Liu, N. (in press). Register shifts in political conference interpreting: A multidimensional analysis. In J. Pan, S. L. Halverson & J. Munday (Eds.), Translating and interpreting political discourse: New trends and perspectives. Leiden: Brill, Chapter 5.

CHAPTER 5

Register Shifts in Political Conference Interpreting

A Multidimensional Analysis

Nannan Liu

Abstract

Register has been identified as one of the most important factors conditioning utterance perception and comprehension. Research in corpus-based interpreting studies has reported a shift towards oral and formal registers in the interpreted language, but few studies have considered the role of source speech interference and the equalising effect on oral-type sources on the basis of systematic variation of linguistic features. This chapter applies a cross-linguistic multidimensional approach to political conference interpreting based on parallel and comparable corpora and a contrastive register analysis between Chinese and English. Quantitative analyses at the feature, dimension, function, and register levels reveal that interpreters have shifted literate source speech to a more oral, attitudinal, less formal and narrative register, and oral source speech to a less oral, narrative, more formal and attitudinal register. The effects of interpreting, source speech, and target register expectations are teased apart. The most important finding is that regardless of source speech registers, interpreting products tend to show more register similarity to each other than to source speech.

Keywords

multidimensional analysis – conference interpreting – register – Chinese and English.

1 Introduction

Register has been identified as one of the most significant predictors of utterance perception and comprehension in a range of settings (for instance, in multilingual courts, see Berk-Seligson, 1989; in signed language interpreting, see Livingston, Singer and Abrahamson, 1994; in academic writing, see Egbert, 2014). Register (also termed 'speech type' or 'text type' in Shlesinger, 1989; Shlesinger and Ordan, 2012) is defined as a variety of language "associated with a particular situation of use" to serve "particular communication purposes" (Biber and Conrad, 2019, p. 6). Register shifts have been reported to occur between source and target speech owing to interpreter mediation, with the interpreted language gravitating towards an oral, formal, and complex register (Shlesinger, 1989; Hu and Tao, 2012; Wang and Zou, 2018).

Studies of such register shifts rely on analyses of authentic interpreting corpora but have been limited in three ways. First, the role of source speech interference has not been accounted for in a principled manner (except by Wang and Zou, 2018). The lacuna is both theoretical and methodological. The theoretical constraint is that corpus-based interpreting studies (CIS) inherit from corpus-based translation studies a focus on product features that set the interpreting modality apart by comparing interpreted (both consecutive [CI] and simultaneous [SI]) with translated and native varieties (e.g. Shlesinger and Ordan, 2012; Kajzer-Wietrzny, 2021). Nonetheless, owing to inherent differences among languages, some shifts (further discussed in Section 2) can be licensed by target language conventions and cross-linguistic differences (Munday, 2016, pp. 150–151), a comprehensive account of which necessitates a contrastive analysis that in itself constitutes a large project. Methodologically, whilst English corpus analysis instruments have been applied in the extant research, for instance, the Multi-Dimensional (MD) Analysis Tagger (used in Kruger and van Rooy, 2016; see Section 3 herein) and Readability Analyser (used in Wang and Zou, 2018), such tools for other languages are still scarce. The original speech should be accorded primacy in an attempt to tease apart the extent to which the product shows interference from the source speech, is constrained by the interpreting mode, and standardised in accordance with target register expectations (Shlesinger, 1998).

Second, the existing literature has treated a few "teddy-bear" features (De Sutter and Lefer, 2019, p. 19) in isolation, rather than systematically assessing the combined effect of a large, meaningful feature set. This can reduce a situational and contextual activity such as interpreting to mere frequency comparison. In that light, a MD view of language varieties at interpreted

conferences based on co-occurring features that typify native registers is desirable.

Third, a major hurdle in CIS is (and will continue to be) data accessibility and availability. Thus, it is not uncommon to make do with limited materials (for instance, Lv and Liang, 2018), which raises the question of "optimal length", given that corpus-based measures are susceptible to text size (Biber, 1993). As text segmentation has been largely arbitrary, I provide some solutions in the present chapter to the text size problem by developing Chinese corpus sampling principles that are amenable to the current enterprise, i.e. MD analysis of register variation.

In the present chapter, I assess both functional dimensions and feature distributions of language varieties at political interpreting events in search of register shifts. The MD analysis approach was utilised to simultaneously account for the co-occurrence of "many lexical, grammatical, and semantic features" in Mandarin Chinese and English to produce a smaller set of functional scales (Friginal and Biber, 2016, p. 73). Register shifts were directly quantified with hierarchical agglomerative cluster analysis. I make use of both parallel and comparable corpora and large-scale register-controlled corpora to tease out the effect of interpreting. The current chapter addresses two descriptive questions: 1. Does the source and target speech in the parallel interpreting corpus under investigation show register similarities? 2. Do the target speech and comparable corpus demonstrate similarities along the dimensions of register variation identified by Biber (1988)?

Section 2 reviews previous research on register shifts in CIS. Section 3 provides an outline of the cross-linguistic MD analysis methodology and Section 4 describes the register-controlled corpora utilised. The results of the intra- and inter-lingual comparisons are given in Section 5, while Section 6 discusses the implications of these findings.

2 Register shifts in conference interpreting

Inspired by Shlesinger's (1989) work, the extant research in CIS has focused on register shifts along the "oral–literate" continuum and on the functions of a handful of features. There remains room for research into the degree of the "equalising effect" on "oral-type" source speech (Shlesinger and Ordan, 2012, p. 54) and into the combined effect of linguistic patterns.

2.1 *Tendency towards orality*

Shlesinger and her collaborator (1989; 2012; see also the discussion of Shlesinger [1989] in Pym [2007]) describe the equalising effect of SI, glossed as the tendency of literate sources being shifted towards more oral target registers. At the same time, the authors admit that the "effect on typically oral texts has yet to be explored" (Shlesinger and Ordan, 2012, p. 54). The "equalising effect", or in precise terms, "a greater tendency towards orality" in the case of literate-type sources (ibid.), was later proposed as the "levelling universal" in Baker (1996, p. 184). Shlesinger and Ordan (2012) emphasise the operationalisation of this strand of research as comparisons of "features which are known to distinguish between text types [and] genres", i.e. features that mark registers (p. 44). For instance, Shlesinger (1989) isolated five parameters relevant to the oral–literate continuum, including the degree of planning, shared context and knowledge, lexis, degree of involvement, and non-verbal features (Chapter 2). To illustrate, the degree of planning was associated with the use of such devices as nominalisations, relative clauses, and attributive adjectives, which were equated with the literate language (Shlesinger, 1989, p. 16)[1]. Pym (2007) pointed out that in this line of inquiry, register has been viewed as the oral versus literate dichotomy, with the role of underlying variables being postulated and tested for English only (p. 11).

Following Shlesinger (1989), Hong and Wang (2011) collected the raw frequencies of ten features in the source and target of two Chinese–English CI events and concluded that in their data there were two-way shifts: CI renders oral sources more literate and literate sources more oral. Owing to data paucity and the absence of justification for feature selection, this research needs to be replicated on a broader scale with adequate theoretical embedding.

2.2 *Lexical patterns relevant to register shifts*

Apart from the "equalising effect", the strand of CIS relevant to register shifts tends to examine changes in lexical patterns, such as those of cohesive devices (Shlesinger, 1995; Kajzer-Wietrzny, 2021), clauses and adverbs (Hu and Tao, 2012), lexical repetitiveness and density (Kajzer-Wietrzny, 2015), and choices of equivalents (Beaton, 2007; Baumgarten, Meyer and Özçetin, 2008). Studies of political conference interpreting have revealed a gravitation towards informational, formal, and complex language, with a mixed weakening or strengthening pattern found (Beaton, 2007; Hu and Tao, 2012). Nonetheless,

[1] A caveat in understanding the proposal of "greater tendency towards orality" in Shlesinger's subsequent work (summarised in Shlesinger and Ordan, 2012) is that such work is oriented towards an identification of features that distinguish SI from translation, and thus the "orality" proposal therein should not be understood as vis-à-vis source speech.

those arguments are made on the basis of frequency comparisons of individual linguistic devices rather than their aggregated functions. A more reliable method is to identify a meaningful set of linguistic features and their role in relation to entire register systems of languages prior to a discussion of register shifts (as in Ji, 2017).

In CIS, linguistic features of the source and target have been examined sporadically. Shlesinger's (1995) investigation of cohesion shifts reveals interpreters' avoidance of "implicit" ellipsis in both English–Hebrew and Hebrew–English parallel corpora, which lends support to the argument that SI "may exert a stronger effect" than target register expectations with regard to explicitation (p. 210). In the Chinese–English combination, Hu and Tao (2012) show political interpreters' penchant for connecting independent clauses in source Chinese using sentence relatives with *which* (see "sentence relatives" in Tables C-3 and D-3 herein), resulting, as they argue, in more complex, formal interpretations than are found in the corresponding sources (pp. 745–746). Their examination of degree adverbs (e.g. *very*, *absolutely*) reveals a general weakening effect owing to mediation.

Research that makes intra-lingual comparisons (e.g. Li and Wang, 2012; Kajzer-Wietrzny, 2015, 2021) is usually interested in examining simplification or explicitation in native and interpreted English. Li and Wang (2012) reported Hong Kong government interpreters' preference for formal and neutral presentation of information in SI, instantiated by their "excessive use" of nominalisations relative to native English speakers (p. 414). Nonetheless, such "excess", as rightly pointed out by Baumgarten, Meyer and Özçeti (2008), can be licensed by the source speech, target register expectations, and communicative conventions (pp. 184–197). Conflating register and the mode of delivery, Kajzer-Wietrzny (2015, 2021) found spoken, impromptu Spanish sources to result in significantly less sophisticated (measured with the proportion of 100 high frequency English words) interpreted English than read-out sources in the European Parliament (2015, p. 249). She recommended a comprehensive source–target comparison to consider language-specific factors and an estimate of "how strong the 'levelling effect'" on spoken and read-out sources is, respectively (2015, pp. 245–252). Wang and Zou (2018) investigated language specificity by integrating parallel and comparable corpora and found that interpreters generate more difficult English output than US Department of State daily press briefings do, and such a pattern was necessitated by cross-linguistic differences between Chinese front-loaded modifying structures marked by the genitive DE and back-loaded structures in English (pp. 71–78).

The extant research also underlines interpreters' choices of equivalents that may have effectuated potential register and perception changes. Beaton (2007) proposed that European Parliament interpreters' intensified use of "European Union" relative to the original speech may have strengthened "the dominant institutional presence". Such choices are highly idiosyncratic and contextualised, but they point to textual properties as manifestations of underlying communication purposes, an association that underpins register research (Biber and Conrad, 2019).

In light of the research reviewed, it is desirable that an analysis of interpreting-induced register shifts should be embedded in both parallel and comparable corpora, include typically oral source speech, and be built upon robust associations of linguistic features with their functions in relation to a language's register system. In this chapter, such a goal is pursued through a cross-linguistic MD approach.

3 Towards cross-linguistic MD analyses of register shifts

The current discussion of register shifts is based on contrastive MD analyses of Chinese and English register systems, or the range of activities that speakers of a language engage in (Biber, 1995, p. 5). MD analysis is a multi-feature, multi-dimensional approach that reduces a large linguistic feature set to a few dimensions of variation (Cf. "parameters" in Shlesinger, 1989). It accomplishes such a reduction through factor analysis, a statistical method that summarises patterns of correlations among individual features by simultaneously grouping co-occurring features (positive features) and features that occur in a complementary pattern (negative features) in the texts (Friginal and Biber, 2016, p. 74). For example, proper nouns, person names, and attributive adjectives typify the press reportage register, wherein simile and onomatopoeic words tend to be absent. Such a grouping of features is called a dimension in MD analysis. Based on the feature structure and distribution of input registers along each dimension, the analyst makes an interpretation about the functions underlying each dimension, which can be "ideational, textual, personal, interpersonal, contextual, processing, [and] aesthetic" (Biber, 1988, p. 34).

The comparison of register systems requires the use of large, register-controlled corpora; the current analyses rely on 15 analogous registers in two Brown family corpora, each containing "500 texts of about 2,000 words" (Biber, 1995, p. 33), as well as other corpora used in Biber (1988). Information about the Chinese and English corpora is given in Table 1.

TABLE 1: Corpora used in Chinese and English (Biber, 1988) MD analyses.

Brown family categories		**Chinese**		**English**		
		upsampled Texts of Recent Chinese 2014 corpus		Lancaster–Oslo/Bergen (LOB) corpus	London–Lund corpus	
Category	Register	Texts	Words	Texts	Register	Texts
A	reportage	44	106,191	44	face-to-face conversation	44
B	editorial	32	78,880	27	telephone conversation	27
C	review	33	81,551	17	public conversations, debates, and interviews	22
D	religion	33	82,768	17	broadcast	18
E	skills/hobbies	36	89,909	14	spontaneous speech	16
F	popular lore	48	128,374	14	planned speech	14
G	biography	75	180,302	14	personal letters*	6
H	official	32	77,189	14	professional letters*	10
J	academic	32	191,547	80		
M	science fiction	32	79,199	6		
P	romance	32	80,516	13		
R	humour	32	80,516	9		
K	general fiction	32	78,572	29		
L	mystery fiction	32	78,698	13		
N	adventure fiction	32	77,182	13		
Total		605	1,489,124	324	960,000 words	157

Note. *collected by Biber and a colleague (1988, p. 66)

As seen in Table 1, a major difference between the Chinese and English corpora is the lack of oral registers in the former. This is not ideal for the current research but no native spoken Chinese corpora are publicly available at the time of this writing. Nonetheless, both corpora include the written–oral register humour (Shlesinger, 1989, pp. 12–13), and as will be explained in Section 4.1.1, one of the Chinese sources under investigation, namely Premier Wen's speech, was argued to epitomise "formal, written, ... [and] planned language" (Chen, 2007, p. 60). Thus a contrastive analysis based on such reference corpora is still valuable. Chinese character sequences were segmented into "words"[2] using the lexical analyser ICTCLAS (Zhang *et al.*, 2003).

MD analysis of the Chinese corpus and contrast with Biber's (1988) English analysis was conducted to discover related dimensions of variation that occur in both languages. Such an approach is inspired by Biber's (1995, 2014) review of dimensions uncovered by MD analyses of nine languages, which reveals two "universal" dimensions: "(1) a fundamental opposition between clausal/'oral' discourse versus phrasal/'literate' discourse, and (2) the opposition between the narrative versus non-narrative discourse" (2014, p. 7). Such dimensions are "universal" because of their similarities in three aspects:

- The co-occurring linguistic features that define the dimensions of variation in each language;
- The functional correlates of the dimensions across languages;
- The linguistic/functional relations among analogous registers across languages (Biber, 2014, p. 21).

Such similarities shall be taken as principles to gauge candidates for related dimensions of register variation in Chinese and English, which are termed "functional scales" hereafter. The MD analysis of Chinese project (henceforth: "MulDi Chinese") is completed by the author and resulted in a Python library (Liu, 2021) based on ICTCLAS. MulDi Chinese follows Biber's (1988) methodology and gathers an exhaustive list of 108 features in the initial stage. Following feature screening and factor analysis, it generates a set of 60 features and four dimensions of variation to describe the 15 registers in Table 1. ANOVA tests revealed all dimensions to be statistically significant discriminators among registers (see post-hoc pairwise comparison results in Liu, 2021). Biber's

[2] A Chinese word can comprise multiple characters with no spaces appearing between them, and ICTCLAS breaks character sequences into individual "words". For instance, the character sequence *nǚshìmen* 女士们 [ladies] is segmented by ICTCLAS into two "words", i.e. 女士 们 [lady plural marker].

(1988) English MD analysis was replicated by Nini (2019) in the Multi-Dimensional Analysis Tagger ("the English tagger" hereafter) based on the Stanford Tagger (Toutanova *et al.*, 2003), with reliability tests indicating the tagger's overall accuracy (Nini, 2019). The English tagger generates standardised frequencies of 67 features and scores of input texts along 6 dimensions. Owing to slight differences in rules employed by Biber (1988) and the Stanford tagger, dimension statistics for the LOB corpus were obtained using the English tagger, instead of using those in Biber's (1988) original analyses (pp. 122–124), which ensures the consistency of comparisons across English registers. In order to situate the Chinese and English registers in comparable systems, the Brown family corpora only (see Table 1) were utilised in the analyses[3]. MulDi Chinese and the English tagger constitute the analytical instruments utilised in this research.

4 Methods

In the present chapter, MD analyses of parallel interpreting and comparable corpora are carried out, enabling four sets of intra- and inter-lingual comparisons, namely (a) within original Chinese (OC) speech, between speakers; (b) between interpreted English (IE) and native English (NE); (c) within IE speech, across interpreters; and (d) between OC and IE along functional scales.

4.1 Corpora

4.1.1 Interpreting parallel corpus

This study utilises a self-built parallel corpus of interpreted premier press conferences that correspond to ten years of Chinese premier press conferences, held after the National People's Congress and Chinese People's Political Consultative Conference sessions. The two Chinese premiers involved are Premiers Wen Jiabao (first-term, 2003–2007) and Li Keqiang (first-term, 2013–2017). During the live stream conferences, the premiers answered questions from Chinese and foreign journalists about government policies and international relations. OC data include the premiers' opening and closing (when available) monologues and responses to reporter questions. The questions were pre-submitted and reporters pre-vetted; thus, as concluded by Yi and Chang (2012), the interpreting events are a political ritual rather than a

[3] The London–Lund corpus was encoded with markup that cannot be read by the English tagger and is removed from the ensuing analyses.

give-and-take exchange (pp. 717–720). Although it was not possible to interview the premiers, both video observation and previous research suggest the highly scripted nature of their utterances (Chen, 2007; Yi and Chang, 2012), especially in the case of Premier Wen Jiabao (see Section 5.1), who had limited interaction with reporters. IE data are consecutive interpretations made by six different staff interpreters of the Chinese Ministry of Foreign Affairs, who worked into their B language. Four different interpreters worked in the sessions of 2003–2007 and the rest worked in 2013–2017. Because my focus herein is the degree of register similarities among source, target, and comparable speech, not interpreter heterogeneity (but see the standard deviation and dispersion values reported in Section 5), the four interpreters who worked for Premier Wen and their interpretations were coded as "IEW" whilst the two for Premier Li and their interpretations as "IEL".

4.1.2 Comparable corpus

A comparable United States State of the Union (SOTU) corpus was created to model political speech in NE. The SOTU address is an annual scripted speech delivered by US presidents and broadcast live as a report and appeal to American Congress and people (Kreiser and Greene, 2022). Both SOTU and Chinese premier speech epitomise ritualistic political performance and serve such functions as policy debriefing, support rallying, and image management (Chen, 2007; Kreiser and Greene, 2022). Such a register is termed "oral performance of scripted political presentation" in this study. The comparable corpus includes transcripts of SOTU addresses in the same period as the parallel corpus, excluding 2017, i.e. addresses by two presidents: George W. Bush (2003–2007) and Barack H. Obama (2013–2016). These serve as baselines to tease apart target register convention– and interpreting-induced effects.

4.2 Corpus sampling

OC, IE, and NE transcripts in the parallel and comparable corpora were divided into chunks following a set of principles for achieving representativeness in corpus sampling. With respect to English, Biber (1993) compared the distributions of ten features across 200-word samples in 110 1,000-word text samples taken from the corpus he used in the MD analysis (1988; see Table 1). He found that common features such as nouns and prepositions follow a linear and stable distribution, whereas rare features, for instance, conditional subordination and relative clauses, demonstrate a curvilinear distribution, that is, they occur with a sharp increase in frequency in the first 200 words but a gradual decay thereafter. Biber concluded that for an investigation of common features, a 200-word segment would suffice, and because "relatively

few additional" occurrences of most rare features were "added after 600 words" (1993, p. 251), a 600-word segment is adequate for drawing frequentist inferences. Biber's (1993) distributional analysis was confirmed by Egbert (2014), who compared mean dimension scores obtained from academic English samples of 500–600 and 5,000 tokens and found almost identical patterns between them (pp. 34–36). To extend MD analysis to the interpreting scenario, I decided on a conservative length of 1,000 tokens for each sample in the NE corpus[4].

With respect to Mandarin, such guidelines of "optimal length" seem absent from the literature. Following Biber (1993), I compared the distributions of 20 features across 200-word segments of OC texts. It was found that such features as first-person pronouns and nouns were common and had a linear distribution, whereas third-person pronouns and amplifiers (e.g. *fēicháng* 'very' and *jídù* 'extremely') showed a curvilinear distribution, with few being added after 600 or 800 words (see Appendix A for an illustration). Further, I ran Welch's *t*-tests and Mann–Whitney *U* tests, as appropriate, to compare dimension scores obtained from 800- ($N = 57$) and 2,000-word segments ($N = 24$) of OC data and found no significant differences between the two groups. On this basis, OC texts were segmented into 800-word samples, and care was taken not to mix Premiers Wen and Li's utterances. The process resulted in 57 files (mean length = 809.26 words, $SD = 57.36$). IE files contain the corresponding interpretations, which, coincidentally, have a comparable token length distribution ($M = 1065.46$, $SD = 136.06$) to the NE files ($M = 1002.64$, $SD = 79.65$). An overview of the triangular set-up of this chapter's data is provided in Table 2.

TABLE 2: Corpora composition.

Corpora	Variety	Speaker(s)	Texts	Words/tokens
Parallel	OC	Wen	24	19,233
		Li	33	26,895
	IE	IEW	24	27,395
		IEL	33	33,336
Comparable	NE	Bush	30	30,009
		Obama	31	31,143
		Total	175	168,011

[4] Welch's *t*-tests and Mann–Whitney *U* tests revealed no significant differences between dimension scores obtained from 1,000- and 2,000-token samples (text length used in Brown family corpora and English MD analyses in Xiao, 2009; Kruger and van Rooy, 2016; $N = 31$ for both IE and NE).

Dimension and feature distribution differences between the two premiers and the two presidents were found (see Sections 5.1, 5.3, and 5.4) via statistical analyses, but as this research is not aimed at analysing the presidents' speech styles (statistical results are shown, nonetheless, in Figures 1–4), the focus is on between-premier and within-IE differences.

4.3 Data analysis

Dimension scores and standardised feature frequencies for OC were generated utilising MulDi Chinese version 0.3.2, and for IE and NE using the English tagger version 1.3.2. The Shapiro–Wilk test was used to assess the normality of response variables "dimension scores" and "feature frequencies". With respect to variables that were not normally distributed, Mann–Whitney *U* tests were used to test for the significant effects of "speaker/interpreter group" (i.e. Premiers Wen and Li; IEW and IEL; Presidents Bush and Obama) on response variables. Regarding response variables that were normally distributed, Welch's *t*-tests were used because the numbers of files are unequal (see Table 2). When non-normal distribution was found, the median was reported as a measure of central tendency and the interquartile range (IQR), i.e. the difference between the 75th and 25th percentile of data, was reported as a measure of dispersion. The conventional 0.05-level was used as the cut-off point for statistical significance. All statistical tests, including hierarchical agglomerative cluster analysis in Section 5.4, were conducted using Python version 3.9.0 (Rossum, 1995). Owing to space constraints, statistical results concerning features are relegated to Appendices B, C and D, grouped by their primary loading dimension. All references to tables denoted with these letters are found in the appendices.

5 Findings

5.1 Dimensions of register variation in OC

MulDi Chinese enables a description of the sources in the events under investigation, which provides an idea of what equivalent IE utterances should look like. The analyses in this section focus on dimensions where statistically significant differences between the two OC speakers, i.e. Premiers Wen Jiabao and Li Keqiang, were found. Table 3 shows mean and standard deviation of dimension scores for each premier's speech and the results of *t*-tests. The full feature structure of each dimension is given in Appendix B and Figures 1a, 2a, and 3a show register distributions along the dimensions.

TABLE 3: Chinese dimension scores by premier.

Chinese dimension	**Premier Wen** *M* [*SD*]	**Premier Li** *M* [*SD*]	**Results**
Dimension 1: Orality/literacy	−2.32 [6.93]	1.68 [5.04]	$t(39.85) = -2.40$, $p = 0.021$, $d = 0.68$, 95% CI [−7.36, −0.64]
Dimension 2: Narration	−1.76 [3.15]	1.28 [4.19]	$t(54.92) = -3.12$, $p = 0.003$, $d = 0.80$, 95% CI [−4.98, −1.09]
Dimension 3: Explicit evaluation	−4.71 [6.43]	3.42 [3.61]	$t(33.45) = -5.50$, $p < 0.001$, $d = 1.63$, 95% CI [−11.09, −5.17]

Chinese Dimension 1 represents the "oral/literate opposition", a "universal" dimension of register variation (see Section 3). This dimension is interpreted as orality versus literacy in that the written–oral register humour received the highest positive scores (see Figure 1a). A Welch's *t*-test revealed a significant difference between Premiers Wen and Li and an association of Premier Li's speech with orality and Premier Wen's with literacy, as Wen's utterances are marked by significantly higher frequencies of literacy features such as longer words, nouns, nominalisations, and phrasal coordination (see Table B-1). Premier Wen's register is illustrated by (1), which is reproduced from his speech in 2005. Literacy features are in bold and nominalisations[5] in italics.

(1)	我	对	中国	**农村**	的	***改革***	**和**	***发展***	是
	I	to	China	rural area	DE	reform	and	development	am
	I	do	have	long-term	considerations	about	reform	and	development.

[5] Note that apart from noun–verbs and noun–adjectives, Mandarin nominalisation also involves placing the genitive particle DE after a verb (Li and Thompson, 1989, pp. 575–576).

(1)	有	*长远*	*考虑*	的，	**可以**	**划分**	为	两	个	**阶段**。
	have	long-term	consideration	DE,	can	classify	as	two	CLF	stages
	It	can	be	classified	into	two	stages.	The	first	is

(1)	第一	个	**阶段**	就	是	**实行**	了	**家庭**
	First	CLF	stage	ADV	is	implement	PFV	family
	to	implement	the	fundamental	economic	institution	of	household

(1)	*经营*	*的*	**基本**	**经济**	**制度，**	给	农民
	management	DE	fundamental	economic	institution	give	farmers
	contract	responsibility	system.	Farmers	were	granted	the

(1)	**以**	*生产*	*经营*	*的*	**自主权**	……
	PREPOSITION	production	management	DE	autonomy	
	autonomy	to	carry	out	production	...

(1) is highly nominal, laced with nine nominalisations as well as a phrasal connector *hé* 'and'. Conversely, Premier Li was significantly more focused on colloquial expression than his predecessor.

Chinese Dimension 2 marks the "universal" narration, comprising perfect and imperfect aspect, and descriptive and onomatopoeic words in Mandarin. There is a significant effect for speaker, as Premier Wen preferred non-narrative abstract nouns (see Table B-1) whereas Premier Li framed concrete actions and vivid scenes with significantly higher frequencies of intransitive verbs and simile (Table B-2).

Chinese Dimension 3 is assessed to reflect explicit attitudes by text producers because it consists of positive features such as evaluative words (amplifiers, hedges, and downtoners; Biber [1988, p. 106]), focusing devices (*shì* 'be' and *yǒu* 'existential *there*'), and cohesive devices, e.g. conditional and concessive conjuncts. There is a highly significant effect for speaker, as Premier Wen's utterances showed a relative lack of quantifiers (see Table B-2), indicative of an absolute and vague register (Biber, 1988, p. 108). In contrast, Premier Li

expressed significantly more personal stance, as is evident in (2), an utterance also about rural reform reproduced from his speech in 2014. Positive features on Chinese Dimensions 1, 2, and 3 are in bold.

(2)	**后来**	**搞**	承包制，	放开	搞活，	农民	**自己**	**决定**
	time	DO	contract system	liberate	relax	farmers	self	decide
	Later	came	contract system,	relax	and	liberate,	farmers	decided for

(2)	**干**	**什么、**	**怎么**	**种，**	**几**	**年**	**时间**	温饱
	DO	what	how	plant	several	CLF	time	food and clothing
	themselves	what and	how to	plant,	just in	a few	years	food and clothing

(2)	**问题**	**解决**	**了。**
	problem	solved	PRF
	was	no longer	a problem.

In sum, significant differences were identified between Premiers Wen and Li in their OC utterances with respect to the degree of orality/literacy, narration, and explicit evaluation. Premier Wen's utterances were assessed to be emblematic of literate, non-narrative, and absolute discourse whereas Premier Li tended to focus on oral, narrative, and evaluative expressions.

5.2 Dimensions of register variation in IE versus NE

In the following section, I compared IE and NE to single out the effect of interpreting on political performances. Dimensions where significant differences between IE and NE scores were found are shown in Table 4.

TABLE 4: English dimension scores by IE and NE.

English dimension	IE	NE	Results
Dimension 1: Involved versus informational production	*M* [*SD*]	*M* [*SD*]	$t(108.19) = -4.5$, $p < 0.001$, $d = 0.82$, 95% CI [−5.66, −2.2]
	−7.66 [3.89]	−3.73 [5.5]	
Dimension 3: Explicit versus situation-dependent reference	*M* [*SD*]	*M* [*SD*]	$t(115.02) = 4.88$, $p < 0.001$, $d = 0.9$, 95% CI [1.45, 3.44]
	7.38 [2.5]	4.93 [2.94]	

Dimension 5: Abstract versus non-abstract information	*Mdn* [*IQR*]	*Mdn* [*IQR*]	$U = 2343$, $p = 0.001$, $r_{rb} = -0.35$
	0.44 [2.35]	−0.88 [2.34]	

English Dimension 1 marks the "universal" oral/literate opposition, indexing "interactional, stance-focused, and generalised content" versus "high informational density and precise word choice" (Biber, 2014, p. 13). IE was significantly more literate and informational than NE, as NE demonstrated a penchant for nine interactive and informal features (see Table C-2).

English Dimension 3 distinguishes between "highly explicit, context-independent reference and nonspecific, situation-dependent reference" (Biber, 1988, p. 110). IE is significantly more "explicit" than NE. Tables C-1 and C-3 reveal IE's preference for "explicitness below the clausal level" (Kruger and van Rooy, 2016, p. 41), owing to significant over-representations of nominalisations (consistent with Li and Wang, 2012) and phrasal coordination (*and* as a phrasal coordinator), while the occurrences of WH clausal features were close to zero in IE (see Tables C-2 and C-3).

English Dimension 5 gauges an abstract and impersonal focus (Biber, 1988, 2014), and IE was shown to be significantly more "abstract" than NE. Table C-1 indicates the abstractness of IE being supported by significantly higher frequencies of conjuncts, predicative adjectives, and agentless passives[6].

To summarise, with the effect of register conventions controlled for, this section foregrounds characteristics of the target speech that are attributed to the source speech or interpreting modality: IE was significantly more informational, explicit, and abstract than NE. Because no significant differences were found between IE and NE on Biber's (1988) Dimensions 2, 4 and 6, the effect of register expectations seems restricted to those dimensions, along which both NE and IE were found to be non-narrative, highly persuasive, and showing "real-time" constraints (pp. 113–114).

5.3 Register shifts on three functional scales

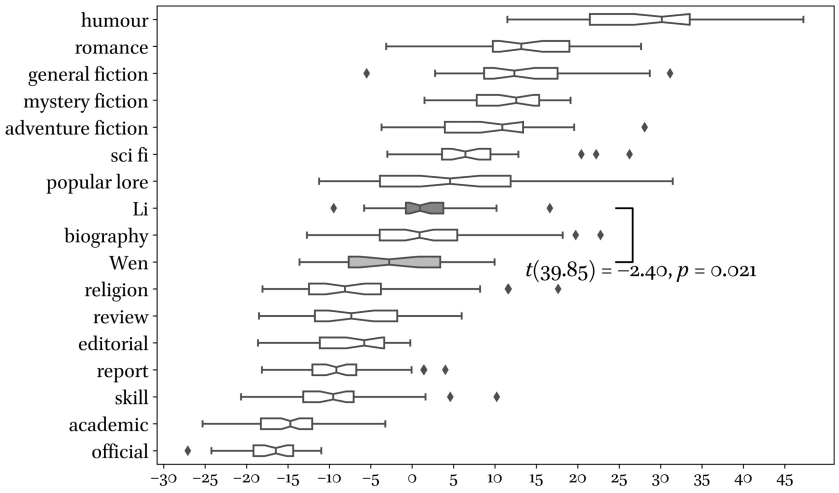
In accordance with the principles set out by Biber (1995, 2014; see Section 3), individual features, dimension functions, and register relations in Chinese and English were scrutinised to uncover related functional scales. I shall address source–target shifts along such scales and the degree to which the shifts, if any, are ascribed to the divergent styles between Premiers Wen and Li (Section 5.1) or to interpreters and the interpreting modality.

[6] The current version of MulDi Chinese does not incorporate dependency parsing, and thus a direct comparison between Chinese and English agentless passives is not possible.

5.3.1 Oral and literate scale

Both Chinese and English Dimension 1 represent the oral/literate opposition and comprise positive features such as verbs, pronouns, discourse markers/particles, WH-words/clauses, and questions, and negative features such as nouns, average word length, and prepositions (see Appendices B and D). The 15 analogous registers have similar score distributions, plotted in Figure 1 in descending order of their mean scores.

a. Chinese Dimension 1



b. English Dimension 1

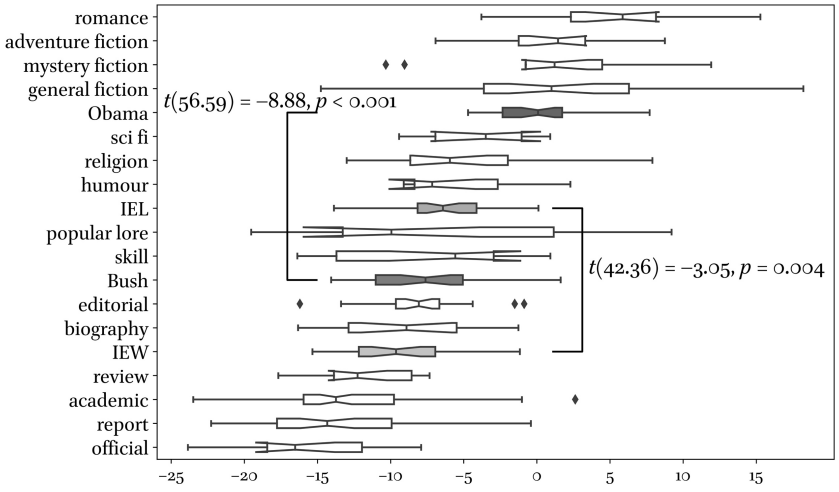


FIGURE 1: Chinese and English Dimension 1 score distributions.

As seen from Figure 1, both Chinese and English popular lore were measured as being more involved than biography, and editorials were more oral than academic writing (although see Ji, 2017, pp. 86–87 for a discussion of potential inherent or sampling differences between Chinese and English humour, skills/hobbies, and biography registers in Brown family corpora). All fiction writing is highly personal, whereas official and report registers receive the lowest scores herein. Thus, there is substantial evidence of similarities in the Dimension 1 of both languages, which reveals an "oral and literate scale".

Figure 1b shows a significant difference between IEW (M = –9.43, SD = 4.08) and IEL's (M = –6.37, SD = 3.23) position on this scale, $t(42.36) = -3.05$, $p = 0.004$, $d = 0.85$, 95% CI [–5.09, –1.04], with IEL being more oral and interactional than IEW. With respect to the literate-type source, as seen from Figure 1, Premier Wen's utterances were assessed along Dimension 1 to be quantitatively akin to religion (M = –6.15, SD = 9.35), whilst IEW was on a par with the less oral English biography (M = –8.82, SD = 5.04). This is a surprising finding given the "equalising effect", which posits that interpreting renders literate speech more oral. Nonetheless, this result should be understood with the foregoing caveat on the inherent differences between some Chinese and English registers. An examination of shifts of co-occurring variables loading on the dimensions considered here reveals evidence for the "equalising effect". Table B-1 shows Premier Wen's statistically significant preference for the orality feature indefinite pronouns (e.g. *shuí* 'who' and *yǒurén* 'someone') compared with Premier Li, but such pronouns were almost completely ignored in both IEW and IEL with no significant differences between them (see Table D-3). At the same time, Premier Wen's penchant for phrasal coordination, longer words, nouns, and nominalisations may well spill over to his interpreters, with IEW showing significantly higher frequencies of phrasal coordination, longer words, and nouns (but not nominalisations) than IEL (see Tables D-1 and D-3). The strength of the "equalising effect" can be exemplified by a comparison of (1) in Section 5.1 and its interpretation in (3). Orality features on English Dimension 1 are in bold and nominalisations in italics.

> (3) **I do have** a long-term plan for rural reform and *development*. **It has** two phases. In the first phase, **we** introduced the basic economic system of family respon contract *responsibility* system, uh **which** in essence was to give greater autonomy to the farmers in *production* and *management*. (IEW 2005)

The original segment (1) features nine nominalisations and one phrasal coordination, while four nominalised words and two phrasal coordination

were utilised in (3), with the additional *and* being used to connect original noun sequence *shēngchǎn* 'production' *jīngyíng* 'management'. Still, fewer nominalisations were used in the interpretations than original speech (Cf. Li and Wang, 2012). As nominalisations characterise formal registers (Kruger and van Rooy, 2016, p. 41), such a pattern suggests a shift towards informality and orality in the case of literate sources. Granted, this result should be understood in light of the encoding differences between Chinese and English nominalisations: verb plus DE, which occurred twice in (1), constitutes nominalisation (see footnote 5) in Chinese, but such a pattern does not exist in English. Taken together, the "greater tendency towards orality" in the interpretation of literate speech is realised by downplaying nominalisations.

With respect to the oral-type source, Premier Li's speech was more oral than biography (M = 1.61, SD = 7.7) whereas IEL resembles popular lore (M = –6.73, SD = 9.93). As the analogous register of English popular lore, i.e. Chinese popular lore, is similar to Premier Li's speech (see Figure 1a), orality–literacy shifts from Premier Li's speech to IEL, if any, may be small. Comparisons by feature show that apart from avoiding indefinite pronouns, IEL also relatively neglected original second-person pronouns, as Table B-2 shows Premier Li's preference for them relative to Premier Wen, a difference that was significantly narrowed down by interpreters (Table D-3). The case of nominalisations in Premier Li's speech is interesting because Premier Wen used more nominalisations than Premier Li, but no significant differences were found between IEW and IEL concerning the use of this device. As nominalisation serves as a grammatical metaphor "that reduces a process into a noun" (Xiao, 2009, p. 429), it can be utilised to pack the significantly more abundant verbs in Premier Li's speech (see Table B-2). This can be illustrated by the rendition of two action verbs *fàngkāi* 'decentralise or deregulate' and *gǎohuó* 'to make (rural economy) dynamic' in (2) as "past restrictions were lifted" in IEL 2014, with *restrictions* being a nominalised verb. Granted, a possible equivalent to *fàngkāi gǎohuó*, inspired by (4), can be "[later came the contract system,] torn down regulations, made villages more dynamic than before," but such colloquial and loose structures were disfavoured by interpreters (Kajzer-Wietrzny, 2015, p. 247).

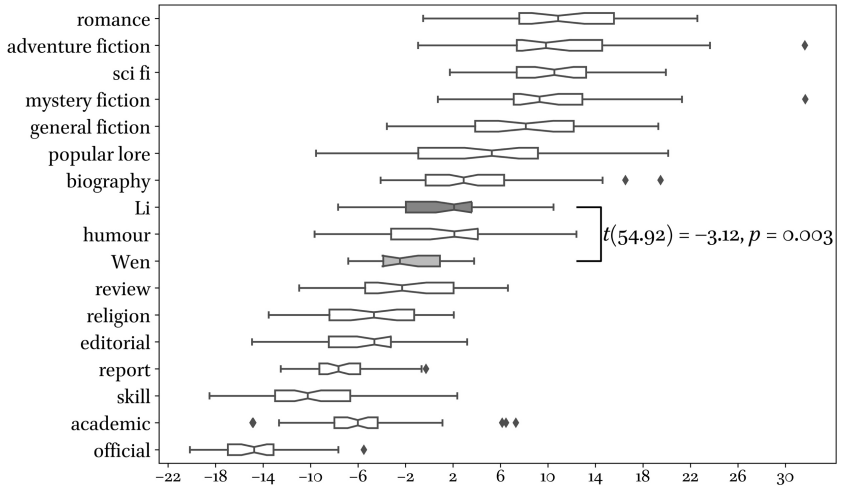
> (4) So let's pass an agenda that helps them [start-ups and small businesses] succeed. Tear down regulations that prevent aspiring entrepreneurs from getting the financing to grow. Expand tax relief to small businesses that are raising wages and creating good jobs. (Obama 2016)

In sum, interpreters may have shifted oral-type utterances (Premier Li's speech) to be more formal and literate by avoiding indefinite and second-person pronouns and packing verbs with nominalisations.

5.3.2 Narrative scale

Both Chinese and English Dimension 2 measure the "universal" narration and are supported by such co-occurring features as perfect aspect markers/verbs and third-person pronouns. The reference registers show analogous dimension score distribution as shown in Figure 2, with fiction receiving the highest scores and non-fiction the lowest scores. Nonetheless, it is evident from Appendices B and D that the two languages employ distinct features for narrative purposes, and so the comparisons herein are rather preliminary.

a. Chinese Dimension 2



b. English Dimension 2

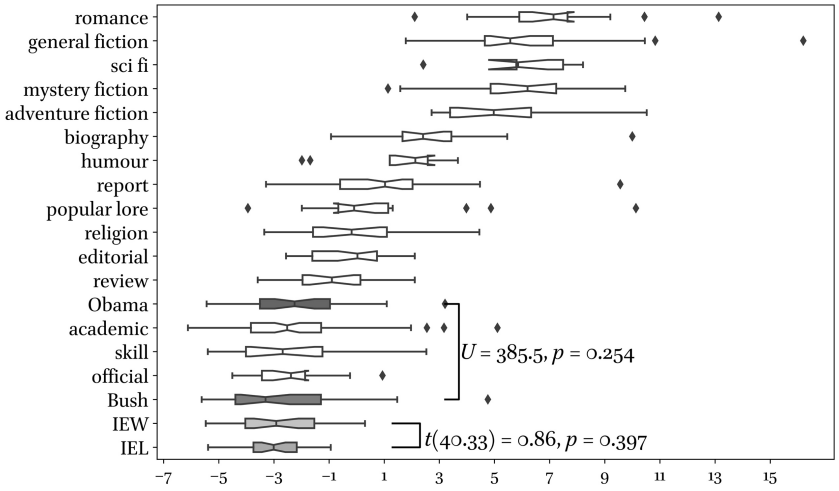


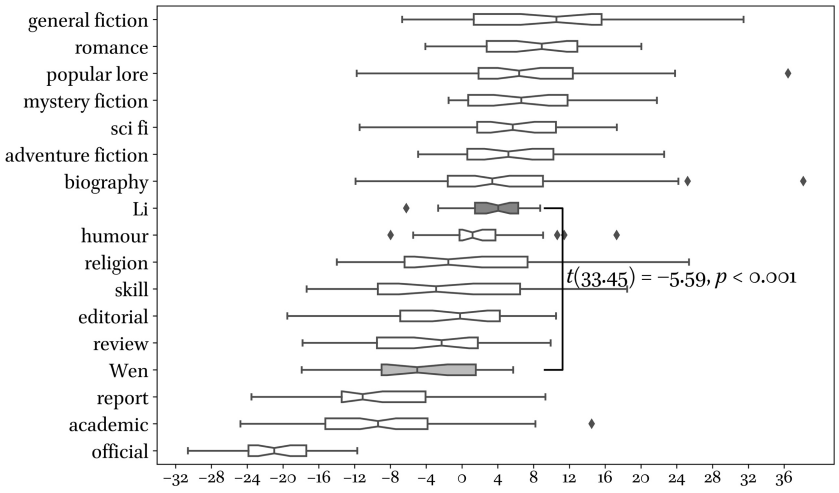
FIGURE 2: Chinese and English Dimensions 2 score distributions.

As seen from Figure 2a, Premier Wen's utterances were moderately narrative as review (M = −2.13, SD = 4.92) whereas his successor Premier Li was more narrative than a humorous writer (M = 0.80, SD = 5.64). In contrast, both IEW and IEL were extremely non-narrative with no significant differences between them, $t(40.33) = 0.86$, $p = 0.397$, $d = 0.24$, 95% CI = [−0.45, 1.1]. Scrutiny of the local features supports the equalising effect of interpreting: Premier Wen favoured perfect aspect markers compared with Premier Li (see Table B-1), but no significant differences between IEW and IWL on the use of perfect aspect verbs were found (Table D-3). The significantly more abundant intransitive verbs in Premier Li's oral-type speech might have been nominalised (Section 5.3.1), resulting in a highly non-narrative interpretation.

5.3.3 Attitudinal scale

An "attitudinal scale" can be generalised from English Dimensions 1 and 4 and Chinese Dimension 3, as they comprise stance-focused devices including amplifiers, hedges, modal verbs, private verbs, conditional conjuncts, and *be*. Register distributions along Chinese Dimension 3 and English Dimension 4 are plotted in Figure 3, showing higher scores of fiction and lower scores of official and report writing.

a. Chinese Dimension 3



b. English Dimension 4

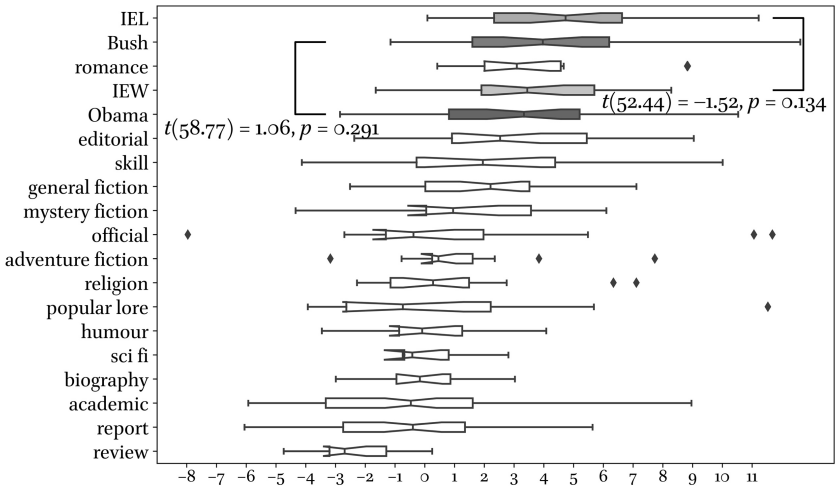


FIGURE 3: Chinese Dimension 3 and English Dimensions 4 score distributions.

A comparison between Figures 3a and 3b shows that the original attitudinal conveyance may have been levelled up as IEW and IEL were among the most stance-focused in the registers examined. The equalising effect also plays out on oral and literate source types to different extents, as significantly higher frequencies of amplifiers, hedges, modal verbs, and *shì* 'be' were found in Premier Li's speech than Premier Wen's speech (see Table B-2), but IEW used more amplifiers than IEL (Table D-1). No significant differences for *be*, hedges,

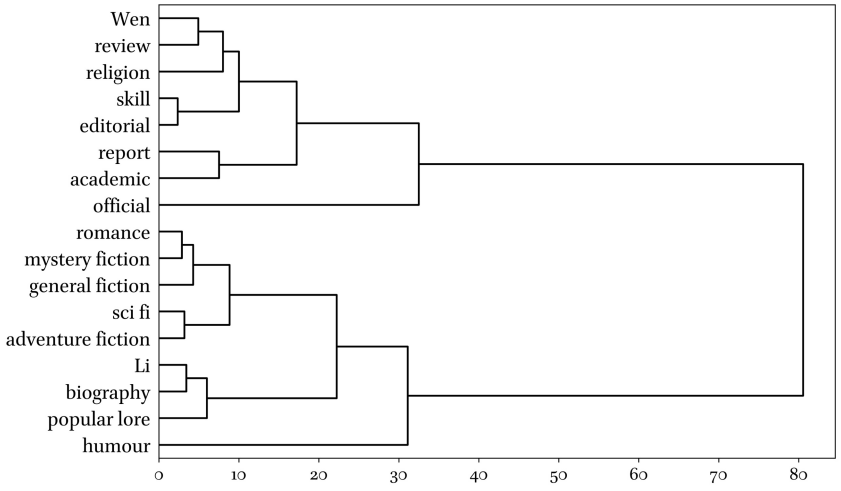
and possibility and necessity modals were found in the two interpretations, with the discrepancy in the use of prediction modals between IEW and IEL being less marked than that in the modals of the original (Tables D-2 and D-3). More interestingly, the occurrences of hedges were close to zero in both interpretations and NE (see Tables C-3 and D-3), suggesting the avoidance of any "fuzziness" (Biber, 1988, p. 240).

5.4 Quantifying register shifts

Whilst earlier subsections singled out related functional scales in English and Chinese, in this section I shall consider IE, NE, and OC's full dimensional ranges and directly quantify the extent of register shifts in the interpreting events.

The statistical method used is hierarchical agglomerative cluster analysis (see Gries, 2008; Ji, 2017), a bottom-up approach that treats each variety/register as a single cluster and then "merges the closest pairs of clusters until only one cluster remains" (Raschka and Mirjalili, 2019, p. 367). The distance between each pair of registers/varieties was measured with Euclidean distances, one of the most straightforward ways to calculate distances between data points in a multidimensional space. To merge small clusters with the shortest Euclidean distances into larger clusters, Ward's method was used, which merges two clusters that lead to the minimum increase of the total within cluster sum of squared distances (Raschka and Mirjalili, 2019, p. 368) and produces compact clusters. The results of such a process are visualised in the dendrograms of Figure 4.

a. Chinese



b. English

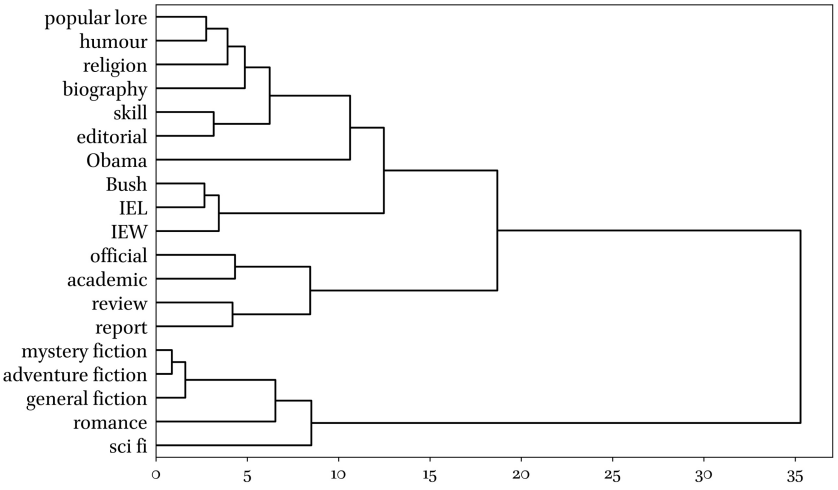


FIGURE 4: Hierarchical agglomerative clustering of OC, IE, NE, and reference registers.

Figure 4 shows the gradual merging of small clusters of registers/varieties that are closest together, such as skills/hobbies and editorials in both Chinese and English, into one large cluster displaying the register system. The Euclidean distance between registers/varieties is represented by the height of the "stem" that connects two clusters: the larger the height, the more dissimilar the two registers/varieties. The most visible distinction in Figure 4 lies in non-fiction versus fiction clusters as 4a shows Chinese non-fiction cluster to contain Premier Wen's speech, review, religion, skills/hobbies, editorial, report, academic, and official writing, whereas Premier Li's utterances belong to the fiction cluster and are most similar with biography. The English non-fiction cluster in Figure 4b encompasses the aforementioned reference registers along with popular lore, humour, and biography, which can be caused by systematic or sampling differences. Most importantly, the non-fiction cluster includes both IEW and IEL and Presidents Bush and Obama's speech.

In other words, Figure 4 indicates that the internal heterogeneity of fiction/non-fiction clusters in both languages being considered, IEL may have shifted Premier Li's speech from a fictional to non-fictional register, which likely results from downplaying indefinite and second-person pronouns (Section 5.3.1), amplifiers, hedges, modals, and *be* (Section 5.3.3), and nominalising verbs (Sections 5.3.1 and 5.3.2). This research thus provides, for the first time, concrete evidence for the features of oral-type source speech that are affected by the "equalising effect" of interpreting (Cf. Shlesinger and Ordan, 2012; Kajzer-Wietrzny, 2015).

Further, the long stem connecting Premiers Wen and Li's speech shows their vast dissimilarity, but their interpretations IEW and IEL are highly similar to each other. This observation can be supported by the pair-wise Euclidean distance values, as Premier Wen's speech was measured to be most similar with review ($d_{Wen,review}$ = 4.92) whilst Premier Li's speech with biography ($d_{Li,biography}$ = 3.42; note that $d_{Wen,Li}$ = 9.95). In contrast, IEW and IEL are closest to President Bush's speech ($d_{IEW,Bush}$ = 2.66, $d_{IEL,Bush}$ = 2.65; Cf. $d_{IEW,Obama}$ = 11.08 and $d_{IEL,Obama}$ = 7.86), which testifies in part to the effect of target register expectations (see Section 5.2). More interestingly, the next closest register of IEW and IEL is each other ($d_{IEW,IEL}$ = 3.76) and their third most similar register is editorial ($d_{IEW,editorial}$ = 4.95, $d_{IEL,editorial}$ = 4.53). As editorial is the most opinionated register in Biber's analysis (1988, p. 148), such a finding likely validates the analysis of the attitudinal scale (Section 5.3.3) that interpreters may have strengthened original stance conveyance.

6 Discussions and limitations

According to interpreting service guidelines by the International Standardization Organization (ISO 18841: 2018), interpreting is "conveying both the register and the meaning of the source language content" (2018). Based on a cross-linguistic multidimensional approach and a case study of political conference interpreting, the interpreters concerned are demonstrated to have shifted literate source speech to a more oral, attitudinal, less formal and narrative register and oral sources to a less oral, narrative, more attitudinal and formal register. This research shows for the first time the "equalising effect" on the register level as regardless of source types, the interpreting products tend to be more similar to each other than to source registers.

Scrutiny of the cumulative effect of linguistic features in relation to each language's register system reveals three functional scales that exist in both Chinese and English. The effects of register expectations, interpreting as a form of mediation, and source speech interference were also teased out. It was found that the register under investigation, i.e. oral performance of scripted political presentation, legitimises a highly persuasive and non-narrative native and interpreted variety. Source speech was indicated to license, to some extent, the manifestation of extensive phrasal coordination, nouns, nominalisations, and long words in the interpreted speech. The effect of interpreting is that ambiguity is avoided, and formal and compact equivalents are preferred in this mode. The strength of the "equalising effect" varies according to source types,

as in the case of literate sources, more amplifiers and fewer nominalisations, indefinite pronouns, and perfect aspect verbs were used in the interpretations, whereas for oral sources, fewer second-person and indefinite pronouns, amplifiers, hedges, *be*, intransitive and modal verbs, and more nominalisations were introduced. Whilst IE's resemblance with President Bush's speech confirms in part the effect of target register expectations, its similarity to editorial irrespective of source register may be indicative of a higher priority in staff interpreters' task hierarchy than interpreting, that is conveying "the PRC's position in major issues" (MFA Department of Personnel, 2018), but other types of evidence are needed to further validate such a claim.

I hasten to note that the register equalising effect can also be attributed to the fact that interpreters considered herein worked into their B language, and only expert users of a language have multiple registers at their disposal and can activate them with relative ease (Hale, 2015). It will hence be valuable to test the MD analyses in the inverse direction (English–Chinese) and among different interpreter groups. Future replication on larger and SI corpora is also desirable. The inclusion of spoken Chinese corpora and dependency parsing can aid the improvement of MulDi Chinese, and the findings of this work should be integrated with process-oriented studies.

Bibliography

Baker, Mona (1996) 'Corpus-based translation studies: The challenges that lie ahead', in Sager, Juan C. and Somers, H. L. (eds) Terminology, LSP, and translation: Studies in language engineering in honour of Juan C. Sager. Amsterdam: John Benjamins, pp. 175–186.

Baumgarten, Nicole, Meyer, Bernd and Özçetin, Demet (2008) 'Explicitness in translation and interpreting: A critical review and some empirical evidence (of an elusive concept)', Across Languages and Cultures, 9(2), pp. 177–203. doi:10.1556/Acr.9.2008.2.2.

Beaton, Morven (2007) 'Interpreted Ideologies in Institutional Discourse: The Case of the European Parliament', The Translator, 13(2), pp. 271–296. doi:10.1080/13556509.2007.10799241.

Berk-Seligson, Susan (1989) 'The role of register in the bilingual courtroom: evaluative reactions to interpreted testimony', International Journal of the Sociology of Language, 1989(79), pp. 79–92. doi:10.1515/ijsl.1989.79.79.

Biber, Douglas (1988) Variation across Speech and Writing. Cambridge: Cambridge University Press.

Biber, Douglas (1993) 'Representativeness in Corpus Design', Literary and Linguistic Computing, 8(4), pp. 243–258.

Biber, Douglas (1995) Dimensions of Register Variation: A Cross-Linguistic Comparison. Cambridge: Cambridge University Press.

Biber, Douglas (2014) 'Using multi-dimensional analysis to explore cross-linguistic universals of register variation', Languages in Contrast, 14(1), pp. 7–34. doi:10.1075/lic.14.1.02bib.

Biber, Douglas and Conrad, Susan (2019) Register, Genre, and Style. Second edition. Cambridge: Cambridge University Press.

Chen, Lijiang (2007) Wenhua yujing yu zhengzhi huayu: zhengfu xinwen fabuhui de huayu yanjiu [Cultural context and political discourse: A discourse analysis of government press conferences]. PhD Thesis. Shanghai International Studies University.

De Sutter, Gert and Lefer, Marie-Aude (2019) 'On the need for a new research agenda for corpus-based translation studies: a multi-methodological, multifactorial and interdisciplinary approach', Perspectives, 28(1), pp. 1–23.

Egbert, Jesse (2014) Reader perceptions of linguistic variation in published academic writing. PhD Thesis. Northern Arizona University.

Friginal, Eric and Biber, Douglas (2016) 'Multi-Dimensional Analysis', in Egbert, Jesse and Baker, Paul (eds) Triangulating Methodological Approaches in Corpus-Linguistic Research. New York, NY: Routledge, pp. 73–89.

Gries, Stefan Th. (2008) 'Exploring variability within and between corpora: some methodological considerations', Corpora, 1(2), pp. 109–151. doi:10.3366/cor.2006.1.2.109.

Hale, Sandra (2015) 'Register', in Pöchhacker, Franz (ed.) Routledge Encyclopedia of Interpreting Studies. Abingdon-on-Thames, UK: Routledge, pp. 338–339.

Hong, Lei and Wang, Binhua (2011) 'Hanying jiaoti chuanyi zhong kouyu xing/shumian xing pianyi de miaoshu xing yanjiu: yi liangchang jizhehui xianchang kouyi wei anli [A descriptive study of shifts in the oral/literate continuum in Chinese-English consecutive interpreting: A case study of two press conferences]', Zhongguo fanyi, (2), pp. 73–77.

Hu, Kaibao and Tao, Qing (2012) 'Jizhe zhaodai hui hanying kouyi jufa caozuo guifan yanjiu [Syntactic operational norms of press conference interpreting (Chinese-English)]', Waiyu jiaoxue yu yanjiu, 44(5), pp. 738–750.

ISO (2018) ISO 18841:2018 Interpreting services -- General requirements and recommendations, ISO. Available at: http://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/35/63544.html (Accessed: 3 March 2022).

Ji, Meng (2017) 'A multidimensional analysis of the translational Chinese genre system', in Ji, Meng et al. (eds) Corpus methodologies explained: An empirical approach to translation studies. London: Routledge, pp. 53–102.

Kajzer-Wietrzny, Marta (2015) 'Simplification in interpreting and translation', Across Languages and Cultures, 16(2), pp. 233–255. doi:10.1556/084.2015.16.2.5.

Kajzer-Wietrzny, Marta (2021) 'An intermodal approach to cohesion in constrained and unconstrained language', Target [Preprint]. doi:10.1075/target.19186.kaj.

Kreiser, Maria and Greene, Michael (2022) History, Evolution, and Practices of the President's State of the Union Address: Frequently Asked Questions. R44770. Washington, DC: Congressional Research Service. Available at: https://sgp.fas.org/crs/misc/R44770.pdf.

Kruger, Haidee and van Rooy, Bertus (2016) 'Constrained language: A multidimensional analysis of translated English and a non-native indigenised variety of English', English World-Wide, 37(1), pp. 26–57. doi:10.1075/eww.37.1.02kru.

Li, Charles N. and Thompson, Sandra A. (1989) Mandarin Chinese: A Functional Reference Grammar. Berkeley, CA: University of California Press.

Li, Dechao and Wang, Kefei (2012) 'Hanying tongchuan zhong cihui moshi de yuliao ku kaocha [A corpus-based study on lexical patterns in simultaneous interpreting from Chinese into English]', Xiandai waiyu, 35(4), pp. 409–415.

Liu, Nannan (2021) Multidimensional analysis tagger of Mandarin Chinese (version 0.3.2). Available at: https://doi.org/10.5281/zenodo.5220449.

Livingston, Sue, Singer, Bonnie and Abrahamson, Theodore (1994) 'Effectiveness Compared: ASL Interpretation vs. Transliteration', Sign Language Studies, 1082(1), pp. 1–54. doi:10.1353/sls.1994.0008.

Lv, Qianxi and Liang, Junying (2018) 'Is consecutive interpreting easier than simultaneous interpreting? – a corpus-based study of lexical simplification in interpretation', Perspectives, 27(1), pp. 91–106. doi:10.1080/0907676X.2018.1498531.

MFA Department of Personnel (2018) Waijiaobu fanyi gongzuo zhe [Ministry of Foreign Affairs Translation and Interpretation Staff], Ministry of Foreign Affairs, China. Available at: https://www.mfa.gov.cn/web/wjb_673085/gbclc_603848/wjsy_660703/fyfc_660713/201309/t20130925_9277461.shtml (Accessed: 3 March 2022).

Munday, Jeremy (2016) Introducing Translation Studies: Theories and Applications. Fourth edition. London: Routledge.

Nini, Andrea (2019) 'The Multi-Dimensional Analysis Tagger', in Berber Sardinha, Tony and Veirano Pinto, Marcia (eds) Multi-Dimensional Analysis: Research Methods and Current Issues. London; New York: Bloomsbury Academic, pp. 67–94.

Pym, Anthony (2007) 'On Shlesinger's proposed equalizing universal for interpreting', in Pöchhacker, Franz, Jakobsen, Arnt Lykke, and Mees, Inger (eds) Interpreting studies and beyond: A tribute to Miriam Shlesinger. Copenhagen: Samfundslitteratur, pp. 175–190.

Raschka, Sebastian and Mirjalili, Vahid (2019) Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2. Third edition. Birmingham, UK: Packt Publishing.

Rossum, Guido (1995) Python Reference Manual. Amsterdam: CWI (Centre for Mathematics and Computer Science).

Shlesinger, Miriam (1989) Simultaneous interpretation as a factor in effecting shifts in the position of texts in the oral–literate continuum. MA Thesis. Tel Aviv University.

Shlesinger, Miriam (1995) 'Shifts in Cohesion in Simultaneous Interpreting', The Translator, 1(2), pp. 193–214. doi:10.1080/13556509.1995.10798957.

Shlesinger, Miriam (1998) 'Corpus-based interpreting studies as an offshoot of corpus-based translation studies', Meta: Journal des traducteurs, 43(4), pp. 486–493. doi:10.7202/004136ar.

Shlesinger, Miriam and Ordan, Noam (2012) 'More spoken or more translated? Exploring a known unknown of simultaneous interpreting', Target, 24(1), pp. 43–60.

Toutanova, Kristina et al. (2003) 'Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network', in Proceedings of HLT-NAACL 2003. HLT-NAACL, Edmonton, Canada, pp. 252–259.

Wang, Binhua and Zou, Bing (2018) 'Exploring Language Specificity as a Variable in Chinese-English Interpreting: A Corpus-Based Investigation', in Russo, Mariachiara,

Bendazzoli, Claudio, and Defrancq, Bart (eds) Making Way in Corpus-based Interpreting Studies. Singapore: Springer, pp. 65–82.

Xiao, Richard (2009) 'Multidimensional analysis and the study of world Englishes', World Englishes, 28(4), pp. 421–450. doi:10.1111/j.1467-971X.2009.01606.x.

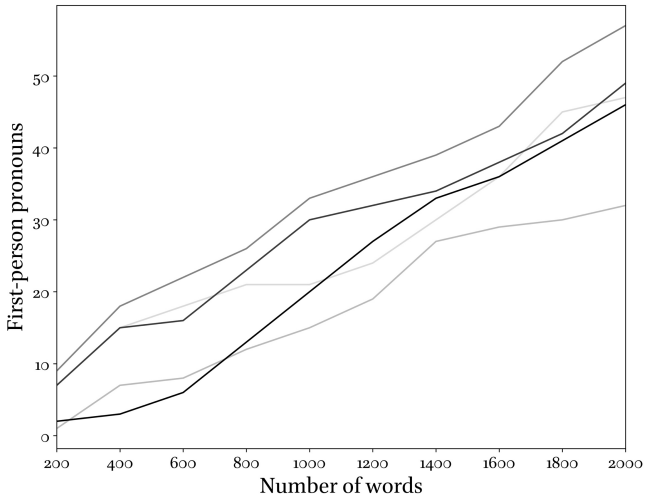
Yi, Yan and Chang, Tsan Kuo (2012) 'Institutionalizing public relations in China: A sociological analysis of the Chinese Premier's Press Conference', Public Relations Review, 38(5), pp. 711–722. doi:10.1016/j.pubrev.2011.12.007.

Zhang, Hua-Ping et al. (2003) 'HHMM-based Chinese Lexical Analyzer ICTCLAS', in Proceedings of the Second SIGHAN Workshop on Chinese Language Processing - Volume 17. ACL2003, Sapporo, Japan: Association for Computational Linguistics (SIGHAN '03), pp. 184–187. doi:10.3115/1119250.1119280.

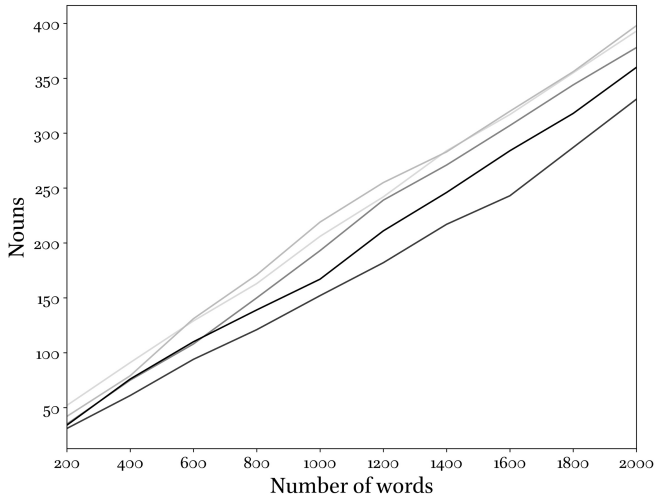
Appendices

Appendix A: Distributions of four features in five 2,000-word samples of OC

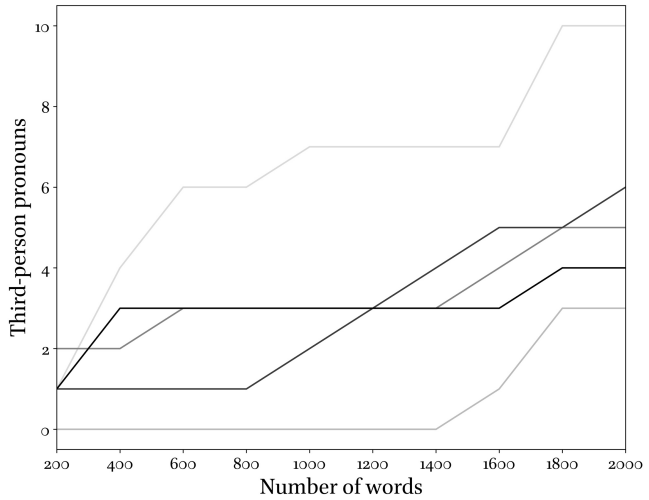
a. First-person pronouns



b. Nouns



c. Third-person pronouns



d. Amplifiers

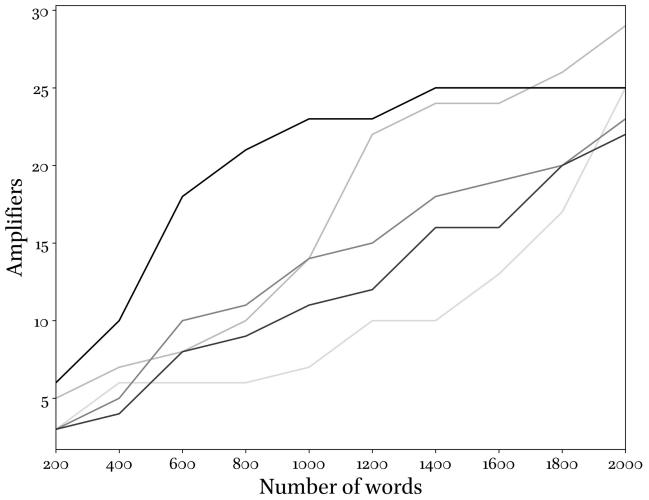


FIGURE A: Distributions of four features in five 2,000-word samples of OC.

Appendix B: Individual feature difference patterns of OC

TABLE B-1: Features of significantly higher frequencies in Premier Wen's speech.

		Wen	**Li**	**Welch's *t*-tests**			
Feature	Primary dimension	*M* [*SD*]	*M* [*SD*]	*df*	*t*	*p*	*d*
Nouns excl. nominalisations	Dimension 1 (literate)	193.47 [16.58]	184.59 [15.36]	47.39	2.06	0.045	0.559
Average word length	Dimension 1 (literate)	1.55 [0.04]	1.51 [0.03]	40.69	5.68	<0.001	1.595
Perfect aspect markers	Dimension 2 (narrative)	9.87 [3.97]	7.53 [3.87]	48.95	2.22	0.031	0.598
Emotion words	Dimension 3 (evaluative)	17.27 [8.43]	13.08 [5.7]	37.83	2.11	0.042	0.6
Auxiliary adjectives	Dimension 4 (abstract writing)	7.66 [3.12]	5.47 [3.3]	51.32	2.56	0.014	0.679
Disyllabic words	Dimension 4 (abstract writing)	57.12 [16.27]	44.68 [9.57]	34.46	3.35	0.002	0.972
Nominalisations	Dimension 4 (abstract writing)	65.42 [12.5]	55.17 [9.24]	40.34	3.4	0.002	0.956
		Wen	**Li**	**Mann–Whitney *U* tests**			
Feature	Primary dimension	*Mdn* [*IQR*]	*Mdn* [*IQR*]	*U*	*p*	r_{rb}	
Indefinite pronouns	Dimension 1 (oral)	4.08 [3.1]	2.46 [2.05]	591.5	0.002	−0.494	

Classical grammatical words	Dimension 1 (literate)	5.08 [5.36]	2.53 [2.56]	567.5	0.006	−0.433	
Classifiers	Dimension 2 (narrative)	24.94 [9.61]	15.97 [6.34]	669	<0.001	−0.689	
Abstract nouns	Dimension 2 (non-narrative)	92.88 [27.93]	74.03 [14.33]	628	<0.001	−0.586	
Concessive conjuncts	Dimension 3 (evaluative)	4.77 [5.79]	3.52 [3.98]	525	0.038	−0.326	
Phrasal coordination	Dimension 4 (abstract writing)	11.23 [7.28]	4.97 [5]	688	<0.001	−0.737	

TABLE B-2: Features of significantly higher frequencies in Premier Li's speech.

		Wen	**Li**	**Welch's *t*-tests**			
Feature	Primary dimension	*M* [*SD*]	*M* [*SD*]	*df*	*t*	*p*	*d*
Monosyllabic verbs	Dimension 1 (oral)	70.43 [8.53]	109.37 [13.5]	54.1	−13.31	<0.001	3.333
Average clause length	Dimension 1 (literate)	6.77 [0.46]	7.15 [0.45]	49.08	−3.05	0.004	0.821
Intransitive verbs	Dimension 2 (narrative)	11.5 [4.32]	16.87 [5.12]	53.74	−4.29	<0.001	1.119
Adverbs	Dimension 3 (evaluative)	54.63 [10.08]	67.31 [8.15]	43.06	−5.07	<0.001	1.407
Existential *yǒu*	Dimension 3 (evaluative)	6.16 [3.34]	9.83 [4.03]	54.01	−3.76	<0.001	0.978
Hedges	Dimension 3 (evaluative)	3.19 [2]	9.53 [4.48]	47.02	−7.2	<0.001	1.734
HSK Level 3 vocabulary	Dimension 3 (evaluative)	143.71 [23.26]	180.02 [17.83]	41.42	−6.4	<0.001	1.791
Modifying adverbs	Dimension 3 (evaluative)	17.61 [5.7]	23.68 [4.72]	43.78	-4.26	<0.001	1.178
shì 'be'	Dimension 3 (evaluative)	18.74 [5.61]	22.31 [3.87]	38.32	−2.68	0.011	0.763

Amplifiers	Dimension 3 (evaluative)	13.69 [4.7]	17.01 [5.72]	54.11	−2.4	0.02	0.623
Modal verbs	Dimension 3 (evaluative)	22.42 [5.96]	32.73 [8.34]	54.99	−5.44	<0.001	1.386
		Wen	**Li**	**Mann–Whitney *U* tests**			
Feature	Primary dimension	*Mdn* [*IQR*]	*Mdn* [*IQR*]	*U*	*p*	r_{rb}	
Particles	Dimension 1 (oral)	15.49 [13.07]	36.36 [6.29]	31	<0.001	0.922	
Second-person pronouns	Dimension 1 (oral)	2 [3.83]	4.78 [3.55]	229	0.007	0.422	
Public verbs	Dimension 1 (oral)	6.55 [4.81]	8.84 [2.79]	231	0.008	0.417	
Average sentence length	Dimension 1 (literate)	17.9 [3.28]	22.54 [4.94]	75	<0.001	0.811	
Simile	Dimension 2 (narrative)	0 [0.55]	1.26 [0.29]	170	<0.001	0.571	
Adversative conjuncts	Dimension 3 (evaluative)	8.11 [7.63]	15.2 [6.19]	134	<0.001	0.662	
Lexical density	Dimension 4 (abstract writing)	523.1 [28]	562.72 [40.39]	89	<0.001	0.775	

TABLE B-3: Features of no significant differences between Premiers Wen and Li's speech.

		Wen	**Li**	**Welch's *t*-tests**			
Feature	Primary dimension	*M* [*SD*]	*M* [*SD*]	*df*	*t*	*p*	*d*
First-person pronouns	Dimension 1 (oral)	27.24 [9.63]	24.42 [6.64]	38.36	1.24	0.224	0.351
Monosyllabic negation	Dimension 1 (oral)	8.01 [5.07]	8.18 [4.1]	43.09	−0.14	0.891	0.038
Classical syntax	Dimension 1 (literate)	28.54 [7.14]	31.81 [7.1]	49.53	−1.71	0.093	0.46

Imperfect aspect markers	Dimension 2 (narrative)	10.23 [2.99]	10.59 [3.81]	54.65	−0.4	0.688	0.104
Unique words ratio	Dimension 2 (narrative)	294.29 [22.36]	303.52 [27.31]	54.16	−1.4	0.167	0.364
HSK Level 1 vocabulary	Dimension 3 (evaluative)	245.66 [23.3]	238.84 [21.86]	47.79	1.12	0.269	0.303
Private verbs	Dimension 3 (evaluative)	16.9 [5.71]	16.65 [4.88]	44.83	0.17	0.862	0.048
Disyllabic verbs	Dimension 4 (abstract writing)	146.58 [17.23]	145.09 [15.46]	46.38	0.34	0.738	0.092
Demonstrative pronouns	Dimension 4 (concrete writing)	15.09 [4.32]	17.14 [4.97]	53.16	−1.66	0.104	0.434
		Wen	**Li**	**Mann–Whitney *U* tests**			
Feature	Primary dimension	*Mdn* [*IQR*]	*Mdn* [*IQR*]	*U*	*p*	r_{rb}	
Questions	Dimension 1 (oral)	0 [2.44]	1.24 [2.43]	343.5	0.377	0.133	
WH-words	Dimension 1 (oral)	1.24 [3.27]	2.47 [2.43]	332	0.297	0.162	
Exclamation	Dimension 1 (oral)	0 [0]	0 [0]	377	0.675	0.048	
Discourse particles	Dimension 1 (oral)	0 [1.39]	1.21 [2.47]	338.5	0.319	0.145	
Chinese person names	Dimension 1 (oral)	1.1 [2.64]	1.16 [1.29]	426.5	0.61	−0.077	
Honorifics	Dimension 1 (oral)	1.25 [2.96]	0 [1.28]	493	0.099	−0.245	
Disyllabic prepositions	Dimension 1 (literate)	3.42 [3.79]	2.46 [2.64]	486	0.147	−0.227	
Descriptive words	Dimension 2 (narrative)	0 [0]	0 [0]	405	0.847	−0.023	
Adverbial marker *di*	Dimension 2 (narrative)	2.67 [2.52]	2.52 [2.65]	365.5	0.627	0.077	
Third-person pronouns	Dimension 2 (narrative)	2.41 [4.97]	2.55 [2.48]	352	0.479	0.111	

Passive marker *bèi*	Dimension 2 (narrative)	0 [0]	0 [0]	348	0.083	0.121	
seem/*appear* words	Dimension 2 (narrative)	0 [0]	0 [0]	384	0.414	0.03	
Prepositions	Dimension 2 (narrative)	1.15 [2.43]	1.3 [1.87]	279	0.056	0.295	
Complement marker *de*	Dimension 2 (narrative)	0 [1.32]	0 [1.15]	466	0.188	−0.177	
Onomatopoeic words	Dimension 2 (narrative)	0 [0]	0 [0]	350	0.167	0.116	
Conditional conjuncts	Dimension 3 (evaluative)	2.46 [4.26]	2.27 [3.72]	426.5	0.624	−0.077	
Other personal pronouns	Dimension 3 (evaluative)	3.76 [4.45]	3.05 [3.68]	446.5	0.419	−0.128	
Downtoners	Dimension 3 (evaluative)	1.23 [1.42]	2.35 [3.84]	298	0.109	0.247	
Disyllabic negation	Dimension 3 (evaluative)	1.23 [1.93]	1.2 [2.48]	405	0.887	−0.023	
Auxiliary verbs	Dimension 4 (abstract writing)	0.6 [1.37]	1.18 [1.3]	380	0.79	0.04	

Appendix C: Individual feature difference patterns of IE versus NE

Note that owing to space constraints, only features loading on dimensions discussed in the text are reported in Appendices C and D.

TABLE C-1: Features of significantly higher frequencies in IE.

		IE	**NE**	**Welch's *t*-tests**			
Feature	Primary dimension	*M* [*SD*]	*M* [*SD*]	*df*	*t*	*p*	*d*
Average word length	Dimension 1 (informational)	4.81 [0.13]	4.66 [0.19]	109.23	5.33	<0.001	0.971

Attributive adjectives	Dimension 1 (informational)	8.59 [1.26]	6.91 [1.27]	115.51	7.2	<0.001	1.326
Private verbs	Dimension 1 (involved)	1.32 [0.48]	1 [0.42]	111.18	3.83	<0.001	0.708
Adverbs	Dimension 3 (situation-dependent reference)	3.14 [0.56]	2.83 [0.85]	104.46	2.4	0.018	0.436
Agentless passives	Dimension 5 (abstract)	0.64 [0.34]	0.52 [0.32]	114.15	2	0.048	0.369
		IE	**NE**	**Mann–Whitney *U* tests**			
Feature	Primary dimension	*Mdn* [*IQR*]	*Mdn* [*IQR*]	*U*	*p*	r_{rb}	
Amplifiers	Dimension 1 (involved)	0.23 [0.27]	0 [0.11]	3112	<0.001	−0.79	
Be as main verb	Dimension 1 (involved)	1.44 [0.6]	1.28 [0.56]	2142	0.03	−0.232	
Demonstrative pronouns	Dimension 1 (involved)	0.55 [0.34]	0.45 [0.46]	2121	0.04	−0.22	
Pro-verb *do*	Dimension 1 (involved)	0.12 [0.22]	0.11 [0.23]	1360	0.038	0.218	
That deletion	Dimension 1 (involved)	0.21 [0.2]	0.11 [0.22]	2216	0.009	−0.275	
Nominalisations	Dimension 3 (explicit reference)	4.82 [1.48]	2.88 [1.3]	3147	<0.001	−0.81	
Conjuncts	Dimension 5 (abstract)	0.2 [0.26]	0 [0.11]	2685	<0.001	−0.544	
Predicative adjectives	Dimension 5 (abstract)	0.63 [0.45]	0.45 [0.45]	2116	0.042	−0.217	

TABLE C-2: Features of significantly higher frequencies in NE.

		IE	**NE**	**Welch's *t*-tests**

Feature	Primary dimension	*M* [*SD*]	*M* [*SD*]	*df*	*t*	*p*	*d*
Total other nouns excl. gerunds and nominalisations	Dimension 1 (informational)	20.74 [1.85]	23.32 [2.03]	115.95	−7.24	<0.001	1.331
Present tense verbs	Dimension 1 (involved)	5.65 [0.88]	6.4 [1.04]	115.01	−4.25	<0.001	0.778
First-person pronouns	Dimension 1 (involved)	4.38 [1.22]	4.99 [1.46]	114.65	−2.49	0.014	0.455
Independent clause coordination	Dimension 1 (involved)	0.8 [0.41]	1.68 [0.49]	114.23	−10.65	<0.001	1.949
Place adverbials	Dimension 3 (situation-dependent reference)	0.27 [0.17]	0.34 [0.21]	114.45	−2.14	0.034	0.393

		IE	**NE**	**Mann–Whitney *U* tests**			
Feature	Primary dimension	*Mdn* [*IQR*]	*Mdn* [*IQR*]	*U*	*p*	r_{rb}	
Type-token ratio	Dimension 1 (informational)	389 [44]	411 [28]	968	<0.001	0.443	
Because	Dimension 1 (involved)	0.1 [0.13]	0.11 [0.27]	1234	0.005	0.29	
Contractions	Dimension 1 (involved)	0.21 [0.25]	0.9 [1.5]	542.5	<0.001	0.688	
Emphatics	Dimension 1 (involved)	0.58 [0.32]	0.74 [0.57]	1326	0.026	0.237	
Indefinite pronouns	Dimension 1 (involved)	0 [0]	0 [0.11]	1353	0.004	0.222	
Possibility modals	Dimension 1 (involved)	0.42 [0.31]	0.66 [0.38]	1214.5	0.005	0.301	
Analytic negation	Dimension 1 (involved)	0.44 [0.41]	0.68 [0.58]	1061	<0.001	0.39	
WH clauses on subject positions	Dimension 3 (explicit reference)	0 [0.11]	0.22 [0.22]	863.5	<0.001	0.503	
Time adverbials	Dimension 3 (situation-dependent reference)	0.22 [0.26]	0.58 [0.49]	461	<0.001	0.735	
Past participial clauses	Dimension 5 (abstract)	0 [0]	0 [0]	1420	0.005	0.183	

TABLE C-3: Features of no significant differences between IE and NE.

		IE	**NE**	**Welch's *t*-tests**			
Feature	Primary dimension	*M* [*SD*]	*M* [*SD*]	*df*	*t*	*p*	*d*
Phrasal coordination	Dimension 3 (explicit reference)	1.44 [0.55]	1.23 [0.6]	115.97	1.97	0.051	0.363
		IE	**NE**	**Mann–Whitney U tests**			
Feature	Primary dimension	*Mdn* [*IQR*]	*Mdn* [*IQR*]	*U*	*p*	r_{rb}	
Prepositions	Dimension 1 (informational)	9.95 [1.25]	9.83 [1.7]	1925	0.316	−0.107	
Discourse particles	Dimension 1 (involved)	0 [0.12]	0 [0.11]	1843	0.532	−0.06	
Hedges	Dimension 1 (involved)	0 [0]	0 [0]	1834.5	0.264	−0.055	
Pronoun *it*	Dimension 1 (involved)	0.7 [0.49]	0.69 [0.66]	1735	0.987	0.002	
Sentence relatives	Dimension 1 (involved)	0 [0.1]	0 [0]	1976	0.117	−0.137	
Second-person pronouns	Dimension 1 (involved)	0.51 [0.38]	0.45 [0.59]	1893.5	0.405	−0.089	
Stranded prepositions	Dimension 1 (involved)	0 [0]	0 [0.11]	1438	0.053	0.173	
Analytic negation	Dimension 1 (involved)	0.11 [0.22]	0.11 [0.33]	1608.5	0.478	0.075	
WH clauses	Dimension 1 (involved)	0 [0]	0 [0.11]	1535.5	0.176	0.117	
Direct WH-questions	Dimension 1 (involved)	0 [0]	0 [0]	1672.5	0.53	0.038	
Pied-piping relative clauses	Dimension 3 (explicit reference)	0 [0]	0 [0]	1803	0.612	−0.037	
WH relative clauses on object positions	Dimension 3 (explicit reference)	0 [0]	0 [0]	1803.5	0.536	−0.037	

By passives	Dimension 5 (abstract)	0 [0.1]	0 [0.11]	1632.5	0.509	0.061	
Other adverbial subordinators	Dimension 5 (abstract)	0.12 [0.25]	0.21 [0.12]	1449.5	0.116	0.166	
Past participial WHIZ deletion relatives	Dimension 5 (abstract)	0.11 [0.14]	0.11 [0.11]	2021.5	0.119	−0.163	
Present participial WHIZ deletion relatives	Dimension 5 (abstract)	0.12 [0.2]	0.11 [0.23]	1930.5	0.295	−0.11	

Appendix D: Individual feature difference patterns of IEW versus IEL

TABLE D-1: Features of significantly higher frequencies in IEW.

		IEW	**IEL**	**Welch's *t*-tests**			
Feature	Primary dimension	*M* [*SD*]	*M* [*SD*]	*df*	*t*	*p*	*d*
Prepositions	Dimension 1 (informational)	10.87 [1.19]	9.54 [0.62]	32.2	5	<0.001	1.47
Total other nouns excl. gerunds and nominalisations	Dimension 1 (informational)	21.45 [1.55]	20.22 [1.9]	54.17	2.7	0.009	0.701
Average word length	Dimension 1 (informational)	4.86 [0.14]	4.78 [0.12]	45.63	2.21	0.032	0.605
Phrasal coordination	Dimension 3 (explicit reference)	1.81 [0.48]	1.17 [0.44]	47.36	5.18	<0.001	1.408
		IEW	**IEL**	**Mann–Whitney *U* tests**			
Feature	Primary dimension	*Mdn* [*IQR*]	*Mdn* [*IQR*]	*U*	*p*	r_{rb}	

Amplifiers	Dimension 1 (involved)	0.3 [0.35]	0.22 [0.19]	536.5	0.024	−0.355	
Present participial clauses	Dimension 2 (narrative)	0.08 [0.18]	0 [0]	524.5	0.018	−0.324	

TABLE D-2: Features of significantly higher frequencies in IEL.

		IEW	**IEL**	**Welch's *t*-tests**			
Feature	Primary dimension	*M* [*SD*]	*M* [*SD*]	*df*	*t*	*p*	*d*
Pronoun *it*	Dimension 1 (involved)	0.62 [0.28]	0.9 [0.44]	54.13	−2.96	0.005	0.742
		IEW	**IEL**	**Mann–Whitney *U* tests**			
Feature	Primary dimension	*Mdn* [*IQR*]	*Mdn* [*IQR*]	*U*	*p*	r_{rb}	
Because	Dimension 1 (involved)	0 [0.1]	0.11 [0.3]	181.5	<0.001	0.542	
Demonstrative pronouns	Dimension 1 (involved)	0.48 [0.44]	0.61 [0.31]	225.5	0.006	0.431	
Discourse particles	Dimension 1 (involved)	0 [0.05]	0.1 [0.3]	229	0.003	0.422	
Stranded prepositions	Dimension 1 (involved)	0 [0]	0 [0.11]	291	0.025	0.265	
Prediction modals	Dimension 4 (overt persuasion)	1 [0.38]	1.44 [0.72]	167	<0.001	0.578	

TABLE D-3: Features of no significant differences between IEW and IEL.

		IEW	**IEL**	**Welch's *t*-tests**

Feature	Primary dimension	*M* [*SD*]	*M* [*SD*]	*df*	*t*	*p*	*d*
Attributive adjectives	Dimension 1 (informational)	8.38 [1.19]	8.74 [1.31]	52.36	–1.08	0.286	0.284
Present tense verbs	Dimension 1 (involved)	5.67 [0.96]	5.63 [0.83]	45.06	0.14	0.89	0.038
Private verbs	Dimension 1 (involved)	1.24 [0.46]	1.38 [0.5]	51.99	–1.09	0.28	0.289
Analytic negation	Dimension 1 (involved)	0.47 [0.27]	0.49 [0.24]	46.92	–0.41	0.687	0.11
Emphatics	Dimension 1 (involved)	0.55 [0.24]	0.62 [0.28]	53.24	–0.95	0.348	0.248
First-person pronouns	Dimension 1 (involved)	4.67 [1.09]	4.16 [1.28]	53.67	1.63	0.109	0.426
Independent clause coordination	Dimension 1 (involved)	0.74 [0.45]	0.84 [0.38]	44.64	–0.84	0.407	0.23
Perfect aspect verbs	Dimension 2 (narrative)	0.84 [0.37]	0.8 [0.39]	51.31	0.37	0.715	0.097
Nominalisations	Dimension 3 (explicit reference)	4.44 [0.86]	4.89 [1.04]	53.97	–1.8	0.078	0.468

		IEW	**IEL**	**Mann–Whitney *U* tests**			
Feature	Primary dimension	*Mdn* [*IQR*]	*Mdn* [*IQR*]	*U*	*p*	r_{rb}	
Type-token ratio	Dimension 1 (informational)	392.5 [44.5]	389 [45]	488	0.139	–0.232	
Be as main verb	Dimension 1 (involved)	1.39 [0.63]	1.54 [0.52]	399	0.968	–0.008	
Contractions	Dimension 1 (involved)	0.2 [0.29]	0.21 [0.25]	398	0.981	–0.005	
Hedges	Dimension 1 (involved)	0 [0]	0 [0]	385	0.75	0.028	
Indefinite pronouns	Dimension 1 (involved)	0 [0]	0 [0]	381	0.66	0.038	
Possibility modals	Dimension 1 (involved)	0.35 [0.26]	0.46 [0.3]	318.5	0.213	0.196	

Pro-verb *do*	Dimension 1 (involved)	0.09 [0.22]	0.12 [0.22]	360	0.552	0.091	
Sentence relatives	Dimension 1 (involved)	0.08 [0.1]	0 [0.1]	486.5	0.096	−0.229	
Second-person pronouns	Dimension 1 (involved)	0.46 [0.44]	0.6 [0.32]	315	0.193	0.205	
Analytic negation	Dimension 1 (involved)	0.1 [0.15]	0.12 [0.13]	313	0.177	0.21	
That deletion	Dimension 1 (involved)	0.22 [0.2]	0.18 [0.2]	447.5	0.408	−0.13	
WH clauses	Dimension 1 (involved)	0 [0.09]	0 [0]	421	0.6	−0.063	
Direct WH-questions	Dimension 1 (involved)	0 [0]	0 [0]	434	0.255	−0.096	
Public verbs	Dimension 2 (narrative)	0.44 [0.19]	0.43 [0.25]	414	0.777	−0.045	
Third-person pronouns	Dimension 2 (narrative)	0.72 [0.71]	0.63 [0.54]	387	0.891	0.023	
Past tense	Dimension 2 (narrative)	0.78 [0.88]	0.86 [0.75]	411	0.815	−0.038	
if and *unless*	Dimension 4 (overt persuasion)	0.07 [0.19]	0.1 [0.15]	358	0.53	0.096	
Necessity modals	Dimension 4 (overt persuasion)	0.36 [0.53]	0.22 [0.3]	473	0.215	−0.194	
Split auxiliaries	Dimension 4 (overt persuasion)	0.68 [0.33]	0.52 [0.37]	488	0.139	−0.232	
Split infinitives	Dimension 4 (overt persuasion)	0 [0.09]	0 [0.12]	316	0.156	0.202	