Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-024-03134-x OPEN

Register-based distribution of expressions of modality in COCA

Jiangping Zhou [1]✉ & Yanhua Xia[1]

This research scrutinizes expressions of modality at both word and clause ranks. Expressions at the former rank encompass modal verbal operators, modal adjuncts, and modal nominalizations, and expressions at the latter rank encompass 'explicit subjective' orientation such as *I think* and 'explicit objective' orientation, such as *it is possible*. These latter expressions are designated as interpersonal metaphors of modality within the framework proposed by Halliday. Previous studies delimited the distribution of expressions of modality either at word rank or at clause rank; in addition, they undertheorized the relationship between this distribution of modality and formality of registers because the register is testified to be an influential factor in the distribution of expressions of modality. This study, drawing on such statistical methods as variation analysis and correlation analysis, explored the distribution of expressions of modality at both word and clause ranks across registers with different degrees of formality by employing the Corpus of Contemporary American English (COCA). The findings are summarized as follows. First, registers in COCA employ expressions of modality with low and/or median values to judge the proposition at issue for the purpose of entertaining other varying voices. Second, in more formal registers, language users seem to prefer modal nominalizations to modal verbal operators and modal adjuncts at word rank, and 'explicit objective' orientation to 'explicit subjective' orientation at clause rank. In doing so, they could efficiently conceal the commentators' subjective meaning, and thus, the objectivity of propositions is highlighted. This study will shed light on the full understanding of the distribution of expressions of modality across registers and the application of such theory as interpersonal metaphor of modality in SFL.

[1] China West Normal University, Nanchong, China. ✉email: zhoujp@pku.edu.cn

Introduction

According to Halliday and Matthiessen (2014), language is a complex semiotic system composed of three strata: phonology/graphology, lexico-grammar, and semantics. In Systemic Functional Linguistics or SFL, the modal meaning in the semantic stratum is realized by expressions of modality in the lexicogrammatical stratum; this realization happens either within the clause (e.g., modal verbal operators)[1] or outside the clause (e.g., 'explicit objective' orientation) (Taverniers, 2003). Consider the following examples,[2] wherein expressions of modality are italicized.

(1) [redacted]

a. One consideration that *might* have played a part is a gender imbalance on the US governing board. (COCA_newspaper_2019)
b. The sex differences in auditory acuity are *possibly* linked to the sex differences in vocal production in rats. (COCA_academic_2019)
c. And a lot of folks are getting very excited about the *possibility* of phase one in the trade deal happening. (COCA_spoken_2019)
d. *I suppose* I have to backtrack on that a little bit. (COCA_magazine_2016)
e. Although the molecular determinants of dynein force production are not well understood, *it is possible* that these mutations specifically affect the ability of dynein to remain bound to microtubules under conditions of high load. (COCA_academic_2019)

Modal meanings in (1a–e) are realized by the modal verbal operator *might* in (1a), the modal adjunct *possibly* in (1b), the modal nominalization *possibility* in (1c), the 'explicit subjective' orientation *I suppose* in (1d), and the 'explicit objective' orientation *it is possible* in (1e) respectively. Expressions of modality in the first three examples (i.e., 1a–c) could be regarded as expressions of modality at word rank[3], realizing the modal meaning within clauses, while those in the last two examples (i.e., 1d-e) could be considered as expressions of modality at clause rank (or interpersonal metaphor of modality in this research), realizing the modal meaning outside clauses.

There are three lines of studies that have examined expressions of modality. The first line pertains to the restricted consideration of expressions of modality based on different theoretical traditions (e.g., Nuyts, 1993; Halliday, 1994; Palmer, 2001; Matthews, 2003; Charnock, 2009). For example, Nuyts (1993) in the traditional sense documented the difference between constructions with modal adverbs, i.e. modal adjuncts at issue, and constructions with predictively used modal adjectives, i.e. 'explicit objective' orientation in question, in terms of three different underlying factors: the discourse functionality of the constructions, the interaction of the epistemic modal qualification with evidential marking, and the performative vs. descriptive nature of the modal expression. In addition, specific adverbs in modal adjuncts, i.e. *maybe* and *perhaps*, were examined with respect to their use in the context (Suzuki, 2018) and the multifunctionality of the 'possible' modal adjuncts by comparing with *conceivably* with *perhaps* (Suzuki and Fujiwara, 2017). The former study uncovered that *maybe* is more prone to subjective use while *perhaps* is a more strongly grammaticalized item; the latter study demonstrated that *conceivably* and *perhaps* display opposite functional characteristics, and the factors influencing the use of these adverbs are strongly associated with the contexts of modality and discourse. Issues in this line are two-fold. First, investigation of subtypes of expressions of modality is restricted. Foci of these studies were either on modal verbal operators (Palmer, 2001), or on predictively used modal adjectives (Nuyts, 1993), or on modal adjuncts (Nuyts, 1993; Suzuki and Fujiwara, 2017; Suzuki, 2018). Second, dynamic nature of these expressions of modality needs to be addressed intricately. Variation or change is the nature of language (de Saussure, 1959; Zhou, 2023a), and thus expressions of modality are of no exception. Therefore, expressions of modality should be explored in the context of different registers.

The second line concerned the distribution of modal verbal operators both diachronically (e.g., Leech, 2011; Leech, et al. 2009) and synchronically (e.g., Biber, et al. 1999; Collins, 2009). In relation to the diachronic studies, Leech, et al. (2009) for example examined the frequency change of modal verbal operators (modal auxiliaries in their research) and documented that some core modal auxiliaries are losing their deontic and epistemic uses and gradually substituted by semi-modal auxiliaries in the process of grammaticalization (*must – have (got)to – want to / need to*). In relation to the synchronic studies, Collins (2009) explored the distribution of modals and quasi-modals in different varieties of English and uncovered that there is strong tendency for American English to be leading the way in the English language developments. Additionally, it is uncovered that quasi-modals flourish in speech while their modal counterparts are favored more in written registers. Apart from the synchronic investigation of the distribution of modals and quasi-modals, some other modal verbal operators such as *should* and *ought to* were also examined in terms of their distribution in present-day British English (Verhulst, et al. 2013). However, a fuller scope of expressions of modality, inclusive of both subtypes of expressions of modality at word rank and those at clause rank, is not systematically concerned, nor is the influencing factor of registers.

The third line of research documented the distribution of expressions of modality in English across registers (e.g., He, 2020; Zhou, 2023a, 2023b, 2023c). For example, He (2020) compared normalized frequencies of specific subtypes of interpersonal metaphor of modality in different registers. It is uncovered that 'explicit objective' orientation is register-sensitive and occurs preferentially in formal registers such as academic texts, while 'explicit subjective' orientation occurs preferentially in informal registers such as spoken texts and fiction. However, he limited his study within modal verbal operators, modal adjuncts, 'explicit subjective' orientation, and 'explicit objective' orientation (i.e., congruent and metaphorical expressions of modality in his study) and left modal nominalizations untouched. In addition, he examined the diachronic distribution of these expressions in such registers[4] as fiction, magazines, newspapers, and non-fiction, leaving the academic register which preferentially accommodates metaphorical expressions (i.e., the 'explicit objective' orientation in this study) in Hallidayan sense (cf. Liardét, 2018) unexamined. Zhou (2023a) restricted his topic within interpersonal metaphor of modality (cf. Section 2), particularly the 'explicit objective' orientation, and he later extended his topic to the relationship between congruent expressions of modality and metaphorical expressions of modality across registers (2023b) or to the attracted expressions of modality to the construction of interpersonal metaphor of modality (2023c). Based on Zhou's studies, the synchronic distribution of all subtypes of expressions of modality in terms of the formality of registers is undertheorized.

Against the backdrops, this study aims to investigate the synchronic distribution trends of a fuller coverage of expressions of modality across registers, specifically the relationship between expressions of modality and the formality of registers, and the motivation underlying this relationship. In fulfilling our aims, the following research questions are proposed.

1. How are expressions of modality synchronically distributed across registers?
2. What is the relationship between expressions of modality and formality of registers, and what explanations can be offered for the relationship?

This paper is structured as follows. Section 2 delineates subtypes of expressions of modality. Section 3 introduces the information of COCA, confirmation of formality of registers, methods, and the way that data are collected. Sections 4 and 5 consider the distribution of subtypes of expressions of modality and the general distribution trend of expressions of modality, respectively. Results are discussed in Section 6, and a conclusion is made in Section 7.

Theoretical foundations

Expressions of modality reflect the speaker's judgment of his/her saying (Fontaine, 2013). They incorporate expressions of modality at word rank and expressions of modality at clause rank (cf. Halliday and Matthiessen, 2014). The former is further subcategorized into modal verbal operators, modal adjuncts, and modal nominalizations; the latter is further classified into 'explicit subjective' orientation and 'explicit objective' orientation (Zhou, 2023c). In this section, key terms in relation to expressions of modality and interpersonal metaphor of modality in SFL, a concept employed to expound the relationship between expressions of modality and formality of registers, are delineated. Subsection 2.1 profiles expressions of modality at word rank, and subsection 2.2 expounds the concept of interpersonal metaphor of modality, which corresponds to expressions of modality at clause rank, i.e., 'explicit subjective' orientation and 'explicit objective' orientation within SFL (Halliday, 1994).

Expressions of modality at word rank. In English, modal verbal operators (Modal auxiliaries in Nuyts' (1993) term, or modal auxiliary verbs in Fontaine's (2013) term, or core modal auxiliaries in Leech's (2011) term), generally include *can, could, may, might, shall, should, will, would*, and *must*. They share a number of typical grammatical properties (e.g., negation and inversion) that make them distinguishable from other modal operators such as peripheral/marginal modal operators or semi-modal operators (e.g., *need, have to, seem to*).[5] These grammatical properties include the main verb in the bare infinitive form and the lack of non-tensed form or person-number agreement. These modal operators, except for *must*, are usually paired as present and past tense forms of single lexemes (i.e., *can/could, may/might, shall/should, will/would*). In this paper it is useful to regard these pairings as individual lexemes, although their relationships are complex, because we only consider their distribution across registers, precluding the semantic differences that lie in each pair. Consider the examples in (2).

(2) [redacted]

a. The storm system *could* affect tens of millions of people in Virginia. (COCA_spoken_2016)
b. The Golden Eagle Inn *might* have a room for you. (COCA_magazine_1997)
c. The warm January *should* bring a big drop in oil and gas prices, right? (COCA_newspaper_1990)
d. The train *would* round the bend, and the rails *would* start to quake a little underfoot. (COCA_fiction_2019)

Could, *might*, *should*, and *would* in (2a-d) exemplify modal verbal operators in each pair. All of them demonstrate the grammatical properties of being followed by a bare infinitive verb form and the lack of person-number agreement between the subject and the modal operator.

Modal adjuncts are adverbials that demonstrate speakers' subjective comment and attitude towards a proposition. Halliday and Matthiessen (2014: 177–178) argued that modality is subcategorized into modalization (referring to the scales of probability and usuality) and modulation (referring to the scales of obligation and inclination). They (2004: 618) expounded that "[i]n philosophical semantics probability is referred to as 'epistemic' modality and obligation as 'deontic' modality". Considering modal adjuncts at issue, those expressing modalization include *certainly, probably, possibly*, and *surely*, and those expressing modulation include *essentially, necessarily*, and *obligatorily*.

Modal nominalization, which is named analogously from Halliday and Matthiessen's (2014: 710) terms such as verbal nominalization and adjectival nominalization, is generally derived from modal adjectives (e.g., from *necessary* to *necessity*) or modal verbs (e.g., from *can* to *ability* or from *may* to *possibility*). Modal nominalizations that express modalization generally include *certainty, probability, possibility*, and *likelihood*, and those that express modulation include *essentiality, necessity*, and *obligation*.

Modal adjuncts and modal nominalizations in relation to expressions of modality at word rank are exemplified in (3) and (4), respectively.

(3) [redacted]

a. She doesn't mean anything by it. She's half-drunk and *probably* has no idea what she's saying. (COCA_academic_2009)
b. So, consumers, when they walk down the grocery aisle, are *necessarily* even going to know what they're buying. (COCA_spoken_1992)

(4) [redacted]

a. The figure shows the *probability* of stroke-free survival in the two treatment groups. (COCA_academic_1991)
b. As a matter of practical *necessity* the state may have to approve appropriations despite foreseeable harm to public trust uses. (COCA_academic_1996)

Value is an important parameter in SFL while denoting the meaning of modality. An expression of modality could be of a high, median, or low value. "Value" used in this sense is synonymous to the more common term in the mainstream literature, i.e., "strength". It should be noted that some minor differences also exist between the two terms. For instance, modal verbal operators *shall* and *will* are regarded as expressions with median modal value in SFL, while they are with high modal value outside SFL. This research follows the SFL framework (underlying the fact that recent studies on the distribution of expressions of modality are generally SFL-based), so they are categorized into modal verbal operators denoting modality with median value. The aforementioned items that are included in the three subtypes of expressions of modality are tabulated in Table 1. It should be also noted that Table 1 only highlights the modal value of subtypes of expressions of modality at word rank, so the semantic difference between *certainly* and *obligatorily* is not addressed. The former expression of modality denotes the epistemic meaning and the latter the deontic meaning, which corresponds to Halliday and Matthiessen's (2004) terms of Probability and Obligation, respectively.

Interpersonal metaphor of modality. Interpersonal metaphor of modality involves the remapping of modal meanings on the semantic stratum onto expressions of modality at different ranks on the lexicogrammatical stratum (Zhou, 2023c). Metaphors of modality, alongside interpersonal metaphors of mood, represent one of two types of interpersonal metaphor related to the ways in which propositions are negotiable or not (Yang, 2019). Modality, a subsystem of interpersonal meaning, includes four main types: probability, e.g., *it is possible, likely*, usuality, e.g., *it is usual, customarily*, obligation, e.g., *it is obligatory, necessarily*, and inclination, e.g., *will*

Table 1 Instances of expressions of modality with various modal values.

		Expressions of modality at word rank		
		Modal operators	**Modal adjuncts**	**Modal nominalizations**
value	high	*must*	*certainly, surely, obligatorily*	*certainty, obligation*
	median	*will/would, should/shall*	*probably, necessarily, essentially*	*probability, necessity, essentiality*
	low	*can/could, may/might*	*possibly*	*possibility, likelihood*

Table 2 Instances of interpersonal metaphor of modality with various values.

		Interpersonal metaphor of modality	
		'explicit subjective' orientation	**'explicit objective' orientation**
values	high	*I/we believe*	*it's certain/sure/obligatory*
	median	*I/we guess*	*it's probable/necessary/essential*
	low	*I/we think/assume/suppose*	*it's possible/likely/vital*

(Halliday and Matthiessen, 2014). Interpersonal meaning, which enacts the interpersonal relationships between interlocutors, is one of the three meanings denoted by a clause. The other two meanings are ideational meaning that construes experiences of the world through language and textual meaning that packages ideational and interpersonal meanings to form texts in contexts. Each of these four types of modality can be realized by means of different orientations that represent different degrees of judgment on the validity of a proposition, either subjective or objective, and either explicit or implicit. Subjective orientation frames the meaning of modality as the speaker's subject judgment on the validity of a proposition, e.g., *I think*, *must*, etc.; objective orientation frames modality as an objective evaluation, e.g., *it's likely*, *possibly*, etc.; explicit orientation directly states the source of judgment in a separate clause, e.g., *I think*, *it's likely*, etc.; while implicit orientation leaves the source of judgment implicit, e.g., *must*, *possibly*, etc. (Halliday and Matthiessen, 2014, p. 181). SFL scholars argue that only 'explicit subjective' and 'explicit objective' orientations realize interpersonal metaphors of modality (cf. Taverniers, 2003; Halliday and Matthiessen, 2014; He, 2020).

'Explicit subjective' orientation is constructed by the formation of a pronoun in its first personal nominative case (i.e., *I* and *we*) and a mental process verb such as *think*, *believe*, *guess*, etc. Specific instances of this type that are employed to explore their register-based distribution are taken from Zhou and Gao's (2021) research, who used them to investigate their diachronic distributions during the past two centuries in American English, including *I/we think* (*assume, believe, guess, suppose*). Consider the following examples in (5). *I think* in (5a) and *we believe* in (5b) are instances of 'explicit subjective' orientation in interpersonal metaphor of modality in that speakers' judgment is positioned outside the projected clause or the proposition i.e., *that culture is changing, and it's changing for the better* in (5a) and *women are going to make the winning difference in this campaign and in this election* in (5b). They are subjective because of the first-person pronoun *I* in (5a) and *we* in (5b) realizing the subject of the projecting clauses respectively.

(5) [redacted]

a. *I think* that culture is changing, and it's changing for the better. (COCA_spoken_2019)
b. *We believe* that women are going to make the winning difference in this campaign and in this election. (COCA_spoken_2000)

'Explicit objective' orientation is structured by the sequence of *it*, any form of the copula *be*, a modal adjective, and a *to/that* clause. This grammatical pattern is also termed by other scholars as "*it*-extraposition" (Kaltenböck, 2005). Consider the examples in (6). *It was likely* in (6a) and *it is obligatory* in (6b) are instances of 'explicit objective' orientation in terms of interpersonal metaphor of modality in that the modal meanings of probability, i.e., *likely*, and obligation, i.e., *obligatory*, are realized outside the propositions. Variations of this grammatical pattern could also be attested. That is, the projecting clause could either have the copula *be* been modified by a modal verbal operator or have the modal adjective been qualified by an adverb. Variations of this kind are instantiated by (7a-b). In (7a), the modal verbal operator *might* is used to modify the copula *be* in the projecting clause *it might be possible*, and in (7b), the adverb *entirely* is interpolated in the projecting clause *it's entirely possible* to qualify the modal adjective *possible*.

(6) [redacted]

a. This told Bosch that *it was likely* that the woman had been a photographer. (COCA_fiction_2012)
b. Because of all of what I have mentioned it seems that *it is obligatory* to pay respect to the tablets of the loudspeaker, that is, to the cylinders. (COCA_academic_2018)

(7) [redacted]

a. *It might be possible* that a respectable case could be made in favor of them. (COCA_magazine_1990)
b. *It's entirely possible* that this year's draft could unfold like last year's. (COCA_newspaper_2019)

Interpersonal metaphors of modality could be also of various values. 'explicit subjective' orientation and 'explicit objective' orientation, with respect to high, median, or low value, are presented in Table 2.[6]

Methodology

Corpus. The on-line corpus COCA (cf. Davies, 2010) is employed in this research.[7] COCA is the only large, genre-balanced corpus of American English,[8] ranging from 1990 to 2019 and containing more than one billion running words of text (words considered in the research are exactly 618,200,644, because we do not consider such registers as the general web, blog, and TV/movies).[9] Registers considered in this study

Table 3 Information of registers in COCA.

Registers	Texts	Words	Description
Spoken texts	44,803	127,396,932	Conversations from different TV and radio programs
Fiction	25,992	119,505,305	Short stories, plays, and fan fiction
Magazines	86,292	127,352,030	Nearly 100 different magazines
Newspapers	90,243	122,958,016	Newspapers from across the US
Academic texts	26,137	120,988,361	More than 200 different peer-reviewed journals
total	273,467	618,200,644	

include spoken texts, fiction, magazines, newspapers, and academic texts. The general information on registers at issue in COCA is tabulated in Table 3.

Formality of registers. With respect to the formality of registers under consideration, they could be either very informal (e.g., spoken texts), or very formal (e.g., academic texts), or somewhere in between (e.g., magazines, newspapers) (cf. overview information on registers in COCA by Mark Davies (2010)). In this research, we employed the formula proposed by Heylighen and Dewaele (1999) and also testified by Zhou (2023b, 2023c) to establish a continuum of formality across registers[10] based on two considerations. One is that the formula counts nouns positively but verbs negatively in establishing the formality of a register, which is in accordance with the argument that nouns are preferred in a formal style while verbs are preferred in an informal style (cf. Heylighen and Dewaele, 1999; Biber and Gray, 2016). Additionally, adjectives, prepositions, and articles are used to specify additional details that nouns contain, while adverbs are used to supplement information to verbs. Pronouns and interjections are closely associated with the context in that their meanings are recovered from the linguistic and situational contexts; context-dependent word classes are generally used in an informal style (cf. Heylighen and Dewaele, 1999). The other is that the formula is practicably operationalizable in that frequencies of each word class could be directly obtained by constructing relevant search queries (e.g., [v*] and [nn*] could be implemented to retrieve all occurrences of verbs and nouns, respectively, in COCA). By implementing the formula, the working continuum of formality of registers is confirmed. Precisely, from the most informal register to the most formal one, these registers in question follow the sequence of spoken texts (normalized frequency or NF = 47.21), fiction (NF = 49.26), magazines (NF = 58.94), newspapers (NF = 59.07), and academic texts (NF = 64.97). The formality of registers confirmed by Heylighen and Dewaele's (1999) formula is also in accordance with that proposed by Biber (1988) and Biber and Conrad (2009) who captured register differences based on a multi-dimensional analysis.

Data collection. For obtaining occurrences of subtypes of expressions of modality in COCA, relevant search queries or SQs were constructed. We first retrieved occurrences of subtypes of expressions of modality at word rank according to SQs 1, 2, and 3. SQ1 reads as a sequence of modal verbal operators at issue and a verb in its bare infinitival form; this query will successfully and exhaustively retrieve such clauses as exemplified in (2). SQ2 was used to retrieve modal adjuncts at issue, and hence, such clauses as exemplified in (3) will be obtained. In a similar vein, modal nominalizations that occur in such clauses as shown in (4) are obtained by SQ3.

SQ1: [vm*] [v?i*]

SQ2: certainly|probably|possibly|surely|essentially|necessarily|obligatorily

SQ3: certainty|probability|possibility|essentiality|necessity|obligation|likelihood

For retrieving occurrences of "explicit subjective" orientation and 'explicit objective' orientation in interpersonal metaphor of modality, SQs 4 and 5 were constructed respectively. Occurrences of the former type were retrieved by SQ4, which reads as a grammatical pattern in the sequence of a punctuation mark, a personal pronoun, a mental process verb, a pronoun or noun, and a verb. This query will successfully retrieve such clauses as exemplified in (5). Those of the latter type as shown in (6) will be retrieved by SQ5, which means a grammatical pattern in the sequence of *it*, any form of copula *be*, any modal adjective, and a *to* or *that* clause. Owing to the fact that SQ5 could not efficiently retrieve such clauses as shown in (7), SQ6, which means the copula *be* is modified by a modal auxiliary, is supplemented to SQ5 to retrieve such clauses as shown in (7a), and SQ7, which reads as the modal adjective is qualified by an adverb, is supplemented to SQ5 for the purpose of retrieving such clauses as shown in (7b).

SQ4: [y*] [p*] think|assume|believe|guess|suppose [pp*]|[nn*] [v*]

SQ5: it [vb*] [j*] to|that

SQ6: it [vm*] [vb*] [j*] to|that

SQ7: it [vb*] * [j*] to|that

Raw frequencies by applying the seven search queries were first normalized to per million words (PMW) for the ease of comparison, and then converted according to the principle of equal totality if necessary (cf. He, 2019; Zhou and Gao, 2022).

Statistical analysis. For the purpose of quantitatively examining the distribution of expressions of modality across registers, this research employed some statistical measures in R language, a project for statistical computing (see R project's webpage: https://www.r-project.org/), to examine the variation or correlation between variables. The variation analysis is employed to answer the first research question and the correlation analysis is used to address the second research question. Specifically, the Shapiro-Wilk normality test (the function in R language is *Shapiro.test*(*x*)) was used for testifying the normal distribution of variables. The Student *T* test (*t.test*(*x,y*)) was used to testify the difference when variables are normally distributed and the Mann-Whitney test (*Wilcox.test*(*x,y*)) for variables that are not normally distributed (cf. Stefanowitsch, 2020; Zhou and Gao, 2022). The Pearson's correlation coefficient test (*cor.test*(*x,y*)) was used to testify the correlation when variables are normally distributed, and the Spearman's correlation coefficient test (*cor.test*(*x,y,method* = "*spearman*") was adopted if one variable is not normally distributed. In addition, figures in this research were visually presented by using the Excel software.

Results

The distribution of subtypes of expressions of modality is shown in Fig. 1, and results of variation analysis between subtypes, inclusive of instances of each subtype, are tabulated in Table 4.[11]

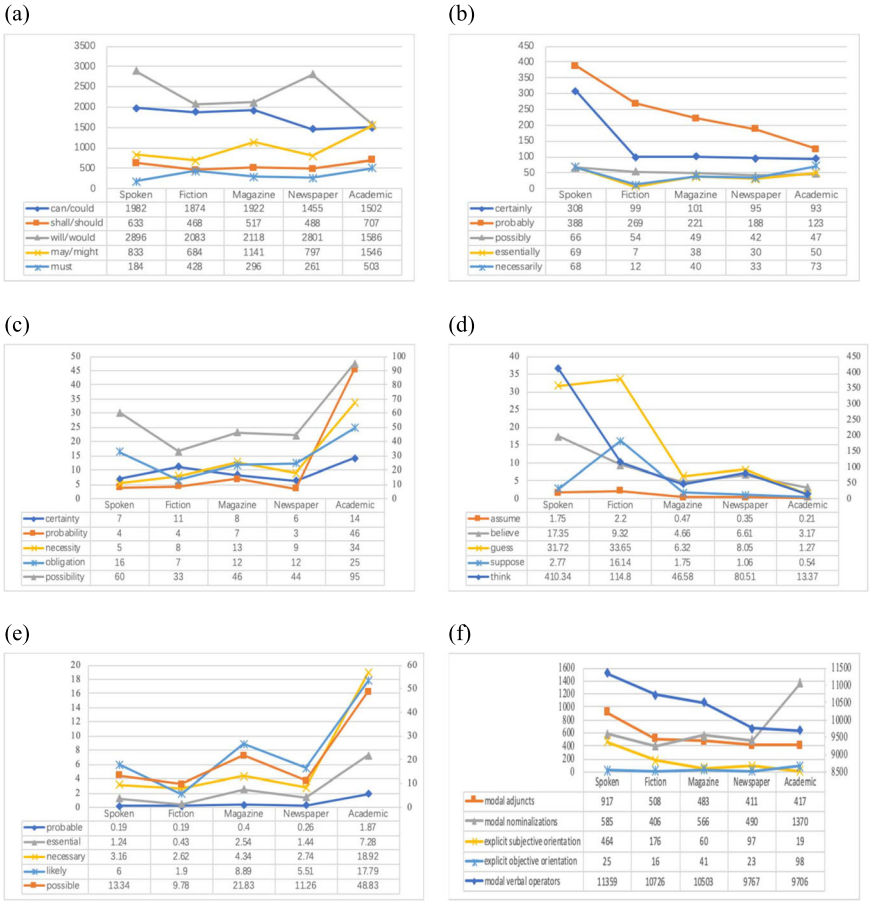

Fig. 1 Normalized frequencies of subtypes of expressions of modality across registers. a Modal verbal operators, (**b**) modal adjuncts, (**c**) modal nominalizations (*possibility* on the right-hand axis), (**d**) explicit subjective orientation (*think* on the right-hand axis) (**e**) explicit objective orientation (*possible* on the right-hand axis), (**f**) expressions of modality (modal verbal operators on the right-hand axis).

According to Fig. 1a, such modal verbal operators with high values as *must* and *shall/should* are not, comparatively, preferentially used by language users while constructing texts at issue. In addition, they are plateauing irrespective of the formality of registers. That is to say, no obvious preferences for the five types of registers could be identified although registers as such are of various degrees of formality. Those with low modal value *can/could* preferentially occur in more informal registers (e.g., normalized frequency or NF = 1982 in spoken texts and 1502 in academic texts). Those with low value *may/might* and median value *will/would* seem to be competing with each other in more formal registers, i.e., the latter occurs less frequently while the former occurs more frequently. Figure 1b shows that such modal adjuncts with high and median values as *certainly* and *probably* occur more frequently in more informal registers (e.g., in spoken texts, 308 for *certainly* and 388 for *probably*) but less frequently in more formal registers (e.g., in academic texts, 93 for *certainly* and 123 for *probably*). Modal adjuncts with median value, such as *essentially* and *necessarily* are particularly disfavored by fictional texts while preferred by academic ones. Figure 1c shows that such modal nominalizations as *possibility* and *obligation* are particularly infrequently used in fictional texts (NF = 33 PMW and 7 respectively), but more frequently in formal registers, peaking in academic texts (95 and 25 respectively). Similarly, such modal nominalizations with high value as *certainty* (7) and with median value as *probability* (4) and *necessity* (5) are most infrequently used in spoken texts, and fluctuate rather minorly in registers like fiction, magazines, and newspapers. However, the fluctuation of these modal nominalizations is greatly broadened in academic texts (14 for *certainty*; 46 for *probability*; 34 for *necessity*).

Figure 1d shows that such 'explicit subjective' orientation as *I/we think* is the most frequently used expression in registers of COCA, particularly in the most informal spoken register (410.34). The second frequent instance of this subtype is *I/we guess*, with *I/we believe*, *I/we suppose*, and *I/we assume* following. It also shows that all instances of 'explicit subjective' orientation are rather preferred in such informal registers as spoken and fictional texts, and generally dispreferred in such formal registers as magazines, newspapers, and academic texts. Figure 1e demonstrates that expressions of 'explicit objective' orientation are less frequently used in fictional (0.19 PMW for *probable*, 9.78 for *possible*, 0.43 for *essential*, 2.62 for *necessary*, and 1.9 for *likely*) and spoken (0.19 for *probable*, 13.34 for *possible*, 1.24 for *essential*, 3.16 for *necessary*, and 6 for *likely*) texts than the other types of registers. That is, they prefer formal texts to informal ones.

Figure 1f demonstrates that modal verbal operators are diminishing in usage drastically from spoken texts (11359 PMW) to academic texts (9706). This diminishing is by and large contributed by the less frequent use of such modal operators with low or median value as *can/could* (1982 in spoken texts and 1502 in academic texts) and *will/would* (2896 in spoken texts and 1586 in academic texts) (cf. Figure 1a) in more formal registers. Modal adjuncts occur frequently in the most informal register like spoken texts (917), but their occurrences do not make any difference in more formal registers such as newspapers (411) and academic texts (417). The register-sensitive nature (He and

Table 4 Results of variation analysis of expressions of modality.[12].

	shall/should	will/would	may/might	must
can/could	$t = 9.8415$ $p = 0.0001^{**}$	$t = -2.0465$ $p = 0.0901$	$t = 3.8977$ $p = 0.0055^{**}$	$t = 11.275$ $p = 2.875e\text{-}05^{**}$
shall/should		$t = -6.9687$ $p = 0.0017^{**}$	$t = -2.6898$ $p = 0.0463^{*}$	$t = 3.0901$ $p = 0.0158^{*}$
will/would			$t = 4.4695$ $p = 0.0031^{**}$	$t = 7.8096$ $p = 0.0009^{**}$
may/might				$t = 4.0018$ $p = 0.01^{**}$

	probably	possibly	essentially	necessarily
certainly	$W = 4$ $p = 0.0952$	$W = 25$ $p = 0.0079^{**}$	$W = 25$ $p = 0.0079^{**}$	$W = 25$ $p = 0.0079^{**}$
probably		$t = 4.1736$ $p = 0.0135^{*}$	$t = 4.3633$ $p = 0.0096^{**}$	$t = 4.2006$ $p = 0.0106^{*}$
possibly			$t = 1.1541$ $p = 0.2985$	$t = 0.5311$ $p = 0.6180$
essentially				$t = -0.4176$ $p = 0.6873$

(a) modal verbal operators

	probability	possibility	necessity	obligation
certainty	$W = 18.5$ $p = 0.2477$	$t = -4.2787$ $p = 0.0119^{*}$	$W = 10.5$ $p = 0.7533$	$t = -1.5538$ $p = 0.173$
probability		$W = 2.5$ $p = 0.0459^{*}$	$W = 6$ $p = 0.2087$	$W = 5.5$ $p = 0.1706$
possibility			$W = 24$ $p = 0.0159^{*}$	$t = 3.6921$ $p = 0.0162^{*}$
necessity				$W = 10$ $p = 0.6752$

(b) modal adjuncts

	assume	believe	guess	suppose
think	$W = 25$ $p = 0.0079^{**}$	$W = 24$ $p = 0.0159^{*}$	$W = 23$ $p = 0.0318^{*}$	$W = 24$ $p = 0.0159^{*}$
assume		$t = -2.8489$ $p = 0.0437^{*}$	$t = -2.2232$ $p = 0.0898$	$W = 5.5$ $p = 0.1732$
believe			$t = -1.0974$ $p = 0.322$	$W = 21$ $p = 0.0952$
guess				$W = 20$ $p = 0.1508$

(c) modal nominalizations

	possible	essential	necessary	likely
probable	$W = 0$ $p = 0.0119^{*}$	$W = 3$ $p = 0.0593$	$W = 0$ $p = 0.0119^{*}$	$W = 0$ $p = 0.0119^{*}$
possible		$W = 25$ $p = 0.0079^{**}$	$W = 22$ $p = 0.0556$	$W = 22$ $p = 0.0556$
essential			$W = 4$ $p = 0.0952$	$t = -1.8424$ $p = 0.1186$
necessary				$W = 9$ $p = 0.5476$

(d) explicit subjective orientation

	adjuncts	nominalizations	explicit subjective	explicit objective
verbal operators	$W = 25$ $p = 0.0079^{**}$	$t = 27.368$ $p = 8.539e\text{-}08^{**}$	$t = 32.055$ $p = 1.643e\text{-}06^{**}$	$t = 33.452$ $p = 4.553e\text{-}06^{**}$
adjuncts		$W = 9$ $p = 0.5476$	$W = 2$ $p = 0.0318^{*}$	$t = 5.3053$ $p = 0.0053^{**}$
nominalizations			$W = 1$ $p = 0.0159^{*}$	$t = 3.6696$ $p = 0.0209^{*}$
explicit subjective				$t = 1.5153$ $p = 0.1997$

(e) explicit objective orientation

(f) expressions of modality

Yang, 2018) is mainly characterized by the frequent use of modal adjuncts with high or median value like *certainly* and *probably* in more informal registers and their infrequent uses in more formal registers (cf. Figure 1b). Modal nominalizations are nearly equally preferred in both informal and formal registers (e.g., 585 in spoken texts and 490 in newspapers) except for the most former register – academic texts (1370). This particular preference in academic texts is mainly caused by such modal nominalizations with median or low value as *probability* and *possibility* (cf. Figure 1c). Pertaining to the two subtypes of interpersonal metaphor of modality, instances of 'explicit subjective' orientation are frequently used in more informal registers but infrequently used in more formal registers, which is by and large contributed by the use of *I/we think* with low value from spoken texts (410.34) to academic texts (13.37) (cf. Figure 1d). Expressions of 'explicit objective' orientation generally keep a minor fluctuation from spoken texts (25) to newspapers (23), but this minor fluctuation is broken up and these expressions are extremely preferred in academic texts (98). This fluctuation is primarily caused by the use of the expression *it is/was possible* with low value from spoken texts (13.34) to academic texts (48.83) (cf. Figure 1e).

Table 4a shows that occurrences of modal verbal operators are either significantly or extremely significantly different from each other except for the one between *can/could* and *will/would* ($t = -2.0465$, $p = 0.0901 > 0.05$), indicating the dominance of *can/could* and *will/would* over the others, with *may/might*, *shall/should*, and *must* following. Table 4b shows that *certainly* ($W = 25$, $p = 0.0079 < 0.01$ with *possibly*, *essentially*, and *necessarily* respectively) and *probably* ($t = 4.1739$, $p = 0.0135 < 0.01$ with *possibly*; $t = 4.3633$, $p = 0.0096 < 0.01$ with *essentially*; $t = 4.2006$, $p = 0.0106 < 0.01$ with *necessarily*) are significantly different from the other modal adjuncts, and this difference is generally contributed by their preferential occurrences in spoken register (cf. Figure 1b). Table 4c shows that occurrences of modal nominalizations *certainty, probability, necessity*, and *obligation* demonstrate no significant difference between each other, indicating no perspicuous preference in these registers of COCA. Concerning the low value modal nominalization *possibility*, it is statistically significantly different from the other four modal nominalizations ($t = -4.2787$, $p = 0.0119 < 0.05$ with *certainty*; $W = 2.5$, $p = 0.0459 < 0.05$ with *probability*; $W = 24$, $p = 0.0159 < 0.05$ with *necessity*; $t = 3.6921$, $p = 0.0162$ with *obligation*), which suggests that language users frequently employ the low modal item in its nominalized form, i.e., *possibility*, to objectify the modal meaning within the proposition.

In relation to the variation between 'explicit subjective' orientation in interpersonal metaphor of modality, Table 4d shows that *I/we think* is significantly different from *I/we assume* ($W = 25$, $p = 0.0079 < 0.01$), *I/we believe* ($W = 24$, $p = 0.0159 < 0.05$), *I/we guess* ($W = 23$, $p = 0.0318 < 0.05$), and *I/we suppose* ($W = 24$, $p = 0.0159 < 0.05$); in addition, *I/we believe* and *I/we assume* ($t = -2.8489$, $p = 0.0437 < 0.05$) are also significantly different from each other. In relation to this variation between 'explicit objective' orientation, it is shown in Table 4e that *probable* and *possible*, *possible* and *essential*, *probable* and *likely*, and *probable* and *necessary* are significantly different from each other, respectively. These differences indicate that when "possibility" in modalization needs to be expressed, 'explicit objective' orientation that denotes low value such as *possible* and *likely* is extremely likely to be employed; when "necessity" in modulation needs to be expressed, the one with median value like *necessary* is likely to be used.

Regarding the overall occurrences of subtypes of expressions of modality, Table 4f shows that modal verbal operators are extremely significantly different from modal adjuncts ($W = 25$, $p = 0.0079 < 0.05$), modal nominalizations ($t = 27.368$, $p = 8.539\text{e-}08 < 0.01$), 'explicit subjective' orientation ($t = 32.055$, $p = 1.643\text{e-}06 < 0.01$), and 'explicit objective' orientation ($t = 33.452$, $p = 4.553\text{e-}06 < 0.01$) respectively, indicating that modal verbal operators are still the most frequently used expressions of modality in registers of COCA, particularly in more informal registers. In addition, modal adjuncts are significantly different from 'explicit subjective' orientation ($W = 2$, $p = 0.0318 < 0.05$) and 'explicit objective' orientation ($t = 5.3053$, $p = 0.0053 < 0.01$), and the significant difference also exists between modal nominalizations and 'explicit subjective' orientation ($W = 1$, $p = 0.0159 < 0.05$) or 'explicit objective' orientation ($t = 3.6696$, $p = 0.0209 < 0.05$)

Results of correlation analysis between subtypes of expressions of modality, inclusive of instances of each subtype, are tabulated in Table 5.

In relation to the relationship between subtypes of expressions of modality at word rank and the formality of registers, Table 5a shows that only the distribution of *will/would* is significantly negatively correlated with that of *must* ($r = -0.9211$, $p = 0.0262 < 0.05$), which suggests that *will/would* occur less frequently in more formal registers while *must* in more informal registers. More importantly, distribution of *will/would* is negatively but not significantly correlated with that of *may/might* ($r = -0.7182$, $p = 0.1717$), which indicates that the compensation of *may/might* for *will/would* does not reach the significant level in more formal registers (cf. Figure 1a). In Table 5b, such modal adjuncts as *probably* and *possibly* are positively significantly correlated ($r = 0.9065$, $p = 0.0338 < 0.05$), indicating that the two modal adjuncts occur more frequently in more informal registers. The positive significant correlation between *essentially* and *necessarily* ($r = 0.9268$, $p = 0.0235 < 0.05$) suggests that both modal adjuncts are preferred in more formal registers (cf. Figure 1b). In Table 5c, these modal nominalizations do not reach a significant level (precluding the correlation between *possibility* and *obligation*) although they are positively correlated. It is also shown that *possibility* and *obligation* are positively extremely significantly correlated with each other ($r = 0.9937$, $p = 0.0006 < 0.01$), indicating that they preferentially occur in more formal registers (cf. Figure 1c).

With respect to the relationship between interpersonal metaphor of modality and the formality of registers, Table 5d shows that *I/we think* and *I/we believe* ($\rho = 1$, $p = 0.0167 < 0.05$), *I/we assume* and *I/we guess* ($t = 0.9855$, $p = 0.0021 < 0.01$), and *I/we assume* and *I/we suppose* ($\rho = 1$, $p = 0.0167 < 0.05$) are positively significantly correlated with each other respectively, indicating their similar distribution trend. That is, the more formal a certain register is, the less frequently 'explicit subjective' orientation will occur (cf. Figure 1d). In Table 5e, it is shown that *probable* and *essential* ($\rho = 9747$, $p = 0.0048 < 0.01$), *possible* and *necessary* ($\rho = 1$, $p = 0.0167 < 0.05$), *possible* and *likely* ($\rho = 1$, $p = 0.0167 < 0.05$), *essential* and *likely* ($t = 0.9860$, $p = 0.0020 < 0.01$), and *necessary* and *likely* ($\rho = 1$, $p = 0.0167 < 0.05$) are significantly correlated with each other respectively, indicating that they share a similar distribution trend that expressions of 'explicit objective' orientation as such are preferentially used in more formal registers (cf. Figure 1e).

It can be seen in Table 5f that three types of correlation between expressions of modality could be considered. First, modal verbal operators and modal adjuncts are positively significantly correlated with one another ($r = 0.8761$, $p = 0.05$), and this positive significant correlation also exists between modal adjuncts and 'explicit subjective' orientation ($r = 0.9677$, $p = 0.0069 < 0.01$); additionally, modal verbal operators and 'explicit subjective' orientation, to some extent, are correlated with each other ($r = 0.8608$, $p = 0.061$) in that this significance is rather close to the threshold of '$a = 0.05$'. The positive significance of correlation between them suggests that the three types of expressions of modality occur frequently in more informal registers but infrequently in more formal registers

Table 5 Results of correlation analysis of expressions of modality.

	shall/should	will/would	may/might	must
can/could	$r = -0.1665$ $p = 0.789$	$r = 0.201$ $p = 0.7459$	$r = -0.3723$ $p = 0.5371$	$r = -0.3898$ $p = 0.5165$
shall/should		$r = -0.3094$ $p = 0.6124$	$r = 0.7356$ $p = 0.1566$	$r = 0.2281$ $p = 0.7121$
will/would			$r = -0.7182$ $p = 0.1717$	$r = -0.9211$ $p = 0.0262^{*}$
may/might				$r = 0.5388$ $p = 0.3488$

	probably	possibly	essentially	necessarily
certainly	$\rho = 0.9$ $p = 0.0833$	$\rho = 0.8$ $p = 0.1333$	$\rho = 0.3$ $p = 0.6833$	$\rho = -0.1$ $p = 0.95$
probably		$r = 0.9065$ $p = 0.0338^{*}$	$r = 0.2957$ $p = 0.629$	$r = -0.0107$ $p = 0.9864$
possibly			$r = 0.4676$ $p = 0.4271$	$r = 0.2716$ $p = 0.6585$
essentially				$r = 0.9268$ $p = 0.0235^{*}$

(a) modal verbal operators

	probability	possibility	necessity	obligation
certainty	$\rho = 0.8208$ $p = 0.0886$	$r = 0.5961$ $p = 0.2887$	$\rho = 0.5$ $p = 0.45$	$t = 0.5064$ $p = 0.3839$
probability		$\rho = 0.6669$ $p = 0.2189$	$\rho = 0.6669$ $p = 0.2189$	$\rho = 0.5263$ $p = 0.3622$
possibility			$\rho = 0.4$ $p = 0.5167$	$r = 0.9937$ $p = 0.0006^{**}$
necessity				$\rho = 0.3501$ $p = 0.5528$

(b) modal adjuncts

	assume	believe	guess	suppose
think	$\rho = 0.8$ $p = 0.1333$	$\rho = 1$ $p = 0.0167^{*}$	$\rho = 0.9$ $p = 0.0833$	$\rho = 0.8$ $p = 0.1333$
assume		$t = 0.7387$ $p = 0.1539$	$t = 0.9855$ $p = 0.0021^{**}$	$\rho = 1$ $p = 0.0167^{*}$
believe			$t = 0.8325$ $p = 0.0802$	$\rho = 0.8$ $p = 0.1333$
guess				$\rho = 0.9$ $p = 0.0833$

(c) modal nominalizations

	possible	essential	necessary	likely
probable	$\rho = 0.8208$ $p = 0.0886$	$\rho = 9747$ $p = 0.0048^{**}$	$\rho = 0.8208$ $p = 0.0886$	$\rho = 0.8208$ $p = 0.0886$
possible		$\rho = 0.9$ $p = 0.0833$	$\rho = 1$ $p = 0.0167^{*}$	$\rho = 1$ $p = 0.0167^{*}$
essential			$\rho = 0.9$ $p = 0.0833$	$t = 0.9860$ $p = 0.0020^{**}$
necessary				$\rho = 1$ $p = 0.0167^{*}$

(d) explicit subjective orientation

	adjuncts	nominalizations	explicit subjective	explicit objective
verbal operators	$r = 0.8761$ $p = 0.05^{*}$	$\rho = -0.3$ $p = 0.6833$	$r = 0.8608$ $p = 0.061$	$r = -0.5587$ $p = 0.3276$
adjuncts		$r = -0.2461$ $p = 0.6898$	$r = 0.9677$ $p = 0.0069^{**}$	$\rho = -0.2$ $p = 0.7833$
nominalizations			$\rho = -0.4$ $p = 0.5167$	$r = 0.9816$ $p = 0.003^{**}$
explicit subjective				$r = -0.5124$ $p = 0.3744$

(e) explicit objective orientation

(f) expressions of modality

(cf. Figure 1f). Second, modal nominalizations are positively significantly correlated with 'explicit objective' orientation ($r = 0.9816$, $p = 0.003 < 0.01$), indicating that expressions of modality as such are preferentially used in more formal registers, particularly in academic texts. Third, modal nominalizations are to some extent negatively correlated with modal verbal operators ($\rho = -0.3$, $p = 0.6833$) and modal adjuncts ($r = -0.2461$, $p = 0.6898$), although they do not reach the significant level, which indicates that the loss of modal verbal operators and modal adjuncts is compensated for by the gaining of modal nominalizations and this compensation happens obviously in academic texts (1370 PMW). In addition, there is a negative insignificant correlation between 'explicit objective' orientation and 'explicit subjective' orientation in terms of interpersonal metaphor of modality ($r = -0.5124$, $p = 0.3744$), suggesting that the gaining of 'explicit objective' orientation is also compensating for the loss of 'explicit subjective' orientation.

Discussion

This section discusses the distribution of expressions of modality across registers in relation to the first research question and the relationship between distribution of expressions of modality and formality of registers in relation to the second research question.

Distribution of expressions of modality across registers. Concerning expressions of modality in each subtype, those with low or median values are preferentially used across registers of COCA by language users to express their subjective judgment of propositions at issue. More specifically, this inclination is instantiated by *can/could*, *will/would*, and *may/might* in modal verbal operators, *probably* and *possibly* in modal adjuncts, *possibility* in modal nominalizations, *I/we think* in 'explicit subjective' orientation, and *possible* in 'explicit objective' orientation (cf. Figure 1a-e and Table 4). Findings of this research are in accordance with earlier studies on the distribution of modal verbal operators across registers (Collins, 2009; Verhulst, et al. 2013; Zhou, 2023b). They also echo the findings documented by Zhou (2023a, 2023b), who examined the distribution of interpersonal metaphor of modality across registers in corpora of COHA, COCA, and BNC. What distinguishes this study from earlier ones is that the current one expanded the investigation of expressions of modality into other domains, such as modal nominalizations and individual expression of modality in each subtype. Reasons for this preferential employment of expressions of modality with low or median value, particularly the former, are basically two-fold. First, language users actively engage with a multiplicity of voices, embracing the fluid and evolving nature of propositions. This approach not only acknowledges the legitimacy of diverse perspectives but also fosters a collaborative spirit of inquiry and negotiation, leading to a richer and more nuanced comprehension of the issues at hand (for the entertainment of varying voices and the arguability and negotiability of propositions, cf. Martin and White, 2005; Yang, 2019). Consider the example in (8a), in which the modal adjunct *probably* denotes a median modal value (Halliday and Matthiessen, 2014). The modification of the proposition *she has no idea what's she's saying* by *probably* with median value of modality is licensed by the preceding proposition *she's half-drunk*, because a drunk person's words are usually uttered unintentionally. In doing so, the use of *probably* leaves sufficient room for hearers to argue the validity of the proposition *she has no idea what's she's saying*. Second, the judicious use of modal expressions serves to mitigate the impact of one's words, ensuring that communication remains respectful and considerate. By opting for language that carries a lower or median degree of assertiveness, speakers navigate the delicate balance between conveying their messages and preserving the dignity and self-esteem of their audience, thereby fostering an environment conducive to open and harmonious exchange (for the warranty of the hearers' or readers' faces by expressions of modality, cf. Zhou, 2023a, 2023c). Consider the example in (8b), wherein *I think* is employed to attenuate the speaker's judgment of the proposition *the Republicans need to start talking about the real issues before the country* so as to accept other varying voices (e.g., *the Republicans do not necessarily need to start talking about the real issues before the country*) and consequently others' faces are protected.

(8) [redacted]

a. She doesn't mean anything by it. She's half-drunk and *probably* has no idea what she's saying. (COCA_academic_2009)
b. *I think* the Republicans need to start talking about the real issues before the country. (COCA_spoken_2004)

Expressions of modality at word rank, i.e., modal verbal operators, modal adjuncts, and modal nominalizations, are used more frequently than expressions of interpersonal metaphor of modality, i.e., 'explicit subjective' orientation and 'explicit objective' orientation, or that modality at clause rank (cf. Figure 1f). This finding is consistent with what was documented in He's (2020) or Zhou's (2023b) research. Both studies uncovered that congruent expressions of modality (modal verbal operators and modal adjuncts in this research) occur more frequently than their metaphorical counterparts across registers. Possible reasons are also two-fold. For one thing, the employed corpus, i.e., COCA, consists of various registers that are constructed for general purposes (e.g., fiction, magazines); that is, their potential readers are the general public. The only register that requires the readers to be trained to understand is academic texts. Academic register generally features grammatical metaphor which incorporates 'explicit objective' orientation (Halliday and Matthiessen, 1999). Metaphorical expressions of modality are more difficult to understand than their congruent counterparts (Halliday, 1994; Halliday and Matthiessen, 1999, 2014). Against this backdrop, registers as such (except for academics) usually employ expressions of modality that are easy for readers to understand, such as modal verbal operators and modal adjuncts. For another, implicitness is typically characteristic of these considered registers of COCA. In SFL, modal verbal operators and modal adjuncts denote their modal meanings implicitly (Zhou, 2023b). Regarding expressions of 'explicit subjective' orientation, their modal meanings are directly projected by the grammatical pattern *I/we think*. Owing to its spoken feature, this grammatical pattern is not particularly favored in other written registers of COCA.

Relationship between distribution of expressions of modality and formality of registers. Regarding certain expressions of modality, those with low or median values are used in more formal registers while those with high value are used in more informal registers. This is in accordance with the finding documented by Zhou (2023a, 2023b). As expounded in the foregoing subsection, those expressions of modality are employed by language users to entertain different voices so as to make the propositions negotiable and arguable (Martin and White, 2005; Yang, 2019; Zhou, 2023a). However, the preferential employment of expressions of modality with high value in more informal registers such as spoken texts is motivated by "the strengthening of propositions when sufficient evidence is provided" (Zhou, 2023c: 175). Consider the example in (9). The expression of modality with high value *we believe* is used in the sense that the validity of the proposition *the majority of doctors will actually get paid either the same or a little bit more* is backed up by the preceding proposition *because we would be covering everybody and we wouldn't*

allow duplicative care, which is how we operate with Medicare right now which functions as the needed evidence.

(9) Because we would be covering everybody and we wouldn't allow duplicative care, which is how we operate with Medicare right now, *we believe* that the majority of doctors will actually get paid either the same or a little bit more. (COCA_spoken_2019)

With respect to the relationship between the general distribution of expressions of modality and formality of registers, two respects are to be discussed. First, modal nominalizations preferentially occur in more formal registers such academic texts, while modal verbal operators and modal adjuncts preferentially occur in more informal registers such as spoken texts if expressions of modality at word rank are concerned. This could be explained by the fact that modal meanings in spoken texts are articulated in a subjective way, while those in academic texts should be realized objectively so as to highlight the credibility and persuasiveness of academic texts (Tian, 2017) and ultimately make the academic texts authoritative (Schleppegrell, 2004). Specifically, modal verbal operators and modal adjuncts realize modal meanings directly in SFL (cf. Halliday and Matthiessen, 2014), whereas those modal meanings are presupposed by the employment of modal nominalizations. In other words, such expressions of modality as modal verbal operators and modal adjuncts are important components of Finite element in Mood system (which consists of Mood and Residue, the former is composed of the subject and the Finite element in a clause and the latter refers to the rest of that clause), realizing interpersonal meaning (i.e., enacting the social relationship between interlocutors). The Finite element is expressed by modal verbal operators which are capable of making propositions arguable or negotiable between interlocutors and, hence, conveying the interpersonal meaning subjectively (Halliday and Matthiessen, 2014). Consider the example in (10a), wherein the subject realized by *this* and the Finite element realized by the modal verbal operator *might* form the Mood of the clause. That is to say, the interpersonal meaning is directly realized by the modal verbal operator, and thus, the speaker's subjective meaning is expressly articulated. On the contrary, a modal nominalization realizing the modal meaning is an element of the proposition. In this sense, the modal meaning is presupposed and the interpersonal meaning is therefore "disguised" as the ideational meaning, which construes the experience in social reality. By doing so, the way to express the modal meaning is made indirectly and thus conveyed objectively. Consider the example in (10b), wherein the writer's subjective meaning is not expressed by a Finite element, but by the modal nominalization *possibility*, which has been turned into an element of the proposition. Thus, the objectivity of expressing the modal meaning is foregrounded.

(10) [redacted]

a. This *might* be my favorite story of the day. (COCA_spoken_2019)
b. Nobody in Los Alamos raised this *possibility*. (COCA_academic_1992)

(11) Although it is extremely difficult to establish legitimacy for an object, *it is often possible* to point out what's wrong with it. (COCA_academic_2003)

Second, expressions of 'explicit objective' orientation are preferred in more formal registers such as academic texts while those of 'explicit subjective' orientation are frequently used in more informal registers such as fictional texts if interpersonal metaphor of modality is concerned. This is corroborated by the argument that grammatical metaphor (i.e., modal nominalizations and 'explicit objective' orientation are categorized into ideational metaphor and interpersonal metaphor of modality, respectively, in Hallidayan sense) is particularly preferred in academic texts (cf. Liardét, 2018; He, 2020; Zhou, 2023a, 2023b). Although the modal meaning, by employing 'explicit subjective' orientation in terms of interpersonal metaphor of modality, is expressed outside the proposition in a projecting clause, the explicit use of the first personal pronouns (i.e., *I* or *we*) foregrounds the commentator who has made the judgment of the proposition. Therefore, this type of expressions of modality denotes a strong sense of subjectivity. However, by means of 'explicit objective' orientation, the modal meaning originally represented by a Finite element is now expressed by another proposition, in which the modal adjective functions as the predicate of the proposition. By doing so, the means of representing the modal meaning is concealed so that the objectivity of the proposition is further highlighted. This is considerably in accordance with the fact that academic texts are constructed in an objective way so as to convey the text information persuasively (Tian, 2017) and authoritatively (Schleppegrell, 2004). Consider the example in (11). The modal meaning is not expressed by a Finite element (e.g., *possibly* or *might*), but by the 'explicit objective' orientation *it is often possible* in which *possible* functions as the predicate of the proposition; hence the means of denoting the modal meaning is concealed and the objectivity of the proposition is highlighted.

Conclusion

This research investigated the register-based distribution of expressions of modality in English. The results show that expressions of modality with low value or occasionally with median value are significantly employed by language users to construct texts in registers of COCA so as to entertain varying voices. Specifically, this significant employment is instantiated by *can/could*, *will/would*, and *may/might* in modal verbal operators, *probably* and *possibly* in modal adjuncts, *possibility* in modal nominalizations, *I/we think* in 'explicit subjective' orientation, and *possible* in 'explicit objective' orientation. These results also show that the gaining of modal nominalizations compensates for the loss of modal verbal operators and modal adjuncts in terms of expressions of modality at word rank, and the gaining of 'explicit objective' orientation compensates for the loss of 'explicit subjective' orientation in terms of interpersonal metaphor of modality. The purpose is to conceal the speakers' or writers' subjective meaning, and the objectivity of the proposition at issue is therefore foregrounded.

This study contributes to existing literature on modality in at least two ways. For one thing, a fuller scope of expressions of modality in English is documented. Instead of solely examining modal verbal operators, modal adjuncts, or modal projecting clauses, this research considered, to a large extent, the distribution of modality across registers in relation to expressions of modality both at word ranks such as modal verbal operators and modal nominalizations and at clause rank or interpersonal metaphor of modality such as 'explicit subjective' orientation and 'explicit objective' orientation. For another, the distribution of modality across registers is theoretically associated with the interpersonal metaphor of modality. This research considered not only the preferential occurrence of expressions of modality across registers, but also the reasons underlying this preference from the

perspective of interpersonal metaphor of modality. Specifically, language users employed different kinds of expressions of modality for the purposes of either entertaining other varying voices or concealing the commentators' subjective meaning to highlight the objectivity of propositions in former registers. This study also contributes to a fuller consideration of the influence of the register variable on expressions of modality in relation to register studies. Inclusive of such registers as spoken texts, fiction, magazines, and newspapers that are investigated in former studies, the current research also considered the academic register regarding the distribution of expressions of modality.

One limitation of the paper is that only the synchronic distribution of expressions of modality in registers of COCA is considered, leaving their diachronic distributions untouched. Another limitation is that COCA is typically characteristic of American English, and thus, it should be to some extent cautious to regard the findings of this research as the common properties of English in general. In addition, as one of the reviewers commented, this research only considered the positive forms of expressions of modality (e.g., *may* or *it is necessary*), excluding the negative formulations of modality. This treatment will arguably make the findings of the study partial. Underlying the three limitations, future studies are therefore suggested to consider the diachronic distribution of expressions of modality (inclusive of both positive and negative formulations of modality) in these registers and in other varieties of English, such as British English and Australian English.

Data availability

All data analyzed in this study were cited in this article and available in the public domain.

Received: 17 December 2023; Accepted: 29 April 2024;
Published online: 21 May 2024

Note

1 The use of key terms in this research is in accordance with the SFL tradition. For example, instead of employing modal auxiliaries and modal adverbs in Nuyts' (1993) or Collins' (2009) terms, modal verbal operators and modal adjuncts are used, respectively.
2 Examples in this study are sourced from the Corpus of Contemporary American English or COCA (Davies, 2010).
3 According to SFL, language generally consists of such ranks as word, group/phrase, clause, clause complex.
4 A register is "a functional variety of language (Halliday, McIntosh & Strevens, 1964; Halliday, 1978) – the patterns of instantiation of the overall system associated with a given type of context (a situation type)" (Halliday & Matthiessen, 2014).
5 The latter modal operators are not considered in this study in that they, though increasingly common, are to some extent amorphous. For instance, different morphological forms of the semi-modal operator *have to* such as *has to*, *have to*, *having to*, and *had to* are frequently used in various registers.
6 Values of these expressions are determined by referring to Halliday and Matthiessen (2014), Van linden (2012), and Zhou (2023a, 2023b, 2023c).
7 Data were collected in June 2021 from https://www.english-corpora.org/coca/.
8 Genre in Davies' (2010) sense corresponds to register in this study.
9 According to Davies (2010) or the webpage of COCA, general web and blog texts were collected in October 2012, and they are more of a "snapshot" of these genres/registers, while the other registers in COCA were collected year by year from 1990 to 2019. For the purpose of avoiding the influence of the "year" variable over the findings, the two registers were not included in my analysis. The precluding of the TV/movie register is because it is similar to the spoken register (data in both registers were collected from TV or radio programs and both are thus to some extent equally informal) in terms of formality of registers.
10 The formula that Heylighen and Dewaele (1999) proposed to compute the formality of registers is "F= (noun frequency + adjective frequency + preposition frequency + article frequency – pronoun frequency – verb frequency – adverb – interjection frequency + 100)/2", which could be exemplified by the academic register, i.e., "F = (26.41 + 9.64 + 12.66 + 8.50-2.25-14.54-4.50-5.99 + 100)/2 = 64.97".
11 Only specific expressions of modality were compared in that others were attested rather scantly (e.g., 0.12 per million words or PMW for *obligatorily* in all registers), which demonstrates that they do not occur frequently in registers under consideration.
12 In tables of this paper, * stands for significance at the 0.05 level and ** stands for significance at the 0.01 level

References

Biber D (1988) Variation across speech and writing. Cambridge University Press, Cambridge
Biber D, Conrad S (2009) Register, genre, and style. Cambridge University Press, Cambridge
Biber D, Gray B (2016) Grammatical complexity in academic English: linguistic change in writing. Cambridge University Press, Cambridge
Biber D, Johansson S, Leech G, Conrad S, Finegan E (1999) The Longman Grammar of Spoken and Written English. Longman, London
Charnock, R. 2009. When *may* means *must*: deontic modality in English statute construction. In, P Busuttil, J van der Auwera, & R Salkie (Eds.), Modality in English: theory and description (pp. 177–198). Berlin/New York: Mouton de Gruyter
Collins P (2009) Modals and quasi-modals in world Englishes. World Englishes 28(3):281–292
Davies M (2010) The corpus of contemporary American English as the first reliable monitor corpus of English. Lit Linguist Comput 25(4):447–464
de Saussure F (1959) Course in general linguistics. Wade Baskin (trans.). New York: Philosophical Library
Fontaine L (2013) Analyzing English grammar: a systemic functional introduction. Cambridge University Press, Cambridge
Halliday MAK (1978) Language as social semiotic: the social interpretation of language and meaning. Edward Arnold, London
Halliday MAK, McIntosh A, Strevens P (1964) The linguistic sciences and language teaching. Longman, London
Halliday MAK (1994) An introduction to functional grammar, 2nd edn. London: Edward Arnold
Halliday MAK, Christian MIMM (1999) Construing experience through meaning: A language-based approach to cognition. London & New York: Cassell
Halliday MAK, CMIM Matthiessen (2014) Halliday's introduction to functional grammar (4th edition). London: Routledge
He Q (2020) A corpus-based study of interpersonal metaphors of modality in English. Stud Neophilol 92:1–22
He Q, Yang B (2018) A corpus-based study of the correlation between text technicality and ideational metaphor in English. Lingua 203:51–65
He, Q. 2019. A corpus-based approach to clause combining in English from the systemic functional perspective. Springer
Heylighen, F J-M Dewaele. 1999. Formality of language: definition, measurement and behavioral determinants. Internal Report. Center "Leo Apostel", Free University of Bruseels
Kaltenböck G (2005) *It*-extraposition in English: a functional view. Int J Corpus Linguist 10(2):119–159
Leech G (2011) The modals ARE declining: reply to Neil Millar's "Modal verbs in TIME: frequency changes 1923-2006". Int J Corpus Linguist 16(4):547–564
Leech G, Hundt M, Mair C, Smith N (2009) Change in contemporary English: A grammatical study. Cambridge University Press, Cambridge
Liardét CL (2018) "As we all know": examining Chinese EFL learners' use of interpersonal grammatical metaphor in academic writing. Engl Specif Purp 50:64–80
Martin JR, White PRR (2005) The language of evaluation: Appraisal in English. Palgrave Macmillan, London & New York
Matthews, R. 2003. Modal auxiliary constructions, TAM and interrogatives. In R Facchinetti, M Krug & F Palmer (Eds.), Modality in contemporary English (pp. 47–70). Berlin/New York: Mouton de Gruyter
Nuyts J (1993) Epistemic modal adverbs and adjectives and the layered representation of conceptual and linguistic structure. Linguistics 31:933–969
Palmer FR (2001) Mood and modality. Cambridge University Press., Cambridge
Schleppegrell, MJ (2004) Technical writing in a second language: The role of grammatical metaphor. In: LJ Ravelli & RA Ellis (Eds.), Analyzing academic writing: contextualized framework (pp. 172–189). London: Continuum
Stefanowitsch A (2020) Corpus linguistics: a guide to the methodology. Language Science Press, Berlin
Suzuki D (2018) Variation between modal adverbs in British English: the case of *maybe* and *perhaps*. Funct Lang 25(3):392–412
Suzuki D, Fujiwara T (2017) The multifunctionality of 'possible' modal adverbs: a comparative look. Language 93(4):827–841
Taverniers, M. (2003) Grammatical metaphor in SFL: a historiography of the introduction and initial study of the term. In A-M Simon-Vandenbergen, M Taverniers & LJ Ravelli (Eds.), Grammatical metaphor: Vviews from systemic functional linguistics (pp. 5–33). Amsterdam: John Benjamins

Tian Y (2017) On the semantic function of objectivity of grammatical metaphor. Mod Foreign Lang 40(2):179–188
Van linden A (2012) Modal adjectives: English deontic and evaluative constructions in synchrony and diachrony. Walter de Gruyter, Berlin & Boston
Verhulst A, Depraetere I, Heyvaert L (2013) Source and strength of modality: An empirical study of root should, ought to and be supposed to in Present-day British English. J Pragmat 55:210–225
Yang B (2019) Interpersonal metaphor revisited: Identification, categorization, and syndrome. Soc Semiot 29(2):186–203
Zhou J (2023a) A corpus-based study of explicit objective modal expressions in English. Stud Neophilol 95(1):100–126
Zhou J (2023b) A corpus-based study of congruent and metaphorical patterns of modality in English. Stud Neophilol 95(3):351–375
Zhou J, Gao Y (2021) A Corpus-based study of grammatical post-metaphorical expressions. J World Lang 7(2):247–282
Zhou J, Gao Y (2022) Diachronic distribution of elemental ordering in English. J Quant Linguist 29(3):350–373
Zhou J (2023c) A collexeme-based study of classes and functions of grammatical metaphor in English. PhD dissertation, Peking University

Author contributions

Jiangping Zhou: conceptualization, data collection, methodology, analysis, first draft, final draft; Yanhua Xia: conceptualization, data collection, analysis, first draft.

Competing interests

The authors declare no competing interests.

Ethical approval

Ethical approval was not required as the study did not involve human participants.

Informed consent

This article did not contain any study with human participants performed by any of the authors.

Additional information

Correspondence and requests for materials should be addressed to Jiangping Zhou.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2024