

1 Introduction

There is an ongoing debate in the literature about whether (im)politeness is solely a matter of context or whether (and to what extent) linguistic form also plays a role in conveying (im)politeness (e.g. Culpeper 2010; Van Olmen et al. 2023). In the first wave of (im)politeness studies, the focus was primarily on linguistic utterances, with contexts of use attracting comparatively little attention (Brown & Levinson, 1987). The situation changed dramatically with the development of the second wave of research, marked by postmodern and discursive approaches to (im)politeness (e.g. Eelen, 2001; Watts, 2003). In their view, no linguistic form is “inherently” (im)polite and the (im)politeness evaluations are entirely context-dependent (Locher & Watts, 2008: 78). Though the importance of contexts is undeniable, completely downplaying the role of linguistic forms “risks throwing the baby out with the bath-water” (Culpeper, 2011: 124). The third wave represents a more balanced perspective, with the work of Terkourafi (2005a, 2005b) and Culpeper (2011) suggesting that linguistic forms *can* acquire associations with (im)politeness through frequent co-occurrences with (im)polite contexts. However, despite the theoretical development, the context-dependent view proposed by discursive approaches remains influential (e.g. Chen and Li, 2023; Kadar and Zhang, 2019: 25).

Recent support for the view of impoliteness as potentially “conventionalized” comes from Van Olmen et al. (2023), who investigate ‘you’ plus a noun phrase functioning as an address in English, Dutch and Polish (e.g. *you idiot!*). They contend that this pattern counts as a distinct construction and find that, in each language, it is used for impolite purposes in more

than two thirds of the cases. The corresponding partial conventionalization of this YOU + NP construction for impoliteness is said to become especially clear with evaluatively neutral noun phrases: it explains why, despite the fact that YOU + NP can serve to compliment someone (e.g. *you cutie!*), nouns like ‘teacher’ and ‘reader’ tend to be forced into an impolite interpretation when they occur in the construction. Van Olmen et al. (2023) thus not only make the case for the existence of (semi-)conventionalized impoliteness constructions but also show that such constructions may be similar across languages.

The present study seeks to examine three constructions in Chinese that appear to be comparable to YOU + NP both structurally and functionally: *nǐ zhè* + NP, *nǐ gè* + NP and *nǐ zhège* + NP. They are composed of three parts, with the second person pronoun *nǐ* followed by the demonstrative *zhè*, the classifier *gè* or the combination of the two *zhège* and then a noun phrase. It has been proposed in the literature that they can be used to express negative attitude toward or a negative evaluation of the addressee (e.g. Zhang & Yin, 2004), which makes them suitable candidates for an investigation into impoliteness constructions in the language. Moreover, there is no consensus about the relationships between these three constructions, with some scholars suggesting that *nǐ zhè* + NP and *nǐ gè* + NP are simply abbreviated forms of *nǐ zhège* + NP (e.g. Zhang & Yin, 2004; Zhang, 2005) but others arguing that they are different constructions (e.g. Fu & Hu, 2020). Examining their association with impoliteness, as the present study aims to do, may shed light on their (dis)similarities. To this end, we will first carry out a multiple distinctive collexeme analysis of the constructions (Gries & Stefanowitsch, 2004) and then an in-depth analysis of their links with impoliteness in usage (à la Van Olmen et al., 2023). If they are indeed found to be associated with impoliteness, we will also try and explain how their components contribute to this association.

The rest of this article is organized as follows. Section 2 will review the debate about the “inherency” in form of (im)politeness as well as the research on YOU + NP constructions and on *nǐ zhè(ge)/gè* + NP. In Section 3, we will introduce the methodology, including our corpus selection, multiple distinctive collexeme analysis and our approach to studying impoliteness in usage. Section 4 will present and discuss the results of our analysis and, in Section 5, we will offer explanations for them. Section 6, finally, will contain our conclusions.

2 Literature review

2.1 Inherency of (im)politeness

No consensus has been reached in the current literature on whether (im)politeness can be inherent to language or, in other words, whether it can be conventionalized in linguistic form (Culpeper, 2011: Chapter 4; Culpeper & Hardaker, 2017: 208-212; Van Olmen et al. 2023).

As Culpeper (2011: 120) notes, almost no mainstream scholars are in support of the view that (im)politeness is “wholly inherent in linguistic expressions”. (Im)politeness is obviously never merely a matter of form, since the same expression can yield different interpretations in different contexts. For example, an expression such as *thank you* can be employed sarcastically (e.g. *thank you so much for ruining my life*) and, likewise, an expression such as *go to hell* can serve to convey intimacy between close friends. Such facts have led postmodern and discursive (im)politeness researchers to argue for a context-dependent analysis of (im)politeness, in which it is regarded as a dynamic social phenomenon and is associated with the evaluation of the interlocutors in specific situations

(e.g. Watts, 2003; Locher & Watts, 2008). This approach can easily explain why some apparently polite or impolite forms can be understood in other ways.

The discursive approaches' emphasis on context, while valuable in countering the form-focused first wave approaches, tends to downplay the role of linguistic structures (Culpeper, 2011: 122). As Terkourafi (2005a: 241) points out, this type of micro-level analysis essentially treats (im)politeness as a particularized implicature, where "no prediction is (or can be) made about the impact of linguistic expressions until one knows the specific context in which they were used". Put differently, every judgment of (im)politeness would require full-blown inferencing, which seems psychologically implausible (Haugh & Culpeper, 2018: 229).

The critics of the second wave research motivated "a general shift in the field towards a middle ground between classic and discursive approaches" (Haugh & Culpeper, 2018: 217). The resulting third wave approaches sought to encompass both participant and analyst perspectives, and take into account both linguistic and contextual factors in evaluating (im)politeness (see Ogiermann & Blitvich, 2019 for recent third wave studies).

The frame-based view of politeness proposed by Terkourafi (2005a) exemplifies this third wave approach, according to which "politeness is achieved on the basis of a generalized implicature when an expression x is uttered in a context with which –based on the addressee's previous experience of similar contexts¹– expression x regularly occurs" (Terkourafi 2005a: 251). That is, if a certain linguistic form is used to convey politeness in specific situations frequently enough, it can be presumed to evoke politeness when occurring in similar contexts. It is probably self-evident that, as Culpeper (2011) argues, this idea can be extended to impoliteness too. There is also evidence that language users are sensitive to this generalized

¹ The definition of "contexts" here follows that of Terkourafi (2005a), which retain information regarding age, gender and class of the interlocutors, their interpersonal relationship and roles and characteristics of the type of interaction.

relationship between (im)politeness and form. Zlov & Zlatev (2024), for instance, study people's reaction time in judging the impoliteness of a range of expressions in controlled contexts. They find that expressions conventionally used for impolite purposes are evaluated more quickly as impolite than expressions that are not. If, as postmodern and discursive approaches imply, any such judgment was made from scratch, we would not expect to see any difference in reaction time.

To our knowledge, there has been little overt discussion about the issue of the inherency of (im)politeness in the context of Chinese. Many studies take the discursive approach for granted and the idea that (im)politeness is a matter of situational evaluations by the interlocutors is considered common sense (e.g. Chen & Li, 2023; but see Wang & Taylor, 2019 for an alternative perspective). Despite the dominance of this view in current research in Chinese (im)politeness, Kádár & Zhang (2019: 26) argue that it may effectively analyze complex interactions, but “it does not help us to capture conventionalised language use which constitute[s] the basis of [(im)]politeness”.

2.2 YOU + NP

Inspired by the frame-based approach to politeness in Terkourafi (2005a), Culpeper (2010, 2011: Chapter 4) tries to establish a list of conventionalized impoliteness formulae in English, which are identified based on an analysis of typically impolite contexts (e.g. army training) and of a hundred diary reports about impolite encounters (see Lai, 2019 for Chinese data and Rabab'ah & Alali, 2020 for Arabic data). Among the impoliteness formulae proposed by Culpeper (2010) is the so-called negative vocative construction, of which *you idiot!* would be an example. This combination of a second person pronoun and a noun phrase seems to be one of the most discussed such formulae in the literature (e.g. Potts & Roeper, 2006), perhaps due

to its high frequency and ease of retrieval from corpora. It also appears to have close equivalents in various other languages (e.g. Corver, 2008 on Dutch; Julien, 2016 on Scandinavian languages), which provides a good opportunity for cross-linguistic comparison.

YOU + NP has been argued to be an evaluative vocative construction (e.g. Corver 2008), serving not only to address someone directly but also to express an attitude toward or evaluation of that addressee. This evaluative meaning is clear from the fact that the construction does not really tolerate non-evaluative noun phrases –as the awkwardness of *you linguist!*, for instance, shows. At the same time, *you linguist!* can be regarded as evidence for YOU + NP’s evaluative nature, in that this example only works if *linguist* is somehow interpreted as an assessment of the addressee’s character. Crucially, YOU + NP can express both positive (e.g. *you angel!*) and negative evaluation (e.g. *you idiot!*) but, as Jain (2022) among others points out, it is biased toward the latter. She asked speakers of English to assess *you deffxigta!* in the absence of context and, despite the nonsense word *deffxigta* clearly not having any (positively or negatively evaluative) meaning, they largely judged it to convey a negative assessment of the addressee in YOU + NP. This fact may be taken to point to the construction’s partial conventionalization for impoliteness. Van Olmen et al.’s (2023) corpus study provides the basis in usage for this default interpretation of YOU + NP: more often than not, the construction is used to insult people rather than to compliment them, not just in English but in Dutch and Polish too. In line with Jain’s (2022) example, Van Olmen et al. (2023) also observe that, in all three languages, evaluatively neutral nouns such as ‘linguist’ tend to be understood as conveying *negative* evaluation in particular when appearing in YOU + NP, further cementing its status as a construction that is partly conventionalized for impoliteness.

2.3 Nǐ zhè(ge)/gè + NP

Like YOU + NP, the *nǐ zhè(ge)/gè* + NP constructions are verbless and can be used to address someone and, at the same time, convey an evaluation of that person (Zhang & Yin, 2004). *Nǐ gè zhū* ‘you pig’, for example, expresses the speaker’s disgust at the addressee for being foolish. Positive evaluation is possible too, as *nǐ gè xiǎo kěài* ‘you cutie’ shows, and it therefore remains to be seen whether the constructions are conventionally associated with impoliteness. *Nǐ zhè(ge)/gè* + NP does differ from YOU + NP as studied in Van Olmen et al. (2023) in two regards.

First, they consist of three parts, with a demonstrative phrase or a classifier between the second singular person pronoun *nǐ* ‘you’ and the noun phrase. It is important to note that, for the two constructions with a demonstrative phrase, only the proximal demonstrative *zhège* ‘this (one)’ or *zhè* ‘this’ can be used (Tao, 1999: 87). With the distal demonstrative *nàge* ‘that (one)’ or *nà* ‘that’, they can only be interpreted as expressing possession (e.g. *nǐ nà hái* ‘your child’). As for *nǐ gè* + NP, *gè* is the only classifier that can appear here. For instance, though *tóu* ‘head’ is the most common classifier for *zhū* ‘pig’, it is ungrammatical to say *nǐ tóu zhū* to mean ‘you pig’.

Second, the *nǐ zhè(ge)/gè* + NP constructions can perform the function of an independent address but they can also occur as the argument of a clause. Example (1)² is a case in point, with *nǐ zhège húndàn* ‘you bastard’ serving as a direct object.³ YOU + NP in Dutch, English and Polish does not allow this (e.g. *?you idiot are ...*), at least in the singular (cf. *you idiots are ...*; Van Olmen et al., 2023: 26-27).⁴

² The following abbreviations are used in the gloss: 1 = first person; 2 = second person; ATT = attributive marker; CLF = classifier; DEM = demonstrative; SG = singular.

³ Our translations will not always be idiomatic English, as we intend them to show the idiosyncrasies of Chinese.

⁴ To be clear, there *are* other languages where (singular) YOU + NP can function as an argument within a clause. German, as in (i), is one of them.

(i) *Gestern hast du kleiner Trottel versagt.*

(1) *Wǒ dǎsǐ nǐ zhège húndàn!*

1SG beat.to.death 2SG DEM bastard

‘I will beat **you bastard** to death!’

Interestingly, Fu & Hu (2020) argue that such syntactically integrated instances of *nǐ zhè(ge)/gè* + NP predate their independent uses and that, with the development of the latter, the constructions have actually become more expressive. For that reason, we will examine whether there is a link between their proportions of integrated versus address uses and their associations with impoliteness.

The preceding paragraphs have discussed the *nǐ zhè(ge)/gè* + NP constructions together. There is, however, a continuing discussion in the literature about the relationships between the three constructions. According to certain researchers (e.g. Zhang & Yin, 2004; Zhang, 2005), *nǐ zhè* + NP and *nǐ gè* + NP are simply abbreviated forms of *nǐ zhège* + NP, even if some of them will still acknowledge potential differences between the three constructions, especially in terms of their association with negative evaluation. Zhang (2005), for instance, claims that *nǐ zhè* + NP carries a stronger sense of criticism than *nǐ gè* + NP. By contrast, Lv (1985: 201-202) suggests that the gravity of offense carried by *nǐ gè* + NP is stronger and that the demonstrative is omitted because the speaker is so emotionally charged. In Hu & Gao’s (2018) view, *nǐ gè* + NP is more evaluative than *nǐ zhè(ge)* + NP. They state that evaluatively neutral nouns like ‘person’ and ‘child’ and proper nouns rarely occur in *nǐ gè* + NP but are perfectly acceptable in the other two constructions. It is important to point out, however, that no empirical support is provided for any of these assertions. Zhang (2005) and Lv (1985) only briefly mention the difference without further elaboration and Hu & Gao’s (2018) claim

‘Yesterday, **you little sucker** screwed up.’
(d’Avis & Meibauer, 2013: 200)

about nouns is mainly based on their own intuitions. Contrary to all these scholars, Fu & Hu (2020) explicitly argue that *nǐ gè* + NP is a construction distinct from *nǐ zhège* + NP. They looked at a historical corpus of Chinese literature and observed that *nǐ gè* + NP already occurs in Later Tang Dynasty texts (907-960) while *nǐ zhège* + NP only first appears in Southern Song Dynasty texts (1127-1279). Given that the constructions are informal and more typical of spoken language, it is of course possible that *nǐ zhège* + NP already existed during the Later Tang Dynasty but was simply not documented. At any rate, there is clearly no consensus on the status of *nǐ zhè(ge)/gè* + NP. The present paper aims to empirically investigate whether *nǐ zhè* + NP and *nǐ gè* + NP are abbreviated forms of *nǐ zhège* + NP or function as independent constructions.

3 Methodology

3.1 Corpus

Our study will make use of the Chinese Web Corpus 2011 (zhTenTen11). This choice of corpus is motivated by several considerations. First, with around 1.7 billion words, it is very large, guaranteeing a sufficient number of attestations for further analysis. Second, the corpus contains a variety of text types –including blogs, online fiction and discussion forums, where the language often approximates speech and the *nǐ zhè(ge)/gè* + NP constructions are more likely to occur. Third, the corpus is linguistically annotated, which enables the relatively easy retrieval of instances of interest. The 2017 version (zhTenTen17) features annotation too but an initial exploration of this more recent corpus revealed problems with the extraction of common nouns, which zhTenTen11 does not have.

3.2 Data selection

The corpus was accessed through the Sketch Engine platform.⁵ The query in (2) was used to conduct the search for *nǐ zhège* + NP (你 *nǐ*, 这 *zhè*, 个 *gè*). For the other two constructions, the second word was replaced by just 这 *zhè* and just 个 *gè* respectively.

(2) [word = “你”] [word = “这个”] [tag = “JJ.*|V.*”]? [word = “的|之”]? [tag = “N.*”]

The queries target cases where the second person pronoun *nǐ* is followed by either the demonstrative phrase *zhège/zhè* or the classifier *gè* and then a noun phrase. This noun phrase is set to include a noun (see [tag = “N.*”]) that may optionally be preceded by a modifying phrase (see [tag = “JJ|V.*”]? [word = “的|之”]?). The possibility of a modified noun expanded the search from just instances like ‘you bastard’ to instances like ‘you stupid bastard’ as well, without opening it up so much that the results would contain an insurmountable amount of false positives. The reason for including V.*, i.e. verbs, as an option in the tag for the modifying phrase, alongside JJ.* for adjectives, is that verbs in Chinese can be placed before the noun to modify it (Li & Thompson, 1989: 116), as in (3).

(3) *nǐ zhège méi lánxīn de jiāhuo*
2SG DEM lack heart ATT guy
‘You ungrateful guy’
(zhTenTen11-1169895078)

⁵ See <https://www.sketchengine.eu/> (last accessed on 28/04/2024).

Here, the verbal phrase *méi lánxīn* ‘lack heart’ modifies *jiāhuo* ‘guy’. The attributive maker *de* (的) and *zhī* (之) mark the modification relationship.

The results were then manually cleaned and corrected.⁶ Cases with superficial similarities to the target constructions were deleted after looking into the concordance lines. For instance, in (4), since the principal is speaking to the parents, *nǐ zhè hái zi* here does not refer to the hearer and does not mean ‘you child’. Rather, it is a possessive phrase, meaning ‘your child’, and is thus not relevant to the current study. In the same vein, given that the study focuses on impolite language, where a specific addressee is typically needed, cases where the second person pronoun is used for generic reference, like (5), were excluded from the data.

- (4) *Yuánzhǎng niǔtóu duì jiāzhǎng shuō, nǐ zhè hái zi bù shì hé lái wǒmen yòuér yuán.*

‘The principal turned to the parents and said, “**Your child** is not suitable for our kindergarten.”’

(zhTenTen11-1106554314)

- (5) *Candidate nàme duō, HR yòu dōu hěn máng, tāmen gāi rúhé lái pànduàn nǐ zhè ge rén shì bù shì fēicháng yōuxiù ne, jiù shì kàn background, nǐ de bèijīng bǐ biérén hǎo, shuō míng nǐ yōuxiù.*

‘There are so many candidates, and HR are often very busy. How can they determine whether **you this person** is truly outstanding? They mainly rely on your background. If your background is better than others’, it indicates that you are excellent.’

(zhTenTen11-372850771)

⁶ There are cases where the corpus mistakenly segments one word into several different tokens. For example, *gǒurìde*, which literally means ‘dog fucking (person)’, is tagged as three characters in the corpus. Our query therefore hit on [*nǐ zhè ge gǒu*] *rì de* instead of [*nǐ zhè ge gǒurìde*] but we manually added *rìde* to *gǒu*, since the three characters together convey the complete meaning.

Table 1 presents the total number of query hits and the total number of relevant cases for each construction.

Table 1. Total and relevant hits of the *Nǐ zhè(ge)/gè* + NP constructions

Construction	Number of query hits	Number of relevant hits
<i>nǐ zhège</i> + NP	7,767	2,402
<i>nǐ zhè</i> + NP	4,307	1,860
<i>nǐ gè</i> + NP	2,000	711

3.3 Multiple distinctive collexeme analysis

We conducted a multiple distinctive collexeme analysis on all relevant hits in Table 1. This technique measures the strength of association between specific lexical items and the constructions, revealing the distinctive lexical preferences of each construction. Based on these preferences, the semantic differences of the constructions can be identified (Gries & Stefanowitsch, 2004). The analysis was performed in RStudio (R Core Team, 2022), using codes provided in Levshina (2015: 248-249). For each construction, a file containing two columns was created, the first one listing the lexical items in the noun phrase slot and the second one displaying their overall frequency in the construction (see the Online Resource 1). Based on the input file, the expected frequency⁷ of each noun phrase can be calculated. Then, a Fisher-Exact test was used to compare the differences between the observed and expected frequencies. For sake of readability, the results of the test are presented in the form of a negative base-10 logarithm of the p-value, termed the distinctiveness value. The attraction between the collexeme and the construction was considered statistically significant if the

⁷ Four frequencies are needed to calculate the expected frequency of the lexical item X in construction Y: the observed frequency of the lexical item X in construction Y (=a), its frequency in other constructions (=b), the sum of all instances of construction Y other than X (=c), and the sum of all instances of other constructions other than X (=d).

distinctiveness value exceeded 1.30 ($p < .05$), with higher distinctiveness values indicating stronger associations.

3.4 Contextual analysis

After randomizing the relevant hits, we selected the first 150 instances of each construction for further in-depth analysis (see the Online Resource 2). Each attestation was then coded for these two aspects: (i) whether it is impolite, non-impolite or unclear and (ii) whether it serves as an address or as an argument.⁸

3.4.1 (Non-)impoliteness

Building on the work by Culpeper (2011: 11-12) and Van Olmen et al. (2023: 29-30), we tagged an instance as impolite when there is sufficient contextual information suggesting that the expression is “meant and/or taken to have negative emotional consequences” for the hearer. Such contextual indications are of various forms. A first type of indication could be the description of a speaker’s aggressive psychological status and/or their non-verbal behavior (Van Olmen et al. 2023: 30). In (6), for instance, it is clear that *nǐ zhège húndàn* ‘you bastard’ is intended to cause offence, as the speaker is described as being angry and shouting at the hearer.

(6) *Qìjībàihuài de Xiǎodōng chòngzhe diànhuà jiù hǎndào: “Gāo Xīn, nǐ zhège húndàn!”*

‘Furiously, Xiaodong shouted into the phone: “Gao Xin, **you bastard!**”’

(zhTenTen11-1239118061)

⁸ Van Olmen et al. (2023) coded their YOU + NP data also for the presence/absence of adjectives and found that cases without adjectives are significantly more frequently impolite than cases with adjectives. This fact was taken to suggest that (positively evaluative) adjectives are needed to counter the default interpretation of the construction as an insult (cf. *you woman!* and *you beautiful woman!*). At the initial stage of our analysis, we too annotated for the presence/absence of modifying phrases but we did not find any differences.

A second type of evidence is the addressee's confrontational verbal and/or emotional reaction. As Culpeper et al. (2003: 1562-1568) note, unlike politeness, impoliteness often elicits responses from the hearer: to accept the insult, to counter it and/or to neglect it. In (7), for example, *nǐ zhège dúfū* 'you poisonous woman' is clearly taken as an insult since the hearer counterattacks by comparing the speaker to a mad dog.

- (7) *Wǒ hěnhěn de tuīkāi tā, zuǐlǐ màdào: "Nǐ zhège dúfū! Dāngchū kànshàng nǐ zhēn shì xiāle yǎn!" Tā yǐyáhuányá: "Nǐ zhè tiáo fēnggǒu! Jiàgěi nǐ zhēnshì hūnle tóu!"*
'I pushed her away violently, cursing: "**You poisonous woman!** I was blind to be attracted to you back then!" She retaliated: "You this mad dog! I lost my mind marrying you!"'
(zhTenTen11-298670357)

The final indication to rely upon is the co-occurrence with other impolite speech acts such as threats, dismissals and negative expressives. In (8), for instance, it is evident that the speaker intends to hurt the feelings of the addressee since, in addition to *nǐ gè chòu bāpó* 'you stinky bitch', the speaker is also cursing and expressing disgust toward the addressee.

- (8) *"Huī! Nǐ bié jiào de nàme ròumá! Wǒ tīngzhe ěxīn! Wǒ gěi nǐ xiě qínghū, nǐ zuòmèng qù ba! Nǐ gè chòu bāpó bù yào nàme zìliàn! Nǐ zǎodiǎn qùsǐ ba!"*
'"Hui! Don't call it so cheesy! It disgusts me! I wrote a love letter for you? In your dreams! **You stinky bitch**, don't be so narcissistic! Just drop dead!"'
(zhTenTen11-668893396)

The impolite instances were further categorized as evaluative or non-evaluative. In

evaluative impoliteness, the noun phrase serves as the speaker's subjective assessment of the addressee (e.g. 'you stinky bitch' in example 8), while in non-evaluative impoliteness, it serves primarily referential functions. For example, in (9), 'journalism teacher' represents a factual description of the occupation of the addressee rather than a subjective evaluation. The speaker's anger and criticism stem from the professional failure of the addressee, rather than negative attributes associated with the occupation itself. Given the noun phrase itself carries no negative assessment (such as invoking pedagogical stereotypes like being pedantic), this example was coded as non-evaluative impoliteness.

- (9) *Yī jìzhě bākāi zhònggrén chōng dào fèngxì chù, xiàng shòukùnzhě dàhǎn: “Zhīdào ma? Nǐ yǐjīng bèi máile 138 gè xiǎoshí! Nǐ zhīdào ma?” Qīde ǎn tàitài tiào qílái dàma ǎn: “Nǐ gè jiāo xīnwén de, nǐ zěnmē jiāo chūlái de xuéshēng? Rénjiā mímídèngdèng zài fèixū lǐ máile 5 tiān 5 yè, zhè rén shàngqù jiù hǎn ‘nǐ máile 138 gè xiǎoshí’, shì xīnlǐ ānwèi háishì xìngzāilèhuò...”*

‘A reporter pushed through the crowd and shouted at the trapped victim: “Do you know? You've been buried for 138 hours! Do you know that?” This angered my wife so much that she jumped up and scolded me: “**You journalism teacher**, how could you train students like this? Someone's been dazed in the ruins for 5 days and 5 nights, and your student shouts about the person being buried for 138 hours? Is that meant to be comforting or just gloating?”’

(zhTenTen11-503676669)

When no impolite indications are available in the context, the hit was coded either as non-impolite or as unclear. The former category includes instances of different types –such as friendly teasing/banter, complimenting and simple identification, like (10) to (12)

respectively. In (10), the speaker may refer to the addressee as a bastard but it is actually intended in a friendly way, which is supported by the interlocutors embracing each other warmly. The context in (11) makes clear that *nǐ zhège xiǎo tiānshǐ* ‘you little angel’ serves to convey a positive evaluation of the child. In (12), there is nothing to suggest that ‘father’ expresses evaluation. Rather, it merely indicates the role of the addressee as a father.

- (10) “*Hāhāhā. Jiāngjūn, děngsǐ nǐ gè chùsheng a. Hāhāhā.*” *Liǎng rén rèqíng de yōngbào zài yìqǐ le.*

“‘Hahaha. Jiangjun, we have waited for **you bastard** for such a long time. Hahaha.’

The two of them warmly embraced each other.’

(zhTenTen11-630595099)

- (11) *Lǎotiānyé zài zhège shíhou bǎ nǐ zhège xiǎo tiānshǐ cìgěi wǒmen, nándào zhè jiùshì tiānyì me? Bǎobèi, nǐ jiùshì luòrù fánjiān de jīnglíng, bàba māma ài nǐ!*

‘At this moment, the heavens bestowed **you little angel** upon us. Could this be fate?

Baby, you are the elf who descended to the world. Mom and dad love you!’

(zhTenTen11-1740098538)

- (12) “*Nǐ zhège lǎobà tǐng rènzhēn de.*” *Tóngshì dǎqù. “Wǒ suànshì mǎhu de, tīng shuō érzi bānlǐ yǒu jǐ wèi jiāzhǎng hái yòng diànnǎo ruǎnjiàn zuòle yǐngjī.”*

“‘**You this father** really take it seriously,” a colleague said teasingly. “I think I’m rather casual about these things. I heard some parents in my son’s class even made photo albums using computer software.”’

(zhTenTen11-1639131886)

The latter category, i.e. that of unclear cases, includes instances for which the contextual information is insufficient to determine whether they are impolite or non-impolite. Example (13) is a case in point.

(13) *Huífù @ Wākào Bái Xiǎobái nǐ gè dà xīguā: Néngli!*

‘Reply to @Wakao Bai Xiaobai, **you big watermelon**: Ability!’

(zhTenTen11-539508247)

It is part of a conversation on an online discussion forum. *Nǐ gè dà xīguā* ‘you big watermelon’ seems to be the nickname of one of the interlocutors but the nature of the rest of the interaction does not allow us to establish whether it functions as simple identification, as a genuine insult or as banter.

The data was analyzed in the above way first by the first author, a first language speaker of Chinese, in several rounds. Difficult cases were discussed with other first language speakers with a major in linguistics and with the second author, through translations into English. To ensure the robustness of the analysis, the first author then trained an external annotator. They independently coded one fifth of the data and the inter-rater reliability was calculated by conducting Cohen’s κ ($\kappa = .748$, $p < .001$), suggesting a substantial agreement between the two raters.

3.4.2 Address or argument

Example (14) was coded as an address usage, since *nǐ gè dà piànzi* ‘you big liar’ occurs on its own, and (15) as an argument usage, since *nǐ zhège zázhǒng* ‘you bastard’ functions as the direct object of the clause. Such decisions were not always so easy to make, though. Take *nǐ*

zhège xiǎo biēsān ‘you worthless nobody’ in (16), for instance. It could be analyzed as the subject of the following sentence, i.e. as the person looking. The presence of the comma after the construction does allow for an alternative interpretation, since it can be seen as reflecting a pause in speech. *Nǐ zhège xiǎo biēsān* would then be an address and the absence of an overt second person subject in the subsequent sentence could be attributed to the fact that, in Chinese, pronouns are often omitted if they are contextually retrievable (Yip & Dong, 2004: 373-374). In such cases, we adopted the second analysis and (16) was coded as an address.

(14) *Xiǎopèi dà hū: “Wénzi, nǐ gè dà piànzi.”*

‘Xiaopei shouted: “Wenzi, **you big liar!**”’

(zheTenTen11-520893952)

(15) *Yán Chéngtǎn: “Wǒ dǎsǐ nǐ zhège zázhǒng!”*

‘Yan Chengtan: “I’ll beat **you bastrad** to death!”’

(zhTenTen11-492550328)

(16) *Jiàn Dàilì zài dǎliáng tā, tā dànù, màdào: “Nǐ zhège xiǎo biēsān, kàn shénme kàn!”*

‘Seeing Dai Li sizing him up, he became furious and cursed: “**You worthless nobody**, what [are you] looking at!”’

(zhTenTen11-691480819)

4 Results

4.1 Multiple distinctive collexeme analysis

Table 2 lists the top fifteen distinctive collexemes of all three constructions in descending order of distinctiveness (see Section 3.3). The collexemes that, at face value, express negative

evaluation are in bold⁹, as they may be indicative of the constructions’ association with impoliteness.

Table 2: Top fifteen distinctive collexemes of the *Nǐ zhè(ge)/gè* + NP constructions

Rank	<i>nǐ zhè(ge)</i> + NP		<i>nǐ zhè</i> + NP		<i>nǐ gè</i> + NP	
	Collexeme	Distinctiveness	Collexeme	Distinctiveness	Collexeme	Distinctiveness
1	<i>péngyǒu</i>	8.6388	<i>háizi</i>	45.1363	<i>gǒurìde</i>	35.7573
	‘friend’		‘child’		‘dog fucker’	
2	<i>tóngzhì</i>	5.2183	<i>xīǎozǐ</i>	24.6307	<i>B</i>	6.7727
	‘comrade’		‘boy’		‘cunt’	
3	<i>húndàn</i>	5.1837	<i>rén</i>	20.7477	<i>gǒuzázhǒng</i>	5.9243
	‘bastard’		‘person’		‘bastard’	
4	<i>érzi</i>	4.3263	<i>yātou</i>	6.3808	<i>xiǎo tùzǎizi</i>	5.2362
	‘son’		‘girl’		‘brat’	
5	(** offensive marker)	2.5310	<i>hóuzi</i>	4.6020	<i>SB</i>	5.0764
			‘monkey’		‘idiot’	
6	<i>fēngzi</i>	2.5310	<i>pōhóu</i>	4.2776	<i>jībā</i>	5.0764
	‘madman’		‘impudent ape’		‘dick’	
7	<i>huàidàn</i>	2.4935	<i>xīǎo jiāhuo</i>	4.2531	<i>lǎo piànzi</i>	4.2876
	‘bad guy’		‘little guy’		‘old lair’	
8	<i>nǚér</i>	2.3073	<i>gǒucái</i>	3.8492	<i>wángbādàn</i>	4.2613
	‘daughter’		‘worthless person’		‘bastard’	
9	(<i>lǎobùsǐde</i> ‘old bastard’)	2.2143	<i>jiāhuo</i>	3.8237	<i>xīǎo B</i>	4.2290
			‘guy’		‘little cunt’	
10	(<i>Dīng Dàhǎi</i>	1.8977	<i>pǐfū</i>	2.9928	<i>xīǎo mápi</i>	4.2290

⁹ We conducted a survey with 121 native speakers to gather their judgments on the sentiment of the NPs (positive, negative, or neutral). Words were classified as negatively evaluative if the majority of respondents judged them to be negative. For example, for the word *pōhóu*, 58.7% of respondents judged it to be negative, 6.6% positive, and 34.7% neutral, which supports its categorization as negative. A corpus-based sentiment analysis as suggested by one of the reviewers could be part of our future analysis to objectively categorize the lexical items.

	proper name)		‘ordinary person’		‘little dick’	
11	<i>nóngmín</i>	1.8977	<i>xiǎo héshang</i>	2.9928	<i>xiǎo jībā</i>	3.3821
	‘peasant’		‘little monk’		‘little dick’	
12	<i>xiǎo tóngzhì</i>	1.8977	<i>xiǎo zéi</i>	2.9928	<i>zhūtóu</i>	3.3709
	‘little comrade’		‘little thief’		‘pig head’	
13	<i>húlijīng</i>	1.8977	<i>xiǎo huǒzi</i>	2.6380	<i>tùzǎizi</i>	3.0329
	‘seductress’		‘little boy’		‘brat’	
14	<i>chǔndàn</i>	1.8977	<i>xiǎo guǐ</i>	2.3780	<i>sǐ rén</i>	3.0329
	‘idiot’		‘little demon’		‘dead person’	
15	<i>bàijiāzǐ</i>	1.8977	<i>nǚzi</i>	2.1370	<i>chòu bīdǎozi</i>	3.0160
	‘spendthrift’		‘girl’		‘stupid bitch’	

A few comments about Table 2 are in order. First, *hóuzi* ‘monkey’ in column four is not bold, as all cases of this collexeme refer to the actual animal. Second, not every collexeme is especially informative. *Dīng Dàhǎi* in column two, for example, is a proper name and appears as the tenth most distinctive collexeme of *nǚ zhège* + NP. But all instances of this word are from the same text, suggesting that we are dealing here with a text-specific preference rather than with a feature of the construction itself. Third, ** and *lǎobùsǐde* ‘old bastard’ in the second column come out as distinctive because of annotation issues with the corpus. Although the offensive marker symbol exists in *nǚ gè* + NP as well, it is presented as *XX*, and the corpus does not recognize it as equivalent to **. As for *lǎobùsǐde*, it is incorrectly tokenized into four characters –*lǎo*, *bù*, *sǐ*, *de*. The first character *lǎo* ‘old’ is annotated as an adjective in *nǚ zhège* + NP, while as adverb in the other two constructions, resulting in the omission of the word in the other constructions. Such “problematic” collexemes are marked with parentheses in Table 2.

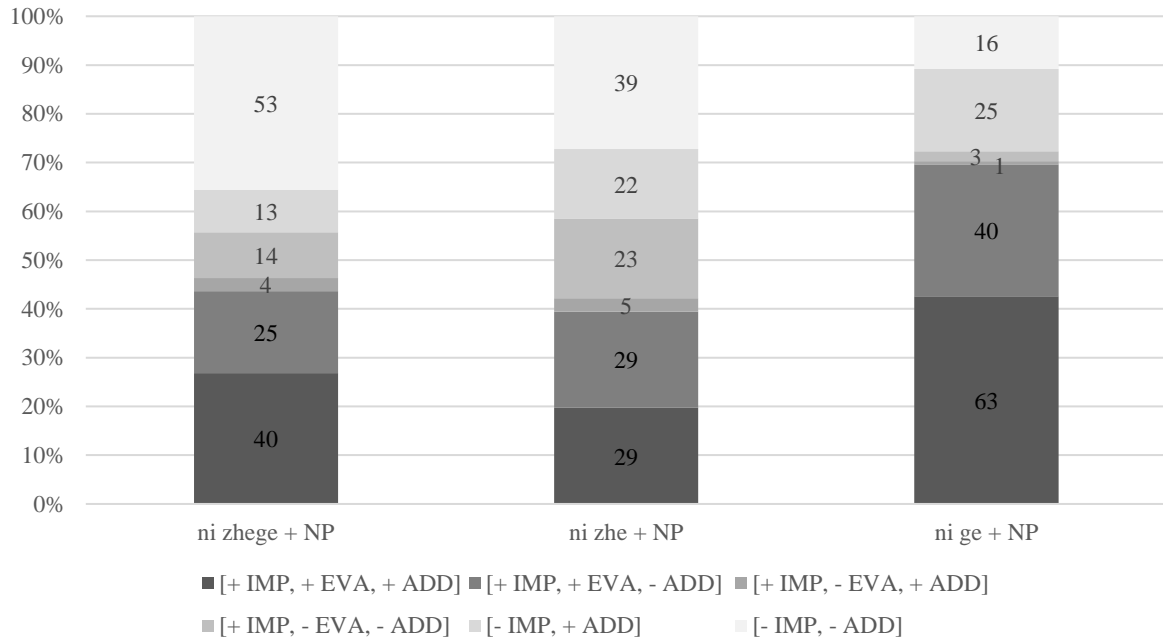
If the three constructions under examination are genuinely interchangeable, as argued by some scholars (see Section 2.3), we do not expect to find any differences between them in the nouns that they might attract. It is evident from Table 2, however, that such differences exist, particularly between *nǐ zhè(ge)* + NP on the one hand and *nǐ gè* + NP on the other hand. Of the twelve relevant distinctive collexemes of *nǐ zhè(ge)* + NP, half are not negatively evaluative in any straightforward way, such as *péngyǒu* ‘friend’ and *tóngzhì* ‘comrade’. The proportion of this type of noun rises to eleven out of fifteen distinctive collexemes for *nǐ zhè* + NP, with *gǒucái* ‘worthless person’ and *pōhóu* ‘impudent ape’ among the exceptions. The collexemes of *nǐ gè* + NP, by contrast, all seem to have negatively evaluative semantics. The most distinctive one, *gǒurìde*, literally means ‘person who fucks dogs’ and is a highly offensive address term in Chinese. This finding empirically confirms Hu & Gao’s (2018) claim that *nǐ gè* + NP is not very compatible with evaluatively neutral noun phrases (see Section 2.3) –unlike *nǐ zhè(ge)* + NP, for which we do see nouns such as *érzi* ‘son’, *háizi* ‘child’, and *rén* ‘person’ among the distinctive collexemes in Table 2. The observed divergence between *nǐ gè* + NP and *nǐ zhè(ge)* + NP is also suggestive of the constructions’ potentially different levels of association with impoliteness and/or of potentially different types of impoliteness linked to the constructions (see Section 3.4.1 on (non-)evaluative impoliteness). We will examine those suggestions in more detail in the next section.

4.2 Contextual analysis

Figure 1 summarizes the results of the contextual analysis of our sample of the three constructions under investigation. Unclear cases (see Section 3.4.1) are not included here, which leaves us with 149 instances of *nǐ zhè(ge)* + NP, 147 of *nǐ zhè* + NP and 148 of *nǐ gè* + NP. Note that, in the legend, [+ IMP] stands for impolite, [- IMP] for non-impolite, [+ EVA]

for evaluatively impolite, [- EVA] for non-evaluatively impolite, [+ ADD] for address usage and [- ADD] for argument usage. For the sake of transparency, Figure 1 also gives the absolute frequencies of the various types in the bars themselves.

Fig. 1 Contextual analysis of the three constructions



4.2.1 *Nǐ zhè(ge)/gè* + NP and impoliteness

Figure 1 reveals a notable association of *nǐ zhè(ge)/gè* + NP with impoliteness assessments. All three constructions are seen functioning as impoliteness triggers in more than half of the cases, with *nǐ gè* + NP having the highest proportion (72.30%), followed by *nǐ zhè* + NP (58.50%) and then *nǐ zhège* + NP (55.70%). When used in this way, the constructions can feature noun phrases that consist of just an evaluative noun, an evaluative adjective and an evaluatively neutral noun, a combination of an evaluative adjective and an evaluative noun or an evaluatively neutral noun, as in (17) to (20) respectively.

(17) *Zài yīpiàn jìngmò zhōng, Āsāo āyí bǐyí de shuō: “Hú Délì, nǐ zhège wángbādàn.”*

‘In the silence, Aunt Asao said disdainfully, “Hu Deli, **you bastard!**”’

(zhTenTen11-849702561)

(18) *Xú dàníang jíle, màdào: “Nǐ zhè chòu xiǎozi gǎnjīn sǐle ba, sǐle bù rě dàhuǒ fán!”*

‘Aunt Xu became impatient and scolded: “**You rotten guy** better die quickly, so that you won’t bother anyone!”’

(zhTenTen11-1448444041)

(19) *“Nǐ gè chòu liúmáng, mǎshàng gěi wǒ gǔn chūqù!” Shuōwán Kē Xuě xiàng fēnggǒu yīyàng pǎo guòlái duì wǒ yòusīyòuyǎo.*

‘“**You stinky rascal**, get out of here!” After saying that, Ke Xue came running at me like a mad dog, scratching and biting.’

(zhTenTen11-622588109)

(20) *Yǎnkàn zhe lǎopó yǎnjīng yòu hóngle, Lín Dùn gǎnjīn tǒngle Hú Xiān yīxià, bùkuài de mà: “Nǐ zhège nǚrén, jiào nǐ bié lái nǐ fēiyào lái. Láile yòu kūkūttī de, yǐngxiǎng lǐngdǎo de qíngxù.”*

‘As his wife’s eyes welled up with tears again, Lin rushed to poke Hu Xian and scolded her: “**You woman!** I told you not to come, but you insisted on coming. Now that you’re here, you’re crying and making a scene.”’

(zhTenTen11-853449399)

As discussed in Section 2.1, if an expression frequently co-occurs with (im)polite contexts, it can become (partially) conventionalized **as impoliteness triggers**. Based on the numbers in Figure 1, we would argue that this is the case for *nǐ zhège/ge/gè* + NP and impoliteness in particular. This conventionalization is noticeable in the constructions’ ability to coerce evaluatively neutral nouns into a negative reading. In (21), for example, *nǐ zhège* + NP can be

said to turn *nǚrén* ‘woman’ into an insult: the negative stereotypes associated with women (e.g. as overly emotional and fragile) are activated and the speaker is criticizing the addressee for exhibiting such traits. The constructions’ conventionalization for impoliteness is even more apparent in an example like (21).¹⁰

(21) “*Wèi, nǚ... chéngpǐnbù zài zhèbiān, nǐ gè... āi, bènsǐ le.*” *Guāngtóu qìde zhí duōjiǎo.*

“‘Hey, you... the department is over here. **You...** oh, so stupid,” the bald guy stamped his feet in frustration.’

(zhTenTen11-911360298)

The noun phrase is absent here but *nǐ gè* + NP still serves as an insult, indicating that the construction itself can convey impoliteness, independent of any lexical content.

One question that needs to be addressed is whether the proportions of impolite instances of *nǐ zhège* + NP and *nǐ zhè* + NP are sufficiently high to argue for (partial) conventionalization. They only marginally exceed 50%, after all. It is important to consider the non-impolite cases in this respect. They include cases of banter, compliments and mere identification, like (22) to (24) respectively. This diversity puts the percentages of 55.70% and 58.50% into perspective, i.e. as still representing the main usage type of *nǐ zhège* + NP and *nǐ zhè* + NP.

(22) “*Zéyīn, nǐ zhège xiǎo shǎguā, kěài de xiǎo shǎguā. Wǒ xiǎng qīnqīn nǐ.*”

“‘Zeyin, **you little fool**, adorable little fool. I want to give you a kiss.’”

(zhTenTen11-1821471534)

(23) *Ò, tàiyáng, tàiyáng, nǐ zhè wěidà de huǒshén, wǒ zhēnde yōngyǒu nǐ le ma?*

¹⁰ This example comes from the corpus but is not actually part of the sample for our contextual analysis.

‘Oh, Sun, Sun, **you great fire god**, do I really possess you?’

(zhTenTen11-2097663926)

(24) *Zài zhèzhǒng huánjìng lǐ, yàngzi xiàng gè xiǎo fūrén de Duōlì jìng shǐyòng de nàxiē cí, shì nǐ zhè wàiguórén kěnéng dōu bùzhīdào huò bùdǒng de, zhè zhēn ràng rén zhènjīng.*

‘In this environment, the words that Duoli, who looks very little, used are likely words that **you this foreigner** may not know or understand. It's truly surprising.’

(zhTenTen11-50532892).

Moreover, cases such as (22) account for a substantial proportion of the non-impolite usage of *nǐ zhè(ge) + NP* (and of *nǐ gè + NP*). Banter cannot be regarded as impolite since it is not meant and/or taken to have negative emotional effects on the addressee. In fact, it serves to signal and create intimacy between people. In our view, however, the way in which banter achieves that crucially depends on the potential for impoliteness (see Leech, 1983: 142-145 too). Support for this position comes from the fact that, as Culpeper (2011: 213-215) points out, some interlocutors may still feel hurt despite knowing that the speaker does not intend to cause offense. The many instances of banter can be seen as a further indication of the association of *nǐ zhè(ge) + NP* (and *nǐ gè + NP*) with impoliteness **assessments**.

4.2.2 *Nǐ zhè(ge) + NP* versus *nǐ gè + NP*

Conventionalization is a matter of degree as it is “a correlate of the (statistical) frequency with which an expression is used in one’s experience in a particular context” (Terkourafi, 2005b: 213). Although all three constructions can be argued to be partly conventionalized for impoliteness, the degree to which varies. *Nǐ gè + NP* is significantly more impolite than both *nǐ zhè(ge) + NP* ($\chi^2 = 8.87$, $p < .005$) and *nǐ zhè + NP* ($\chi^2 = 6.20$, $p < .05$). The proportion of

impolite usage of *nǐ zhè* + NP is higher than that of *nǐ zhè(ge)* + NP but not significantly so ($\chi^2 = 0.24$, $p > .05$). The difference between *nǐ zhè(ge)* + NP on the one hand and *nǐ gè* + NP on the other hand is relevant for the ongoing debate about the relationships between the constructions (see Section 2.3), in that it supports the view that *nǐ gè* + NP at least is a construction distinct from the other two (see also Section 4.1).

Nǐ gè + NP also differs from *nǐ zhè(ge)* + NP in the (non-)evaluative nature of the impoliteness expressed, i.e. whether or not certain negative traits are attributed to the addressee with the noun phrase (see Section 3.4.1). Of the impolite cases of *nǐ gè* + NP, 96.26% are evaluative, which is significantly higher than both *nǐ zhè(ge)* + NP (78.31%; $\chi^2 = 14.71$, $p < .0005$) and *nǐ zhè* + NP (67.44%; $\chi^2 = 28.63$, $p < .0005$). The difference between the latter is again not statistically significant ($\chi^2 = 2.52$, $p > .05$). In other words, of the three constructions, *nǐ gè* + NP is the one most similar to YOU + NP as described by Van Olmen et al. (2023), i.e. as a construction expressing (typically negative) addressee evaluation. Another way that *nǐ gè* + NP resembles YOU + NP more than *nǐ zhè(ge)* + NP does is the use as an address rather than an argument: *nǐ gè* + NP serves as an address in 60.14% of the cases, compared to 38.26% for *nǐ zhè(ge)* + NP ($\chi^2 = 14.22$, $p < .0005$) and 38.10% for *nǐ zhè* + NP ($\chi^2 = 14.33$, $p < .0005$), which once more do not differ from each other ($\chi^2 = 0.00$, $p > .05$).

As mentioned in Section 2.3, Fu & Hu (2020) contend that *nǐ zhè(ge)/gè* + NP's argument use predates its address use and that the constructions became more expressive as they gained independence (a process that the authors regard as a case of subjectification). Our finding that *nǐ gè* + NP, which has the highest proportion of address uses, is the construction most conventionalized for impoliteness and evaluative impoliteness in particular is in line with this hypothesis. Further support for Fu & Hu's (2020) argument comes from a comparison of *nǐ zhè(ge)* + NP's impolite address and argument usage: *nǐ zhè(ge)* + NP is evaluative in 90.91% of its address uses but in just 64.10% of its argument uses ($\chi^2 = 8.75$, p

< .005) while, for *nǐ zhè* + NP, the respective proportions are 85.29% and 55.77% ($\chi^2 = 8.16$, $p < .005$). Put differently, cases where the speaker explicitly expresses an assessment of the addressee against certain qualities indeed appear to be more common in the address usage of *nǐ zhè(ge)* + NP.

5 Discussion

We hope to have shown thus far that all three constructions under investigation can be regarded as partly conventionalized for impoliteness. The question that we wish to explore here is why, even though expressions like *nǐ gè xiǎokěài* ‘you cutie’ are possible, *nǐ zhè(ge)/gè* is so well-suited for impolite purposes, by looking at the different components of the constructions.

5.1 Second person pronoun *nǐ*

One feature that the three constructions share is *nǐ* ‘you’ and, interestingly, second person pronouns have repeatedly been argued to have some kind of link with impoliteness. For instance, Giomi & Van Oers (2022), who study grammatical structures dedicated to insulting people in the world’s languages, note that such structures often feature ‘you’, possibly in its possessive form (e.g. Swedish *din idiot!* ‘you idiot!’, literally ‘your idiot!’). Van Olmen et al. (2023: 38) compare YOU + NP and ‘my’ + NP structures in Dutch, English and Polish and observe that, while the former tend to **be an antecedent of impoliteness evaluations**, the latter are typically employed for politeness (e.g. ‘my friend’). One explanation for this compatibility of the second person with impoliteness is that the use of ‘you’ increases the directness of the expression by pointing out the target explicitly (Culpeper & Haugh, 2014:

170). In ‘you bastard’, the second person pronoun is redundant in a sense, as ‘bastard’ on its own can be used to address someone. The overt reference to the addressee serves the purpose of explicitly associating them with unfavorable characteristics (Culpeper, 2005: 41). Given that Chinese is often considered to prefer indirect communication (e.g. Chen & Wang, 2021), the mere presence of *nǐ* in *nǐ zhè(ge)/gè* + NP may already gear the constructions toward impoliteness.

Using *nǐ* ‘you’ has also been argued to **evoke** a face-to-face communication environment, allowing speakers to confront a target even when the target itself is not present (Zhang, 2005). In (25), for example, with the first-person inclusive pronoun *zánmen* ‘us’, it is clear that *nǐ gè sǐdōngxi* ‘you dead thing’ does not refer to the addressee but to an absent individual, i.e. *tā* ‘he’ in this case. Instead of using *tā gè sǐdōngxi* ‘he dead thing’ to accuse the third person, the speaker opts for *nǐ* ‘you’, enabling them to accuse him more directly, as if to his face.

(25) *kàn tā hái guǎn zánmen bù guǎn nǐ gè sǐdōngxi*

‘Let's see if he still cares about us, **you dead thing!**’

(Zhang, 2005: 81)

In addition, as mentioned in previous literature (e.g. Lv, 1985: 34), *nǐ* is often used by individuals of higher social status to address those of lower social status. As a result, it may carry “a sense of superiority over others” (Cui, 2000: 50), which obviously matches well with impoliteness.

5.2 Proximal demonstrative *zhè(ge)*

Like *nǐ* ‘you’, *zhè(ge)* ‘this’ often appears to be optional. In (26), for instance, provided there are no other individuals called David in the context that this person has to be distinguished from, the truth-conditional meaning of the sentence would remain the same if the demonstrative was omitted.

- (26) *zhège* *David* *zhēn* *tǎoyàn*.
 DEM David really annoying
 ‘**This David** is so annoying.’

If Rybarczyk (2015: 51) is right in writing that “non-obligatory items at the level of syntax can be seen as indicators of implicated meanings”, the occurrence of the demonstrative must contribute to the meaning of the sentence in some way. One likely account is that the demonstrative here performs the function of an attitudinal marker, expressing the emotion of the speaker toward the referent, a phenomenon which has been observed in many languages (e.g. Lakoff, 1974 for English; Rybarczyk, 2015 for Polish). For German, for instance, Averintseva-Klish (2016) points out that the optional use of proximal demonstrative *dies-* ‘this, these’ evokes affective meaning, which is largely negatively biased. The proximal demonstrative *zhè(ge)* ‘this’ in Chinese is similar to the German one in the sense that it also shows a tendency toward a pejorative interpretation (Zhang, 2005). This negative bias may be related to the basic function of proximal demonstratives, i.e. to directly point at the referent, which has the potential to be face-threatening (Averintseva-Klish, 2016; Zhang, 2005).

When combined with *nǐ* ‘you’, the referent of the demonstrative and that of the addressee are the same. Consequently, the speaker’s negative attitude toward the referent evoked by *zhè(ge)* ‘this’ gets directed toward the addressee and this can be said to contribute to *nǐ zhè(ge)* + NP’s well-suitedness for impoliteness **assessments**. This negative affective

function of *zhè(ge)* ‘this’ also explains the frequent occurrence of evaluatively neutral nouns such as *rén* ‘person’ and *érzi* ‘son’ in the two constructions (see Section 4.1). Their impoliteness **assessments** need not derive from an evaluation of the addressee against certain qualities expressed in the noun phrase but can simply come from the demonstrative.

5.3 Classifier *gè* ¹¹

Researchers have pointed out that classifiers can be employed to convey speaker attitude toward a referent (e.g. Contini-Morava & Kilarski, 2013: 277; Deng et al., 2020). Song & Allasonniere-Tang (2021: 122), for instance, note that, in Chinese, “referring to a human via different classifiers expresses diverse levels of respect towards the referent”. Consider the classifiers *gè* and *wèi* in (27) and (28).

(27) *yī gè lǎoshī*

one CLF teacher

‘one teacher’

(28) *yī wèi lǎoshī*

one CLF teacher

‘one teacher’

Example (27) is more informal than (28), and conveys less respect toward the referent (Song & Allasonnier-Tang, 2021: 122). An explanation for this difference lies in the fact that the

¹¹ We acknowledge the debate regarding the status of *gè* in *nǐ gè* + NP, with some researchers classifying it as classifier (e.g. Lv, 1985: 201-202), while others as demonstrative (Fu & Hu, 2020). In this study, we have analyzed it as a classifier, as its generic function aligns well with impoliteness. In addition, future research could collect prosodic data to explore the effect of prosody (e.g. tone neutralization) on (im)politeness assessments, as suggested by one of the reviewers.

classifier *wèi* is specialized for human reference whereas *gè*, one of the most frequent classifiers in Chinese, can combine with many types of nouns, including those denoting animals and general things (Biq, 2004). The generic character of the latter may produce an effect of lack of formality and/or respect when used with human referents. Its presence in *nǐ gè* + NP then obviously makes the construction especially convenient for impoliteness.

6 Conclusion

Drawing on corpus data and analyzing it automatically through multiple distinctive collexeme analysis as well as qualitatively through careful consideration of context, this article has contributed to the discussions about the potential inherency of impoliteness and about the relationships between the three constructions under investigation.

Our conclusion that *nǐ zhè(ge)/gè* + NP should be regarded as partially conventionalized for impoliteness is supported by the sheer frequency with which, in actual usage, the constructions serve impolite purposes. It is the predominant function of each construction, particularly if one considers the diversity of the other uses and the fact that their (non-impolite) usage for banter relies on the potential for impoliteness. *Nǐ zhè(ge)/gè* + NP's partial conventionalization for impoliteness is also evident from the fact that the constructions have the ability to impose a negative reading on evaluatively neutral noun phrases. Overall, our results support the claim that impoliteness is not solely contextual but can also be conventionalized in linguistic form (Culpeper, 2011; Van Olmen et al., 2023).

Our findings also indicate that, contrary to Zhang (2005) and others but in line with Fu & Hu (2020), *nǐ gè* + NP should not be regarded as a reduced form of *nǐ zhè(ge)* + NP but as a construction in its own right that, of the three, most closely resembles YOU + NP in typically expressing negative addressee evaluation. It differs significantly from both *nǐ zhège*

+ NP and *nǐ zhè* + NP in terms of (i) the types of nouns that are attracted to the construction, (ii) its greater proportion of impolite instances, (iii) its higher number of evaluative impolite ones and (iv) its larger proportion of address uses. *Nǐ zhège* + NP and *nǐ zhè* + NP, by contrast, are very similar to each other in all these respects, which might be taken to suggest that the latter is indeed a reduced variant of the former, with *gè* being omitted perhaps for reasons of economy.

Future research into *nǐ zhè(ge)/gè* + NP could involve questionnaire asking for judgments about well-formedness –for instance, to test how compatible *nǐ gè* + NP is with non-evaluative noun phrases – and/or about degrees of (im)politeness –for example, to assess the respective contributions of *nǐ*, *zhè(ge)* and *gè* or to see to what extent the construction coerce an insulting interpretation on noun phrases that are not negatively evaluative.

Declarations

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Averintseva-Klisch, M. (2016). Demonstrative pejoratives. In R. Finkbeiner, J. Meibauer & H. Wiese (Eds.), *Pejoration* (pp. 119-142). Benjamins.
<https://doi.org/10.1075/la.228.06ave>
- Biq, Y. O. (2004). Construction, reanalysis, and stance: ‘V yi gè N’ and variations in Mandarin Chinese. *Journal of Pragmatics*, 36(9), 1655-1672.
<https://doi.org/10.1016/j.pragma.2003.11.009>

- Chen, X., Li, M. (2023). Personality and (im)politeness: Evidence from WeChat/QQ group chats. In Xie, C. (Ed.), *Advances in (im)politeness studies* (pp.51-71). Springer, Cham.
https://doi.org/10.1007/978-3-031-37064-9_4
- Chen, X., & Wang, J. (2021). First order and second order indirectness in Korean and Chinese. *Journal of Pragmatics*, 178, 315-328.
<https://doi.org/10.1016/j.pragma.2021.03.022>
- Contini-Morava, E., & Kilarski, M. (2013). Functions of nominal classification. *Language Sciences*, 40, 263-299. <https://doi.org/10.1016/j.langsci.2013.03.002>
- Corver, N. (2008). Uniformity and diversity in the syntax of evaluative vocatives. *The Journal of Comparative Germanic Linguistics*, 11, 43-93.
<https://doi.org/10.1007/s10828-008-9017-1>
- Cui, X. (2000). Rénchēng dàcí jí qí chēngwèi gōngnéng (Personal Pronouns and their appellative functions). *Language Teaching and Linguistic Studies*, 1, 46-54.
- Culpeper, J. (2005). Impoliteness and entertainment in the television quiz show: *The Weakest Link*. *Journal of Politeness Research*, 1(1), 35-72. <https://doi.org/10.1515/jplr.2005.1.1.35>
- Culpeper, J. (2010). Conventionalised impoliteness formulae. *Journal of Pragmatics*, 42(12), 3232-3245. <https://doi.org/10.1016/j.pragma.2010.05.007>
- Culpeper, J. (2011). *Impoliteness: Using language to cause offence*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511975752>
- Culpeper, J., Bousfield, D., & Wichmann, A. (2003). Impoliteness revisited: With special reference to dynamic and prosodic aspects. *Journal of Pragmatics*, 35(10-11), 1545-1579.
[https://doi.org/10.1016/S0378-2166\(02\)00118-2](https://doi.org/10.1016/S0378-2166(02)00118-2)

- Culpeper, J., & Haugh, M. (2014). *Pragmatics and the English language*. Bloomsbury.
<https://doi.org/10.48548/pubdata-197>
- d'Avis, F., & Meibauer, J. (2013). Du Idiot! Din idiot! pseudo-vocative constructions and insults in German (and Swedish). In B. Sonnenhauser & P. N. A. Hanna (Eds.), *Vocative! Addressing between system and performance* (pp. 189-218). De Gruyter Mouton.
<https://doi.org/10.1515/9783110304176.189>
- Deng, Y., Yap, F. H., & Chor, W. (2020). Negative attitudinal uses of quantifying classifier di45 in Wugang Xiang. *Journal of Pragmatics*, 160, 14-30.
<https://doi.org/10.1016/j.pragma.2020.02.010>
- Eelen, G. (2001). *A critique of politeness theory*. Manchester: St Jerome.
<https://doi.org/10.4324/9781315760179>
- Fu, H., & Hu, J. (2020). “Nǐ zhè(ge) + NP” de xíngchéng yǔ fāzhǎn (The formation and development of “nǐ zhè(ge) + NP” . *Journal of Chinese Historical Linguistics*, 1, 73-83.
- Giomi, R., & van Oers, D. (2022). Insultive constructions: A crosslinguistic perspective. Paper presented at *55th Annual Meeting of the Societas Linguistica Europaea*, Bucharest.
- Gries, S. T., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective on ‘alternations’. *International Journal of Corpus Linguistics*, 9(1), 97-129.
<https://doi.org/10.1075/ijcl.9.1.06gri>
- Hu, Q., & Gao, Q. (2018). Rènzhī cānzhàodiǎn yǔ “nǐ zhè(ge) + NP” gòushì (Cognitive reference points and the “nǐ zhè(ge) + NP” construction). *Chinese Language Learning*, 2, 44-54.
- Jain, K. H. (2022). You Hoboken! Semantics of an expressive label maker. *Linguistics and Philosophy*, 45(2), 365-391. <https://doi.org/10.1007/s10988-021-09333-y>

- Julien, M. (2016). Possessive predication vocatives in Scandinavian. *Journal of Comparative Germanic linguistics*, 19, 75-108.
<https://doi.org/10.1007/s10828-016-9081-x>
- Kádár, D. Z., Zhang, S., 2019. Approaches to (Chinese) linguistic politeness. *Foreign Languages and Their Teaching*, 6, 18–28. <https://doi.org/10.13458/j.cnki.flatt.004630>
- Lakoff, R. (1974, April). Remarks on this and that. *Proceedings of the Chicago Linguistics Society*, 10, 345-356.
- Leech, G.N. (1983). *Principles of pragmatics* (1st ed.). Routledge.
<https://doi.org/10.4324/9781315835976>
- Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. Benjamins. <https://doi.org/10.1075/z.195>
- Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. University of California Press.
- Locher, M. A., & Watts, R. J. (2008). Relational work and impoliteness: Negotiating norms of linguistic behavior. In D. Bousfield, & M. A. Locher (Eds.), *Impoliteness in language: Studies on its interplay with power in theory and practice* (pp. 77-99). Mouton de Gruyter.
<https://doi.org/10.1515/9783110208344.2.77>
- Lv, S. X. (1985). *Jìndài Hànyǔ zhǐdàicí (Modern Chinese referential pronouns)*. Xuelin.
- Potts, C. & Roeper, T. (2006). The narrowing acquisition path: From expressive small clauses to declarative. In L. Progovac, K. Paesani, E. Casielles, & E. Barton (Eds.), *The syntax of nonsententials: Multi-disciplinary perspectives* (pp. 183-201). Benjamins.
<https://doi.org/10.1075/la.93.09pot>

- R Core Team (2022). *R: A language and environment for statistical computing*. R foundation for statistical computing. <https://www.R-project.org/>
- Rybarczyk, M. (2015). *Demonstratives and possessives with attitude: An intersubjectively-oriented empirical study*. Benjamins. <https://doi.org/10.1075/hcp.51>
- Song, N., & Allasonnière-Tang, M. (2021). The diversity of classifier inventory in Mandarin dialects: A case study of Baoding. *Faits de Langues*, 52(2), 115-132. <https://doi.org/10.1163/19589514-05202001>
- Tao, H. (1999). The grammar of demonstratives in Mandarin conversational discourse: A case study. *Journal of Chinese Linguistics*, 27(1), 69-103.
- Terkourafi, M. (2005a). Beyond the micro-level in politeness research. *Journal of Politeness Research*, 1(2), 237-262. <https://doi.org/10.1515/jplr.2005.1.2.237>
- Terkourafi, M. (2005b). Pragmatic correlates of frequency of use: The case for a notion of “minimal context”. In S. Marmaridou, K. Nikiforidou & E. Antonopoulou (Eds.), *Reviewing linguistic thought: Converging trends for the 21st century* (pp. 209-234). De Gruyter. <https://doi.org/10.1515/9783110920826.209>
- Van Olmen, D., Andersson, M., & Culpeper, J. (2023). Inherent linguistic impoliteness: The case of insulting you+ np in Dutch, English and Polish. *Journal of Pragmatics*, 215, 22-40. <https://doi.org/10.1016/j.pragma.2023.06.013>
- Wang, J., & Taylor, C. (2019). The conventionalisation of mock politeness in Chinese and British online forums. *Journal of Pragmatics*, 142, 270-280. <https://doi.org/10.1016/j.pragma.2018.10.019>
- Watts, R. J. (2003). *Politeness*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511615184>

- Yip, P.-C., & Rimmington, D. (2004). *Chinese: A comprehensive grammar*. Routledge.
<https://doi.org/10.4324/9780203880722>
- Zhang, H., & Yin, H. (2004). “Nǐ zhè(ge) + NP” jiégòu de duō jiǎodù kǎochá (A multi-perspective study of “nǐ zhè(ge) + NP”). *Journal of Xuzhou Normal University (Philosophy and Social Sciences Edition)*, 2, 75-78.
- Zhang, X. (2005). “Nǐ zhège + NP!” de biǎodá gōngnéng yánjiū (A study on the expressive function of “nǐ zhège + NP!”). *Chinese Teaching in the World*, 4, 79-84.
- Zlov, V., & Zlatev, J. (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.
<https://doi.org/10.1515/pr-2022-0017>