Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-024-03967-6 OPEN

Linguistic variation in mediated diplomatic communication: a full multi-dimensional analysis of interpreted language in Chinese Regular Press Conferences

Yao Yao [1,2], Dechao Li [2✉], Yingqi Huang[2] & Zhonggang Sang[1]

The integration of corpus linguistics within translation studies has revolutionised our understanding of mediated language. This study endeavours to advance this burgeoning field by employing a full multi-dimensional analysis to investigate linguistic variation in interpreted language within the specialised context of diplomatic discourse. Specifically, the research examines the co-occurring patterns of linguistic features in interpreted diplomatic language vis-à-vis its non-interpreted counterpart. Employing a multivariate statistical technique, this investigation conducted a factor analysis of 113 linguistic variables, yielding five distinct linguistic dimensions: (1) Involved vs. Informational Production, (2) Objective vs. Addressee-focused Narration, (3) Literate-Oral Continuum, (4) Information Elaboration, and (5) Narrative vs. Non-narrative Concerns. The resulting patterns demonstrate that interpreted diplomatic language tends to be more informative, objective, less elaborated, non-narrative, and aligns more closely with formal registers compared to its non-interpreted counterpart, although both navigate the literate-oral continuum. This study delineates the prevailing co-occurrence patterns in interpreted and non-interpreted diplomatic languages and seeks to elucidate the potential factors shaping these linguistic variations by situating these patterns within the context of diplomatic communication. In doing so, it contributes to a nuanced understanding of how specialised contexts influence mediated language use. The findings have significant implications for corpus-based interpreting studies, shedding light on the multi-dimensional nature of interpreted language and informing the development of targeted pedagogical approaches for diplomatic interpreter training.

[1] School of Foreign Studies, Xi'an Jiaotong University, Xi'an, Shaanxi, China. [2] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR. ✉email: dechao.li@polyu.edu.hk

Introduction

The past three decades have witnessed a surge in corpus-based translation and interpreting studies, aiming to characterise language produced in mediated contexts (Baker, 1993; Shlesinger, 1998). This burgeoning field posits that interpreted language constitutes a distinct language variety, diverging from both its source and target counterparts and exhibiting features such as simplification, explicitation, and normalisation, collectively referred to as "translation universals" (Olohan and Baker, 2000; Laviosa, 2002; Kruger and Van Rooy, 2016; Xu and Li, 2022; Liu et al. 2023). Recent scholarship, however, has critically reflected on this trajectory, highlighting several limitations and advocating for new avenues of exploration (De Sutter and Lefer, 2020; Gu, 2024; Wu et al. 2024; Calzada Pérez and Ramos, 2021). A primary critique targets the over-reliance on "translation universals" to explain differences between interpreted and non-interpreted language, potentially overlooking the influence of other factors like communicative contexts, which dynamically shape language use. Additionally, the manual selection of linguistic features in most studies raises concerns about achieving a holistic understanding of the multi-dimensional nature of interpreted language. There is also a recognised need to adopt advanced statistical methods to more effectively investigate linguistic patterns and norms in interpreted language.

Addressing these limitations entails shifting from identifying isolated linguistic features to uncovering how language operates holistically within specific communicative contexts. This approach aligns with the understanding that interpreting is inherently context-dependent, necessitating consideration of the interplay between linguistic features and context (Wadensjö, 1998). This is particularly pertinent in institutional settings like press conferences, which serve as platforms for politicians and public figures to communicate with the media and the public represented by journalists (Ekström et al. 2017; Sandrelli, 2021). Diplomatic interpreting—the translation of speeches by government officials at these diplomatic events—exemplifies the need to move beyond analysing isolated features (Yang, 2012). The interpreted language in these high-stakes encounters is carefully crafted to achieve specific political objectives and carries significant weight, directly impacting international relations, shaping public perception, and potentially influencing policy decisions. While existing research has made valuable contributions to understanding interpreted diplomatic language, a closer examination reveals a tendency to focus on general aspects of norms, agency, or surface-level linguistic features. For instance, several studies have explored interpreting norms and the mediating role of interpreters in managing cross-cultural communication (Schäffner, 2004, 2008; Bhatia, 2006; Wang, 2012; Gu and Wang, 2021). Others have examined textual features, such as the use of hedging, modal verbs, and other stylistic elements in interpreted diplomatic language (Russo et al. 2006; Li and Hu, 2013; Dayter, 2018; Fu and Wang, 2022; He, 2022). These studies often rely on manually selected linguistic features, limiting their scope to predetermined categories and potentially overlooking subtle yet significant linguistic variations. Although the importance of context is acknowledged in general terms, existing research often falls short of systematically investigating how specific contextual factors interact with linguistic choices to shape the dynamics of diplomatic communication. This gap is particularly evident in the lack of studies employing robust corpus-based methods and multi-dimensional analysis to uncover co-occurring patterns of linguistic features and their relationship to contextual variables. It is worth noting that the term "co-occurring" and its variants used in this study specifically refer to the statistical correlation of frequency distributions of linguistic features across texts, as revealed through factor analysis. This usage differs from the concept of "co-occurrence" in corpus linguistics, which typically indicates the probability of linguistic items appearing within a specific linguistic unit, such as a clause or sentence.

In response to these limitations, this study proposes a multi-dimensional analysis (MDA) of linguistic variation in interpreted diplomatic language during Chinese Regular Press Conferences, comparing it to its non-interpreted counterpart by identifying co-occurring linguistic patterns. This approach aligns with calls for a more nuanced understanding of linguistic variation in mediated language (Kruger and Van Rooy, 2018; Hu et al. 2016; Xu, 2021), recognising that examining co-occurrence patterns across multiple dimensions provides richer insights into the complex interactions between language, context, and communicative purpose (Biber, 1995, 1988). The context of Chinese Regular Press Conferences organised by the Ministry of Foreign Affairs (MFA) provides a unique opportunity to investigate interpreted language in a high-stakes setting. As one of the most representative press briefings in terms of history and influence in China (Fu and Wang, 2022), these conferences offer a direct window into China's diplomatic stances and policies conveyed through spokespersons' responses to journalists' inquiries during Q&A sessions[1]. Analysing interpreted language in this context reflects the linguistic norms and practices upheld by highly skilled interpreters from the MFA's Department of Translation and Interpretation, considered a benchmark for interpreting excellence in China (Li, 2018, p. 4).

This research aims to provide a comprehensive understanding of the linguistic variation inherent in mediated diplomatic communication. To achieve this, it investigates the co-occurring patterns of linguistic features in interpreted diplomatic language compared to non-interpreted counterparts. By illuminating the multi-dimensional nature of interpreted language within this specialised context, this research holds significant implications for corpus-based interpreting studies and pedagogical approaches for diplomatic interpreter training. This study will address the following research questions:

(1) What are the similarities and differences between interpreted diplomatic language and its non-interpreted counterpart in terms of the co-occurring patterns of linguistic features?
(2) What contextual factors contribute to the co-occurring linguistic patterns in interpreted diplomatic language?

Literature review

Interpreted diplomatic language in press conferences. The linguistic turn in the 1990s initiated scholarly investigations into language use within diplomatic settings (Guo, 1990; Schäffner, 2004). Early research concentrated on the stylistic features of original diplomatic language, identifying characteristics such as precision, formality, ambiguity, and preparedness (Ren, 2000; Hu and Wang, 2001; Guo and Wang, 2002; Schäffner, 2008; Kadrić et al. 2021). According to Kadrić et al. (2021, p. 30), diplomatic communication "transcends cultural boundaries and becomes a new form of communication that not only conveys messages, but also charms, convinces, and persuades the interlocutor to find an agreement with the speaker". This emphasis on strategic language use underscores the critical role of interpreting in diplomatic contexts, rendering the study of interpreted diplomatic language a pivotal area within interpreting studies.

While existing research has explored interpreted diplomatic language through the lenses of interpreting norms and agency (Schäffner, 2004, 2008; Bhatia, 2006; Wang, 2012; Gu and Wang, 2021), a significant gap remains in understanding its linguistic

patterns as a distinct variety compared to non-interpreted counterparts. Initial investigations have examined specific lexical and syntactic features (Russo et al. 2006; Sandrelli et al. 2010; Li and Hu, 2013; Li, 2018; Hu and Tian, 2018; Dayter, 2018; Fu and Wang, 2022; Zhang and Cheung, 2022; He, 2022). For instance, Russo et al. (2006) and Sandrelli et al. (2010), analysing the Spanish monolingual comparable corpus of the European Parliament Interpreting Corpus (EPIC), found that interpreted language is characterised by higher lexical density, corroborating Dayter's (2018) findings using the Russian-English Simultaneous Interpreting (SIREN) corpus. In the context of Chinese press conferences, Li and Hu (2013) compared the use of modal verbs in Chinese-English interpreted language with non-interpreted English employed in American press conferences. They observed a greater reliance on strong modals and reduced use of weak modals in the interpreted language, attributing this to differences in communicative tenor, Chinese linguistic structures, and the interpreter's mediating role. Adopting a broader perspective, He (2022) examined lexical and syntactical features, comparing interpreted diplomatic language with both source and non-interpreted target languages. This analysis provided evidence for the coexistence of normalisation towards the target language and the "shining through" of source language features in the interpreted language. Collectively, these studies highlight the complexity of interpreted language in high-stakes political environments such as press conferences, demonstrating that it is not merely a product of linguistic transfer but also reflects interpreters' strategic choices within specific communicative frameworks.

Despite these advancements, three key limitations persist in the literature. First, insufficient attention has been paid to the impact of communicative context on the linguistic features of interpreted diplomatic language. Halliday's (1978) register theory posits that language use varies across situational contexts, shaped by the interplay of field (subject matter or topic), mode (channel or medium), and tenor (relations among participants). The linguistic nuances of interpreted diplomatic language are likely influenced by various situational factors, including the nature of the content (e.g. diplomatic policies and the need for clarifying stances), the dynamic nature of communication (characterised by question-and-answer exchanges between spokespersons and journalists), and the interplay among participants (encompassing spokesperson-journalist and triadic spokesperson-interpreter-journalist relationships). Second, existing literature often focuses on a limited set of linguistic features, overlooking a comprehensive examination of co-occurring linguistic elements that would provide a more holistic understanding of interpreted language as a linguistic variety. Finally, previous studies have predominantly relied on difference tests to compare interpreted diplomatic language with its source or non-interpreted equivalents in the target language, an approach that does not fully capture the differences and similarities in language use between these varieties.

Therefore, a multi-dimensional analysis (MDA) that considers the interplay of various linguistic features within specific communicative contexts is crucial for fully capturing the complexities of interpreted diplomatic language. This approach will enrich the existing literature on the multi-dimensional nature of interpreted language and deepen our understanding of the linguistic attributes that typify interpreted diplomatic discourse.

Multi-dimensional analysis and its application in mediated language. To transcend the limitations of analysing isolated linguistic features, multi-dimensional analysis (MDA) emerges as a robust framework for examining the interplay of multiple linguistic elements. MDA is a multivariate statistical method systematically exploring differential linguistic patterns of language use across registers. It is predicated on the theoretical assumption that "strong co-occurrence patterns of linguistic features mark underlying functional dimensions" (Biber, 1988, p. 13). By condensing a constellation of annotated variables (i.e. linguistic features) into derived variables (i.e. factors) that demonstrate high shared variance through factor analysis, MDA facilitates nuanced comparisons of linguistic variation across diverse communicative contexts (Friginal and Weigle, 2014). Two primary approaches to MDA are prevalent in linguistic research: standard MDA and full MDA. Standard MDA, pioneered by Biber (1988), utilises a predefined set of 67 linguistic features and six dimensions, initially developed to distinguish between spoken and written English. Full MDA, as defined by Brezina (2018), adopts an inductive approach involving feature selection, factor analysis, and dimension interpretation tailored to the specific discourse domain under investigation. This method has been employed to illuminate the distinctive linguistic features of a wide range of discourses, including academic writing (Cao and Xiao, 2013; Friginal and Weigle, 2014; Jin, 2021), editorials (Huang and Ren, 2019), and corporate annual reports (Bu et al. 2020).

Despite the growing recognition of MDA in translation and interpreting studies (Hu, 2010; Xiao and Hu, 2015; Hu et al. 2016; Su and Liu, 2022; Xiao, 2015; Xu, 2021; Wang and Liu, 2024), its application to interpreted diplomatic language remains surprisingly limited. Two recent studies highlight the existing gaps and the potential in this area. Zou and Wang (2021) employed full MDA to investigate variations in translated and interpreted Chinese diplomatic discourse compared to their non-mediated counterparts. Their findings indicated that interpreted diplomatic language aligns more closely with professional letters rather than oral speeches. Specifically, interpreted language is distinct from its non-interpreted counterpart, navigating along the oral-literate and planned-unplanned continuum. The study's reliance on Biber's (1988) predefined 67 grammatical features may have limited the scope of their analysis, potentially neglecting linguistic features crucial for understanding mediated language. Additionally, the interpretation of dimensional findings could be strengthened by incorporating more robust statistical evidence, such as dimension scores, to further validate their conclusions. In a more recent study focusing on consecutive interpreting at Chinese Premier press conferences, Sheng and Li (2024) employed standard MDA to compare interpreted language to its non-interpreted counterpart in American government press conferences. Their analysis revealed that interpreted language exhibits higher information density and reference clarity, more non-narrative and abstract information, and a lower degree of involvement and information elaboration compared to non-interpreted language. However, the replication of Biber's (1988) framework may have restricted their analysis to general linguistic dimensions, potentially masking nuanced differences specific to interpreted language in this context. Furthermore, using a parametric *t-test* for significance testing, which assumes a normal distribution of data, may not be appropriate for all linguistic features and could impact the reliability of their findings. Given the different speakers within each sub-corpus, the potential influence of speaker identity on language use also warrants further consideration and control in future research.

The empirical evidence from these studies supports the notion that the MDA framework is applicable and insightful when applied to mediated languages. While pioneering studies have laid the foundation by analysing translated texts and their inherent linguistic features, research on interpreted language, particularly within diplomatic contexts, remains limited. Moreover, these studies underscore the need for a more tailored and rigorous

Table 1 Description of SI and NS corpora.

	SI	NS
Number of texts	515	242
Number of word tokens	1,140,302	1,062,135
Mean length of text	797	4389
Time period	202009–202211	202102–202211

application of MDA in the study of interpreted diplomatic language. Consequently, this study seeks to further the understanding of MDA's potential by focusing on interpreted diplomatic language, a register that poses unique challenges due to its intricate blend of formal and relational language within political and cultural contexts. By conducting a full MDA, we aim to identify specific linguistic variations that distinguish interpreted diplomatic language from its non-interpreted counterparts and contribute to a more nuanced understanding of how language functions in high-stakes cross-cultural communication.

Methodology

Corpus description. For this study, we compiled a comparable corpus using transcripts from two sources: Chinese-English simultaneous interpreting during Chinese Regular Press Conferences (SI corpus) and non-interpreted speeches from U.S. Department Press Briefings (NS corpus). The transcripts were obtained from the official websites of the Ministry of Foreign Affairs (English version)[2] and the Bureau of Global Public Affairs of the U.S. Department of State.[3] The corpora encompass spoken language used to elucidate foreign policies and engage both domestic and international audiences through responses to queries from journalists globally. Thus, the interpreted and non-interpreted diplomatic languages are comparable in communicative setting, speaker identity, and the delivery mode of spontaneous speeches (Sheng and Li, 2024). As of December 1, 2022, the SI corpus comprises 1,140,302 tokens, while the NS corpus incorporates 1,062,135 tokens. During data cleaning, we focused on preserving only the spokespersons' replies based on the premise that these responses most accurately represent the diplomatic language utilised in both Chinese and American GPCs. All paralinguistic elements, abbreviations, and bracketed notes were removed using EditPad Pro, resulting in the final corpus presented in Table 1.

Multi-dimensional analysis of SI and NS corpora. The present study adhered to the procedure framework established by Biber (1988) for the multi-dimensional analysis (MDA) of the SI and NS corpora. The analysis began with collecting texts and identifying relevant linguistic features. Firstly, for each text in the corpus, the frequency of each linguistic feature was calculated and then normalised to frequencies per 100 words to ensure comparability across texts of varying lengths. Following this normalisation process, standardised scores—commonly referred to as z-scores—were calculated for each linguistic feature. This standardisation facilitates comparative analysis across texts by reducing the influence of variance in text length and feature dispersion. Secondly, factor analysis was performed on the set of normalised frequency data. Factor analysis clusters linguistic features based on their co-occurrence patterns, thereby identifying dimensions representing core functional aspects of the language used in the corpora. Thirdly, dimension scores for each text were computed by summing the z-scores of all features with positive loadings on a given factor (indicating a direct relationship) and subtracting the sum of z-scores for features with negative loadings on the same factor (indicating an inverse relationship). The computed dimension scores quantify the degree to which each text aligns with the linguistic tendencies represented by each factor. The factors were then subjected to interpretive analysis to ascertain their functional implications, a critical step in discerning how the factors relate to actual communicative practices within the corpora. Each factor was then labelled to reflect the functional domain of language use it most strongly represents. The final analytical step involves the application of non-parametric Mann-Whitney *U* tests to detect any statistically significant differences between the two corpora. These tests were implemented across multiple levels: on each dimension, the individual linguistic features comprising each dimension, and the aggregated overall scores. The Mann-Whitney *U* test was selected for its ability to manage non-normally distributed data and its suitability for comparing two independent samples. This statistical testing establishes whether the differences observed between the interpreted SI corpus and non-interpreted NS corpus are statistically significant, providing empirical evidence for conclusions regarding the linguistic characteristics of the two sets of diplomatic discourse.

Selection of linguistic features. The meticulous selection of linguistic features is essential for detecting and interpreting linguistic patterns and understanding the fundamental dimensions they represent. Informed by Biber's (1995) seminal work, the current study adopted an inclusive approach to incorporating linguistic features, ensuring that an extensive array of features is considered while deliberately avoiding preconceived assumptions about their significance. This strategy facilitates an unbiased exploration of the data, allowing for a more authentic and revealing analysis. A review of previous research aided in identifying potential linguistic features (Biber, 1988; Gries et al. 2011; Cao and Xiao, 2013; Hu et al. 2016; Xu, 2021), resulting in an initial selection of 113 linguistic features, including 67 grammatical features, 40 n-grams, and 6 mediated language-related textual features. Fifty-four grammatical features were retained after grouping those with similar communicative functions (Jin, 2021), such as *wh-* and *that-* relative clauses, and removing 10 punctuation-related features due to their potential to confound analysis of spoken language (Brugman et al. 2021). Drawing on the methodologies of Gries et al. (2011) and Cao and Xiao (2013), this study incorporated n-grams for their proven efficacy in distinguishing between various linguistic registers. A carefully curated set of 31 trigrams and quadrigrams have been chosen for their higher informational value compared to bigrams, offering rich insights into linguistic patterns and facilitating a more nuanced differentiation of register variation. Additional mediated language-related textual features were included following Hu et al. (2016) and Xu (2021), derived from previous research on translated and interpreted languages.

Upon the selection of pertinent linguistic features, a suite of statistical instruments was employed to quantify their occurrences. These tools facilitate the accurate computation of frequency counts, providing a foundational dataset for further analysis. The frequency list of 54 grammatical features was generated by Nini's (2019) Multidimensional Analysis Tagger 1.3.3 (MAT), the frequency of 40 n-grams was counted by PatConc (Liang and Xiong, 2008), and WordSmith 7.0 (Scott, 2016) was adopted to calculate shorter words, longer words, and average sentence length. MATLAB R2020b was applied for counting the Top 10 vocabulary coverage and Textalyser[4] for extracting lexical density. Raw frequencies of each linguistic feature were normalised, and three features with an average normalised frequency below 0.003 per 100 tokens were excluded (Bu et al. 2020). Finally, the normalised frequencies of 88

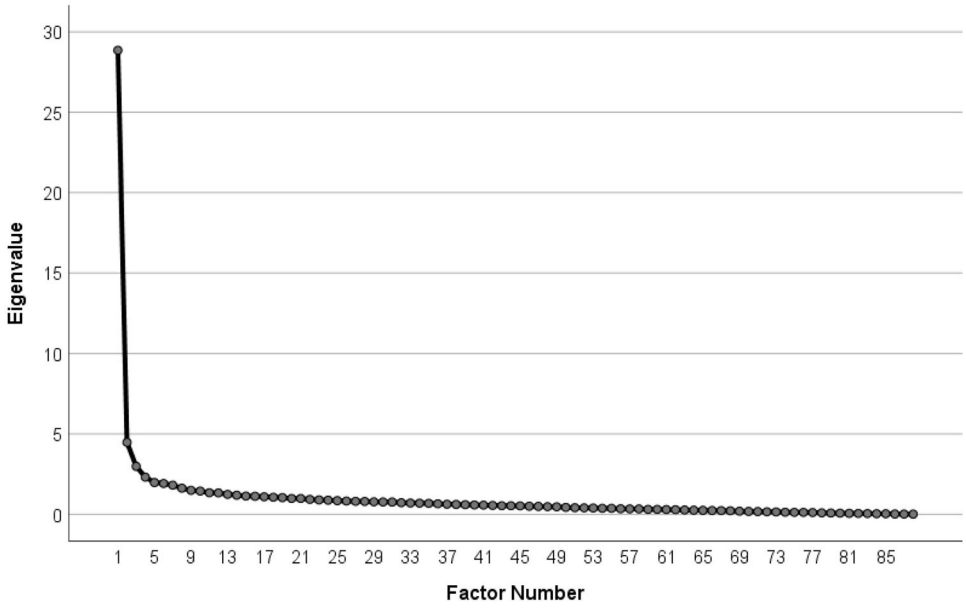

Fig. 1 Scree plot. The inflection point illustrated herein suggests a five-factor solution is optimal for extraction.

Table 2 Total variance explained by the first five factors.

Factor	Initial Eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	Variance %	Sums %	Total	Variance %	Sums %	Total	Variance %	Sums %
1	28.854	32.788	32.788	28.854	32.788	32.788	27.797	31.588	31.588
2	4.476	5.087	37.875	4.476	5.087	37.875	2.576	2.928	34.515
3	2.984	3.390	41.265	2.984	3.390	41.265	2.575	2.926	37.442
4	2.313	2.628	43.893	2.313	2.628	43.893	2.473	2.810	40.252
5	1.976	2.245	46.139	1.976	2.245	46.139	2.245	2.551	42.803

Extraction Method: Principal Axis Factoring.

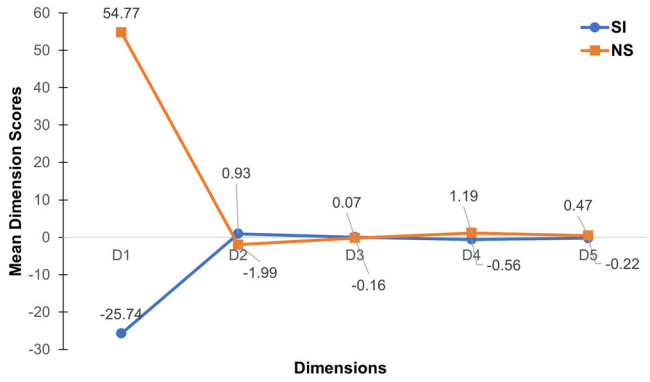

Fig. 2 Mean dimension scores of SI and NS corpora. This figure compares mean dimension scores in simultaneously interpreted versus non-interpreted diplomatic languages, providing an overview of their disparities across each dimension.

linguistic features were subject to factor analysis via SPSS 26 (Appendix A)[5].

Factor analysis. Factor analysis serves as the primary statistical tool in MDA for linguistic variation (Biber, 1988, p. 79). This method facilitates the extraction of underlying co-occurring patterns based on normalised frequency counts of linguistic features, simplifying the interpretation of resulting dimensions compared to analysing each linguistic feature separately. Prior to factor analysis, the goodness-of-fit of data was validated via Kaiser's Measure of Sampling Adequacy (KMO), resulting in a value of 0.94, and Bartlett's Test of Sphericity, which demonstrated significance ($\chi^2 = 55569.586$, $p < 0.001$). Principal factor analysis (PFA) was employed to extract co-occurring patterns among linguistic features, followed by Varimax rotation to ensure each factor is represented by a minimal number of features (Biber, 1988, p. 102). Absolute factor loadings above 0.30 were retained, adhering to the commonly accepted threshold for significant loadings (Biber, 1988, p. 87). The scree plot indicated that a five-factor solution was optimal due to minimal cross-loadings, no factors with fewer than three items, and ease of interpretation (Fig. 1). Altogether, these five factors account for 42.80% of the shared variance, comparable to other MDA studies, such as 40% for the eight-factor solution in Xu (2021) and 48.6% for the seven-factor solution in Hu et al. (2016) (Table 2).

Results and analysis

A comprehensive comparison of the mean dimension scores for simultaneous interpreted (SI) and non-interpreted spoken diplomatic languages (NS) is visually represented, as shown in Fig. 2. The Mann-Whitney U tests for dimension scores in Table 3 reveal significant differences between SI and NS regarding D1, D2, D4, and D5, with no significant difference observed on D3. The most substantial variation between SI and NS occurs on D1. The interpretation of the factors is based on the theoretical assumption that "these co-occurring linguistic patterns indicate an underlying communicative function shared by the features" (Biber, 1988, p. 101). Accordingly, the linguistic features grouped

Table 3 Mann-Whitney U tests for dimension scores between SI corpus and NS corpus.

	SI		NS		z	P	H_0
	Mean	SD	Mean	SD			
D1	−25.74	11.82	54.77	11.55	−22.209	<0.001	×
D2	0.93	2.65	−1.99	5.26	6.651	<0.001	×
D3	0.07	3.06	−0.16	2.67	0.675	0.500	√
D4	−0.56	2.68	1.19	2.98	−7.888	<0.001	×
D5	−0.22	2.58	0.47	3.46	−4.489	<0.001	×

"x" means the null hypothesis H0 is rejected; "√" means H0 is retained.
Significant at $p < 0.05$ (2-tailed).

Table 4 Linguistic features on D1 (See full version in Appendix B).

	Positive	Loading	Negative	Loading
D1	First-person pronouns	0.925	Lexical density	−0.975
	Demonstratives	0.904	Average word length	−0.968
	Demonstrative pronouns	0.889	Longer words	−0.951
	Contractions	0.882	Total other nouns	−0.920
	Present tense	0.877	Attributive adjectives	−0.920
	Analytic negation	0.819	Phrasal coordination	−0.905
	Relative clauses	0.814	N+ and/or +N	−0.863
	Total adverbs	0.807	Nominalizations	−0.817

on each factor were interpreted as a dimension by analysing the communicative functions most widely shared by the features and referring to dimensions in prior MDA studies. The results for each dimension are elaborated upon in the following sections.

Dimension 1: involved vs. informational production. D1, akin to Biber's (1988, p. 104) "Involved vs. Informational Production" dimension, represents a salient factor embodying linguistic variation, with 33 features on the positive end and 26 features on the negative end (See features with loadings over 0.80 in Table 4). Features with positive weights represent a reduced surface form, a generalised or uncertain presentation of information, and a generally fragmented production of texts (Biber, 1988, p. 106). Pronominal forms, such as first-, second-, and third-person pronouns, pronoun *it*, and demonstrative pronouns, suggest "a relatively low informational load, a lesser precision in referential identification, or a less formal style" (Biber, 1988, p. 225). The reduced surface form is marked by demonstrative pronouns, contractions, and shorter words. Analytic negation, *be* as main verb, and possibility modals indicate a fragmented or uncertain presentation of information. Three subordination features—causative subordination, conditional subordination, and *wh*-clause—indicate a relatively loose presentation of information and relate to affective functions associated with personal attitudes or feelings (Biber, 1988, p. 107).

Features with the highest negative loadings include lexical density, average word length, and longer words, which signify a high informational focus, careful integration of information, and precise lexical choice (Biber, 1988, p. 105). Additionally, noun-centred colligation patterns, including $N + and/or + N$, $V + art/adj./adv./N + N$, $N + prep + V\text{-}ing$, and $adj. + N + V$ further contribute to the expansion of idea units (Huang and Ren, 2019). Therefore, D1 can be characterised as "Informational vs. Involved Production".

Table 5 Linguistic features on D2.

	Positive	Loading	Negative	Loading
D2	Perfect aspect	0.668	Second person pronouns	−0.441
	Other adverbial subordinators	0.562	Public verbs	−0.423
	Downtoners	0.326	Predictive modals	−0.341
	Pronoun it	0.325		

Table 3 indicates significant variation between SI and NS on D1 ($p < 0.001$), with SI loading on the negative side and NS on the positive side (−25.74 vs. 54.77). As illustrated in Extract 1a[6], SI employs a series of total other nouns (e.g. efforts), attributive adjectives (e.g. mutual), and $V + art/adj./adv./N + N$ (e.g. *promote mutual learning)* to convey China's awareness of diplomatic responsibility and contributions to the world. Colligation pattern $N + and/or + N$ (e.g. *state governance and development)* and phrasal coordination (e.g. and) are also adopted to enhance the integration of information in the texts. Comparatively, NS in Extract 1b evinces a greater prevalence of first-person pronouns (e.g. I), demonstratives (e.g. that), contractions (e.g. you've), present tense (e.g. think), and analytic negation (e.g. wouldn't), aiming to underscore the importance that America places on upholding its own values and principles in all diplomatic relationships. Biber (1988, p. 231) notes that demonstratives are often found in conjunction with interactive and spontaneous discourse features, such as first-person pronouns and contractions. The frequent co-occurrence of these elements highlights a pattern of language use designed to facilitate interpersonal interaction. Such usage indicates an involved and interactive communicative objective, typically manifested in speech production contexts where immediacy and relational engagement are prioritised.

Extract 1a:

We are accelerating efforts to build a Global Knowledge Network for Development and engaging actively in experience-sharing on *state governance and development* to *promotemutual learning* among countries. The GDI is an example of China contributing its solutions to addressing the global *governance deficit and other global challenges*. It is a vivid illustration of building a community with a shared future for mankind that demonstrates the sense of responsibility of China's diplomacy in the new era. (SI_20221020)

Extract 1b:

I wouldn't want to **go** beyond what is in the readout. What **I** would **say**, however, is **I think** something **you've** probably heard **me say** before is **that** in every relationship, whether it is one with **our** closest allies, **our** closest partners, and with **our** closest security partners, **we** will **never check our** values, **we** will **never check our** principles at the door. (NS_20210210)

Dimension 2: objective vs. addressee-focused narration. D2 contrasts the co-occurrence of the perfect aspect, other adverbial subordinators, downtoners, and pronouns *it* on the positive side with second-person pronouns, public verbs, and predictive modals on the negative side (Table 5). Perfect tense verbs, adverbials, and pronouns all fall into the stereotypically narrative

Table 6 Linguistic features on D3.

	Positive	Loading
D3	art + N + of	0.855
	prep + art + N + of	0.846
	Top 10 vocabulary coverage	0.616
	Total prepositional phrases	0.531

linguistic features. Specifically, other adverbial subordinators include various causative, concessive, and conditional ones (Huang and Ren, 2019). Perfect aspect verbs are predominantly employed to describe past events, and when combined with adverbials and pronouns, they create an objective narrative about a specific situation (Friginal, 2009, p. 81). Downtoners indicate the implementation of hedging tactics to reduce the force of verbs and avoid stance-taking expressions.

The heaviest loading of second-person pronouns at the negative end suggests that "the transfer of information is highly addressee-focused" (ibid). Public verbs typically involve actions observable by others and are primarily used to introduce indirect statements (Biber, 1988, p. 242). Predictive modals express the speaker's volition or prediction regarding future events or outcomes and their epistemic stance (ibid, p. 241). When combined with second-person pronouns, these features indicate the speaker's intention to inform the addressee about the possibility of specific events. Collectively, these linguistic features lead to the interpretation of D2 as "Objective vs. Addressee-focused Narration".

Compared to SI, NS displays a statistically significant increase in the frequency of second-person pronouns, public verbs, and predictive modals on D2 ($p < 0.001$), with SI having a positive weight and NS having a negative weight (0.93 vs. −1.99). As demonstrated in Extract 2b, NS employs several second-person pronouns (e.g. you) to engage the journalist directly in interactive communication. In responding to questions, NS incorporates predictive modals (e.g. won't) and public verbs (e.g. comment) to provide an indirect and hedging answer regarding the government's position on certain issues, thereby indicating its stance towards the death of an activist. On the contrary, SI in Extract 2a is characterised by a lower frequency of second-person pronouns juxtaposed with greater usage of perfect aspect (e.g. has gone up), other adverbial subordinators (e.g. whereas), pronoun *it*, and especially the downtoners (e.g. may, nearly). These features collectively contribute to an objective narration of China's achievements in protecting people from COVID-19 and increasing their average life expectancy during the pandemic. The markedly higher frequency of downtoners in SI suggests an intent to express an objective stance and present probability more neutrally, thereby illustrating China's prudent decision to fight over the pandemic persistently.

Extract 2a:

Whereas the average life expectancy in some developed countries, such as the US, declined during the pandemic, China's average life expectancy **has** steadily **gone** up in the past two years. A research paper of Nature Medicine estimated that if China stops putting up a tough fight over COVID as some other countries did, **it may** lead to more than 112 million infections and **nearly** 1.6 million deaths. (IS_20220617)

Extract 2b:

So I'm going to start with your second question first. For that, we simply won't comment on the holds, and for that I'd have to send you back to Senator Risch's office for any specific commentary. To your first question, I'll just say that we're deeply disturbed by the death of Palestinian activist Nizar Banat and the information that has been reported regarding the circumstances surrounding his death. (NS_20210624)

Dimension 3: Literate-Oral Continuum. The prominent linguistic features co-occurring at the positive end of D3 encompass elements related to information density, such as total prepositional phrases and two noun-centred colligation patterns (Table 6). Total prepositional phrases, commonly found in written texts, facilitate information integration into idea units and expand the information conveyed within each idea unit (Biber, 1988, p. 237). Biber et al. (2006) note that "the reliance on nouns and complex noun phrases results in a style of the text with dense informational content packed into relatively few words", exemplified by structures such as *art + N + of* and *prep + art + N + of*. These linguistic features collectively characterise a literate style prioritising information integration and density. Notably, the high loading of the Top 10 vocabulary coverage indicates the use of repetitive and simplified spoken language, attributable to the constrained nature of speech production (Xu, 2021). Considering all these linguistic features, D3 is thus labelled as "Literate-Oral Continuum".

Although D3 shows no significant difference ($p = 0.500$) between SI and NS, SI demonstrates a significant increase in the usage of *art + N + of* (e.g. *the message of*) and total prepositional phrases (e.g. from, for) ($p < 0.05$, respectively) to express China's gratitude and firm commitment to establishing strategic partnership with Pakistan. These features contribute to a high information density and a literate-oriented production in SI. Similarly, *prep + art + N + of* (e.g. *with a number of*) in NS also points to an informative production regarding the American commitment to promoting democracy in Venezuela by collaborating with various allies and partners (Extract 3b). However, the significantly higher Top 10 vocabulary coverage ($p < 0.05$) lends itself to a more impromptu, unplanned oral production within the constraints of speech. Combining the mean dimension scores (SI vs. NS = 0.07 vs. −0.16), the observed features imply a tendency towards literateness in SI, whereas NS leans towards orality, despite both falling on the literate end of D3.

Extract 3a:

Having received *the message of* congratulation **from** Prime Minister Imran Khan, the Chinese side expresses its thanks and appreciation. CPC stands firmly **for** developing a friendly relation **with** Pakistan, which is also a consistent position of the Chinese Government. ***With*** *the efforts of* our two leaders and supports **of** our two peoples, China and Pakistan have become all-weather strategic partners. (SI_20210701)

Extract 3b:

We will work *with a number of* allies and partners to bring about progress **towards** democracy **in** Venezuela. We will do that **with** our partners **in** the region, we will do that **with** our European partners and allies, we'll do that **with** the OAS, we will do that **through** the Lima Group, ***through*** *a number of* fora. (NS _20210203)

Dimension 4: information elaboration. D4 builds on three colligation patterns of adjectives and *be* verbs, together with amplifiers (Table 7), all situated on the positive end, similar to

Table 7 Linguistic features on D4.

	Positive	Loading
D4	V + adv + adj	0.910
	be + adv + adj	0.797
	adv + adj + N	0.485
	Amplifiers	0.480

Table 8 Linguistic features on D5.

	Positive	Loading	Negative	Loading
D5	Past tense	0.685	Split auxiliaries	−0.398
	Third person pronouns	0.388	PP + prep + V-ing	−0.317
	Seem/appear	0.373		
	Average sentence length	0.311		

D3. Adjectives, serving as the foci in these colligation patterns, stand for information elaboration and expansion (Biber, 1988, p. 237). The three adjective-related colligation patterns are associated with evaluation, conveying the speakers' stance while indicating informational or detailed concepts (Huang and Ren, 2019). Amplifiers intensify the force of verbs, signalling solidarity with the listener and expressing certainty or conviction towards the proposition (Holmes, 1984; Quirk et al. 1985, p. 623). The co-occurrence of these linguistic features reflects high informational density with specialised description, thus this dimension can be tagged as "Information Elaboration".

A significant difference exists between SI and NS on D4 ($p < 0.001$), with NS displaying a positive loading and SI a negative one (1.19 vs. −0.56). This suggests that NS employs language in a more conceptually elaborative manner. As shown in Extract 4a, NS intends to convey its belief in the significance of maintaining international exchanges. By endorsing the ambassador's viewpoint using an amplifier (e.g. completely), NS further elaborates its stance through several adjective-related patterns, such as *be + adv. + adj.* (e.g. *are always important, are especially important, been very vocal*).

Extract 4a:

So that said, the ambassador's point is a **completely** valid one and one we believe in. We believe that lines of communication, lines of dialogue, *are always important*, but they *are especially important* at – during times of increased tension or, in this case, even conflict or war. We want to see those lines preserved. It's why we have *been**very** vocal* in speaking out against the unjustified steps that the Russian Government had taken vis-à-vis our diplomatic presence in Moscow. (NS_20220606)

Dimension 5: narrative vs. non-narrative concerns. Features with positive weights on D5 include past tense verbs, third-person pronouns, seem/appear, and average sentence length (Table 8). The two most salient positive features—past tense verbs and third-person pronouns—are major components of narrative discourse, describing past events and referring to animate entities other than the speaker and addressee, thereby indicating a narrative and reported style (Biber, 1988, p. 109). Perception verbs like seem/appear provide additional evidence of narrative concerns, functioning as downtoners related to impression or appearance rather than asserting concrete past occurrences, which frames the storytelling aspect.

Negatively weighted features on D5 include split auxiliaries and the three-gram construction *pp + prep + V-ing*. Split auxiliaries occur when auxiliaries are followed by an adverb(s) and a verb, often co-occurring with information-rich linguistic features such as passives and prepositions (ibid, p. 111). The presence of the passive construction *pp + prep + V-ing* contributes to thematic discourse cohesion and emphasises the patient over the agent (Cao and Xiao, 2013). The co-occurrence of these features suggests an obscured identification of the agent alongside a strong information-focused description characteristic of non-narrative discourse. The positive and negative loadings thus represent two ends of narrative expression, leading to the designation of D5 as "Narrative vs. Non-Narrative Expression".

Table 3 indicates a significant difference between SI and NS concerning D5, with SI exhibiting a negative loading of −0.22 and NS displaying a positive loading of 0.47 ($p < 0.001$). In Extract 5b, NS employs typical narrative features—such as past tense verbs (e.g. heard, were) and third-person pronouns (e.g. they, their)—to recount statements by the United States and Iran regarding bilateral negotiations. Seem/appear (e.g. seems like) is adopted to reflect on the events being discussed and convey the impression of recent occurrences. In contrast, SI is characterised by the use of split auxiliaries, exemplified by phrases such as "has always strictly observed", and constructions involving a past participle followed by a preposition and a gerund (*pp + prep + V-ing*), like "*committed to maintaining*". These linguistic choices are instrumental in articulating a non-narrative style that foregrounds factual information, establishing a formal tone and projecting a sense of authority. Together, these stylistic elements underscore the precision of China's stance and reinforce its commitment to the principles governing China-India relations, effectively conveying a non-narrative focus that emphasises the clarity and firmness of China's diplomatic position.

Extract 5a:

China's position on the China-India boundary issue is consistent and clear. China has always strictly observed the relevant agreements signed by the two countries and is *committed to maintaining* peace and stability in the China-India border areas, while firmly safeguarding national territorial sovereignty and security. (SI_20200916)

Extract 5b:

Well, we have **heard** similar statements from the Irani Government at various levels over the past couple weeks. If you recall, we **were** talking about this in New York, which **seems** like it was just last week, a couple weeks ago now. And we have **heard** from the Iranians that **they** expect negotiations to resume soon. We hope **their** definition of soon matches our definition of soon. (NS_20211007)

Overall style comparison. To assess the general stylistic differences between the SI and NS corpora, we employed the method proposed by Huang and Ren (2019) to compute an overall formal-informal style score for each text. This quantitative measure contrasts formal and informal styles and mitigates the impact of extreme dimension scores, particularly in short texts, by weighting the scores according to the percentage of variance explained by the factors. The overall score F_i of each text is calculated using the formula:

$$F_i = \sum_{j=1}^{n} a_j D_{ji} (i = 1, 2, \ldots, N; j = 1, 2, \ldots, n)$$

Table 9 Normality test for overall scores.

		Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	p	Statistic	df	p
Overall score	NS	0.081	242	0.001	0.964	242	0.000
	SI	0.050	515	0.004	0.981	515	0.000

Table 10 Mann-Whitney U test for overall scores.

	SI		NS		z	p	H_0
	Mean	SD	Mean	SD			
Overall score	−16.71	7.77	35.55	7.82	−22.209	< 0.05	×

"×" means the null hypothesis H_0 is rejected; "√" means H_0 is retained.
Significant at $p < 0.05$ (2-tailed).

where F_i is the overall score of text i, a_j is the weighting of factor j (*i.e.* the percentage of the factor's squared loading relative to the total explained variance), and D_{ji} is the dimension score of text i on factor j. Here, n represents the number of factors extracted, and N is the number of texts in the corpus.

In this calculation, the scores for D3 are reversed to align the positive or negative orientation with the other four dimensions, since a positive dimension score indicates an oral/informal style, while a negative score signifies a literate/formal style (Westin and Geisler, 2002; Huang and Ren, 2019). Due to the non-normal distribution of the overall scores, as shown in Table 9, it is inappropriate to compare the general stylistic characteristics of the two corpora using a *t*-test or by simply averaging the overall scores. Therefore, the non-parametric Mann-Whitney U test is a suitable alternative for comparing the overall style of the two corpora. The results reveal a significant difference between the two corpora ($z = -22.209$, $p < 0.05$), suggesting that the SI corpus generally exhibits a more formal style compared to the NS corpus (Table 10).

Discussion

By employing a full MDA framework, this research aims to investigate the linguistic variations in simultaneously interpreted language (SI) in contrast to its non-interpreted counterpart (NS) in government press conferences. The comparative analysis reveals that interpreted diplomatic language, to a large extent, significantly diverges from non-interpreted language, exhibiting higher informativeness (low D1 scores), greater objectivity (high D2 scores), reduced elaboration (low D4 scores), a non-narrative style (low D5 scores), and a formal tone (low overall scores). Both SI and NS, however, display similar tendencies along the literate-oral continuum (similar D3 scores).

The distinctions in linguistic style observed in D1 and D2 reveal a significant trend towards a more informative and objective approach in interpreted diplomatic language compared to non-interpreted speech. This trend can be attributed to the communicative aims, interpreting norms, and participant dynamics inherent in Chinese press conferences. The subject matter and purpose of these events play a central role in shaping the linguistic style of the interpreted language. As Chen (2011), Zou and Wang (2021), and Sheng and Li (2024) pointed out, the necessity to convey crucial diplomatic information efficiently within time constraints prompts spokespersons to craft statements that are dense with information and maintain an objective tone. This approach ensures accurate and authoritative communication on diplomatic matters, avoiding unnecessary embellishment or subjectivity. The observed difference in informativeness and objectivity can thus be linked to specific norms of language use within the diplomatic context of Chinese press conferences. As highlighted by Wang (2012), based on the analysis of similar events, interpreting norms in this setting prioritise adequacy, explicit logical relations, rich informational content, and clarity of meaning. This emphasis on faithful conveyance aligns with the stringent nature of interpreting in Chinese press conferences, where interpreters have limited latitude due to the high-stakes political context (Ren, 2000). Given these constraints, interpreters may prioritise accuracy and completeness in their renditions, even if this results in a more formal style than the non-interpreted counterpart, thereby contributing to the differences observed between the two diplomatic languages.

The interaction dynamics among participants at these conferences also influence the observed linguistic style. Chinese press conferences typically adhere to a strict question-and-answer format, devoid of interruptions or interactive exchanges during each segment. This structured procedure, mediated by interpreters, fosters a somewhat detached communication relationship. In stark contrast, American press conferences, as described by Schäffner (2008) and Fu and Wang (2022), promote a more conversational, interactive, and informal environment that encourages interruptions, follow-ups, and cross-questioning. The observation that English translations of China's diplomatic language often employ content-heavy words—thereby underpinning an impression of solidity and practicality—aligns with the finding of Hu and Tian (2018). This linguistic usage supports broader communicative goals of projecting a grounded and pragmatic image in international relations. This result stands in contrast to the findings of Xu (2021), who posits that interpreted language typically exhibits characteristics of involvement and informality. The discrepancy points to the possibility that the attributes of interpreted language, such as informativeness and formality, may vary considerably depending on the context. Xu's (2021) research focused on legislative environments, whereas the present study concentrates on diplomatic settings. This context-dependent nature of language use highlights the adaptability of interpreted language to meet the specific requirements and expectations of different institutional settings.

The absence of a significant difference along the literate-oral continuum (D3) between interpreted and non-interpreted diplomatic languages suggests that both forms share characteristics, placing them similarly in this dimension. Diplomatic speeches and statements inherently convey seriousness and carry significant weight. Although delivered orally, they are typically scripted, imbuing them with qualities more commonly associated with written language. This aligns with the observations of Guo (1990) and Guo and Wang (2002), who propose that the oral delivery of diplomatic communications is underpinned by a written structure emphasising formality and careful composition. The tendency for interpreted language to exhibit features associated with a literate style, as Shlesinger and Ordan (2012) noted, is reflected in the lack of significant differences observed in this study. Interpreters may unconsciously introduce literate constructs into their output, possibly due to the cognitive processes involved in interpreting or as a deliberate choice to enhance clarity and formality in the target language. Additionally, these findings are consistent with previous research (Ren, 2000; Xiao, 2015; Xiao and Hu, 2015; Zou and Wang, 2021), which provides empirical support for the notion that the literate-oral dimension is a fundamental and universal aspect of interpreted language. Interpreted language, regardless of the original speech's style, often shifts towards the literate end of the spectrum. This tendency may result from interpreters' training—which typically emphasises accuracy and the preservation of the original

message's formality—and the context of diplomatic settings, which often require a formal register and careful word choice to avoid misunderstandings or diplomatic incidents. The convergence of interpreted and non-interpreted diplomatic language on the literate-oral continuum suggests that interpreting does not substantially alter the inherent literate characteristics of diplomatic discourse. It also underscores the complex nature of interpreted language, balancing the spontaneity of speech with the deliberateness of writing.

Considering D4, which pertains to the degree of information elaboration, interpreted diplomatic language tends to exhibit less elaboration than non-interpreted speech. This phenomenon can partly be attributed to the presence of shared background knowledge among participants, as articulated by Biber (1988) and Hu and Wang (2001). In diplomatic contexts, spokespersons often address questions by implying information presumed to be commonly understood, thereby avoiding unnecessary repetition or detailed explanations of familiar concepts. The more condensed nature of interpreted language suggests that interpreters may operate under the assumption that their audience possesses significant background knowledge. This is likely influenced by the intrinsic constraints of the interpreting process, such as limited time, the cognitive load associated with memory retention, and disparities in shared knowledge between source and target audiences. As Gumul (2017, p. 140) indicates, these constraints may lead interpreters to "filter" information from the source language, concentrating on transmitting novel information while relying on implicatures to convey shared knowledge. In relation to D5, concerning narrative versus non-narrative style, interpreted diplomatic language is characterised by a lower incidence of narrative elements compared to its non-interpreted counterpart. Non-narrative discourse is often associated with expository and procedural information that is descriptive or argumentative rather than narrating events or individual experiences (Biber, 1988, p. 137). Diplomatic language typically serves functions such as clarifying policies, expressing governmental stances on international issues, and facilitating understanding among domestic and global audiences. Guo (1990) underscores the importance of these communicative goals within diplomatic language. Interpreters, acting as mediators in diplomatic communication, are thus expected to focus on conveying essential information rather than engaging in narrative storytelling or expressing personal sentiments. The priority lies in accurately transmitting the core content of diplomatic messages, which often necessitates a formal and structured linguistic approach aligned with expository and argumentative styles. This ensures that diplomatic discourse remains focused on its primary objectives of policy clarification and stance communication, thereby supporting the overarching goals of diplomacy.

The observation that interpreted diplomatic language is generally more formal than its non-interpreted counterpart aligns with the expectations and conventions of diplomatic communications, particularly within the Chinese context. Several factors contribute to the formal tone maintained by Chinese diplomatic interpreters: (1) Seriousness of Subject: Diplomatic discourse often involves matters of national and international importance. The gravity of these subjects necessitates a formal tone to convey the appropriate level of seriousness and respect (Ren, 2000); (2) Sensitivity of Language: Diplomatic language is scrutinised for nuance and implication. Interpreters must exercise precision and caution in their choice of words to avoid misinterpretation or diplomatic incidents; (3) Awareness of the Audience: Diplomats and interpreters recognise that their audience includes both domestic and international listeners, including government officials, policymakers, and the general public. A formal tone ensures that the message is taken seriously by all parties and remains accessible to non-native speakers who may be listening. In contrast, American press conferences are noted for their more informal and conversational style (Schäffner, 2008). This difference in formality is evident in the transcripts of press conferences from the two countries. American transcripts are likely to capture the full spectrum of oral communication, including colloquialisms, interjections, hesitations, and other features of spontaneous speech that contribute to the informal tone of the interactions. Chinese transcripts tend to record proceedings in a more edited and structured format, presenting only the questions and responses. This method of documentation can strip away the informal aspects of oral interaction, resulting in transcripts that appear more formal. The "minimum flexibility" afforded to Chinese interpreters regarding formality reflects the constraints under which they operate. These constraints are not solely linguistic but are also influenced by political and cultural expectations. Consequently, the formal tone in interpreted diplomatic language serves a dual purpose: it ensures clarity and respect in communication and maintains the desired image and posture of the state in international relations.

Conclusions

This study represents one of the first attempts to identify the distinctive co-occurring linguistic patterns that typify interpreted diplomatic language as a linguistic variation compared with its non-interpreted counterpart, employing a full multi-dimensional analysis (MDA). From the initial 113 linguistic features, five functional dimensions were identified: (1) Involved vs. Informational Production, (2) Objective vs. Addressee-focused Narration, (3) Literate-Oral Continuum, (4) Information Elaboration, and (5) Narrative vs. Non-narrative Concerns. The comparative analysis uncovers prominent discrepancies between interpreted and non-interpreted diplomatic speeches across four distinct dimensions, with the exception of D3. This illustrates that interpreted language tends to be informative, objective, succinct, non-narrative, and formal. These results underscore the importance of contextual influences in shaping the linguistic characteristics of interpreted language.

The significance of this study is twofold. Academically, it advances corpus-based interpreting studies (CBIS) by delving into the complex multi-dimensionality of interpreted language through a multi-dimensional approach. This enhancement to the academic literature provides deeper insights into the subtleties and variations inherent in interpreted language. Pedagogically, it holds considerable pedagogical value for the training and practice of interpreters specialising in diplomatic contexts. The identified linguistic patterns within professionally interpreted diplomatic language are invaluable for the development of educational materials. They serve as foundational resources for textbook creation and classroom instruction, assisting teachers in their educational roles. Moreover, these patterns can function as benchmarks for training interpreters, improving their skills and enabling them to execute diplomatic interpreting tasks with greater fluency and accuracy. This study, therefore, contributes to raising the standard of interpreter training and enhancing the professional competency of practitioners in the field.

Indeed, the study is subject to certain limitations. Methodologically, using a fixed threshold for factor loadings, instead of a sample-size-based calculated threshold, may have overlooked significant loadings that fell below the predetermined threshold. While screen plots provide a visual representation of the data, parallel analysis could offer a more robust and objective method for identifying the optimal number of factors. Furthermore, the variance explained by the factorial solution does not meet the 50% threshold, suggesting the need for a more refined analytical method to capture a more significant proportion of the variance. Additionally, the cross-loadings suggest an overlap or correlation between different dimensions, which may result from the multi-dimensional nature of language, the multifunctionality of

linguistic features, and sample-specific characteristics. To address this issue, further studies should consider employing oblique rotation methods, such as Promax or Direct Oblimin, which allow for correlations among factors. Furthermore, conducting a sensitivity analysis that compares results from different rotation methods could also help assess the robustness of the factor structure. By building upon these findings and addressing the identified limitations, future research can continue to advance our understanding of the intricacies of interpreted language and its role in facilitating effective communication in diplomatic settings.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author upon reasonable request.

Received: 29 March 2024; Accepted: 15 October 2024;
Published online: 24 October 2024

Notes

1 See https://www.fmprc.gov.cn/mfa_eng/wjb_663304/zzjg_663340/xws_665282/ for the introduction of the Information Department of the Ministry of Foreign Affairs
2 https://www.fmprc.gov.cn/mfa_eng/xwfw_665399/s2510_665401/2511_665403/
3 https://www.state.gov/department-press-briefings/?coll_filter_year&coll_filter_month&coll_filter_speaker&coll_filter_country&coll_filter_release_type=393&coll_filter_bureau&results
4 http://textalyser.net/
5 For detailed definitions and explanations of the linguistic features adopted in this study, please refer to Appendix II in Biber's (1988: 223–245) work.
6 Examples are listed to contrast texts with positive and negative dimension scores, thereby assisting in the interpretation of the dimensions. In these examples, positive-loading features are highlighted in bold, negative-loading features are underlined, and n-grams are italicised.

References

Baker M (1993) Corpus linguistics and translation studies: Implications and applications. In: Baker M, Francis G, Tognini-Bonelli E (eds) Text and Technology: In Honour of John Sinclair. Benjamins, Amsterdam, p 233-250
Bhatia A (2006) Critical discourse analysis of political press conferences. Discourse Soc 17:173–203
Biber D (1988) Variation Across Speech and Writing. Cambridge University Press, Cambridge
Biber D (1995) Dimensions of Register Variation: A Cross-Linguistic Comparison. Cambridge University Press, Cambridge
Biber D, Davies M, Jones JK, Tracy-Ventura N (2006) Spoken and written register variation in Spanish: a multi-dimensional analysis. Corpora 1(1):1–37
Brezina V (2018) Statistics in Corpus Linguistics: A Practical Guide. Cambridge University Press, Cambridge
Brugman BC, Burgers C, Beukeboom CJ, Konijn EA (2021) From the daily show to last week tonight: a quantitative analysis of discursive integration in satirical television news. Journal Stud 22(9):1181–1199
Bu H, Connor-Linton J, Wang L (2020) Linguistic variation in the discourse of corporate annual reports: a multi-dimensional analysis. Discourse Stud 22(6):647–677
Calzada Pérez M, Ramos MMS (2021) MDA analysis of translated and non-translated parliamentary discourse. In: Ji M, Michael O (eds) Corpus Exploration of Lexis and Discourse in Translation. Routledge, London/New York, p 26-55
Cao Y, Xiao R (2013) A multi-dimensional contrastive study of English abstracts by native and non-native writers. Corpora 8(2):209–234
Chen N (2011) The evolving Chinese government spokesperson system. In: Wang J (ed) Soft Power in China: Public Diplomacy through Communication. Palgrave Macmillan, New York, p 73-93
Dayter D (2018) Describing lexical patterns in simultaneously interpreted discourse in a parallel aligned corpus of Russian-English interpreting (SIREN). Forum 16(2):241–264
De Sutter G, Lefer MA (2020) On the need for a new research agenda for corpus-based translation studies: a multi-methodological, multifactorial and inter-disciplinary approach. Perspectives 28(1):1–23
Ekström M, Eriksson G (2017) Press conferences. In: Wodak R, Forchtner B (eds) The Routledge Handbook of Language and Politics. Routledge, London, p 342-354
Friginal E (2009) The Language of Outsourced Call Centers: A Corpus-Based Study of Cross-Cultural Interaction. John Benjamins, Amsterdam
Friginal E, Weigle S (2014) Exploring multiple profiles of L2 writing using multi-dimensional analysis. J Second Lang Writ 26:80–95
Fu R, Wang K (2022) Hedging in interpreted and spontaneous speeches: a comparative study of Chinese and American political press briefings. Text Talk 42(2):153–175
Gries ST, Newman J, Shaoul C (2011) N-grams and the clustering of registers. Empir Lang Res J 5(1) http://ejournals.org.uk/ELR/article/2011/1
Gu C (2024) One-third of a century on: the state of the art, pitfalls, and the way ahead relating to digital humanities approaches to translation and interpreting studies. Digit Scholarsh Humanit 39(1):154–161
Gu C, Wang B (2021) Interpreter-mediated discourse as a vital source of meaning potential in intercultural communication: The case of the interpreted premier-meets-the-press conferences in China. Lang Intercult Commun 21(3):379–394
Gumul E (2017) Explicitation in Simultaneous Interpreting: A Study into Explicitating Behaviour of Trainee Interpreters. Wydawnictwo Uniwersytetu Śląskiego, Katowice
Guo H (1990) The linguistic features of diplomatic language. Mod Foreign Lang (4):9–14
Guo L, Wang H (2002) On the preciseness and vagueness of diplomatic language. J Foreign Aff Coll (4):80–84
Halliday MAK (1978) Language as Social Semiotic: The Social Interpretation of Language and Meaning. Edward Arnold, London
He L (2022) A Corpus-Based Research on Normalization and SL Shining Through in Chinese-English Diplomatic Interpreting—A Case Study of Chinese Foreign Minister's Press Conferences (2016–2020). Dissertation, Beijing Foreign Studies University
Holmes J (1984) Hedging your bets and sitting on the fence: some evidence for hedges as support structures. Te Reo 27(1):47–62
Hu G, Wang J (2001) The analysis of the language use in Sino-foreign press conferences. J Tsinghua Univ (Philos Soc Sci) 16(3):83–88
Hu K, Tian X (2018) A corpus-based study of the image of Chinese diplomacy in the English translations of Chinese discourse on diplomacy. Foreign Lang China 6:79–88
Hu X (2010) A corpus-based multi-dimensional analysis of the stylistic features of translated Chinese. Foreign Lang Teach Res 42(6):451–458
Hu X, Xiao R, Hardie A (2016) How do English translations differ from non-translated English writings? A multi-feature statistical model for linguistic variation analysis. Corpus Linguist Linguist Theo 15(2):347–382
Huang Y, Ren W (2019) A novel multidimensional analysis of writing styles of editorials from China Daily and the New York Times. Lingua 235
Jin B (2021) A multi-dimensional analysis of research article discussion sections in an engineering discipline: Corpus explorations and scientists' perceptions. SAGE Open 11(4)
Kadrić M, Rennert S, Schäffner C (2021) Diplomatic and Political Interpreting Explained. Routledge, London/New York
Kruger H, Van Rooy B (2016) Syntactic and pragmatic transfer effects in reported-speech constructions in three contact varieties of English influenced by Afrikaans. Lang. Sci. 56:118–131
Kruger H, Van Rooy B (2018) Register variation in written contact varieties of English: a multidimensional analysis. Engl World-Wide 39(2):214–242
Laviosa S (2002) Corpus-Based Translation Studies: Theory, Findings, Applications. Approaches To Translation Studies. Rodopi, Amsterdam/New York
Li X (2018) The Reconstruction of Modality in Chinese-English Government Press Conference Interpreting: A Corpus-Based Study. Springer, Singapore
Li X, Hu K (2013) A corpus-based study of modal verbs in Chinese-English government press conference interpretation. Technol Enhanc Foreign Lang Educ 3:26–32+74
Liu Y, Cheung AK, Liu K (2023) Syntactic complexity of interpreted, L2 and L1 speech: a constrained language perspective. Lingua 286:103509
Liang M, Xiong W (2008) Applications of PatCount in foreign language teaching and research. Comput -Assist Foreign Lang Educ (5):71–76
Nini A (2019) The Multi-Dimensional Analysis Tagger. In: Sardinha TB, Pinto MV (eds) Multi-Dimensional Analysis: Research Methods and Current Issues. Bloomsbury Academic, London, p 67–94
Olohan M, Baker M (2000) Reporting that in translated English: evidence for subconscious processes of explicitation. Across Lang Cult (2):141–158
Quirk R, Greenbaum S, Leech G, Svartvik J (1985) A Comprehensive Grammar of the English Language. Longman, London
Ren X (2000) Flexibility in diplomatic interpretation. Chin Translators J 21(5):40–44
Russo M, Bendazzoli C, Sandrelli A (2006) Looking for lexical patterns in a trilingual corpus of source and interpreted speeches: extended analysis of EPIC (European Parliament Interpreting Corpus). Forum 4(1):221–254

Sandrelli A, Claudio B, Mariachiara R (2010) European Parliament Interpreting Corpus (EPIC): methodological issues and preliminary results on lexical patterns in simultaneous interpreting. Int J Transl 22(1-2):165–203
Sandrelli A (2021) Conference interpreting at press conferences. In: Albl-Mikasa M, Tiselius E (eds) The Routledge Handbook of Conference Interpreting. Routledge, London, p 80–89
Schäffner C (2004) Political discourse analysis from the point of view of translation studies. J Lang Polit 3(1):117–150
Schäffner C (2008) 'The prime minister said…': voices in translated political texts. Synaps 22:3–25
Scott M (2016) The WordSmith Tools [version 7.0]. Lexical Analysis Software, Stroud
Sheng D, Li X (2024) A multi-dimensional analysis of interpreted and non-interpreted English discourses at Chinese and American government press conferences. Hum Soc Sci Commun 11(1):472
Shlesinger M (1998) Corpus-based interpreting studies as an offshoot of corpus-based translation studies. Meta 43(4):486–493
Shlesinger M, Ordan N (2012) More spoken or more translated?: Exploring a known unknown of simultaneous interpreting. Target 24(1):43–60
Su Y, Liu K (2022) Orality in translated and non-translated fictional dialogues. In: Moratto R, Li D (eds) Advances in Corpus Applications in Literary and Translation Studies. Routledge, London/New York, p 119-137
Wadensjö C (1998) Interpreting as Interaction. Longman, London/New York
Wang B (2012) A descriptive study of norms in interpreting: Based on the Chinese-English consecutive interpreting corpus of Chinese premier press conferences. Meta 57(1):198–212
Wang Z, Liu K (2024) Linguistic variations between translated and non-translated English chairman's statements in corporate annual reports: a multi-dimensional analysis. SAGE Open 14(2)
Westin I, Geisler C (2002) A multi-dimensional study of diachronic variation in British newspaper editorials. ICAME J 26:133–152
Wu K, Lei VL, Li D (2024) Charting the trajectory of corpus translation studies: Exploring future avenues for advancement. Corpus-based Stud across Humanit (0)
Xiao R, Hu X (2015) Corpus-Based Studies of Translational Chinese in English-Chinese Translation. Springer, Heidelberg
Xiao X (2015) On the oral-literate continuum: a corpus-based study of interpretese. Dissertation. Xiamen University
Xu C (2021) Identification of L2 interpretese: a corpus-based, intermodal, and multidimensional analysis. dissertation. The Hong Kong Polytechnic University
Xu C, Li D (2022) Exploring genre variation and simplification in interpreted language from comparable and intermodal perspectives. Babel 68(5):742–770
Yang M (2012) The principles and tactics on diplomatic translation: a Chinese perspective. Babel 58(1):1–18
Zhang Y, Cheung AK (2022) A corpus-based study of modal verbs in Chinese-English governmental press conference interpreting. Front Psychol 13
Zou B, Wang B (2021) A corpus-driven multi-dimensional analysis of interpreted discourses in political settings. In: Wang C, Zheng B (eds) Empirical studies of translation and interpreting. Routledge, London/New York, p 156–178

Acknowledgements

The research was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU 15602621).

Author contributions

Y Yao: Conceptualization; data curation; formal analysis; investigation; methodology; writing—original draft; writing—review & editing. DC Li: Conceptualization; funding acquisition; methodology; supervision; writing—review & editing. YQ Huang: Data curation; investigation; methodology; Writing—review & editing. ZG Sang: Conceptualization; Writing—review & editing.

Competing interests

DC Li was a member of the Editorial Board of this journal at the time of acceptance for publication. The manuscript was assessed in line with the journal's standard editorial processes, including its policy on competing interests. Y Yao, YQ Huang, and ZG Sang declare no potential conflict of interest.

Ethical approval

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

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-024-03967-6.

Correspondence and requests for materials should be addressed to Dechao Li.