’ [lit. ‘Oh! That to you may come the good.’]
 (4_TEATRO_GOLDONI_LE SMANIE PER LA VILLEGGIATURA)

Deh or *Doh* can express “rimprovero, disapprovazione” [reproach, disagreement] according to the GDLI (s.v.). The three cases all come from Ariosto’s *La Lena* as seen in examples (1), (6), and (7). *Olà* is an attention getter, but the GDLI (s.v.) notes its use to emphasise threats and warnings, as in (41).

²⁵ The punctuation raises the issue of the delimitation of the curse. I aimed to retain full sentences, which start with a capital letter and end with a full stop, a question mark, an exclamation mark, making an exception for the semi-colon.

- (41) *Olà, che venga fuoco dal cielo che t' arda.*
 hey COMP come.SBJV.3SG fire from the sky REL 2SG.ACC burn.SBJV.3SG
 'Hey, may a fire come from heaven and burn you!' [lit. 'Hey, may fire come from the sky that burns you.']
 (3_PROSA_BANDELLO_NOVELLE)

Emotive interjections occur in 11 curses.

Insults

Vocative insults appear in (4), (6), (20),²⁶ (24), (27), (31), (35), (36). In (31) *pezzo di* 'piece of' exemplifies impolite categorisation (Alfonzetti and Spampinato Beretta 2012: 3), but it also contains the blasphemy *tra le sacristie* [amongst the sacristies]. In (42) the insult itself is intensified by a crescendo of derogatory adjectives:

- (42) *Tu farai il gavoccio che ti venga,*
 2SG.NOM do.IND.FUT.2SG the plague sore COMP/REL 2SG.DAT come.SBJV.3SG
 'A plague sore take you,' [lit. 'You'll get the plague sore, that may come to you,']
sozzo cane, unto, bisunto!
 dirty dog.VOC greasy double greasy
 'dirty, disgusting, right filthy dog!'
 (3_PROSA_BANDELLO_NOVELLE)

15 curses in CODIT contain insults. Sometimes, the derogatory comment is not a vocative. In (34) and (38) it refers to a third party and is a criticism (*reimproperatio*) rather than an insult.

Swearing

Swearing is blasphemous (see Section 2.2) and can be considered impolite because it breaches moral and religious expectations. Culpeper (2018) notes how taboo words are largely used to intensify other forms of impoliteness, as seen in (31) which is blasphemous. (43) contains swearing:

- (43) *Al corpo di Cristo, io ci vo' entrar dentro e far questione con questo parmegiano tirasassi,*
 'By the body of Christ, I want to enter inside and deal with this stone-throwing Parmesan,'

²⁶ In (20) the honorific title is mock-polite.

che gli vengano mille cacasangui!
 COMP 3SG.DAT.M come.SBJV.3PL A thousand bloody fluxes
 ‘the bloody flux take him a thousand times over!’ [lit. ‘that to him may come a
 thousand bloody fluxes!’]
 (3_PROSA_BANDELLO_NOVELLE)

Overall, only 16 examples out of 61 are not intensified in one form or another:

- (44) *Papagallo volesti dire,*
 ‘Parrot you meant,’
che ti venga il grosso.
 COMP 2SG.DAT come.SBJV.3SG the big one
 ‘a plague sore take you.’ [lit. ‘that to you may come the plague sore.’]
 (3_TEATRO_ARETINO_CORTIGIANA)

As an intermediate conclusion, the most frequent pattern in CODIT is *che ti venga* NP with some form of intensification.

3.2.5 The Noun Phrase

The diseases used in curses strongly relate to the taboo sphere. A common curse invokes the plague, which was extremely dangerous. Curses use infectious diseases and parasitic infestations like dysentery, cholera, malaria, scabies. Symptoms such as fever, rashes, bloody diarrhoea, vomiting, purulent, festering sores, severe itchiness are particularly unpleasant and stigmatising. Some curses invoke veterinary diseases – *l'anticuore*, *la pipetola*²⁷ –, reducing the cursed to the state of an animal. Others wish for sustained misfortune: *malanno*, *ritenso*, *rovello* mean, respectively, a year of bad luck, an accident, torment. Others bring on inevitable death, by *fuoco*, *tuono*, *saetta da cielo*, resp. ‘heavenly fire’, ‘thunder’, ‘lightning’. Over time, the link to the actual disease may have weakened: *canchero* and *rogna*, the GDLI (s.v.) observes, were also used as metaphors for misfortune. Figure 1 shows the raw figures for the distribution of diseases and (mis)fortune²⁸ over the five historical periods in CODIT:

²⁷ *L'anticuore* or ‘anticor’ is a tumour in proximity of the heart in four-legged domestic animals, e.g. horses. *La pipetola* is Neapolitan dialect for *pipita* (see <http://www.vesuvioweb/>, accessed 8 January 2025), a disease of the tongue in birds, especially in chickens.

²⁸ There is, indeed, one benedictive use.

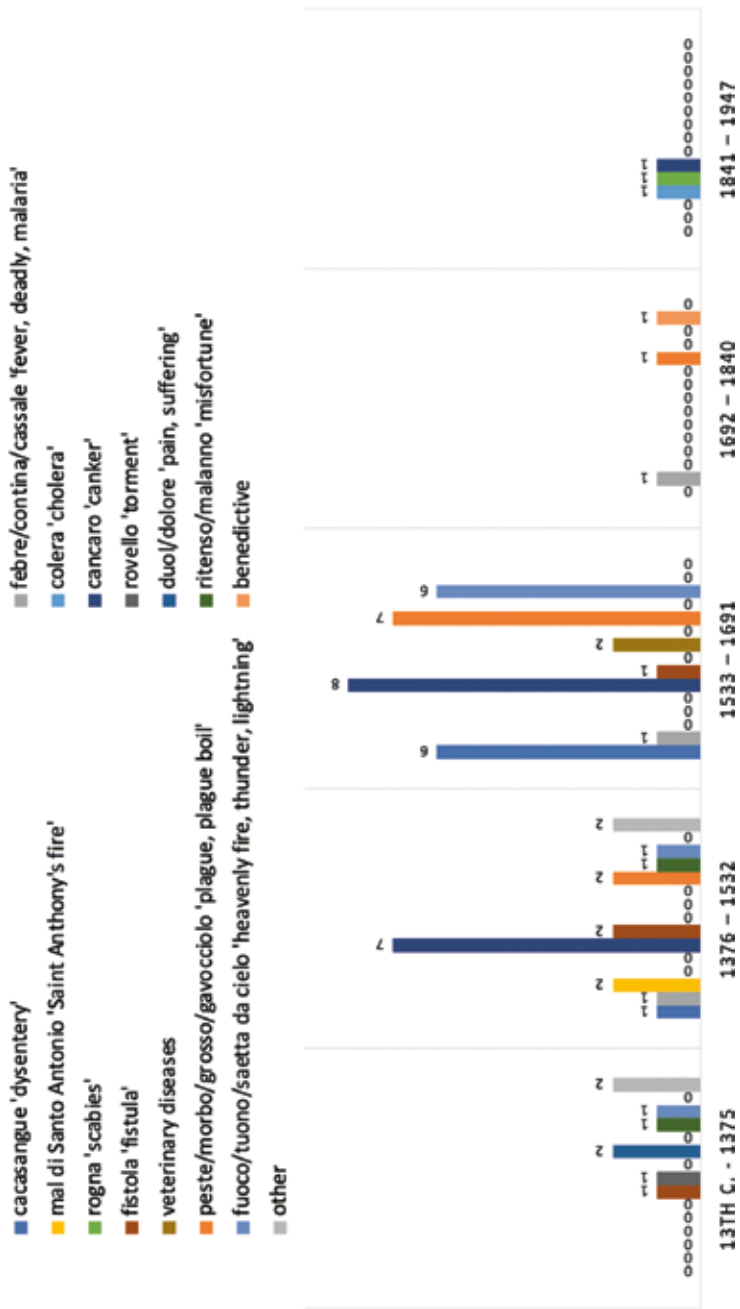


Figure 1: Distribution of different diseases and (mis)fortune in the curse *che ti venga* NP over the five historical periods in CODIT²⁹.

²⁹ Note that there are slightly more diseases than curses: example (38) includes two diseases and (39) three.

In Figure 1, the most frequent condition is a festering canker, with 7 occurrences in period 2 and 8 in period 3, followed by 7 cases of the plague in period 3. Dysentery has 6 occurrences in period 3, as do maledictions wishing for a heavenly fire, thunder or lightning. The disease choice roughly corresponds to epidemics: the plague causes two epidemics in Italy, in the fourteenth and the seventeenth centuries, and the first cholera epidemic occurred in the 1830s, with an 1881 cholera curse in CODIT. Generally, the impoliteness strategy is to curse someone with the most dangerous, debilitating and humiliating disease of the time.

Regardless of the type of NP used, Figure 1 shows an increase in disease curses until the seventeenth century followed by a drop from 1692 onwards. This trend in the occurrences of curses across CODIT is visualised in Figure 2 with normalised figures for the total amount of curses per historical period:³⁰

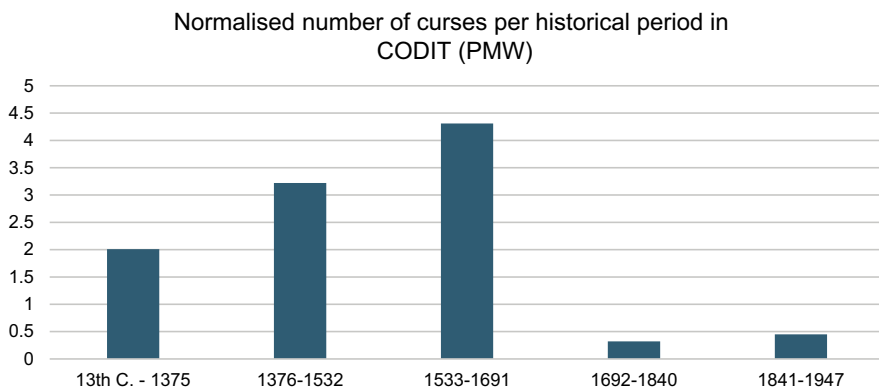


Figure 2: Distribution of number of curses using '*che ti venga* NP' per historical period in CODIT (normalised frequencies per million words).

From the thirteenth century to 1691, curses steadily increase, peaking in Period 3. There is some risk of bias here: Trifone (2000: 135, 2019) and D'Onghia (2011), leading scholars of Italian drama, point out that sixteenth- and seventeenth-century drama made ample use of impoliteness to entertain audiences: consequently, the numbers in periods 2 and 3 may be overrepresented, at least in part. Note that over the five periods a third of examples occur in theatre texts, 23 out of 61. From 1692 curses sharply decline. Within period 3 (see raw figures in Figure 1), most cases, 24, take place in the sixteenth century, and only 5 in the seventeenth century, all before 1615. After 1615, curses virtually disappear: there are no cases between 1615 and

³⁰ I provide normalised frequencies given that the first period in CODIT contains a much lower number of tokens compared to the later periods, ca. 4 million instead of 6 million.

1691, the entire second half of period 3, with only 3 in both periods 4 and 5. The rigorous prosecution of blasphemy by the Inquisition in the sixteenth and seventeenth century (Nash 2007) clearly played a key role in reducing the prevalence of curses in written texts intended for publication, but it is equally true that high-brow theatre evolved towards tragedy, with high-ranking protagonists using less scurrilous language. The numerous curses in Goldoni's plays (see Section 3.3), despite being subject to censorship by the *Esecutori contro la bestemmia* [Executors Against Blasphemy], can be attributed to the diminishing persecution of blasphemy in the second half of the eighteenth century, influenced by the growing impact of Enlightenment ideals, which allowed such language to be somewhat tolerated in theatrical contexts.

3.2.6 Maledictive context

The expectation is for maledictions to appear in contexts where characters are arguing. Almost all appear in fictional dialogues. The contexts range from angry rants to heated arguments and isolated outbursts. Example (30) is uttered in an angry rant, (45), by a cleric angry at his superior, a priest who has lost a valuable coin:

- (45) *E narrò al chierico la perdita del doppio ducato. Come il giovine sentì la pazzia del messere, se gli rivolse con il più brutto viso che poté e disse: – Oimè, che sento! che vi vengano tremila cacasanguì! E ch' avete voi voluto fare, uomo da poco e da meno assai ch'io non dico? Voi adunque avete restituito un doppione perchè non era così di peso come la vostra avara ingordigia avrebbe voluto, [. . .].* 3_PROSA_BANDELLO_NOVELLE [And he told the cleric of the loss of the double ducat. As soon as the young man heard the man's madness, he turned to him with the ugliest face he could muster and said: – O my goodness, what do I hear! May the bloody flux get you three thousand times! And what have you wanted to do, man of little worth and far less than I say? You have returned a double ducat because it was not as heavy as your avaricious greed would have wished]

The priest remains silent. Other curses form part of heated arguments, here in (46) where a husband scolds his wife:

- (46) – *che questa traditora pisana sarà venuta a Lucca per volermi governare. Che fussi io stato in letto con la quartana quel dì che mi venne voglia di prender moglie pisana, chè tutti tutti, uomini e donne, sète traditori! Che venga il fuoco dal cielo che t'arda, rea femina che tu sei! – A questo, Beatrice che del marito teneva poco conto, per più farlo adirare gli rispose: – A la croce di Dio che*

avete una gran ragione a dir questo [. . .]. 3_PROSA_BANDELLO_NOVELLE [– That this traitorous Pisan woman will have come to Lucca to govern me. If only I had stayed in bed with a bad fever the day I felt like taking a Pisan wife, because all of you, all men and women, are traitors! May the heavenly fire come to burn you, wicked woman that you are! – At this, Beatrice, who cared little for her husband, to make him angrier, answered him: – By the cross of God, you have a good reason for saying this]

Beatrice swears and replies with sarcasm. At times, the curse is a mere angry outburst, in an otherwise non-conflictual exchange, as in (47):

- (47) [. . .] *che vogliono sapere?*” domandò comare Grazia. “*Vogliono sapere se è vero che la Lia se la intendeva con don Michele, e che suo fratello ‘Ntoni abbia voluto ammazzarlo per tagliarsi le corna; me l’ha detto l’avvocato.*” “*Che vi venga il colera! – soffiò loro lo speziale facendo gli occhiacci. – Volete che andiamo tutti in galera? Sappiate che colla giustizia bisogna dir sempre di no, e che noi non sappiamo niente.*” Comare Venera si rincantucciò nella mantellina, ma seguì a borbottare.” 5_PROSA_VERGA_MALAVOGLIA [what do they want to know?” cousin Grazia asked. “They want to know if it’s true that our Lia was carrying on with Don Michele, and that her brother ‘Ntoni wanted to kill him to revenge himself; the lawyer told me so.” “Cholera take you both! – the chemist hissed with an angry glare. – “Do you want us all to go to prison? Know that with the men of law you always have to say no, and that we don’t know anything.” Cousin Venera curled up in her cape, but continued to mutter.]

The curse aims at boosting the warning not to reveal anything.

Two usages are mock-impolite. The antiphrastic use depends on the hearer (or reader) recognising the routine impoliteness, thereby reinforcing the disease curse’s status as a conventional expression:

- (48) [. . .] *raccontar le genealogie veneziane esser un Tullio, ma nel resto dimostrarsi il maggior sciocco del mondo. A la fine il vecchio mutata la voce e il modo di parlare, ridendo disse: – Io so che sète galanti uomini a non riconoscer il vostro Girolamo pittore. Che vi venga il gavocciolo, “poëtis quae pars est”.³¹ – Fu subito riconosciuto, e risolvendosi il tutto in riso, egli se n’andò in una camera, e spogliatosi l’abito da comedia si rivestì i suoi panni e ritornò in sala*

³¹ *Poëtis* corrects *poftis*: the original in CODIT appears to misrender a letter with a trema, likely due to scanning and OCR issues.

[. . .]. 3_PROSA_BANDELLO_NOVELLE [telling Venetian genealogies like a real Cicero, but otherwise proving himself the world's greatest fool. At the end the old man changed his voice and manner of speaking and, laughing, he said: – I know that you gentlemen do not recognise your painter Girolamo. A plague sore take you, some “poets” you are. – He was immediately recognised, and the whole thing ending in laughter, he went into a room, and removing his comic garb, he put on his own clothes and returned to the hall]

The painter in (48) is familiar to his audience. He laughs before cursing them, to signal the joking intent, adding a mock-impolite insult. The ensuing laughter confirms the uptake of banter.

The second example, (49), appears in a poem with the title *A scusa d'un francesismo scappato nel precedente sonetto* [Apology for a Frenchism That Escaped me in the Previous Sonnet] in which the young Carducci jokingly addresses the purist linguists, defenders of Italian:

- (49) [. . .] *balii de la lingua, affeddiddio Che questo a punto a punto è il vostro caso, E voi potete pur darmi di naso Menando gran rumor del fatto mio. Guardivi sant'Anton come rimaso D'un franciosismo al laccio or son anch'io; E cancher venga al nemico di Dio Che pria la rima n'arrecò in Parnaso. Ch'io veggio correr fuori a gran baldanza, Pur me ammiccando con un risolino, Molti linguisti di molta importanza. E' vanno per consigli a l'Ugolino.* [. . .] 5_POES_CARDUCCI_JUVENILIA [minders of our language, for God's sake, This precisely is your case, And you may rightly mock me, Making a great noise about my mistake. May Saint Anthony protect you, For now I too have fallen into the trap of a Frenchism; And may the enemy of God rot Who first brought rhyme to Parnassus. For I see many linguists of great importance, Running out boldly, though they Wink at me with a little smile, And going for advice to Ugolino.]

Being a poet himself, Carducci's curse against whoever invented rhymed poetry cannot be taken too seriously.

Out of the 61 curses found, one is a mention rather than a usage, one is an example in a grammar book,³² one is part of a song,³³ two are mock-impolite and one is a polite, benedictive usage, which will be discussed in Section 4. The routinised context is one of impoliteness.

³² Example (24) also figures as a grammar example regarding preposition use in 2_PROSA_BEMBO_PROSE DELLA VOLGAR LINGUA, from 1525.

³³ [cantando] *Venir vi possa el diavolo allo letto, Dapoi ch'io non vi posso venir io!* 2_TEATRO_MACHIAVELLI_MANDRAGOLA [(singing) May the devil visit you in bed, as I myself am not allowed there].

3.2.7 Conclusion

In CODIT, the phrase *che ti venga* NP, followed by an exclamation mark, is a recurrent, stable form, hence a formula. In contrast, the formula occurs with additional, optional slots, which allow for ample creativity as regards intensification. Curses target equals and social inferiors using *ti*, but respectful or deferential pronouns are also possible. The frequency of the disease curses peaks from the fifteenth to the seventeenth century, favouring lethal and debilitating diseases or severe ill-fortune. Disease nouns and intensification with insults, body orifices and swearwords place the curse into the taboo sphere. The context is largely maledictive and impolite, hence *che ti venga* NP it is a conventionalised impoliteness expression.

3.3 *Che ti venga* NP in ItalianLISCortIta XVIII secolo – commedie Goldoni

With 23 CODIT curses in drama texts, exploring an all-theatre corpus is logical. Carlo Goldoni's (1707–1793) extensive oeuvre is ideal for pragmatic investigation due to its interactive nature and representation of all social classes. Goldoni transformed the commedia dell'arte into more realistic plays with rounded characters and complete scripts, using more natural language (Matarrese 2010).

ItalianLISCortIta XVIII secolo – commedie Goldoni³⁴ comprises 71 occurrences of *che ti venga* NP.

3.3.1 Connective *che* 'that'

In CODIT, sentence-initial *che* appeared in about two thirds of the curses. In Goldoni, this percentage is 80%:

- (50) *Che ti venga la rabbia!*
 COMP 2SG.DAT come.SBJV.3SG the rabies
 'Rabies get you!' [lit. 'That to you may come the rabies!']
 (*Pamela nubile*, 1750)

Only 13 out of 71 curses have no *che*:

34 Hosted on <https://www.sketchengine.eu/> (accessed 8 January 2025).

- (51) *Venga la stizza, venga la rabbia*
 come.SBJV.3SG the fury come.SBJV.3SG the rabies
 ‘May the fury, may rabies get’ [lit. ‘May come the fury, may come the rabies’]
a chi m’ ha fatto metter in gabbia
 to REL 1SG.ACC have.IND.PRS.3SG do.PST.PTCP put.INF in cage
 ‘whoever has had me put in a cage’ [lit. ‘to who has had me put in a cage.’]
 (*Arcifanfano re dei matti*, 1750)

Example (51) is part of a libretto and matches the verse and rhyme scheme. Mid-sentence *che* is not found.

3.3.2 The augurative subjunctive *venga* ‘come’

Goldoni uses the present subjunctive *venga* for augurative clauses. Only example (52) has *vegna*, from the Bergamasque dialect spoken by servant Truffaldino:

- (52) *Mo se l’ è qua, in casa, in sala,*
 ‘But if he is here, in the house, in the drawing room,’
che ve venga el malanno
 COMP 2PL.POL.DAT come.SBJV.3SG the misfortune
 ‘bad luck to you!’ [lit. ‘that to you may come the misfortune.’]
 (*Il servitore di due padroni*, 1745)

One curse, (53), uses modal *possa*:

- (53) *Sposa la pastorella, vaga, gentile e bella,*
 marry.IMP.2SG the shepherdess fair gentle and pretty
 ‘Marry the shepherdess, Faire, gentle and pretty,’
Che ti possa venir la caccarella.
 COMP 2SG.DAT may.SBJV.3SG come.INF the diarrhoea
 ‘May you get the shitty.’ [lit. ‘that to you may come the squirts.’]
 (*Le virtuose ridicole*, 1752)

In each and every case, the disease (Subject) follows the Verb *venga*; it never precedes.

Based on the presence of *che*, the Verb form *venga* and the consistent word order, disease curses in Goldoni appear even more formulaic compared to those in CODIT. This uniformity in Goldoni can be attributed to the distinct composition of the two corpora. While CODIT is a compilation of sources of diverse origins and his-

torical periods, the Goldoni corpus represents the language use of a single author within a specific text type and period.

3.3.3 The personal pronoun

In the eighteenth century, *voi* is the unmarked pronoun: it expresses respect, even in top-down relationships. *Ella* or *Lei* express formality and deference, whereas *tu* is particularly marked to express reduced distance (Molinelli 2018). 50 curses use this marked *ti* form. Only 11 curses have *vi*: 8 express respect, and 3 refer to a plural recipient. In (52) Truffaldino addresses a social superior with *vi*; in (54) *vi* is used for an inferior, Brighella:

- (54) *Non mi parlate di perdere,*
 NEG 1SG.DAT speak.IMP.2PL of lose.INF
 ‘Do not speak to me of losing,
che vi venga il malanno.
 COMP 2PL.POL.DAT come.SBJV.3SG the misfortune
 ‘bad luck to you!’ [lit. ‘that to you may come bad luck.’]
(Il giocatore, 1750)

Brighella, who runs a casino, addresses the gambler Florindo with *Lei*. Florindo uses *voi* for Brighella, maintaining his respectful address in curse (54). In (55), *vi* is a plural and refers to multiple women:

- (55) *Vi venga la saetta, che siate maledette.*
 2PL.DAT come.SBJV.3SG the lightning COMP be.SBJV.2PL damned
 ‘May you all be struck by lightning, may you be damned.’ [lit. ‘To you may come the lightning, may you be damned.’]
(Il filosofo inglese, 1753)

In 9 examples a third party is cursed, as in (56):

- (56) *Venga il canchero all’avaraccio.*
 come.SBJV.3SG the canker to the miser
 ‘May the miser rot.’ [lit. ‘May come the canker to the miser.’]
(Il ritorno dalla villeggiatura, 1761)

One curse has no indirect object:

- (57) *Venga* *lo cancro*.
 come.SBJV.3SG the canker
 ‘Go rot.’ [lit. ‘May come the canker.’]
 (*La cameriera brillante*, 1753)

Appearing within a substantial quote, curse (57) does not interfere with the interpersonal dynamic of the characters. No curse uses deferential *Le*.

Whereas CODIT has 31 cases with *ti* out of 61, Goldoni uses *ti* in 50 cases out of 71. Both Molinelli (2018) and Ghezzi (2023) analyse eighteenth-century personal pronouns based on Goldoni, noting that *tu* is used in asymmetrical exchanges, with a higher-class member addressing someone of lower class, say master to servant, and among lower classes. Characters can temporarily switch from *voi* to *tu* in close interactions, when distance is reduced for a positive or a negative reason (Molinelli 2018; Ghezzi 2023). This aligns with the use of *ti* in curses, however, its use is influenced by a dramatic device that was hardly present in the CODIT curses: the aside. Most uses of *ti* occur in an aside, addressing the audience, while the characters on the stage cannot hear. Therefore, it is useful to compare the pronoun use in the asides with on-record uses between the same characters: the aside may serve as a safe space because an on-record curse would be inappropriate given its impoliteness. For this comparison, I slightly adapt my method and make use of the entire play.³⁵

Out of 50 cases of *ti*, in 6 a master addresses a servant, as in (58) and (67), where *ti* is used out- and inside curses, which are all on record:

- (58) *Va* *ora, che ti venga* *il malanno*
 go.IMP.2SG now COMP 2SG.DAT come.SBJV.3SG the misfortune
 ‘Go now, bad luck to you’ [lit. ‘Go now, that to you may come the misfortune.’]
 (*gli dà una spinta, lo fa muovere, e vede in terra lì due zecchini*).
 ‘(he gives him a push, makes him move, and sees the two coins lying on the ground).’
 (*Il giocatore*, 1764)

The stage direction shows Florindo pushing his servant to the side.

³⁵ The context panel in the Sketch Engine concordancer provides access to extensive extracts of the script, but the formatting makes navigating turn-taking arduous. Metadata provide a link to a digital version of the play in the Ortolani edition on <https://liberliber.it/autori/autori-g/carlo-goldoni/> (accessed 8 January 2025)

Three other cases of on-record *ti* regard symmetrical exchanges between servants (59) or friends:

- (59) *Io? che ti venga il fistolo!*
 1SG.NOM COMP 2SG.DAT come.SBJV.3SG the fistula
 ‘Me? A running fistula take you!’ [lit. ‘Me? That to you may come the fistula!’]
 (*La vedova spiritosa*, 1757)

In (60) the friends are aristocrats:

- (60) *Che ti venga la rabbia nel dorso;*
 COMP 2SG.DAT come.SBJV.3SG the rabies in the back
 ‘Rabies get you in your back;’ [lit. ‘That to you may come the rabies in the back;’]
 (*La favola de’ tre gobbi*, 1749)

A count and a marquess dispute the love of a gentlewoman. Outside arguments, they use *voi*; during arguments, they switch to *tu*. 6 on-record curses, indeed, ‘downgrade’ *vi* to *ti* (as in (73), to express increased closeness). Middle-class Lelio addresses gondolier Tita with respectful *voi*, switching to *ti* in curse (61):

- (61) *Che ti venga la rabbia, barcaiuolo del diavolo.*
 COMP 2SG.DAT come.SBJV.3SG the rabies boatman.VOC of the devil
 ‘Rabies get you, boatman of the devil.’ [lit. ‘That to you may come the rabies, boatman of the devil.’]
 (*La putta onorata*, 1748)

On-record *ti* (15 cases out of 50) is used either for a master cursing a servant or among equals. The pronoun may be ‘downgraded’ from *vi* to *ti*. Put differently, these are the contexts allowing for an on-record curse with *ti* and there are no cases where an inferior curses a superior with *ti*.

The most frequent use of *ti* (an intriguing 35 cases out of 50) is off record: in asides, monologues, or when the cursed is hard of hearing or off-stage. Except for three cases (cursing two servants and a cat), the 32 remaining off-record curses target someone usually addressed with respectful *voi* or deferential *Ella/Lei*:

- (62) *(Che ti venga la rabbia) (da sé)*
 COMP 2SG.DAT come.SBJV.3SG the rabies (to himself)
 ‘(Rabies get you!) (to himself)’ [lit. ‘(That to you may come the rabies) (to himself)’]
 (*Il padre di famiglia*, 1750)

Earlier, Ottavio addressed the cursed, Florindo, with *voi*.

- (63) *(Oh, che ti venga la rabbia!)*
 oh COMP 2SG.DAT come.SBJV.3SG the rabies
 ‘(Oh, rabies get you!)’ [lit. ‘(Oh, that to you may come the rabies!)’]
 (*Le avventure della villeggiatura*, 1761)

Costanza addresses Vittoria with *Lei*, but in the aside she curses her with *ti*.

Nearly half of Goldoni’s curses occur off-record, involving a downgrading of the address pronoun. While these asides signal the curser’s anger to the public, the recipient remains unaware, and there is no impoliteness uptake. That the usual pronoun is *voi* or *Ella/Lei* indicates that the respect or the deference for the target likely prevents an on-record curse. The presence of curses in asides confirms their conventionalised use for impoliteness: Goldoni relies on the audience recognising the curse as too impolite when targeting a character usually addressed with a respectful or deferential pronoun. Because the curse is conventionalised for impoliteness, it can only go on-record under certain conditions: with inferiors or amongst equals (whether servants or aristocrats). Nevertheless, some on-record curses – examples (52) and (70) – use *vi* from a servant to a middle-class person. No curse uses deferential *Le*, a finding identical for CODIT. Note that an offended aristocrat may challenge the speaker to a duel.

On or off record, the dominating pronoun is *ti* and *che ti venga* is the most frequent formula.

3.3.4 Intensification

Intensification by having the disease in front of the Verb is not found in Goldoni, neither is hyperbole.

Body part

Body parts only appear in two cases. Interestingly, one is part of a play set in the sixteenth century:

- (64) *Che te venga lo canchero in mezzo dello core.*
 COMP 2SG.DAT come.SBJV.3SG the canker in middle of the heart
 ‘The canker strike you in the middle of your heart’ [lit. ‘That to you may come the canker in the middle of the heart.’]
 (*Torquato Tasso*, 1755)

The intensification recalls example (36) and Goldoni probably tries to give a Renaissance ‘flavour’ to the curse. Example (60) uses *dorso* ‘back’. The protagonists are hunchbacks (the *gobbi* of the title). The curse targets their deformity and the addition of *nel dorso* is both referential and an intensification.

Double curses

In CODIT one curse used two diseases. In Goldoni, examples (51) and (55) contain two curses.

Insults

Insults are found in (61) and here:

- (65) *Che ti venga la rabbia, lacchè del diavolo!*
 COMP 2SG.DAT come.SBJV.3SG the rabies valet.VOC of the devil
 ‘Rabies get you, valet of the devil!’ [lit. ‘That to you may come the rabies, valet of the devil!’]
(La vedova scaltra, 1748)

- (66) *Che ti venga la rabbia, ragazzo impertinente!*
 COMP 2SG.DAT come.SBJV.3SG the rabies boy.VOC insolent
 ‘Rabies get you, insolent boy!’ [lit. ‘That to you may come the rabies, insolent boy!’]
(Il medico olandese, 1756)

Not only can masters curse servants openly (see Section 3.3.3), but they can also add insults without fear of an impolite countermove. Example (69) brings the total to 4.

Emotive interjections

Interjections were frequent in CODIT; in Goldoni they occur in a small minority. 7 cases use *Oh!*:

- (67) *(Oh! che ti venga la rovella!)*
 oh COMP 2SG.DAT come.SBJV.3SG the torment
 ‘(Oh! May you be tormented!).’ [lit. ‘(Oh! That to you may come the torment!).’]
(Il ritorno dalla villeggiatura, 1761)

The only other interjection is *ah*, occurring once:

- (68) *Ah che ti venga la rabbia!* (verso il palazzino)
 ah COMP 2SG.DAT come.SBJV.3SG the rabies towards the house
 ‘Ah rabies get you! (in the direction of the house)’ [lit. ‘Ah that to you may come the rabies! (towards the house)’]
 (*Il ventaglio*, 1764)

Exclamation marks

The exclamation mark occurs in 29 cases, e.g. examples (65) to (68). While in CODIT over half of the curses had an exclamation mark, here the proportion is reduced to just over 40%.

Swearing

There are 3 examples of a swearword. In examples (61) and (65) *diavolo* is a swearword inside an insult, like in (69):

- (69) *Che ti venga la rabbia, Dottor del diavolo!*
 COMP 2SG.DAT come.SBJV.3SG the rabies doctor.VOC of the devil
 ‘Rabies get you, Doctor of the devil!’ [lit. ‘That to you may come the rabies, Doctor of the devil!’]
 (*Il cavaliere e la dama*, 1749)

Overall, intensification by means other than the exclamation mark is infrequent and nearly half of the curses, 34, have no intensification. This contrasts with CODIT, where most curses are intensified. The trend seen above for Goldoni is continuing: *che ti venga* is highly recurrent as a formula, showing little variation. There is also less creativity at the level of intensification compared to CODIT.

3.3.5 The Noun Phrase

The trend towards a reduced palette of formal choice is continuing when considering the disease nouns, as shown in figure 3:

Rabies predominates, appearing as *rabbia* and *stizza*. Given the CODIT strategy that curses invoke the most dangerous, debilitating and humiliating disease of their time, the history of rabies epidemics gives interesting insights:

In 1691, then again in 1693, a major epidemic of animal rabies (especially among dogs) was reported in several Italian provinces [. . .]. In 1779, a number of people and domesticated

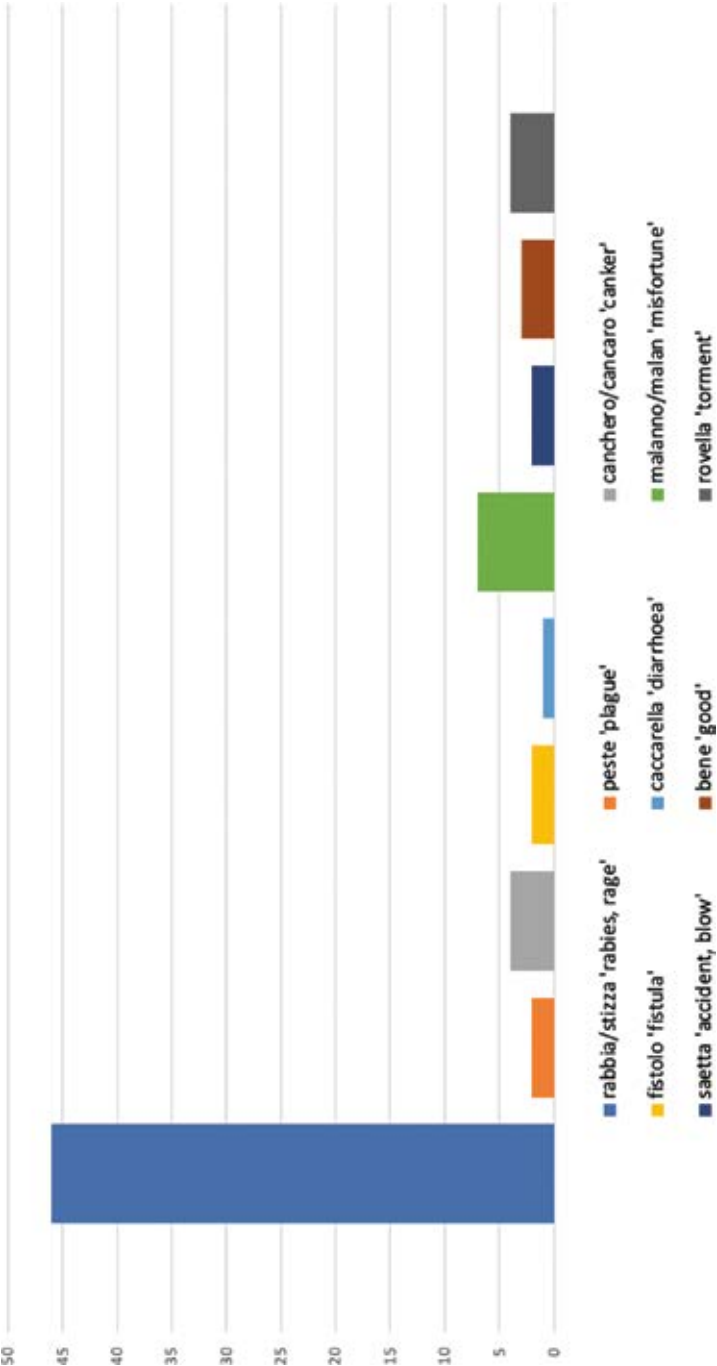


Figure 3: Distribution of different diseases and (mis)fortune in the curse *che ti venga NP* in Italian LIS. Cortina XVIII secolo – commedie Goldoni.

animals were bitten in the district of Belluno [*sic*] by a mad wolf, and many died. (King et al. 2004: 17; see also Müller et al. 2015).

Belluno was Venetian territory. Rabies, always fatal, caused mass hysteria during seventeenth- and eighteenth-century European epidemics (see Pastoureau 2018; Rao 2018 on the fear of wolves in modern times). Goldoni's dramatic oeuvre, therefore, follows the CODIT impoliteness strategy: curse with the deadliest disease available. Rabies has been described since the ancient world (King et al. 2004), and the Italian word for anger, *rabbia*, derives from the disease. For the lemma *rabbia* the GDLI (s.v.) includes the *che ti venga* curse within the first meaning of the term, 'rabies'; however, it can also be used "nel signif. generico di accidente, malanno" [in the generic meaning of accident, bad luck]. In *stizza* the metaphorical meaning of 'rage', 'fury' prevails. New is *caccarella* 'diarrhoea'.

3.3.6 Maledictive context

Outside asides and monologues, where an impolite uptake is excluded, *che ti venga* NP has a maledictive use. As in CODIT, it appears in longer arguments (70) or as an isolated angry outburst (71):

- (70) FLAMINIO *Venite qui, acchetatevi. Consento che Zelinda vi dica tutto.*
 LINDORO *Non vo' sentir altro.*
 TOGNINA *(a Lindoro) Ma ascoltateli, che vi venga la rabbia.*
 ZELINDA *(a Lindoro) Il signor Don Flaminio. . .*
 LINDORO *È un cavaliere indegno.*
 FLAMINIO *Ah temerario! se non rispettassi Zelinda!!!*
 TOGNINA *(a Flaminio) Fermatevi! (a Lindoro) Andate via!*
 LINDORO *Non crediate di spaventarmi. . . Ma saprò farmi conoscere. (parte)*
La gelosia di Lindoro, 1764³⁶

[FLAMINIO Come here, calm down. I allow Zelinda to tell you everything.
 LINDORO I don't want to hear any more. TOGNINA (to Lindoro) But listen to them, rabies get you. ZELINDA (to Lindoro) Sir Don Flaminio. . . LINDORO Is an unworthy gentleman. FLAMINIO Ah reckless man! if it weren't for Zelinda!!! TOGNINA (to Flaminio) Stop! (to Lindoro) Go away! LINDORO Don't think you scare me. . . But I'll make you respect me. (leaves)]

³⁶ See footnote 35 for the bibliographical reference.

Lindoro is jealous of Don Flaminio's attentions to his wife Zelinda, though Don Flaminio loves Barbara. Lindoro burst in on Zelinda and Don Flaminio in the presence of Tognina, Barbara's maid. Don Flaminio wants to explain, but Lindoro refuses to listen. Tognina urges Lindoro to listen, boosting her request with a curse. Zelinda starts explaining, but is interrupted by Lindoro, who insults Don Flaminio. Don Flaminio replies with another insult and threatens violence. Tognina tries to stop Flaminio, and requests Lindoro leave. Lindoro leaves, threatening revenge.

In (71) the scene represents two aristocrats. Madama di Bigné is bored and host don Alessandro brings her an almanac:

- (71) ALESSANDRO *Qui non ho che il lunario.*
 MADAMA *Oh sì, sì, quest'è un libro che divertir mi suole.*
Presto si legge, e presto si lascia, se si vuole.
 ALESSANDRO *Eccol, per obbedirvi.*
 MADAMA *Dov'è il corrente mese?*
Che vi venga la rabbia: un lunario francese?
 ALESSANDRO *Madama, non intende?*
 MADAMA *La lingua l'ho studiata*
Quindici o venti giorni, poi mi sono annoiata. Il cavalier
giocondo, 1755

[ALESSANDRO Here I have but an almanac. MADAMA Oh yes, yes, now there's a book that tends to amuse me. Soon you read it, and soon you leave it, if you want. ALESSANDRO Here it is, to obey you. MADAMA Where is the current month? Rabies get you: a French almanac? ALESSANDRO Madama, you do not understand? MADAMA I have studied the language Fifteen or twenty days, then I got bored]

When *Madama* notes the almanac is in French, she is irritated and curses her host (using respectful pronoun *vi*). Alessandro shows attentiveness with his question, and *Madama* replies cooperatively. Her anger is limited to the curse. If Don Alessandro is offended, he does not show it. In sum, when on the record, the curse appears in arguments or in isolated outburst, rarely in angry rants. It is not used for the purpose of banter.

3.3.7 Conclusion

Compared to CODIT, Goldoni uses more sentence-initial *che*, he almost always uses *venga*, and he never puts the disease before the verb. The use of *ti* is also up. There

are fewer types and occurrences of intensification and fewer disease nouns. In other words, the phrase *che ti venga* followed by a disease noun, mainly *rabbia*, is recurrent and highly formulaic in Goldoni. The predominance of *rabbia* supports the CODIT cursing strategy: rabies is always lethal, with epidemics in Italy at the end of the seventeenth century, and frequent rabid wolf attacks in the eastern Dolomites in the eighteenth century.

Goldoni's asides act as safe spaces for cursing with the downgraded pronoun *ti* indicating a temporary lapse in respect or deference. The aside is evidence that the audience is expected to recognise the curse as too impolite for the context. *Che ti venga* NP is mainly maledictive, however, the Goldoni corpus contains three benedictive cases: *che ti venga il/del bene* 'bless you', one of which also appears in CODIT. These examples will be subjected to a qualitative analysis in Section 4.

4 Benedictive examples

All examples use *che ti venga il/del bene* 'bless you', literally, 'may good come to you'.³⁷ The analysis follows a chronological order.

4.1 *La sposa sagace*, 1758, Act III, Scene 3

Donna Barbara, the daughter of Don Policarpio, has secretly pledged to marry the Count of Altomare. The servant Moschino reveals Barbara's secret to her father Policarpio:

- (72) MOSCHINO [...]
 *Finge di non curarsi di ritrovar marito;
 Eppure il matrimonio l'ha messa in appetito.
 E sa con artificio l'amante aver presente,
 E burlasi di tutti, e alcun non sa niente.*
 POLICARPIO *Oh che ti venga il bene! non lo credeva mai.*
 MOSCHINO *Ascoltate, signore, che cosa io penetrai.*

³⁷ An Italian grammar for English speakers (Roster 1875: 292) includes *Che ti venga il bene*, translated as 'Bless the man!' in a list of "Interjections", suggesting the phrase is more a discourse marker than a genuine blessing.

[MOSCHINO She pretends not to care about finding a husband; Yet she has worked up an appetite for marriage. And knows with artifice to have her lover present, And mocks all, and some know nothing. POLYCARPIO Oh bless you! I never believed it. MOSCHINO Listen, sir, to what I understood]

The revelation catches Policarpio by surprise. *Oh che ti venga il bene!* expresses gratitude towards the servant. Nevertheless, it can also be an interjection to express surprise or perplexity ('Heaven!'), without invoking a real benediction. Both interpretations are polite, whether a genuine blessing, or a sign of attentiveness towards the servant.

4.2 *Le smanie per la villeggiatura*, 1761, Act I, Scene 5

This example, (73), also figures in CODIT. Leonardo and his sister Vittoria prepare to depart for their summer residence, discussing Ferdinando. They agree he is a parasite, but Leonardo argues he enhances the family's reputation by being a perfect guest. Ferdinando, visiting, gossips about last summer, revealing he was snubbed by the aging hostess, whom he was serving as *cicisbeo*, a young man acting as a devoted companion or escort to a married woman, especially in eighteenth-century Italian society. This revelation surprises Vittoria:

- (73) VITTORIA *Oh! che ti venga il bene. Con un giovanetto di ventidue anni?*
 FERDINANDO *Sì, e mi piace di dire la verità; era un biondino, ben cincinato, bianco e rosso come una rosa.*
 LEONARDO *Mi maraviglio di lui, che avesse tal sofferenza.*
 FERDINANDO *Sapete, com'è? È uno di quelli che non hanno il modo, che si appoggiano qua e là, dove possono; e si attaccano ad alcuna di queste signore antichette, le quali pagano loro le poste, e danno loro qualche zecchino ancor per giocare.*
 VITTORIA *(È una buona lingua per altro).*
 [VITTORIA Oh! bless you. With a young man of twenty-two? FERDINANDO Yes, and to be honest, he was a blond chap, well-coiffed, all rosy and cream. LEONARDO I am surprised, that he tolerated as much. FERDINANDO You know how it is. He's one of those who can't pay their way, who lean in here and there, wherever possible; and they attach themselves to some aging ladies who pay their trips and give them a few coins to gamble with. VITTORIA (He is well-spoken, by the way).]

Che ti venga il bene expresses Vittoria's surprise and empathy, indicated by the pronoun change from *vi* to *ti*. However, like in (72), it can also function as a discourse marker signalling attentiveness. The public knows Vittoria considers Ferdinando a necessary evil: there is an element of *Schadenfreude*, that this parasite was publicly snubbed. That this may not be a genuine blessing is confirmed by Vittoria's aside to her brother, where she coldly evaluates Ferdinando's conversational skills. Nonetheless, even as a discourse marker it is still polite.

4.3 *Il ventaglio*, 1764, Act II, Scene 1

Il Signor Evaristo, in love with la Signora Candida, buys her a new fan from Susanna's haberdashery. Evaristo asks Giannina, a peasant woman, to secretly deliver it. In her monologue Susanna is frustrated: although she considers herself middle-class, the villagers do not distinguish between her and local peasant women. Susanna is especially jealous of Giannina, having seen Evaristo give her the fan. The passage is sarcastic:

- (74) *Gli hanno donato un ventaglio! Cosa vuol fare una contadina di quel ventaglio? Oh, farà la bella figura! Si farà fresco. . . la. . . così. . . Oh, che ti venga del bene! Sono cose da ridere; ma cose che qualche volta mi fan venire la rabbia. Son così, io che sono allevata civilmente, non posso soffrire le male grazie. (siede e lavora)* [They gave her a fan! What does a peasant girl want to do with that fan? Oh, she'll cut a nice figure! She'll far herself. . . so and. . . so. . . Oh, may it bring you good! It's ridiculous; but these things sometimes make me angry. That's me. I am well-educated and cannot suffer bad manners. (sits and works)]

Oh, che ti venga del bene! follows an "unpalatable question", a challenging rhetorical question (Culpeper 2011: 135). The implied answer to *Cosa vuole fare una contadina di quel ventaglio?* is 'nothing, she will have no use for it at all'. *Oh farà la bella figura!* is sarcastic, mocking the idea of a peasant woman fanning herself. Susanna then mimics Giannina, pronouncing the sarcastic blessing 'may it bring you good!' meaning 'damn you!'. Susanna finds the situation laughable and is angry about Giannina crossing class boundaries. By changing from the third-person reference to second-person address *ti*, Susanna makes the curse more engaging (as a mock dialogue within a monologue).

4.4 Conclusion

Che ti venga il/del bene can be a genuine blessing expressing gratitude or commiseration, or a discourse marker indicating attentiveness after a revelation. Both are polite. However, it can also be sarcastic, expressing the speaker's anger. The rarity of *che ti venga il/del bene* (3 out of 71 Goldoni curses), with one being clearly sarcastic, supports the hypothesis that *che ti venga* NP is strongly biased for impoliteness (see also Dobrushina, this volume, who finds strong constructional biases in Nakh-Daghestanian curses too).

5 Metapragmatic evidence of conventionalisation

Metapragmatic evidence of conventionalisation comes from historical textbooks for learners of Italian as a foreign language. The late-seventeenth-century French-Italian textbook *Le Maître italien* by Giovanni Veneroni enjoyed “numerous further editions and adaptations until the end of the eighteenth century” (Betsch 2019: 41–42). The manual includes a section called *Pour souhaiter du mal* ‘to wish ill’ (Veneroni 1687⁵: 327) listing 11 formulae, three of which include (*che*) *ti venga*: *ti venga la rabbia*; *che ti venga il canchero*; *che ti venga la cacarella* ‘may you get rabies, a canker, the quirts’, which are proposed as translations for the French *La rage te puisse venir*; *qui te puisse venir un chancre*; *puisse tu avoir la foire*. Notably, Veneroni’s list of blessings lacks the *che ti venga* phrase. Similarly, a textbook for British learners of Italian has no *che ti venga* phrases in blessings, but the list of formulae *To wish Ill* includes three occurrences of *che ti venga* with a disease: “*Ti venga la peste* or *la rabbia* pox or plague take ye”; “*Ti venga il canchero* mayst thou rot or be rotted” (Palermo 1755: 342–343).³⁸

These historical textbooks confirm the impoliteness interpretation of *che ti venga* NP: “wish ill”; “souhaiter du mal”. The presence of the formula further underscores its status as a conventionalised expression. Moreover, the remarkable editorial success of Veneroni’s manual may have contributed to its wider conventionalization. For seventeenth- and eighteenth-century learners of Italian these textbook entries served as prescriptive mentions of linguistic usages, providing the kind of indirect experience Culpeper identifies as crucial to the conventionalisation process of impoliteness formulae (§ 2.1).

38 Palermo’s grammar is partially based on Veneroni’s (De Gasperin 2016).

Similarly, the use of *che ti venga* NP in theatre may have had a comparable conventionalising effect on audiences, as well as on readers of the novels and poetry in which the expression appears.

6 Concluding remarks

This essay demonstrated that Italian *che ti venga* NP is a conventionalised impoliteness formula for disease curses between the fourteenth and the twentieth century. These curses were not only impolite but also blasphemous, sinful, and legally punishable, posing risks with the cursed, the law, and God. Evidence comes from legislative, judicial, theological, and linguistic metasources. Quantitative data from two historical corpora show that most of the 132 occurrences are maledictive: appearing in arguments, they invoke the disease that is most lethal and debilitating at the time; they are exacerbated by further use of taboo words, in insults, swearwords, body orifices, etc. Very few uses are mock-impolite or polite. The corpora produced slightly different findings. In Goldoni, the phrase *che ti venga* NP appears more formulaic and less intensified. Goldoni's curser uses asides with a 'downgraded' address pronoun. I took the asides to be evidence of the conventionalisation for impoliteness with the audience. One limitation of this study is the difference in textual genre between the two corpora analysed. The higher degree of formulaicity observed in Goldoni's works may be attributed to the fact that this corpus consists only of theatrical texts by a single author, whereas the CODIT corpus encompasses a broader range of genres over a significantly longer time span. This discrepancy in corpus composition may influence the results, making direct comparisons more challenging. The CODIT includes alternative disease curses showing a closer association between one verb and one disease. Of 13 curses with *nascere* 'to be born' in CODIT, 9 have *vermocane* 'gid, staggers' – further examples are (15) and (16), which originate outside CODIT –, whilst *venire* allows for a large variety of Noun Phrases. Ultimately, to curse someone with a disease, the go-to phrase is *che ti venga*. It is a construction specialised for this purpose.

Historical contrastive pragmatics is a promising research avenue (e.g. López-Simo 2023). The English Historical Book Collection (EEBO, ECCO, Evans) corpus on Sketch Engine returns curses such as: 'the pox take you'; 'that the gallows take you'; 'the devil take you'; 'a vengeance take you all'; 'a murrain take you'; 'plague take you all'. The verb *to take* has the disease as the Subject with the cursed undergoing the action as Direct Object. The curser is, grammatically speaking, not part of the curse, just as with Italian *venire*.

This applies for present-day disease curses, offering scope for further contrastive research. Italian disease curses still use *che ti venga* NP. The Italian Web 2020 corpus (itTenTen20) returns, amongst others, these curses: *Che ti venga un attacco di tourette davanti al vescovo mentre fai la cresima a tuo figlio*. [May you have an attack of tourette's in front of the bishop when you present him your son to be anointed]; *Che ti venga un accidente, brutta puttana* [An accident take you, ugly whore]; *Che ti venga un infarto bastardo bastardoBASTARDO!!!* [A heart attack take you, bastard bastardBASTARD!!!!]. Note the following metapragmatic comments, also from Italian Web 2020:

- (75) *C'at vegna un Ocse (che ti venga un Ocse, come fosse una malattia incurabile)*
[May an OECD take you (as if it were an incurable disease)]

When the OECD (Organisation for Economic Co-operation and Development) met in Bologna in 2000, a local protester wore a T-shirt with this slogan. Using *che ti venga* with a Noun Phrase turns the noun *ipso facto* into a disease. I also found *che ti venga del bene*. Spooked by a dog, a man utters a curse:

- (76) rientro rapidamente con “c'at vena un cancher” (trad. dal dialetto: “che ti venga del bene!”) [I quickly go back inside with “may you rot” (translated from dialect: “bless you!”)]

Che ti venga del bene is clearly impolite and a pragmatic reversal may have taken place, as is the case for uses with *benedetto* ‘blessed’ as in *benedetto ragazzo* ‘foolish boy’.

The Dutch Wikipedia page on *verwensingen* ‘curses’ states that disease curses are “typisch Nederlands” [typically Dutch].³⁹ The Dutch Web 2020 corpus (nlTenTen20) contains numerous examples (search string *krijg de* ‘get the’). Here is a handful: *Krijg de tyfus* [get typhoid fever]; *krijg de tering* [get consumption]; *krijg de kanker* [get cancer]; *krijg de pest* [get the plague]; *krijg de jicht* [get gout]. With the verb *krijgen* ‘to get, receive’ the cursed is undergoing the action, and the curser is not part of the curse.

Disease curses prefer a dangerous disease in a syntactic structure that avoids giving agency to the curser. Contrastive pragmatics may yet bring up more parallels between more languages.

39 <https://nl.wikipedia.org/wiki/Verwensing> (accessed 8 January 2025)

Acknowledgements: I am very grateful to the reviewers and editors, especially Jonathan Culpeper and Daniel Van Olmen, for their insightful comments, which greatly improved the paper (the usual disclaimers apply). I specifically thank Federico Barbierato and Pierroberto Scaramella for providing me the opportunity to engage with historians of blasphemy in Italy during the workshop *Cursing God, Blasphemy and its Repression in Early Modern Italy (15th-19th Century)* organised at the University of Bari, December 2024.

References

Sources

- Aquinas, Thomas. 1920–1942. *The “Summa Theologica” of St. Thomas Aquinas*. Literally translated by Fathers of the English Dominican Province. 2nd and revised edn. London: Burns Oates and Washbourne. 22 vols. Retrieved from <https://archive.org/details/in.ernet.dli.2015.506372/page/n5/mode/1up>.
- Ariosto, Ludovico. 2007 [1528]. *La Lena*. Edited by Stefano Bianchi. Milan: Rizzoli.
- Goldoni, Carlo. 1955. *Tutte le opere*. Edited by Giuseppe Ortolani. Milan: Mondadori.
- Maria, Gaetano. 1732. *L'uomo apostolico istruito nella sua vocazione al confessionario colle avvertenze de Sacri canoni, e de Santi padri, massimamente di San Carlo Borromeo*. Brescia: Rizzardi.
- Palermo, Evangelist. 1755. *A Grammar of the Italian Language*. London: A. Millar.
- Roster, P.L. 1875. *A new Italian grammar on a plan which will greatly facilitate a practical as well as a theoretical knowledge of the language*. Florence: Bencini.
- Veneroni, Giovanni. 1687⁵. *Le maître italien* [. . .]. Cinquième édition. *Revue, corrigée et augmentée par l'Auteur*. Paris: Étienne Loyson.

Dictionaries

- Battaglia, Salvatore. 1961–2002. *Grande dizionario della lingua italiana*. Turin: UTET. <https://www.gdli.it/>.
- Florio, John. 1598. *A worlde of wordes, or, Most copious, and exact dictionarie in Italian and English*. London: Arnold Hatfield for Edward Blount. Retrieved from <https://archive.org/details/worldofwordesor00flor/page/168/mode/2up>.

Corpora and Datasets

- Corpus diacronico dell'italiano (CODIT). Ústav Českého národního korpusu FF UK. 2021. Compiled by M. Silvia Micheli, Prague, <http://www.korpus.cz>.
- Dutch Web (nlTenTen20). 2020. on <https://www.sketchengine.eu/>.

- English Historical Book Collection (EEBO, ECCO, Evans) on <https://www.sketchengine.eu/>.
 Gutenberg Italian. 2020. on <https://www.sketchengine.eu/>.
 Italian Web (itTenTen20). 2020. on <https://www.sketchengine.eu/>.
 ItalianLISCortIta XVIII secolo – commedie Goldoni. Compiled by Chiara Ghezzi, shared on <https://www.sketchengine.eu/>.
 Paternoster, Annick. 2025. The conventionalised impoliteness formula 'che ti venga NP' in two historical corpora, CODIT (13th century to 1947) and Goldoni's drama works (2nd half 18th Century) (Version 1.0) [Data set]. LaRS – Language Repository of Switzerland.

Secondary literature

- Alfonzetti, Giovanna. 2009. Gli insulti: alcuni criteri di categorizzazione. In Salvatore Carmelo Trovato (ed.), *Studi linguistici in memoria di Giovanni Tropea*, 67–78. Alessandria: Edizioni dell'Orso.
- Alfonzetti, Giovanna. 2017. *Questioni di (s)cortesia. Complimenti e insulti*. Avellino: Sinestesie Edizioni.
- Alfonzetti, Giovanna & Margherita Spampinato Beretta. 2012. Gli insulti nella storia dell'italiano. Analisi di testi del tardo medioevo. In Barbara Wehr & Frédéric Nicolosi (eds.), *Pragmatique historique et syntaxe: actes de la section du même nom du XXXIe Romanistentag allemand, Bonn, 27 septembre-1er octobre 2009*, 1–21. Frankfurt-Berlin: Peter Lang.
- Arnovick, Leslie K. 1999. *Diachronic Pragmatics: Seven Case Studies in English Illocutionary Development*. Amsterdam: Benjamins.
- Baker, Eleonor. 2023. *Book Curses*. Oxford: Bodleian Library.
- Betsch, Michael. 2019. Address forms in grammars and textbooks of West and South Slavic languages in the eighteenth and nineteenth century. In Annick Paternoster & Susan Fitzmaurice (eds.), *Politeness in Nineteenth-Century Europe*, 39–74. Amsterdam/Philadelphia: John Benjamins.
- Biasiori, Lucio. 2024. Taking the Name of the Lord in Vain in Early Modern Italy. *Church History and Religious Culture* 104. 340–358.
- Bremmer, Rolf H. Jr. 1998. Insults hurt: verbal injury in late medieval Frisia. In Rolf H. Bremmer Jr, Thomas S.B. Johnston & Oebele Vries (eds.), *Approaches to Old Frisian Philology*, 89–112. Amsterdam/ New York: Rodopi.
- Breschi, Giancarlo. 1994. Atti criminali recanatesi, edizione commentata. In Francesco Bruni (ed.), *L'italiano nelle regioni. Testi e documenti*, 484–486. Turin: UTET.
- Burke, Peter. 1987. Insult and blasphemy in Early Modern Italy. In Peter Burke, *The Historical Anthropology of Early Modern Italy: Essays on Perception and Communication*, 95–109. Cambridge: Cambridge University Press.
- Casagrande, Carla & Silvana Vecchio. 1987. *I peccati della lingua. Disciplina ed etica della parola nella cultura medievale*. Rome: Istituto della Enciclopedia italiana.
- Culpeper, Jonathan. 2011. *Impoliteness. Using language to cause offence*. Cambridge: Cambridge University Press.
- Culpeper, Jonathan. 2018. *Taboo language and impoliteness*. In Keith Allan (ed.), *The Oxford Handbook of Taboo Words and Language*, 28–40. Oxford: Oxford University Press.
- Culpeper, Jonathan & Michael Haugh. 2021. The metalinguistics of offence in (British) English: A corpus-based metapragmatic approach. *Journal of Language Aggression and Conflict* 9(2). 185–214.
- Culpeper, Jonathan & Merja Kytö. 2010. *Early Modern English Dialogues. Spoken Interaction as Writing*. Cambridge: Cambridge University Press.

- Culpeper, Jonathan & Elena Semino. 2000. Constructing witches and spells. Speech acts and activity types in Early Modern England. *Journal of Historical Pragmatics* 1(1). 97–116.
- Danet, Brenda & Bryna Bogoch. 1992. “Whoever alters this, may God turn his face from him on the day of Judgement”, Curses in Anglo-Saxon legal documents. *Journal of American Folklore* 105. 132–65.
- Dardano, Maurizio, Claudio Giovanardi & Massimo Palermo. 1992. Pragmatica dell’ingiuria nell’italiano antico. In Giovanni Gobber (ed.), *La linguistica pragmatica. Atti del XXIV Congresso della Società di Linguistica Italiana* (Milano, 4–6 settembre 1990), 3–37. Rome: Bulzoni.
- Dean, Trevor. 2007. *Crime and Justice in Late Medieval Italy*. Cambridge: Cambridge University Press.
- De Gasperin, Vilma. 2016. Rules and Grammars of Italian in Eighteenth-century England: The Case of Giuseppe Baretti. *The Italianist* 36(3). 429–446.
- Delumeau, Jean (ed.). 1989. *Injures et blasphèmes*. Paris: Imago.
- Domaneschi, Filippo. 2020. *Insultare gli altri*. Turin: Einaudi.
- D’Onghia, Luca. 2011. Aspetti della lingua comica di Giovan Battista Andreini. *La lingua italiana. Storia, strutture, testi* 7. 57–80.
- Duranti, Alessandro. 2004. Agency in language. In Alessandro Duranti (ed.), *A Companion to linguistic anthropology*, 451–473. Malden, MA: Blackwell.
- Evans, Nicholas. 2007. Insubordination and its uses. In Irina Nikolaeva (ed.), *Finiteness: Theoretical and Empirical Foundations*, 366–431. Oxford: Oxford University Press.
- Fantappiè, Renzo (ed.). 2000. *Nuovi testi pratesi dalle origini al 1320*, 2 vols. Florence: Accademia della Crusca.
- Fisher, John. 1908. Cursing. In *The Catholic Encyclopedia*. Vol. 4. New York: Robert Appleton Company. Retrieved via <https://www.newadvent.org/cathen/04573d.htm>.
- Gager, John G. 1992. Introduction. In John G. Gager (ed.), *Curse Tablets and Binding Spells from the Ancient World*, 3–41. New York/Oxford: Oxford University Press.
- Gonthier, Nicole. 2007. «Sanglant Coupaul!» «Ordre Ribaude!» *Les injures au Moyen Âge*. Rennes: Presses Universitaires de Rennes.
- Ghezzi, Chiara. 2021. Verticality and horizontality in Italian address forms. In Chiara Ghezzi, Piera Molinella & Luis Unceta Gómez (eds.), *Politeness between cognition and culture from Latin to Italian and Spanish*. [Special Issue] *Studi Italiani di Linguistica Teorica e Applicata* 50(3). 648–666.
- Greco, Angelo. 2021. “Augurare la morte sui social è reato?” https://www.laleggepertutti.it/483685_augurare-la-morte-sui-social-e-reato (accessed 8 January 2025).
- Jucker, Andreas H. 2020. *Politeness in the History of England. From the Middle Ages to the Present Day*. Cambridge: Cambridge University Press.
- Jucker, Andreas H. & Irma Taavitsainen. 2000. Diachronic Speech Act Analysis: Insults from Flyting to Flaming. *Journal of Historical Pragmatics* 1(1). 67–95.
- Kitz, Anne Marie. 2007. Curses and Cursing in the Ancient Near East. *Religion Compass* 1. 615–627.
- Kitz, Anne Marie. 2014. *Cursed Are You! The Phenomenology of Cursing in Cuneiform and Hebrew Texts*. Winona Lake: Eisenbrauns.
- Kádár, Dániel Z. & Andrea Szalai. 2020. The socialisation of interactional rituals: A case study of ritual cursing as a form of teasing in Romani. *Pragmatics* 30. 15–39.
- King, Arthur A., Anthony R. Fooks, Michel Aubert & Alexander I. Wandeler (eds.). 2004. *Historical Perspective of Rabies in Europe and the Mediterranean Basin*. Paris: OIE.
- Lagorgette, Dominique. 1994. Termes d’adresse, acte perlocutoires et insultes: la violence verbale dans quelques textes des XI^e, XV^e et XVI^e siècles. *Senefiance* 36. 317–332.

- Lagorgette, Dominique. 2003. Les syntagmes nominaux d'insulte et de blasphème: analyse diachronique du discours marginalisé. [Special Issue]. *Thélème: Revista complutense de estudios franceses*. 171–188.
- Larson, Pär. 2004. Ingiurie e villanie dagli Atti podestarili pistoiesi del 1295. *Bollettino dell'Opera del Vocabolario italiano* 9. 349–354.
- Lesnick, Daniel R. 1991. Insults and threats in medieval Todi. *Journal of Medieval History* 17(1). 71–89.
- Lett, Didier. 2018. *Crime, genre et châtement: Hommes et femmes face à la justice au Moyen Age (XIIe–XVe siècle)*. Paris: Armand Colin.
- Limberg, Holger. 2009. Impoliteness and threat responses. *Journal of Pragmatics* 41(7). 1376–1394.
- Little, Lester K. 1993. *Benedictine Maledictions: Liturgical Cursing in Romanesque France*. Ithaca/London: Cornell University Press.
- López-Simo, Mireia. 2023. Formules votives négatives en français et en espagnol: de la malédiction à l'expulsion. *Estudios de Traducción* 13: 21–35.
- Madero, Marta. 1992. *Manos violentas, palabras vedadas. La injuria en Castilla y León (siglos XIII–XV)*. Madrid: Taurus Humanidades.
- Marcheschi, Daniela. 1983. *Ingiurie, impropri, contumelie ecc. Saggio di lingua parlata del Trecento cavato dai libri criminali di Lucca per opera di Salvatore Bongi. Nuova edizione rivista e corretta*. Lucca: Pacini Fazzi.
- Mattarese, Tina. 2010. Goldoni, Carlo. In Istituto della Enciclopedia Italiana (ed.), *Enciclopedia dell'italiano*, [https://www.treccani.it/enciclopedia/carlo-goldoni_\(Enciclopedia-dell'Italiano\)/](https://www.treccani.it/enciclopedia/carlo-goldoni_(Enciclopedia-dell'Italiano)/) (accessed 8 January 2025).
- Micheli, M. Silvia. 2022. CODIT. A new resource for the study of Italian from a diachronic perspective: Design and applications in the morphological field. *Corpus* 23.
- Molinelli, Piera. 2018. Sociocultural and Linguistic Constraints in Address Choice from Latin to Italian. In Kate Beeching, Chiara Ghezzi & Piera Molinelli (eds.), *Positioning the Self and Others. Linguistic Perspectives*, 51–79. Amsterdam: John Benjamins.
- Müller, Thomas, Conrad Martin Freuling, Patrick Wysocki, Micha Roumiantzeff, Jean Freney, Thomas Christoph Mettenleiter & Adriaan Vos. 2015. Terrestrial rabies control in the European Union: historical achievements and challenges ahead. *The Veterinary Journal* 203(1). 10–77.
- Nash, David. 2007. *Blasphemy in the Christian World. A History*. Oxford: Oxford University Press.
- Pastoreau, Michel. 2018. *Le loup. Une histoire culturelle*. Paris: Le seuil.
- Paternoster, Annick. 2015. *Cortesi e scortesi. Percorsi di pragmatica storica da Castiglione a Collodi*. Rome: Carocci.
- Radtke, Edgar. 1994. *Gesprochenes Französisch und Sprachgeschichte. Zur Rekonstruktion der Gesprächskonstitution in Dialogen französischer Sprachlehrbücher des 17. Jahrhunderts unter besonderer Berücksichtigung der italienischen Adaptionen*. Tübingen: Max Niemeyer.
- Ramos, Melissa Dianne. 2015. *Spoken Word and Ritual Performance: The Oath and the Curse in Deuteronomy 27–28*. Los Angeles: University of California Los Angeles dissertation.
- Rao, Riccardo. 2018. *Il tempo dei lupi*. Turin: UTET.
- Raveggi, Sergio. 2018. Il lessico delle ingiurie contro le donne. In Anna Esposito, Franco Franceschi & Gabriella Piccinni (eds.), *Violenza alle donne. Una prospettiva medievale*, 129–150. Bologna: Il mulino.
- Salvi Giampaolo & Lorenzo Renzi (eds). 2010. *Grammatica dell'italiano antico*. 2 vols. Bologna: Il Mulino.
- Searle, John R. 1979. *Expression and Meaning*. Cambridge: Cambridge University Press.
- Szalai, Andrea. 2023. When a curse is not impolite: Insistence in rites of hospitality in Romani interaction. Paper presented at SymPol 2023, 14th Symposium on Politeness. (Im)Politeness studies: new frontiers, synergies and perspectives, Universidade da Madeira, 3–5 July, 2023.

- Tardivel, Chloé. 2019. Le délit d'injure verbale d'après les statuts communaux de la Marche d'Ancône (Italie, XIVe–XVe siècles). *Questes* 41. 85–105.
- Tardivel, Chloé. 2021. Giudicare la violenza verbale alla fine del Medioevo. In Didier Lett (ed.), *I registri della giustizia penale nell'Italia dei secoli XII–XV*, 301–320. Rome: Publications de l'École française de Rome.
- Tardivel, Chloé. Forthcoming. Words and Deeds: Gender and Social Practices of Verbal Violence in Bologna's Trial Records (1350–1385). Gender, Networks, and Communities in Medieval Europe, <https://hal.archives-ouvertes.fr/hal-02483828v2> (accessed 8 January 2025).
- Terkourafi, Marina. 2001. *Politeness in Cypriot Greek: A Frame-Based Approach*. Cambridge: University of Cambridge dissertation.
- Terkourafi, Marina. 2005. 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/New York: Mouton de Gruyter.
- Terkourafi, Marina & Dániel Kádár. 2017. Convention and Ritual (Im)Politeness. In Jonathan Culpeper, Michael Haugh & Dániel Kádár *The Palgrave Handbook of Linguistic (Im)Politeness*, 171–195. London: Palgrave Macmillan.
- Trifone, Pietro. 2000. *L'italiano a teatro. Dalla commedia rinascimentale a Dario Fo*. Pisa/Roma: Istituti editoriali e poligrafici internazionali.
- Trifone, Pietro. 2019. Il teatro tra dire e fare. In Nicola De Blasi & Pietro Trifone, *L'italiano sul palcoscenico*, 7–19. Firenze: Accademia della Crusca.
- Vecchio, Silvana. 2014. Légitimité et efficacité de la malédiction dans la réflexion théologique médiévale. In Nicole Bériou, Jean-Patrice Boudet & Irène Rosier-Catach (eds.), *Le pouvoir des mots au Moyen Âge: Études réunies*, 349–361. Turnhout: Brepols.
- Vise, Melissa E. 2015. *The Threat of the Tongue: illicit speech in Late Medieval Italy, 1250–1450*. Evanston (IL): Northwestern University dissertation.
- Vise, Melissa E. 2025. *The Unruly Tongue. Speech and Violence in Medieval Italy*. Philadelphia: University of Pennsylvania Press.

Rita Finkbeiner

6 *Wehe* ‘woe’ + verb-second conditional clause in German: A conventionalized threat construction?

Abstract: This paper is concerned with the German interjection *wehe* (‘woe’) and its potential role in indicating the illocution of a threat. While it is well-known that *wehe* can have a lamenting usage and a usage as prediction of calamity, also attested for English *woe*, less is known about the usage of German *wehe* in threats. The paper starts with the observation that *wehe* has various usage patterns, one of them being the usage as a matrix element embedding a conditional clause. This usage can receive either a prediction or a threat reading. Based on a discussion of the functional and formal properties of threatening in German and a close description of the grammatical properties of *wehe*, a corpus study is carried out to investigate whether there is evidence for a more particular pattern of *wehe* embedding conditional clauses that is conventionally linked to a threat reading. The results of the study provide initial evidence for the existence of a syntactic construction *wehe* + verb-second conditional clause with 2nd person subject that is reserved for threats. Still, the interpretation of instances of this construction as threats is context-dependent to some degree, suggesting an analysis as illocutionary force indicator in the sense of Searle, i.e. a linguistic means signaling a threatening illocution by default, which can be overridden in context.

Keywords: corpus study, interjection, illocutionary indicator, syntactic construction, threat

1 Introduction

It is a standard assumption in research on the interaction between grammar and pragmatics that there are certain default relations between (types of) linguistic means and (types of) illocutions. Searle (1969: 30) introduced the notion of “explicit illocutionary force indicators” to capture this insight. The notion of “indicator” reflects that this relationship is conceived to be relatively stable but can still be overridden in context. The most prototypical explicit illocutionary force indicator, according to

Rita Finkbeiner, Johannes Gutenberg University Mainz, Jakob-Welder-Weg 18, 55128 Mainz, Germany, e-mail: finkbeiner@uni-mainz.de

Searle (1969), is the explicit performative construction (e.g., uttering *I promise, I'll do my homework* to make a promise). While there is a relatively stable relationship between the explicit performative construction *I promise* and promises, this relationship can be overridden in context, e.g. if a speaker uses *I promise* not in a promise, but in an affirmation, as in *I promise, it wasn't me*. Beyond performative verbs, other means such as sentence types, verbal mood, modal particles, and prosody have been shown to play an important role in signaling illocutionary force. For example, most languages avail themselves of an interrogative sentence type signaling the illocutionary force of a question (Wunderlich 1976; Siemund 2018). Similarly, there are systematic links in the world's languages between imperative sentences and directive speech acts, as well as between declarative sentences and assertions. On the other hand, it has long been taken for granted that most expressive speech acts, particularly those used in impolite interaction, such as insults or reproaches, do not have specific grammatical correlates, but receive their interpretation primarily based on socio-pragmatic and interactional factors (Mills 2003; Locher and Watts 2008).

More recently, though, particular grammatical constructions have been identified in various languages that seem to be conventionally associated, to varying degrees, with impolite speech acts (cf. Finkbeiner, Meibauer, and Wiese 2016). For example, Van Olmen, Andersson, and Culpeper (2023) investigate the insulting construction *You*+NP in English, Dutch, and Polish, finding a great degree of conventionalization across these three languages (see also Van Olmen et al., this volume, and Van Olmen and Andersson, this volume). Van Olmen (2018) examines the relationship between imperative constructions and the function of reproaches in Dutch and other languages of Europe, finding evidence for a hybrid reproachative construction. As for threats in English, Muschalik (2018) finds empirical evidence for the existence of a set of recurrent linguistic means including formulae like *you better*, swear words, 'violent verbs' (e.g., *hit, beat, smash*), and conditionals (cf. also Culpeper 2011), rebutting Limberg's (2009: 1378) categorical claim that "[t]hreats do not come in a standardized linguistic format" (see also Dobrushina, this volume, on curses in Nakh-Daghestanian languages).

For German, the literature on threats is sparse, and very little is known as to the potential grammatical correlates of threats. Existing work (e.g., Apeltauer 1977, Falkenberg 1992) mostly starts out by listing constructed examples taken to illustrate prototypical threat utterances. Interestingly, these lists often include utterances containing the interjection *wehe* ('woe'), cf. (1)-(2).

- (1) *Wehe, wenn du das machst.*
 woe if you that make.PRS.2SG
 'Don't you dare do that'
 (Apeltauer 1977: 188)

- (2) *Wehe, du lässt dich hier noch einmal blicken!*
 woe you let.PRS.2SG you.ACC here else once see
 'Don't you dare show your face here once again'
 (Falkenberg 1992: 178)

In contrast to German, the English interjection *woe* cannot immediately embed a conditional clause. This is only possible if *woe* is part of a complete matrix clause *woe betide* (+ object) [+ if-clause], with the usage without object being described in the Oxford English Dictionary (henceforth OED) as "archaic" and "rare" (cf. reading P.3. in the OED). In contrast, the Modern German examples (1)–(2) are completely unmarked choices if one wants to express a threat. Furthermore, the usage of *woe betide* + object is characterized in the OED as "colloquial with weakened sense 'you (he, etc.) will get into trouble (if . . .)'"'. The Modern German examples (1)–(2), however, are clear, unattenuated threats. The most adequate translation for (1)–(2) therefore seems to be one that uses an English threat expression such as 'don't you dare'.

While Apeltauer (1977) does not say anything about the status of *wehe*, Falkenberg (1992: 179), in a side-remark, suggests that *wehe* might be one of few existing lexical indicators of threats in German. However, this suggestion so far lacks systematic empirical substantiation. What is more, from the perspective of a "grammar of impoliteness" research agenda, examples (1)–(2) raise the question whether an analysis of *wehe* as a *lexical* illocutionary indicator of threats is on the right track. Crucially, *wehe* does not unequivocally indicate threats, but has – as will be demonstrated below – a range of other usages, e.g., in lamentations. On the other hand, if one takes a closer look at (1) and (2), one finds that in both cases, *wehe* is used as a sentence-initial matrix element embedding a clause, and that in both (1) and (2), the embedded clause – a *wenn*-clause in (1), a verb-second (henceforth V2) clause in (2) – receives a conditional reading. This suggests that the threat reading of (1) and (2), rather than being indicated by *wehe* in itself, might be indicated by the larger syntactic construction it appears in, namely, the syntactic construction '*wehe* + embedded conditional clause'.

However, one can easily find uses of *wehe* + embedded conditional clause that are not threats. By uttering (3), the speaker does not threaten anyone but merely foreshadows negative consequences of the possible event that the radioactive cloud moves into the direction of Tokyo.

- (3) *Wehe, wenn der Wind dreht und die radioaktive*
 woe if the wind turns and the radioactive
Wolke damit Richtung Ballungszentrum Tokio weht
 cloud therewith direction congested-area Tokyo waves

‘Woe betide if the wind turns and the radioactive cloud waves into the direction of the metropolitan area of Tokyo’
(St. Galler Tagblatt, 03/22/2011)

The embedded clause in (3) is a *wenn*-clause. Based on the examples (1)–(3), one might thus hypothesize that the utterance of *wehe* + embedded *wenn*-clause may convey a threat reading, cf. (1), or a predictive reading, cf. (3), while utterances of *wehe* + embedded V2 clause might be more restricted in their illocutionary potential, preferably conveying threats.

Starting from these initial observations, the question arises whether there is indeed a particular syntactic construction consisting of *wehe* as a matrix element and an embedded V2 conditional clause that functions as a *grammatical* illocutionary indicator of threats, and if so, how its linguistic properties should be specified. To find an answer to this question, I will empirically investigate the usage of *wehe* in a large, annotated corpus of *wehe*-utterances compiled from DeReKo (German Reference Corpus). As to the theoretical framework used, I adopt a Construction Grammar approach (Goldberg 2013), which regards syntactic constructions as complex, conventionalized form-function pairings. Conventionalization can be defined with Terkourafi (2005: 213) as “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”. To assess the conventionalization status of a potential threat construction, a feasible way is to show that the majority of utterances containing this particular construction in a given corpus are threats. Note that this definition of a conventionalized construction still allows for the existence of utterances of the construction that are not threats. It is thus compatible with the notion of illocutionary indicator as a default grammatical means for a particular illocutionary function that can be overridden in context.

The structure of the chapter is as follows. In Section 2, I go into the basic functional properties of threats, with a focus on German, and into their potential grammatical correlates. In Section 3, I take a closer look at the grammatical properties of *wehe*. In Section 4, I present the results of the corpus study on *wehe*, providing a quantitative and qualitative analysis and arguing for the existence of a conventionalized syntactic threat construction ‘*wehe* + V2 clause with 2nd person subject’. Section 5 concludes.

2 Threats: From function to form

Threatening can be defined as “a linguistic strategy that is used to manipulate or even coerce the addressee into (not) doing something which has an undesirable outcome for him/her [i.e., the speaker], where “[i]n case of the addressee’s non-compliance, the threatener may initiate negative consequences directly or indirectly as a kind of punishment for non-cooperation” (Limberg 2009: 1378). More generally, the communicative function of threats has been characterized in pragmatics both “as a means of manipulation, [. . .] a form of (verbal) impoliteness and [. . .] as an exercise of power by a speaker” (Muschalik 2018: 20). If one takes impoliteness to refer to “language that is used to cause offence” (Culpeper 2018), or to “strategies that are oriented towards attacking face” (Culpeper 1996: 350), threats are impolite speech acts in that they involve “coercive action that is not in the interest of the target, and hence involves both the restriction of a person’s action environment and a clash of interests” (Culpeper 2011: 226; cf. also Falkenberg 1992: 182).

In most pragmatic approaches to threats, the fact is acknowledged that threats combine both a commissive aspect, i.e., the speaker’s conditional commitment to punish the addressee in case of non-compliance, and a directive aspect, i.e., the speaker’s attempt to make the addressee do something (or refrain from doing something). What is sometimes neglected, however, is the expressive aspect of threats, explicitly put by Falkenberg (1992: 187) as “*Drohungen bestehen im Einflößen von Furcht oder von ‘negativ’ besetzten affektiven Einstellungen*” (‘Threats consist in instilling fear or ‘negatively’ loaded affective attitudes’, my translation; cf. also Fraser 1998: 161). Rolf (1997: 223) proposes an account according to which expressive speech acts essentially are attempts to affect the emotional state of the addressee. The expression of a certain psychological state by the speaker – Searle’s (1978) original essential condition of expressive speech acts – is modeled by Rolf as the sincerity condition of expressive speech acts. Under Rolf’s definition, the expressive aspect of threats can be seen in them being attempts to instill fear into the addressee. It seems that it is exactly this expressive aspect which relates threats to impoliteness – in contrast, for example, to warnings, which resemble threats in many ways, but lack an expressive aspect, and usually are not taken to be impolite speech acts. The expressive aspect of threats is also reflected, e.g., in the fact that threats can be assigned to “output strategies” such as “frighten[ing]” and “invading the other’s space” in Culpeper’s list of negative impoliteness strategies (cf. Culpeper 1996: 358).

Current research into threatening in English suggests that there are certain linguistic means that are conventionally associated, to a greater or lesser degree, with the interpretation of the utterance as a threat (e.g., Culpeper 2011; Muschalik 2018; Culpeper, Van Dorst and Gillings, this volume). For example, Culpeper

(2011: 136) lists the following conventionalized impoliteness formula for threats in English:

- (4) [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]
- (5) [you'd better be ready Friday the 20th to meet with me/do it] [or] [else] [I'll] [X]
- (6) [X] [before I] [hit you/strangle you]

In (4)–(6), several structural features become apparent that have been systematically examined on a broad empirical basis by Muschalik (2018). She starts out with the following list of alleged linguistic properties of threats which she then checks against a corpus of utterances meta-linguistically classified as threats (Muschalik 2018: 52–53):

- (a) conditional language
- (b) futurity
- (c) “violent” verbs
- (d) expressions of speaker agency
- (e) use of personal pronouns (1st and 2nd person)
- (f) swear words
- (g) the mention of weapons

Muschalik's findings on the frequency and combinatory patterns of these properties suggest a higher degree of predictability and systematicity regarding the form side of threats – e.g., syntactic, morpho-syntactic and lexical features – than assumed in earlier studies (Muschalik 2018: 181–182). Thus, empirical work has contributed to a more differentiated view of the form-function relationship with respect to threatening in English. At the same time, it is uncontroversial that threatening can also be achieved in the absence of particular grammatical means and that apparent threat utterances can be interpreted differently, depending on context.

Threatening in German is so far largely under-researched. One of the few papers dealing with the linguistics of threatening in German is Falkenberg's (1992) very concise study on the notion of threatening. He develops the following list of characteristic formal and functional properties of threats in German:

- (i) There is no preferred sentence type for threats.
- (ii) There are only few lexical indicators of threats (with *wehe* as a candidate).
- (iii) Threats do not allow for illocutionary explicitness (*drohen* ‘to threaten’ cannot be used in an explicit performative construction).

- (iv) Threats can be performed quasi-explicitly by other illocutionary verbs such as *versprechen* ('promise') or *empfehlen* ('recommend').
- (v) The basic propositional structure of a threat is conditional: In case the addressee B does X (antecedent), the speaker A will do Y (consequent).
- (vi) By means of a threat, the speaker (conditionally) commits themselves to a future action.
- (vii) Threats can be propositionally implicit in that the consequent may be missing.
- (viii) The goal of a threat is to make the addressee refrain from an action X (deterrent threat) or to make the addressee perform an action X (extortionate threat).
- (ix) There is something like a "threatening intonation".
- (x) Threats can be performed non-verbally, without much contextual effort.
- (xi) Threats consist in instilling fear or 'negatively' loaded affective attitudes.
(Falkenberg 1992: 1992: 179–187; my translation)

While a close examination of properties (iii), (iv), (x) and (xi) is beyond the scope of this chapter, the properties (i) and (ii) will come under scrutiny in the corpus study (see Section 4). As to property (ix), I assume that phonological features of threat constructions are part of their inherent properties, without being able to substantiate this claim based on my (written) data. Falkenberg's property (vii) should, more properly, be regarded as an addendum to (v), clarifying that the conditional basic structure does not have to be realized overtly. As becomes clear, Falkenberg's properties (v), (vi), and (viii) are reminiscent of Muschalik's (2018) properties (a), (b), (d), and, more indirectly, (e), rendering these properties particularly interesting. Therefore, the present study will take these properties as a starting point, deriving from them a number of basic linguistic criteria to be used as a heuristics to identify, as a first methodological step, potential candidates of threatening *wehe*-utterances in the corpus. Before we get there, a short discussion of these properties is in order.

Let us start with (v). According to Falkenberg (1992), threats are necessarily propositionally conditional, as they express a conditional intention. The "basic propositional structure" consists in that the speaker makes the action denoted by the consequent conditional to the addressee's (non-) compliance with the action denoted by the antecedent. Propositional conditionality can be realized at the syntactic surface in various ways, including, for example, *wenn*- ('if') clauses, disjunctives (*entweder-oder* 'either-or'; *sonst* 'or else'), and pseudo-imperatives (e.g., *Sag noch ein Wort und ich schreie* 'Say one more word and I'll scream'; cf. Dacyngier 1998: 188–192). Conditionality can also be purely implicit, as in *Hat sonst noch jemand Lust auf eine Abreibung?* ('Anybody else who likes to be beaten up?', Falkenberg 1992: 179). In the next sections, we will see that also *wehe*-threats display a conditional basic structure, typically realized syntactically by a conditional clause

embedded by *wehe*, which is interpreted as the antecedent in the conditional structure.

Property (vii) relates to the potential implicitness of conditionality in threats. To perform a threat, it is, in appropriate contexts, sufficient to mention the antecedent and leave the consequent open to be inferred. Such cases are, on Falkenberg's account, still propositionally conditional. In Muschalik's (2018: 14) study, this is modeled by the concept of gradable pragmatic explicitness. While some authors, such as Limberg (2009: 1379), also allow for semantically non-conditional threats (e.g., "I will do X"), Falkenberg (1992) convincingly argues that utterances such as *Ich werde dir meine Leute auf den Hals schicken!* ('I will send my people after you'), said by a convict towards the judge at the end of a court hearing, or, similarly, utterances like *Das wird dir noch einmal leid tun!* ('You will regret this someday'), are not threats, as they are uttered after the action from which the speaker wanted to prevent the addressee. They thus do not fulfil the central illocutionary point of a threat, as defined by Falkenberg in (viii). Falkenberg suggests treating such utterances, instead, as predictions or as affirmations of the consequent of a previously performed threat.

Property (vi) comprises a futurity aspect. The speaker's commitment to a future action may be modeled, in Searlean terms, as the sincerity condition of threats, and is therefore closely related to the aspect of intentionality, i.e., the expression of the speaker's intention to act in a certain way (Muschalik 2018: 53, cf. also Fraser 1998: 168). Because of the semantic futurity/intentionality aspect of threats, one may expect that the consequent will not contain past tense verbs (*?If you don't do your homework, I punished you*). Furthermore, threats are about a potential (non-factive) action of the addressee which is taken to be the condition for a subsequent speaker action. The potential addressee action is denoted by the antecedent. Therefore, one may expect the antecedent of a threat not to include verbs in the past tense. Apparent counterexamples such as *Wenn du mich angelogen hast, dann gnade dir Gott!* ('If you lied to me, then may God help you') do not denote a factive event in the past (that the addressee has lied to the speaker), but are interpreted as containing an epistemic aspect, resulting in a potential (non-factive) reading: 'If it turns out that you lied to me, then God may help you'. Such examples would not count as threats under Falkenberg's definition, as they violate property (viii) according to which the goal of a threat is to make the addressee refrain from an action.

Property (viii) is about speaker agency and addressee-orientation. In a threat utterance, the speaker (agent) imposes some (non-)action upon the addressee (target). From this, one can expect that threateners will regularly refer to themselves and their addressees in uttering threats. We may thus expect that person-deictic pronouns will be used consistently in threat utterances. It goes without saying that the use of pronouns may vary along with the degree of explicitness of the threat. For example, in

threats in which there is no explicit consequent also the agent of the inferred future action (the 1st person) will remain implicit. While Muschalik (2018) also takes into consideration utterances that do not target the addressee, but a third person, it is questionable whether such utterances are threats. On Falkenberg’s definition, an utterance like *If Reagan came to Sheridan, I would shoot him* (Muschalik 2018: 103) does not count as a threat towards Reagan. This is because the speaker cannot coerce the referent of the 3rd person subject, Reagan, into (not) doing something, by uttering this sentence towards an addressee different from Reagan, say, Peter. Neither can the speaker, in uttering this sentence towards Peter, instill fear into Reagan.

By way of interim conclusion, from the general illocutionary function of threats as outlined above, whose illocutionary point is to make the addressee refrain from an action or to perform an action, in that the speaker conditionally commits themselves to a future action, we can derive a number of linguistic features that may be indicative of threat utterances in German: the basic conditional structuring of the propositional content of the utterance (including implicit conditionality), choices of verbal tense that are consistent with a future action in the consequent and a potential action in the antecedent, as well as the use of 1st and 2nd person pronouns in order to refer to agent and target of the threat. In the present corpus study, these characteristics will be used to search for relevant occurrences of threats, i.e., as heuristic criteria, not as definitional ones. As they can be taken to be characteristic – but neither necessary nor sufficient – for threats, they may be used to pre-select hits from the corpus, which then must be checked manually in context to determine whether they indeed are used as threats. As the study focuses on potentially threatening *wehe*-utterances, the general threat characteristics developed here will be complemented by characteristics derived from the grammatical properties of *wehe*, to be used, in the same methodological way, as pre-selectors. These will be developed in the next section.

3 *Wehe* (‘woe’): Grammatical properties

Etymologically, Modern German *wehe* and its variant *weh* have developed from Old High German (OHG) *wah* and Middle High German (MHG) *wē* (Kluge 2011: 879), with correlates in other Germanic languages such as Old English *wā*, English *woe*, Old Norse-Icelandic *vei*, Middle Dutch *wee* (*Wörterbuch der deutschen Gegenwartssprache*, henceforth WDG). Barðdal et al. (2013) reconstruct in detail the syntactic properties of Proto-Indo-European **wai* (‘woe’) to which also *wehe* – as well as its cognates from Indo-Iranian, Italic, Baltic, Slavic and Germanic – can be traced back.

There seems to be consensus across sources that already in OHG and MHG, *weh(e)* was used as an interjection (Kluge 2011: 879), instantiating (i) an exclamation of dismay (a lamentation) or (ii) a prediction of something hideous or malign. The readings (i) and (ii) are provided by both the *Deutsches Universalwörterbuch* (henceforth DUW) and the WDG. By contrast, Barðdal et al. (2013: 327) assign to **wai* only reading (i): “[It] was generally used when something bad happened to people. It is exploited to convey anguish and consternation, basically functioning as ‘*Ausdruck des Jammers*’ [‘expression of lament’].” Furthermore, they mention, in passing, the functions of curse and insult (Barðdal et al. 2013: 327). While there is no mention of either a predicting or a threatening function of **wai* in Barðdal et al. (2013), both the DUW and the WDG point to the threatening potential of the prediction reading (ii) of *wehe*. As argued in Section 2, threats always include the announcement of a (conditional) future action of the speaker. Thus, the alleged threatening use of *wehe* can be regarded as a subtype of the predicting use.

Apart from the interjection, the dictionaries also list entries of *weh* as a neuter noun (*das Weh*, *das Wehe* ‘pain’, ‘distress’) and as adjective (sometimes also classified as adverb) *weh* (‘painful’, ‘wistful’) (cf. WDG, DUW). It is not quite clear how interjectional, nominal and adjectival *weh(e)* are interrelated. While the WDG suggests that the noun developed in the 9th century based on the OHG interjection, which is said to occur since 800, Kluge (2011: 879) lists both the interjection and the noun as occurring alongside each other since the 7th century. The typical adjectival context for *weh(e)* is the predicative usage with a copular verb as in *mir ist weh ums Herz* (‘me.DAT is woe round the heart’; DUW), which is structurally analogous to something like *mir ist kalt* (‘me.DAT is cold’). This predicative use of *weh* sounds rather archaic in modern German but can still be found in poetic language. The attributive use of *weh* (e.g., *seine wehen Füße* ‘his aching feet’), on the other hand, developed only later, according to WDG.

As to the interjection *weh(e)*, it is important to note that according to the dictionaries, readings (i) and (ii) can be distinguished from early on. That is, already in OHG, there is a lamenting and a predicting usage. The two usages can be illustrated by the following examples taken from WDG.

- (7) a. *O weh!*
 oh woe
 ‘Alack!’
 b. *Weh mir!*
 woe me.DAT
 ‘Woe betide me!’

- (8) a. *Weh(e) dem, der das tut*
 woe DEM.DAT who.NOM this does
 'Woe betide the person who does this'
- b. *Wehe, wenn du nicht pünktlich nach Hause kommst*
 woe if you not on-time toward home come
 'Woe betide you if you don't get home on time'

Lamenting (7) and predicting (8) *wehe* can also be distinguished on formal grounds. Firstly, as the WDG notes, the lamenting use preferably features the form *weh*, while the predicting use preferably features *wehe*. However, in combination with a dative complement, the form *weh* can also be used in predictions (8a). Secondly, it seems that lamenting *weh* is often combined with other (lamenting) interjections such as *o* or *ach* (both corresponding, roughly, to English *oh*), cf. (7a). By contrast, predicting *wehe* cannot be combined with these interjections, cf. (8b').

- (8) b.' ?*O/?Ach wehe, wenn du nicht pünktlich*
 oh woe if you not on-time
nach Hause kommst
 toward home come

Thirdly, while lamenting *weh* (7a) can be replaced by other interjections such as *oje* ('oh dear') or *au wei(a)*¹, this replacement is hardly acceptable with predicting *wehe*, cf. (8b'').

- (8) b.'' ?*Oje/?Au weia wenn du nicht pünktlich*
 oh dear if you not on-time
nach Hause kommst
 toward home come

These observations clearly support the view that there are two different interjections *weh(e)* with two different meanings.

However, there is also a certain syntactic overlap between the two usages. For example, as becomes clear from (7b) and (8a), both usages of *weh(e)* can function as the head of an interjection phrase in which *wehe* takes a dative complement denoting an experiencer or maleficiary. This dative complement may be internally complex, containing a relative clause, as in (8a). More generally, the capacity of interjections to take NP complements (which may denote varying semantic roles)

¹ *Au wei(a)* can be regarded as a more colloquial variant of (*oh*) *weh* in German.

is well-known from interjections such as, e.g., *o* ('oh') and *pfui* ('fie'), as shown in a series of works on German interjections by Fries (1996, 2002). Some examples from older stages of German are (9) and (10), (9) dating back to the 17th century (Early New High German), (10) to the 18th century (New High German).

- (9) *o dem verhängnis*
 oh the.DAT doom.DAT
 'Oh this doom'
 (Opitz; cf. Fries 1996: 321)

- (10) *pfui des bösewichts*
 fie the.GEN evildoer.GEN
 'Ugh this evildoer'
 (Herder; cf. Fries 1996: 321)

In modern German, the NP complement of interjections such as *o* and *pfui* typically takes the nominative case (*O diese Politiker* 'Oh those.NOM politicians.NOM'; *Pfui das Schwein* 'Ugh this.NOM pig.NOM'). *Weh(e)* can be regarded as a relict in this respect, having kept dative case with its NP complement.

On the other hand, the dictionary entries suggest that lamenting *weh*, as opposed to predicting *weh(e)*, cannot embed a clausal complement, cf. the contrast between (7a-b) and (8b). The different syntactic behavior of the two *weh(e)* interjections can be taken to reflect the differences in their speech act potential. A lamentation, in the sense of an exclamation of dismay, is an expressive speech act by which the speaker expresses a feeling of pain, distress or regret about some event, which is presupposed as factual.² This explains the incompatibility with a conditional *wenn*-clause. On the other hand, a prediction is a future-oriented speech act which may be conditional on some event, which explains the compatibility with a *wenn*-clause. At the same time, both speech acts feature an experiencer, someone that is affected by a past or future event. This explains the compatibility of both with a dative complement as the target.

As to the form of the clausal complement of predicting *wehe*, the entries in the WDG and DUW only list verb-final *wenn*-clauses (e.g., 8b). By contrast, Falkenberg's (1992: 178) example, see (2), features an embedded verb-second (V2) clause. The observation that *wehe* can embed a V2 clause is not trivial. Only particular classes

² It is in this sense that the WDG and the DUW describe lamenting *weh(e)* (German *Klage*); cf. also Barðdal et al. (2013) on **wai*. In a wider sense, the term "lamentation" may also refer to things like whining and moaning (German *Jammern*), which do not necessarily presuppose a factual event, but may also refer to prospective events or just describe the expression of a speaker's psychological state with no relation to any actual situation. This wider reading is not the one relevant here.

of verbs in German, such as *verba dicendi* and preference predicates, are taken to be licensors of embedded V2 clauses, cf. (11)–(12).

- (11) *Peter sagt, du kommst alleine*
 Peter says you come alone
 'Peter says that you come alone'
 (Sode 2023: 3)

- (12) *Es ist besser, du kommst alleine*
 it is better you come alone
 'It is better if you come alone'
 (Sode 2023: 2)

As Sode (2023) shows, V2 clauses such as (11) are declarative and can be replaced by *dass*- ('that') clauses without change of meaning. By contrast, V2 clauses such as (12) can "always be replaced by *wenn*- ('if') clauses without obvious change of meaning", but can only be replaced by *dass*- ('that') clauses "under certain conditions" (Sode 2023: 6). This indicates that V2 clauses of the type (12) are semantically not declarative, but conditional. It seems that the interjection *wehe* fits well into the class of licensors of conditional V2 clauses, cf. (13).

- (13) *Wehe, du kommst alleine*
 woe you come alone
 'Woe betide you if you come alone'

Firstly, a V2 clause embedded by *wehe* can always be replaced by a *wenn*- ('if') clause, indicating its status as a conditional clause. Secondly, similarly to the licensor in (12), *wehe* involves an evaluative component, with the difference that it does not mark the event in the embedded clause as something positive (preferred), as in (12), but as something dispreferred.

4 Corpus study

4.1 Corpus, search query, and annotation categories

For the study, the DeReKo (German Reference Corpus) was used, which can be searched via the COSMAS search platform provided by *Leibniz-Institut für deutsche Sprache*. DeReKo is the largest available collection of written texts of contempo-

rary German, comprising 57.6 billion words from a variety of text types, mainly newspapers, but also fiction, academic prose as well as conceptually or medially spoken text types (cf. Koch and Oesterreicher 1985) such as Wikipedia discussions, fictional dialogues, or plenary debate protocols. The strength of DeReKo is its sheer size, which is especially important here because *wehe*, overall, is an infrequent phenomenon. A weakness of DeReKo is its unbalanced composition with regard to the shares of different text types, as it contains predominantly written genres such as journalistic texts.

The corpus was searched for sentences with either sentence-initial *wehe* or one of the sentence-initial clusters *doch wehe* ('yet woe'), *aber wehe* ('but woe'), *und wehe* ('and woe'), *oder wehe* ('or woe'), *denn wehe* ('because woe'), *sonst wehe* ('else woe'), *nur wehe* ('only woe'), while excluding sentences in which *wehe* co-occurred with *wohl/Wohl* ('well'), *Wind* ('wind'), or *losgelassen* ('let-go'), for reasons I will go into below. The search string used is specified in (14).

- (14) ((wehe:sa %0s,Max (wohl or (wind or losgelassen))) or (((doch:sa or (aber:sa or (und:sa or (oder:sa or (denn:sa or (sonst:sa or nur:sa)))))) /+1w,Max wehe %0s,Max (wohl or (wind or losgelassen))))

Restricting *wehe* to the sentence-initial position reduced the number of non-target hits to a minimum, as it excluded, for example, sentences containing the noun *Wehe* ('contraction') in descriptions of giving birth. While the interjection *wehe* typically occurs sentence-initially, it may be preceded by a connector, as listed in the clusters above. The exclusion of co-occurrences with *wohl/Wohl* ('well'), *Wind* ('wind'), and *losgelassen* ('let-go') further reduced the number of non-target hits. In combination with *Wind*, one is most likely dealing with a form of the verb *wehen* ('to blow'). In combination with *Wohl*, one is often dealing with the high-frequent formulae *Wohl und Wehe* ('weal and woe'). Both were irrelevant to the present study. In combination with *losgelassen* ('released'), the attestations correspond to the quotation *Wehe, wenn sie losgelassen* ('woe betide if they are released') from Schiller's poem *Die Glocke*, which occurs frequently in the corpus and whose inclusion would have skewed the search results.

The search resulted in 12.868 hits, of which 1.000 were randomly extracted. Of these, the first 500 were selected to form the data set of the study. The data set was represented by COSMAS as a list of key words in context (KWIC), the context comprising (mostly) one sentence to the left and one sentence to the right. The data set was manually annotated according to the categories listed in Table 1.

Table 1: Annotation categories.

Category	Parameter	Specification	Example
Type of complement	Ø	–	– <i>Wehe!</i> 'Woe'; <i>Wehe, wehe!</i> 'Woe, woe!'
	Phrasal complement (dative)	–expansion	– <i>Wehe dem!</i> 'Woe [betide] DEM.DAT.SG.M' – <i>Wehe den Besiegten!</i> 'Woe [betide] the.DAT.PL defeated.DAT.PL'
		+expansion by RC ³	– <i>Wehe dem, der erwischt wird</i> 'Woe [betide] DEM.DAT.SG.M who gets caught' – <i>Wehe dem Minister, der kein Geld herausrückt</i> 'Woe [betide] the.DAT secretary.DAT who does not fork out money'
		+expansion by CC ⁴	– <i>Wehe den Protokollbeamten, wenn der Füller kleckst</i> 'Woe [betide] the.DAT.PL keepers.DAT.PL of the minutes if the fountain pen blots'
	Clausal complement (CC)	wenn-clause	– <i>Wehe, wenn die irgendwo hinpinkeln</i> 'Woe [betide] if they pee anywhere'
		V2-clause	– <i>Wehe, die Personalzahlen sinken weiter</i> 'Woe [betide] if the personnel numbers continue to decrease'
Person marking	Dative argument	1 st	– <i>Wehe mir/uns</i> 'Woe [betide] me.DAT/us.DAT'
		2 nd	– <i>Wehe dir/euch/Ihnen</i> 'Woe [betide] you.DAT.SG/you.DAT.PL/you.DAT.POL'
		3 rd	– <i>Wehe ihm/ihr/ihnen/dem/der/denen/NP</i> 'Woe [betide] him.DAT.SG.M/her.DAT.SG.F/them.DAT.PL/DEM.DAT.SG.M/DEM.DAT.SG.F/DEM.DAT.PL/NP
	Subject of CC	1 st	– <i>Wehe, [wenn] ich/wir . . .</i> 'Woe [betide] [if] I.NOM/we.NOM'
		2 nd	– <i>Wehe, [wenn] du/ihr/Sie . . .</i> 'Woe [betide] [if] you.NOM.SG/you.NOM.PL/you.NOM.POL'
		3 rd	– <i>Wehe, [wenn] er/sie/es/sie/der/die/das/die/NP . . .</i> 'Woe [betide] [if] he.NOM.M/she.NOM.F/it.NOM.N/they.NOM/DEM.NOM.M/DEM.NOM.F/DEM.NOM.N/DEM.NOM.PL/NP

³ RC=relative clause⁴ CC=conditional clause

Table 1 (continued)

Category	Parameter	Specification	Example
Tense marking	Clausal complement (CC, RC)	–past tense	see examples above
		+past tense	– <i>Wehe dem, der nicht die Wahrheit preisgab</i>
		(preterite, perfect, pluperfect)	‘Woe [betide] DEM.DAT.SG.M who did not reveal [lit. not revealed.PRET] the truth’
			– <i>Wehe aber, wenn einer sein Vertrauen missbraucht hat</i> ‘But woe [betide] if someone abused [lit. has abused.PF] his trust’ – <i>Wehe, wenn er [. . .] selber etwas gefunden hatte</i> ‘Woe [betide] if he himself had found [PLUPF] something’

For the analysis, the distribution of the 500 hits was explored with respect to the chosen annotation categories. Starting with the distribution as to types of complements, the exploration was refined in several steps, taking into account the distribution of person marking and tense marking. This procedure allowed zooming into those hits that displayed all the linguistic properties that were established above as heuristic characteristics of threats. Relevant attestations were then considered in their co-texts to assess their intended meanings. Additionally, selected clusters of attestations lacking one or more relevant linguistic characteristics of threats were cross-checked in their co-texts as to their intended meanings.

4.2 Results

In what follows, the quantitative results are presented for each of the selected categories.

4.2.1 Type of complement

Table 2 shows the distribution of attestations with respect to the type of complement embedded by *wehe*.

Table 2: Type of complement.

Parameter	Specification	Number	Total number	Percentage
Ø	–	10	10	2.0
Phrasal complement (dative)	–expansion	12	143	28.6
	+expansion by RC	124		
	+expansion by CC	7		
Clausal complement (CC)	<i>wenn</i> -clause	140 ⁵	335	67.0
	V2-clause	195		
Other		12	12	2.4
Total			500	100

In roughly two thirds (67%) of the cases, *wehe* is used with a conditional clause complement. Within this class, V2 clauses occur more frequently (58.2%) than *wenn*-clauses (41.8%). Phrasal (dative) complements are used in less than one third (28.6%) of cases. Of these, the vast majority (86.7%) features dative nouns or dative demonstrative pronouns expanded by a relative clause. Bare uses of the interjection, without complement, occur only very rarely (2%). The class of “other cases”, which is also very small (2.4%), subsumes some individual cases that did not fit into any of the other classes.⁶

4.2.2 Person marking

Table 3 zooms into the three largest classes in the corpus, which together make 478 of 500 hits: (i) *wehe* + phrasal (dative) complement, (ii) *wehe* + *wenn*-clause, and (iii) *wehe* + V2 clause. The table shows the distribution of person marking (1st, 2nd, 3rd) on the dative argument (i) or the subject of the conditional clause (ii), (iii), respectively.

⁵ Of the 140 examples of *wehe* + *wenn*-clause, 6 are elliptical (*Wehe, wenn nicht* ‘Woe if not’).

⁶ For example, there is one attestation of a temporal *als*- (‘when’) clause, cf. *Wehe aber, als die Außenwelt eindringt* ‘Woe betide when the outer world is intruding’ as well as one attestation with *über*-PP instead of a dative, cf. *Wehe dann über dich Elenden* ‘Woe betide upon you wretched’. As to examples of the type *Wehe dem, einer hatte gemogelt*, which were also assigned to this class, see (20) below.

Table 3: Person marking.

Category	Person marking	Number	Total number	Percentage
phrasal complement (dative)	1 st	2	143	1.4
	2 nd	3		2.1
	3 rd	138		96.5
				100
wenn-clause subject	1 st	12	140	8.6
	2 nd	2		1.4
	3 rd	120		85.7
	elliptic ⁷	6		4.3
				100
V2-clause subject	1 st	12	195	6.2
	2 nd	20		10.2
	3 rd	163		83.6
				100
Total			478	

In all three classes, *wehe*-utterances predominantly feature 3rd person marking. Of the cases with phrasal (dative) complement, 96.5% display a 3rd person dative. Of the cases with conditional clause complement, 85.7% of *wenn*-clauses and 83.6% of V2 clauses display 3rd person subjects. For illustration, cf. (15)-(19).

- (15) *Wehe dem, der nicht über die nötigen*
woe DEM.SG.DAT who.NOM not over the necessary
Beziehungen verfügt, um an die begehrten Jobs
relations disposes to at the sought-after jobs
zu kommen
to come
‘Woe betide those who do not dispose of the necessary contacts to get the hot jobs’
(Rhein-Zeitung, 06/24/2000)
- (16) *Wehe, der Gegner verliert den Ball und ist*
woe the opponent loses the ball and is
eine Sekunde lang unsortiert
one second long unsorted
‘Woe betide if the opponent loses the ball and is confused for one second’
(die tageszeitung, 01/31/2011)

7 Elliptical cases omit subject and finite verb, e.g., *Wehe, wenn nicht*.

- (17) *Wehe, wenn Per den Ball trifft: Der überzeugende*
 woe if Per the ball hits the convincing
Abwehrspieler trifft zum 2:1
 defender scores to-the 2:1
 'Woe betide if Per hits the ball: The impressive defender scores to make it 2:1'
 (Süddeutsche Zeitung, 06/17/2005)
- (18) *Wehe, wenn die Kurse fallen*
 woe if the stock-prices fall
 'Woe betide if the stock prices fall'
 (Rhein-Zeitung, 02/24/2000)
- (19) *Wehe, die Heizung fällt länger als eine Stunde aus*
 woe the heating falls longer than one hour out
 'Woe betide if the heating breaks down for more than one hour'
 (Frankfurter Rundschau, 01/11/1997)

By uttering (15)-(19), the speaker signals their foreshadowing of some future calamity resulting from the property or event denoted by the complement of *wehe*. The utterances thus convey the reading of prediction of calamity. In usages with a phrasal dative complement, cf. (15), the foreshadowed calamity targets the (human) referent denoted by the dative pronoun or noun. In usages with a conditional clause complement, the target of the calamity must be inferred from the utterance in context. While in (16), the target can be inferred to be identical with the human subject of the conditional clause (it is the opponents who suffer from their losing the ball), in (17), the target is not the subject of the clause. It is not the football player Per but the opponent team who suffers from Per's scoring. In (18)-(19), with no human subject contained in the clause, the target must be inferred as being some human individual who suffers from the fall of the stock prices or the breakdown of the heating, respectively.

A peculiar usage, which was assigned to the class of "other" uses as indicated in Table 2 (n=12), is illustrated by (20)-(21).

- (20) *Wehe dem, ein Stück bleibt aus Versehen im Karton*
 woe DEM.SG.DAT a piece remains out mistake in-the box
 'Woe betide if a piece by mistake remains in the box'
 (Nordkurier, 12/24/2011)
- (21) *Wehe dem, wenn diese Umleitung durch einen*
 woe DEM.SG.DAT if this redirection through a

Verkehrsunfall blockiert wird
traffic-accident blocked becomes
'But woe betide if this redirection gets blocked by a traffic accident'
(Rhein-Zeitung, 03/29/2008)

What is peculiar about these cases is that the dative demonstrative pronoun *dem* is referentially empty. In this respect, (20)-(21) differ from uses such as (15), in which *dem* refers to the experiencer. As to their interpretation, (20)-(21) convey the same meaning as they would without *dem*. While one might assume that these are simply “erroneous” uses, the fact that there are four instances of this pattern among the “other” uses might be taken as some initial evidence for a process of reanalysis being under way, in which *dem* loses its referential meaning, being amalgamated with *wehe* to form some kind of complex interjection *wehedem* with the same meaning as *wehe*.

4.2.3 Tense marking

Generalizing from (15)-(19) that *wehe*-utterances with 3rd person referents (i.e., 3rd person embedded clause subjects and most 3rd person datives) typically convey a prediction reading with no threatening involved, one may conclude from Table 3 that the vast majority of attestations of *wehe* + complement in the corpus are predictions. Interestingly, this is also true for 3rd person uses in which the embedded (conditional or relative) clause displays a verb in a past tense. Table 4 provides an overview of the distribution of past versus non-past tense forms in the embedded clauses (n=466) in the corpus. 14% of the conditional clauses and 8.9% of the relative clauses display past tense marking.

Table 4: Tense marking.

Category	Tense marking	Number	Total number	Percentage
Conditional clause	–past tense	294	342	86,0
	+past tense	48		14,0
				100
Relative clause	–past tense	113	124	91,1
	+past tense	11		8,9
				100
Total			466	

While past tense marking seems, at first glance, to be inconsistent with a prediction reading, these uses can be explained as narrative (reported) predictions, cf. (22).

- (22) *Der Schwiegersohn soll [...] das Atelier [...] nur unter
the son-in-law shall the artist-studio only under
strengsten Auflagen benutzt haben dürfen. Der Boden
strictest conditions used have may.PTCP the floor
musste abgedeckt sein. Wehe, irgendwo blieb ein
must.PST covered be woe anywhere remained a
Farbfleck zurück.
color-stain back*

'It is said that the son-in-law was allowed to use the artist studio only under the strictest conditions. Woe betide if some color stain was left behind.'

(Süddeutsche Zeitung, 03/23/2002)

Utterances like (22) occur in contexts in which a narrator foreshadows negative consequences of an event, of which the narrator has knowledge at the time of utterance, but of which the person narrated about does not have knowledge at the time of the narrated world. These specific narrative conditions allow for prediction readings in past tense.

While most 3rd person uses in the data set convey prediction readings with no threatening involved, it is clear that the mere appearance of a 3rd person subject does not prevent an utterance from expressing a threat. For example, one can find utterances such as (23)-(24), which under certain conditions may be perceived as threats.

- (23) *Wehe, wenn die irgendwo hinpinkeln, dann behalte
woe if DEM.PL somewhere pee then keep
ich die nicht
I them not*

'Woe betide if they pee somewhere, then I won't keep them'

(Rhein-Zeitung, 07/18/2003)

- (24) *Wehe, die Mitgift stimmt nicht
woe the dowry is-right not
'Woe betide if the dowry is not adequate'
(Rhein-Zeitung, 07/23/1997)*

In (23), the demonstrative 3rd person plural pronoun refers to two abandoned kittens the speaker was persuaded to accommodate. The utterance can be inter-

puted as a threat towards the speaker's interlocutors who persuaded the speaker to accommodate the kittens, i.e., who are responsible for them. While for (24), there is no further context provided by DeReKo, if uttered towards the bride's father by the groom, one may interpret the utterance as a threat towards the father, as the one responsible for the size of the dowry, despite its subject being a 3rd person NP.

4.2.4 Uses with person deixis

Let us now zoom into person-deictic uses, i.e., *wehe*-utterances with 2nd person or 1st person embedded clause subjects. Our heuristic criteria suggest that these should be good candidates for threats. Overall, 2nd person uses are quite rare in the corpus. However, Table 3 shows that the class of *wehe* + V2 clausal complement contains a comparatively large share of 2nd person (*du* 'you.SG', *ihr* 'you.PL', *Sie* 'you.SG/PL.POL') subjects (10.2%), as compared to only 2.1% for *wehe* + phrasal complement and 1.4% for *wehe* + *wenn*-clause. Examining these cases within their co-texts in more detail, it turns out that the majority of uses of *wehe* + V2 clausal complement with 2nd person subject (17 of 20) indeed convey a threat reading. These are listed in (25)-(41).

- (25) *Muttern passte gut auf: „Wehe, ihr esst*
 mother watched well up woe you.PL eat
sie nicht, dann gibt's eben gar nichts.
 them not then is-there PRT PRT nothing
 'Mother was very attentive: You better eat them, or you will get nothing instead'
 (Berger, Rudi W.: *Spitzenrausch*. Föritz 2006)

- (26) *„Wehe, du bist Punkt elf nicht zu Hause“⁸*
 woe you.SG are point eleven not to home
 'You better be at home at eleven o'clock sharp'
 (Nordkurier, 05/08/2004)

- (27) *Sie [. . .] drohte nur: „Wehe, du verlässt dieses Haus.“*
 she threatened only woe you.SG leave this house
Kaum war sie gegangen, stürzte ich hinaus.
 barely was she gone rushed I out
 'She just threatened: "Don't you dare leave this house." No sooner had she left than I rushed out'
 (NZZ am Sonntag, 10/26/2014)

⁸ DeReKo provides no further context for this example.

- (28) *Wehe, du erhöhst schon wieder die Umlage, mein Lieber!*⁹
 woe you.sg raise already again the levy my
 dear
 'Don't you dare raise the levy once again, my friend'
 (*Rhein-Zeitung*, 02/09/2013)
- (29) *Petra H. hatte auch gesehen, wie die beiden ein Klebeband mitnahmen, um den Automann zu fesseln. Und schließlich hatten die Räuber gedroht: „Und wehe, du sagst was!“*
 Petra H. had also seen how the both
 a tape along-took for the car-man to tie
 and finally had the robbers threatened and woe
 you.sg say what
 'Petra H. had also observed how the two took a sticky tape to tie up the car dealer. And finally, the robbers threatened: "Don't you dare say anything"'
 (*die tageszeitung*, 07/06/2001)
- (30) *Der NPD-Politiker behauptet in seiner Anzeige, die Männer hätten ihn wüst bedroht. „Wehe, du gibst einem Schüler die CD – dann schlag ich dir den Schädel ein!“*
 the NPD-politician claims in his complaint the
 men had him wildly threatened woe You.sg
 give any pupil the CD then hit
 I you.DAT the skull in
 'The NPD politician claims in his complaint that the men ranted and raved at him. "Don't you dare give the CD to a pupil – or I'll bash your head in!"'
 (*die tageszeitung*, 09/27/2007)
- (31) *„Wehe, du erzählst es mir nochmal. Was meinst du, was ich mir den halben Tag habe anhören müssen?“ „Aber warum hast du heute nicht kommen?“*
 woe you.sg tell it me again what
 mean you.sg what I me the half
 day have listen must.PTCP but why have

9 DeReKo provides no further context for this example.

du das denn nicht verhindert?“, rief Uwe.
 you.SG that PRT not impeded exclaimed Uwe
 “Don’t you dare tell me again. What do you think I have been listening to half
 of the day?” “But why didn’t you impede this?”, Uwe exclaimed.’
 (Braunschweiger Zeitung, 06/19/2006)

- (32) „Wehe, du heulst“, sagte ich. Arme, alte
 woe you.SG cry said I poor old
 Frau, sagte mein Gewissen [. . .]. Ich hörte nicht
 woman said my conscience I heard not
 auf mein Gewissen.
 on my conscience
 “Don’t you dare cry”, I said. Poor, old woman, my conscience said. I didn’t
 listen to my conscience.’
 (Zander, Wolfgang: *Hundeleben*. Meßkirch 2011)

- (33) Wehe, du vögelst meine Mom“, feuert Dyrus
 woe you.SG fuck my mom fires Dyrus
 John entgegen. „Besser, du gewöhnst dich dran“,
 John toward better you get-used you.ACC to-it
 zischt der zurück.
 spits this-one back
 “Don’t you dare fuck my mom”, Dyrus hurls at John. “You better get used to
 it”, John spits as a reply.’
 (Süddeutsche Zeitung, 12/02/2010)

- (34) Es waren Polizisten auf Streife [. . .].
 it were policemen on patrol
 Snowmobiler haben sich an die
 snowmobilers have themselves on the
 Verkehrsregeln zu halten. „Habt ihr
 traffic-rules to keep have you.PL
 denn nicht das Snowmobiler-Handbuch bekommen?
 PRT not the snowmobile-handbook received
 Diesmal geht es ohne Ticket
 this-time goes it without ticket
 ab. Aber wehe, ihr lest
 off but woe you.PL read

das Büchlein nicht.“

the booklet not

‘They were policemen on the beat. Snowmobile drivers must obey the traffic regulations. “Didn’t you receive the snowmobile handbook? This time, you’ll get away without a ticket, but you better read the booklet.”’

(*Die Zeit*, 01/14/1983)

- (35) *Vor einigen Jahren [...] wurde mir von*
 before some years was me.DAT by
einer Therapeutin empfohlen, einen indianischen Traumfänger
 a therapist recommended an Indian dreamcatcher
ins Fenster zu hängen. Ich sollte
 in-the window to hang I should
aber unbedingt zwei meiner Haare hineinflechten.
 though absolutely two my.GEN hair in-braid
[...] Gott sei Dank wusste ich es
 god be thanks knew I it
aufgrund meines Studiums besser. Ich habe
 because my.GEN study better I have
sie später angerufen und ihr gesagt:
 her later phoned and her said
„Wehe, Sie wagen es, für diesen
 woe you.SG.POL dare it for this
Unsinn eine Rechnung zu stellen.“
 bullshit an invoice to issue
 ‘Some years ago, I got a therapist’s advice to put an Indian dreamcatcher in front of my window. It allegedly was very important to also braid into it two hairs of mine. Luckily, I knew better because of my academic education. I called her later and told her: “Don’t you dare invoice this bullshit.”’
 (*Zeit Wissen*, 08/02/2011; Letters to the editor)

- (36) *Manchmal wird auch ein hitziger Wortwechsel*
 sometimes becomes also a fierce quarrel
daraus. So wie [...] bei den Miami
 out-of-it so as by the Miami
Open zwischen der Spanierin Garbiñe Muguruza
 Open between the Spanish Garbiñe Muguruza
und ihrem französischen Coach Sam Sumyk.
 and her French coach Sam Sumyk

Muguruza maulte, und Sumyk konterte: „Wehe,
Muguruza grumbled and Sumyk countered woe
du sagst mir noch einmal, dass
you.SG say me.DAT again once that
ich, verdammt noch mal, den Mund
I damn again PRT the mouth
halten soll.“
keep shall

‘Sometimes it develops into an angry argument. Like during the Miami Open between the Spanish Garbiñe Muguruza and her French coach Sam Sumyk. Muguruza grumbled, and Sumyk countered: “You better don’t tell me again to shut-damn it-up.”

(*Berliner Zeitung*, 03/30/2017)

- (37) *Warum fragt ihr nicht mal den*
why ask you.PL not PRT the
Experten für historische Frankfurter Gebäude, gerne
expert for historical Frankfurt buildings gladly
auch auf dem nächsten Frankfurter Stammtisch?
also on the next Frankfurt stem-table
Und wehe, du sagst jetzt, das
and woe you.sg say now this
sei doch nicht belastbar –Haselburg-müller
be.SBJV PRT not sound –Haselburg-müller

‘Why don’t you ask the expert for historical Frankfurt buildings, preferably also at the next Frankfurt regular’s table? And you better not tell me now that this is not sound.’

(Wikipedia, 08/27/2011 – http://de.wikipedia.org/wiki/Diskussion:Hessischer_Hof)

- (38) *„Wehe, du kotzt mir auf die Theke!“*
woe you.SG puke me.DAT on the counter
ist die erste ganz und gar eigene
is the first wholly and PRT own
Cartoonveröffentlichung des Zeichners.
cartoon-publication the.GEN drawer.GEN

“‘Don’t you dare puke on my counter” is the first completely own cartoon publication of the drawer.’

(*Spiegel-Online*, 09/09/2009)

- (39) „Ich bin um fünf in der Abtei.
 I am about five in the abbey
Und wehe, du lässt mich warten.“
 and woe you.SG let me wait
 'I'll be at five in the abbey. And don't you dare let me wait.'
 (Regnier, Sandra: *Das Flüstern der Zeit*, Band 1. Hamburg 2015)
- (40) Sarah blieb derweil vor mir stehen und
 Sarah kept meanwhile before me stand and
deutete auf ihre Haare. Doch statt mich
 pointed on her hair but instead me.acc
nach meiner Meinung zu fragen, sagte sie
 after my opinion to ask said she
grimmig: „Wehe, du sagst auch nur ein
 fiercely woe you.SG say also only one
Wort zu meiner Frisur...“
 word to my haircut
 ‚Meanwhile, Sarah kept standing in front of me, pointing to her hair. But instead of asking me for my opinion, she fiercely said: “Don't you dare say one single word about my haircut...”'
 (Riemer, Martina: *Road to Hallelujah*. Hamburg 2015)
- (41) Sie sagen, Sie würden helfen, die
 you.SG.POL say you.SG.POL would help the
Probleme zu lösen. Anschließend drohen Sie
 problems to solve subsequently threaten you.SG.POL
aber: Wehe, ihr macht etwas im
 though woe you.PL make something in.the
sozialpolitischen Bereich, dann werden wir euch
 socio-political area then will we you.PL.ACC
auf die Finger klopfen.
 on the fingers knock
 'You claim that you want to assist with solving the problems. But then you threaten: You better don't do anything within the area of social politics, otherwise we will rap your knuckles.'
 (Plenary speech, Parlament Hamburgische Bürgerschaft, 01/23/2002)

In all these examples, the *wehe*-utterances are represented as part of a real or fictional conversation in direct speech. They appear within (youth) novels (25), (32), (39), (40), news reports (26)-(31), (33)-(34), (36), in computer-mediated communica-

tion (37), letters to the editor (35) or political speeches (41). A special case is (38), where the *wehe*-utterance is a quotation that is used as the title of a cartoon book. Within its original context, (38) is an utterance of a bar keeper who threatens a restaurant guest. The interactions are often characterized by familiarity and by hierarchical power relations between the interlocutors, cf. the threats uttered by parents towards children (26), policemen towards teenagers (34), a tennis coach towards their coachee (36), or threats between teenagers, e.g. (39). In all examples, the (original) speaker of the *wehe*-utterance intends to make the addressee refrain from an action, e.g. (28), (32), (33), or to perform an action, e.g., (25), (26), (34), conditionally committing themselves to a future action. This future speaker action remains implicit in most cases; except (30) and (41). The utterances can be regarded, in many cases, as attempts to instill fear in the addressee. While, e.g., (29) and (30) will be perceived as highly intimidating, (25) and (31) seem to be less offensive. As to the contextual embedding, one finds illocutionary verbs characterizing the utterances, meta-linguistically, as threats, cf. (27), (29), (30), (41). Also, the ironic vocative *mein Lieber!* ('my friend') (28) is indicative of a threat, on a par with *Freundchen* ('buster'), which is a typical threat indicator according to Falkenberg (1992: 179). Furthermore, there are 'violent' verbs in some of the contexts, cf. (30), as well as *verba dicendi* that denote an aggressive verbal exchange, cf. (33), (36), (40).

The remaining three attestations of *wehe* + V2 clause with 2nd person subject do not convey threats. In these cases, the 2nd person pronoun is used generically, not deictically, cf. (42)–(44). Such cases can be regarded as predictions of calamity towards a generic group of people, where the calamity is not caused by the speaker.

- (42) *Auf den Rolltreppen zum Beispiel. In*
 on the staircases to.the example in
München heißt es rechts stehen, links
 Munich means it right stand left
gehen. Und wehe, du stehst links.
 go and woe you.SG stand left
Da wirst du sofort zusammengeputzt.
 then become you.SG immediately blasted
 'For example, on moving staircases. In Munich, they say stand right, go left.
 And woe betide if you stand left. You will be blasted immediately.'
 (Nürnberger Nachrichten, 03/23/2013)

- (43) *Twitter und Facebook haben die sogenannte*
 Twitter and Facebook have the so-called
Timeline zum Maß des Medienkonsums gemacht [. . .].
 timeline to.the measure the.GEN media.use.GEN made

*Wehe, du schaust auf dein Smartphone,
 woe you.SG look on your smartphone
 und seit einer Minute hat niemand
 and since one minute has nobody
 dich auf deiner Timeline kontaktiert. Dann
 you.SG.ACC on your timeline contacted then
 weißt du, was Einsamkeit ist.
 know you.SG what loneliness is*

'Twitter and Facebook turned the so-called timeline into a measure of media use. Woe betide if you look at your smartphone and nobody has been contacting you for one minute. Then you know what loneliness is.'

(Weltwoche, 06/26/2014)

- (44) *Jeder hier hat eine Geschichte zu
 Everybody here has a story to
 erzählen über die Flieger, die den
 tell about the planes that the
 Tod abwerfen. Wenn du leben willst,
 death drop if you.SG live want
 sagt einer, musst du ein Loch
 says one must you.SG a hole
 graben [...]. Und bleib weg von den
 dig and stay away from the
 Bäumen. Wehe, du bist unter einem
 trees woe you.SG are under a
 Baum, wenn die Bomben kommen.
 tree when the bombs come*

'Everybody here has a story to tell about the planes that drop death. If you want to live, one of them says, you must dig a hole. And stay away from the trees. Woe betide if you are under a tree when the bombs come.'

(Süddeutsche Zeitung, 11/02/2001)

As Table 3 shows, there are also 12 instances of *wehe* + V2 clause with 1st person subject. An example is (45).

- (45) *Meine Söhne fragen regelmäßig: „Wann gibt's
 My sons ask regularly when gives.there
 mal wieder Heimat?“ Damit meinen sie*

PRT again homeland therewith mean they
Königsberger Klopse. Und wehe, ich verändere
 Königsberg meatballs and woe I change
was am Rezept. Das muss immer
 what on.the recipe that must always
gleich schmecken.
 same taste

'My sons keep asking me: "When will we have homeland again?" By this, they mean meatballs 'Königsberg'. Woe betide me if I change something in the recipe. The taste must always be the same.'

(Frankfurter Rundschau, 04/01/1997)

Examples such as (45) do not convey a threat, as the antecedent denotes an action from the speaker, not a (dispreferred) action from the addressee. While the consequent in (45) is implicit, it can be inferred that the speaker conditionally predicts some calamity to affect her- or himself. The relevant 1st person attestations can therefore be regarded as predictions of calamity affecting the speaker, or as self-warnings.

Similarly, a check of the *wehe* + *wenn*-clause attestations with 1st person subject (n=12, cf. Table 3) in context reveals that they do not receive a threat reading, at least not under the definition adopted in this paper, cf. (46).

- (46) „*Wehe, wenn ich einen erwische, der*
 woe if I one catch who
nur einen Millimeter abweicht, den werde
 only one millimeter deviates whom will
ich eigenhändig aus dem Verein schmeißen.“
 I own-handed out the association throw

“Woe betide if I catch a anyone who deviates only one millimeter [from this requirement], I will fire him from the association.”

(Frankfurter Rundschau, 03/13/1998)

For examples such as (46), it is disputable whether one may speak of threats. While on Muschalik's account, (46) would count as a threat, on Falkenberg's account, it clearly would not, because the utterance does not target the addressee, but a 3rd person, indefinite, non-specified individual.

Typically, attestations of *wehe* + conditional clause (V2 or *wenn*-clause) with 1st person subject receive a prediction reading, including narrative predictions such as (47).

- (47) *Da saß ich dann ganz stolz*
 there sat I then wholly proud
im Sattel auf dem linken Hinterpferd
 in.the saddle on the left back-horse
und musste achtgeben, dass die Vorderpferde
 and must.PST watch-out that the front-horses
auf Zuruf vorrückten. Aber wehe, wenn
 on call proceeded but woe if
ich nicht aufgepasst hatte. Dann wurde
 I not looked-out had.IND then was
gewaltig geflucht.
 heavily cursed

'There I sat very proudly in the saddle of the left back horse and had to look out for the front horses to move forward on call. But woe betide if I did not look out. Then people cursed heavily.'

(Braunschweiger Zeitung, 08/28/2007)

4.3 Discussion

This study aimed at exploring the question whether one can identify, based on a quantitative and qualitative examination of utterances containing the interjection *wehe*, a particular, conditional clause embedding pattern of *wehe*-usages that is, to a high degree of conventionality, associated with a threat reading.

The results indicate, first, that in most cases, the interjection *wehe* is used as a matrix element that embeds a conditional clause or a (relative-clause embedding) dative pronoun or noun. In most of these uses, the pattern does not convey a threat reading, but the more general reading of a prediction of calamity. Thus, it is clearly not the case that German *wehe* is, univocally, a threat marker, as Falkenberg suggested. An important indicator for the interpretation of a given *wehe* + embedded conditional clause utterance as a (mere) prediction or a threat is person marking. The analysis shows that uses with a 3rd person subject in the conditional clause typically convey the speech act function of a prediction. By contrast, if the embedded clauses display a 2nd person subject, the chances are very high that the utterances convey threats.

Furthermore, it becomes clear from the analysis that there are distributional differences with respect to the two subtypes of conditional clauses embedded by *wehe*. On the one hand, both variants predominantly feature 3rd person subjects, indicating a default usage as prediction. On the other hand, among the (few) uses

with 2nd person subjects, there is a clear contrast between the two syntactic variants. While more than 10% of the V2 clauses display 2nd person subjects, only 1.4% of the *wenn*-clauses do. This contrast can be taken as an indication of the existence of a highly specific syntactic construction *wehe* + V2 clause with 2nd person subject – albeit relatively rarely used in the corpus – that is to a high degree conventionalized for threats. Of the 20 uses of *wehe* + V2 clause with 2nd person subject, 17 convey a threat reading. In the sense of Terkourafi (2005), it can be argued, based on this observation, that in German speakers' linguistic experience, there is a particularly close association between utterances of *wehe* + V2 clause with 2nd person subject and threatening contexts.

At the same time, the analysis shows that the interpretation of *wehe* + V2 clause with 2nd person subject is still context-dependent to some degree. Cases in which the 2nd person pronoun is used generically are not interpreted as threats. Note that the observation that the interpretation of instances of a syntactic construction may vary depending on context is not *per se* a problem for a construction analysis, as it is a basic tenet in Construction Grammar that conventionalization is a matter of degree (Goldberg 2013).

Overall, the study reveals that threat usages, as well as lamentation usages, are rather infrequent in the corpus. By far the most frequent usage is the one as a prediction of calamity. Among the predicting uses there are also narrative cases, in which predictions are performed by means of utterances with a past tense verb. Interestingly, this usage is not mentioned in the dictionaries. The findings may be taken to suggest that since the OHG and MHG times, the usage of *wehe* has changed such that in Modern German, the lamenting use with or without dative experiencer, which can be traced back to Proto-Indo-European (Barðdal et al. 2013), decreased in favor of the usage in predictions and (to a lesser extent) threats. However, it must be kept in mind that the corpus used in the present study mainly contains written (newspaper) texts, something that may have skewed the distribution of the various usage types in the data set. It may well be that, e.g., lamenting usages are more frequent in spoken dialogues (or monologues).

Further research therefore should examine the use of *wehe* in spoken data. Furthermore, it would be desirable to complement the findings on *wehe* + V2 clause with 2nd person subject by studies that take a perspective from function to form, i.e. that investigate the distribution of a broader variety of formal means to express threats in German. This would allow for an assessment of the status of *wehe* + V2 clause with 2nd person subject within the functional domain of threatening in German in comparison with other linguistic means of threatening.

5 Concluding remarks

The present study adds new evidence to the growing research on the grammar of impoliteness by describing a specific impolite construction in German, and providing empirical evidence for its existence, which has been neglected in the research literature so far. A theoretical implication that can be drawn from the study is that impoliteness is a phenomenon at the interface between grammar and pragmatics. To comprehensively understand this phenomenon, one needs to take into account both grammatical and pragmatic factors and to systematically describe their interaction in the process of meaning constitution. Studying utterances containing the interjection *wehe* (‘woe’) in a large corpus of German, it was shown that while the interjection cannot be regarded as a linguistic means univocally indicating the impolite speech act function of a threat, there does exist a specific syntactic construction *wehe* + V2 clause with 2nd person subject in German that is to a high degree conventionalized for threats. The fact that the construction is still context-dependent to a certain degree supports the assumption that we are dealing with a constructional illocutionary force indicator whose “literal” illocution can be overridden in context.

Acknowledgments: I would like to thank Riccardo Giomi for his insightful comments and advice on an earlier version of this paper. Thanks go also to the audiences at ICCG13, University of Gothenburg, and at the Faculty of Philology, Translation and Communication, University of València, for valuable feedback, and to Ruben Solms for assistance with the corpus study.

Abbreviations not included in the Leipzig Glossing Rules

POL polite
PRT particle

References

- Apeltauer, Ernst. 1977. Drohen. In Konrad Sprengel (ed.), *Semantik und Pragmatik. Akten des 11. Linguistischen Kolloquiums: Aachen 1976*, Vol. 2, 187–198. Tübingen: Niemeyer.
- Barðdal, Jóhanna, Valgerdur Bjarnadóttir, Serena Danesi, Tonya Kim Dewey, Thórhallur Eythorsson, Chiara Fedriani & Thomas Smitherman. 2013. The Story of ‘Woe’. *Journal of Indo-European Studies* 41(3–4). 321–377.

- Culpeper, Jonathan. 1996. Towards an anatomy of impoliteness. *Journal of Pragmatics* 25. 349–367.
- Culpeper, Jonathan. 2011. *Impoliteness: Using Language to Cause Offence*. Cambridge: Cambridge University Press.
- Culpeper, Jonathan. 2018. Taboo language and impoliteness. In Keith Allan (ed.), *The Oxford Handbook of Taboo Words and Language*, 28–40. Oxford: Oxford University Press.
- Dancygier, Barbara. 1998. *Conditionals and Prediction. Time, Knowledge, and Causation in Conditional Constructions*. Cambridge: Cambridge University Press.
- Duden Deutsches Universalwörterbuch (2007). Online-Ressource. Berlin: Langenscheidt. <https://owb.langenscheidt.com> (accessed 27 June 2024)
- Falkenberg, Gabriel. 1992. Drohen. In Gabriel Falkenberg, Norbert Fries & Jadwiga Puzynina (eds.), *Sprachliche Bewertung polnisch und deutsch*, 177–192. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.
- Finkbeiner, Rita, Jörg Meibauer & Heike Wiese (eds.). 2016. *Pejoration*. Amsterdam: Benjamins.
- Fraser, Bruce. 1998. Threatening revisited. *The International Journal of Speech, Language and the Law* 5 (2). 159–173.
- Fries, Norbert. 1996. Interjektionen, Interjektionsphrasen und Satzmodus. In Inger Rosengren (ed.), *Satz und Illokution*, Vol. 1, 307–342. Tübingen: Niemeyer.
- Fries, Norbert. 2002. Die Wortart ‚Interjektionen‘. In D. Alan Cruse (ed.), *Lexikologie: Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen*, Vol. 1, 654–657. Berlin/New York: De Gruyter.
- Goldberg, Adele E. 2013. Constructionist approaches. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, 15–31. Oxford: Oxford University Press.
- Kluge, Friedrich. 2011. *Etymologisches Wörterbuch der deutschen Sprache*, 25th edn. Berlin: De Gruyter.
- Koch, Peter & Wulf Oesterreicher. 1985. Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36. 15–43.
- Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- Limberg, Holger. 2009. Impoliteness and threat responses. *Journal of Pragmatics* 41(7). 1376–1394.
- Locher, Miriam A. & Richard J. Watts. 2008. Relational work and impoliteness: negotiating norms of linguistic behaviour. In Derek Bousfield & Miriam A. Locher (eds.), *Impoliteness in Language. Studies on its Interplay with Power in Theory and Practice*, 77–99. Berlin: Mouton de Gruyter.
- Mills, Sara. 2003. *Gender and Politeness*. Cambridge: Cambridge University Press.
- Muschalik, Julia. 2018. *Threatening in English: A mixed method approach*. Amsterdam: Benjamins.
- Oxford English Dictionary 2001–. Oxford: Oxford University Press. <https://www.oed.com> (accessed 27 June 2024)
- Rolf, Eckard. 1997. *Illokutionäre Kräfte: Grundbegriffe der Illokutionslogik*. Opladen: Westdeutscher Verlag.
- Searle, John R. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.
- Searle, John R. 1978. *Expression and meaning: Studies in the theory of speech acts*. Cambridge: Cambridge University Press.
- Siemund, Peter. 2018. *Speech acts and clause types: English in a cross-linguistic context*. Oxford: Oxford University Press.
- Sode, Frank. 2023. On the conditional nature of V2-clauses in desire reports of German. In Katrin Axel-Tober, Lutz Gunkel, Jutta M. Hartmann & Anke Holler (eds.), *On the nominal nature of propositional arguments*. Hamburg: Buske, 227–262.

- Terkourafi, Marina. 2005. 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: De Gruyter.
- Van Olmen, Daniel. 2018. Reproachatives and imperatives. *Linguistics* 56. 115–162.
- 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.
- Wörterbuch der deutschen Gegenwartssprache*. 1964–1977. Kuratiert und bereitgestellt durch das Digitale Wörterbuch der deutschen Sprache. <https://www.dwds.de/d/wb-wdg> (accessed 27 June 2024).
- Wunderlich, Dieter. 1976. *Studien zur Sprechaktttheorie*. Frankfurt: Suhrkamp.

Gerhard B. van Huyssteen, Adri Breed and Suléne Pilon

7 ‘What the hell?!’ vs. ‘Wat de hel?!’: Contrasting the intensifying WHX construction in English and Afrikaans

Abstract: Certain constructions in a language (or even across languages) are conventionally associated with impoliteness; one such construction is the WHX construction, e.g., *what the hell* in English, or *wat de hel* in Afrikaans. It is claimed in this contribution that the latter is not merely a constructional calque from English, but that it is starting to take on a life of its own. Our methodology is in essence contrastive corpus linguistics: we firstly summarise our existing knowledge of the English construction (based on a previous corpus-driven study), before contrasting aspects of the English construction with Afrikaans corpus data. We focus specifically on the characteristics of the different slots in the construction in the two languages, and add an investigation of the constructional relations (attractions or repulsions) between specific WH pronouns (PN.WH) and noun phrases (NP) in the Afrikaans construction. Our investigation finds that: (a) only eight Afrikaans WH pronouns occur in the construction; (b) monosyllabic WH pronouns occur with a much higher frequency than disyllabic ones, but – unlike in English – disyllabic and even polysyllabic items are often found as noun phrases; (c) unlike in English, there is no mentionable difference between the frequency of *wat* ‘what’ and *hoe* ‘how’; (d) similar to English, *hel* ‘hell’ consistently occurred with the highest frequency in the construction; (e) bare noun phrases occur with the highest frequency in the Afrikaans construction; (f) non-morphemic word-formation strategies, and phonological and graphemic modifications are used in leetspeak to by-pass online moderators; (g) in addition to the semantic domains RELIGION and COSMOGONY activated in the English construction, the Afrikaans construction extends to RELIGION, ANIMALS, NATURAL PHENOMENA, and NEGATIVE EVENTS, but not COSMOGONY; and (h) while the English construction primarily functions as an intensifier, we find that the Afrikaans construction additionally functions as a vehicle for humour, and that this aspect could also be investigated further for English.

Gerhard B. van Huyssteen, North-West University, 11 Hoffman St, 2531 Potchefstroom, South Africa, e-mail: gerhard.vanhuysteen@nwu.ac.za

Adri Breed, North-West University, 11 Hoffman St, 2531 Potchefstroom, South Africa, e-mail: adri.breed@nwu.ac.za

Suléne Pilon, University of Pretoria, Lynnwood Rd, 0132 Hatfield, Pretoria, e-mail: sulene.pilon@up.ac.za

Keywords: Afrikaans, constructional calque, contrastive corpus linguistics, covarying collexeme analysis, taboo language

Warning: This chapter contains examples of language that might be offensive to some users.

Disclaimer: Our classification of racist, homophobic, and other offensive expressions is based on our subjective understanding and analysis of their context and intended meaning, and does not imply endorsement or agreement with these views.

1 Introduction

(Im)politeness is typically associated with emotive psychological states of mind, emotively motivated human behaviour, perceptions and expectations of what is appropriate or not, (dis)agreeable social interactions and relationships, cultural identity, etc. (see for instance Hickey and Stewart 2005: 1; Kádár 2017: 1; Spencer-Oatey 2005: 91). These and other aspects of (im)politeness is most often realised in linguistic behaviour – from gestures and body language to overt expressions in politeness formulae (like greetings), to subtle expressions through sarcasm, humour, and other indirect forms of communication. As such, we view specifically impolite linguistic behaviour as the (semi-/non-)intentional usage of various contextually determined expressive/emotive language acts with some kind of negative (e.g., abusive, rude, dysphemistic, etc.) perlocutionary effect (see, among others, Culpeper 2011: 23; Jay 2008: 268; Jay 2020: 39; see Hoeksema 2019 for a variety of examples).

Hence, language and linguistic analyses more often than not occupy a central position in (im)politeness studies. The predominant approaches in such studies promulgate the idea that impoliteness is not inherent in linguistic constructions, but rather a contextual judgement. However, recent research (see also Van Olmen et al., this volume) started to prove that, in addition to the importance of context, certain constructions in a language are conventionally associated with impoliteness. The argument that follows is that linguistic form and meaning pairings (i.e., constructions) deserve a more prominent place in impoliteness research. For example, Van Olmen, Andersson and Culpeper (2023: 38) conclude that “there do exist constructions in language that are (to a large extent) conventionalised for impoliteness and that impoliteness can be strongly conventionalised across languages”. However, to see how widespread such inherently impolite constructions are, more research is of course required.

The aim of this chapter is to contribute to this endeavour by investigating the intensifying WHX construction (e.g., *what the hell* . . ? in English) in Afrikaans (see example (1) below), contrastively with English (example (2) below). We adapted the term “WHX construction” from Hugou (2017), who explained it as follows: “The letter ‘W’ designates a wh- word (*what, where, when, who*, etc.), or a word which has a similar syntax to wh- words, such as *how*. The letter ‘H’ refers to a noun phrase such as *the hell*, and the letter ‘X’ stands for the rest of the interrogative clause, i.e. the proposition.” In our adaptation, WH represents a WH pronoun (e.g., *what, where, how*, etc. in English), and x the noun plus its pre- and postmodifiers in either a noun phrase (e.g., *the hell*), or prepositional phrase¹ (e.g., *in the hell*). As can be seen in (3), we therefore leave “the rest of the interrogative clause” out of the construction name, since for us WH = WH word, and x = noun plus pre-/postmodifiers in a noun phrase (NP) or prepositional phrase (PP).

- (1) *Wat de fok rook die ou?*
 what the fuck smoke the guy
 ‘What the fuck is he smoking?’
 (OSMO-2023)
- (2) *Hey, this guy doesn’t know what the fuck consciousness is either.*
 (enTenTen21-2021)

(3) **Intensifying WHX construction**

[_{Z_{PN,WH}} [_{NP} _{Y_{ART.DEF}} _{X_N}]] AND [_{Z_{PN,WH}} [_{PP} _{P_{PREP}} _{Y_{ART.DEF}} _{X_N}]]
 where x = a taboo word (e.g., *fuck*), or euphemism (e.g., *flip* ‘freck’)

Despite it being indisputably a highly frequent, productive and ubiquitous construction in both Afrikaans and English,² research on the Afrikaans WHX construction is non-existent, while Hugou’s (2017) corpus study is – to our knowledge – the only comprehensive study of the construction in English. In fact, Hugou (2017: 1)

1 Although prepositional phrases can also be used in the x part of the Afrikaans WHX construction (for example, *wat in die wêreld* ‘what in the world’), we narrow our focus in this chapter to look only at noun phrases in the x part (also see footnote 9).

2 Preliminary investigations regarding the same construction in Dutch indicate that it might not be as frequent, productive, or ubiquitous as in Afrikaans and English. For instance, in the Dutch nlTenTen20 corpus (Sketch Engine 2020), comprising 5,890,009,964 words, *wat de hel* (‘what the hell’) only appears 80 times (relative frequency of 0.0136 per million words; for comparative frequencies in English, see Hugou (2017), and Afrikaans, see Section 6 of this chapter). The Dutch construction has also not been investigated properly, bar some preliminary remarks (Taalprof 2009), and a brief corpus investigation (Van der Wouden 2019).

shows that “no more than spotty attention has been paid to this construction in the existing literature”. Recent literature searches using WorldCat³ and Litmaps⁴ also did not reveal any new literature on the construction since Hugou’s article. This chapter therefore addresses a gap in the existing literature on this construction as an example of an inherently impolite grammatical construction.

Our research is done contrastively with Hugou’s (2017), implying that we focus on those aspects where the Afrikaans construction might differ from its English counterpart. For example, while Hugou is primarily interested in the expression of intensity, using the WHX construction as a case study to explore whether lexical and syntactic variation in a construction correlate with a variation in intensity, our focus is more only on the constructional variation of the construction. We accept his findings on the expression of intensity as a priori for our study (and therefore for Afrikaans as well), which gives us the opportunity to rather focus on the variation in the Afrikaans construction, and to what extent it is similar to, or different from the English construction (i.e., in contrast with the English construction). Such a contrastive approach is pertinent to make the central claim of our research, namely that the Afrikaans construction is a constructional calque that it is starting to take on a life of its own.⁵

We define a constructional calque as a construction schema (i.e., a construction with less specified, schematic slots) from a donor language, which is realised in a recipient language by means of forms (e.g., typically phonological and/or lexical items) that are analogous to the forms in the donor language. Importantly, the donor language’s original syntactic pattern, meaning, and function are by and large preserved in the constructional calque. In other words, the construction in the recipient language mimics a construction in the donor language by directly translating each component (slot) in the construction. As Michaelis and Haspelmath (2020: 1121) put it: “the meaning of a composite form in the donor language is rendered by an analogous composite form in the recipient language.”

These and other kinds of interlingual “contamination” phenomena are aptly explained from a diasystematic construction grammar (DiaCxG) perspective – a sub-theory in construction grammar (CxG) that explains various emergent multi-/bilingual phenomena (Höder 2018, 2019; Höder, Prentice, and Tingsell 2021). A central notion in DiaCxG is that diasystematic (i.e., language-unspecific) schemas could instantiate specific lower-level idiosyncratic (i.e., language-specific) realisations. From this emerges the view of a multilingual construction that contains

³ <https://search.worldcat.org/>

⁴ <https://app.litmaps.com/>

⁵ See Taalprof (2009) and Van der Wouden (2019) for similar views on the development of the equivalent Dutch construction.

both language-unspecific diasystematic constructions (diacxns), and language-specific idiosyncratic constructions (idiocxns), which can all be active and entrenched on various levels of schematicity. Höder (2019) illustrates this with the case of the Danish idiocxn *fin* [[fi:n] ⇔ [FINE, DELICATE]]_{DAN} and the German idiocxn *fein* [[fəi:n] ⇔ [FINE, DELICATE]]_{GER} that result in the diacxn [[f] . . [n] ⇔ [FINE, DELICATE]]_{DIACXN}. Similarly, constructional calques can be explained as diacxns that emerged due to analogous change based on formal similarities between idiocxns in two (or more) languages. We will return to this perspective in the conclusion of the chapter.

To support our central claim, we want to answer the following specific questions:

1. Which similarities in, and differences between the definite-article part, WH part, and noun phrase part of the English and Afrikaans construction can be identified?
2. What are the constructional relations (attractions or repulsions) between specific WH pronouns, definite articles, and nouns in the Afrikaans construction?

In the next section, we commence with a rather detailed overview of Hugou (2017), in order to firstly summarise those cognitive-pragmatic aspects of the construction that we do not investigate (and hence take as axiomatic for the Afrikaans construction), and secondly to identify the formal characteristics of the English WHX construction (which will be used for comparative purposes throughout the article). Section 3 contains a brief overview of our methodology and our Afrikaans dataset. Using simple frequency counts, research question 1 above is addressed in Sections 4 (the definite-article slot), 5 (the WH pronoun slot), and 6 (the noun phrase slot). By means of a covarying collexeme analysis (Gries and Ellis 2015; Stefanowitsch and Gries 2003, 2005), we investigate in Section 7 whether certain Afrikaans pronouns attract specific nouns (research question 2 above). In Section 8 we summarise and draw conclusions.

2 The English WHX construction

In his article, *The WHX construction (what the hell. . .?) and intensity: A corpus-based study*, Hugou (2017) is primarily interested in the expression of intensity, using the WHX construction as a case study to explore whether lexical and syntactic variation in a construction correlate with a variation in intensity. From a constructionist perspective, it is taken as an axiom that changes in form impose a particular construal, and therefore always have conceptual import (see for example Langacker 1991: 302). Hence, if there is variation in the construction, it should have some effect on the semantics or pragmatics of the construction. He therefore explores how the

WHX construction “is motivated by the meaning of its parts and [how] the way in which they are combined is inherently meaningful” (Hugou 2017: 5), and hypothesises that “there might be a correlation between varying degrees of intensity and the many sub-types of the WHX construction, be they lexical (*what the hell / fuck / heck / holy hell* . . .) or grammatical in nature (*what in the hell / in hell / the hell* . . .)” (Hugou 2017: 12).

He argues that the WHX construction inherits features from, among others, the subject-auxiliary inversion construction in counterfactual conditionals (*Had I known that before!*), and exclamatory-inversion constructions (*Did I know little before!*), as well as WH question constructions (e.g., *What didn't I know before?*). In speech acts, the WHX construction therefore does not express assertiveness, but rather inquiry, i.e., it expresses “an informational need and therefore an unsatisfied emotion from someone, namely a doubt, or any notion, as distinguished from fact” (Hugou 2017: 6). Because the construction “is structurally similar to an interrogative clause, but [. . .] conveys a similar illocutionary force to an exclamatory clause”, it is used with a “predominantly emotive function to express different feelings such as anger, surprise or dismay; [it has] a strong subjective quality, in that [it communicates] the speaker's emotional reaction to a situation” (Hugou 2017: 7). Since it is such an emotive construction, the intensity effect “is a mere corollary, something that naturally follows or results from the expression of a strong emotion” (Hugou 2017: 25).

This emotive function makes it especially difficult (if not close to impossible) to pinpoint the exact intensity and range of emotions expressed by examples in text corpora (i.e., without pronunciation, gestures, facial expressions, etc.). He therefore investigates users' perception of intensity in various variants of the WHX construction by means of two questionnaires and follow-up email interviews – a method that falls outside the scope of our current methodology and research. His findings based on these questionnaires include:

- Frequency of use influences the perceived intensity: the more familiar a construction, the higher the intensity attributed to it. Subsequently, the [_{PN.WH} *the hell*] construction, compared to [_{PN.WH} *in the hell*] and [_{PN.WH} *in hell*], is perceived as the most intense due to its high frequency and familiarity.
- The perceived intensity of the WHX construction is closely related to the offensiveness of [x]: taboo words like *fuck* and *hell* enhance the intensity, while euphemisms like *freak*, *heck*, or *blankety blank* diminish it. Somewhat contrary to the first finding, constructions with *fuck* are perceived as more intense than those with *hell*, while *what the hell* is more frequent than *what the fuck*. This is of course unsurprising, since *fuck* has a higher taboo value than *hell*.
- Perceived intensity is also linked to the length of the realisations of the construction: shorter forms of the construction (*what the hell*) are more intense than longer forms (e.g., *wherever the hell*, or *what the thrice cursed hell*). In

fact, longer versions have “an archaic or humorous flavor” (Hugou 2017: 23), similar to other so-called attention-seeking euphemisms (Burridge and Benczes 2019: 189). He points out: “By using a longer variant (e.g. *worldwide hell*, *flying blue fuck*), the speaker, or rather the hearer is invited to reconceptualize his or her experience, that is view everything from a different perspective” (Hugou 2017: 23). It therefore requires more processing effort, which might distract the hearer from the actual message (Hugou 2017: 24).

- Factors related to the linguistic, socio-cultural, and situational context also play a role in the perceived intensity – as is the case in all other constructions that are heavily influenced by socio-pragmatic factors.

For the description of the constructional variation and the constraints on such variation, he uses corpus data from the *Corpus of Contemporary American English* (COCA). His observations can be summarised in terms of the WH part of the construction; the NP part of the construction; and the construction as a whole. (For ease of reference, these observations are numbered sequentially with small letters.)

2.1 The WH part of the English construction

a) Based on his analyses (Hugou 2017: 13), we can divide English WH (compounded) words according to their frequency rank order into three groups, as in (4).⁶

- (4) *what* (48,328)
 > *how* (1,532) > *why* (1,445) > *who* (1,367) > *where* (1,264)
 > *whatever* (269) > *when* (77) > *whoever* (30) > *wherever* (28) > *however* (4) > *whenever* (1)

b) It is clear that WH compounded words rarely appear in the construction, while *which* and *whichever* never appear in the construction in Hugou's (2017) data.

2.2 The noun phrase part of the English construction

c) The noun phrase part of the construction can either be a bare noun phrase (*what the hell*), or an NP in a prepositional phrase (*what in the hell*), which in turn can be internally modified by adjectives (*what the bloody hell*; *what the flying blue fuck*).

⁶ Words are listed from most frequent to least frequent; in parentheses are summations of the frequency counts in Hugou's (2017: 14–16) Table 1 and 2.

In addition, the noun can be remodelled by means of non-morphemic word-formation strategies (*what the eff*; *what the f. . .*; *WTF*; *what the . . .*), morphemic word-formation strategies,⁷ and phonological / graphemic modifications (*what the fudge*; *what the %!**; *what the fuuuck*).

d) The noun phrase part of the construction shows “great lexical diversity, ranging from high-frequency variants [*what the hell*; *who the fuck*], to on-the-spur-of-the-moment creations [*why in the sweet world of wonder*; *what the piss*] which, for the most part, die aborning” [our additions, from his examples] (Hugou 2017: 16).

e) The noun in the noun phrase is mostly monosyllabic (*hell*, *heck*; *fuck*, *freak*, *fudge*; *shit*; *earth*, *world*), similar to other taboo words (Allan and Burridge 1991: 145–146). Exceptions of course do occur, with *devil* and *God’s name* the most noticeable.

f) Words from the Greek or Latin stratum (i.e., “learned words” – (Hugou 2017: 18)) are not likely to occur in the construction, also similar to other taboo constructions.

g) Longer, more ornate variants of the noun phrase (*the worldwide hell*; *the flying blue fuck*) dilute the intensity of the construction, as was pointed out above. Unsurprisingly, longer noun phrases tend to have an “archaic or humorous flavor” (Hugou 2017: 23).

h) With regards to the semantic fields of nouns, Hugou (2017: 18) concludes that the construction favours words and phrases belonging to RELIGION / COSMOGONY, such as *hell*, *devil*, *God’s name*, *world*, and *earth*. In all cases the noun is used as profanity (i.e., “with careless irreverence” – see Allan and Burridge (1991: 37)), and not as blasphemy (i.e., that which “vilifies or ridicules the deity” – see Allan and Burridge (1991: 37)).

i) The two exceptions to the semantic constraints are *fuck* and *shit*, which, as Hugou (2017: 18) correctly points out, have by and large lost the connections to their original semantic fields (SEX and BODILY EFFLUVIA), and have become all-purpose intensifiers.

j) He finds that, as the noun in the noun phrase part of the construction, *hell* and its variants (N=10,127) are used consistently more frequently than *fuck* and its variants (N=1,769). He ascribes this to two potential factors, viz. that:

– His data comes from written genres (novels, magazines, newspapers);⁸ and

⁷ No examples provided by Hugou (2017: 17).

⁸ A search in The Movie Corpus (containing 200 million words of English data from movie dialogues – see Davies (2019)) supports this suspicion of Hugou (2017: 15). Whereas the ratio between *hell/fuck* is 85:15 in COCA, in The Movie Corpus it is 57:43 (with N for *what the hell*=30,954; *what the fuck*=23,348). Further research is required to confirm the hypothesis that frequencies of the

- there might be a “need to use more appropriate language in certain social situations” (Hugou 2017: 15) (as reflected by the formerly mentioned written genres). Since both *hell* and *fuck* are regarded as general intensifiers in English, it is unsurprising that they occur with high frequency in the construction. Hugou (2017: 18) argues that their high frequency should be explained in terms of their general promiscuity (Taylor 2002: 266) as intensifiers in English.

k) While *hell* and *fuck* have many variants in the data, *devil* has only two: *dickens* (an established euphemism for *devil*), and *deuce*. Hugou (2017: 18) notes that their “dated character may leave little need for creativity in present-day English.” He also ascribes *hell*’s survival in the construction to this “dated character” (in addition to *hell*’s aforementioned status a general intensifier)

l) Based on Bybee’s (2010) definition and operationalisation of productivity (i.e., a greater variety of items appearing in the free slots of a construction is an indication that the construction is entrenched and conventionalised, and therefore more likely to be productive), he concludes that the construction is “very productive” (Hugou 2017: 16), since it has a rather high type frequency. The large number of hapax legomena of the construction is also an indication of the construction’s high hapax-conditioned degree of productivity, as defined and operationalised by Baayen (1993).

m) However, since the construction favours certain semantic areas (RELIGION / COSMOGONY), as well as monosyllabic words, Hugou (2017) concludes that the construction resists full productivity.

n) Note that Hugou (2017) does not investigate variants of the article in the noun phrase, such as *what da fuck* (N=2 in COCA).

2.3 The English construction as a whole

o) He identifies nine different syntactic variations of the English construction (see (5) to (13) below, based on Hugou (2017: 13)), of which the equivalents of only (5) and (9) are attested in our Afrikaans data.⁹ Of all these sub-constructions, (5) dis-

construction in written and spoken data will differ significantly. Also see Hugou’s (2017: 26) suggestions for further research in this regard.

⁹ Although sentences such as (6) and (7) have been excluded from our dataset due to our search criteria (see Section 3), they do occur in Afrikaans, but – as far as we know – only in the constructions [z_{PN.WH} [pp op deeske aarde] . . . ?] and [z_{PN.WH} [pp op aarde] . . . ?] (‘WH on (this) earth . . . ?’). Since these cases fall outside the scope of this chapter (see footnote 1), it is left for future research.

plays by far the highest frequency in the *Corpus of Contemporary American English*, followed by (6) and (7) with much lesser frequencies (Hugou 2017: 14–15), and the others trailing far behind.

- (5) [z_{PN.WH} [NP *the* x_N] . . . ?] (*Who the hell is this guy?*)
- (6) [z_{PN.WH} [PP *in the* x_N] . . . ?] (*Where in the hell have you been?*)
- (7) [z_{PN.WH} [PP *in* x_N] . . . ?] (*Why in hell can he never be on time?*)
- (8) [*the* [x_N] . . . ?] (*The hell you doing?*)
- (9) [z_{PN.WH} [NP *the* x_N]?] (*“What the hell?” Johnny said with surprise.*)
- (10) [z_{PN.WH} [NP *the* x_N q_N] . . . ?] (*What the hell difference does that make?*)
- (11) [z_{PN.WH} [ADV *the* x_N q_{ADV}] . . . ?] (*How the hell long do I have to keep doing this?*)
- (12) [z_{PN.WH}’s [NP *the* x_N q_N] . . . ?] (*What’s the fuck are you talking about?*)
- (13) [how’d [NP *the* x_N] . . . ?] (*Well, how’d the hell it get there?*)

r) The construction in (9) is a case of aposiopesis, where a sentence is not being completed by the speaker. Hugou (2017: 13) argues that this construction might have developed into a separate construction, especially in cases where it expresses dismissal (*I wasn’t supposed to tell you yet, but what the hell!*). It is, however, most often not possible or easy to differentiate between such cases, and it is therefore categorised as a sub-construction of the WHX construction.

s) The construction has a constructional preference for the modal auxiliary *would*, as in *Why the fuck would I want to kill Dominic?* This ties in with the idea that WH questions typically “presents an incomplete proposition” (Hugou 2017: 5), even though the construction as a whole is used to “overtly express [speakers’] stance toward a certain state of affairs: the unsatisfied emotion” (Hugou 2017: 6).

t) While the construction most often appears in direct WH questions (*What the fuck are you doing here?*), it is also found in indirect questions (*I didn’t know what the hell I was talking about.*). Nevertheless, the construction most often appears in initial rather than clause-final position, as first noted by Huddleston and Pullum (2002: 916).

u) Since so-called telling questions (Thompson, Fox and Couper-Kuhlen 2015) often exhibit the powerlessness of the speaker, the construction is often found in such questions (*Where the hell did you meet Virgil's wife at? Hell, I've lived next door and I've only caught a few glimpses of her myself.*). The same also applies to rhetorical questions (*How the hell would I know?*), which often “testifies to the attitudinal stance of the speaker” (Hugou 2017: 10). On the other hand, the construction is not associated with so-called reprise questions (*Tracy likes who (*the hell)?*) (Hugou 2017: 11).

v) In nominal relative clauses the construction only appears if it is introduced with a *WH-ever* compound (*I won't sit still for extortion or manipulation or... whatever the hell she has in mind.* vs. *What (*the hell) I like is chocolate.*) (Hugou 2017: 9).

w) In short, Hugou (2017: 21) concludes that “the WHX construction functions like expletives”.

3 Afrikaans data

Against this existing knowledge of the English construction, we are particularly interested in the following aspects of the Afrikaans equivalent of the construction:

- variation in the definite-article slot of the construction;
- the frequency with which WH words occur in the construction;
- the lexical items that can occur after the definite article to fill the x part of the construction; and
- the relationship or interaction between the various WH words and x.

While Hugou's description is based on COCA data and questionnaires, our description of the WHX construction is based on data from two corpora¹⁰ that are inherently different from his – both in size and nature. He used the *Corpus of Contemporary American English*, which contains more than 1 billion words of text (Davies n.d.), while the two corpora that we use, together contain only about 61 million words (see Table 1). Despite these differences, we will compare our findings about the

¹⁰ In the initial phase of our investigation of the Afrikaans WHX construction, we also analysed data from a third corpus, namely the *Language Commission Corpus* (TK; Taalkommissie van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns 2011), which contains only edited data from Afrikaans publications. We found that the construction occurs with a very low relative frequency, and with very little variation in this corpus, most probably due to the intervention of editors of these kinds of texts. We therefore did not include the results from this corpus further in our investigation.

Afrikaans construction with Hugou’s (2017) discussion of the English construction throughout, without making any claims that would require comparable corpus data.

Table 1: Corpus size, number of instances (N), and frequency per million (fpm) of the WHX construction.

Corpus	Size	N	fpm
ModCor	9,024,829	382	42
ComCor	51,693,425	2,031	39
Total	60,718,254	2,413	40

It is not clear whether Hugou (2017: 25) used the full *Corpus of Contemporary American English*, or only “written genres (novels, magazines, newspapers)”; we therefore assume that he used the full corpus. Our corpora cover the following:

- The *NWU Corpus of Moderated Comments* (ModCor; Centre for Text Technology (CTeX) 2020) contains commentary posts from a newspaper website, which were deleted by the newspaper’s team of moderators because they were deemed to be impolite or inappropriate. This corpus is used for the initial description, since we expect to find good examples of impolite constructions here, given that the data in this corpus was already categorised as impolite/inappropriate by external respondents (the moderators). 382 instances of the construction were found, resulting in 42 instances per million words (see Table 1).
- The *NWU Commentary Corpus* (ComCor; Centre for Text Technology (CTeX) 2023) contains unedited data from the comment section of news websites and blogs. This corpus is therefore in nature very similar to ModCor, but unlike ModCor, the data in ComCor is not exclusively impolite/inappropriate. We therefore expect to find in this corpus innovative uses of the construction, where language users mask impoliteness by not using lexical items with a high taboo value in the noun phrase part of the construction. For this reason, data from ComCor is used to expand and refine the initial description of the construction. We found in total 2,031 instances of the construction, which is proportionally similar to that of ModCor, viz. 39 instances per million words (see Table 1).

To extract data, we used the CQL search string in (14); the strings to the right of the concordance results were considered to be candidates for x – i.e., nouns (*wat de hel* ‘hell’), nouns with premodifiers (*wat de flippen hel* ‘what the flipping hell’),

or nouns with postmodifiers (*wat de hel op aarde* ‘what the hell on earth’). As will transpire in the next section, the definite article *de* ‘the’ occurs much more frequently than any other variations (e.g., *die* ‘the’) in the definite-article slot of the construction; we therefore excluded variations of the definite article from our dataset, and only use instances with *de* as definite article.¹¹

- (14) [word="(?)wat.*|hoe.*|wie|waar.*|hoekom|wanneer"] [word="se"]?
[word="(?)de"]

In our dataset, all instances of grawlix (e.g., %\$&#@) were regarded as a single type, since it is mostly impossible to determine which specific taboo item was intended by the author. Other subterfuge phenomena, like leetspeak (e.g. *f@k* ‘f@ck’), redacting (e.g. *f*k* ‘f*ck’), and abbreviation (e.g. *f < fok* ‘fuck’), were normalised to the intended taboo item. The total number of tokens also includes three instances where the x part was left empty. Once again, these instances are regarded as belonging to a single type, although different taboo items might have been omitted. Spelling errors and typos (e.g. *vok* instead of *fok* ‘fuck’), and spelling variants (e.g. *donner*, which is a spelling variant of *donder* ‘thunder’) were also normalised. However, instances of leetspeak with valid words (e.g. homophones like *vlok* ‘flake’ instead of *fok* ‘fuck’), were left unchanged, since we regard these as examples of creative extensions of the construction. Incorrect spellings of valid words (e.g. **flok* instead of *vlok* ‘flake’), were normalised to the conventional spelling.

4 The definite-article part of the WHX construction

As was shown in (14) above, we only included *de* ‘the’ in the definite-article part of our main corpus searches regarding the WH part, and the x part of the construction. To investigate variation in the definite-article part, we restricted our search to ComCor, and to those two instances of x that occur most in the construction, viz., *hel* and *fok* (see Section 6). The search string is presented in (15).

- (15) [word="(?)wat.*|hoe.*|wie|waar.*|hoekom|wanneer"] [word="se"]?
[word=".*"] [word="(?)hel|fok"]

¹¹ Hugou (2017) also restricted his investigation to only *the* in the definite-article part of the English construction.

After filtering out irrelevant cases (e.g., *mense wat **die** hel se vure laat brand* ‘people who stoke the fires of **the** hell’), the frequencies in ComCor are displayed in Table 2. It is quite surprising to note that *de* appears in the construction with a frequency more than ten times higher than *die*. One would have expected *die* to be the most frequent, since it is the prototypical definite article in Afrikaans, and also the highest-frequency word in Afrikaans. The word *de*, on the other hand, is usually only found in archaic fixed expressions from Dutch (e.g., *om **de** dood nie* ‘by no means’), surnames (e.g., ***De** Lange*), and loan phrases (e.g., ***de** facto*). This unexpected occurrence can be explained quite easily from a diasystematic construction grammar (DiaCxG) perspective.

Table 2: Frequency of lexical items in the definite-article slot of the WHX construction in ComCor.

Rank	Definite article	N
1	<i>de</i>	1,091
2	<i>die</i>	89
3	<i>te</i>	29
4	<i>di</i>	4
5	<i>da</i>	1
6	<i>the</i>	1
Total		1,215

As was pointed out in Section 1, DiaCxG posits that multi-/bilingual speakers have access to both language-specific idiosyncratic constructions (idiocxns) and language-unspecific diasystematic constructions (diacxns), which emerge based on formal similarities between idiocxns in two (or more) languages. To apply these ideas to the Afrikaans WHX construction, it is important to know that most adult Afrikaans speakers can also speak English, albeit often accented, and with various idiosyncrasies (Van Rooy and Wasserman to appear). For example, it is a well-known fact that English [ð, θ] is often substituted for dental(ised) plosives [d, t] by some Afrikaans speakers.

Based on this general observation, we can now propose that in the multi-/bilingual constructicon of Afrikaans speakers, the English idiocxn *the* [[ðə] ⇔ [THE]]_{ENG} and the Afrikaans idiocxn *die* [[di] ⇔ [THE]]_{AFR} results in the diacxn *de* [[də] ⇔ [THE]], so that *the* is then realised as [də] in the English of many Afrikaans speakers. In a loan construction like the WHX construction, the strong phonological similarities in the immediate textual context of ⟨*the*⟩ also strengthen this diacxn schema, e.g., on the lefthand side [[*wat*] ⇔ [WHAT]]_{ENG} vs. [[*vat*] ⇔ [WHAT]]_{AFR}; and on the righthand side [[*həl*] ⇔ [HELL]]_{ENG} vs. [[*ɦəl*] ⇔ [HELL]]_{AFR}, or [[*fΔk*] ⇔ [FUCK]]_{ENG} vs.

[[fɔk] ⇔ [FUCK]]_{AFR}. During the process of calquing, written English ⟨*the*⟩ is then realised in writing as ⟨*de*⟩ instead of the expected ⟨*die*⟩. The fact that ⟨*die*⟩ does occur in the Afrikaans construction (see Section 3), might be a sign of a newer, more recent development, so that we can postulate the following developmental path (where round parentheses indicate non-entrenched items) as hypothesis for future research:

([vat di hɛl]_{AFR}) <
 (⟨*wat die hel*⟩) <
 ⟨*wat de hel*⟩ <
 [vat də hɛl]_{AFR} <
 [wat də hɛl] <
 [wat ðɪ hɛl]_{ENG}

Therefore, the frequent use of *de* 'the' instead of the more prototypical *die* 'the', sets the scene for our assertion that the Afrikaans construction is a constructional calque. In the next sections, we investigate this assertion further when discussing the WH part and X part of the Afrikaans construction.

5 The WH part of the WHX construction

5.1 Identification and description (ModCor)

Table 3 presents the raw and relative frequencies¹² of WH pronouns used in the WH part of the construction in ModCor, while the relative frequency results are also visualised in Figure 1. Seven out of a potential total of ten WH pronouns (see Kirsten and Breed (2020) for an overview) are used in the construction; (16) to (22) below contain an example of each).¹³ As is the case for the English construction (see Section 2, point a, they can be grouped into three frequency categories, viz.:

wat 'what' (N=153), *hoe* 'how' (N=141)
wie 'who' (N=49), *waar* 'where' (N=32)
hoekom 'why' (N=5), *wanneer* 'when' (N=1), *waarom* 'why' (N=1).

¹² Unless indicated otherwise, relative frequency in this chapter is the frequency per million (fpm) words.

¹³ For the sake of brevity, we don't provide glosses for the examples, but rather only translations, since in most cases the Afrikaans words relate almost directly to the English, e.g., *hel* 'hell', or *fok* 'fuck', and *duiwel* 'devil'.

Also like in the English construction, the four WH pronouns occurring with the highest frequency are all short (i.e. contain only one syllable) and morphologically simplex (see Section 2, point b). Longer, morphologically complex WH pronouns like *waarom* ‘why’ occur only once in the construction, and other morphologically complex WH pronouns like *waarop* ‘on what’, and *waarvoor* ‘for what’ do not occur at all. Also, none of the so called attributive interrogative pronouns (Kirsten and Breed 2020), like *watter* and *wat se* ‘which’, occurs in the construction.

Table 3: Frequency of WH pronouns in the WHX construction in ModCor.

Rank	WH pronoun	Translation	N	fpm
1	<i>wat</i>	what	153	16.95
2	<i>hoe</i>	how	141	15.62
3	<i>wie</i>	who	49	5.43
4	<i>waar</i>	where	32	3.55
5	<i>hoekom</i>	why	5	0.55
6	<i>wanneer</i>	when	1	0.11
7	<i>waarom</i>	why	1	0.11
Total			382	

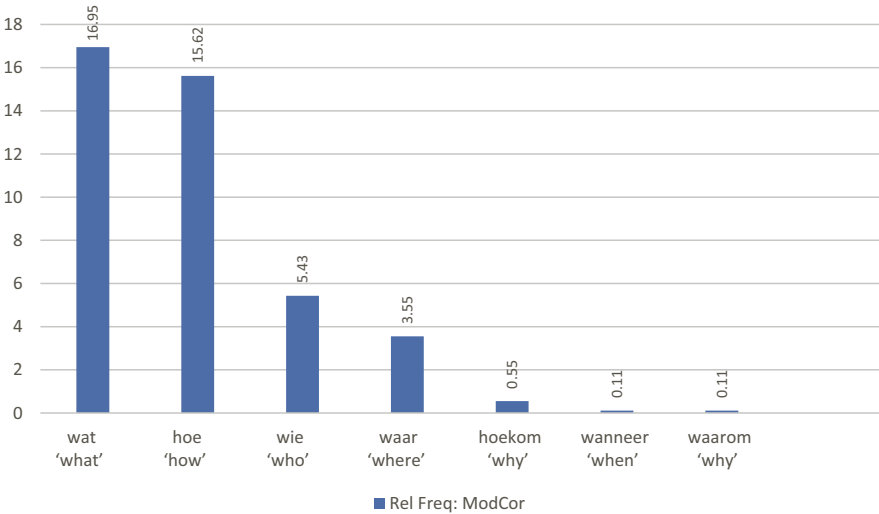


Figure 1: Relative frequency of WH pronouns in WHX in ModCor.

- (16) ***Wat** de dinges is fout met jou man!!!!????* (ModCor-2018)
'What the thingamajig is wrong with you man!!!!????'
- (17) ***Hoe** de duivel kry mens dit reg . . .* (ModCor-2019)
'How the devil does one manage that . . .'
- (18) ***Wie** de heng noem jou kind "Success"?* (ModCor-2018)
'Who the heck calls your child "Success"?'
- (19) ***Waar** de f%????% werk dit so* (ModCor-2019)
'Where the f%????% does it work like that'
- (20) *As dit dan common is **hoekom** de joos lees jy dit?* (ModCor-2016)
'If it is common then why the devil do you read it?'
- (21) ***Wanneer** de duivel praat iemand due waarheid??* (ModCor-2019)
'When the devil does someone speak the truth??'
- (22) . . . ***waarom** de donner is daar al die wette om hulle in blanke besighede en boerderye in te forseer.* (ModCor-2018)
' . . . why the hell [thunder] are there all these laws to force them into white businesses and farms.'

The Afrikaans WH pronouns occurring in the construction also differ from those in the English construction in two ways. Firstly, *wat* 'what' and *hoe* 'how' have very similar frequencies in the Afrikaans construction, while in English *what* has a much higher frequency than *how*. Secondly, *why* has the third highest frequency in the English construction, while the Afrikaans translation equivalents, *hoekom* 'why' and *waarom* 'why', have a much lower frequency in this construction. Since Hugou (2017) pointed out that the construction favours short lexical items, this discrepancy is probably due to the length of the Afrikaans lexical items, making them less likely to occur in the construction. This indicates that, as is the case for the English construction, the Afrikaans construction also prefers shorter lexical items in the WH part.

5.2 Verification and extension (ComCor)

Table 4 presents the raw and relative frequencies of WH pronouns in the WH part of the construction in ComCor. For comparison, the relative frequencies of the WH

pronouns in the construction in ComCor and in ModCor are visualised in Figure 2. From this visualisation it is clear that the distribution of WH pronouns used in the construction is very similar in the two corpora. The relative frequencies of the WH pronouns are also very similar in the two corpora, indicating that the construction occurs with similar frequency in both corpora.

Table 4: Frequency of WH pronouns in the WHX construction in ComCor.

Rank	WH pronoun	Translation	N	fpm
1	<i>wat</i>	what	812	15.71
2	<i>hoe</i>	how	750	14.51
3	<i>wie</i>	who	264	5.11
4	<i>waar</i>	where	171	3.31
5	<i>hoekom</i>	why	26	0.50
6	<i>wanneer</i>	when	4	0.08
7	<i>waarom</i>	why	3	0.06
8	<i>waarvoor</i>	what for	1	0.02
Total			2,031	

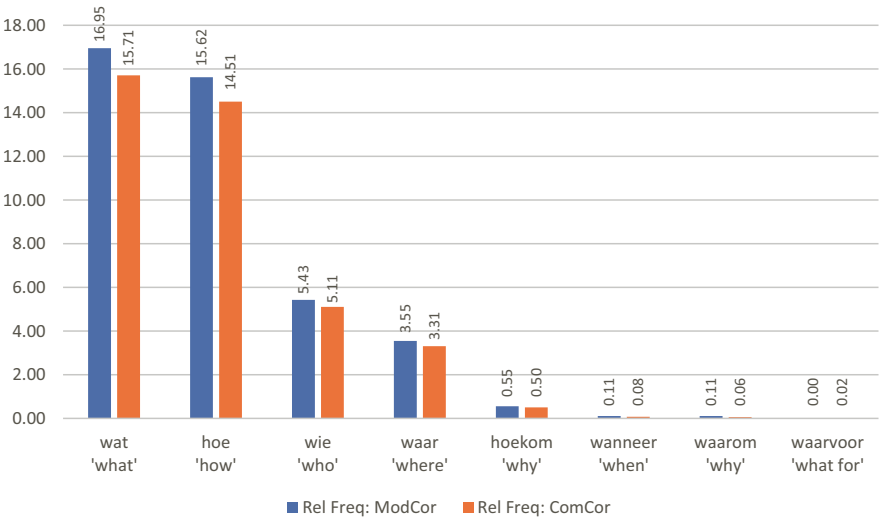


Figure 2: Relative frequencies of WH pronouns in the WHX construction in ModCor and ComCor.

Once again the WH pronouns group together in three frequency categories with *wat* ‘what’ (N=812) and *hoe* ‘how’ (N=750) occurring with a much higher frequency than the other WH pronouns. The second frequency grouping again consists of *wie* ‘who’ (N=264), and *waar* ‘where’ (N=171), and the disyllabic WH pronouns *hoekom* ‘why’ (N=26), *wanneer* ‘when’ (N=4), and *waarom* ‘why’ (N=3) occur with the lowest frequency. We found one additional WH pronoun, namely *waarvoor* ‘what for’, occurring once (see (23)) in the WH part of the construction.

- (23) *Nou as alles dit deur die staat betaal word, **waarvoor** de F@K is sy salaris dan????* (ComCor-2019)
 ‘Now if everything is paid for by the government, what the f@ck is his salary for then????’

6 The x part of the WHX construction

6.1 Identification and description (ModCor)

Table 5 contains the lexical items that are used in the x part of the WHX construction in ModCor.¹⁴ The 382 instances consist of only 30 types, which include a type marked [GRAWLIX] (e.g. *hoe de #\$\$@&&&* ‘how the #\$\$@&&&’ – 28 instances), and a type consisting of [EMPTY] slots (3 instances). Examples with lexical items occurring with a frequency N>10 are presented in (24) to (31).

Table 5: Frequency of lexical items in the x part of the WHX construction in ModCor (N>10).

Rank	x	Translation	N
1	<i>hel</i>	hell	182
2	<i>donder</i>	thunder; interjection	29
3	[GRAWLIX]		28
4	<i>duiwel</i>	devil	25
5	<i>moer</i>	coffee sediment; beat up; interjection	24
6	<i>fok</i>	fuck	22
7	<i>joos</i>	devil	14
8	<i>ongeluk</i>	accident	12

¹⁴ Given that the ns are in principle an open list, across corpora, we do not provide relative frequencies for ns. Where comparisons are made, we report on the rank order in the respective corpora, based on raw frequencies.

Table 5: continued

Rank	x	Translation	N
9	<i>heng</i>	heck	7
10	<i>dinges</i>	thingamajig	5
11	<i>bliksem</i>	lightning; interjection	4
12	[EMPTY]		4
13	<i>vlok</i>	flake < fuck	4
14	<i>donkie</i>	donkey	3
15	<i>hoender</i>	chicken	2
16	<i>drommel</i>	idiot	2
17	<i>flip</i>	flip < fuck	2
18	<i>blikskottel</i>	tin plate < <i>bliksem</i> (see above)	1
19	<i>blou dinges</i>	blue thingamajig	1
20	<i>dêrn</i>	damn	1
21	<i>vonk</i>	spark < fuck	1
22	<i>hek</i>	gate; heck < hell	1
23	<i>hemel</i>	heaven	1
24	<i>hoenders</i>	chickens	1
25	<i>otter</i>	otter	1
26	<i>poes</i>	pussy (person); vulva; interjection	1
27	<i>vak</i>	subject < fuck	1
28	<i>vloek</i>	curse (potentially < fuck)	1
29	<i>water</i>	water	1
30	<i>wetter</i>	yard scoundrel	1
Total			382

- (24) ... want hoe de **hel** gaan jy aan na so iets? (ModCor-2017)
 ‘... because how the hell do you go on after something like that?’
- (25) Hoe de **donder** kan dit veilig wees vir kinders ... (ModCor-2019)
 ‘How the heck [thunder] can it be safe for children’
- (26) ek kyk na sy foto en wonder- hoe de *****kry jy so iets reg? (ModCor-2017)
 ‘I look at his picture and wonder – how the ***** do you manage something like that?’
- (27) Met alle respek gesê, hoe de **duiwel** is jy 16 in graad 7. (ModCor-2019)
 ‘With all due respect, how the devil are you 16 in grade 7.’

- (28) *Hoe de **moer** redeneer jy?* (ModCor-2018)
 'How the hell [beat up] do you reason?'
- (29) *hoe de **fok** neem die sap iemand in diens wat op parool is ...* (ModCor-2018)
 'How the fuck does the SAP employ someone who is on parole ...'
- (30) ... *hoe de **joos** kry die regstelsel dit reg?????* (ModCor-2019)
 '... how the devil does the legal system manage that????'
- (31) *Hoe de **ongeluk** verduidelik jy dit dán?* (ModCor-2017)
 'How the heck [accident] do you explain it then?'

Like in English (compare Section 2, point c), the x part of the Afrikaans WHx construction mostly contains only a bare noun with only one example of an NP with an adjective as modifier, viz. *blou dinges* 'blue thingamajig'. As is the case for the English construction, several of the high frequency Afrikaans nouns are also remodelled by non-morphemic word-formation strategies,¹⁵ e.g., *f.* < *fok* 'fuck'; *m.* < *moer* 'beat up'; and *d.* < *donder* 'thunder'. We also see instances of phonological or graphemic modification of the noun in the x part, e.g., *vlok* 'flake' < *fok* 'fuck'; *vonk* 'spark' < *fok* 'fuck'; *vak* 'subject' < *fok* 'fuck'; etc.

Regarding the lexical diversity of the Ns, the Afrikaans construction is also like the English one: lexical items range from high-frequency items (e.g., *hel* 'hell'), to once-off creations (e.g., *water* 'water'; *vloek* 'curse'). We also do not see any Ns in the ModCor list that originate from the Greek or Latin stratum (see Section 2, point f). A number of nouns on the ModCor list are also, similar to English, used as all-purpose intensifiers and have largely lost the connections to their original semantic fields. These include *fok* 'fuck'; *heng* 'heck'; *bliksem* 'thunder'; *flip* 'flip'; *dêmn* 'damn'; and *poes* 'pussy' (see Section 2, points d and i).

Regarding the use of euphemisms, several interesting examples can be identified in the Afrikaans ModCor dataset. Like English, the word *duiwel* 'devil' has a euphemised form, namely *joos* 'devil' (N=14). A number of other nouns also have euphemistic variants, namely *hek* 'heck' for *hel* 'hell'; *vak* 'subject', *vlok* 'flake', *vloek* 'curse', *vonk* 'spark', and *flip* 'flip' for *fok* 'fuck'; and *donkie* 'donkey' and *drommel* 'rascal' for *donder* 'thunder' (also compare Section 2, points c and k).

Unlike the English construction, we also do not see any longer, ornate examples of noun phrases that might add an "archaic or humorous flavor" (Hugou 2017: 23)

¹⁵ While several of these remodelled examples occur in our data, only the normalised types are presented in Table 5.

to the Afrikaans construction. It is possible that these types of noun phrases are absent in the ModCor results because the humorous nature of these words might not have been regarded as impolite by the moderators, and were therefore not removed from the comments (see Section 2, point g).

According to Hugou (2017), most of the nouns in the English construction are monosyllabic, with some exceptions (such as *devil* and *God's name* as the most noticeable). The results for the Afrikaans construction differ in this regard. Only about half of the nouns on this list (i.e., 14 out of the 30 noun types) are monosyllabic (such as *hel* 'hell', *moer* 'beat up' and *joos* 'devil'), while 12 are disyllabic (such as *duiwel* 'devil', *bliksem* 'lightning' and *hoender* 'chicken'), and three are even polysyllabic (namely *ongeluk* 'accident', *blikskottel* 'tin plate' and *blou dinges* 'blue thing') (see again Section 2, point e).

With regards to semantic fields, Hugou (2017) determined that the English construction favoured words belonging to the category RELIGION (such as *hell*, *devil*, and *God's name*), and COSMOGONY (such as *earth* and *world*) (see Section 2, point h). The set of Afrikaans nouns in ModCor also include words belonging to the category of RELIGION, such as *hel* 'hell', *hemel* 'hemel', *duiwel* 'devil', and *jissis* 'Jesus'. However, no clear examples of nouns related to COSMOGONY have been attested. We also observe three additional semantic categories, namely ANIMALS (e.g., *otter* 'otter' and *donkie* 'donkey'), NATURAL PHENOMENA (e.g., *donder* 'thunder' and *bliksem* 'lightning'), and NEGATIVE EVENTS (e.g., *moer* 'beat up' and *ongeluk* 'accident'). Note that, as is the case in English, many of the words in the latter two categories are often used as interjections, as illustrated for *bliksem* 'lightning' in example (32), which implicates that the original semantic fields (NATURAL PHENOMENA and NEGATIVE EVENTS) are perhaps not part of the nouns' construal anymore – even more so when used in the construction.

(32) *O bliksem, waar val jy uit die bus uit?*

'O damn [lightning], where are you from?'

In his English dataset, Hugou (2017: 15) found that *hell* and its variants are used consistently more frequently than *fuck* and its variants (see Section 2, point j). He ascribes this to two potential factors: his data comes from written genres (novels, magazines, newspapers), and that there might be a "need to use more appropriate language in certain social situations." We see a similar phenomenon in the Afrikaans ModCor dataset, where *hel* 'hell' (N=182) is by far the most frequently used noun in the construction. However, even though *fok* 'fuck' is the most frequent Afrikaans swearword in most contexts (Van der Merwe 2022), *fok* 'fuck' is only the fourth most frequent noun on the ModCor list, and the frequency difference between *hel* 'hell' and the second most frequent noun, namely *donder* 'thunder'

(N=29), is substantial. We therefore see similar results to those of Hugou (2017), but the frequency difference between *hel* in Afrikaans and the rest of the nouns on the list is notable. We must keep in mind that our Afrikaans dataset and Hugou's dataset differ, as the ModCor dataset comprises unedited language, while Hugou's dataset has edited language. However, the reason for the large difference in frequency between *hel* 'hell' and the rest of the Afrikaans nouns on the list may be that contributors to the ModCor dataset were aware that their comments might be removed due to impolite noun choices, and might therefore have opted for the more appropriate, less offensive *hel* 'hell', rather than the inappropriate, offensive *fok* 'fuck'.

Lastly, based on Hugou's (2017) interpretation of Bybee's (2010) definition and operationalisation of productivity (see Section 2, point l), we can also conclude that the WHX construction in Afrikaans is productive. About half of the nouns on the ModCor list are hapax legomena, which serves as an indication of the construction's high hapax-conditioned degree of productivity (see again Baayen 1993). Nonetheless, even though the Afrikaans nouns shows greater variation in terms of semantic categories, the construction in Afrikaans, similar to English, resists full productivity – compare Section 2, points m and l.

6.2 Verification and extension (ComCor)

To verify and extend our list of possible noun phrases, we again compare it with data from ComCor. Table 6 presents the frequencies of noun phrases found in the x part of the WHX construction in ComCor. In total, 2,031 tokens (of 72 different types) were found, with more than 40 new lexical items that can be used in the x part of the WHX construction.

Table 6: Frequency of x in the WHX construction in ComCor.

	x	Translation	N
1	<i>hel</i>	hell	801
2	<i>fok</i>	fuck	381
3	<i>duiwel</i>	devil	151
4	<i>joos</i>	devil	93
5	[GRAWLIX]		87
6	<i>moer</i>	coffee sediment; beat up; interjection	86
7	<i>donder</i>	thunder; interjection	75
8	<i>ongeluk</i>	accident	60
9	<i>dinges</i>	thingamajig	39
10	<i>poes</i>	pussy (person); vulva; interjection	39
11	<i>heng</i>	heck	37

Table 6 (continued)

	x	Translation	N
12	<i>hoenders</i>	chickens	26
13	<i>drommel</i>	idiot	20
14	<i>swernoot</i>	scoundrel	16
15	[EMPTY]		8
16	<i>donkie</i>	donkey	8
17	<i>hoender</i>	chicken	8
18	<i>bliksem</i>	lightning; interjection	7
19	<i>hek</i>	gate; heck < hell	6
20	<i>vlok</i>	flake < fuck	6
21	<i>wetter</i>	yard scoundrel	6
22	<i>flip</i>	flip < fuck	5
23	<i>vy</i>	fig < fuck	4
24	<i>otter</i>	otter	3
25	<i>dēm</i>	damn	2
26	<i>flippen hel</i>	flipping hell	2
27	<i>heck</i>	heck < hell	2
28	<i>hemel</i>	heaven	2
29	<i>hoppende fok</i>	hopping fuck	2
30	<i>jissis</i>	Jesus	2
31	<i>Moses</i>	Moses	2
32	<i>vet</i>	fat < fuck	2
33	<i>vloek</i>	curse (potentially < fuck)	2
34	<i>vrek</i>	die (like an animal) < fuck	2
35	<i>zir</i>	zir	2
36	<i>actual hel</i>	actual hell	1
37	<i>bliksemstraal</i>	lightning bolt	1
38	<i>blikskottel</i>	tin plate < <i>bliksem</i> (see above)	1
39	<i>clue</i>	clue	1
40	<i>dammit</i>	damn it	1
41	<i>deksels</i>	lids < <i>donder</i> (see above)	1
42	<i>diekens</i>	Dickens < devil	1
43	<i>dooie dinges</i>	dead thingamajig	1
44	<i>falala</i>	falala (nonsensical, perhaps < fuck)	1
45	<i>fokken hel</i>	fucking hell	1
46	<i>fokietie fok</i>	fuckity fuck fuck	1
	<i>fok</i>		
47	<i>frieken hel</i>	freaking hell	1
48	<i>grote griet</i>	great Griet < <i>grote God</i> 'great God'	1
49	<i>heilige fok</i>	holy fuck	1
50	<i>hel op aarde</i>	hell on earth	1
51	<i>herrie</i>	Harry < hell	1
52	<i>hoeders</i>	keepers (potentially < <i>hoenders</i>)	1

Table 6 (continued)

	x	Translation	N
53	<i>iets</i>	something	1
54	<i>kees</i>	monkey	1
55	<i>kont</i>	cunt	1
56	<i>kriewel</i>	squirm	1
57	<i>liewe goeie</i> <i>heilige fok</i>	dear good holy fuck	1
58	<i>moeder</i>	mother < <i>moer</i> (see above)	1
59	<i>oertel</i>	asshole	1
60	<i>os</i>	ox	1
61	<i>pienk</i>	pink < pussy	1
62	<i>poefies</i>	shits	1
63	<i>pok</i>	small pox < fuck	1
64	<i>te hel</i>	too hell	1
65	<i>vak</i>	subject < fuck	1
66	<i>vakbond</i>	trade union < fuck	1
67	<i>vonk</i>	spark < fuck	1
68	<i>vrommel</i>	crumple	1
69	<i>vrug</i>	fruit (potentially < fuck)	1
70	<i>water</i>	water < <i>wetter</i> (see above)	1
71	<i>wharra</i>	Wharra (nonsensical)	1
72	<i>Yaris</i>	Yaris	1
Total			2,031

The results from the ComCor *verify* that Afrikaans nouns are modified phonologically and graphemically (such as *vlok* ‘flake’ and *fonk* ‘spark’ for *fok* ‘fuck’), as well as being remodelled by non-morphemic word-formation strategies (such as *d@nner* for *donder* ‘thunder’, and *m..r* for *moer* ‘beat up’). We also do not see any words on the ComCor list that originate from the Greek or Latin stratum. Furthermore, in the ModCor dataset, we noted that, unlike English, the Afrikaans construction does not have such a strong preference for monosyllabic nouns, as several of the Afrikaans nouns in the ModCor dataset were disyllabic or even polysyllabic. This observation is even more relevant to the ComCor dataset, wherein 29 of the nouns (phrases) occurring in the x part of the construction are disyllabic and 15 are polysyllabic. In other words, most of the nouns in the list are actually non-monosyllabic words, and the preference for these types of nouns is apparently not a feature of the Afrikaans construction. Lastly, based on the number of hapax legomena in the ComCor list (37 types out of the total 72 types), we can confirm that, in terms of the operationalised definition of productivity mentioned in 2m above, this construction could be considered productive in Afrikaans.

Additionally, we can *expand* our initial description of the construction. While only one case of a noun modified by an adjective was found in ModCor (namely *blou dinges* ‘blue thing’), numerous examples can be found in ComCor. Compare for instance *flippen hel* ‘flipping hell’; *hoppende fok* ‘hopping fuck’; *actual hel* ‘actual hell’; *dooie dinges* ‘dead thingamajig’; *fokken hel* ‘fucking hell’; *grote griet* ‘Great Griet’; *heilige fok* ‘holy fuck’; and *lieuwe goeie heilige fok* ‘dear good holy fuck’. Furthermore, in addition to the same high-frequency nouns (such as *hel* ‘hell’), a number of additional, on-the-spur-of-the-moment creations have been attested, such as *yaris* ‘Yaris car’; *falala* ‘falala’ (nonsensical); *vakbond* ‘trade union’; and *kriewel* ‘squirm’.

In addition to the nouns belonging to the category RELIGION that we have already identified in ModCor, we can now add several proper names and/or (euphemistic) modifications of such religious proper names, e.g., *Moses*; *jissis* ‘Jesus’; *grote Griet* ‘great Griet’ < ‘good God’; *diekens* ‘Dickens’ < ‘devil’; *Joos/Josie* < ‘devil’; and *Yaris* < (probably) ‘Jesus’. We also see numerous intensifiers of specifically *hel*, such as *fokken hel* ‘fucking hel’; *frieken hel* ‘freaking hell’; *flippen hel* ‘flipping hell’. While we did not find any noun phrases belonging to the category of COSMOGONY in ModCor, we found one example in ComCor, namely *hel op aarde* ‘hell on earth’. Two new nouns belonging to the category of ANIMALS could also be added, namely *os* ‘ox’ and *kees* ‘baboon’, and one new noun belonging to the category of NATURAL PHENOMENA, namely *bliksemstraal* ‘thunderstrike’. Additionally, we see one new noun in the category of NEGATIVE EVENTS, namely *vrek* ‘die like an animal’. On the other hand, it could also possibly simply be a modification of *fok* ‘fuck’, since we find numerous examples of leetspeak with valid words – compare for *fok* ‘fuck’ the following potential candidates: *vakbond* ‘trade union’; *vet* ‘fat’; *vrommel* ‘crumple’; *vrug* ‘fruit’; *vy* ‘fig’ (all starting with [f]); and *pok* ‘pox’ (rhyming with *fok*). For *hel* ‘hell’ we found the modifications *heck*, *heng*, and *herrie*; and for *duiwel* ‘devil’, among others, *deuvel*. While one might be tempted to add a new semantic category EXCREMENT for the noun *poefies* ‘shit’ (diminutive, plural), we consider it to rather fit under the larger umbrella of euphemistic words being used in written conversations in the public domain.

While we did not see any nouns in the ModCor list that could be regarded as adding an “archaic or humorous flavor” (Hugou 2017: 23) to the WHX construction, we do find in the ComCor dataset several noun phrases that might be construed as humorous, for example *hoppende fok* ‘hopping fuck’, *dooie dinges* ‘dead thingamajig’, *foketie fok fok* ‘fuckity fuck fuck’, *grote griet* ‘Great Griet’, and *lieuwe goeie heilige fok* ‘dear good holy fuck’.

7 Constructional relations between the WH part and x part

We now turn to our second research question, namely, to investigate the attraction between specific WHs and specific Xs in the Afrikaans WHX construction, using covarying collexeme analysis. Whereas the results in the previous sections could be compared easily to Hugou’s (2017) results for English, the results below cannot, since he did not investigate this aspect of the English construction.

Covarying collexeme analysis is a technique developed by Stefanowitsch and Gries (2003), and Gries and Stefanowitsch (2004a, 2004b) as part of the larger set of techniques called collostructional analysis. Collostructional analysis uses a statistical method to measure the strength of association within or between constructions (also see Gilquin 2010: 195), while covarying collexeme analysis is specifically used to measure the association strength between variables in a construction. Stefanowitsch and Gries (2005: 9) explain that this method is suitable for investigating constructions with “two (or more) slots which may be associated with sets of items whose semantic properties we want to investigate with respect to each other”.

To investigate the attraction between the WHs and the Xs in the Afrikaans WHX construction, we made use of a combined dataset (ComCor and ModCor). We used this dataset to analyse all individual instances of the 8 WH pronouns in the combined 2,414 occurrences of the construction, assessing all possible collocations with all 74 types of X items in this combined dataset. The analyses have been performed using the statistical software R, using the *Coll.analysis* 4.0 script (Gries 2022).

Table 7 presents the results from the covarying analysis of the combined ComCor and ModCor data. Only significant attractions (i.e., collocations with a likeliness ratio $LLR > 3.84$) are included in the table. We also did not include items where the noun phrase appears only once in the construction, and thus also only once in the combination, because it will obviously appear statistically significant.

Table 7: Results of covarying analysis of the combine ComCor and ModCor data.

WH_PN	Translation	RudeNP	Translation2	Freq	FREQOFSLOT1	FREQOFSLOT2	RELATION	LLR
hoe	how'	moer	beat up'	64	891	110	Attr.	21.411
hoe	how'	duiwel	devil'	94	891	176	Attr.	21.334
waar	where'	fok	fuck'	58	203	403	Attr.	19.623
wie	who'	fok	fuck'	78	313	403	Attr.	15.917
wat	what'	hel	hell'	437	965	983	Attr.	13.737
hoe	how'	hoenders	chickens'	19	891	27	Attr.	12.564
hoe	how'	ongeluk	accident'	41	891	72	Attr.	12.230
hoekom	why'	fok	fuck'	13	31	403	Attr.	11.131
waar	where'	flippen_hel	flipping hell'	2	203	2	Attr.	9.920
wie	who'	blikskottel	tin plate'	2	313	2	Attr.	8.181
hoe	how'	hoender	chicken'	8	891	10	Attr.	7.809
waarom	why'	donder	thunder'	2	4	104	Attr.	7.242
hoe	how'	dinges	thingamajig'	24	891	44	Attr.	5.732
waar	where'	poes	pussy'	8	203	40	Attr.	5.317
waar	where'	grote_griet	great Griet'	1	203	1	Attr.	4.955
waar	where'	os	ox'	1	203	1	Attr.	4.955
waar	where'	vrommel	crumple'	1	203	1	Attr.	4.955
wanneer	when'	fok	fuck'	3	5	403	Attr.	4.753
wie	who'	dên	damn'	2	313	3	Attr.	4.638
wie	who'	hemel	heaven'	2	313	3	Attr.	4.638
wie	who'	hel	hell'	145	313	983	Attr.	4.605
wanneer	when'	duiwel	devil'	2	5	176	Attr.	4.213
wie	who'	doolie_dinges	dead thingamajig'	1	313	1	Attr.	4.088
wie	who'	kont	cunt'	1	313	1	Attr.	4.088
hoe	how'	heck	heck'	2	891	2	Attr.	3.988
hoe	how'	Moses	Moss'	2	891	2	Attr.	3.988

<i>hoe</i>	how'	<i>water</i>	water'	2	891	2	Attr.	3.988
<i>hoe</i>	how'	<i>zir</i>	zir'	2	891	2	Attr.	3.988
<i>hoe</i>	how'	<i>poes</i>	pussy'	9	891	40	Rep.	-3.913
<i>wie</i>	who'	<i>ongeluk</i>	accident'	4	313	72	Rep.	-4.445
<i>wanneer</i>	when'	<i>hel</i>	hell'	0	5	983	Rep.	-5.239
<i>wat</i>	what'	<i>swernoot</i>	scoundrel'	2	965	16	Rep.	-5.942
<i>wie</i>	who'	<i>duivel</i>	devil'	13	313	176	Rep.	-6.015
<i>wie</i>	who'	<i>drommel</i>	idiot'	0	313	22	Rep.	-6.143
<i>wie</i>	who'	<i>moer</i>	beat up'	6	313	110	Rep.	-7.097
<i>wat</i>	what'	<i>hoenders</i>	chickens'	4	965	27	Rep.	-8.252
<i>hoe</i>	how'	<i>vlok</i>	flake'	0	891	10	Rep.	-9.241
<i>wat</i>	what'	<i>moer</i>	beat up'	29	965	110	Rep.	-9.392
<i>wat</i>	what'	<i>duivel</i>	devil'	51	965	176	Rep.	-9.967
<i>waar</i>	where'	<i>joos</i>	devil'	1	203	106	Rep.	-12.436
<i>hoe</i>	how'	<i>hel</i>	hell'	321	891	983	Rep.	-13.067
<i>hoekom</i>	why'	<i>hel</i>	hell'	3	31	983	Rep.	-15.136
<i>hoe</i>	how'	<i>fok</i>	fuck'	89	891	403	Rep.	-48.764

The results clearly indicate that some WH pronouns and specific nouns attract each other within the WHX construction. For example, there is a very strong attraction between *hoe* 'how' and *moer* 'beat up' (LLR=21.411), stronger than any other WH pronoun with any other noun in this construction. It is also clear that we cannot assume that all highly frequent noun phrases (such as *hel* 'hell', *fok* 'fuck', and *duiwel* 'devil') show a strong collocational preference for all pronouns that occur in the construction. For example, one might assume that *hel* 'hell' (N=983) would show a strong attraction to all or most WH pronouns in this construction, but there is in fact only an attraction relationship between *wat* 'what' (LLR=13.737) and *hel* 'hell', and *wie* 'who' and *hel* 'hell' (LLR=4.605). While combinations with other WH pronouns do occur and are used frequently, the attraction is never statistically significant. In fact, there is even a strong collocational repulsion between *hel* 'hell' and *hoe* 'how' (LLR=-13.067) and *wanneer* 'when' (LLR=-5.239). The strong repulsion between *hel* and *hoe* is despite the high frequency of this collocation (N=321). From this result, we see again that the covarying collexeme analysis is not suitable to identify prototypical and entrenched constructions.

Similarly, we observe that there is a strong attraction between *fok* 'fuck' and *waar* 'where' (LLR=19.623) and *wie* 'who' (LLR=15.917), and a weak (but still significant) attraction with *wanneer* 'when' (LLR=4.753). However, *fok* 'fuck' shows an extremely strong repulsion with *hoe* 'how' (LLR=-48.764, the strongest collocational value in our results). *Duiwel* 'devil' and *hoe* 'how' (LLR=21.334), on the other hand, are strongly attracted, while *duiwel* 'devil' and *wie* 'who' (LLR=-6.015) and *wat* 'what' (LLR=-9.967) are in a repulsion relationship. *Joos* 'devil', the fourth most frequent noun in our ComCor dataset, does not show any significant attractions, except for a repulsion with *waar* 'where' (LLR=-12.436).

Another interesting observation is that there are strong attractions between *hoe* 'how' and *moer* 'beat up' (LLR=21.411), *hoenders* 'chickens' (LLR=12.564), and *hoender* 'chicken' (LLR=7.809). This could be due the assonance effect created by combining these words in the construction, contributing to a humorous effect. The fact that the *wat* 'what' and the nouns *moer* 'beat up' and *hoenders* 'chicken' show a repulsive relation (LLR=-9.392 and LLR=-8.252 respectively) strengthens our assertion.

As noted, the aim of the covarying collexeme analysis is primarily to enable the identification of semantic patterns in the way items in the two slots collocate with each other. Although we were able to identify strong attractions and repulsions for various collocations, we do not see any clear semantic patterns emerging from our results. However, this is not entirely surprising, given the nature of the two items that can fill these slots. The WH pronouns are function words that carry little semantic weight (serving mainly as an interrogative to request specific information), while the x items function to convey taboo value or humour. The con-

struction's function is ultimately to attract attention rather than to convey semantic content, and it may thus be better considered a pragmatic construction.

8 Conclusion

Given the fact that the Afrikaans WHX construction had not been previously described in the literature, we set out to provide such a description to get supporting evidence for our assertion that the Afrikaans construction is a constructional calque, based on the English WHX construction (Hugou 2017), but it is starting to take on a life of its own. By conducting a corpus linguistic investigation of the construction in two Afrikaans corpora, we focused specifically on the characteristics of the various open slots in the construction, and the constructional relations (attractions or repulsions) between specific WH pronouns and noun phrases. Throughout, we contrasted our findings with those of Hugou (2017) for English. Our investigation finds the following:

- a. Only eight Afrikaans WH pronouns occur in the construction, viz. *wat* 'what', *hoe* 'how', *wie* 'who', *waar* 'where', *hoekom* 'why', *wanneer* 'when', *waarom* 'why' and *waarvoor* 'what for'.
- b. In both corpora the four monosyllabic WH pronouns occur with a much higher frequency than the disyllabic ones, with *wat* 'what', followed by *hoe* 'how' consistently occurring with the highest frequency. The fact that the Afrikaans construction clearly favours monosyllabic WH pronouns and the fact that *wat* 'what' occurs in this part most frequently, correlates well with what Hugou (2017) reported for the English construction. However, unlike in English, disyllabic and even polysyllabic items are often found as noun phrases. This indicates that Afrikaans speakers are elaborating on the construction, proving that the construction is taking on a life of its own.
- c. One major difference between the Afrikaans and English construction is the frequency difference between *wat* 'what' and *hoe* 'how': while Hugou (2017) reports a very large difference between the frequency of *what* and *how*, in Afrikaans there is no mentionable difference between the frequency of the two WH pronouns. Nonetheless, distributional matters aside, this illustrates the existence of a diacxn schema clearly: the English idiocxn *what* $[[wɒt] \Leftrightarrow [\text{WHAT}]]_{\text{ENG}}$ and the Afrikaans idiocxn *wat* $[[vat] \Leftrightarrow [\text{THE}]]_{\text{AFR}}$ results in the diacxn $[[w/v \dots t] \Leftrightarrow [\text{WHAT}]]$ (and so forth for the other pronouns).
- d. We identified 76 unique noun phrases that can occur in the x part of the Afrikaans construction. In addition, *gawlix* was also used in this part often, and in

some instances the x part was left empty. Identical to English, *hel* ‘hell’ consistently occurred in this part of the construction with the highest frequency. Five other lexical items make up the 6 most frequently occurring nouns in both our corpora, viz. *donder* ‘thunder’; *duiwel* ‘devil’; *fok* ‘fuck’; *joos* ‘devil’; and *moer* ‘beat up’. Apart from high-frequency items, in-the-spur-of-the-moment nouns (e.g. *vakbond* ‘trade union’ and *Yaris*) also occur, as is the case in the English construction. The high number of hapax legomena that occur in the x part of the Afrikaans construction indicates that, just like the English construction, this construction could be considered productive in Afrikaans (specifically in unedited, informal, written contexts).

- e. While bare noun phrases occur with the highest frequency in the Afrikaans construction, nouns can also be modified, e.g. *hoppende fok* ‘hopping fuck’.
- f. Noun phrases in the Afrikaans construction can also contain lexical items that are remodelled by non-morphemic word-formation strategies, e.g. *.f.* < *fok* ‘fuck’, and *MMMMMMMMMMMM* < *moer* ‘beat up’. Instances of phonological and graphemic modification were also attested quite frequently, e.g., *vlok* ‘flake’ < *fok* ‘fuck’, and *pok* ‘pox’ < *fok* ‘fuck’. This is most probably the result of the context of our data, where users might resort to leetspeak to by-pass moderators removing their comments.
- g. Hugou (2017) identified only two semantic categories, viz. RELIGION and COSMOGONY, for the x part of the English construction. Our analysis of the Afrikaans counterpart reveals a broader range of categories, including RELIGION, ANIMALS, NATURAL PHENOMENA, and NEGATIVE EVENTS, but not COSMOGONY (bar one disputable example). This testifies to the elaboration of the construction schema in Afrikaans.
- h. Lastly, while Hugou (2017) posits that the English construction primarily functions as an intensifier, our findings suggest an additional function for the Afrikaans construction, namely as a vehicle for humour. Examples such as *hoppende fok* ‘hopping fuck’, *fokietie fok fok* ‘fuckity fuck fuck’, *zir, falala, Yaris, grote griet* ‘great Griet’, *hoenders* ‘chickens’, *otter* ‘otter’, *oertel* ‘asshole’, and *vakbond* ‘trade union’ indicate a strong tendency towards comedic expression. This highlights an important function of the Afrikaans construction – a feature that was not explicitly identified by Hugou (2017), given his specific focus on intensification. This therefore does not mean that this is a unique feature of the Afrikaans construction, because humorous examples of the English construction abound, e.g. *what the flying blue fuck* (an example taken from Hugou 2017: 16).

The large number of telling similarities between the Afrikaans and English construction confirm our thesis that the Afrikaans construction is a constructional calque, based on the English WHX construction. However, while the construction might be borrowed, it is already developing into new directions, taking on a life of its own. This is specifically noticeable in the occurrence of disyllabic and even polysyllabic noun phrases in the Afrikaans construction, a broader range of semantic categories of the noun phrases, and the emergence of humour as an important function of the construction. However, we should note that we suspect that one might also see similar effects in comparable English genres, and in the WHX construction in other languages – something that should be examined in future research.

Lastly, as native speakers of Afrikaans, we know that the WHX construction is sometimes univerbated, e.g., *watdefok* < *wat de fok*. In addition, such univerbations could also be phonologically and/or orthographically deformed, e.g., *warrefok* < *waddefok* < *watdefok* < *wat de fok*, where *waddefok* is a progressively assimilated form of *watdefok*, and *warrefok* is a rhotacized form of *waddefok* (see Wissing 2020). Although such changes (univerbations and deformations) fall outside the scope of our current research (and therefore our original searches in the corpora), we did an exploratory investigation of changes to *wat de hel* and *wat de fok* only (i.e., the most frequent WH pronoun, with the two most frequent nouns). The results in Table 8 clearly show a strong tendency to change those constructions with *fok* in the x part of the construction – a somewhat surprising result, given that *hel* occurs much more frequently in the WHX construction (see Table 6). We surmise, on the one hand, that since *fok* is generally considered to be much more taboo than *hell*, users might be trying to euphemise the construction with *fok*, or otherwise simply to trick profanity checkers (i.e., univerbation as a kind of leetspeak). It could, on the other hand, also be an indication of lexicalisation in the making: the phrasal *wat de fok* is slowly becoming a lexical item *waddefok*, which might in the future develop a life of its own. Further research into other such forms, e.g., *hoe-de-fok* 'how-the-fuck', could potentially shed light on the development of new taboo words.

Table 8: Frequency of contracted forms in the WHX construction in ComCor.

Rank	Contracted form with <i>wat</i> and <i>fok</i>	N	Contracted form with <i>wat</i> and <i>hel</i>	N
1	<i>waddefok</i>	82	<i>warrehel</i>	6
2	<i>watte fok</i>	8	<i>waddehel</i>	5
3	<i>watdefok</i>	7	<i>wate hel</i>	1
4	<i>warrefok</i>	5	<i>wattehel</i>	1
5	<i>warre fok</i>	4		
6	<i>wadde fok</i>	3		

Table 8 (continued)

Rank	Contracted form with <i>wat</i> and <i>fok</i>	N	Contracted form with <i>wat</i> and <i>hel</i>	N
7	<i>watefok</i>	2		
8	<i>wattefok</i>	2		
9	<i>waddiefok</i>	1		
10	<i>watde fok</i>	1		
11	<i>wate fok</i>	1		
Total		116	Total	13

Ethical matters

Ethical clearance for the research project was obtained through the Language Matters Ethics Committee of the North-West University (ethics number: NWU-00632-19-A7).

Declaration of interests

In the table below, the roles of the authors are defined in terms of the Contributor Roles Taxonomy (CRediT; <https://credit.niso.org/>).

Role	Van Huyssteen	Breed	Pilon
<i>Conceptualisation</i>	X	X	X
<i>Data curation</i>	X	X	X
<i>Formal analysis</i>	X	X	X
<i>Funding acquisition</i>	–	X	–
<i>Investigation</i>	X	X	X
<i>Methodology</i>	X	X	X
<i>Project administration</i>	X	–	–
<i>Resources</i>	X	X	X
<i>Software</i>	–	–	–
<i>Supervision</i>	–	–	–
<i>Validation</i>	X	X	X
<i>Visualization</i>	–	–	–
<i>Writing (original draft)</i>	X (§ 1–3, 8)	X (§ 4–7)	X (§ 4–7)
<i>Writing (review and editing)</i>	X	X	X

This research is partially funded by the Suid-Afrikaans Akademie vir Wetenskap en Kuns, and partially made possible through barter agreements with BlueTek Computers, Afrikaans.com, and WatKykJy.co.za. The *Woordeboek van die Afrikaanse Taal* (WAT), *Handwoordeboek van die Afrikaanse Taal* (HAT) and Centre for Text Technology (CTeXT) of the North-West University are hereby also acknowledged for generously supplying the project with material from their respective databases.

Van Huyssteen is a director of the not-for-profit company Viridevert NPC (CIPC registration number: 2016/411799/08), who owns and manages the website vloek.co.za. This website was developed specifically for this project, and this conflict of interest has been approved by NWU.

References

- Allan, Keith & Kate Burridge. 1991. *Euphemism and dysphemism: Language used as shield and weapon*. New York & Oxford: Oxford University Press.
- Baayen, Harald. 1993. On frequency, transparency and productivity. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1992*, 181–208. Dordrecht: Springer Netherlands.
- Burridge, Kathryn & Reka Agnes Benczes. 2019. Taboo as a driver of language change. In Keith Allan (ed.), *The Oxford handbook of taboo words and language*, 180–199. Oxford: Oxford University Press.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Centre for Text Technology (CTeXT). 2020. NWU Korpus Gemodereerde Kommentaar 1.0 [NWU Corpus of Moderated Comments 1.0]. In North-West University (ed.) 1.0 edn. Potchefstroom: Queried on Sketch Engine: app.sketchengine.eu.
- Centre for Text Technology (CTeXT). 2023. NWU Kommentaarkorpus 2.2 [NWU Commentary Corpus 2.2]. In North-West University (ed.) 2.2 edn. Potchefstroom: Queried on Sketch Engine: app.sketchengine.eu.
- Culpeper, Jonathan. 2011. *Impoliteness: Using language to cause offence*. Cambridge: Cambridge University Press.
- Davies, Mark. 2019. The Movie Corpus. <https://www.english-corpora.org/movies/>.
- Davies, Mark. n.d. Corpus of Contemporary American English (COCA). <https://www.english-corpora.org/coca/>.
- Gilquin, Gaëtan. 2010. *Corpus, cognition and causative constructions*. Amsterdam: Benjamins.
- Gries, Stefan Th. 2022. Coll.analysis 4.0. A script for R to compute perform collostructional analyses. <https://www.stgries.info/teaching/groningen/index.html>.
- Gries, Stefan Th. & Nick C. Ellis. 2015. Statistical measures for usage-nased linguistics. *Language Learning* 65 (S1). 228–255.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004a. Co-varying collexemes in the into-causative. In Michel Achard & Suzanne Kemmer (eds.), *Language, culture, and mind*, 225–236. Stanford: CSLI.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004b. Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9(1). 97–129.

- Hickey, Leo & Miranda Stewart. 2005. Introduction. In Leo Hickey & Miranda Stewart (eds.), *Politeness in Europe*, 1–12. Clevedon: Multilingual Matters.
- Höder, Steffen. 2018. Grammar is community-specific: Background and basic concepts of Diasystematic Construction Grammar. In Hans C. Boas & Steffen Höder (eds.), *Constructions in contact: Constructional perspectives on contact phenomena in Germanic languages*, 37–70. Amsterdam: Benjamins.
- Höder, Steffen. 2019. Phonological schematicity in multilingual constructions: A diasystematic perspective on lexical form. *Word Structure* 12(3). 334–352.
- Höder, Steffen, Julia Prentice & Sofia Tingsell. 2021. Additional language acquisition as emerging multilingualism: A Construction Grammar approach. In Hans C. Boas & Steffen Höder (eds.), *Constructions in contact: Language change, multilingual practices, and additional language acquisition*, 309–338. Amsterdam: Benjamins.
- Hoeksema, Jack. 2019. Taboo terms and their grammar. In Keith Allan (ed.), *The Oxford handbook of taboo words and language*, 160–179. Oxford: Oxford University Press.
- Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Hugou, Vincent. 2017. The WHX construction (what the hell. . .?) and intensity. A corpus-based study. *Lexis* (10).
- Jay, Timothy B. 2020. Ten issues facing taboo word scholars. In Nico Nassenstein & Anne Storch (eds.), *Swearing and cursing*, 37–52. Berlin: De Gruyter Mouton.
- Jay, Timothy B. & Kristin Janschewitz. 2008. The pragmatics of swearing. *Journal of Politeness Research* 4 (2). 267–288.
- Kádár, Dániel Z. 2017. *Politeness, impoliteness and ritual: Maintaining the moral order in interpersonal interaction*. Cambridge: Cambridge University Press.
- Kirsten, Johanita & Adri Breed. 2020. Interrogative pronouns. *Taalportaal*. Taalportaal Consortium: <https://taalportaal.org/taalportaal/topic/pid/topic-20190516105546177>. Last Modified May 2020. Accessed 20 December 2024.
- Langacker, Ronald W. 1991. *Concept, image, and symbol: The cognitive basis of grammar*. Berlin: Mouton de Gruyter.
- Michaelis, Susanne Maria & Martin Haspelmath. 2020. Grammaticalization in creole languages: Accelerated functionalization and semantic imitation. In Walter Bisang & Andrej Malchukov (eds.), *Volume 2 Grammaticalization scenarios from Africa, the Americas, and the Pacific*, 1109–1128. Berlin: De Gruyter Mouton.
- Sketch Engine. 2020. Dutch Web corpus 2020 (nlTenTen20). <https://www.sketchengine.eu/nltenten-dutch-corpus/>.
- Spencer-Oatey, Helen. 2005. (Im)politeness, face and perceptions of rapport: Unpackaging their bases and interrelationships. *Journal of Politeness Research* 1(1). 95–119.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1(1). 1–43.
- Taalkommissie van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns. 2011. Taalkommissiekorpus 1.1. Potchefstroom: North-West University.
- Taalprof. 2009. Wat voor den duivel is dat nou weer? [What the devil is that again?]. *Taalprof Plein*. <https://taalprof.blogspot.com/2009/07/wat-voor-den-duivel-is-dat-nou-weer.html>.
- Taylor, John R. 2002. *Cognitive grammar*. Oxford: Oxford University Press.

- Thompson, Sandra A., Barbara A. Fox & Elizabeth Couper-Kuhlen. 2015. *Grammar in everyday talk: Building responsive actions*. Cambridge: Cambridge University Press.
- Van der Merwe, Mart-Mari. 2022. *Prototipiese Afrikaanse taboewoorde [Prototypical Afrikaans taboo words]*. Pretoria: University of Pretoria.
- Van der Wouden, Ton. 2019. Hoe de fuk. *Neerlandistiek: Online Tijdschrift voor Taal- en Letterkunde*. <https://neerlandistiek.nl/2019/11/hoe-de-fuk/>.
- 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.
- Van Rooy, Bertus. 2021. A Diasystematic Construction Grammar analysis of language change in the Afrikaans and English finite verb complement clause construction. In Hans Boas & Steffen Höder (eds.), *Constructions in contact 2: Language change, multilingual practices, and additional language acquisition*, 109–137. Amsterdam: Benjamins.
- Van Rooy, Bertus & Ronel Wasserman. to appear. English of Afrikaans speakers. In Raymond Hickey (ed.), *New Cambridge history of the English language*. Cambridge: Cambridge University Press.
- Wissing, Daan. 2020. Consonant related processes. *Taalportaal*. <https://taalportaal.org/taalportaal/topic/pid/topic-20170830162004808>. Last Modified May 2020. Accessed 20 December 2024.

8 Such an impoliteness: Evidence for the ‘evaluative *such* construction’

Abstract: There has been much debate regarding whether impoliteness can be inherently associated with particular linguistic structures. Adopting a usage-based, interactionalist approach to impoliteness, we conducted four questionnaire studies to find evidence for structurally embedded impoliteness. Specifically, we investigate the evaluative *such* construction [*such* + (article) + nominal]. Our studies examine the construction embedded in the form [PRON BE *such* ART N] in two closely related languages, English and German, using the 2nd and 3rd person singular pronouns, respectively. Our results confirm that predicative statements of the form [PRON BE (*such*) ART N] are more likely to be rated as evaluative, predominantly as negative, when they contain *such/so*. This effect is stronger in both German studies than in the English data. We also show that evaluations differ across various nouns used in the nominal slot, and that the construction has the power to even switch the interpretation of the (otherwise) same utterance from positive to negative meaning for some nouns. Chi-squared tests show statistically significant associations between the presence and absence of *such/so* and experimental ratings. We conclude that non-evaluative nouns in the nominal slot tend to be coerced into an evaluative reading. We identify the item *such/so* as a major contributor to this coercion effect.

Keywords: impoliteness, construction, conventionalized impoliteness formulae, coercion, evaluation study

1 Introduction

When encountering the string *You are such a . . .*, we have good reason to believe the completed utterance will be an insult. This intuition is based on our prior knowledge of similar utterances, specifically of the form [PRON BE *such* ART N]. The form is best attested with negatively connoted nouns (see Section 2.2.2). Compare similar sentences like *You’re such a jerk*, *I’m such an idiot*, and *He became such an asshole*.

Angela Queisser, Heidelberg University, Interdisciplinary Center for Scientific Computing, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany, e-mail: angela.queisser@iwr.uni-heidelberg.de.

Monika Pleyer, Heidelberg University, English Department, Kettengasse 12, 69117 Heidelberg, Germany, e-mail: monika.pleyer@as.uni-heidelberg.de

Examples with *you* as the subject have been described by Culpeper (2011: 133, 135) as a ‘personalized negative assertion’, a type of ‘conventionalized impoliteness formula’ (see also Culpeper et al, this volume). This means it is used for explicit orientation to a target and negative evaluation thereof.

We adopt here a perspective on impoliteness informed by usage-based and interactionalist approaches (Bousfield 2008; Culpeper 2011). We view the proposed formula as a ‘construction’, a form-meaning pair as described in Construction Grammar (e.g. Goldberg 1995). We extend the scope of said construction by narrowing its form to [*such* + (article) + nominal], thereby allowing it to occur with different subjects, (copula) verbs, and tenses, as illustrated by the three example sentences above. We refer to the construction as the ‘evaluative *such* construction’ and propose that a non-evaluative noun in the nominal slot is coerced into an evaluative, predominantly negative, reading. Consequently, the utterance is read as making a negative statement about the subject, i.e. it is open to an interpretation as impolite. We identify the lexically fixed item *such* as a major contributor to this coercion effect.

To compile evidence for our hypotheses, we draw on previous studies of related phenomena as well as corpus evidence, and conduct our own studies. In a corpus study, Van Olmen, Andersson, and Culpeper (2023) found corroborating evidence for the similar construction [*you* + noun phrase]. For instance, *you theoretician* is attested with an insulting meaning. Our study complements the work by Van Olmen and colleagues, but also departs from it. Their aim was to attest the existence of their construction in corpora; we use our previously compiled corpus evidence¹ as the foundation for four questionnaire studies (but see also Van Olmen and Andersson, this volume, for their questionnaire-based approach). Specifically, we are interested in coercion effects attributable to *such* in otherwise non-evaluative nouns attested in English and German. The German equivalent of the construction proposed above, [*so* (ART) Nom], is attested in German corpora (Hirschmann 2024: 199). Hirschmann (2024: 202–203) notes that an intensifying use as in *so ein Idiot* ‘such an idiot’ is comparatively rare. He also lists usages with neutral nouns, e.g. *Hans ist so ein Student* ‘Hans is such a student’; however, these are described not as evaluative, but as the subject having the properties of the noun to a high degree (compare Section 2.2.1). In investigating evaluative usages of the construction in the German data, our paper helps close a research gap.

We aim to show that conventionalization in impoliteness structures may be stable across two closely related languages. Our research questions are:

¹ A substantial part of the current contribution is based on our first author’s Master’s thesis (Queisser 2024), available in the Heidelberg University document repository.

- Do speakers evaluate otherwise identical statements conforming to the described pattern differently in the presence and absence of *such*?
- Do evaluations differ across various neutral nouns used in the nominal slot?
- Of which polarity (positive/negative) are the evaluations, and is there a discernible pattern?

In Section 2 of this chapter, we present the theoretical framework that serves as the foundation for our studies. We begin by briefly discussing previous findings on the conventionalization of impoliteness (2.1) before turning to constructions and commenting on the role of *such* (2.2.1) and our current construction (2.2.2). Section 3 presents our method; we present the stimuli used in the questionnaires (3.1) as well as respondent choice (3.2). Section 4 discusses the results for the two English studies (4.1) and the complementary German studies (4.2), followed by the description of the statistics (4.3) as well as a summary and general discussion (4.4). This is followed by a brief conclusion and suggestions for future research (5).

2 Theoretical framework

This section discusses the theoretical background for our studies presented in Sections 3 and 4 below. We employ a combined approach to situate our studies, drawing on research from sociological and interactional approaches to im/politeness studies (e.g. Terkourafi 2005; Culpeper 2010, 2011) and Construction Grammar (e.g. Goldberg 1995).

2.1 Conventionalization of impoliteness

Impoliteness² has been described as a term with fuzzy boundaries (Kleinke and Bös 2015), which has drawn forth a plethora of definitions. We follow Culpeper’s (2011: 23) definition, which situates impoliteness in interpersonal interactions. We wish to highlight two key aspects of Culpeper’s definition: first, impoliteness as an intentional negative evaluation of the hearer, and second, its connection to (social) norms. Our proposed construction [*such* (ART) Nom] can be classified as the crucial

² We use impoliteness as a technical term (rather than rudeness) as it mirrors the linguistic notion of politeness (see Culpeper 2011).

part of an insult. In insults, the speaker addresses the hearer by an epithet,³ or states characteristics of the hearer in such a way that the hearer may perceive them as inappropriate and intentional (Jucker and Taavitsainen 2000: 73). That is, insults as evaluative forms threaten the hearer's face.

There has been much debate regarding whether impoliteness can be inherently associated with particular linguistic structures (see Van Olmen et al., this volume). One position holds that impoliteness is inherent in particular linguistic expressions; see first-wave accounts, e.g. by Leech (1983: 83) or Brown and Levinson's (1987: 65) notion that certain acts intrinsically threaten face. The following second-wave position (e.g. Fraser and Nolen 1981: 96) holds that "no sentence is inherently polite or impolite"; in short, impoliteness is determined by contextual factors.⁴

Our own third-wave position acknowledges the relevance of contextual factors; however, we believe that certain linguistic structures are predisposed to an impolite reading. Normative aspects are of relevance regarding the conventionalization of particular expressions. We understand norms here as "regularities of co-occurrence between linguistic expressions and their extra-linguistic contexts of use" (Terkourafi 2005: 247). Impoliteness is normative insofar as there are certain recognizable expressions that signal the speaker's intentional negative evaluative stance: "I can only be rude to you in a way that you recognize as being rude. Otherwise, no matter how rude I think I am being, unless you concur with this evaluation, I have not been rude to you" (Terkourafi 2005: 249).

Manes and Wolfson (1981: 123) show that the structure [NP is/looks really ADJ] has become conventionalized for compliments in American English in multiple contexts. In short, politeness is created by a particular linguistic form regularly co-occurring with particular context types (Terkourafi 2005: 248). Through repeated use, speakers acquire "a knowledge of which expressions to use in which situations" (Terkourafi 2002: 197); this knowledge is bleached from particulars and generalizable to a 'default context' (see Section 3.1). Similar formulaic expressions regularly co-occur with contexts which are predisposed for impoliteness, and come to be conventionalized (Culpeper 2010: 3243), i.e. perceived as impolite in almost all contexts of use based on the hearer's previous experience of similar such contexts and usage (see Culpeper 2011: 135–136 for a non-exhaustive list of formulae).

³ We use "epithet" as described by Huddleston and Pullum (2002: 380–381): "[A]n emotive expression which serves to indicate annoyance with the individual concerned rather than to give an objective description".

⁴ See also Kienpointner (1997: 225) for a supportive view, and further discussions in Van Olmen, Andersson, and Culpeper (2023) and Culpeper (2011: 117–126).

Taylor (2012: 243) discusses default contexts in connection with entrenchment. While impoliteness overall is much less frequent in everyday interactions than politeness, and may be “rather marginal to human linguistic behaviour in normal circumstances” (Leech 1983: 105), we understand it as the more salient phenomenon. Some contexts may be especially prone to impoliteness or even license the use of impoliteness, such as exploitative TV shows (Bousfield 2008; Culpeper and Holmes 2013) or hate speech in online interactions (Kienpointner 2018). That is, speakers are assumed to have knowledge of impoliteness formulae outside of personal use.

Culpeper (2011: 134) grouped conventionalized impoliteness formulae in five languages, among them German and English, “according to structural commonalities”, here referred to as “patterns”. We align with Culpeper and understand such patterns as constructions (Goldberg 1995; see Section 2.2). We conclude by investigating one such construction, which is comprised in the personalized negative assertion “[you] [are] [so/such a] [shit/stink/[. . .]/bitch/hypocrite/disappointment [. . .]]” identified by Culpeper (2011: 135).⁵ The speaker states their opinion on the subject, and makes a predication or evaluation about the subject that is negative, i.e. it is an insulting form. Our proposed construction [*such* (ART) Nom] is an abstraction of this formula.

As Culpeper investigates impoliteness, his examples only contain negatively connoted nouns (e.g. *shit*, *stink*, *bitch*, *hypocrite*, *disappointment*), however other elements may enter the noun slot; see the positively evaluative noun in *You’re such a sweetheart*. We also acknowledge the possibility that negatively evaluative nouns may be used as banter (Leech 1983), e.g. to express solidarity between very close friends.

As “insults can also be creatively modified [. . .] to intensify their offensiveness” (Van Olmen, Andersson, and Culpeper 2023: 25), we assume a large pool of possible candidates for the noun slot; see, for instance, neutral-valence nouns like *linguist* or *theoretician* (Van Olmen, Andersson, and Culpeper 2023: 27, 33), which might be coerced into an evaluative reading in our construction. Culpeper (2011) does not comment explicitly on the role of *such* in the personal negative assertion; we address this further in Section 2.2.

⁵ Personalized negative assertion in the form ‘you are such a NP’ is also attested in corpus data for the nouns *bitch* and *bastard* in Van Olmen, Andersson, and Culpeper (2023: 26), albeit with a lower frequency than the corresponding *you* + NP construction.

2.2 Constructions

We conduct our analysis within the framework of Construction Grammar, thereby taking a cognitive-linguistic, usage-based approach. While theories vary⁶ and we do not adopt a specific sub-approach, we generally follow Goldberg (e.g. 1995, 2006, 2019). Construction Grammar views linguistic expressions as pairings of form/structure and meaning/function (Fillmore, Kay, and O'Connor 1988: 507; Goldberg 2006: 3; Hilpert 2014: 2). Crucially, it emphasizes the significance of conventionalized, idiomatic expressions, as “idiomaticity in a language includes a great deal that is productive, highly structured, and worthy of serious grammatical investigation” (Fillmore, Kay, and O'Connor 1988: 501, 534).

Generally, constructions come as (a) lexically fixed, e.g. idioms, (b) partially schematic with some lexically fixed material, and (c) fully schematic templates, such as argument structure constructions (cf. Taylor 2012: 84; Goldberg 2019). The evaluative *such* construction is partially schematic as only the lexical item *such* is fixed. It matches Taylor's (2012: 84) “constructional idioms” and “formal idioms” as described by Fillmore, Kay, and O'Connor (1988: 505–506). Construction Grammar proposes that linguistic items are stored in the mind within a *network* in which they cluster, connect, overlap, and get co-activated (Goldberg 2019: 6). Novel constructions enter the network via repeated exposure, which leads to *entrenchment* (Taylor 2012: 122; Goldberg 2019: 54). As constructions are motivated by higher-level constructions, they *inherit* formal and/or semantic features from their parent constructions (Goldberg 1995: 72–81).

Two central concepts are *non-compositionality* and *coercion*. Constructions may express meaning beyond the sum of their lexical constituents. This non-compositional, idiosyncratic meaning may lead to coercion: the meaning imposed by the construction overrides the denotational meaning of certain lexical items occurring in the construction (Taylor 2012: 95, 279; Hilpert 2014: 17; Goldberg 2019: 37). As a result of a construction's formal and functional requirements, it may impose *constraints* that limit the lexical material that is admissible in its schema (Hilpert 2014: 18–20; Goldberg 2019: 51–73). Example (1) illustrates some of these concepts.

- (1) *He sneezed the napkin off the table.* (Goldberg 1995: 224)

Through coercion, the intransitive verb *sneeze* acquires a transitive, three-argument sense (Goldberg 1995: 225). Extending Goldberg's example, *He coughed the napkin off the table* sounds plausible, while *?He breathed the napkin off the table*

6 For overviews see Ungerer and Hartmann (2023) and Haspelmath (2023).

does not. Sufficient force, which breathing lacks, seems to be a constraint on the verb licensed by the construction (cf. Goldberg 1995: 29, 2006: 100).

2.2.1 Aside: on *such*

As Van der Auwera and Sahoo (2020: 2) point out, “despite the rich grammatical tradition, English grammarians do not know what to do with *such*”. Prototypically, *such* is a referential expression. The Oxford English Dictionary describes it as “a demonstrative word used to indicate the quality or quantity of a thing by reference to that of another or with respect to the effect that it produces or is capable of producing. Thus, syntactically, *such* may have backward or forward reference”. Huddleston and Pullum (2002: 1546) identify *such* as a modifier in a noun phrase structure that concerns either degree or kind, and connect it to a “scalar comparison of equality” (Huddleston and Pullum 2002: 1130). Others (Altenberg 1994: 229–230; Spinillo 2003: 197–200) call this “comparative reference” (cf. Halliday and Hasan 1976: 76–87). As Bolinger (1972: 62) puts it, *such* “identifies a quality rather than an object”. However, often there is no identifiable (comparative) co-referent in the discourse. A solution comes from Van der Auwera and Sahoo (2020), who outline a “demonstrative similative”: *such* creates an ad hoc, context-dependent category in discourse. For the utterance *I want such a cat*, “[t]he speaker [. . .] might well stand in front of a cat and point at it. [. . .] The speaker has just created an *ad hoc* category and the cat that (s)he wants is an indefinite exemplar of this new category” (Van der Auwera and Sahoo 2020: 2). This is similar to what Altenberg (1994: 231) calls an “exophoric”, or situational, reference (cf. Halliday and Hasan 1976: 31–37). Huddleston and Pullum (2002: 1546) agree that for *I’ve never had to wait such a long time before*, “the secondary term is retrieved from the situation of utterance: ‘such a long time as this, i.e. as the time I’m currently having to wait’”.

Several authors have also investigated *such* as an intensifier (e.g. Bolinger 1972: 61–77; Altenberg 1994; Ghesquière and Van de Velde 2011). They typically analyze *such* as either identifying or intensifying, depending on context. For noun phrases like “*such a X*”, Altenberg (1994: 234) states that the interpretation of *such* depends on whether there is a possible co-referent in the context or a gradable element within the noun phrase. *Such* is seen as identifying if there is a possible co-referent but no gradable element, but as intensifying if the opposite is the case. Compare *such a snob* and *such a telescope*, where *snob* is seen as gradable and a *telescope* as either there or not (Spinillo 2003: 207). A sub-entry in the Oxford English Dictionary goes further and lists a colloquial use for *such* as “an absolute intensive, the implied clause of comparison being indeterminate and quite lost sight of (‘[W]e stayed the night in *such* an inn!’)”. The absolute intensive specifies the type of mod-

ification – one to the highest possible degree. The missing clause of comparison, which still presupposes a co-referent, echoes the exophoric/situational reference mentioned. Absolute intensive further implies an exclamatory character of *such*, which has long been noted (Bolinger 1972: 68; Altenberg 1994: 233; Huddleston and Pullum 2002: 923). Bolinger (1972: 91–93) and Altenberg (1994: 239) argue that the function of *such* has diachronically shifted towards intensification, and Altenberg makes a connection to a cline from propositional via textual to expressive meaning described by Traugott (1982). Ghesquière and Van de Velde (2011), adopting a data-driven, constructional view, found supporting corpus evidence for this.

Merging these analyses and viewing *such* as simultaneously identifying and intensifying may explain what the evaluative *such* construction does. *Such* creates an ad hoc category of e.g. an idiot, one that behaves in a particular way, based on the discourse situation. It identifies the subject as a specimen of this ad hoc category. Due to the absolute intensive character of *such*, this specimen is situated on the upper extreme of the scale, meaning it displays the attributes inherent or associated with the category to the highest possible degree.

2.2.2 The evaluative *such* construction

Our proposed construction, exemplified by *You are [such an idiot]*, is hypothesized to have negative evaluation as its prototypical function. It is characterized by the modifier *such* as a lexical *pivot*⁷ and by a tendency to feature epithets. While *such* has been linked to intensification and it could be argued that the negative evaluation resides in the epithet and is merely intensified by *such*, we propose that the evaluation is also caused by the construction itself, as outlined above. The construction has been linked to negative evaluation by authors such as Culpeper (2011: 135, see Section 2.1) and Taylor (2012: 90), who even calls it an “epithet construction”. However, they do not analyze it in detail.

Formally, the construction is a noun phrase comprising *such*, the indefinite article (except with plurals and non-count nouns), and a nominal, typically a noun: [*such* (ART) Nom]. The nominal can include an attributive adjective⁸ modifying the

⁷ We adopt *pivot* from *pivot schemas* in language acquisition (cf. Hilpert 2014: 164), meaning a fixed item accompanied by open slots.

⁸ The role of the adjective is not trivial. An adjective can be disambiguating, and it would be easy to propose that the adjective becomes obligatory in the construction if coercion is to be avoided. Van Olmen, Andersson, and Culpeper (2023) observe this effect in their corpus data for *you* + NP, and it is also apparent in the data for Queisser (2024). However, this raises the question whether the utterance remains an instantiation of the construction, i.e. if we are dealing with the same

noun, as exemplified in (02) below, or a phrasal noun like *pain in the ass*. Due to its scope, the current discussion is limited to plain nouns. The construction occurs in various syntactic contexts (see Table 1):

Table 1: The evaluative *such* construction across clause types (cf. also Huddleston and Pullum 2002: 923).

no.	example	clause type
(02)	<i>You're [such a fucking idiot].</i> (The Handmaid's Tale series)	declarative
(03)	<i>You're not [such an idiot].</i> (The Simpsons series)	declarative negated
(04)	<i>Do you have to be [such an idiot]?</i> (blog post)	closed interrogative
(05)	<i>When did you become [such an idiot]?</i> (movie review)	open interrogative
(06)	<i>Don't be [such an idiot].</i> (opinion piece)	imperative
(07)	<i>[Such an idiot!]</i> (The Dead Zone series)	exclamative ⁹
(08)	<i>Who hired [such an idiot]?!</i> (discussion forum)	non-predicative interrogative

The construction typically occurs in a predicative context. However, the copula *be* is not part of the construction. Other copular verbs occur, as shown by (05). Examples (07) and (08) are not formally predicative. However, we argue that predication is implied in (07) and that in (08), it is presupposed that the person who was hired is *an idiot*. Predication is thus a central feature of the construction's use. As predication ascribes characteristics (Huddleston and Pullum 2002: 251–252), this feeds into the evaluative character of the construction, especially with the pronoun *you* as the subject. In this specific context, the speaker tells the addressee what the addressee is. This is not information-giving. What the speaker is really doing is stating an *opinion* about the addressee. This is in line with the expressive function of *such* outlined above.

A corpus analysis was conducted for Queisser (2024).¹⁰ The construction was expected to occur most frequently in informal spoken language (cf. Altenberg 1994: 235), with predominantly negative, person-denoting nouns (epithets). The find-

form-meaning pair. Compare *You are such a girl* with *You are such a pretty girl*. Goldberg's (1995: 31–39) ‘constructional polysemy’ and Hilpert's (2014: 181) “many-to-many mappings” offer good solutions for this.

⁹ Huddleston and Pullum (2002: 923) use the term “non-exclamative exclamation” due to their more restrictive definition of exclamatives.

¹⁰ Corpora used: Corpus of Contemporary American English (COCA), TV Corpus, Movie Corpus, Corpus of American Soap Operas, News on the Web (NOW) Corpus. The COCA was used as a default, mixed-genre corpus, the entertainment corpora as examples of informal spoken language modelled after naturally occurring discourse, and the NOW as a control corpus featuring more formal, non-interpersonal language.

ings confirmed both expectations. They further showed that the subject is most frequently a personal pronoun while nouns and proper nouns are rare. Of the personal pronouns, singular *you* is the most frequent, followed by *I*. This provides evidence for frequent interpersonal use of the construction. While the construction occurs most frequently with negative nouns – in close to 80% of the data obtained, depending on the corpus and subject – there are instances with positive and neutral nouns. The positive nouns indicate that the construction is indeed evaluative on a general level rather than only impolite, while the abundance of negative nouns underscores the construction's prototypical, impolite force. The neutral nouns provide evidence for a coercion effect. They should be interpreted as evaluative, mostly negative, due to being used in the construction. A small type-token ratio and many unique instantiations (hapax legomena) indicate that the construction is very productive and that speakers use it creatively.

We briefly illustrate this. For the search string [you BE *such* ART N] in the COCA¹¹, the top ten nouns are: *asshole* (96 instances), *liar* (89), *jerk* (79), *bitch* (63), *idiot* (61), *dick* and *loser* (52 each), *baby* (50), *inspiration* (46), *child* (31), and *gentleman* (31). There are 2122 total instances, featuring 604 different nouns, with 379 nouns occurring only once, and 172 occurring twice. While the top ten nouns account for 30% of all occurrences, nouns occurring only once or twice account for 26%. Conventionalized expressions and more novel/creative examples thus occur in comparable shares, exemplifying Goldberg's (2006: 89) "cognitive anchoring": "a high-frequency type of example act[ing] as [. . .] a salient standard of comparison".

3 Method

To investigate coercion effects attributable to *such* (German *so*) we conducted four questionnaire studies. The studies investigate the construction embedded in the form [PRON BE *such* ART N] in English and German, using the 2nd and 3rd person singular pronouns, respectively.

The third-person pronouns *he/she* and *er/sie* were used in one set of studies; we shall refer to these as E3P (English 3rd person) and G3P (German 3rd person). Third person was used intentionally to avoid respondents feeling addressed by the statements. The objective was for respondents to focus on the communicative intention of the speaker rather than their own interpretation as an imagined target. Further, in conditions without *such/so*, the third-person pronoun implies neutral informa-

¹¹ Accessed 06 June 2024. This very brief analysis is purely quantitative, and context was not controlled for. Individual erroneous hits may not represent the construction.

tion-giving, which served as a control for evaluative meaning arising from factors other than *such*.

Another set of studies used the second-person singular pronouns *you* and *du*, as *you* is the most frequent subject in the corpora (see Section 2.2.2).¹² We refer to these studies as E2P (English 2nd person) and G2P (German 2nd person). Note that German uses formal and informal second-person pronouns; in G2P, we selected the informal pronoun *du* instead of the formal *Sie*. While the formal pronoun is possible in impoliteness contexts (e.g. *Sie arschloch* (‘you-V asshole’)), it is marked, and might be read as having a humorous effect. The non-binary singular pronoun *they* was not used as its prototypically plural meaning may have affected the perceived grammaticality of the stimuli.

For each study, two separate questionnaires were compiled. In Condition 1, respondents were presented with predicative sentences without *such* (*so* in the German studies); Condition 2 comprises largely the same sentences with *such* (*so*).

3.1 Stimuli and presentation

Each study featured ten experimental sentences and 20 filler items. The predicative nouns used in the experimental sentences are presumed to have either positive (2), negative (2), or neutral (6) valence (see Table 2). However, we acknowledge that no person-denoting predicative noun will be perceived as neutral in 100% of the cases. The positive and negative nouns served to verify that prototypical (i.e. frequent in the corpora) examples of the construction would indeed be rated as expected, while a higher number of neutral nouns was selected to obtain more data relevant to the study of coercion effects.

The nouns were chosen based on their frequency in the corpus data obtained for Queisser (2024). Frequent nouns (*child, girl, man*) are contrasted with likely less conventionalized but attested low-frequency nouns (*perfectionist, student, teacher*). Three of the neutral stimuli used in E3P and G3P (*child, girl, man*) were considered unsuitable for a non-evaluative context with *you*, as stating something so obvious about the addressee may imply ‘more’ meaning. For E2P and G2P, these were replaced with nouns that we considered suitable (*neighbor, guest, shareholder*). In E2P and E3P, the stimuli were randomly assigned a gender for the pronoun; the gender is retained in G2P and G3P.¹³ We acknowledge the possible influence of

¹² *You* may also increase the directness of reference to the hearer and strengthen the separation between interlocutors (see Van Olmen, Andersson, and Culpeper 2023: 38).

¹³ Note that German has obligatory gender suffixation in nouns, with *-in* being used for female referents.

gender and stereotypical ascriptions of (professional) roles to certain genders on participants' ratings (see also our concluding remarks in Section 5).

Table 2: Experimental stimuli with and without *such* in English and German.

	S/he is (such) a(n)	You are (such) a(n)	Sie/er ist (so) ein(e)	Du bist (so) ein(e)
positive		inspiration		Inspiration
		sweetheart		Schatz
negative		asshole		Arschloch
		idiot		Idiot
neutral		perfectionist		Perfektionistin
		student		Studentin
		teacher		Lehrer
	girl	neighbor	Mädchen	Nachbar
	man	guest	Mann	Gast
	child	shareholder	Kind	Aktionär

It proved challenging to identify suitable nouns, and for some of the resulting stimuli it may be less easy to imagine a speech situation in which they occur naturally. However, the nouns occur in the corpus data for Queisser (2024), albeit with a low frequency and with a disambiguating adjective. We selected these nouns to investigate the ratings in the absence of the disambiguating adjective and hoped to present the respondents with novel statements that would increase the reliance on constructional over lexical meaning. Thus, we do not see the peculiarity of these stimuli as a weakness of our studies. Respondents were expected to draw on their prior linguistic experience with the construction, comparable to studies which use nonce words to achieve this effect.

To mask the purpose of the study, 20 fillers were included (see Appendix). They did not share the form of the construction and were identical across studies wherever possible. Some fillers were changed to be used with *you*. The fillers were designed in such a way that there was a total of ten positive, negative, and neutral statements each in the experiment. An equal share of feminine and masculine pronouns/nouns was used.

As shown in Table 2 above, the studies used context-less examples. While we acknowledge the role (social) context plays for impoliteness judgments (see Section 2.1), we chose not to provide contexts for the stimuli for the following reasons:

First, previous research has shown that speakers can judge impoliteness outside of particular contexts; see, e.g., Jain (2022: 389) and Van Olmen, Andersson,

and Culpeper (2023: 37), who note that the *you* + NP construction may be interpreted as impolite without a specific context. Second, we assume that “people acquire a knowledge of impoliteness formulae that far exceeds *their own direct experience of usage*” (Culpeper 2010: 3238, emphasis in original; see also a similar point in Kleinke and Bös 2015: 25 on respondents’ first- and second-order understanding of impoliteness). Third, our stimuli contain pronouns, i.e. deictic expressions which by their very nature assume a speech situation with shared knowledge between speaker and hearer (cf. Traugott 1982: 248). Fourth and final, a specified context may be leading and therefore counterproductive. Based on these points we assume that speakers will draw on their pre-existing knowledge of impoliteness contexts and shared conventions about face-threats to construe for themselves a context of use for the stimuli.

In all studies, the conditions (with and without *such*) were run between-group, meaning that each respondent saw only one version of the questionnaire. Respondents were not aware of this. The questionnaires were assigned (50:50) by a random generator when the link was opened. The 30 stimuli were displayed one per page in a randomized order. The prompt accompanying each statement was “Do you think the speaker is making a positive, negative, or neutral statement about the person?” The prompt aimed to make the respondents focus on the communicative intention of the speaker. Respondents rated each statement on a five-point Likert scale with the options ‘very negative’, ‘negative’, ‘neutral’, ‘positive’, and ‘very positive’. Smiley faces accompanied each option to avoid a reverse reading of the scale.

“You are such an asshole.”

Do you think the speaker is making a positive, negative, or neutral statement about the person?



Figure 1: Screenshot of an experimental stimulus in E3P.

3.2 Respondents and data

Respondents were recruited by disseminating the survey link via social media. We asked potential respondents with a background in linguistics to refrain from taking part, as the studies’ aims might have been overly transparent for them.

Some sociodemographic data were gathered: gender (female, male, non-binary), age in years, highest educational achievement, self-identified level of proficiency in English/German with the options ‘basic’, ‘intermediate’, ‘advanced’, ‘fluent’, ‘close to native’, and ‘I am a native speaker of English/German’, and the variety of English/German spoken (optional answer). Only data sets of respondents who rated their proficiency in English/German as ‘fluent’ or higher were used for the analysis. A small number of respondents was excluded as their overall rating vastly differed from other respondents, especially in the filler items. These respondents might have miscomprehended the rating scale, or might have operated under a particular bias. We believe that exclusion is warranted as these answers are not representative of the general population we tested.

4 Results and discussion

4.1 Studies on English

E3P, conducted for Queisser (2024), used *he* and *she* as subject pronouns; E2P, conducted for the current work, used *you*. We discuss both studies together. Data on the respondents are available in Table 3. All have at least a high school diploma or equivalent, most have a university degree. Gender, age, education, and variety of English spoken¹⁴ had no discernable impact on ratings. However, due to the small sample size and the heterogeneity of the respondents, it was not possible to form meaningful groups for an in-depth assessment. Due to prematurely aborted surveys and individual questionnaires that had to be excluded (see Section 3.2), the number of responses per condition differs.

Table 3: Respondent data for both studies: subject pronoun, total number of responses, number by condition, age information, gender (female/male/non-binary), and self-reported level of English (native/close to native/fluent).

subj.	tot.	w/o such	with such	ages	mean age	gender	level
<i>he/she</i>	48	27	21	22–69	37	27/20/1	31/10/7
<i>you</i>	33	17	16	24–73	40	19/8/3	19/6/8

¹⁴ Variety of English is not reported here for reasons of brevity.

With some exceptions, all fillers and control sentences with overt positive or negative valence were rated as expected. Although the fillers were not part of the experimental conditions, their mostly uniform ratings show that there is a high degree of consensus across the respondents. Some interesting ratings of the control sentences with overt positive or negative nouns occurred. *You are such an idiot* was rated as positive by two native speakers, who may have read it as banter. One native speaker rated *She is such an asshole* as positive, which most likely happened by mistake. Isolated neutral ratings occurred for *You are such an inspiration* (2) and *You are such a sweetheart* (1).

4.1.1 Overall results

We focus on the neutral stimuli for which coercion was predicted. If otherwise neutral nouns adopt an evaluative meaning when occurring in the evaluative *such* construction, these stimuli should be rated as neutral without *such* and as positive or negative with *such*. Predominantly, they should be rated as negative. These predictions are confirmed by the results of both studies, however to different degrees. The effect is strongest in E3P. Figures 2 and 3 illustrate the shares of positive, negative, and neutral ratings in both studies. Detailed results follow in Section 4.1.2. As is clearly visible, the white (neutral) area of the graphs is smaller, and the dark gray (negative) area outweighs the light gray (positive) area when *such* is present.

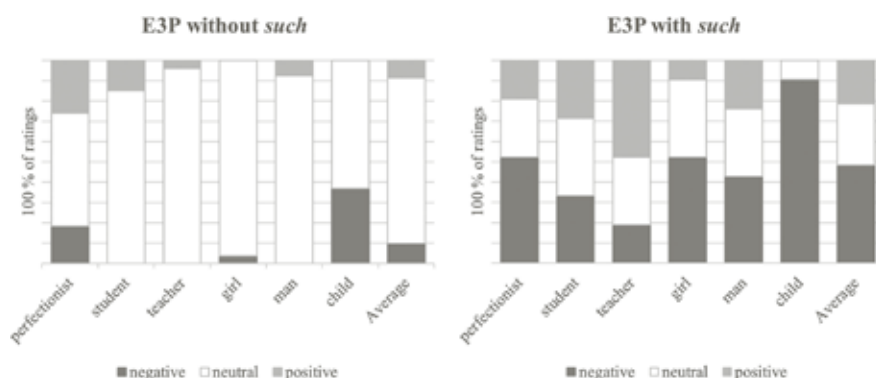


Figure 2: E3P: shares of positive, negative, and neutral ratings per noun and on average.

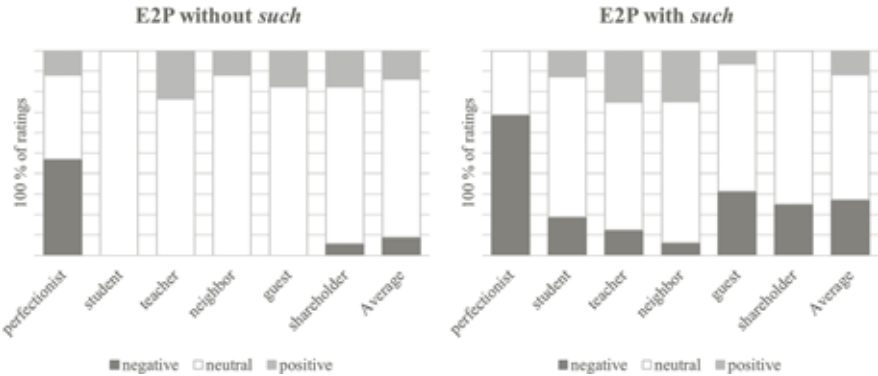


Figure 3: E2P: shares of positive, negative, and neutral ratings per noun and on average.

4.1.2 Detailed results

The results in percentages are available in Tables 4 and 5. We first discuss E3P. As the column ‘Average’ shows, positive and negative ratings account for under 10% each (18.5% total) without *such*. However, there is an interesting variation between the stimuli, indicating that some of the nouns do lean towards a positive or negative evaluation. This is not surprising and could be due to a conventionalized or individual interpretation. *Perfectionist* is the only item to receive both positive and negative ratings without *such*. None of the stimuli are rated as ‘very positive’ or ‘very negative’. The majority (81.5%, on average) of ratings are neutral, as was predicted.

Table 4: E3P: percentages of positive, negative, neutral, and total evaluative ratings with and without *such*, and the difference (Δ) of percentages. Potential discrepancies in sums are caused by mathematical rounding to one decimal.

	<i>perfectionist</i>			<i>student</i>			<i>teacher</i>			<i>girl</i>		
	w/o	w	Δ	w/o	w	Δ	w/o	w	Δ	w/o	w	Δ
pos.	25.9	19.0	−6.9	14.8	28.6	+13.8	3.7	47.6	+43.9	0.0	9.5	+9.5
neg.	18.5	52.4	+33.9	0.0	33.3	+33.3	0.0	19.0	+19.0	3.7	52.4	+48.7
neut.	55.6	28.6	−27.0	85.2	38.1	−47.1	96.3	33.3	−63.0	96.3	38.1	−58.2
eval.	44.4	71.4	+27.0	14.8	61.9	+47.1	3.7	66.7	+63.0	3.7	61.9	+58.2

Table 4 (continued)

	<i>man</i>			<i>child</i>			Average			Legend
	w/o	w	Δ	w/o	w	Δ	w/o	w	Δ	
pos.	7.4	23.8	+16.4	0.0	0.0	0.0	8.6	21.4	+12.8	w/o: without <i>such</i> w: with <i>such</i> Δ: delta/difference
neg.	0.0	42.9	+42.9	37.0	90.5	+53.4	9.9	48.4	+38.5	
neut.	92.6	33.3	-59.3	63.0	9.5	-53.4	81.5	30.2	-51.3	
eval.	7.4	66.7	+59.3	37.0	90.5	+53.4	18.5	69.8	+51.3	

This changes clearly when *such* is present. The average share of neutral ratings drops to 30.2%, while positive (21.4%) and negative (48.4%) ratings increase. Individual statements are also rated as stronger, meaning 'very positive' or 'very negative'. For better legibility and to meet the sample size requirements of our statistical test (see Section 4.3), we include 'very negative' ratings in the 'negative' ratings and 'very positive' ratings in the 'positive' ratings in all experiments. Negative ratings now account for almost half of all ratings. This is crucial as it provides evidence for the hypothesis that the evaluative *such* construction has a particularly negative connotation, and that coercion will work in favor of negative polarity. *Teacher* is an interesting exception that may have to do with prestige. The delta (Δ) columns show the differences between the two conditions¹⁵. The final column best illustrates the overall effect. The increase in total evaluative ratings, quite logically, corresponds to the decrease in neutral ratings.

In E2P, the results without *such* are comparable. The average share of positive ratings (13.7%) is slightly larger and the share of negative ratings (8.8%) slightly smaller than in E3P. Only two stimuli are rated as negative, while five are rated as positive. The average share of neutral ratings is 77.5%, which is not much lower than in E3P (81.5%). An even smaller share may have been expected, as using *you* as a predicand should not be interpreted as informational but implies opinion-giving. For the stimuli used in both studies, a direct comparison is possible. *Perfectionist* is rated more negatively with *you*, *teacher* is rated more positively. While an analysis of the lexical meaning of the items is beyond the scope of the current discussion, it is interesting that the interpretation of the stimuli also varies with the subject pronoun.

¹⁵ The difference is calculated by subtraction of the percentages. It should be noted that the number of respondents per condition differs (27 vs. 21). This means that fewer individuals account for the percentage obtained for the condition with *such*. In E2P, the numbers were almost even (16 vs. 17). This slight imprecision is remedied by the statistical test (see Section 4.3).

Table 5: E2P: percentages of positive, negative, neutral, and total evaluative ratings with and without *such*, and the difference (Δ) of percentages. Potential discrepancies in sums are caused by mathematical rounding to one decimal.

	<i>perfectionist</i>			<i>student</i>			<i>teacher</i>			<i>neighbor</i>		
	w/o	w	Δ	w/o	w	Δ	w/o	w	Δ	w/o	w	Δ
pos.	11.8	0.0	-11.8	0.0	12.5	+12.5	23.5	25.0	+1.5	11.8	25.0	+13.2
neg.	47.1	68.8	+21.7	0.0	18.8	+18.8	0.0	12.5	+12.5	0.0	6.3	+6.3
neut.	41.2	31.3	-9.9	100	68.8	-31.3	76.5	62.5	-14.0	88.2	68.8	-19.5
eval.	58.8	68.8	+9.9	0.0	31.3	+31.3	23.5	37.5	+14.0	11.8	31.3	+19.5

	<i>guest</i>			<i>shareholder</i>			Average			
	w/o	w	Δ	w/o	w	Δ	w/o	w	Δ	
pos.	17.6	6.3	-11.4	17.6	0.0	-17.6	13.7	11.5	-2.3	Legend
neg.	0.0	31.3	+31.3	5.9	25.0	+19.1	8.8	27.1	+18.3	w/o: without <i>such</i>
neut.	82.4	62.5	-19.9	76.5	75.0	-1.5	77.5	61.5	-16.0	w: with <i>such</i>
eval.	17.6	37.5	+19.9	23.5	25.0	+1.5	22.5	38.5	+16.0	Δ : delta/difference

Like in E3P, we see an increase in evaluative ratings when *such* is present. Surprisingly, the overall effect is smaller, with neutral ratings still accounting for 61.5% in E2P (versus 30.2% in E3P). Interestingly, native speakers were roughly 10% more likely to rate the stimuli as neutral than non-native speakers. Unlike in E3P, only average negative ratings increase, while average positive ratings *decrease*. One ‘very negative’ rating was obtained for *guest*. *Perfectionist* is rated particularly negatively. As the delta (Δ) columns show, there is a decrease in neutral ratings for each stimulus, despite the smaller effect. With the exception of *neighbor*, all stimuli gain more negative than positive ratings. While in E3P only *perfectionist* changes from being rated more positively to more negatively, this effect of switching polarity is observed for two stimuli in E2P: *guest*, and *shareholder*. The results of this study therefore confirm the effects observed in E3P, albeit to a lesser degree.

4.2 Studies on German

In line with the studies discussed in Section 4.1, this section presents the German data on the evaluative *such* construction. G3P used the German third-person pronouns *er* and *sie* ‘he/she’ as subjects; in G2P, the second-person pronoun *du* ‘you’ was used. Both studies will be discussed together below. Respondent data are dis-

played in Table 6. As in the English studies, all respondents have at least a high school diploma or equivalent, with most having completed a university degree. The variables gender, age, education, and variety of German spoken had no discernable impact on ratings.

Table 6: Respondent data for both studies: Subject pronoun, total number of responses, number by condition, age information, gender (female/male/non-binary), and self-reported level of English (native/close to native/fluent).

subj.	tot.	w/o <i>so</i>	with <i>so</i>	ages	mean age	gender	level
<i>er/sie</i>	101	48	53	19–65	39	73/27/1	99/2/0
<i>du</i>	57	23	34	20–84	38	34/23/0	54/1/2

The mostly uniform ratings of the fillers show that there is a high degree of consensus across the respondents. While some exceptions occurred in filler ratings, they have no bearing on respondents’ performance in the experimental conditions. Of the control sentences with overt positive or negative nouns, one native speaker rated *Sie ist ein arschloch* ‘She is an asshole’ as positive, which may have been a mistake. Isolated neutral ratings occurred for the lexeme *Idiot* ‘idiot’, specifically in *Er ist ein Idiot* ‘He is an idiot’ (2), *Du bist ein Idiot* ‘You are an idiot’ (1) and *Du bist so ein Idiot* ‘You are such an idiot’ (1).

4.2.1 Overall results

We focus here on the stimuli including nouns with neutral valence for which we predicted coercion effects. Both studies confirm that these otherwise neutral statements adopt an evaluative meaning in the condition with *so*. Figures 4 and 5 illustrate the shares of positive, negative, and neutral ratings in both studies; detailed results follow in Section 4.2.2. As is clearly visible, the white (neutral) area of the graphs is smaller, and the dark gray (negative) area outweighs the light gray (positive) area when *so* is present. The effect is stronger in G3P, just like in the English set (see Section 4.1); overall, the coercion effect is stronger in both German studies than in the English data.

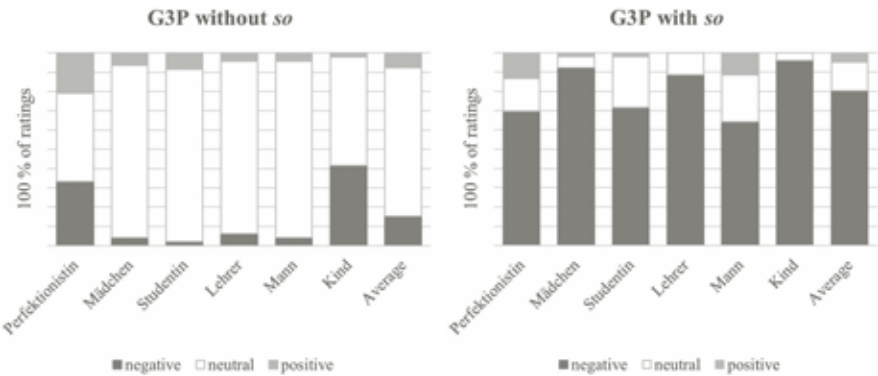


Figure 4: G3P: shares of positive, negative, and neutral ratings per noun and on average.

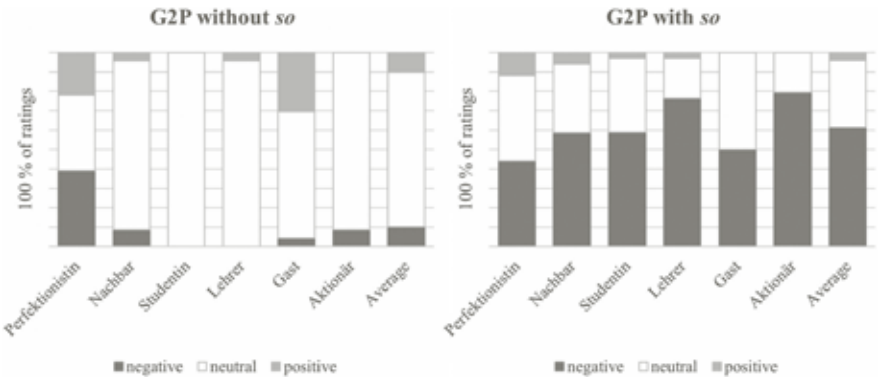


Figure 5: G2P: shares of positive, negative, and neutral ratings per noun and on average.

4.2.2 Detailed results

The results in percentages are given in Tables 7 and 8. We first discuss G3P. As the column ‘Average’ shows, evaluative ratings account for 22.9% in the condition without *so*, which is similar to the average in E3P (18.5% evaluative ratings). In the German data, we also find variation between the stimuli. The stimulus *Kind* ‘child’ has a negative rating of 41.7%, while *Perfektionistin* ‘perfectionist’ is again the only item to receive both positive and negative ratings over 20%; this could be due to a conventionalized or individual interpretation of the stimuli’s semantic content. In contrast to E3P, some of the stimuli received few ratings as ‘very positive’ or ‘very negative’ (*Perfektionistin* ‘perfectionist’, *Mädchen* ‘girl’, *Mann* ‘man’, *Kind* ‘child’). The majority (77.1%, on average) of ratings is neutral, as was predicted.

Table 7: G3P: percentages of positive, negative, neutral, and total evaluative ratings with and without *so*, and the difference (Δ) of percentages. Potential discrepancies in sums are caused by mathematical rounding to one decimal.

	<i>Perfektionistin</i>			<i>Studentin</i>			<i>Lehrer</i>			<i>Mädchen</i>		
	w/o	w	Δ	w/o	w	Δ	w/o	w	Δ	w/o	w	Δ
pos.	20.8	13.2	-7.6	8.3	1.9	-6.4	4.2	0.0	-4.2	6.3	1.9	-4.4
neg.	33.3	69.8	+36.5	2.1	71.7	+69.6	6.3	88.7	+82.4	4.2	92.5	+88.3
neut.	45.8	17.0	-28.9	89.6	26.4	-63.2	89.6	11.3	-78.3	89.6	5.7	-83.9
eval.	54.2	83.0	+28.9	10.4	73.6	+63.2	10.4	88.7	+78.3	10.4	94.3	+83.9

	<i>Mann</i>			<i>Kind</i>			<i>Average</i>			
	w/o	w	Δ	w/o	w	Δ	w/o	w	Δ	
pos.	4.2	11.3	+7.2	2.1	0.0	-2.1	7.6	4.7	-2.9	Legend w/o: without <i>so</i> w: with <i>so</i> Δ : delta/difference
neg.	4.2	64.2	+60.0	41.7	96.2	+54.6	15.3	80.5	+65.2	
neut.	91.7	24.5	-67.1	56.3	3.8	-52.5	77.1	14.8	-62.3	
eval.	8.3	75.5	+67.1	43.8	96.2	+52.5	22.9	85.2	+62.3	

Confirming our expectations, we can observe a clear change in the condition with *so*. Neutral (14.8%) and positive ratings (4.7%) both drop, while we see a sharp increase in negative ratings (80.5%). There is also a slight increase in ‘very negative’ ratings; no statement was rated ‘very positive’ in the condition with *so*. The overall increase in negative ratings ($\Delta +65.2\%$) is stronger than for the comparable English study ($\Delta +38.5\%$).¹⁶ We see the sharpest increase in negative ratings for the stimuli *Lehrer* ‘teacher’ and *Mädchen* ‘girl’; here we see a difference to the English data, where *teacher* is evaluated more positively. *Perfektionistin* ‘perfectionist’ shows the smallest increase in negative ratings, and retains the highest share of positive ratings of all stimuli.

In G2P, the results without *so* are comparable (see Table 8). Average shares of evaluative ratings account for 20.3%, with an equal share of positive and negative evaluations (10.1% each). The majority of ratings (79.7%) are neutral, which is as predicted. A comparison of the stimuli used in both studies shows that *Perfektionistin* ‘perfectionist’ behaves like in G3P, with 39.1% negative and 21.7% positive ratings. *Studentin* ‘student’ has a 100% neutral rating in G2P.¹⁷ *Lehrer* ‘teacher’ was rated slightly more positively in G2P; it receives no negative ratings. Of the neutral stimuli only used in this study, *Gast* ‘guest’ has a positive rating of 30.4%; the informal second-person pronoun invites the reading of ‘house guest’, i.e. a person one

¹⁶ Calculations of percentages follow the procedure discussed in Section 4.1.2.

¹⁷ See a similar rating of this item in E2P.

voluntarily invites to one's residence, thus respondents may attribute more positive value to the item. Two stimuli received one rating each as 'very negative' (*Perfektionistin* 'perfectionist', *Aktionär* 'shareholder'); *Gast* 'guest' was rated 'very positive' once.

Table 8: G2P: percentages of positive, negative, neutral, and total evaluative ratings with and without *so*, and the difference (Δ) of percentages. Potential discrepancies in sums are caused by mathematical rounding to one decimal.

	<i>Perfektionistin</i>			<i>Studentin</i>			<i>Lehrer</i>			<i>Nachbar</i>		
	w/o	w	Δ	w/o	w	Δ	w/o	w	Δ	w/o	w	Δ
pos.	21.7	11.8	-10.0	0.0	2.9	+2.9	4.3	2.9	-1.4	4.3	5.9	+1.5
neg.	39.1	44.1	+5.0	0.0	58.8	+58.8	0.0	76.5	+76.5	8.7	58.8	+50.1
neut.	39.1	44.1	+5.0	100	38.2	-61.8	95.7	20.6	-75.1	87.0	35.3	-51.7
eval.	60.9	55.9	-5.0	0.0	61.8	+61.8	4.3	79.4	+75.1	13.0	64.7	+51.7

	<i>Gast</i>			<i>Aktionär</i>			Average			Legend
	w/o	w	Δ	w/o	w	Δ	w/o	w	Δ	
pos.	30.4	0.0	-30.4	0.0	0.0	0.0	10.1	3.9	-6.2	w/o: without <i>so</i> w: with <i>so</i> Δ : delta/difference
neg.	4.3	50.0	+45.7	8.7	79.4	+70.7	10.1	61.3	+51.1	
neut.	65.2	50.0	-15.2	91.3	20.6	-70.7	79.7	34.8	-44.9	
eval.	34.8	50.0	+15.2	8.7	79.4	+70.7	20.3	65.2	+44.9	

Like in G3P, evaluative ratings increase in the condition with *so*. 65.2% of ratings are evaluative, of which 61.3% are negative evaluations; 34.8% of ratings are neutral. No stimulus was rated 'very positive' in this condition. The overall share of evaluative ratings is comparable in the two German studies. However, the overall effect in negative ratings is higher in G3P (compare Δ +51.1% to Δ +65.2%), which also happened in the English equivalent.

Contrasting the items used in both studies, we see that for *Perfektionistin* 'perfectionist', the share of positive ratings decreases to 11.8%, with only a small increase in negative and neutral ratings. *Lehrer* 'teacher' has a roughly equal share for evaluative ratings, but retains a small amount of positive ratings (2.9%) in G2P. For *Studentin* 'student', which had a 100% neutral rating without *so*, negative ratings increase to 58.8 %; the stimulus also received a small share of positive ratings. For items used only in G2P, we observe an increase in negative ratings for *Nachbar* 'neighbor' and *Aktionär* 'shareholder'. *Gast* 'guest', which received 30.4% positive ratings in the condition without *so*, now has an equal share of negative and neutral ratings (50% each), i.e. the stimulus loses all positive ratings in the condition with *so*.

4.3 Statistical significance

We performed a separate chi-squared test for association for each of the four experiments. To conduct the analyses, we used JASP¹⁸, due to its user-friendly interface. Our null hypothesis was that there would be no association between the condition (with or without *such*) and the ratings. Our alternative hypothesis was that there would be an association between the two variables. Only the experimental stimuli (see Table 2) were included, not the controls with overt positive or negative nouns. All experimental stimuli feature nouns with supposedly neutral valence. The noun was not used as an additional variable, but all sentences under the same condition were treated as identical (*‘no such’* versus *‘such’*). This was done to test the average effect of the construction across all neutral nouns, while also improving the sample sizes. We detail the variation between nouns in Section 4.4.

The chi-square value and p-value for each experiment are provided in Table 9. We found a statistically highly significant association between the condition and the ratings for experiments E3P, G3P, and G2P. We found a slightly weaker but still statistically significant association for experiment E2P. Due to the smaller sample size and the overall smaller effect of *such* in the second-person experiments, this is not surprising. Therefore, there is strong evidence to suggest a significant influence of the presence of *such* on ratings for items with neutral valence.

Table 9: Chi-square value at two degrees of freedom χ^2 (2) and p-value for all four experiments.

Experiment	χ^2 (2)	p-value
E3P	79.1	< 0.001
E2P	11.3	0.003
G3P	261.4	< 0.001
G2P	89.3	< 0.001

4.4 Summary and general discussion

The comparison performed in the studies shows that predicative statements of the form [PRON BE (*such*) ART N] are more likely to be rated as evaluative when they contain *such/so*. Moreover, there is also a tendency for the ratings to be negative rather than positive.

¹⁸ JASP is open-source and available at <https://jasp-stats.org/>.

As shown in Tables 4 and 5 for English and Tables 7 and 8 for German, positive ratings increase comparatively little or even decrease when *such* is present. In German, only one noun, *Mann* ‘man’, increases its positive rating in the condition with *so*.

For English, all stimuli except *teacher* and *neighbor* increase more strongly in their negative rating than in their positive rating. This effect is even stronger in G3P, where all neutral stimuli experience an increase in negative ratings with *so*. *Child*, which is already rated quite negatively without *such*, is rated so in 90% of the cases with *such* in E3P; in the German data, its negative rating is increased to 96.2%. In G2P, negative ratings exceed positive ratings for all stimuli in the condition with *so*. In the English data, no negative ratings at all occur for *student*, *teacher*, *man*, *neighbor*, and *guest* without *such*. With *such*, however, all these stimuli receive negative ratings. In G2P, only two nouns, *Studentin* ‘student’ and *Lehrer* ‘teacher’, receive no negative ratings at all in the condition without *so*; in the condition with *so*, they follow the pattern of the nouns in the English data and receive a majority of negative ratings.

Perfectionist, *guest*, and *shareholder*, the stimuli losing positive ratings in the English data, are even more interesting. This observation indicates that *such* may not just tip a neutral noun towards negative polarity, but may even override an otherwise positive reading. This also holds for *Perfektionistin* ‘perfectionist’ and *Gast* ‘guest’ in German; in addition, *Lehrer* ‘teacher’ also loses positive ratings in the condition with *so*.

The different outcomes across the stimuli within each study indicate that the lexical meaning and the constructional meaning interact. This is not surprising and has been noted by Goldberg (1995: 224–225) for verbs. Overall, the tendency towards negative evaluation remains apparent and supports the hypothesis that the coercion effect proposed for the evaluative *such* construction does indeed enforce a negative reading. The use of different pronouns as subjects introduces another dimension. It is interesting that while the construction occurs most frequently with singular *you* across the corpora examined in the preliminary corpus study in Queisser (2024), the proposed coercion effect is smaller in E2P and G2P. A likely reason is the choice of noun. A predicative statement of the type *You are...* can be expected to be understood as evaluative even without *such*. Hence, it was even more important than in E3P and G3P to use nouns that could be read as neutral. As these nouns may represent marginal examples of the construction (see Section 3.1), respondents may have chosen a neutral meaning because they were more unsure in their evaluation. However, as the data still confirm the predictions, we believe that we were successful in making our respondents rely on the constructional meaning. Finally, as Culpeper (2011: 113) points out, “it is not, of course, the case that any particular linguistic form guarantees an evaluation of impoliteness in all contexts”. The evalu-

ative *such* construction therefore describes a tendency of how these utterances are likely to be used and understood. It does not constitute a general rule.

There are several limitations to the current studies. The sample of experimental stimuli comprised only six sentences each. The selection of nouns was by design deliberate, and the possible influence of personal biases should be noted. A conventionalized connotation of the nouns may have influenced the ratings. For instance, the Oxford English Dictionary lists negative uses for *child* and *girl*. Different nouns may have elicited different ratings while a larger number of stimuli may have yielded a more fine-grained overall result. However, to motivate respondents to complete the survey, a short duration was considered essential. The sample size poses a further limitation. Although 48 and 33 respondents in the English studies may be satisfying for a small-scale project, a larger sample would have made the results more robust. With 101 and 57 respondents, the German studies were more successful. In general, individual interpretations and biases towards the nouns used may have influenced the results, so that the effects observed may not be attributable solely to the proposed constructional effects.

5 Conclusion

As the results of all four studies have shown, there is strong evidence for the proposed coercion effect of the evaluative *such* construction. We answered our research questions positively by illustrating that otherwise identical statements are rated more evaluatively and more negatively when they conform to the evaluative *such* construction. We also showed that evaluations differ across various nouns used in the nominal slot, and that the construction has the power to even switch the interpretation of the (otherwise) same utterance from positive to negative meaning for some nouns. One aspect that could not be investigated is the presence or absence of an adjective in the nominal slot of the construction. As was briefly mentioned in Footnote 8, the function of the adjective is not a trivial question. However, the significance of the adjective should maybe not be overestimated. The proposed construction is a model to help conceptualize linguistic knowledge, after all. As a model, it must allow for flexibility to accommodate emerging evidence. The constructional polysemy described by Goldberg (1995: 31–39) and the many-to-many mappings mentioned by Hilpert (2014: 181) can easily accommodate varying manifestations of similar-enough patterns.

The adjective, the noun, and the subject pronoun are interesting targets for further research. As the construction is very productive, as shown by the low type-token ratio and large number of unique instances in the corpora, a corpus or

survey approach may not be suitable. Instead, a more qualitative approach such as a detailed discourse analysis is needed. Ideally, this will incorporate not only a wider discourse context in writing, but also auditory information (e.g. prosody) and visual (extralinguistic) information to identify speaker intent, recipient reaction, sarcasm, or banter. Imagine, for instance, *You are such a man* being uttered by a chauvinist versus a radical feminist. As many corpora feature material from movies and television whose original video sequences should be accessible, this is a feasible yet work-intensive task. To obtain more quantitative or quantifiable data, studies such as the ones conducted here could be repeated using methods such as recording reaction times or eye tracking. Reaction times may differ depending on the frequency with which nouns occur in the construction, and thereby their entrenchment and acceptability. Eye tracking may help discover which parts of the utterance respondents orient to in different configurations, and if it is indeed the word *such*. The use of nonce words is a further option that was hinted at above. Subsequently, the resulting toolbox can be extended to other formally and/or functionally related constructions, and to the current construction in other languages (see Ghesquière and Van de Velde 2011 for Dutch). Ultimately, compiling evidence that is as comprehensive as possible may help establish with some certainty that meaningful structures like the one(s) investigated here have cognitive reality and are therefore of significance for various forms of social interaction. This may require the involvement of other disciplines such as psycholinguistics and neuropsychology.

Acknowledgements: We are grateful to Michael Pleyer and two anonymous reviewers for their valuable feedback throughout the writing process, and to Suhasini Subba Rao for her support in the statistical analysis.

Appendix

Table 10: Experimental stimuli with and without *such/so* in English and German.

	E3P	G3P	E2P	G2P
positive	She is (such) an inspiration.	Sie ist (so) eine Inspiration.	You are (such) an inspiration.	Du bist (so) eine Inspiration.
	He is (such) a sweetheart.	Er ist (so) ein Schatz.	You are (such) a sweetheart.	Du bist (so) ein Schatz.
negative	She is (such) an asshole.	Sie ist (so) ein arschloch.	You are (such) an asshole.	Du bist (so) ein arschloch.
	He is (such) an idiot.	Er ist (so) ein Idiot.	You are (such) an idiot.	Du bist (so) ein Idiot.

Table 10 (continued)

	E3P	G3P	E2P	G2P
neutral	She is (such) a perfectionist.	Sie ist (so) eine Perfektionistin.	You are (such) a perfectionist.	Du bist (so) eine Perfektionistin.
	She is (such) a student.	Sie ist (so) eine Studentin.	You are (such) a student.	Du bist (so) eine Studentin.
	He is (such) a teacher.	Er ist (so) ein Lehrer.	You are (such) a teacher.	Du bist (so) ein Lehrer.
	She is (such) a girl.	Sie ist (so) ein Mädchen.	You are (such) a neighbor.	Du bist (so) ein Nachbar.
	He is (such) a guest.	Er ist (so) ein Mann.	You are (such) a guest.	Du bist (so) ein Gast.
	He is (such) a shareholder.	Er ist (so) ein Kind.	You are (such) a shareholder.	Du bist (so) ein Aktionär.

Table 11: Filler items used in the studies.

	E3P	G3P	E2P	G2P
positive	She makes great art.	Sie macht großartige Kunst.	You make great art.	Du machst großartige Kunst.
	She quickly solved our problem.	Sie hat unser Problem schnell gelöst.	You quickly solved our problem.	Du hast unser Problem schnell gelöst.
	She brings the best gifts.	Sie bringt die besten Geschenke mit.	You bring the best gifts.	Du bringst die besten Geschenke mit.
	She was there when I needed her.	Sie war da, als ich sie brauchte.	You were there when I needed you.	Du warst da, als ich dich brauchte.
	He makes the best pizza.	Er macht die beste Pizza.	You make the best pizza.	Du machst die beste Pizza.
	He gave a great speech.	Er hat eine großartige Rede gehalten.	You gave a great speech.	Du hast eine großartige Rede gehalten.
	He helped me move.	Er hat mir beim Umziehen geholfen.	You helped me a lot.	Du hast mir sehr geholfen.
	He inspired me to write a book.	Er hat mich dazu inspiriert, ein Buch zu schreiben.	You inspired me to write a book.	Du hast mich dazu inspiriert, ein Buch zu schreiben.

Table 11 (continued)

	E3P	G3P	E2P	G2P
negative	She has bad breath.	Sie hat Mundgeruch.	You have bad breath.	Du hast Mundgeruch.
	She stole my idea.	Sie hat meine Idee gestohlen.	You stole my idea.	Du hast meine Idee gestohlen.
	She stood me up again.	Sie hat mich schon wieder versetzt.	You stood me up again.	Du hast mich schon wieder versetzt.
	She is always late.	Sie kommt immer zu spät.	You are always late.	Du kommst immer zu spät.
	He is balding.	Er wird kahl.	You are balding.	Du wirst kahl.
	He can't keep deadlines.	Er kann Deadlines nicht einhalten.	You can't keep deadlines.	Du kannst Deadlines nicht einhalten.
	He drinks too much.	Er trinkt zu viel.	You drink too much.	Du trinkst zu viel.
	He is annoying.	Er ist anstrengend.	You are annoying.	Du bist anstrengend.
neutral	She is from Australia.	Sie kommt aus Australien.	You are in group B.	Du bist in Gruppe B.
	She submitted the report.	Sie hat den Bericht eingereicht.	You are in charge of part 2 of the report.	Du übernimmst Teil 2 des Berichts.
	He works as an accountant.	Er arbeitet als Buchhalter.	You work the early shift.	Du arbeitest in der Frühschicht.
	He lives in Denver.	Er lebt in Denver.	You are in room 19.	Du übernachtetest in Zimmer 19.

References

- Altenberg, Bengt. 1994. On the functions of such in spoken and written English. In Nelleke Oostdijk & Pieter de Haan (eds.), *Corpus-based Research into Language*, 223–239. Amsterdam/ Atlanta: Rodopi.
- Bolinger, Dwight. 1972. *Degree Words*. The Hague: Mouton.
- Bousfield, Derek. 2008. *Impoliteness in Interaction*. Amsterdam/New York: John Benjamins.
- Brown, Penelope & Stephen C. Levinson. 1987. *Politeness. Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Culpeper, Jonathan. 2010. Conventionalised impoliteness formulae. *Journal of Pragmatics* 42. 3232–3245.
- Culpeper, Jonathan. 2011. *Impoliteness. Using Language to Cause Offence*. Cambridge: Cambridge University Press.
- Culpeper, Jonathan & Oliver Holmes. 2013. Impoliteness and exploitative TV in Britain and North America: *The X Factor* and *American Idol*. In Nuria Lorenzo-Dus & Pilar Garcés-Conejos Blitvich

- (eds.), *Real Talk: Reality Television and Discourse Analysis in Action*, 169–198. London: Palgrave Macmillan.
- Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language: Journal of the Linguistic Society of America* 64(3). 501–538.
- Fraser, Bruce & William Nolen. 1981. The association of deference with linguistic form. *International Journal of The Sociology of Language* 27. 93–109.
- Ghesquière, Lobke & Freek van de Velde. 2011. A corpus-based account of the development of English *such* and Dutch *zulk*: Identification, intensification and (inter)subjectification. *Cognitive Linguistics* 22(4). 765–797.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago/London: The University of Chicago Press.
- Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Goldberg, Adele E. 2019. *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton/ Oxford: Princeton University Press.
- Halliday, Michael Alexander Kirkwood & Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
- Haspelmath, Martin. 2023. On what a construction is. *Constructions* 15 (1). No pagination.
- Hilpert, Martin. 2014. *Construction Grammar and Its Application to English*. Edinburgh: Edinburgh University Press.
- Hirschmann, Hagen. 2024. Das pränominal so und seine lexikalischen Varianten im Deutschen – eine syntaktische, konstruktions- und variationslinguistische Beschreibung. In Dagobert Höllein, Klaus Welke & Marc Felfe (eds.), *Regelbasierte Konstruktionsgrammatik. Musterbasiertheit vs. Idiomatizität*, 183–230. Berlin/Boston: de Gruyter.
- Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Jain, Kate H. 2022. You Hoboken! Semantics of an expressive label maker. *Linguistics and Philosophy* 45(2). 365e391.
- Jucker, Andreas H. & Irma Taavitsainen. 2000. Diachronic speech act analysis: insults from flying to flaming. *Journal of Historical Pragmatics* 1(1). 67e95.
- Kienpointner, Manfred. 1997. Varieties of rudeness: Types and functions of impolite utterances. *Functions of Language* 4(2). 251–287.
- Kienpointner, Manfred. 2018. Impoliteness online. Hate speech in online interactions. *Internet Pragmatics* 1(2). 329–351.
- Kleinke, Sonja & Birte Bö. 2015. Intergroup rudeness and the metapragmatics of its negotiation in online fora. *Pragmatics* 25(1). 47–71.
- Leech, Geoffrey. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- Manes, Joan & Nessa Wolfson. 1981. The compliment formula. In Florian Coulmas (ed.), *Conversational Routine: Explorations in Standardized Communication Situations and Prepatterned Speech*, 116–132. Mouton: The Hague.
- Oxford English Dictionary, s.v. “child (n.),” March 2024, <https://doi.org/10.1093/OED/5645033063>. (accessed 25 May 2024)
- Oxford English Dictionary, s.v. “girl (n.),” March 2024, <https://doi.org/10.1093/OED/7744946133>. (accessed 25 May 2024)
- Oxford English Dictionary, s.v. “man (n.1),” February 2024, <https://doi.org/10.1093/OED/1034670116>. (accessed 25 May 2024)

- Oxford English Dictionary, s.v. “such (adj. & pron.),” March 2024, <https://doi.org/10.1093/OED/4195345364>. (accessed 21 April 2024)
- Queisser, Angela. 2024. *Such a construction! Evidence for an exclamatory evaluative construction with pivotal such*. Heidelberg: Heidelberg University master thesis.
- Spinillo, Mariangela. 2003. On *such*. *English Language and Linguistics* 7(2): 195–210.
- Taylor, John R. 2012. *The Mental Corpus: How Language Is Represented in the Mind*. Oxford: Oxford University Press.
- Terkourafi, Marina. 2002. Politeness and formulaicity: evidence from Cypriot Greek. *Journal of Greek Linguistics* 3(1). 179–201.
- Terkourafi, Marina. 2005. Beyond the micro-level in politeness research. *Journal of Politeness Research* 1(2). 237–262.
- Traugott, Elizabeth Closs. 1982. From propositional to textual and expressive meanings: Some semantic-pragmatic aspects of grammaticalization. In Winfred P. Lehmann & Yakov Malkiel (eds.), *Perspectives on Historical Linguistics*, 245–271. Amsterdam: John Benjamins.
- Ungerer, Tobias & Stefan Hartmann. 2023. *Constructionist Approaches: Past, Present, Future*. Cambridge: Cambridge University Press.
- Van der Auwera, Johan & Kalyanamalini Sahoo. 2020. Such similatives: A cross-linguistic reconnaissance. *Language Sciences: An Interdisciplinary Forum* 81. No pagination.
- 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.

Daniel Van Olmen and Marta Andersson

9 Conventionalized impoliteness in English and Polish: The case of ‘you idiot!’

Abstract: This study argues against the dominant view in the current research that linguistic forms cannot be conventionalized for (im)politeness. As a case study, we examine a construction in English and Polish typically characterized as expressing addressee evaluation, i.e. ‘you idiot/beauty!’. However, recent work has shown that this construction is heavily biased toward genuine insults in usage and has therefore claimed that it exhibits a high level of conventionalization for impoliteness, possibly due to the pragmatic explicitness and directness of adding the second person pronoun to an address. We put this claim to the test, through a questionnaire that asks first language speakers to rate the well-formedness and (im)politeness of addresses featuring different types of nouns with ‘you’ or without it. Our results confirm the construction’s overall conventionalization for impoliteness. Addresses with evaluatively neutral nouns such as ‘reader’, for example, are found not only to be less well-formed when combining with the second person pronoun but also to be forced into an evaluative and, more specifically, impolite interpretation with ‘you’. Yet, our results contain little evidence for the hypothesis in the previous work that the second person pronoun would increase the impoliteness of negatively evaluative addresses like ‘(you) idiot!’ or for the idea in the earlier work that the Polish construction would be more conventionalized for impoliteness.

Keywords: conventionalization, English, insult, impoliteness, Polish, questionnaire

1 Introduction

Impoliteness may be characterized as (linguistic) behavior that is assessed negatively – in specific situations and against a range of different ideals (cf. Section 2.1) – and has (often intentional) offensive effects or, put differently, causes emotional ramifications such as anger and hurt for some person/people (Culpeper 2011: 23). Contrary to the prevailing view in the literature, this chapter seeks to show

Daniel Van Olmen, Lancaster University, Bailrigg, LA1 4YW, Lancaster, United Kingdom,
e-mail: d.vanolmen@lancaster.ac.uk

Marta Andersson, Uppsala University, 753 10 Uppsala, Sweden,
e-mail: marta.andersson@engelska.uu.se

that impoliteness understood in this manner *can* be conventionalized. To make our case, we will look at the English pattern instantiated by *you idiot!* and its equivalent in Polish, which are typically intended and/or perceived as genuine direct insults and can thus be said to have an offensive or negative emotional effect on the addressee(s). In this introduction, we will first review the debate about conventionalized (im)politeness (see also Van Olmen et al., this volume), then introduce the particular pattern under examination and conclude with the aims of our study.

1.1 Conventionalized (im)politeness

Classic theories of politeness (e.g. Lakoff 1974; Brown and Levinson 1987) acknowledged that politeness may be “relative to norms in a given society, group, or situation” but they maintained at the same time that it also exists “in terms of the lexicogrammatical form and semantic interpretation of an utterance” (Leech 2014: 88). In fact, their focus was very much on how specific linguistic forms relate to various maxims or principles of politeness and on how the choice of such forms can be affected by extra-linguistic factors – typically treated as invariable – like social distance and power. The field has, however, witnessed a significant shift since the discursive and post-structuralist turn in politeness studies (e.g. Eelen 2001; Mills 2003). The dominant view nowadays is that “no utterance is inherently polite” (Locher 2006: 251) and politeness is thus seen as a purely situational judgment by the interlocutors. Accordingly, the focus at present is mostly on how speech participants themselves construe politeness and construct it through discourse.

It will probably come as little surprise that the debate about “formal inherency” just presented (in an admittedly simplified manner) is present in the literature on impoliteness too (e.g. Culpeper 1996; Locher and Watts 2008). Culpeper (2011: 120–121) offers a useful evaluation of the two positions. On the one hand, it would obviously be wrong to assume that impoliteness is just a matter of form. Speakers can easily cause offence without resorting to ostensibly impolite expressions. The way in which such forms are perceived may also vary between cultures, situations and/or individuals and they are often used ironically or as banter as well (e.g. Lagorgette and Larrivée 2004 on insults as markers of solidarity). On the other hand, impoliteness is not simply a matter of discourse either. People can and do assess the (level of) impoliteness of expressions out of context. As Van Olmen and Grass (2023) show, for instance, French speakers judge the pseudo nouns¹ *plauche*

¹ We use this term to refer to words that look like real nouns in a language but do not actually exist. The way that such words are interpreted within structures, like *plauche* and *galpon* in *espèce*

and *galpon* to be offensive, even with no situational information, when they occur in *espèce de . . . !* (lit. ‘species of . . . !’). This fact suggests that there is something intrinsically impolite about the expression.

A way to reconcile the conflicting positions can be found in Terkourafi’s (2005a, 2005b) work. She argues that linguistic forms may indeed be associated with politeness: if they repeatedly combine with “particular types of context . . . as the *unchallenged* realisations of particular acts”, they can establish frames together “that create the perception of politeness” (Terkourafi 2005a: 248). Crucially, however, these frames have a variable degree of conventionalization in Terkourafi’s (2005b: 213) view, correlating with “the (statistical) frequency with which an expression is used in one’s experience of a particular context”. As experiences may diverge, the linguistic forms in such frames need not be polite for all speakers and/or in all situations. Moreover, what happens – in Neo-Gricean terms – when faced with a specific expression is that, “rather than engaging in full-blown [particularized] inferencing about the speaker’s intention, the addressee draws on that previous experience (represented holistically as a frame) to derive the proposition that ‘in offering an expression *x* the speaker is being polite’ as a generalised implicature of the speaker’s utterance” (Terkourafi 2005a: 251). Politeness would thus be this form’s assumed or preferred interpretation. Yet, this reading would still be cancelable.

Culpeper (2011: 113–154) shows that this framework can be applied to impoliteness too and adopts it to identify a range of relevant formulae in British English. One of them involves *you* plus a noun phrase functioning as an insulting address (e.g. *you idiot!*), which is the topic of the next section.

1.2 YOU+NP in English and Polish

Van Olmen, Andersson and Culpeper (2023) point out – with reference to, among others, Potts and Roeper (2006) and Corver (2008) – that this formula counts as a construction in English as well as in Polish and Dutch.² That is, what we call YOU+NP is a “conventionalized” pairing “of form and function” (Goldberg 2006: 3), combining unique grammatical properties with a distinct meaning (see Van Olmen,

de . . . !, can reveal important characteristics of those structures. For instance, if pseudo nouns receive a particular interpretation in some structure, that interpretation can only really be assumed to come from the structure itself.

² As the present chapter deals with English and Polish, Dutch will not be discussed further. Let it suffice to mention here that it behaves in more or less the same manner as the other two languages (see Van Olmen, Andersson and Culpeper 2023: 31–33).

Andersson and Culpeper 2023: 26–27). At first glance, for instance, (1a) and (1b) may look similar to the appositive patterns in (1c) and (1d) in that they are made up of a pronoun and a noun phrase. There are differences, however. The appositives in (1c) and (1d) are integrated into the clausal syntax, as subjects, in both Polish and English while (1a) and (1b) stand on their own. Moreover, YOU+NP can be singular or plural in the two languages, as (1a) and (1b) make clear, but the appositives can only be plural, as shown by (1c) versus (1e) and (1d) versus (1f). In Polish, YOU+NP also requires the noun phrase to be in vocative rather than nominative case, as the comparison of (1g) to (1a) reveals – but note that, as in (1b) to (1d), the two cases are syncretic for plural nouns.³

- (1) a. *Ty* *idioto!*
 2SG.NOM/VOC idiot.VOC.M.SG
 ‘You idiot!’
- b. *Wy* *idioci!*
 2PL.NOM/VOC idiot.NOM/VOC.M.PL
 ‘You idiots!’
- c. *My* *studenci* *jesteśmy* *inteligentni.*
 1PL.NOM student.NOM/VOC.M.PL be.1PL.PRS intelligent.NOM.PL
 ‘We students are intelligent.’
- d. *Wy* *studenci* *jesteście* *inteligentni.*
 2PL.NOM/VOC student.NOM/VOC.M.PL be.2PL.PRS intelligent.NOM.PL
 ‘You students are intelligent.’
- e. **Ja* *student* *jestem* *inteligentny.*
 1SG.NOM student.NOM.M.SG be.1SG.PRS intelligent.NOM.M.SG
 ‘I student am intelligent.’
- f. **Ty* *student* *jesteś* *inteligentny.*
 2SG.NOM/VOC student.NOM.M.SG be.2SG.PRS intelligent.NOM.M.SG
 ‘You student are intelligent.’
- g. **Ty* *idiota!*
 2SG.NOM/VOC idiot.NOM.M.SG
 ‘You idiot!’ (intended meaning)

³ We will provide glosses for Polish just in (1), since it is the only place where such grammatical information is relevant. For our other examples, translations should suffice. The abbreviations used in (1) are: 1 first person, 2 second person, M masculine, NOM nominative, PL plural, PRS present, SG singular and VOC vocative. Note also that (1e) and (1f) would be more acceptable if *student* was separated intonationally – or by commas in writing – from the rest of the clause but that the noun would then be parenthetical rather than part of an appositive structure.

YOU+NP’s semantics can be described as conveying addressee evaluation. This meaning manifests itself clearly in (2). Non-evaluative noun phrases like *rowerzysto* ‘cyclist’ in (2a) do not seem very compatible with the construction – unless the noun is modified by evaluative adjectives such as *głupi* ‘stupid’ and *dzielny* ‘brave’ in (2b).

- (2) a. ? *Ty rowerzysto!*
 ? ‘You cyclist!’
 b. *Ty głupi/dzielny rowerzysto!*
 ‘You stupid/brave cyclist!’

Example (2b) also shows that addressee evaluation need not actually be negative in YOU+NP. This fact raises the question why Culpeper (2011) and numerous others (e.g. Teleman, Andersson and Hellberg 1999: 797; Ooms and Van Keymeulen 2005: 63–64; Finkbeiner, Meibauer and Wiese 2016: 4) nevertheless regard the construction as an impoliteness formula.

In line with the above understanding of conventionalization, Van Olmen, Andersson and Culpeper (2023: 28) argue that the answer lies in the use of YOU+NP: “If we can establish that, in actual usage, the construction (most) frequently serves impolite purposes, it is not unreasonable to assume that language users generalise over such instances and there exists a schema [or frame] – alongside a more abstract evaluative one – where the form YOU+NP is associated with the ‘function’ of impoliteness.” To this end, they examine 200 random corpus attestations of the construction for each language. A quick look at the nouns appearing in their data is already highly suggestive (see Van Olmen, Andersson and Culpeper 2023: 32–33). The five most common ones, presented in (3),⁴ are clearly all negatively evaluative.⁵

- (3) a. *idiot* (16), *bastard* (11), *bitch* (5), *fucker* (5), *moron* (5)
 b. *idiot(k)a* ‘idiot’ (13), *świnia* ‘swine’ (8), *chuj* ‘fucker’ (7), *drań* ‘bastard’ (7), *dupek* ‘asshole’ (4)

4 When different Polish forms are given with the same translation, like *idiot(k)a* ‘idiot’ in (3) and *sąsiedzie/sąsiadko* ‘neighbor’ in (5), they are simply the masculine and feminine variants of the noun.

5 One reviewer wishes to know how many of the 200 cases per language contain negatively evaluative nouns. Van Olmen, Andersson and Culpeper (2023: 31), however, do not look at nouns in isolation but consider whole noun phrases (*girl* on its own may not be overtly evaluative but, when combined with *dumb*, for instance, it does convey negative evaluation). Moreover, they explicitly refrain from giving frequencies for evaluative versus non-evaluative noun phrases because of difficulties in distinguishing them. It is therefore only possible to provide the rough estimate that, in both languages, circa 80% of the noun phrases are negatively evaluative.

The study goes further, though, and analyzes the co-text of every hit to determine whether it is truly impolite or, in words reminiscent of Culpeper (2011: 11–12), whether there is sufficient evidence that it is intended and/or taken to have negative emotional ramifications for the addressee. An in-depth discussion of this co-textual evidence is beyond the scope of the present chapter but the underlined parts in (4) should give the reader a good idea of the types of indications considered (see also Van Olmen, Andersson and Culpeper 2023: 29–30).⁶

(4) a. meta-linguistic comments

*usmiechnol sie do mnie szyderczo gdy widzial ze policja mnie powstrzymuje, krzyknołem “ty pedale!” a on do mnie “**ty heteryku!**” :/ od kiedy heteryk to cos zlego? :| ale skoro on mnie tka obraza to uwaza ze to cos zlego* (plTenTen19-390136)

‘he smirked at me when he saw that the police were stopping me, I shouted “you faggot!” and he said to me “**you heterosexual!**” :/ since when is being straight something wrong? :| but since he is insulting me in this way, he must think it’s wrong’

b. impoliteness responses (e.g. challenging; Culpeper, Bousfield and Wichmann 2003: 1563)

*I’m surprised at your arrogant post hasn’t gotten you flamed yet; you certainly deserve to be, **you dolt**. – I don’t see how I would be considered a dolt and the post was not arrogant.* (enTenTen18-35133812)

c. narrative insights into the interlocutors’ intent and/or mental state

*Bassam explained that the Border Police soldiers were driving by the school in Anata, taunting the children by saying, “Come out, **you heroes**.” . . . They routinely use the loudspeakers to yell profanity at homes while on patrol.* (enTenTen18-13452138)

d. co-occurring acts (e.g. threats)

Zostawcie tą biedną dziewczynę w spokoju! Albo pokażę wam, co to jest prawdziwy BÓL wy chorzy degeneraci! (plTenTen19-1264337)

‘Leave that poor girl alone! Or I’ll show you what real PAIN is, you sick degenerates!’

The analysis reveals that English YOU+NP exhibits a strong predisposition in usage to impoliteness and its Polish counterpart an even stronger one, with 75% of the former’s hits and 92% of the latter’s resembling those in (4) (these numbers do not

⁶ The corpus examples in this chapter all come from Van Olmen, Andersson and Culpeper (2023), who relied on the multilingual TenTen corpus family (Jakubíček et al. 2013) for their investigation.

even include the respective 12% and 4% of attestations that contain negatively evaluative noun phrases but are employed in a non-impolite way, to “talk dirty” in sexually charged interactions or as banter; Van Olmen, Andersson and Culpeper 2023: 33–36) (see also Culpeper, Van Dorst and Gillings, this volume, for new corpus data on English YOU+NP).

These findings justify calling YOU+NP an impoliteness construction in Van Olmen, Andersson and Culpeper’s (2023) view. The frequency data does not explain, however, *why* this addressee evaluation construction, which *may* be positive too after all (e.g. *you extraordinary beauty!*), is so biased toward impoliteness in English and even more so in Polish. Van Olmen, Andersson and Culpeper (2023: 37) believe that the presence of the second person pronoun plays a crucial role here. While, strictly speaking, it may be somewhat redundant in an address (already marked by an intonation break, for instance), it does serve to openly ascribe the noun phrase’s meaning to the addressee. Making the target clear in this way can be said to make the address pragmatically more explicit and thus direct (see Culpeper and Haugh 2014: 170). Such directness is often eschewed on account of politeness (cf. Brown and Levinson’s 1987: 131 strategies to “impersonalise S[peaker] and H[earer] and “avoid the pronouns ‘I’ and ‘you’”). If one wishes to insult someone, by contrast, “explicitly associat[ing]” them “with a negative aspect” (Culpeper 2005: 41), like ‘you’ does in YOU+NP, is very effective (see Dobrushina, this volume, for a similar argument about ‘you’ in Nakh-Daghestanian curses). Van Olmen, Andersson and Culpeper (2023: 37) go as far as hypothesizing that the construction “does the job of hurting the addressee’s feelings better than an offensive address that does not contain a second person pronoun” but concede that “this conjecture obviously needs to be tested”. They also discuss possible reasons why the impact of ‘you’ may be especially pronounced in Polish. For one, as it is a pro-drop language, the nominative-vocative pronouns *ty* ‘you’ and *wy* ‘you all’ do not appear very often and their overt expression immediately evokes strong interpersonal emphasis, which lends itself to impoliteness. Moreover, Polish makes a distinction between informal *ty/wy* and formal *Pan* and variants (see Piskorska 2023). The fact that YOU+NP features the former pronouns means that it may be comparatively acceptable for addressing people close to the speaker but, when directed at others, the construction may come with an extra layer of irreverence or even condescension.

1.3 Present study

As mentioned in the preceding section, Van Olmen, Andersson and Culpeper’s (2023) hypothesis about the difference between ‘idiot!’ and ‘you idiot!’ deserves to

be checked. The same is true for some of the predictions that they make based on their results (see Van Olmen, Andersson and Culpeper 2023: 37). They expect, for instance, that *YOU+NP* will tend to be perceived as impolite even with evaluatively neutral nouns (e.g. ‘reader’) and pseudo nouns (e.g. ‘sprim’) out of context. In other words, the frequency-determined frame associated with the construction will force not just an evaluative interpretation on such nouns but a negatively evaluative one in particular (cf. Jain 2022: 389). They also suggest, given that *YOU+NP* seems almost exclusively impolite in Polish, that these tendencies may be stronger in this language than in English and, implicitly, that the effects could even arise with positively evaluative nouns (e.g. ‘angel’).

Putting these claims to the test is what the present chapter seeks to do after this introduction. More specifically, we will examine, through a questionnaire, how compatible *YOU+NP* in English and Polish is with different types of nouns and what impact adding *YOU* to such nouns has on their interpretation. The methodology for this study will be discussed in Section 2 (see also Queisser and Pleyer, this volume, for a very similar approach). Section 3 will present our results. In Section 4, finally, we will give our conclusions.

2 Methodology

The present section will first describe the design of the questionnaire. We will move on to the instructions given to the participants next and then to the way in which the data was collected. The section ends with the details of the statistical analysis.

2.1 Design

The questionnaire consists of thirty scenarios.⁷ Since we are explicitly interested in judgments on *YOU+NP* out of context, as they can reveal much about the construction (see Section 1.1), each provides the same minimal situational information of “imagine that someone addresses you in this way”. The scenarios do vary, of course, in what the second person is actually called. They feature – both with and without ‘you’ – the negatively evaluative nouns in (5a), the positively evaluative

⁷ The English and Polish surveys in their entirety can be accessed at <https://forms.gle/PXmA-FDSGxmBgBMkK7> and <https://forms.gle/rC9sBrjtSy8gcxnB7> respectively (both last accessed on 09/10/2024).

ones in (5b), the evaluatively neutral ones in (5c), the pseudo nouns in (5d) and the inanimate nouns in (5e). This nominal diversity will allow us to see, for example, whether YOU+NP goes as well with positive *and* negative evaluation in Polish as in English, whether the construction forces an impolite reading onto non-evaluative and pseudo nouns or whether it can even create addresses out of nouns that do not normally characterize people (see Section 1.3). The choice of the specific nouns results from extensive deliberation, based in part on the findings of Van Olmen, Andersson and Culpeper (2023), about terms that have comparable de- and connotations in the two languages.

- (5) a. *(ty) debilu!* – *(ty) degeneracie/degeneratko!*
 (you) moron! – *(you) degenerate!*
 b. *(ty) aniele!* – *(ty) słodziaku*
 (you) angel! – *(you) sweetie!*
 c. *(ty) czytelniku/czytelniczko* – *(ty) sąsiedzie/sąsiadko*
 (you) reader! – *(you) neighbour!*
 d. *(ty) sprimie* – *(ty) wabie*
 (you) sprim! – *(you) wabe!*
 e. *(ty) butelko* – *(ty) garnku*
 (you) bottle! – *(you) pot!*

The list in (5) accounts for twenty scenarios. The other ten contain filler pairs – like *Wasza/Moja Wysokość!* ‘Your/My Highness!’, *(mój) panie/(moja) pani!* ‘(my) Sir/Madame!’ and *mój/drogi głupku!* ‘my/dear fool!’ – to obscure the focus of the questionnaire to some degree. In addition, all of these scenarios appear in an order that makes any direct comparison of the members of a pair more difficult. For example, the first half of the survey includes one negatively evaluative noun with ‘you’ (*ty degeneracie/degeneratko!* ‘you degenerate!’ in scenario 11) and one without it (*debilu!* ‘moron!’ in scenario 2) and the second half their counterparts (*degeneracie/degeneratko!* ‘degenerate!’ in scenario 24; *ty debilu!* ‘you moron!’ in scenario 19).

For each scenario, our survey has two questions, as the mock example in (6) shows. The first one asks, like in (6a), for an assessment, on a seven-point scale, of the well-formedness of the address. It will enable us to test how compatible YOU+NP is with different types of nouns in English and Polish and thus check intuitions like

that in (2a), i.e. that non-evaluative nouns are somewhat strange in the construction (see Section 1.2).

- (6) Imagine that someone addresses you in this way: “You idiot!”
- a. How natural/well-formed is *you idiot!* in this case? (Remember that 4 stands for “neither unnatural/ill-formed nor natural/well-formed”.)
- | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|--------------------------------|---|---|---|---|---|---|---|-------------------------------|
| very unnatural/very ill-formed | 0 | 0 | 0 | 0 | 0 | 0 | 0 | very natural/very well-formed |
- b. How unkind/hurtful/etc. or kind/complimentary/etc. is the individual saying *you idiot!* in this case? (Remember that 4 stands for “neither unkind/hurtful/etc. nor kind/complimentary/etc.”)
- | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|--|---|---|---|---|---|---|---|---|
| very unkind/very hurtful/very bad-natured/very uncivil/very impolite | 0 | 0 | 0 | 0 | 0 | 0 | 0 | very kind/very complimentary/very good-natured/very civil/very polite |

The second question, like in (6b), asks for a judgment, on a seven-point scale, about the (im)politeness of the address. It will allow us to test, for instance, whether negatively evaluative nouns are considered (even) more impolite when they occur in YOU+NP or whether evaluatively neutral nouns are seen as (more) impolite in the construction (see Sections 1.2 and 1.3). What is important to note here, though, is the insight from the discursive and post-structuralist research that the interpretation of terms like “polite” and “impolite” is not stable at all. For that reason, our survey does not just use these labels in the questions (and the instructions; see Section 2.2). Following Oliver (2023: 134) and others, the questionnaire also captures (im) politeness: (i) with ‘kind/unkind’ as an assessment of behavior for signaling a close/distant relationship (cf. Brown and Levinson 1987 and Culpeper 1996 on positive (im)politeness); (ii) with ‘complimentary/hurtful’ as an appraisal of the costs and benefits of conduct to others (cf. Culpeper and Tantucci 2021 on the principle of (im)politeness reciprocity); (iii) with ‘good-natured/bad-natured’ as an evaluation of a person’s innate character (cf. Kádár 2017 on (im)politeness and morality); and, lastly, (iv) with ‘civil/uncivil’ as an appraisal of the adherence to some conventional code of conduct (cf. Sifianou 2019 on the connection between (im)politeness and (in)civility).

2.2 Instructions

The participant information page of the survey warns people that it contains language that some may find offensive. The page also alerts potential participants of the fact that there are a number of optional demographic questions about age, gender, education, employment and languages. This information is gathered for two reasons. On the one hand, it enables us to remove individuals who compromise the comparability of the English and Polish participant groups (see Section 2.3). On the other hand, it allows us to delete the data of anyone who wishes to withdraw from the study within four weeks of taking part (no such request was received, however).

The instructions themselves firstly inform participants that they will be presented with a range of short scenarios and illustrate them with one featuring *szczęśliwiarzu!* ‘lucky duck!’ as the address in the format of (6). Participants are also told that, “if a scenario includes multiple gendered forms (e.g. ‘waiter/waitress!’), [they] are encouraged to consider only the form that [they] think applies to [them] for [their] answers”. Then, we introduce, through our illustration, the questions and the ways to respond to them. The first one is said to ask participants “to assess how well-formed ‘lucky duck!’ is linguistically as a way of addressing someone in [their] language”. For the sake of clarity, we also rephrase the question: “How natural do you think it is in your language to call someone ‘lucky duck!’ when you talk to them?”. The answer is described as requiring the selection of “a score on a 7-point scale, where 1 stands for ‘very unnatural/very ill-formed’, 7 for ‘very natural/very well-formed’ and 4 for ‘neither unnatural/ill-formed nor natural/well-formed’”. The second question is said to ask participants “to assess to what extent the person saying ‘lucky duck!’ to [them] is being kind/complimentary/good-natured/civil or unkind/hurtful/bad-natured/uncivil”. It too gets rephrased, as “how polite do you think the speaker of ‘lucky duck!’ is?”. The answer is characterized as expecting participants “to pick a score from 1, which means ‘very unkind/very hurtful/. . .’, to 7, which means ‘very kind/very complimentary/. . .’”, with 4 standing “for ‘neither unkind/hurtful/. . . nor kind/complimentary/. . .’”.

The final guidelines are of a more general nature. Participants are told that there are no correct or incorrect answers: “This questionnaire is not a test of your knowledge of any rules of [your] language or . . . culture. We are interested in your own linguistic intuitions and judgments, not in what other people, institutions, style guides or grammars might think.” They are also instructed to respond as instinctively as possible and to avoid changing any scores given. We furthermore stress that there is no time limit to the survey and encourage participants to use the whole seven-point scale: “For instance, if you believe that a particular expression is not especially well-formed but not impossible either, you may want to consider

assigning a score of 2 or 3. Similarly, if you believe that a certain expression is more kind than unkind but not especially kind, you may want to consider assigning a score of 5 or 6.”

2.3 Data collection

The data for English was collected in two ways.⁸ The first author invited his own undergraduate students at Lancaster University – the 2022 first- and second-year cohorts in May 2023 and the 2023 first-year cohort in October 2023 – to complete the questionnaire. He also asked linguist-colleagues at other British universities in June 2023 to circulate a call among their undergraduates. All students were told that the survey was looking for judgments on the well-formedness and degree of (im)politeness of a range of English expressions by first language speakers. We also stressed that they were entirely free to take part or not, that the questionnaire was anonymous and that non-participation would not affect their studies or grades in any way. The survey received a total of fifty-seven responses. Five were removed, however: two because the participant did not identify as a first language speaker of English, three because the participant reported that they were not a student. This last decision was mainly taken to make sure that the English and Polish groups were as similar as possible. The fifty-two remaining participants were all born between 2000 and 2005, 71.15% of them identified as female and 88.46% said that they were studying linguistics and/or a modern language.

To get the data for Polish,⁹ the second author asked colleagues at universities in Poland in March 2023 to distribute an invitation to take part in the survey among their undergraduates. The message said that we were interested in judgments by first language speakers on the well-formedness and degree of (im)politeness of

⁸ For which ethical approval was obtained by the first author from the Faculty of Arts and Social Sciences Research Ethics Committee at Lancaster University in April 2023. Thanks are due to João Almeida (University of Glasgow), Federica Formato (University of Brighton), Robbie Love (Aston University), Carmen Ríos García (University of Liverpool) and Ellen Smith-Dennis (University of Warwick) for their help in disseminating the English survey.

⁹ In accordance with Swedish law at the time of data collection, given that the study does not directly deal with or process potentially sensitive personal data in Sweden, alongside the intended storage of data at Lancaster University, the survey was deemed exempt from ethics clearance by the second author's affiliations of Umeå University and Uppsala University. Thanks are due to Agata Rozumko (University of Białystok), Adam Gład (Maria Curie-Skłodowska University in Lublin), Adam Wojtaszek (University of Silesia), Sylwia Karolak (Adam Mickiewicz University in Poznań), Dariusz Szczukowski (University of Gdańsk) and Łukasz Książyk (University of Warsaw) for their help in disseminating the Polish questionnaire.

a range of Polish expressions. We also again emphasized that participation was voluntary, anonymous and unconnected to the student’s studies. The questionnaire received one hundred and twenty-six responses, of which twenty-eight were excluded: two participants did not identify as first language speakers of Polish, three were born in the 1970s and 1980s and twenty-three stated that they were not students. The other ninety-eight had an overall profile comparable to that of our English participants. They were all born between 1995 and 2004, 73.47% of them identified as female and 94.90% reported that they were studying linguistics and/or a modern language.

For the sake of clarity, Table 1 summarizes and compares the information of our Polish and English participants.

Table 1: Comparison of questionnaire participants.

Language	# Respondents	# Excluded	Year of birth	% Female-identifying	% Language degree
English	57	5	2000–2005	71.15	88.46
Polish	126	28	1995–2004	73.47	94.90

2.4 Analysis

We will provide the following descriptive statistics for our questionnaire results: means (μ) and standard deviations (σ). The former capture the average score for well-formedness or level of (im)politeness of a specific type of noun with or without the second person pronoun. The latter measure the variation between the scores given by all participants for this type of noun with or without ‘you’: if the standard deviation is low, those scores tend to be close to the mean; if it is high, they are more dispersed (see Rasinger 2013: 134–136).

For the comparison of two scores (e.g. the (im)politeness of positively evaluative nouns with or without the second person pronoun), we will use two-tailed paired t-tests. They assess whether the mean scores differ significantly from each other or not, also taking into consideration their standard deviations. Our t-tests are two-tailed because we do not always have clear expectations about the direction of the difference between scores (see Baayen 2008: 81). They are paired because we always compare data from the same group of participants (see Rasinger 2013: 200). When contrasting one score to multiple others (e.g. in the post-hoc analysis after an analysis of variance; see below), our standard level of significance will undergo Bonferroni correction, dividing it up by the number of comparisons conducted. This adjustment minimizes the likelihood of overvaluing any particular test result

with a p-value below 0.05, since it may simply arise by chance amidst numerous tests (see Baayen 2008: 114).

For well-formedness in particular, we will also need to contrast several scores at the same time (e.g. different kinds of nouns with or without the second person pronoun) and we can use an analysis of variance – ANOVA, for short – to do so (see Rasinger 2013: 209). Our ANOVAs are of the type with repeated measures since we always compare data from the same group of participants (see Baayen 2008: 264). They are also of the two-way variety since we wish to test two separate variables (i.e. noun type and presence/absence of ‘you’; see Rasinger 2013: 210–217).¹⁰

3 Results

3.1 Overview

Table 2 gives, for both English and Polish, the means and the standard deviations for the well-formedness and the degree of (im)politeness of the five different types of nouns (see Section 2.1) *with* the second person ([+you]) and *without* it ([–you]). In the rest of this section, we will first examine the findings for well-formedness and then those for (im)politeness.

Table 2: All questionnaire results.

Noun type	[±you]	Well-formedness				(Im)politeness			
		English		Polish		English		Polish	
		μ	σ	μ	σ	μ	σ	μ	σ
negatively evaluative	[–you]	5.06	1.77	4.65	2.03	1.90	0.97	1.73	0.89
	[+you]	5.47	1.61	5.15	1.93	1.92	0.96	1.69	1.16
positively evaluative	[–you]	5.27	1.53	5.22	1.72	5.58	1.24	5.96	1.11
	[+you]	4.92	1.76	4.68	1.97	5.94	1.21	6.00	1.19
evaluatively neutral	[–you]	3.85	1.73	5.71	1.51	4.33	0.83	5.18	1.18
	[+you]	2.05	1.32	2.74	1.89	3.62	0.93	3.61	1.29
pseudo	[–you]	1.85	1.16	1.65	1.11	3.45	0.88	3.08	1.15
	[+you]	2.47	1.62	1.81	1.26	3.17	0.97	2.96	1.19
inanimate	[–you]	2.03	1.26	1.81	1.19	3.48	0.80	3.07	0.99
	[+you]	2.95	1.58	1.94	1.22	2.97	0.84	2.90	1.03

¹⁰ The statistical analysis was performed in SPSS 29 (IBM Corp 2022).

3.2 Well-formedness

Let us begin with the results for English, for which Figure 1 presents the mean scores in a more accessible way (note, again, that 1 stands for ‘very ill-formed’, 7 for ‘very well-formed’ and 4 for ‘neither ill-formed nor well-formed’).

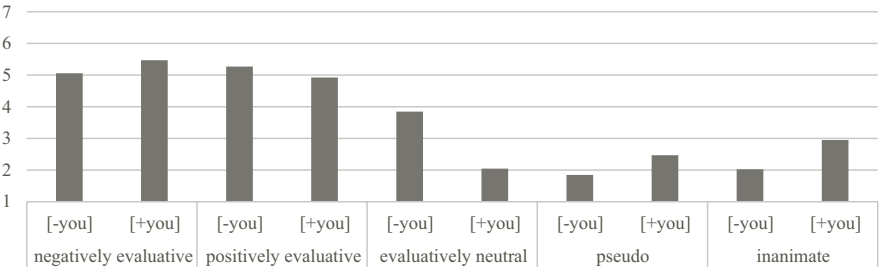


Figure 1: Well-formedness in English.

An initial observation concerns the pseudo and inanimate nouns. With scores ranging between 1.85 ($\sigma = 1.16$) and 2.95 ($\sigma = 1.58$), they are clearly judged to be ill-formed by virtually every participant, whether *you* is present or not. They differ in this regard from the other nouns and will therefore be treated separately here. It is, of course, hardly surprising that any address with words like *sprim* and *bottle* is regarded as unnatural: referring to someone with a fake term or as a thing is just strange. Interestingly, however, when such nouns occur in *YOU+NP* in English, their well-formedness as addresses does improve in a statistically significant way ($p < 0.001$ for the two t-tests). The pseudo ones go from 1.85 ($\sigma = 1.16$) to 2.47 ($\sigma = 1.62$) and the inanimate ones from 2.03 ($\sigma = 1.26$) to 2.95 ($\sigma = 1.58$). While they continue to be seen as ill-formed, the construction can be argued to make an address interpretation at least somewhat more plausible. This phenomenon may simply be due to the presence of the second person pronoun, explicitly assigning the word’s meaning to the other person. It could also be attributed to the semantics of *YOU+NP*, though, with the construction implying that a pseudo noun such as *sprim* must have some evaluative sense and that an inanimate noun like *bottle* is intended as a kind of assessment of the addressee.

For the negatively and positively evaluative and evaluatively neutral nouns, our two-way ANOVA (see Section 2.4) allows us to compare the six relevant means at the same time and to test the impact and potential interplay of the factors along which the data varies, i.e. noun type and presence/absence of *you*. The results reveal that both the former ($F(2,206) = 122.03$, $p < 0.001$) and the latter ($F(1,103) = 28.49$, $p < 0.001$) have an independent effect and that their interaction is significant

too ($F(2, 206) = 35.78, p < 0.001$). Subsequent t-tests contrasting the multiple pairs of means enable us to identify where the specific differences lie. As discussed in Section 2.4, these so-called post-hoc t-tests do need to meet a higher level of significance, through Bonferroni correction, since one mean score ends up being compared separately to several other scores.

The post-hoc t-tests show that the evaluatively neutral nouns are seen as less well-formed than the evaluative ones, whether the second person pronoun is there or not. Without *you*, the former appear to be neither ill-formed nor well-formed – i.e. a score of 3.85 ($\sigma = 1.73$) – while the latter are clearly quite natural – i.e. scores of 5.06 ($\sigma = 1.77$) and 5.27 ($\sigma = 1.53$) ($p < 0.001$ for all t-tests). With *you*, the negatively and positively evaluative nouns remain well-formed – i.e. respectively 5.47 ($\sigma = 1.61$) and 4.92 ($\sigma = 1.76$) – whereas the evaluatively neutral ones become ill-formed – i.e. 2.05 ($\sigma = 1.32$) ($p < 0.001$ for all t-tests). Especially the first difference between the noun types, when the second person pronoun is absent, was not expected. We acknowledge, however, that, out of context, calling someone *moron* or *sweetie* is more natural than addressing someone as *reader*. Such an evaluatively neutral noun may need more specific circumstances to truly work (e.g. a writer appealing to their audience) while our survey tried to keep the scenarios constant with minimal situational information. The second difference between the noun types, when *you* is present, confirms the intuition (and judgments in the literature) that *YOU+NP*, as an addressee evaluation construction, is not very compatible with evaluatively neutral noun phrases.

The post-hoc t-tests provide further support for this last statement in that the evaluatively neutral nouns are the only ones where we see a statistically significant drop in well-formedness, from 3.85 ($\sigma = 1.73$) to 2.05 ($\sigma = 1.32$), when the second person pronoun is added ($p < 0.001$). The positively evaluative nouns do not seem affected by the presence or absence of *you*, with scores of 4.92 ($\sigma = 1.76$) and 5.27 ($\sigma = 1.53$) respectively ($p > 0.05$). The negatively evaluative nouns, by contrast, appear to be seen as more well-formed addresses in *YOU+NP* than on their own, with respective scores of 5.47 ($\sigma = 1.61$) and 5.06 ($\sigma = 1.77$) – although, with a p-value of 0.008, the difference does not reach the required Bonferroni-corrected significance level here. It is still tempting, of course, to interpret this result as reflecting some kind of special connection between negative evaluation (and thus impoliteness) and *YOU+NP*.

For well-formedness in Polish, Figure 2 presents the means scores in Table 2 in a reader-friendly fashion. From the chart, it is immediately clear that, like in English, the pseudo nouns and the inanimate ones stand out: regardless of the presence or absence of *ty*, they are regarded as ill-formed. For that reason, they are discussed separately here too.

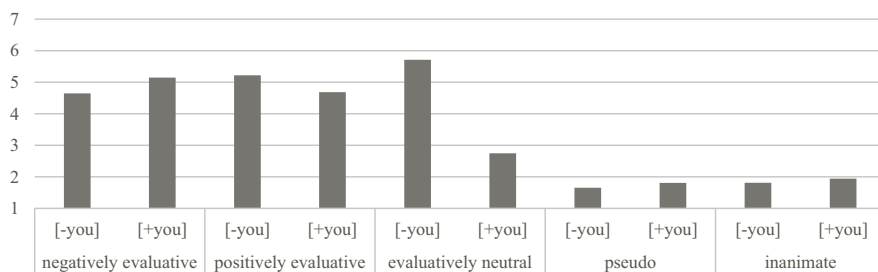


Figure 2: Well-formedness in Polish.

To explain the low scores, we can again appeal to the inherent strangeness of addressing someone with a pseudo noun like *wabie* or a noun referring to an object such as *garunku* 'pot'. The situation in Polish is not entirely the same, however. Unlike in English, there are no significant differences between addresses with or without *ty*. The pseudo nouns are rated as 1.65 ($\sigma = 1.11$) on their own and as 1.81 ($\sigma = 1.26$) in *YOU+NP*; the inanimate ones as 1.81 ($\sigma = 1.19$) and 1.94 ($\sigma = 1.22$) respectively ($p > 0.05$ for both t-tests). This result may tell us something about our two proposed accounts of the variation in English: it is probably not *YOU+NP*'s evaluative meaning that makes the pseudo and inanimate nouns in this language somewhat more well-formed addresses, since one would then expect their Polish equivalents to exhibit similar behavior. The more likely explanation is therefore *YOU+NP*'s function of explicitly attributing the noun to the addressee. A possible reason why it does not increase the well-formedness of the Polish pseudo and inanimate nouns is that they are already overtly assigned to the other person on their own, through vocative case.

For the other nouns, the ANOVA indicates that there are independent effects for the presence/absence of *ty* ($F(1,195) = 204.83$, $p < 0.001$) as well as the type of noun ($F(2,390) = 14.27$, $p < 0.001$), which also interact significantly ($F(2,390) = 185.68$, $p < 0.001$). The post-hoc Bonferroni-corrected t-tests make clear that the evaluatively neutral nouns are again unusual. Like in English, they are essentially judged to be ill-formed when occurring in *YOU+NP* – i.e. 2.74 ($\sigma = 1.89$) – and significantly less well-formed than all other addresses ($p < 0.001$ for all t-tests). This result shows that the construction is not particularly compatible with non-evaluative nouns in Polish either. Unlike in English, however, the evaluatively neutral nouns on their own score substantially higher for well-formedness – i.e. 5.71 ($\sigma = 1.51$) – than any other address. An explanation for this finding is that the vocative probably forces people to imagine circumstances in which something like *czytelniku/czytelniczko!* 'reader!' would work, even if no actual context is provided.

The post-hoc t-tests also reveal interesting differences *between* the evaluative nouns in Polish. With scores reliably above 4, they may all be seen as well-formed addresses, whether the second person pronoun is present or not, but the positively evaluative nouns display a significant decrease in well-formedness, from 5.22 ($\sigma = 1.72$) to 4.68 ($\sigma = 1.97$), when *ty* is inserted ($p < 0.001$) while the negatively evaluative ones exhibit a significant increase, from 4.65 ($\sigma = 2.03$) to 5.15 ($\sigma = 1.93$), when combining with *ty* ($p < 0.001$). Polish resembles English in the latter (though the difference does not meet the required level of significance there) but is distinct when it comes to the former. In other words, negatively evaluative nouns seem especially well-suited for *you+NP* in Polish but, relatively speaking, this appears to be less the case for positively evaluative ones in the language. This finding could be taken to reflect the construction's extremely high degree of conventionalization for impoliteness in Polish. In this regard, it is probably also worth pointing out that addresses like *ty aniele* 'you angel' are deemed less well-formed than addresses like *ty debilu* 'you moron' – i.e. 4.68 ($\sigma = 1.97$) and 5.15 ($\sigma = 1.93$) respectively – even if the p-value of 0.006 does not reach the necessary Bonferroni-corrected level.

3.3 (Im)politeness

Figure 3 charts the means for (im)politeness in English (recall that 1 stands for 'very impolite', 7 for 'very polite' and 4 for 'neither impolite nor polite').

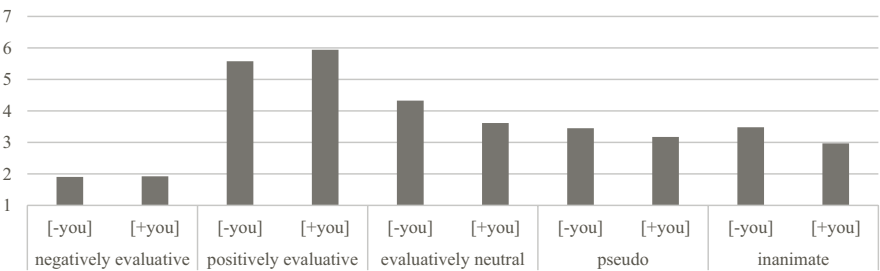


Figure 3: (Im)politeness in English.

Let us again start with the pseudo and inanimate nouns. Perhaps unsurprisingly, they are judged to be impolite, irrespective of the presence or absence of *you*: addressing a person with a would-be word and especially as a thing is likely to be interpreted as offensive. Interestingly, though, these nouns *are* seen as even more impolite when they appear in *you+NP*. The pseudo ones go from 3.45 ($\sigma = 0.88$) to

3.17 ($\sigma = 0.97$) ($p < 0.01$ for the t-test); the inanimate ones from 3.48 ($\sigma = 0.80$) to 2.97 ($\sigma = 0.84$) ($p < 0.001$). These facts may be taken to reflect the construction’s usage-based conventionalization: the offensiveness of calling someone *wabe* or *pot* is strengthened by YOU+NP’s link with impoliteness.

This association comes to the fore with the evaluatively neutral nouns in particular. On their own, they score 4.33 ($\sigma = 0.83$) but, in YOU+NP, they drop significantly to 3.62 ($\sigma = 0.93$), on the impolite side of the spectrum ($p < 0.001$). The way in which we understand this observation is as follows: (i) although the construction is not very compatible with evaluatively neutral nouns (see Section 3.1), it forces an evaluative reading onto them to make them fit; and (ii) while, strictly speaking, this reading could be either positive or negative, people rely on their knowledge/experience of YOU+NP and thus tend to assign a negative one.

For the evaluative nouns, finally, it is hardly remarkable, of course, that the negative ones are deemed impolite and the positive ones polite. They exhibit some unexpected behavior too, however. For one, the hypothesis that the directness of *you* would increase the impoliteness of insults (see Section 1.2) is not borne out: the negatively evaluative nouns, such as *moron*, have similar scores with or without the second person pronoun – i.e. respectively 1.92 ($\sigma = 0.96$) and 1.90 ($\sigma = 0.97$) ($p > 0.05$). The positively evaluative nouns, such as *angel*, do see a change but it is an increase in politeness when they feature in YOU+NP, from 5.58 ($\sigma = 1.24$) to 5.94 ($\sigma = 1.21$) ($p < 0.001$). We have no immediate explanation for this result. At one point, Van Olmen, Andersson and Culpeper (2023: 37) implicitly suggest that the pragmatic explicitness of ‘you’ might be able to enhance the politeness of positively evaluative nouns too: “It is entirely imaginable that someone wishing to evaluate another person in a positive way may also want . . . to overtly ascribe their assessment to their addressee.” It still remains unclear then why no equivalent effect is observable with negatively evaluative nouns. Perhaps, the fact that, comparatively, they are rated as more impolite in any case than the positively evaluative nouns are rated as polite simply means that any effect of the second person pronoun is bound to be minimal. It is not entirely inconceivable that we might still see an effect of ‘you’ with negatively evaluative nouns that are less strong than *moron* and *degenerate* (e.g. *dumbo*).

To conclude the present section, consider the means for (im)politeness in Polish in Figure 4. We can discern a number of tendencies in this language that are similar to those in English. First, the pseudo and inanimate nouns are generally deemed impolite, whether *ty* is present or not. Second, the evaluatively neutral nouns are forced into an impolite interpretation when they occur in YOU+NP: they score 5.18 ($\sigma = 1.18$) on their own but 3.61 ($\sigma = 1.29$) when combining with the second person pronoun ($p < 0.001$). Third, the negatively and positively evaluative nouns are judged to be impolite and polite respectively. Fourth, no evidence exists for a

directness effect of ‘you’ in the negatively evaluative nouns: they are rated as 1.73 ($\sigma = 0.89$) without *ty* and as 1.69 ($\sigma = 1.16$) with *ty* ($p > 0.05$).

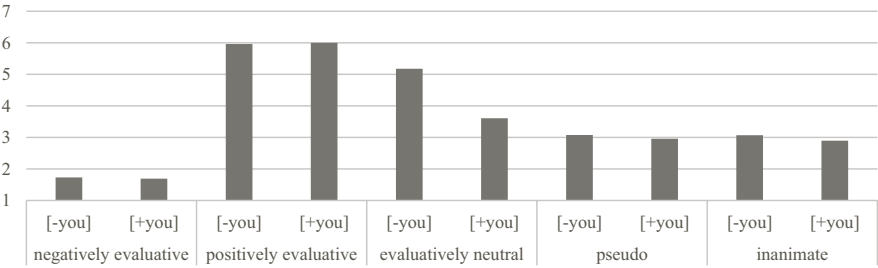


Figure 4: (Im)politeness in Polish.

There are also two important differences with English. First, the level of politeness of the positively evaluative nouns does not vary with the second person pronoun’s presence or absence – i.e. respectively 6.00 ($\sigma = 1.19$) and 5.96 ($\sigma = 1.11$) ($p > 0.05$). In other words, we have no consistent evidence for a possible politeness-boosting effect of ‘you’ with nouns like ‘angel’ either, which may be due to the lack of politeness associated with *ty* (see Section 1.2). Moreover, one might have expected that Polish *YOU+NP*’s especially strong association with impoliteness could influence positively evaluative nouns too, perhaps evoking ironic or sarcastic interpretations when they combine with *ty* (cf. Van Olmen and Grass’s 2023 results for French *ange!* ‘angel!’ versus *espèce d’ange!* ‘you angel!’). But they appear to be resistant to any such hypothetical pressure. Second, unlike in English, the pseudo nouns do not become significantly more impolite in *YOU+NP*: they are rated as 3.08 ($\sigma = 1.15$) without *ty* and 2.96 ($\sigma = 1.19$) with *ty* ($p > 0.05$). For some reason, the construction only affects the inanimate nouns here, which go from 3.07 ($\sigma = 0.99$) on their own to 2.90 ($\sigma = 1.03$) in *YOU+NP* ($p < 0.01$). A potential reason is that, in Polish, coupling a pseudo noun with vocative marking makes the noun’s semantic oddness feel more nonsensical than insulting. In English, by contrast, the lack of overt morphology may more easily activate the insult interpretation, even when the noun is opaque.

4 Conclusions

YOU+NP has been described in the literature as a construction conveying addressee evaluation. Our results support this characterization for English as well as for Polish. In both languages, the construction is found to be well-formed as an address

with positively and negatively evaluative nouns but ill-formed with evaluatively neutral ones. Our English data in particular may be taken to point to *YOU+NP*'s nature as an address in yet another way. The fact that the pseudo and inanimate nouns become somewhat more well-formed when combining with *you* suggests that the construction has at least some ability to create or improve an address interpretation of unlikely terms of address. This ability is probably due to *YOU+NP*'s function of overtly assigning the noun to the addressee, through the second person pronoun, rather than to its evaluative nature. Otherwise, the pseudo and inanimate nouns in Polish, which already achieves explicit addressee attribution by way of the vocative in any case, would be expected to exhibit the same behavior as their English equivalents.

However, according to Van Olmen, Andersson and Culpeper (2023), *YOU+NP* is not *just* a construction expressing addressee evaluation. They argue, based on usage data from corpora, that it is strongly conventionalized specifically for impoliteness in English and even more so in Polish. Our results confirm the general claim about *YOU+NP* as an impoliteness construction in these languages in three ways. For one, when evaluatively neutral nouns combine with 'you', they are found to lose their (a)polite reading and to acquire an impolite one in both English and Polish. One would anticipate such nouns to be made compatible with the (evaluative) construction somehow but the fact that they tend to be interpreted as impolite can, in our view, only be accounted for by (speakers' knowledge and experience of) *YOU+NP*'s usage-based conventionalization for impoliteness. We believe that it also explains the result that addresses with inanimate nouns in English and Polish (as well as with pseudo nouns in English) are seen as even more impolite when they contain 'you'. A last finding relevant here is that, in both languages, the well-formedness as addresses of negatively evaluative nouns – unlike that of positively evaluative ones, for example – actually increases when they occur in the construction.

That said, the more specific claim about Polish *YOU+NP*'s stronger association with impoliteness gets little support from our findings. One might have hypothesized, for instance, that even positively evaluative nouns would acquire a (sarcastic) impolite reading when combining with *ty* but they are, in fact, deemed as polite with the second person pronoun as without it. Moreover, given how rarely Polish *YOU+NP* features such nouns in corpus data, one might have expected to find evidence of a certain incompatibility between the construction and positively evaluative nouns. The fact that addresses with these types of nouns are seen as somewhat less well-formed with the second person pronoun than without it in Polish (but not in English) could be an indication. We should bear in mind, though, that any address with a positively evaluative noun is still very much well-formed in the language. In other words, while the partiality in usage suggests that there is a very strong schema or frame associating *YOU+NP* with negative evaluation and

impoliteness in Polish, the language still also has a schema of the construction for “unbiased” addressee evaluation.

Finally, to explain YOU+NP’s conventionalization for impoliteness, Van Olmen, Andersson and Culpeper (2023) appeal to the notion of pragmatic explicitness. In their view, spelling out the target of an evaluation, through ‘you’, makes it more direct and such directness is well-suited for impoliteness but may be avoided for politeness. They even hypothesize that insults with the second person pronoun would therefore be more impolite than those without it. There are, however, no signs of such an effect in our findings: addresses with negatively evaluative nouns are as impolite with ‘you’ as without it in both English and Polish. Importantly, we do not believe that this result necessarily invalidates the general argument about the relationship between directness and impoliteness (see also Culpeper 2011: 183–193). It just means that ‘you idiot!’ is not more impolite than ‘idiot!’. The question what the actual difference between the two then is remains to be answered, of course.¹¹

References

- Baayen, R.H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- Brown, Penelope & Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Corver, Norbert. 2008. Uniformity and diversity in the syntax of evaluative vocatives. *Journal of Comparative Germanic Linguistics* 11(1). 43–93.
- Culpeper, Jonathan. 1996. Towards and anatomy of impoliteness. *Journal of Pragmatics* 25(3). 349–367.
- Culpeper, Jonathan. 2005. Impoliteness and entertainment in the television quiz show: The Weakest Link. *Journal of Politeness Research* 1(1). 35–72.
- Culpeper, Jonathan. 2011. *Impoliteness: Using Language to Cause Offence*. Cambridge: Cambridge University Press.
- Culpeper, Jonathan & Michael Haugh. 2014. *Pragmatics and the English Language*. Basingstoke: Palgrave.

¹¹ To our knowledge, few other attempts have been made to offer an explanation. d’Avis and Meibauer (2013: 197) do rightly point out (for German and Swedish) that YOU+NP differs from a regular address like ‘(hey,) dad, . . .’ in that it cannot really be used to get someone’s attention and serves instead “to confirm the addressee-status [in the social world] of the person spoken to”. This distinction does not seem to be especially relevant for ‘(you) idiot!’, however: it is difficult to see how such a negatively evaluative noun could function as a way to obtain the attention of a particular addressee in any situation.

- Culpeper, Jonathan & Vittorio Tantucci. 2021. The principle of (im)politeness reciprocity. *Journal of Pragmatics* 175. 146–164.
- Culpeper, Jonathan, Derek Bousfield & Anne Wichmann. 2003. Impoliteness revisited: With special reference to dynamic and prosodic aspects. *Journal of Pragmatics* 35(10/11). 1545–1579.
- d'Avis, Franz & Jörg Meibauer. 2013. *Du Idiot! Din idiot!* Pseudo-vocative constructions and insults in German (and Swedish). In Barbara Sonnenhauser & Patrizia Noel Aziz Hanna (eds.), *Vocative! Addressing between System and Performance*, 189–217. Berlin: De Gruyter Mouton.
- Eelen, Gino. 2001. *A Critique of Politeness Theories*. Manchester: St Jerome.
- 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: Benjamins.
- Goldberg, Adele. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- IBM Corp. 2022. *IBM SPSS Statistics for Windows, Version 29.0*. Armonk: IBM Corp.
- Jain, Kate H. 2022. You Hoboken! Semantics of an expressive label marker. *Linguistics and Philosophy* 45(2). 365–391.
- Jakubíček, Miloš, Adam Kilgariff, Vojtech Kovár, Pavel Rychlý & Vit Suchomel. 2013. The TenTen Corpus Family. *Proceedings of the International Corpus Linguistics Conference 7*. 125–127.
- Kádár, Dániel Z. 2017. *Politeness, Impoliteness and Ritual: Maintaining the Moral Order in Interpersonal Interaction*. Cambridge: Cambridge University Press.
- Lagorgette, Dominique & Pierre Larrière. 2004. Interprétation des insultes et relations de solidarité. *Langue française* 144. 83–103.
- Lakoff, Robin. 1973. The logic of politeness: Or, minding your p's and q's. In Claudia Corum, T. Cedric Smith-Stark & Ann Weiser (eds.), *Papers from the Ninth Regional Meeting of the Chicago Linguistic Society*, 292–305. Chicago: Chicago Linguistic Society.
- Leech, Geoffrey. N. 2014. *The Pragmatics of Politeness*. Oxford: Oxford University Press.
- Locher, Miriam A. 2006. Polite behavior within relational work: The discursive approach to politeness. *Multilingua* 25(3). 249–267.
- Locher, Miriam A. & Richard J. Watts. 2008. Relational work and impoliteness: Negotiating norms of linguistic behaviour. In Derek Bousfield & Miriam A. Locher (eds.), *Impoliteness in Language: Studies on Its Interplay with Power in Theory and Practice*, 77–99. Berlin: Mouton de Gruyter.
- Mills, Sara. 2003. *Gender and Politeness*. Cambridge: Cambridge University Press.
- Oliver, Samuel J. 2023. *A Corpus-based Approach to (Im)Politeness Metalanguage: The Case of Shakespeare's Plays*. Lancaster: Lancaster University PhD dissertation.
- Ooms, Miet & Jacques Van Keymeulen. 2005. *Vlaams Brabants en Antwerps*. Tielt: Lannoo.
- Piskorska, Agnieszka. 2023. Has madam read Wilson (2016)? A procedural account of the T/V forms in Polish. *Pragmatics* 33(3). 486–504.
- Potts, Christopher & Thomas Roeper. 2006. The narrowing acquisition path: From expressive small clauses to declaratives. In Ljiljana Progovac, Kate Paesani, Eugenia Cassiellas & Ellen Barton (eds.), *The Syntax of Nonsententials: Multi-Disciplinary Perspectives*, 183–201. Amsterdam: Benjamins.
- Rasinger, Sebastian M. 2013. *Quantitative Research in Linguistics: An Introduction*. London: Bloomsbury.
- Sifianou, Maria. 2019. Im/politeness and in/civility: A neglected relationship? *Journal of Pragmatics* 147. 49–64.
- Teleman, Ulf, Erik Andersson & Staffan Hellberg. 1999. *Svenska Akademiens Grammatik*. Stockholm: Norstedts.

- Terkourafi, Marina. 2005a. Beyond the micro-level in politeness research. *Journal of Politeness Research* 1(2). 237–262.
- Terkourafi, Marina. 2005b. 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: De Gruyter.
- 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.
- Van Olmen, Daniel & Delphine Grass. 2023. *Espèce de linguiste!* An impoliteness construction in French? Paper presented at the 18th International Pragmatics Conference, Brussels, 9–14 July, 2023.

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\ ") ("you")* </u>) (<u>"you" [pos="JJ".*"]* [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	39.4	0.667
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.

Conclusion

Marina Terkourafi

11 *What's in a word?* Reflections about impoliteness and future directions

Abstract: This short piece, prompted by the chapters in this volume, considers the question whether impoliteness is more a matter of the form of particular expressions (what the expression looks like) or of the meaning that they convey (what the expression means). A new distinction between ‘conventions of form’ and ‘conventions of meaning’ is proposed to capture these two possibilities, which can be accounted for under two different understandings of conventionalization, as lexicalization (verbatim repetition of a form across contexts) or as constructionalization (possibility of replacement by synonymous expressions leading to abstraction away from specific lexical items). Both options have been entertained in the literature and the chapters in this volume offer the tools to approach this fascinating question at the heart of im/politeness research with renewed theoretical and methodological rigor.

Keywords: conventionalization, constructionalization, lexicalization, conventions of form, conventions of meaning

1 Reflections

During an online lecture at the start of the academic year in August 2020, University of Southern California professor of clinical business communication Greg Patton was explaining to students the concept of pause or filler words, that is, words such as *err*, *umm* and *you know* in English, which speakers use to buy themselves time while they are planning what to say next. Patton, who had by that time been coordinating the university's programme in China/Korea relations for several years, went on to illustrate the concept with an example from Chinese, saying: “In China, the common pause word is ‘that, that, that’. So in China, it might be *na-ge, na-ge, na-ge*.” The phonological proximity of the Chinese word to the N-word in English, despite the avowed lack of any semantic/meaning relationship between the two, was enough to raise protests by students, in response to which the professor stepped down from teaching the course and the University issued a statement that: “It is

Marina Terkourafi, Leiden University, Reuvensplaats 3–4, 2311 BE Leiden, the Netherlands, e-mail: m.terkourafi@hum.leidenuniv.nl

simply unacceptable for the faculty to use words in class that can marginalise, hurt and harm the psychological safety of our students”.¹

While the chronological proximity of this incident to the murder of George Floyd by US police in May 2020 and the country-wide protests that followed in the wake of the Black Lives Matter movement certainly heightened recipients’ sensitivity to this particular phonological sequence, the occurrence of the incident alone raises questions whose ramifications extend beyond the narrow interests of im/politeness researchers: to what extent can impoliteness be inscribed in particular words, such that merely producing a similar sounding phonological sequence can be perceived as offensive and impolite ‘by association’? Far-fetched as it may sound,² the possibility that linguistic form, separated from meaning, can directly cause offence cannot be excluded, as this incident suggests.

Incidents such as this fly in the face of the recent trend in im/politeness studies to insist that politeness and impoliteness do not lie in forms but in how speakers use them. And while the last part of this statement, “how speakers use [linguistic forms]”, can be understood in at least two ways – namely, as emphasizing individual speaker intentions (and their recognition by listeners) or as highlighting the (more or less abstract) contexts of occurrence of linguistic forms – the question of whether im/politeness is a matter of semantics or pragmatics looms large in this discussion.

The question is not new. In a way, it spans the history and even the pre-history of the field, in that the earliest treatises on the topic – manuals dating back to the 15th century – typically associated politeness and impoliteness with particular forms, creating the impression that, so long as certain words are used or not used, politeness or impoliteness is guaranteed. This attention to form survives to this day in the casual teaching of politeness to children, who are often told to use the “magic word” (*please*, in English), as if the word by itself is enough to guarantee positive reception of their request. It took the field decades, if not centuries, to move away from this understanding of politeness as something inscribed in specific linguistic forms (which had the added effect of creating the impression that a closed set of such forms can be identified, the rest of the language being irrelevant to im/politeness assessments) and inject growing doses of contextual sensitivity into our analyses. And now, the pendulum is swinging back . . . ?

Not exactly. The volume’s answer to the thorny question whether im/politeness is a property of linguistic expressions (semantic) or a matter of how people

¹ Source: <https://www.bbc.co.uk/news/world-asia-china-54107329> (last accessed 30/05/2025).

² Admittedly, the professor did not lose his position and, after internal investigation, it was concluded that he had not acted in bad faith, suggesting that the university’s response was more a knee-jerk reaction in a tense situation than recognition of intentional wrong-doing.

use them (pragmatic) is a nuanced one – pretty much like the distinction between semantics and pragmatics itself. Indeed, this distinction has become more nuanced in recent years, with increasing attention paid to the middle ground between semantics and pragmatics, which is where notions such as generalized conversational implicatures, explicatures, and conversational defaults reside. However, in delivering this nuanced answer, the preceding chapters break new ground, in that they emphasize what is linguistic (rather than psychological or social) about impoliteness, and, crucially, how we can use linguistic theories and techniques to study it. The volume can thus be seen as a manifesto ‘reclaiming’ im/politeness for linguistic theory, and in so doing reminding us of the intellectual roots of the field in early attempts, such as Robin Lakoff’s (1973) rules of politeness, to explain what makes utterances pragmatically “well-formed”, this phrasing alone echoing the prevailing syntactic concerns of the time. Half a century later, the question remains current, with generative syntax being represented in the current volume by the chapter by Davis and Jang on the anti-honorific prefix *-che* in Korean. Their analysis places impoliteness (or: anti-honorification, as they call it) firmly on the semantic side, although they also allow for the possibility of sarcastic honorification handled by pragmatics, a matter they leave to future research.

Other chapters adopt a Construction Grammar approach, which suits well both their quantitative methodologies and the gradient nature of their findings. In Construction Grammar and similar usage-based or lexicalist frameworks, there is no strict division between semantics and pragmatics. Rather, the unit of analysis is the sign, which may be smaller or larger than a conventional word and carries several features at the same time. These features include information about the sign’s combinatorial possibilities, be those with other signs (syntax and semantics) or with aspects of the context, including the speaker’s illocutionary goals and extra-linguistic/social context (cf. Copestake and Terkourafi 2010). Signs combine with other signs through a process of unification, which is where eventual conflicts between their features are resolved. In other words, a sign is a combination of form (at its most basic, a string of sounds) and meaning (understood as the sign’s combinatorial possibilities with other signs and with extra-linguistic contexts). On the meaning side, it makes no difference whether these properties are called ‘syntax’, ‘semantics’, or ‘pragmatics’ and no one feature has priority over the other. Moreover, because signs carry meaning of their own that goes beyond the meanings of their component parts, lexicalist frameworks are well suited to formalizing the patterns discovered by many of the chapters, making it no surprise that they are the framework of choice for several of them (see chapters by Finkbeiner, Van Huyssteen, Breed and Pilon, and Queisser and Pleyer).

Besides being theoretically diverse, the volume is also methodologically diverse, featuring data drawn from corpora, metapragmatic comments found in textbooks, ethnographic fieldwork, and online experiments. These methodological choices are dictated in part by the wide range of language varieties represented, which include both well documented and less studied varieties as well as different periods of these varieties, making different resources and methods appropriate for studying them. Beyond its descriptive interest for the varieties studied, this typological diversity is an asset in itself, as it enables answers to some very interesting questions raised by the juxtaposition of studies in this volume – a task to which I turn next.

2 Future directions

Given this theoretical, methodological and typological richness, it is not surprising that the volume also opens up a wealth of new directions to explore. A first direction relates to the role of coercion in generating impolite meanings and in helping us identify the relevant constructions. Since first popularized by Michaelis (2004), the notion of coercion captures the idea that multi-word expressions can ‘force’ the meaning of component words that appear in them to align with the expression’s overall meaning, no matter what the component word’s own stored meaning might otherwise be. In impolite or insulting constructions, then, a word will become negative ‘by association’. Convincing empirical evidence for this is offered in the chapter by Queisser and Pleyer, and it has been proposed that coercion can be used as a criterion to decide whether an expression is an impoliteness formula in the first place (Giomi, Van Olmen and Van Oers 2025). This claim, in turn, generates a further, very interesting question, namely, whether the same applies to politeness. Is politeness equally lexicalized (carried by fixed lexical forms) or is lexicalization primarily a feature of impoliteness? And if so, why? Although adversarial contexts (such as army training; Culpeper 1996) exist in which impoliteness is the norm, politeness and impoliteness are definitely not symmetrical in many respects, including their cognitive salience and overall frequency of occurrence; so there is no reason to expect that they will be symmetrical when it comes to lexicalization either. The question, however, whether impoliteness is more lexicalized and thus more encoded in specific lexemes than politeness, is beyond the current volume’s focus in impoliteness and necessitates a comparative treatment of the two phenomena to be answered.

The notion of coercion is related to conventionalization, a theme that runs through the volume as a whole. In my work on politeness, I defined conventionalization as a three-way relationship between an expression, a user, and a context

of use (Terkourafi 2001: 130), from which it follows that conventionalization is a matter of degree that varies by user and for the same user across different contexts and over time: an expression which is conventionalized for me relative to a context (*qua* frequent in that context in my experience until now), may not be conventionalized for me at this time in a different context. As such, encountering that expression in a different context will require extra processing effort from me to decide on its import. As a next step in this process, an expression can spread across contexts, gaining frequency in a 'lateral' way. This process of expansion of the range of contexts in which an expression is used, described in Terkourafi (2015a) using the example of *My bad!* for (urban, cool-sounding) apologies in US English, is also the result of conventionalization but this time at a higher (supra-individual) level. Crucially, on this community-level understanding of conventionalization, the expression becomes increasingly detached from both nonce (one-off or local) contexts and individual users. This second step in the conventionalization process corresponds to Morgan's (1978) "conventions of usage", which can in turn generate, as a third step, "conventions of the language". One possible path for this is grammaticalization realized as phonological reduction (Terkourafi 2001: 158–160; think of the specialization of American English *wanna VP?* for invitations/suggestions, or of the English greeting *goodbye* originating in the wish *God be with you*). What is important for our purposes is that the lexical material in these cases fuses but is not replaced by semantically synonymous material – so the expression remains stable or fixed in its lexical make-up. Conventionalization as described in this paragraph may thus be considered a matter of "conventions of form".

That is, however, not the only way in which conventionalization can be understood. Further to lexicalization leading to more formulaic realisations (seen in repetition of the same form, as above), conventionalization can also be understood as generalisation *away* from specific lexemes leading to more *abstract* constructions. On this second understanding, what becomes conventional (fixed) is not the form that realizes a certain meaning (or illocutionary function) but rather the meaning (or illocutionary function) itself – "conventions of content" is thus a more suitable label to describe the outcome of conventionalization in this second sense. This is the way in which, for instance, the illocutionary repertoires of different communities can be seen to emerge.

Interestingly, this second understanding of conventionalization can also be diachronically explained using a Construction Grammar account: conventionalization of content would then be akin to "constructionalization" (Traugott and Trousdale 2013), which occurs when not (only) contexts but (also) particular lexical forms are gradually abstracted away from and replaced by abstract grammatical categories such as Verb, Noun etc. that come to fill (at least some) slots in the construction. Which of these aspects of conventionalization is responsible for creating the impo-

lite effect each time? Since both can be explained using a Construction Grammar approach, the use of this framework by several of the chapters in the volume cannot help us advocate between these two options.

The volume puts its finger on this problem right from the start, when, on the first page of the Introduction, we read that “no account of (im)politeness can be complete without a thorough understanding of the role of actual linguistic form in it”. But what exactly *is* the role of form in im/politeness assessments? Is a form impolite because of *what it looks like*, as a “badge” signaling belonging in a community of practice, this signaling alone being enough to produce the positive or negative evaluation that is im/politeness (as argued in Terkourafi 2002: 186; 2005a: 293)? Or is a form impolite because of *what it means*, that is, because it encodes certain values such as a general directive to Raise Other and Lower Self as predicted, for instance, in Leech’s (2014: 90) General Strategy of Politeness?

Both views have been defended in the literature. Regarding the first possibility, explaining the frame-based approach to politeness, in Terkourafi (2019) I argued that, when it comes to the linguistic expressions that are paired with certain extra-linguistic categories in a frame,

what matters is not so much what they mean but the sheer fact that they are uttered and how, including their intonational contour and potentially other phonetic detail. [...] What does the work in such cases is their form (the *signifier*) rather than their meaning (*what is signified*) so we don’t need to be actively communicating solidarity or avoiding imposition through the semantic meaning of our words in order to be perceived to be polite. Rather, what we have is an interaction between frequency of occurrence and face-constituting, as exemplar theorists describe it, such that each time the expression-context combination, the frame, is accessed in either encoding or decoding, its face-constituting potential is reconfirmed. [...] This treating of linguistic expressions as forms opens up fascinating new avenues for investigating the power of words as signifiers [...] – quite independently of the speaker’s intention *and* of the expression’s semantic meaning. This is a dimension that is often overlooked in the mentalizing literature on implicatures and speech acts, and is only occasionally captured in anthropological linguistic work on indexicality. Awareness of this dimension is sometimes found in analyses of magical language, which can be characterized as being performative (bringing about changes in the world), while its propositional /referential functions can remain obscure to its direct users (Tambiah 1968).

On the other hand, the possibility of forms being impolite because of what they encode/mean (their *signifieds*) is captured in Geoffrey Leech’s maxim-based view, whose content-full nature I highlighted in Terkourafi (2015b: 959–960). “By attributing to the maxims explicit semantic content,” I argued,

Leech is claiming (as he did in 1983) that politeness is about maximizing the expression of polite beliefs, and correspondingly minimizing the expression of impolite beliefs. This is reiterated in the current formulation of the general principle of politeness, which states that ‘in

order to be polite, S [i.e. self, speaker] expresses or implies meanings that associate a favorable value with what pertains to O[ther] or associate an unfavorable value with what pertains to S[elf] (2014: 90). While this makes the current scheme flexible enough to account for impoliteness as well [. . .], it also makes politeness a matter of what we say, rather than, as is often claimed, of how we say it. [. . .]. While the maxims of politeness can motivate the conventionalization of particular expressions, conventionalization would be driven by their semantics rather than their grammatical form, allowing also for nonconventionalized instances of politeness to be accounted for under the same scheme. And while Leech is clearly not suggesting that the polite beliefs encapsulated in the maxims should be sincerely held [. . .], this is a most interesting proposition that can also be tested crosslinguistically.

Put in a nutshell, the two possibilities outlined above might be summarized as follows: are linguistic forms actually processed semantically and is it their meaning that does the face constituting (Leech 1983, 2014)? Or are they merely registered as forms achieving belonging and recognition as a member of a group not through what they mean but through what they look like (whereby synonymous forms do not have the same politeness import) in a quasi-magical fashion (Terkourafi 2015a, 2019)? The contributions in this volume give us renewed impetus to tackle this question.

A first possibility is to look to the diachrony of grammatical expressions of impoliteness. With respect to this, we may ask: considering that conventionalization is a cline (a matter of degree), how is the impoliteness potential of expressions affected as they become more conventionalized? Do we observe a process comparable to semantic bleaching (whereby greater conventionalization translates to less/attenuated offence) or is the opposite the case (with greater conventionalization resulting in greater/aggravated offence)? These matters can be empirically investigated and the answer could help advocate between the two options outlined earlier and which of the two, conventions of content or conventions of form, carries the burden of im/politeness work: if greater conventionalization of form bleaches the impoliteness effect, then impoliteness may be carried more by contents/signifieds (which are taken to be less sincerely expressed the more formulaically they are worded, since less effort is expended in producing/processing them); if, conversely, greater conventionalization of form aggravates the impoliteness effect, then impoliteness may be carried more by forms/signifiers (in that the signifier alone produces offence directly, unmediated by what it means or even the ability to process what it means).

A preliminary answer to this question may be found in the chapter by Culpeper, Van Dorst and Gillings, who found that “the more productive, more abstract formulae (more dependent on grammar as opposed to specific words) attract slightly lower impoliteness scores”. While this might be specific to impoliteness (see the discussion of coercion above), the finding that the more abstract the formula, the

less impoliteness it carries, may be seen as supporting the view that impoliteness is inscribed in forms (understood as *signifiers*) rather than their contents – confirming the “conventions of form” or ‘magical’ (locutionary) view above.

However, other findings in the volume can be seen to point in the opposite direction. The common finding, for instance, in previous research as well as in many of the chapters in this volume (see chapters by Dobrushina, Paternoster, Queisser and Pleyer, and Van Olmen and Andersson), that explicit mention of the target of the offence (‘you’) heightens its impoliteness import suggests the existence of cross-linguistic tendencies that could be seen as supporting the second, “conventions of content” or *signified*-based view. While in some cases we could be simply dealing with more surface-like phenomena due to borrowing, a possibility explored in the chapter by Van Huyssteen, Breed and Pilon, a more fascinating prospect is that of a universal grammar of impoliteness. More typological/comparative research on impoliteness is clearly needed to uncover further cross-linguistic tendencies that would justify looking for the underlying reasons for this, if that turns out to be the case. If cross-linguistic tendencies are discovered that go beyond the explicit mention of ‘you’, we could indeed be dealing with the third-level of im/politeness research envisaged by Terkourafi (2005b), at which theories such as Brown and Levinson’s (1987) may be seen to operate. Such theories would then be capturing underlying motivations (in Brown and Levinson’s theory: positive and negative face) predicting the range of expressions that can do im/politeness work, with different possibilities becoming conventionalized (in the ‘conventions of form’ sense) in different languages. Indeed, herein lies a possible compromise position between the two conventionalization scenarios outlined above.

Yet, even the question of a universal grammar of impoliteness does not mark the end of the line for impoliteness research. Because, even if a more semantic, signified-based, “conventions of content” view of impoliteness ends up being supported by the typological evidence, we could still ask whether linguistic forms (such as certain moods, grammatical persons, numbers etc.) encode impoliteness directly or rather what they encode is an altogether different meaning from which impoliteness (but also politeness) can be generated. Such a possibility is proposed, for instance, for the semantics of the diminutive by Jurafsky (1996), who used cross-linguistic evidence to organize the meanings of the diminutive in a radial category from which both endearing (polite) and denigrating (impolite) meanings can be generated. To come full circle to the example that opened these reflections, even the N-word, when used by members of the in-group, in certain interactional moments and with certain phonological features, can serve solidarity functions (Rahman 2011), as has also been shown for other expressions, such as *re malaka* (lit. ‘hey, wanker’) in Greek (Vergis and Terkourafi 2015). As Dobrushina and Paternoster, in the respective chapters in this volume, emphasize, the findings regarding impo-

liteness are no more than tendencies and even a curse such as Italian *che ti venga . . .*, ‘may you get . . .’ has “rare benedictive uses”. Impoliteness, in other words, continues to sit stubbornly on the fence between semantics and pragmatics and the questions it raises are sure to keep us busy for years to come.

Acknowledgements: I would like to thank the editors for the original invitation to serve as discussant to the panel on the grammar of impoliteness that they organized at the 56th Annual Meeting of the Societas Linguistica Europaea, and subsequently to contribute to the volume that resulted from that meeting, which gave me the opportunity to reflect on questions that I had been thinking about for some time.

References

- Brown, Penelope & Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Copestake, Ann & Marina Terkourafi. 2010. Conventional speech act formulae: From corpus findings to formalization. In Peter Kühnlein, Anton Benz & Candace Sidner (eds.), *Constraints in Discourse 2*, 125–140. Amsterdam: Benjamins.
- Culpeper, Jonathan. 1996. Towards an anatomy of impoliteness. *Journal of Pragmatics* 25(3). 349–367.
- Giomì, Riccardo, Daniel Van Olmen & Denise van Oers. 2025. Insults as a sentence type: A cross-linguistic perspective. Paper presented at the 10th International Workshop on Functional Discourse Grammar, Schoorl, 2–5 July, 2025.
- Jurafsky, Dan. 1996. Universal tendencies in the semantics of the diminutive. *Language* 72. 533–578.
- Lakoff, Robin. 1973. The logic of politeness or minding your p's and q's'. In Claudia Corum & T. Cedric Smith-Stark (eds.), *Papers from the Ninth Regional Meeting of the Chicago Linguistic Society*, 292–305. Chicago: Chicago Linguistic Society.
- Leech, Geoffrey. 1983. *Principles of Pragmatics*. London: Longman.
- Leech, Geoffrey. 2014. *The Pragmatics of Politeness*. Oxford: Oxford University Press.
- Michaelis, Laura. 2004. Type shifting in construction grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics* 15(1). 1–67.
- Morgan, Jerry L. 1978. Two types of convention in indirect speech acts. In Peter Cole (ed.), *Syntax and Semantics 9: Pragmatics*, 261–280. New York: Academic Press.
- Rahman, Jacquelyn. 2011. The N word: Its history and use in the African American community. *Journal of English Linguistics* 40(2). 137–171.
- Tambiah, Stanley J. 1968. The magical power of words. *Man* 3(2). 175–208.
- Terkourafi, Marina. 2001. *Politeness in Cypriot Greek: A Frame-based Approach*. Cambridge: Cambridge University PhD dissertation.
- Terkourafi, Marina. 2002. Politeness and formulaicity: Evidence from Cypriot Greek. *Journal of Greek Linguistics* 3. 179–201.
- Terkourafi, Marina. 2005a. Identity and semantic change: Aspects of T/V usage in Cyprus. *Journal of Historical Pragmatics* 6(2). 283–306.

- Terkourafi, Marina. 2005b. Beyond the micro-level in politeness research. *Journal of Politeness Research* 1(2). 237–262.
- Terkourafi, Marina. 2015a. Conventionalization: A new agenda for im/politeness research. *Journal of Pragmatics* 86. 11–18.
- Terkourafi, Marina. 2015b. Review of *The Pragmatics of Politeness*, by Geoffrey Leech, Oxford University Press, 2014. *Language* 91(4). 957–960.
- Terkourafi, Marina. 2019. The frame-based approach to politeness: Rationale and future directions. Plenary talk at the 16th *International Pragmatics Conference*, Hong Kong, 12 June 2019.
- Traugott, Elizabeth C. & Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.
- 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.

Index

- address 21, 45, 53–54, 56, 100–101, 105, 109, 114,
123, 133, 143–144, 158–161, 169, 171, 256,
262, 283, 285, 289–293, 297–300, 302–304
- addressee evaluation 283, 287, 289, 298,
302–304, 312
- affixoid 32–33, 35–37, 40
- Afrikaans 215–219, 223, 225–226, 228–229, 231,
235–237, 239, 241, 245–246
- Avar 103–104, 107, 109–111, 113, 116, 121–123
- banter 4, 155, 166, 257, 267, 278, 284, 289, 309,
317–318, 326–327, 329–332
- banter principle 317
- benediction 20, 133, 150, 155, 167–168, 347
- blasphemy 129, 136–137, 149, 153, 171, 222
- blessing 14, 18, 95–97, 99–114, 116, 118–124, 133,
167–170
- British National Corpus 2014 307–309, 315, 319
- coercion 15, 20, 253–254, 258, 260, 262–263, 267,
269, 271, 276–277, 342
- combining form 33, 35, 37
- compound constituent 32–33, 35, 37–39, 57
- concordance analysis 326
- conditional 179–188, 190–191, 195–198, 206,
208–209, 220
- construction 11–13, 38, 171, 180–182, 210–211,
217–225, 227–231, 233–237, 241–247, 254,
258–259, 260–262, 277, 285–289, 301–304,
312, 343
- construction, syntactic 38, 179, 181–182, 210–211
- Construction Grammar 11–12, 182, 210, 218, 228,
254–255, 258, 341, 343–344
- Construction Grammar, Diasystematic 12, 22,
218, 228
- constructional calque 215, 218–219, 229, 245, 247
- context-dependency 179, 184, 210–211, 259
- conventionalization 13–15, 18, 20–22, 96, 124,
132–133, 138, 170–171, 182, 210, 254–256,
283–285, 287, 300–301, 303–304, 308–309,
311, 331–332, 339, 343–346
- conventionalized impoliteness formula 96, 131,
171, 184, 254, 257, 307–308, 312, 318–320,
325, 327, 329
- corpus 14–16, 20–21, 33, 39, 53, 96, 102–103,
129, 138–139, 156, 171–172, 182, 191–192,
220–222, 225–226, 261, 278, 307–309,
314–316, 332–333
- Corpus Diacronico dell'Italiano 129, 138–152,
154, 156–157, 159, 161–163, 165–168, 171
- corpus linguistics 14, 215, 245, 311
- corpus linguistics, contrastive 215
- corpus search query 20, 103, 191, 307, 315,
320–322
- corpus tagging 319–320, 332
- covarying collexeme analysis 219, 241–244
- CQP syntax 319–325
- creativity 156, 163, 222–223, 227, 235, 240,
257, 262
- criticism 32–33, 48, 51, 53, 57, 131, 149, 313, 321
- curse 4, 10, 14, 17–18, 21, 95–97, 99–124,
129–142, 144–167, 169–172, 188, 234–235,
238, 314, 323, 347
- definite article 219, 225, 227–228
- deixis, person 200
- derision 32–33, 41, 48–50, 53, 57
- desemanticization 36, 38
- diachrony 13, 18, 21–22, 116, 260, 343, 345
- Discourse Grammar 11–12
- drama 139, 152, 156
- embedded clause 181–182, 186, 190–191, 198,
200, 209
- emotive 6, 148–149, 162, 216, 220, 256
- English 4–5, 10, 14, 17–18, 31–35, 37–39, 41, 53,
55, 57–58, 72–73, 82, 95–96, 171, 179–181,
183–184, 187, 215–219, 221–226, 228–231,
235–237, 239, 241, 245–247, 253–257, 259,
264, 266, 271, 273–274, 276–277, 283–286,
288–291, 293–295, 296–304, 307–308, 312,
314–315, 333, 339–340, 343
- ergative 99, 104, 108, 115, 119
- evaluative 11, 14–15, 18–19, 33–34, 36, 54, 124,
191, 253–254, 256–258, 260–263, 267–277,
283, 287, 289–292, 295–304, 312, 344
- expressive 19, 34, 36, 134, 180, 183, 190, 216,
260–261, 314, 322–325, 328–329, 331, 333

- frame 133, 285, 287, 290, 303, 311, 344
 frequency 13–14, 19, 32–33, 35–36, 40–47, 103,
 156, 182, 184, 216, 219–226, 228–237, 240,
 244–247, 262–264, 278, 285, 289–290, 309,
 311, 314, 317, 321–322, 328, 342–344
 Functional Discourse Grammar 10, 12, 19
 futurity 10, 134, 136, 184–188, 190, 197, 206

 Generative Grammar 9–10, 12, 19, 22, 64, 72, 341
 German 5, 33–34, 36, 179–182, 184, 187–192,
 209–211, 219, 253–255, 257, 262–264, 266,
 270–272, 274, 276–278, 304
 grammar/pragmatics interface 211
 grammaticalization 13, 21–22, 96, 115, 327, 343
 grawlix 227, 233, 237, 245

 honorificity 61–68, 71, 78–81, 83, 85, 87–91,
 149, 341
 honorificity, anti- 61–62, 66–67, 78, 85–89,
 91, 341
 humor 216, 221–222, 235–236, 240, 244,
 246–247, 263, 318

 illocutionary indicator 179, 181–182
 imperative 17, 58, 99, 100, 102, 108, 111, 114,
 118–121, 123, 180, 185, 261
 impoliteness, definition of 5–8, 35–36, 67–68,
 97, 133, 183, 216, 255–256, 283–284,
 292, 310
 impoliteness, inherent 11, 68, 95, 97, 216, 218,
 254, 256, 284, 307–308, 310, 331
 impoliteness, mock 154–155, 171, 317–318
 impoliteness and context 4, 13–14, 50, 106,
 124, 133, 179, 182, 184, 210–211, 221, 256,
 264–265, 284–285, 290, 310–311, 316, 318,
 341, 342–343
 impoliteness effects 49–52
 impoliteness formula 14, 96, 131, 133, 170–171,
 184, 253–254, 257, 265, 287, 308–309, 312,
 314, 316–333, 342
 impoliteness judgement 20, 216, 264, 284, 290,
 292–294, 308, 310, 325, 327, 331
 insult 4–5, 15–18, 32, 34, 58, 111, 123, 131–138,
 146, 149, 155–156, 162–163, 166, 171, 180,
 188, 253, 256–257, 283–284, 289, 301–302,
 304, 308, 312–314, 318, 323, 329
 insulting 15–19, 96, 180, 254, 257, 285, 342

 intensification 42, 63, 133, 139, 141, 146–150, 156,
 161–163, 167, 171, 216–217, 246, 254, 257,
 259–260, 329
 intensifier 58, 129, 138, 146, 216, 222–223, 235,
 240, 246, 259
 interjection 21, 138, 148–149, 162, 167–168,
 179–181, 188–192, 195, 198, 209, 211,
 233–234, 236–238
 interpersonal 5, 10–11, 21, 133, 139, 159, 255,
 261–262, 289
 interrater reliability 14, 308–309, 325, 327–332
 Italian 5, 18, 31, 33–35, 38, 41, 53–58, 129,
 131–132, 136, 138–139, 141–145, 152, 155,
 165, 167, 170–172, 347

 Korean 10, 61–66, 68–69, 71–78, 80, 82–84,
 86–88, 90–91, 341

 Lak 101–103, 116–123
 lamentation 179, 181, 188–190, 210
 leetspeak 215, 227, 240, 246–247

 malediction 17, 129, 132–136, 152–153
 maledictive 102, 132, 139, 153, 156, 165, 167, 171
 metalinguistic 288, 308–309, 316–317, 326–327,
 329–330, 332
 metapragmatic 20, 129, 132, 147, 170, 172, 342
 modification 215, 221–222, 235, 239–240, 246,
 257, 259–260, 287, 312, 329
 morality 7, 149, 292
 morphology 8–10, 18, 20, 22, 31–38, 41, 48, 50,
 52, 54, 57–58, 61–66, 71, 75, 80–87, 89,
 91, 95, 97, 99–100, 102–104, 121, 123–124,
 230, 302
 morphology, transitional 31–33, 36–37
 Morphopragmatics 11, 31–35, 40, 46, 48–49

 Nakh–Daghestanian 20, 95–100, 102–103,
 110–112, 118, 122–124
 non-morphemic word formation 215, 222, 235,
 239, 246
 noun phrase 11, 72, 131, 150, 163, 172, 217, 219,
 221–223, 226, 241, 254, 259–260, 285–286

 offense 4, 8, 14, 20, 31–35, 39–41, 43, 45–46,
 48, 51–53, 57–58, 62, 65, 96, 131, 133, 136,
 146, 183, 206, 220, 237, 257, 283–285,

- 289, 300–301, 310, 312, 315, 318, 329, 340, 345–346
- off-record 160–161
- optative 14, 95, 99–108, 111–114, 116–118, 120, 123–124, 139–141
- Optimality Theory 12
- Parallel Architecture 9
- participle 55, 105, 112, 114–116, 123
- Pattern Grammar 208
- Polish 284–295, 298–304, 312
- politeness 5–7, 14, 17–18, 61–68, 80, 88, 95–97, 133, 140, 216, 255–257, 284–285, 289, 301–302, 304, 308–311, 314–317, 330, 340–342, 344–346
- politeness, mock 149, 171
- politeness formula 216, 314, 330
- pragmatic explicitness 17, 111, 186, 283, 289, 301, 304
- predication 257, 261
- prediction 10, 179, 186, 188–190, 197–199, 206, 208–210
- prefix 55–56, 61–66, 69–71, 73–78, 82, 84–85, 88–89, 114
- prepositional phrase 38, 54–57, 217, 221
- productivity 12–13, 31–32, 35–36, 39–40, 57, 69, 76, 217, 223, 237, 239, 246, 258, 262, 277, 309, 329, 332, 345
- profitability 39–40, 57
- pronoun 21, 107, 109–111, 115–116, 121–124, 129, 131, 138, 140, 143–145, 158–161, 166–167, 169, 171, 197–199, 206, 209–210, 217, 219, 230, 232–233, 244, 247, 261–263, 266, 269–271, 273, 277, 283, 286, 289, 295–298, 300–301, 303–304, 320–324
- pronoun, personal 110, 131, 138–139, 143, 158–159, 184, 262
- pronoun, second person 16–17, 21, 95, 107, 109–111, 114–118, 121–124, 144, 146, 169, 263, 270, 273, 275, 283, 289–290, 295–298, 300–304, 320
- pronoun, wh 216–217, 219, 229–233, 241, 244–245, 247
- prosody 10, 148, 180, 278, 329–330
- questionnaire 15, 19–20, 220, 253–255, 262–263, 265–266, 283, 290–296
- rating 70, 254, 264, 266–277, 329
- Rutul 100, 102–103, 111–116, 121–123
- sarcasm 18, 20, 34, 49–50, 66, 87–88, 148, 154, 169–170, 216, 278, 302–303, 310, 341
- schema 218, 228, 245–246, 258, 287, 303–304
- secretion 37
- semantic domain 216
- sentence type 5, 10, 15, 19, 180, 184
- slang 8, 31–35, 39–41, 43, 49, 51–53, 57–58
- speech act 10, 132–135, 180, 183, 190, 209, 211, 220, 344
- strategy 17–18, 32, 34, 54, 58, 62, 66, 78, 82, 145, 152, 163, 165, 167, 183, 215, 222, 235, 239, 246, 289, 344
- suffix 17, 34, 36–37, 39, 48, 54–56, 61–62, 65, 67, 74–75, 78–79, 82, 85, 97, 104, 111–112, 114, 116–119
- survey 70, 265, 277–278, 291–294, 298
- swearing 132–133, 136, 149, 163, 311
- syntax 9–10, 61–64, 72, 81, 217, 286, 341
- Systemic Functional Grammar 9
- taboo 147, 150, 156, 220, 226, 244, 247
- taboo language 15, 133, 149, 171, 216–217, 220, 222, 227, 247, 307–308, 311–312, 328–329, 331
- theater 129, 138, 152–153, 156, 171
- threat 4, 10, 134, 148, 179–188, 194, 198–200, 206, 208–211, 288, 308, 314, 323
- usage-based 22, 253–254, 258, 301, 303, 341
- vocative 117, 149, 206, 286, 289, 299, 302–303, 307, 312–314, 319, 324, 328–332
- well-formedness 15, 20, 80, 283, 291–300, 302–303, 341
- wish 52, 95–97, 99–104, 106–116, 118–123, 131, 135–136, 139, 343

