to appear in 2021. In J. Pan, S. L. Halverson, & J. Munday (Eds.), *Translating and Interpreting Political Discourse: New Trends and Perspectives*. Brill.

CHAPTER 1

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. The extant 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 concerning oral-type source based on the systematic variation of linguistic features. This chapter applies a cross-lingual multidimensional approach to political conference interpreting parallel and comparable corpora based on a contrastive register analysis between Chinese and English. Quantitative analyses at the feature, dimension, function, and register levels reveal interpreters to have shifted literate source speech to a more oral, attitudinal, and less formal register, and oral source speech to a less oral, more formal, and attitudinal register. The effects of interpreting, source speech, and target register expectations were teased out. The most important finding is that regardless of source speech registers, interpreting products tend to show more similar registers to each other than to source speech.

Keywords: multidimensional analysis, political conference interpreting, register shifts, 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 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 for Wang and Zou, 2018). The lacuna is both theoretical and methodological. The theoretical constraint is that corpus-based interpreting studies (CIS) inherit a focus from corpus-based translation studies on product features that set the modality apart by comparing interpreted (both consecutive [CI] and simultaneous [SI]) with translated and native varieties (e.g. Shlesinger and Ordan, 2012; Kajzer-Wietrzny, 2015). 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 Multidimensional (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 has been interfered by source speech, constrained by the interpreting mode, and standardised according to target register expectations (Shlesinger, 1998).

Second, the existing literature has treated a few "cherry-picked" features (Setton, 2011, p. 45) in isolation, rather than systematically assessed the combined effect of a large, meaningful feature set. This can reduce a situational and contextual activity such as interpreting to mere frequency comparison (suggested by Malmkjaer, 1998). In that light, a MD view of language varieties at conferences based on co-occurring features that typify native registers is desirable. Third, a major hurdle in CIS, as acknowledged in various overviews (e.g. Shlesinger, 1998; Setton, 2011), 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, 2019), which raises the question of "optimal length" given that corpus-based measures are susceptible to text size (Biber, 1993; Kajzer-Wietrzny, 2015). 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 distribution of language varieties at political interpreting events in search of potential register shifts. The MD analysis approach was utilised to simultaneously account for the co-occurrence of "many lexical, grammatical, and semantic features" in two languages, i.e. Mandarin Chinese and English, and 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 apart the effect of interpreting. The current chapter addresses two descriptive questions.

- Do the source utterances (SU) and target utterances (TU) in the interpreting corpora under investigation show similarities in their registers?
- Do the TU and comparable corpus demonstrate similarities in the dimensions of register variation identified by Biber (1988)?

Section 2 reviews previous research of register shifts in CIS. Section 3 provides an outline of the cross-lingual MD analysis methodology and register-controlled corpora utilised, followed by the results of intra- and inter-lingual comparison in Section 5. Section 6 discusses the implications of these findings.

2 Register shifts in conference interpreting

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

2.1 Tendency towards orality

Shlesinger and her collaborators (1989; 2012; see also the discussion of Shlesinger [1989] in Pym [2007]) underline the equalising effect of SI, which means literate, written SU are shifted to more oral registers, whilst admitting that findings for oral-type SU were inconclusive (Shlesinger and Ordan, 2012,

p. 54). The "equalising effect", or in precise terms, "a greater tendency towards orality" in the case of literate-type SU (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" between SU and TU, 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 nominalisation, relative clauses, and attributive adjectives, which were equated with 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 written dichotomy, with the role of underlying variables being postulated and tested for English only (p. 11).

Following Shlesinger (1989), Hong and Wang (2011) collected raw frequencies of ten features in the SU and TU of two Chinese–English CI events and concluded two-way shifts that CI renders oral SU more literate and literate SU 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 lexical pattern changes, such as those of cohesive devices (Shlesinger, 1995), clauses and adverbs (Hu and Tao, 2012), high frequency words and list heads (Kajzer-Wietrzny, 2015), and choices of equivalents (Beaton, 2007; Baumgarten, Meyer and Özçetin, 2008). Such studies have reported a gravitation towards informational, formal, and complex language, with a mixed weakening or strengthening pattern found. Granted, such findings are more forcefully represented by shifts of linguistic feature distribution than a generalisation based on a comprehensive examination of register shifts. Therefore, a meaningful set of linguistic features and their role in relation to entire register systems of languages should be identified prior to a discussion of register shifts (as in Ji, 2017).

[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 set SI from translation apart, and thus the "orality" proposal therein should not be understood as vis-à-vis source speech.

In CIS, linguistic features have been examined far more sporadically. Shlesinger's (1995) investigation of cohesion shifts reveals interpreters' avoidance of "implicit" ellipsis in both English–Hebrew and Hebrew–English parallel corpora (Shlesinger, 1989, pp. 171–172), which lends support to the argument that SI "may exert a stronger effect" than target register expectations with regard to explicitation (1995, p. 210). In the Chinese–English combination, Hu and Tao (2012) show interpreters' penchant for connecting independent clauses in Chinese SU using sentence relatives with *which* (see "sentence relatives" in Tables C-3 and D-3), resulting, as they argue, in more complex, formal TU than SU (pp. 745–746). Their examination of degree adverbs (e.g. *very*, *absolutely*) reveals a general weakening effect owing to mediation (Cf. Beaton, 2007; Beaton-Thome, 2010).

Research that makes intra-lingual comparisons (e.g. Li and Wang, 2012; Kajzer-Wietrzny, 2015) is usually interested in examining lexical simplification or explicitation in interpreting, and in the Chinese–English direction has reported a preference for formal and neutral presentation of information, instantiated by the "excessive use" of nominalisation (Li and Wang, 2012, p. 414). Nonetheless, such "excess", as rightly pointed out by Baumgarten et al. (2008), can be licensed by the source text, target register expectations, and communicative conventions (pp. 184–197). Kajzer-Wietrzny (2015) refers to SU types as the mode of delivery and found oral, spoken SU to result in significantly more repetitive (measured with list heads) and less sophisticated (measured with high frequency words) interpretation than scripted SU (pp. 245–249). She also recommended a comprehensive inter-lingual contrastive analysis of SU and TU to consider language specific factors and an estimate of "how strong the 'levelling effect'" on oral and scripted SU types is, respectively (2015, pp. 245–252). Wang and Zou (2018) investigated language specificity by integrating parallel and comparable corpora and found interpreters to generate more difficult English output than US Department of State daily press briefings, and such a pattern was necessitated by cross-linguistic differences between Chinese front-loaded modifying structures marked by 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 change. In the institutional setting, Beaton-Thome's series of research (e.g. 2007; 2010) illuminate the vague boundaries between professional and institutional ideology among EU simultaneous interpreters, as an intensified use of "European Union" and "inclusive we" relative to the original speech can be suggested to strengthen "the salience of EU institutional hegemony" (Beaton, 2007, pp. 280–281). 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 texts, and is built upon robust associations of linguistic features with their functions in a language's register system. In this chapter, such a goal is pursued through a cross-lingual MD approach.

3 Towards cross-lingual MD analyses of register shifts

The current discussion of register shifts is based on contrastive MD analyses of Chinese and English registers (Biber, 1988, 1995, 2014). 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 correlation among individual features by simultaneously grouping co-occurring (positive features) or co-disappearing features (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 interjections 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/texts 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).

MD analysis can be applied to describe the register system of a language, or the range of activities that speakers of a language engage in (Biber, 1995, p. 5), which requires the use of large, register-controlled corpora. The present comparison between Chinese and English relies on 15 analogous registers in two Brown family corpora, each containing "fifteen registers, and 500 texts of about 2000 words" (Biber, 1995, p. 33), as well as other corpora used in Biber's (1988) MD analysis. Information about the Chinese and English corpora can be found in Table 1.

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

Brown family		**Chinese**		**English**		
		upsampled Texts of Recent CHinese 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	romantic fiction	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 Chinese and English corpora is the lack of spoken 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 oral written register humour (see Table 1), and as will be expounded in Section 4.1.1, the Chinese SU under investigation, especially Premier Wen's speech, are argued to epitomise "formal, written, … [and] planned language" (Chen, 2007, p. 60), thus a contrastive analysis based on such reference corpora can still be valuable.

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") has been completed by the author and resulted in a Python programme[2]. MulDi Chinese follows Biber's (1988) methodology and generates a set of 54 features and five 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 at the hyperlink in footnote 2). Biber's (1988) English MD analysis was replicated by Nini (2018) in the Multidimensional 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, 2018). Nonetheless, owing to slight differences in rules employed by Biber's (1988) analyses and the Stanford tagger, dimension statistics of the LOB corpus were obtained using the English tagger, instead of using those in Biber's (1988)

[2] All code, data, and feature description available at https://github.com/Nannan-Liu/Multidimensional-Analysis-Tagger-of-Mandarin-Chinese.

original analyses (pp. 122–124), which ensures the consistency of comparisons among IE, NE, and English reference registers. To situate OC, IE, and NE in comparable register systems, Brown family Chinese and English corpora only (see Table 1) were utilised in the analysis[3]. The English tagger generates standardised frequencies of 67 features and scores of input texts along 6 dimensions. MulDi Chinese and the English tagger constitute 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; (b) between interpreted English (IE) and native English (NE); (c) within IE speech; and (d) between OC and IE along functional scales.

4.1 Corpora

4.1.1 Interpreting parallel corpora

This research relies on self-built parallel corpora of interpreted premier press conferences (Cf. Wang and Zou, 2018) that correspond to ten years of Chinese premier press conferences after 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 premier opening and closing (when available) monologues and responses to reporter questions. The questions were pre-submitted and reporters pre-vetted, and thus as concluded by Yi and Chang (2012), the interpreting events are a political ritual rather than a 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 interpretation made by staff interpreters of China's Ministry of Foreign Affairs, who worked into their B

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

language. Chinese character sequences in OC were segmented into "words"[4] and tagged with segmentation software ICTCLAS.

4.1.2 Comparable corpus

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, 2020). Both SOTU and Chinese premier speech epitomise ritualistic political performance and serves such functions as policy debriefing, support rallying, and image management (Chen, 2007; Kreiser and Greene, 2020), and such a register is termed "oral performance of scripted political presentation" in this research. The comparable corpus includes transcripts of US presidential SOTU addresses in the same period as the parallel corpora excluding 2017, i.e. addresses by two presidents: George W. Bush (2003–2007) and Barack H. Obama (2013–2016), which serve as baselines to tease apart register- and interpreting-induced effect.

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 distribution of ten features across 200-word samples in 110 1000-word text samples taken from the corpus he used in the MD analysis (1988; see Table 1), finding common features such as nouns and prepositions to follow a linear and stable distribution, whereas rare features, for instance, conditional subordination and relative clauses, a curvilinear distribution, i.e. their occurrences showed a sharp increase in the first 200 words but a gradual decay thereafter. He 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" (Biber, 1993, p. 251), a 600-word segment can be 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 5000 tokens and found almost identical patterns between them

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

(pp. 34–36). To extend MD analysis to the interpreting scenario, I decided on a conservative length of 1000 tokens for each sample in the NE corpus[5].

With respect to Mandarin, such guidelines of "optimal length" seem absent from the literature. Following Biber (1993), I compared the distribution of 20 features across 200-word segments of OC texts. It was found that such features as *shì* 'be', first-person pronouns, and nouns had a linear distribution, whereas imperfect aspect markers (e.g. durative *zhe* and progressive *zài*), third-person pronouns, and amplifiers (e.g. *fēicháng* 'very' and *jídù* 'extremely') showed a curvilinear distribution, with very few being added after 600 or 800 words (see Appendix A for an illustration). Further, I ran t-tests and Mann-Whitney *U* tests, as appropriate, to compare dimension scores obtained from 800- (N = 57) and 2000-word segments (N = 24) of OC data and found no significant differences between the two groups. In that light, OC texts were segmented into 800-word samples, and care was taken not to split or mix Premiers Wen and Li's utterances. The process resulted in 57 files (mean length = 814.07 words, SD = 52.21). IE files contain corresponding interpretation, which, coincidentally, have a comparable token length distribution (M = 1061.91, SD = 143.74) with NE files (M = 986.37, SD = 79.62). 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	19,302
		Li	33	26,989
	IE	IE1	24	27,259
		IE2	33	33,270
Comparable	NE	Bush	31	29,803
		Obama	31	31,355
		Total	176	167,978

As shown in Table 2, interpreters who worked for Premier Wen at the events and their interpretation were coded as "IE1" whilst those for Premier Li as "IE2". Dimension and feature distribution differences between the two premiers and

[5] T-tests and Mann-Whitney *U* tests revealed no significant differences between dimension scores obtained from 1000- and 2000-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).

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 1–2 and 4), the focus shall be placed on between-premier and within-IE differences.

4.3 Data analysis

Dimension scores and standardised feature frequencies of OC were generated utilising MulDi Chinese, and those of IE and NE using the English tagger. Shapiro-Wilk tests and Levene's test were used to assess the normality and variance equality 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 significant effect of speaker/interpreter group (i.e. Premiers Wen and Li; IE1 and IE2; Presidents Bush and Obama) on response variables. Regarding variables that were normally distributed but did not show equal variance between groups, Welch's t-tests were used. With respect to variables that were normally distributed and had equal variance, independent samples t-tests were used. 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. All statistical tests, including hierarchical agglomerative cluster analysis in Section 5.4, were conducted using Python version 3.6.3 (Rossum, 1995). Owing to space constraints, statistical results concerning features are relegated to Appendices B, C and D, grouped by their primary loading dimension.

5 Findings

5.1 Dimensions of register variation in OC

MulDi Chinese enables a description of the SU in the events under investigation, which can provide an idea of what equivalent IE utterances should look like. The analyses herein focus on dimensions where statistically significant differences between two OC speakers, i.e. Premiers Wen Jiabao and Li Keqiang, were found. Table 3 shows mean/median and standard deviation/IQR of dimension scores for each premier's speech and the results of t-tests or Mann-Whitney *U* tests for significant differences between premiers.

TABLE 3: Chinese dimension scores by premier.

Chinese dimension	**Premier Wen**	**Premier Li**	**Results**
Dimension 1: Orality–literacy	*M* and [*SD*]	*M* and [*SD*]	$t(55) = -2.31$, $p = 0.025$, $d = -0.62$, 95% CI [−1.38, 1.38]
	−1.80 [5.35]	1.31 [4.76]	
Dimension 3: Explicit evaluation	*Mdn* [*IQR*]	*Mdn* [*IQR*]	$U = 100.00$, $p < 0.001$, $r_{rb} = 0.75$
	−4.31 [7.28]	2.35 [7.93]	

Chinese Dimension 1 represents the “oral/literate opposition”, a “universal” dimension of register variation (see Section 3). This dimension (see Appendix B for full feature structure and Figure 1 for register distribution along this dimension) is interpreted as marking a personal colloquial focus in that the oral written register humour received the highest scores herein. Independent samples t-test revealed a significant difference between Premiers Wen and Li and an association of Premier Li’ 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, phrasal coordination, and nominalisation[6] (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.

(1)	我	對	中國	**農村**	的	***改革***	**和**	***發展***	是	有	***長遠***	***考慮***	***的，***	**可以**	**劃分**	為	兩	個	**階段**。
	I	to	China	rural area	DE	reform	and	development	am	have	long-term	consideration	DE,	can	classify	as	two	CLF	stages
	I	do	have	long-term	considerations	about	reform	and	development.	It	can	be	classified	into	two	stages.	The	first	is	

[6] 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), e.g. *kǎolǜ* 考慮 ‘consider’ 的 DE in (1).

(1)	第一	個	**階段**	就	是	**實行**	了	**家庭**	***經營***	***的***	**基本**	**經濟**	**制度，**	給	農民	**以**	***生產***	***經營***	***的***	**自主權……**
	First	CLF	stage	ADV	is	implement	PFV	family	management	DE	fundamental	economic	institution	give	farmers	PREPOSITION	production	management	DE	autonomy
	to	implement	the	fundamental	economic	institution	of	household	contract	responsibility	system.	Farmers	were	granted	the	autonomy	to	carry	out	production…

(1) is highly nominal and laced with ten nominalisations (highlighted in italics) as well as a phrasal connector *hé* 'and'. Conversely, Premier Li's utterances are significantly more focused on colloquial expression than his predecessor.

Chinese Dimension 3 is assessed to reflect explicit attitudes by text producers because it consists of positive features (see Appendix B and Figure 3) 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 conjuncts and demonstrative pronouns. There is a highly significant effect for speaker, as Premier Wen's utterances showed a lack of quantifiers (see Table B-1), indicative of an absolute and vague register (Biber, 1988, p. 108). In contrast, Premier Li expressed significantly more personal stance and affect than his predecessor, as is evident in (2), an utterance also about rural reform reproduced from his speech in 2014. Positive features on Chinese Dimensions 1 and 3 are in bold.

(2)	後來	搞	承包制，	放開	搞活，	農民	自己	決定	幹	什麼、	怎麼	種，	幾	年	時間	溫飽	問題	解決	了。
	time	DO	contract system	liberate	relax	farmers	self	decide	DO	what	how	plant	several	CLF	time	food and clothing	problem	solved	PRF
	Later	came	contract system,	relax	and	liberate,	farmers	decided for	themselves	what and	how to	plant,	just in	a few	years	food and clothing	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 literacy/orality and explicit evaluation. Premier Wen's utterances were assessed to be emblematic of literate discourse whereas Premier Li tended to focus on oral 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 wherein 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* and [*SD*]	*M* and [*SD*]	$t(117) = -3.97$, $p <$ 0.001, $d = -0.72$, 95% CI [−6.72, −4.98]
	−7.57 [3.93]	−4.26 [5.03]	
Dimension 3: Explicit versus situation-dependent reference	*M* and [*SD*]	*M* and [*SD*]	$t(117) = 5.14$, $p < 0.001$, $d = 0.94$, 95% CI [5.60, 6.65]
	7.43 [2.37]	4.93 [2.87]	
Dimension 5: Abstract versus non-abstract information	*Mdn* and [*IQR*]	*Mdn* and [*IQR*]	$U = 2365.5$, $p = 0.001$, $r_{rb} = -0.34$
	0.37 [2.62]	−0.9 [2.1]	

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 11 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 and Table C-1 reveals IE's strong preference for "explicitness below the clausal level" (Kruger and van Rooy, 2016, p. 41) owing to significant over-representations of phrasal coordination (*and* as a phrasal coordinator) and nominalisation than NE (consistent with Li and Wang, 2012), whilst 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, agentless passives, and past participial WHIZ deletion relatives (e.g. *enjoyed* in "they have safeguarded well the freedom and rights enjoyed by Hong Kong compatriots", IE1 2003)[7].

To summarise, this section foregrounds symptoms of TU independent of the register at hand, i.e. oral performance of scripted political presentation. As IE was found to be significantly more informational, explicit, and abstract than NE, the effect of register expectations seems restricted to dimensions not presented herein, i.e. Biber's (1988) Dimensions 2, 4 and 6, 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 shifts between OC and IE along such scales and the degree to which such shifts, if any, is attributable to the divergent styles between Premiers Li and Wen (Section 5.1) or to interpreters.

5.3.1 Oral and personal scale

Both Chinese and English Dimension 1 represent the oral/literate opposition and comprise positive features such as verbs, pronouns, WH-words/clauses,

---

[7] MulDi Chinese, in its current shape, did not incorporate dependency parsing, and thus a direct comparison between Chinese and English agentless passives is not possible.

questions, and negation (see Appendices B and C). The 15 analogous registers have similar score distribution herein, which is plotted in Figure 1 in descending order of their mean scores from top to bottom.

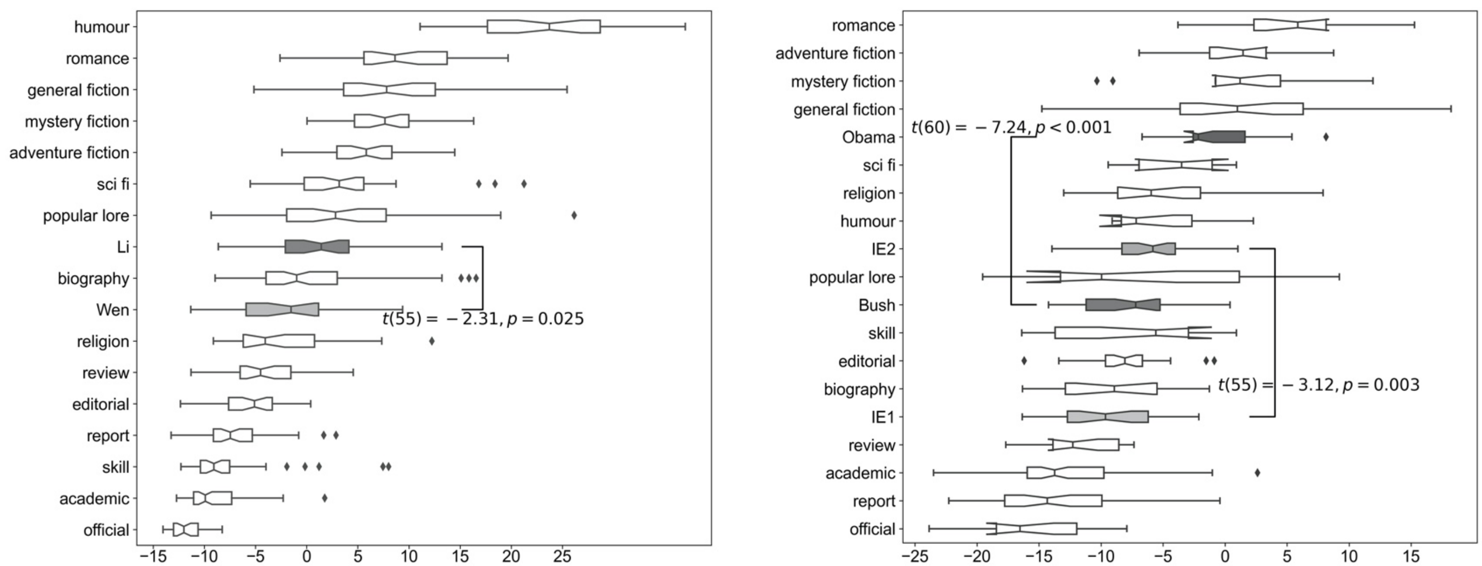


FIGURE 1: Chinese and English Dimension 1 score distribution.

As can be seen from Figure 1, both Chinese and English popular lore was measured as being more involved and oral than biography writing, whereas official, academic and report writing hardly personal (although see Ji, 2017, pp. 86–87 for a discussion of potential inherent differences between Chinese and English humour samples in Brown family corpora). Thus, there is substantial evidence of similarities in positive poles of Chinese and English Dimension 1, which reveals a tentatively named "oral and personal scale" in both languages.

Figure 1 shows a significant difference between IE1 (M = −9.34, SD = 4.25) and IE2's (M = −6.29, SD = 3.17) position on this scale, $t(55) = -3.12$, $p = 0.003$, $d = -0.84$, 95% CI [−8.61, −6.53], with IE2 being more personal and interactional than IE1. As can be seen from Figure 1, whilst Premier Wen's utterances were assessed along Dimension 1 to be akin to the moderately oral register religion (M = −2.25, SD = 5.26), IE1 was on par with the less oral English biography (M = −8.83, SD = 5.04), a surprising finding given the "equalising effect" on literate speech (see Section 2.1; and Sections 5.3.2 and 5.4 for a discussion of potential inherent differences between Chinese and English biography samples). Conversely, Premier Li's speech was more oral than biography (M = 0.52, SD = 5.83) whereas IE2 resembles popular lore (M = −6.73, SD = 9.93), and as the analogous register of English popular lore, i.e. Chinese popular lore, is similar to Premier Li's speech (see Figure 1), oral and personal functional shifts from Premier Li's speech to IE2, if any, may be very small.

An examination of shifts of co-occurring variables loading on the dimensions considered herein reveals the strength of the "equalising effect". With respect to literate speech, Table B-1 shows Premier Wen's statistically significant preference for indefinite pronouns (e.g. *shuí* 'who' and *yǒurén* 'someone') compared with Premier Li, but such pronouns were almost completely ignored in IE1 and IE2 with no significant differences between them (see Table D-3). Table B-2 shows Premier Li's penchant for WH-words (e.g. *zěnme* 'how' and *wèishíme* 'why'; note that the two premiers did not differ significantly in their question use as shown by Table B-3), nonetheless, IE1 demonstrates an "overuse" of direct WH-questions relative to IE2 (see Table D-1). No significant differences between Premiers Wen and Li's first-person pronoun use were found (see Table B-3), but IE1 shows significantly higher frequencies of first-person pronouns than IE2 (see Table D-1), as illustrated by the interpretation of (1) in (3). Positive features on English Dimension 1 are in bold whilst nominalisation 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*. (IE1 2005)

In (3) the interpreter added **we** compared with (1) and "wandered" from **I** to **we**, indicative of a strong identification with Premier Wen's statements (Beaton-Thome, 2010, pp. 130–134). Taken together, the "greater tendency towards orality" in the interpretation of literate speech is realised by avoiding ambiguity (indefinite pronouns) and strengthening speaker statements and affect expression (first-person pronouns and direct WH-questions).

In the case of oral-type speech, apart from avoiding indefinite pronouns and WH-questions, IE2 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 (see Table D-3). Overall, the "tendency towards orality" exerts a stronger effect on Premier Wen's literate speech than Premier Li's oral speech (Cf. Kajzer-Wietrzny, 2015).

5.3.2 Literate and informational scale

A scrutiny of Chinese Dimensions 1, 4, and 5, with the latter two measuring different facets of literacy, and English Dimensions 1 and 3 suggests a comparable "literate and informational scale" as they share a reliance on average word length, attributive/auxiliary adjectives, nouns, prepositions, nominalisation, and phrasal coordination (see Appendices B and D), which marks "a literate style" of "high informational density and precise word choice" (Biber, 2014, pp. 12–32). The validity of such a scale is further borne out by the similar dimension score distribution of reference registers, as exemplified by Figures 1 and 2, such as official, review, and academic registers' low scores on Chinese and English Dimension 1 and high scores on English Dimension 3.

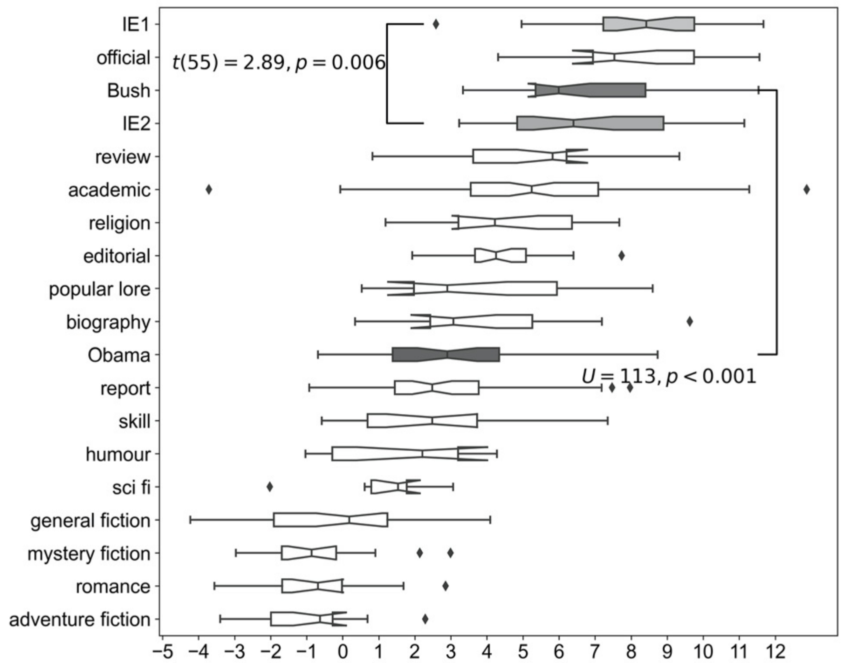


FIGURE 2: English Dimension 3 score distribution.

Regarding the degree of literacy, Premier Wen's utterances were in league with religion along Chinese Dimension 1, and skills/hobbies writing on both Chinese Dimensions 4 and 5 (on Dimension 4, $M = 0.28$, $SD = 3.49$ for Wen and $M = 0.53$, $SD = 3.01$ for skills; on Dimension 5, $M = -0.41$, $SD = 3.94$ for Wen and $M = -1.09$, $SD = 2.33$ for skills), whilst his successor Premier Li's speech was similar to biography on Dimension 1, editorial on Dimension 4 ($M = -0.20$, $SD = 3.08$ for Li; $M = -0.28$, $SD = 2.07$ for editorial), and review on Dimension 5 ($M = 0.30$, $SD = 2.29$ for Li; $M = 0.65$, $SD = 1.24$ for review). Referring to Figure 1, IE1 and IE 2, respectively, were akin to comparatively more oral registers biography and popular lore. It appears that interpreters have rendered both premiers' speech less literate and informational, but three caveats to this observation are (a) dimension scores were obtained by adding z-scores of standardised frequencies of positive features and subtracting z-scores of negative features (Biber, 1988, p. 81), and thus when IE1 was shifted towards involved production (Section 5.3.1), lower scores on information production were generated; (b) there can be inherent differences between some Chinese and English registers as Chinese biography is assessed to be personal but its English counterpart rather informational (see also the discussion in Ji, 2017, pp.

84–87; and Section 5.4 herein); (c) an examination of potential informational shifts should be conducted based on fine-grained annotated information units rather than a lexical-grammatical analysis utilising the MD approach.

English Dimension 3 can further complement the analysis of the literate and informational scale. An independent samples t-test reveals a significant difference between IE1 (M = 8.43, SD = 2.13) and IE2 (M = 6.70, SD = 2.30) scores herein, $t(55) = 2.89$, $p = 0.006$, $d = 0.78$, 95% CI = [6.80, 8.05]. As can be seen from Figure 2, IE1 is more "explicit" than official writing (M = 7.91, SD = 2.23) whereas IE2 more so than review (M = 5.36, SD = 2.36), which can be explained by the finding that Premier Wen's speech is more literate than Premier Li's (Section 5.1). Such a judgment is validated by local comparisons: Premier Wen's statistically significant preference for phrasal coordination, longer words, and nominalisation may well spill over to his interpreters, with IE1 showing significantly higher frequencies of phrasal coordination, nouns (but not nominalisation), longer words, and larger type-token ratio than IE2 (see Tables D-1 and D-3). The spill-over can be exemplified by a comparison of (1) in Section 5.1 and (3) in Section 5.3.1: ten nominalisations and one phrasal coordination were used in the original segment (1), whilst four nominalised words and two phrasal coordination were utilised in its interpretation (3), with the additional *and* being used to connect original noun sequence *shēngchǎn* 'production' *jīngyíng* 'management'. Still, fewer nominalisations were used in the interpretation than original speech (Cf. Li and Wang, 2012). As nominalisations characterise formal registers (Kruger and van Rooy, 2016, p. 41), such a pattern further suggests a shift towards informality and orality in the case of literate SU. Granted, such a result should be understood in light of the encoding differences between Chinese and English nominalisations: verb plus DE, which occurred three times in (1), constitutes nominalisation (see footnote 6) in Chinese, but such a pattern does not exist in English.

The case of nominalisation in "oral-type" SU (Premier Li's speech) is interesting because Premier Wen used more nominalisations than Premier Li, but no significant differences were found between IE1 and IE2 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 IE2 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 packing verbs with nominalisation.

5.3.3 Attitudinal scale

An "attitudinal scale" can be generalised from English Dimension 1 and Chinese Dimension 3, as both comprise stance-focused devices including amplifiers, emphatics, hedges, and *be* (Biber, 1988, p. 106). Register distribution along Chinese Dimension 3 is plotted in Figure 3. Coupled with Figure 1, they show higher scores of fictions on the attitudinal scale in both Chinese and English and low scores of official and report writing.

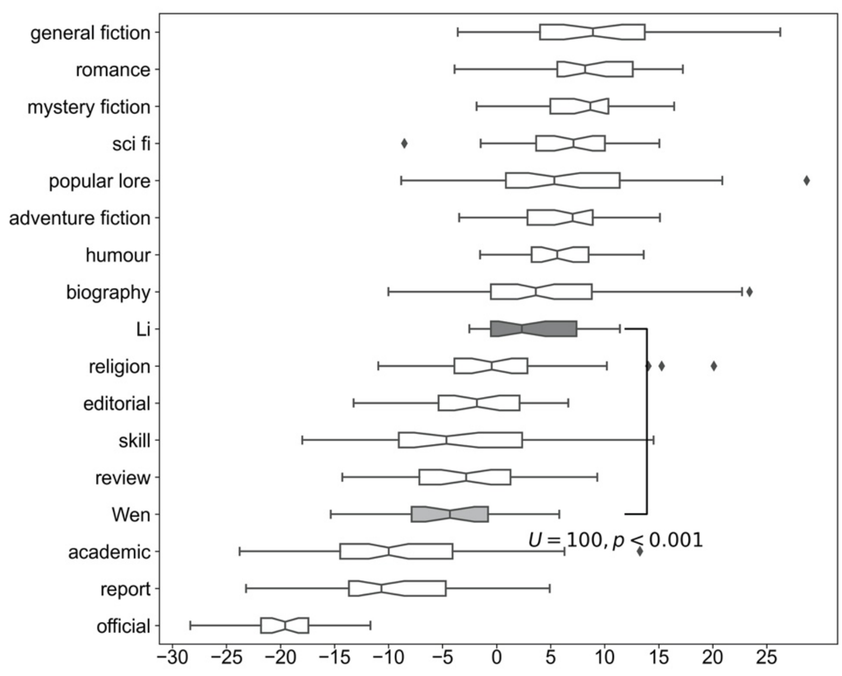


Figure 3: Chinese Dimension 3 score distribution.

Along Chinese Dimension 3, Premier Wen's speech was less opinionated than press review (*Mdn* = −2.82, *IQR* = 8.43) whilst Premier Li's as explicit as biography writers (*Mdn* = 3.64, *IQR* = 9.34). A comparison of Figures 1 and 3 shows that original attitudinal conveyance may have been levelled up as English biography and popular lore are more stance-focused than Chinese review and biography, respectively. The equalising effect also plays out on oral and literate speech types to different extents, as significantly higher frequencies of amplifiers, hedges, downtoners, and *shì* 'be' were found in Premier Li's speech than Premier Wen's speech (see Table B-2), but IE1 used significantly more amplifiers and downtoners than IE2 (see Table D-1). No significant differences for *be* and hedges were found between IE1 and IE2 (see Table D-3). More interestingly, the occurrences of hedges were close to zero in both interpretation and NE (see Table C-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 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 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). The results of such a process are visualised in the dendrograms of Figure 4.

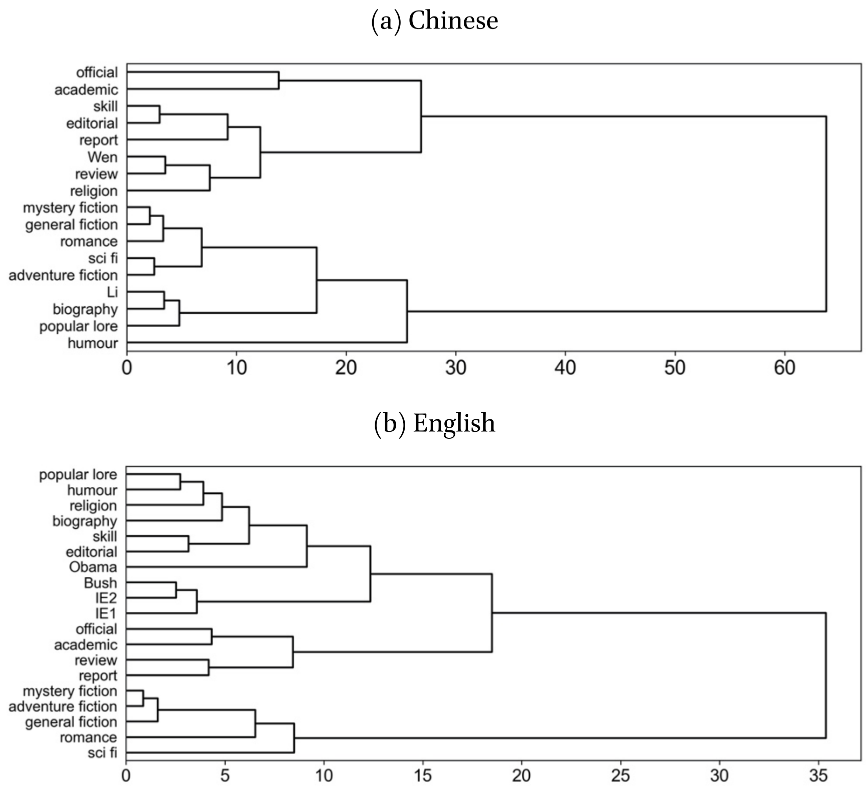


FIGURE 4: Hierarchical agglomerative clustering of OC, IE, NE, and reference registers.

Figure 4 shows the gradual merge of small clusters of registers/varieties that are closest together, such as official and academic registers in both Chinese and English, into one large cluster displaying the register system. The 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 clusters. The most visible distinction in Figure 4 lies in non-fiction versus fiction clusters as (a) shows Chinese non-fiction cluster to contain official, academic, skills/hobbies, editorial, report, Premier Wen’s utterances, review, and religion, whereas Premier Li’s utterances belong to the fiction cluster and are most similar with biography. English non-fiction cluster in (b) encompasses the aforementioned reference registers along with popular lore, biography, and humour, which can be caused by systematic or sampling differences between Chinese and English popular lore, biography and humour registers (Ji, 2017, pp. 86–87). Most importantly, the non-fiction cluster includes both IE1 and IE2 and Presidents Bush and Obama’s speech.

In other words, Figure 4 points to that internal heterogeneity of fiction/non-fiction clusters in both languages being considered, IE2 may have shifted

Premier Li's speech from a fictional to non-fictional register, which likely results from downplaying indefinite and second-person pronouns, WH-questions (Section 5.3.1), amplifiers, hedges, downtoners, and *be* (Section 5.3.3), and nominalising verbs (Section 5.3.2). This research thus provides, for the first time, concrete evidence for the features of oral-type source language 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 interpretation IE1 and IE2 are highly similar to each other. This observation can be supported by the pair-wise Euclidean distance values (see Appendix E for the distance matrices), as Premier Wen's speech was measured to be most similar with review ($d_{Wen,review}$ = 3.51) whilst Premier Li's speech with biography ($d_{Li,biography}$ = 3.41; note that $d_{Wen,Li}$ = 8.75). In contrast, IE1 and IE2 are closest to President Bush's speech ($d_{IE1,Bush}$ = 2.86, $d_{IE2,Bush}$ = 2.53; Cf. $d_{IE1,Obama}$ = 10.08 and $d_{IE2,Obama}$ = 6.89), which testifies in part to the effect of target register expectations (see Section 5.2 and Figures 1-2). More interestingly, the next closest register of IE1 and IE2 is each other ($d_{IE1,IE2}$ = 3.79) and their third most similar register is editorial ($d_{IE1,editorial}$ = 5.01, $d_{IE2,editorial}$ = 4.5). 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 have strengthened the salience of 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-lingual multidimensional approach and a case study of political conference interpreting, the interpreters concerned are demonstrated to have shifted literate SU to a more oral, attitudinal, and less formal register and oral SU a less oral, more attitudinal, and formal register. This research shows for the first time the "equalising effect" on the register level as regardless of SU types, the interpreting products tend to be more similar to each other than SU registers.

A 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 English variety. Source speech was indicated to license, to some extent, the manifestation of extensive phrasal coordination and nominalisation in IE. The effect of interpreting is that ambiguity is avoided, and formal equivalents are preferred in this mode. The strength of the "equalising effect" varies according to SU types, as in the case of literate SU, more WH-questions, first-person pronouns, amplifiers, and downtoners, as well as fewer nominalisations were used in the interpretation, whereas for oral SU, fewer WH-questions, second-person pronouns, amplifiers, and downtoners, 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 speech 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 equalising effect concerning register 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 would hence be valuable to test the MD analyses in the inverse direction (English–Chinese; see Shlesinger, 1995) and among different interpreter groups. Future replication on larger and SI corpora is also desirable. The inclusion of spoken Chinese corpora can aid the improvement of MulDi Chinese, and the findings of this work should also be integrated with processed-oriented studies.

Bibliography

Baker, M. (1996) ‘Corpus-based translation studies: The challenges that lie ahead’, in Sager, J. 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, N., Meyer, B. and Özçetin, D. (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, M. (2007) ‘Interpreted Ideologies in Institutional Discourse’, *The Translator*, 13(2), pp. 271–296. doi: 10.1080/13556509.2007.10799241.

Beaton-Thome, M. (2010) ‘Negotiating Identities in the European Parliament: The Role of Simultaneous Interpreting’, in Baker, M., Olohan, M., and Calzada-Pérez, M. (eds) *Text and Context: Essays on Translation and Interpreting in Honour of Ian Mason*. Manchester: St. Jerome, pp. 117--138.

Berk-Seligson, S. (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, D. (1988) *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. (1993) ‘Representativeness in Corpus Design’, *Literary and Linguistic Computing*, 8(4), pp. 243--258.

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

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

Biber, D. and Conrad, S. (2019) *Register, Genre, and Style*. 2nd edn. Cambridge: Cambridge University Press.

Chen, L. (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.

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

Friginal, E. and Biber, D. (2016) 'Multi-Dimensional Analysis', in Egbert, J. and Baker, P. (eds) *Triangulating Methodological Approaches in Corpus-Linguistic Research*. New York, NY: Routledge, pp. 73–89.

Gries, S. T. (2008) 'Exploring variability within and between corpora: some methodological considerations', *Corpora*, 1(2), pp. 109–151. doi: https://doi.org/10.3366/cor.2006.1.2.109.

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

Hong, L. and Wang, B. (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, K. and Tao, Q. (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: 11 September 2019).

Ji, M. (2017) 'A multidimensional analysis of the translational Chinese genre system', in Ji, M. et al. (eds) *Corpus Methodologies Explained: An empirical approach to translation studies*. London: Routledge, pp. 53--102.

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

Kreiser, M. and Greene, M. (2020) *History, Evolution, and Practices of the President's State of the Union Address: Frequently Asked Questions*. R44770. Washington, DC: Congressional Research Service, p. 8.

Kruger, H. and van Rooy, B. (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, C. N. and Thompson, S. A. (1989) *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, CA: University of California Press.

Li, D. and Wang, K. (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.

Livingston, S., Singer, B. and Abrahamson, T. (1994) 'Effectiveness Compared: ASL Interpretation vs. Transliteration', *Sign Language Studies*, 1082(1), pp. 1–54. doi: 10.1353/sls.1994.0008.

Lv, Q. and Liang, J. (2019) '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.

Malmkjaer, K. (1998) 'Love thy Neighbour: Will Parallel Corpora Endear Linguists to Translators?', *Meta: Journal des traducteurs*, 43(4), pp. 1–8. doi: 10.7202/003545ar.

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.fmprc.gov.cn/mfa_chn/wjb_602314/gbclc_603848/wjsy_660703/t1080232.shtml (Accessed: 9 September 2019).

Munday, J. (2016) *Introducing Translation Studies: Theories and Applications*. Fourth edition. London and New York: Routledge.

Nini, A. (2018) *Multidimensional Analysis Tagger* (*Version 1.3.1*). Available at: http://sites.google.com/site/multidimensionaltagger.

Pym, A. (2007) 'On Shlesinger's proposed equalizing universal for interpreting', in Pöchhacker, F., Jakobsen, A. L., and Mees, I. (eds) *Interpreting studies and beyond: A tribute to Miriam Shlesinger*. Copenhagen: Samfundslitteratur, pp. 175–190.

Raschka, S. and Mirjalili, V. (2019) *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2, 3rd Edition*. Birmingham, UK: Packt Publishing.

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

Setton, R. (2011) 'Corpus-based interpreting studies (CIS): Overview and prospects', in Kruger, A., Wallmach, K., and Munday, J. (eds) *Corpus-Based Translation Studies: Research and Applications*. London and New York: Contiuum International

Publishing Group, pp. 33–75.

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

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

Toutanova, K. *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, B. and Zou, B. (2018) 'Exploring Language Specificity as a Variable in Chinese-English Interpreting: A Corpus-Based Investigation', in *Making Way in Corpus-based Interpreting Studies*. Singapore: Springer Singapore, pp. 65–82.

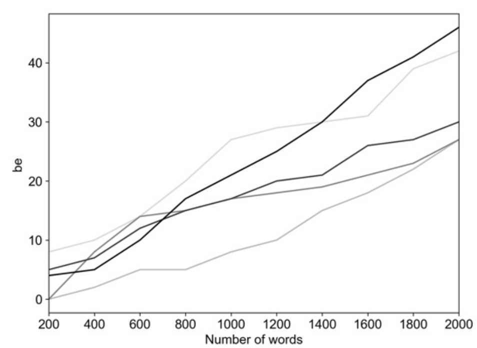
Xiao, R. (2009) 'Multidimensional analysis and the study of world Englishes', *World Englishes*, 28(4), pp. 421–450. doi: 10.1111/j.1467-971X.2009.01606.x.

Yi, Y. and Chang, T. K. (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.

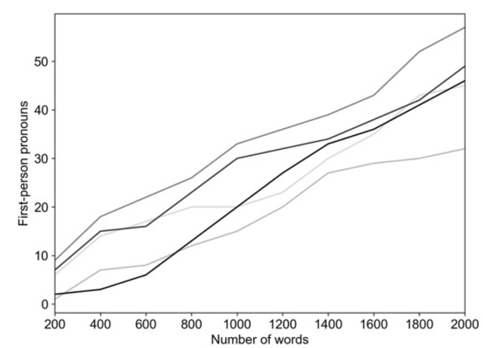
Appendices

Appendix A: Distribution of six features in five 2000-word samples of OC

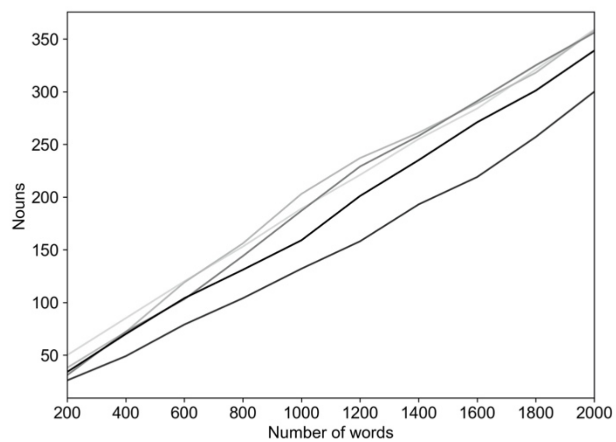
(a) *shì* 'be'



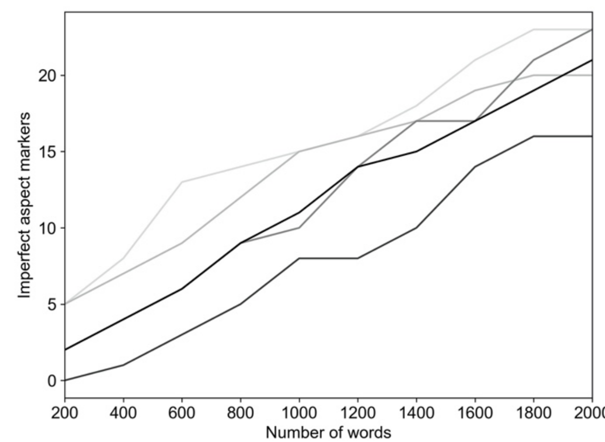
(b) First-person pronouns



(c) Nouns



(d) Imperfect aspect markers



(e) Third-person pronouns

(f) Amplifiers

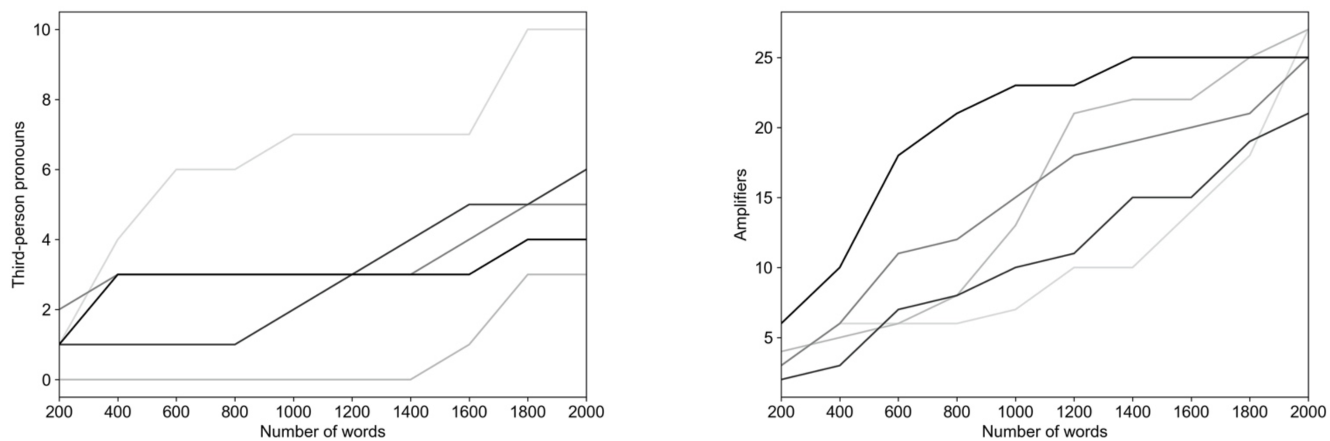


FIGURE A: Distribution of six features in five 2000-word samples of OC.

Appendix B: Individual feature difference patterns of OC

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

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

		M and [SD]		**Independent samples/Welch's t-tests**			
Feature	Primary dimension	Wen	Li	*df*	*t*	*p*	*d*
Average word length	Dimension 1 (literate)	1.57 [0.05]	1.51 [0.02]	55	5.3	<0.001	1.423
Phrasal coordination	Dimension 4 (long/abstract writing)	17.23 [5.32]	8.68 [3.73]	55	7.14	<0.001	1.916
Disyllabic words	Dimension 5 (modern writing)	57.58 [16.22]	44.61 [10.08]	55	3.46	0.001	0.928
Auxiliary adjectives	Dimension 5 (modern writing)	7.72 [3.22]	5.43 [3.15]	55	2.67	0.01	0.717

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

		Median and [IQR]		**Mann-Whitney *U* tests**	
Feature	Primary dimension	Wen	Li	*U*	*p*
Indefinite pronouns	Dimension 1 (oral)	3.66 [3.26]	2.3 [2.09]	204.5	0.001
Classical function words	Dimension 5 (classical writing)	3.9 [5.81]	2.38 [2.3]	253	0.011

Nominalisation	Dimension 5 (modern writing)	63.72 [18.68]	52.75 [11.13]	230.5	0.004

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

		M and [SD]		**Independent samples/Welch's t-tests**			
Feature	Primary dimension	Wen	Li	*df*	*t*	*p*	*d*
Public verbs	Dimension 1 (oral)	5.93 [3.04]	8.71 [3.27]	55	-3.27	0.002	-0.877
Hedges	Dimension 3 (explicit evaluative)	3.22 [2.24]	9.44 [4.15]	55	-7.28	<0.001	-1.953
HSK 3	Dimension 3 (explicit evaluative)	141.69 [23.20]	177.91 [14.83]	55	-6.71	<0.001	-1.801
Adverbs	Dimension 3 (explicit evaluative)	55.2 [10.82]	67.29 [8.04]	55	-4.85	<0.001	-1.3
Modifying adverbs	Dimension 3 (explicit evaluative)	17.72 [5.93]	23.82 [4.26]	55	-4.52	<0.001	-1.214
yǒu 'existential *there*'	Dimension 3 (explicit evaluative)	6.25 [3.69]	9.72 [3.8]	55	-3.45	0.001	-0.924
shì 'be'	Dimension 3 (explicit evaluative)	18.92 [5.94]	22.27 [4.16]	38.78	-2.37	0.023	-0.637
Amplifiers	Dimension 3 (explicit evaluative)	13.87 [5.21]	17.07 [5.27]	55	-2.27	0.027	-0.61
Intransitive verbs	Dimension 4 (short/concrete writing)	11.47 [4.45]	16.90 [5.37]	55	-4.04	<0.001	-1.08
Lexical density	Dimension 5 (modern writing)	783.82 [36.13]	802.27 [26.28]	55	-2.23	0.030	-0.599

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

		Median and [IQR]		**Mann-Whitney *U* tests**	
Feature	Primary dimension	Wen	Li	*U*	*p*

Second-person pronouns	Dimension 1 (oral)	1.76 [2.68]	3.82 [2.25]	230.5	0.004
WH-words	Dimension 1 (oral)	0.0 [0.0]	1.19 [1.27]	218	0.001
Modal particles	Dimension 1 (oral)	16.06 [12.56]	36.23 [13.08]	28	<0.001
Downtoners	Dimension 3 (explicit evaluative)	1.21 [1.31]	2.37 [3.89]	264.5	0.016

TABLE B-3: Features of no significant between-premier differences.

		Median and [IQR]		Mann-Whitney *U* tests	
Feature	Primary dimension	Wen	Li	*U*	*p*
Emotion words	Dimension 1 (oral)	10.9 [5.39]	9.13 [3.71]	329	0.141
Honorifics	Dimension 1 (oral)	1.18 [2.48]	0.0 [1.29]	319.5	0.09
Interrogatives	Dimension 1 (oral)	1.15 [3.2]	2.34 [3.39]	331.5	0.145
Mono-negation	Dimension 1 (oral)	8.4 [6.75]	7.57 [6.01]	394	0.49
Questions	Dimension 1 (oral)	0.0 [2.4]	1.17 [2.34]	370	0.331
Exclamations	Dimension 1 (oral)	0.0 [0.0]	0.0 [0.0]	364.5	0.248
Conditional conjuncts	Dimension 3 (explicit evaluative)	2.48 [3.14]	1.36 [1.97]	0.723	0.473
Demonstrative pronouns	Dimension 3 (explicit evaluative)	14.28 [5.32]	17.65 [7.49]	302.5	0.066
Disyllabic negation	Dimension 3 (explicit evaluative)	1.28 [2.37]	1.26 [2.39]	376	0.374
Other personal pronouns	Dimension 3 (explicit evaluative)	1.33 [4.28]	2.18 [3.82]	356	0.26
Disyllabic prepositions	Dimension 4 (long/abstract writing)	2.98 [3.57]	2.39 [2.49]	308.5	0.079
Classical syntax	Dimension 5 (classical writing)	3.76 [3.51]	3.75 [4.76]	368.0	0.328

TABLE B-3 (CONT.): Features of no significant between-premier differences.

		M and [SD]		Independent samples t-tests	
		Wen	Li	$t(55)$	*p*

First-person pronouns	Dimension 1 (oral)	24.77 [9.64]	23.19 [6.35]	0.747	0.458
Private verbs	Dimension 3 (explicit evaluative)	13.20 [5.15]	3.69 [3.95]	-0.409	0.684
HSK 1	Dimension 3 (explicit evaluative)	240.13 [22.93]	235.71 [23.45]	0.708	0.482
Nouns excl. nominalisation	Dimension 3 (inexplicit un-evaluative)	183.71 [20.00]	176.24 [17.97]	1.478	0.145
Unique words ratio	Dimension 4 (short/concrete writing)	297.1 [45.84]	307.59 [27.79]	-1.241	0.22
Disyllabic verbs	Dimension 4 (short/concrete writing)	45.24 [11.63]	41.97 [8.26]	1.243	0.219

Appendix C: Individual feature difference patterns of IE versus NE

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

		M and [SD]		**Independent samples/Welch's t-tests**			
Feature	Primary dimension	IE	NE	*df*	*t*	*p*	*d*
Attributive adjectives	Dimension 1 (informational)	8.59 [1.38]	6.91 [1.31]	117	6.8	<0.001	1.248
Average word length	Dimension 1 (informational)	4.8 [0.14]	4.63 [0.19]	110.53	5.39	<0.001	0.988
Private verbs	Dimension 1 (involved)	1.33 [0.5]	1.01 [0.43]	117	3.75	<0.001	0.689
Be as main verb	Dimension 1 (involved)	1.49 [0.45]	1.24 [0.5]	117	2.87	0.005	0.527
Adverbs	Dimension 3 (situational)	3.16 [0.63]	2.85 [0.84]	112.33	2.29	0.024	0.42

Phrasal coordination	Dimension 3 (explicit)	1.45 [0.51]	1.23 [0.59]	117	2.17	0.032	0.398

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

		Median and [IQR]		Mann-Whitney *U* tests	
Feature	Primary dimension	IE	NE	*U*	*p*
Amplifiers	Dimension 1 (involved)	0.27 [0.34]	0.0 [0.11]	408	<0.001
That deletion	Dimension 1 (involved)	0.21 [0.18]	0.11 [0.2]	1205.5	0.001
Demonstrative pronouns	Dimension 1 (involved)	0.56 [0.29]	0.46 [0.4]	1421	0.033
Nominalisation	Dimension 3 (explicit)	4.81 [1.39]	2.88 [1.42]	342.5	<0.001
Conjuncts	Dimension 5 (abstract)	0.18 [0.26]	0.0 [0.12]	820.5	<0.001
Predicative adjectives	Dimension 5 (abstract)	0.64 [0.42]	0.44 [0.5]	1275.5	0.004
Agentless passives	Dimension 5 (abstract)	0.6 [0.52]	0.46 [0.45]	1394	0.024
Past participial WHIZ deletion relatives	Dimension 5 (abstract)	0.11 [0.21]	0.11 [0.12]	1457	0.046

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

		M and [SD]		Independent sample t-tests		
Feature	Primary dimension	IE	NE	$t(117)$	*p*	*d*
Total other nouns excl. gerunds and nominalisation	Dimension 1 (informational)	20.73 [1.79]	23.47 [2.02]	-7.808	<0.001	-1.433

Present tense verbs	Dimension 1 (involved)	5.69 [0.9]	6.15 [0.87]	-2.882	0.005	-0.529
Possibility modals	Dimension 1 (involved)	0.47 [0.28]	0.61 [0.33]	-2.55	0.012	-0.468
Place adverbials	Dimension 3 (situational)	0.27 [0.17]	0.35 [0.21]	-2.255	0.026	-0.414

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

		Median and [IQR]		**Mann-Whitney *U* tests**	
Feature	Primary dimension	IE	NE	*U*	*p*
TTR	Dimension 1 (informational)	389.0 [50.0]	406.0 [38.0]	1272.5	0.004
Independent clause coordination	Dimension 1 (involved)	0.69 [0.52]	1.7 [0.8]	335.5	<0.001
Indefinite pronouns	Dimension 1 (involved)	0.0 [0.0]	0.0 [0.11]	1387	0.002
Because	Dimension 1 (involved)	0.1 [0.15]	0.12 [0.24]	1247.5	0.002
First-person pronouns	Dimension 1 (involved)	4.13 [1.44]	5.01 [1.86]	1283.5	0.005
Pro-verb *do*	Dimension 1 (involved)	0.12 [0.2]	0.14 [0.28]	1293.5	0.005
Emphatics	Dimension 1 (involved)	0.54 [0.38]	0.74 [0.66]	1332	0.01
Contractions	Dimension 1 (involved)	0.21 [0.22]	0.44 [0.98]	1356.5	0.014
Stranded prepositions	Dimension 1 (involved)	0.0 [0.0]	0.0 [0.12]	1440.5	0.018
Analytic negation	Dimension 1 (involved)	0.43 [0.41]	0.5 [0.53]	1439.5	0.041
Time adverbials	Dimension 3 (explicit)	0.24 [0.25]	0.57 [0.48]	479.5	<0.001
WH clauses on subject positions	Dimension 3 (explicit)	0.0 [0.1]	0.21 [0.21]	833.5	<0.001

Past participial clauses	Dimension 5 (abstract)	0.0 [0.0]	0.0 [0.0]	1473	0.004
Other adverbial subordinators	Dimension 5 (abstract)	0.11 [0.24]	0.21 [0.13]	1452	0.046
Downtoners	Other	0.15 [0.25]	0.22 [0.22]	1404.5	0.026

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

		Median and [IQR]		**Mann-Whitney *U* tests**	
Feature	Primary dimension	IE	NE	*U*	*p*
Prepositions	Dimension 1 (informational)	9.81 [1.5]	9.91 [1.62]	1618.5	0.216
WH clauses	Dimension 1 (involved)	0.0 [0.0]	0.0 [0.11]	1554.5	0.081
Sentence relatives	Dimension 1 (involved)	0.0 [0.1]	0.0 [0.0]	1570.5	0.1
Hedges	Dimension 1 (involved)	0.0 [0.0]	0.0 [0.0]	1668.5	0.128
Second person pronouns	Dimension 1 (involved)	0.55 [0.37]	0.46 [0.71]	1597	0.184
Direct WH-questions	Dimension 1 (involved)	0.0 [0.0]	0.0 [0.0]	1675	0.184
Discourse particles	Dimension 1 (involved)	0.0 [0.12]	0.0 [0.11]	1629	0.208
Pronoun *it*	Dimension 1 (involved)	0.73 [0.58]	0.74 [0.67]	1758.5	0.483
Pied-piping relative clauses	Dimension 3 (explicit)	0.0 [0.0]	0.0 [0.0]	1722	0.366
WH relative clauses on object positions	Dimension 3 (explicit)	0.0 [0.0]	0.0 [0.0]	1730	0.368
By passives	Dimension 5 (abstract)	0.0 [0.1]	0.0 [0.12]	1673	0.283
Present participial WHIZ deletion relatives	Dimension 5 (abstract)	0.18 [0.28]	0.12 [0.29]	1712	0.385

Appendix D: Individual feature difference patterns of IE1 versus IE2

TABLE D-1: Features of significantly higher frequencies in IE1.

		M and [SD]		Independent samples/Welch's t-tests			
Feature	Primary dimension	IE1	IE2	*df*	*t*	*p*	*d*
Prepositions	Dimension 1 (informational)	10.96 [1.17]	9.6 [0.62]	32.28	5.18	<0.001	1.388
Total other nouns excl. gerunds and nominalisation	Dimension 1 (informational)	21.47 [1.63]	20.2 [1.73]	55	2.81	0.007	0.753
TTR	Dimension 1 (informational)	395.38 [27.26]	378.21 [33.17]	55	2.07	0.043	0.557
Average word length	Dimension 1 (informational)	4.84 [0.14]	4.77 [0.13]	55	2.01	0.05	0.538
Phrasal coordination	Dimension 3 (explicit)	1.83 [0.39]	1.17 [0.39]	55	6.32	<0.001	1.694
Adverbs	Dimension 3 (situational)	3.39 [0.52]	2.99 [0.65]	55	2.44	0.018	0.654

TABLE D-1 (CONT.): Features of significantly higher frequencies in IE1.

		Median and [IQR]		Mann-Whitney *U* tests	
Feature	Primary dimension	IE1	IE2	*U*	*p*
Amplifiers	Dimension 1 (involved)	0.3 [0.35]	0.24 [0.21]	277.5	0.028
First-person pronouns	Dimension 1 (involved)	4.28 [1.69]	3.96 [1.04]	279.5	0.03

Direct WH-questions	Dimension 1 (involved)	0.0 [0.0]	0.0 [0.0]	342.5	0.04
Past participial WHIZ deletion relatives	Dimension 5 (abstract)	0.19 [0.14]	0.1 [0.13]	279.5	0.029
Downtoners	Other	0.2 [0.18]	0.11 [0.23]	282.5	0.032

TABLE D-2: Features of significantly higher frequencies in IE2.

		M and [SD]		**Welch's t-tests**		
Feature	Primary dimension	IE1	IE2	$t(52.59)$	p	d
Pronoun *it*	Dimension 1 (involved)	0.64 [0.27]	0.89 [0.48]	-2.5	0.016	-0.671

TABLE D-2 (CONT.): Features of significantly higher frequencies in IE2.

		Median and [IQR]		**Mann-Whitney *U* tests**	
Feature	Primary dimension	IE1	IE2	U	p
Because	Dimension 1 (involved)	0.0 [0.08]	0.11 [0.19]	162	<0.001
Discourse particles	Dimension 1 (involved)	0.0 [0.04]	0.11 [0.22]	226.5	0.001
Demonstrative pronouns	Dimension 1 (involved)	0.43 [0.36]	0.61 [0.34]	211.5	0.001

TABLE D-3: Features of no significant differences between IE1 and IE2.

		Median and [IQR]		**Mann-Whitney *U* tests**	
Feature	Primary dimension	IE1	IE2	U	p

Sentence relatives	Dimension 1 (involved)	0.04 [0.11]	0.0 [0.09]	313	0.061
Stranded prepositions	Dimension 1 (involved)	0.0 [0.1]	0.1 [0.13]	309	0.065
Second person pronouns	Dimension 1 (involved)	0.49 [0.24]	0.58 [0.33]	304	0.07
Independent clause coordination	Dimension 1 (involved)	0.6 [0.55]	0.74 [0.49]	311	0.086
Pro-verb *do*	Dimension 1 (involved)	0.1 [0.2]	0.12 [0.2]	370	0.335
Contractions	Dimension 1 (involved)	0.21 [0.32]	0.21 [0.19]	383	0.42
Analytic negation	Dimension 1 (involved)	0.45 [0.41]	0.4 [0.45]	380	0.401
WH clauses	Dimension 1 (involved)	0.0 [0.08]	0.0 [0.0]	372.5	0.311
Hedges	Dimension 1 (involved)	0.0 [0.0]	0.0 [0.0]	384	0.364
Be as main verb	Dimension 1 (involved)	1.45 [0.6]	1.53 [0.53]	385.5	0.436
Indefinite pronouns	Dimension 1 (involved)	0.0 [0.0]	0.0 [0.0]	381	0.33
That deletion	Dimension 1 (involved)	0.22 [0.26]	0.21 [0.16]	350.5	0.232
WH relative clauses on object positions	Dimension 3 (explicit)	0.0 [0.0]	0.0 [0.0]	357	0.151
WH clauses on subject positions	Dimension 3 (explicit)	0.08 [0.11]	0.0 [0.1]	348.5	0.207
Pied-piping relative clauses	Dimension 3 (explicit)	0.0 [0.0]	0.0 [0.0]	383.5	0.393
By passives	Dimension 5 (abstract)	0.0 [0.15]	0.0 [0.1]	347	0.182
Present participial WHIZ deletion relatives	Dimension 5 (abstract)	0.14 [0.31]	0.2 [0.24]	376.5	0.378
Past participial clauses	Dimension 5 (abstract)	0.0 [0.0]	0.0 [0.0]	387.5	0.369
Conjuncts	Dimension 5 (abstract)	0.18 [0.26]	0.19 [0.28]	382.5	0.417

Other adverbial subordinators	Dimension 5 (abstract)	0.11 [0.16]	0.12 [0.28]	395.5	0.5

TABLE D-3 (CONT.): Features of no significant differences between IE1 and IE2.

		M and [SD]		Independent samples/Welch's t-tests		
		IE1	IE2	*df*	*t*	*p*
Attributive adjectives	Dimension 1 (informational)	8.34 [1.32]	8.77 [1.41]	55	-1.17	0.248
Private verbs	Dimension 1 (involved)	1.25 [0.48]	1.39 [0.52]	55	-1.03	0.308
Emphatics	Dimension 1 (involved)	0.56 [0.23]	0.62 [0.28]	55	-0.88	0.381
Possibility modals	Dimension 1 (involved)	0.43 [0.26]	0.49 [0.3]	55	-0.87	0.386
Present tense verbs	Dimension 1 (involved)	5.7 [1.0]	5.68 [0.84]	55	0.07	0.941
Nominalisation	Dimension 3 (explicit)	4.47 [0.77]	4.91 [1.05]	55	-1.75	0.086
Place adverbials	Dimension 3 (situational)	0.28 [0.2]	0.26 [0.14]	38.54	0.38	0.706
Predicative adjectives	Dimension 5 (abstract)	0.71 [0.3]	0.62 [0.31]	55	1.15	0.256
Agentless passives	Dimension 5 (abstract)	0.57 [0.28]	0.7 [0.3]	55	-1.43	0.16

Appendix E: Euclidean distance matrices of OC, IE, and reference register dimension scores

Most similar register pairs' Euclidean distances are in bold and underlined.

TABLE E-1: Euclidean distance matrix of OC and Texts of Recent CHinese corpus registers.

	academic	adventure fiction	biography	editorial	general fiction	humour	mystery fiction	official	popular lore
academic	0	29.5	22.34	12.39	31.73	38.73	31.04	**<u>13.86</u>**	25.24
adventure fiction	29.5	0	8.43	20.9	4.67	19.54	2.82	38.99	6.23
biography	22.34	8.43	0	12.98	10.51	23.08	9.9	32.23	**<u>3.8</u>**
editorial	12.39	20.9	12.98	0	22.51	31.02	22.11	20.54	15.77
general fiction	31.73	4.67	10.51	22.51	0	16.6	**<u>2.09</u>**	41.22	7.02
humour	38.73	19.54	23.08	31.02	16.6	0	17.35	46.59	20.26
mystery fiction	31.04	2.82	9.9	22.11	**<u>2.09</u>**	17.35	0	40.42	6.84
official	13.86	38.99	32.23	20.54	41.22	46.59	40.42	0	35.07
popular	25.24	6.23	3.8	15.77	7.02	20.26	6.84	35.07	0

lore									
religion	15.54	16.82	9.01	8.46	18.59	27.19	18.15	26.52	11.94
report	12.87	25.35	18.34	8.35	27.66	34.98	26.85	15.73	21.43
review	12.62	18.23	10.59	**5.06**	20.6	29.63	19.84	22.06	13.9
romance	33.52	4.62	12.49	24.68	3.42	**17.02**	2.67	42.72	9.49
sci fi	28.12	**2.5**	6.8	19.19	5.52	21.04	4.3	37.64	4.76
skill	**10.7**	22.6	14.5	2.99	24.53	33.36	24.03	19.42	17.63
Wen	13.99	16.85	10.25	7.09	19.33	27.62	18.41	22.45	13.19
Li	20.3	10.48	**3.41**	11.01	11.8	22.5	11.45	30.23	5.07

TABLE E-1 (CONT.): Euclidean distance matrix of OC and Texts of Recent CHinese corpus registers.

	religion	report	review	romance	sci fi	skill	Wen	Li
academic	15.54	12.87	12.62	33.52	28.12	10.7	13.99	20.3
adventure fiction	16.82	25.35	18.23	4.62	**2.5**	22.6	16.85	10.48
biography	9.01	18.34	10.59	12.49	6.8	14.5	10.25	**3.41**
editorial	8.46	8.35	5.06	24.68	19.19	**2.99**	7.09	11.01
general fiction	18.59	27.66	20.6	**3.42**	5.52	24.53	19.33	11.8
humour	27.19	34.98	29.63	17.02	21.04	33.36	27.62	22.5

mystery fiction	18.15	26.85	19.84	2.67	4.3	24.03	18.41	11.45
official	26.52	15.73	22.06	42.72	37.64	19.42	22.45	30.23
popular lore	11.94	21.43	13.9	9.49	4.76	17.63	13.19	5.07
religion	0	14.01	5.82	20.74	15.59	8.76	7.63	7.74
report	14.01	0	9.09	29.16	23.98	7.88	9.58	17.06
review	**5.82**	9.09	0	22.35	16.8	5.31	**3.51**	9.22
romance	20.74	29.16	22.35	0	6.44	26.61	20.79	14.09
sci fi	15.59	23.98	16.8	6.44	0	20.89	15.69	8.98
skill	8.76	**7.88**	5.31	26.61	20.89	0	8.16	12.89
Wen	7.63	9.58	**3.51**	20.79	15.69	8.16	0	8.75
Li	7.74	17.06	9.22	14.09	8.98	12.89	8.75	0

TABLE E-2: Euclidean distance matrix of IE, NE, and LOB corpus registers.

	academic	adventure fiction	biography	editorial	general fiction	humour	mystery fiction	official	popular lore
academic	0	17.92	7.49	6.99	18.03	10.12	18.16	**4.33**	8.05
adventure	17.92	0	11.54	12.48	**1.62**	8	**0.87**	21.25	10.32

fiction									
biography	7.49	11.54	0	4.66	11.45	4.5	11.63	10.25	3.03
editorial	6.99	12.48	4.66	0	12.39	5.28	12.56	9.44	3.61
general fiction	18.03	1.62	11.45	12.39	0	8.3	1.26	21.19	10.42
humour	10.12	8	4.5	5.28	8.3	0	8.22	13.48	**2.75**
mystery fiction	18.16	**0.87**	11.63	12.56	1.26	8.22	0	21.37	10.46
official	**4.33**	21.25	10.25	9.44	21.19	13.48	21.37	0	11.11
popular lore	8.05	10.32	**3.03**	3.61	10.42	**2.75**	10.46	11.11	0
religion	9.09	10.04	5.41	4.4	10.05	3.95	10.29	12.32	3.34
report	5.99	16.11	5.47	7.2	16.25	8.68	16.2	7.65	7.18
review	5.36	15.82	5.44	6.81	16.02	8.18	15.97	7.46	6.05
romance	22.81	5.64	16.48	16.8	5.19	13.12	5.32	25.9	15.18
sci fi	13.86	5.65	6.65	9	5.44	4.9	5.6	16.83	6.62
skill	6.93	12.49	5.91	**3.16**	12.8	5.05	12.72	10.12	4.06
IE1	7.52	16.49	8.28	5.01	16.32	9.61	16.5	8.03	7.39
IE2	9.37	14.24	8.67	4.5	14.11	8.3	14.3	11.04	6.93
Obama	13.46	9.05	10.02	7.76	9.35	6.51	9.27	16.47	7.36

Bush	9.12	14.52	8	4.3	14.4	8.17	14.49	10.33	6.51

TABLE E-2 (CONT.): Euclidean distance matrix of IE, NE, and LOB corpus registers.

	religion	report	review	romance	sci fi	skill	IE1	IE2	Obama	Bush
academic	9.09	5.99	5.36	22.81	13.86	6.93	7.52	9.37	13.46	9.12
adventure fiction	10.04	16.11	15.82	5.64	5.65	12.49	16.49	14.24	9.05	14.52
biography	5.41	5.47	5.44	16.48	6.65	5.91	8.28	8.67	10.02	8
editorial	4.4	7.2	6.81	16.8	9	**3.16**	5.01	4.5	7.76	4.3
general fiction	10.05	16.25	16.02	**5.19**	5.44	12.8	16.32	14.11	9.35	14.4
humour	3.95	8.68	8.18	13.12	**4.9**	5.05	9.61	8.3	6.51	8.17
mystery fiction	10.29	16.2	15.97	5.32	5.6	12.72	16.5	14.3	9.27	14.49
official	12.32	7.65	7.46	25.9	16.83	10.12	8.03	11.04	16.47	10.33
popular lore	**3.34**	7.18	6.05	15.18	6.62	4.06	7.39	6.93	7.36	6.51
religion	0	9.78	8.17	14.38	7.54	4.88	7.56	5.89	**5.52**	6.63
report	9.78	0	**4.19**	21.26	11.25	7.49	9.18	10.9	13.78	9.59
review	8.17	**4.19**	0	20.91	11.4	6.8	7.56	9.5	12.32	8.27
romance	14.38	21.26	20.91	0	10.53	17.15	20.31	17.66	12.18	18.17

sci fi	7.54	11.25	11.4	10.53	0	9.75	13.15	11.96	9.43	11.67
skill	4.88	7.49	6.8	17.15	9.75	0	6.54	5.61	7.23	5.62
IE1	7.56	9.18	7.56	20.31	13.15	6.54	0	3.79	10.08	2.86
IE2	5.89	10.9	9.5	17.66	11.96	5.61	3.79	0	6.89	**2.53**
Obama	5.52	13.78	12.32	12.18	9.43	7.23	10.08	6.89	0	7.92
Bush	6.63	9.59	8.27	18.17	11.67	5.62	**2.86**	**2.53**	7.92	0