Exploring grammatical colloquialisation in non-native English: a case study of Philippine English[1]

X I N Y U E   Y A O
Renmin University of China

and

P E T E R   C O L L I N S
University of New South Wales
(Received 14 January 2016; revised 14 December 2016)

Colloquialisation, a process by which 'writing becomes more like speech', has been identified as a powerful discourse-pragmatic mechanism driving grammatical change in native English varieties. The extent to which colloquialisation is a factor in change in non-native varieties has seldom been explored. This article reports the findings of a corpus-based study of colloquialisation in Philippine English (PhilE), alongside its 'parent variety', American English (AmE). Adopting a bottom-up approach, a comprehensive measure was derived to determine the degree to which a text prefers grammatical features typical of speech and disprefers those typical of writing. This measure was then used to compare and contrast texts in a parallel, multi-register corpus of PhilE and AmE sampled for the 1960s and 1990s. Evidence for colloquialisation was found to vary across registers. While Philippine press editorials and American fiction show a clear colloquialising tendency, learned writing does not show remarkable changes irrespective of variety. The evolution of PhilE registers cannot be explained by a simple process involving emulation of AmE. The patterns uncovered reflect the uniqueness of the sociohistorical circumstances in which PhilE has evolved.

1 Introduction

As a powerful discourse-pragmatic mechanism of linguistic change, colloquialisation has been explored in many corpus-based diachronic studies of English (e.g. Biber & Finegan 1989; Mair & Hundt 1995; Hundt & Mair 1999; Biber 2003; Leech *et al.* 2009). The term is often used to refer to the shift of writing from a more formal, literary style to a more conversational, speech-like style (Leech *et al.* 2009: 239). Colloquialisation is seen as a driving force in the changing patterns of use of a number of grammatical features in English. In his study of long-term historical drift in English writing styles, Biber comments:

> [I]n the course of the nineteenth and twentieth centuries, popular written registers like letters, fiction, and essays have … evolved to become more similar to spoken registers,

often becoming even more oral in the modern period than in the seventeenth century. These shifts result in a dispreference for certain stereotypically literate features, such as passive verbs, relative clause constructions and elaborated noun phrases generally. (Biber 2003: 169)

Empirical support for colloquialisation in contemporary English is provided in Leech *et al.*'s (2009) investigation of more recent grammatical change in English, namely, that which happened between the early 1960s and 1990s. Based on the four parallel members of the Brown family of corpora, their study shows that the frequency rise-and-falls of a range of grammatical features such as the modal auxiliaries, the progressive, genitive phrases and relative clauses are in line with the prediction of the colloquialisation hypothesis. Similar findings for an overlapping range of grammatical features were reported in the studies on recent, short-term change presented in Aarts *et al.* (2013), in which a wider range of diachronic corpora, both spoken and written, were used to represent contemporary English.

To this day, the study of colloquialisation has been largely restricted to native English varieties, in particular British and American English (BrE and AmE). Until recently, possible manifestations of colloquialisation in non-native English varieties have not been investigated in detail. This is due in part to, and is reflective of, the shortage of historical data representing these varieties, which empirical investigations of stylistic change crucially rest upon. The research gap has been addressed in a recent issue of the *Journal of English Linguistics* compiled by Nöel, Van der Auwera and Van Rooy. Focusing on expressions of modality in English, the contributions to this issue examine convergences and divergences between Philippine English (PhilE) and its historical input variety (Collins *et al.* 2014), and between Black South African English and its native counterpart in the same contact setting (Van Rooy & Wasserman 2014). Further extending our understanding of the evolution of non-native Englishes is the volume edited by Collins (2015a), which presents diachronic analyses of selected grammatical features in several other non-native varieties, such as Caribbean English, Hong Kong English and Indian English.

While the corpus-based studies mentioned above do not focus specifically on colloquialisation, the comparisons between speech and writing made in a number of them are suggestive of a distinction between the role of colloquialisation in non-native varieties and in their native counterparts: namely, that colloquialisation is a less powerful driver of grammatical change in the former than it is in the latter. This observation is essentially in line with the findings of several synchronic studies of non-native Englishes based on the *International Corpus of English* (Collins 2009; Xiao 2009; Mair & Winkle 2012; Collins & Yao 2013). A pattern emerging from these synchronic comparisons is that grammatical features more typical of speech are more frequent in AmE and BrE than in non-native varieties, and within the latter group, more so in South East Asian varieties than in Indian and East African varieties.

Another notable finding of these studies is that the results are very mixed depending on the individual linguistic features that are analysed. For example, focusing on newspaper language in the Caribbean in the past half century, Hackert

& Deuber (2015) examined the distribution of three linguistic features well known to be associated with colloquialisation: contractions, *that-* vs *which*-relatives, and the *be*-passive. While the first two features showed patterns consistent with the colloquialisation hypothesis, the *be*-passive evidenced little frequency change in the data. Conflicting tendencies have also been reported by Collins and his associates for Philippine English of around the same time period (Collins *et al.* 2014, 2015b). Their studies revealed, on the one hand, increases in the frequencies of semi-modal expressions (e.g. *have to*, *need to*) and relative *that*, which are driven arguably by colloquialisation, and on the other hand, stable and even falling frequencies of contracted and present-tense progressives, indicating that changes concerning the progressive cannot be accounted for by colloquialisation.

The mixed results obtained so far concerning individual linguistic features highlight the necessity for a more comprehensive, empirically grounded analysis of colloquialisation in non-native English. As such, the analysis should take into account not only a few opportunely selected features, but a much wider range of features possibly involved in the process. At the conceptual level, taking such an approach allows us to extend our understanding of colloquialisation and to see it not simply as an explanatory mechanism, but more importantly, as an empirically attestable phenomenon. At a more applied level, it enables us to determine whether and to what extent typical spoken features have permeated written language over time in non-native varieties across a range of registers, and how these varieties are similar to, or different from, native varieties in this respect.

In the present study, accordingly, we regard colloquialisation as a process affecting the overall shape of (at least) written English and involving a wide array of lexical, grammatical and discourse features.[2] We believe that colloquialisation can and is best measured using a data-driven approach and relying on a set of empirically defined COLLOQUIAL FEATURES and ANTI-COLLOQUIAL FEATURES (see section 3 below). In the following sections we present a case study of colloquialisation using this approach and focusing on its manifestations in grammar. The non-native variety examined is PhilE, with comparisons made with its input variety, AmE. Our analysis covers the period from the early 1960s to the early 1990s, as well as three written registers – press editorials, learned writing and fictional writing. In the next section we turn to a brief overview of PhilE. We then move on to introduce a general measure of colloquialisation derived from quantitative analysis of a large body of spoken and written data. The findings of an investigation of colloquialisation in PhilE and AmE are then presented.

---

[2] Elsewhere we have argued that colloquialisation should be defined in more general terms, i.e. as the spread of features typical of casual conversations to not only writing, but also more formal, detached types of speech such as scripted monologues (Collins & Yao 2013). However, as the present study does not take into account colloquialisation in speech, we restrict ourselves to a more established definition of colloquialisation.

2 Philippine English

English was introduced to the Philippines by American soldiers in 1898 after around three hundred years of Spanish occupation. It then began to spread throughout the country via the public school system established by the Americans, at a speed 'unprecedented in colonial history' (Gonzalez 1997: 28). By 1948, census results indicated that around 37.2 per cent of the population claimed to be able to speak English, an increase of 10.6 per cent over the figures in 1939 (Bureau of Census and Statistics 1954: 304). The popularity of English continued to rise steadily during the post-war period. Under the 1974 Bilingual Education Policy, English became the medium of instruction for science, mathematics and economics at all levels of education. Nowadays English is recognised as an official language alongside Filipino (the basis of which is Tagalog), and is used in many controlling social domains in the Philippines, including government administration, mass media, commerce, science and technology, and international relations (Sibayan 1994). Despite a recent purported decline in English proficiency in the country, census figures show that as of 2006 about two-thirds of Filipino adults 'understand spoken English'; another two-thirds 'read English'; about half 'write English'; about a third 'speak English'; only 14 per cent of the respondents say that they 'are not competent in any way when it comes to the English language' (Social Weather Stations 2006).

As English became more deeply rooted in the new context, it began to exhibit distinctive features in terms of phonology, grammar, lexicon and discourse. Since the publication of Llamzon's (1969) pioneering monograph, considerable work has been conducted to document aspects of PhilE, in particular those that differ greatly from Standard AmE (Alberca 1978; Gonzalez 1992; Bautista 2000, 2011; Bautista & Bolton 2008; Dayag 2012). With regard to grammar, distinctive uses have been noted in word order, subject–verb agreement, plural marking, tense–aspect–modality expressions and so on. While some of the distinctive uses can be traced to analogy and substrate transfer, others appear to reflect an inclination for polite and formal style. Examples include the frequent use of -*ly* adverbs as disjuncts (e.g. *essentially*, *frankly*, *unfortunately*) in everyday speech (Dita 2011) and the preference for modal *would* in non-past, non-conditional contexts where AmE would require *will* (Bautista 2004). Gonzalez (2004: 12) comments that Filipinos have a tendency to speak as they write and to transform features characteristic of formal written English to speech and less formal registers, so that their English is 'a monostylistic variety of English'. The monostylisticism hypothesis may serve as an explanation for the tendency discussed above, that PhilE makes less use of typical spoken features compared to native varieties of English. It remains to be seen whether the findings of the present large-scale study – albeit one whose focus is on writing – will provide any support for this hypothesis. In non-controlling, informal settings such as the home and the neighbourhood, communication often takes place in Filipino (Sibayan 1994), and it is common to find codeswitching among English, Filipino and Taglish, a mixed code of English and Tagaolog elements (Thompson 2003).

In recent years there has been growing awareness and acceptance of PhilE as a legitimate variety of English among local elites. Borlongan's (2009) survey of private university students showed that PhilE is functionally native for them and a representation of their Philippine identity. However, in other social groups English proficiency is lower and the attitude more ambivalent. Martin (2010, 2014a, 2014b) argues that PhilE seems to have found its place among the educated class; among less privileged members of the society, English may be largely inaccessible, and where it is, the preferred model of teaching and learning is still Standard AmE. The changing linguistic landscape in the Philippines, along with its complex sociohistorical context, provides a tempting subject for empirical diachronic investigation.

3   Defining colloquiality

As suggested in the introduction, the present study seeks to establish an analytical method that allows us to characterise the degree of colloquialisation across distinct varieties and registers of English. To this end, we rely on the term COLLOQUIALITY, using it in a more general and technical sense to refer to a combination of the degree of preference for linguistic features more typical of speech, and the degree of dispreference for linguistic features more typical of writing. We call the two opposing groups of features colloquial and anti-colloquial features respectively. This of course does not mean that colloquial features can be found only in speech and anti-colloquial features only in writing. Rather, the two terms are reflective of distributional patterns which are shaped by relative frequencies of linguistic forms as opposed to their mere presence or absence in a given register.

It can be seen that our data-driven approach to colloquialisation resembles the multidimensional approach developed by Biber (1988). In both approaches, the operationalised measure of a construct is identified quantitatively and in a bottom-up fashion. However, the primary foci of the two approaches are quite different. The multidimensional approach focuses on the so-called 'dimensions' such as 'informational vs involved production' and 'elaborate vs situation-dependent reference'. These stylistic oppositions highlight co-occurrence patterns of groups of linguistic features which are not equivalent to the more general speech/writing opposition that we are concerned with here.

3.1   The data

Our first goal is to determine the precise nature of colloquial and anti-colloquial features using a large collection of naturally occurring data. The texts collected for this purpose should cover a broad range of spoken and written texts, and ideally, represent the full range of situational variation across speech and writing. Moreover, as we aim to compare varieties of English at different time periods, it is necessary to take into account variation along the parameters of variety and time.

With these considerations in mind, we drew data from the following corpora:

(i) The Philippine component of the *International Corpus of English* (ICE-Phil). Sampled mainly for the 1990s, ICE-Phil is by far the most comprehensive multi-register corpus of contemporary PhilE. It contains around 1 million words divided into 32 text categories, 15 spoken and 17 written.

(ii) The American component of the *International Corpus of English* (ICE-US). In its current form, ICE-US contains around 400,000 words of written texts. These texts are parallel to the written section of ICE-Phil and represent AmE of roughly the same time period.

(iii) The *Santa Barbara Corpus of Spoken American English* (SBC; Du Bois *et al*. 2000–5). SBC was built with a view to being included in the spoken section of ICE-US. The corpus was also sampled for the 1990s and contains approximately 249,000 words. Although SBC does not provide a precise match of the spoken section of ICE-Phil, it also represents many spoken registers, for example, face-to-face conversations, telephone conversations, classroom lectures and sermons.

(iv) The Brown Corpus (Francis & Kučera 1964, 1971, 1979). As a 1 million-word corpus, Brown represents written AmE of the 1960s. Three of the text categories – B (press editorials), J (learned) and K (general fiction) – were selected as the basis for the diachronic analysis (see further section 3.4). The other written categories were omitted from our dataset, in order to avoid creating a major imbalance between spoken and written texts.

(v) The Phil-Brown Corpus (Borlongan in progress). Also sampled for the 1960s, Phil-Brown was designed to be the Philippine counterpart of Brown. Around two-thirds complete, Phil-Brown has been used to study individual grammatical features of PhilE (Collins *et al.* 2014, 2015b). For current analytical purposes, the same three categories, B, J and K, were selected to match the Brown texts.

Table 1 presents a summary of the corpus data employed in this study. In total our dataset contains slightly over 2 million words, around two-thirds of which are written and one-third spoken. Although texts representing the 1960s are outnumbered by those representing the 1990s, as a whole the current dataset provides us with a useful point of departure for studying speech-and-writing differences in the two varieties. It should be borne in mind that, unlike the AmE corpora (which feature people from a wide variety of social backgrounds), both ICE-Phil and Phil-Brown feature the English used by educated Filipino speakers and writers. This means that the demographic makeup of the language users represented by the two PhilE corpora is arguably different from that for the AmE ones. Such difference is inevitable due to the disparate roles of English in monolingual and multilingual ecologies: as previously suggested, English is mainly used in controlling domains in the Philippines, making PhilE an acrolectal variety.

Table 1. *Corpus data used in this study[a]*

		PhilE		AmE	
		Written	Spoken	Written	Spoken
1960s	Source	B, J and K in Phil-Brown	–	B, J and K in Brown	–
	No. of texts	15, 45 and 45 texts	–	27, 80 and 20 texts[b]	–
1990s	Source	ICE-Phil written	ICE-Phil spoken	ICE-US written	SBC
	No. of texts	200 texts	300 texts	200 texts	60 texts

[a]Each text is approximately 2,000 words in length, except for those in SBC.
[b]The compilation of Phil-Brown is still incomplete. There are therefore some variations between Phil-Brown and Brown in the number of texts available for the individual categories.

To prepare for frequency counts, all texts were part-of-speech tagged with CLAWS (C7 tagset) after having their original markups removed (see http://ucrel.lancs.ac.uk/claws/ for an introduction to CLAWS).

3.2   The linguistic features

The next step is to decide on the linguistic features to be used as a basis for determining colloquiality. Our primary focus is on grammar, an area which has triggered much recent scholarly interest in relation to colloquialisation. A valuable resource of grammatical features potentially involved in colloquialisation is presented in Biber's (1988) multidimensional study of English registers. In his model, 67 features were used to uncover textual relations among spoken and written text types. We had to exclude five of these features because they do not yield very reliable results with CLAWS-tagged texts.[3] The remaining ones are then modified and further supplemented by other features derived from a survey of the literature on register variation in English worldwide with a focus on grammar (e.g. Biber *et al.* 1999; Leech *et al.* 2009; Kortmann & Lunkenheimer 2013).[4] In selecting the features we seek to cover as many

---

[3] These include direct *wh*-questions, present participial clauses (e.g. *Holding the hair-dryer in her left hand, she cut her hair with her right hand*), past participial clauses (e.g. *Made in China in the 14th century, this vase is very valuable*), discourse particles (e.g. *well* and *anyway* in utterance-initial positions) and sentence relatives (e.g. *The girl is wearing a yellow hat, which is the funniest thing I have ever seen*). The reason for the low precision rates is that the retrieval of these features relies on tone-unit boundaries (as defined in Quirk *et al.* 1985). Unfortunately, CLAWS does not have a separate code for tone-unit boundaries, unlike the part-of-speech tagger used in Biber (1988). In addition, we had to exclude gerunds (participial forms serving nominal functions) because they cannot be automatically identified with CLAWS either, and manual editing of the results is not realistic given the large number of irrelevant instances retrieved by searching simply for *-ing*.

[4] Modification involves the following features: first, Biber coded relative clauses in subject and object positions as two separate features, but we combined them into one single feature due to the lack of a reliable method for distinguishing sentence objects from subjects. Second, we merged into one group (clausal subordination)

possibly implicated features as we can. Our main consideration was feasibility, i.e. whether the frequencies of a given feature in the data can be determined with an automatic or semi-automatic procedure, since the size of our dataset does not lend itself to manual coding or close reading of the linguistic contexts of each hit. Our second consideration was that the features should be known to be shared by, or at least relevant to, the two varieties of English, particularly in the written mode, because the corpus texts used for the diachronic comparison in this study only contain written material (see further below). As is commonly known, corpus linguistic methodologies are not well suited for investigating very rare phenomena, and when an item can hardly be found in the data, the analysis based on its frequencies cannot be statistically robust (Szmrecsanyi 2013: 36–7). We therefore set a frequency of 100 tokens (i.e. around 0.04 tokens per thousand words) in the entire dataset as the minimum threshold frequency for a feature to be added into the feature list. Features unique to PhilE were also excluded because applying them to published AmE writing as represented in the Brown family corpora where their absence is fairly categorical runs the risk of over-reporting colloquiality in PhilE. As a result, we had to rule out many interesting grammatical features known to be sensitive to register variation, such as the following:

(i) Interrogative/exclamative sentence types, cleft sentences (e.g. *It is he who left us first.*), zero relative clauses (e.g. *the way he walks*) and *yes–no* questions, which are features of standard English but can only be manually identified;

(ii) Loss of singular inflections of verbs (e.g. *He go away.*), use of definite article where standard English has indefinite article (e.g. *I had the toothache.*) and vice versa (e.g. *A sun is shining.*), which have often been suggested to be distinctive of non-native varieties but also need to be manually identified;

(iii) Borrowings from Tagalog, including the conjunction *kasi* ('because') and enclitic particles with pragmatic functions such as *no*, *ba* and *pa* (Lim & Borlongan 2011), which are common in our spoken PhilE data but non-existent in AmE. Although ICE-Phil was built with a view to involve 'minimum Tagalog insertions' (Bautista 2011: 7), it is not hard to find unnaturalised indigenous linguistic elements with grammatical functions in the corpus, especially in spoken private conversations, as one reviewer of this article has correctly pointed out. In total the corpus yields 6,552 instances of indigenous single words and word strings (annotated as <indig></indig> by the corpus compilers), reflecting the close intertwining of English and Tagalog. Notably, 81.1 per cent

the four types of adverbial subordinators in Biber (1988), because when treated separately, their frequencies in individual texts are sometimes almost negligible. Moreover, we have divided Biber's category of attributive adjectives into descriptors (e.g. *a nice girl*) and classifiers (e.g. *foreign countries*) in view of their different distributional patterns across spoken and written registers (see Biber *et al.* 1999: 508–15). Finally, we replaced Biber's original four specialised verb classes by the more comprehensive and empirically attested taxonomy in Biber *et al.* (1999: 360–402). This includes common single-word activity verbs (e.g. *bring*, *hold*), mental verbs (e.g. *know*, *hear*), communication verbs (e.g. *talk*, *answer*) and existence verbs (e.g. *seem*, *remain*).

of the indigenous elements appear in the spoken texts. Users of these elements account for 51.7 per cent of the 710 sampled speakers, and each user produces an average of 14.5 instances, corresponding to a textual frequency of around 8.9 instances per thousand words. By contrast, only 34.5 per cent of the 400 sampled writers produce such elements, with an average of 9.0 instances per writer and a textual frequency of 3.1 instances per thousand words. This clear contrast between the spoken and written texts suggests that some indigenous elements may qualify as colloquial features by our definition. However, they are not examined in this study given the present comparative purposes.

(iv) *Me* instead of *I* in coordinate subjects (e.g. *My brother and me were late*), plural forms of non-count nouns (e.g. *staffs*, *advices*) and no number distinction in demonstratives (e.g. *this children*). These features have been reported to be existent in PhilE (Kortmann & Lunkenheimer 2013) but their frequencies in the data do not reach the minimum threshold.

The resultant feature list contains a total of 86 features divided into seven major categories: (A) the noun phrase; (B) the verb phrase; (C) adjectival, adverbial and prepositional phrases; (D) subordination; (E) other phrasal and clausal elements; (F) reduced forms and dispreferences; (G) lexical complexity (see table 2 below).[5]

3.3 Determining colloquial and anti-colloquial features

Frequencies of the 86 features were retrieved with PowerGREP, a grep software that enables complex searches using regular expressions. The results were then normalised to frequencies per thousand words for each text.[6] Means and standard deviations were computed based on normalised frequencies. Eighty-seven ANOVAs were then conducted to determine which of these features are preferred in speech or writing. Each ANOVA provides two useful statistics: p-value, which indicates whether the distributional difference between speech and writing is significant (set at the level of 0.05 in this study); and r^2, which indicates the importance or strength of the difference.[7] An r^2 higher than 0.20 is often interpreted as showing an important relationship between the independent and dependent variables (see Biber & Finegan 1989: 498). These two statistics enable us to arrive at our operational definitions of colloquial and anti-colloquial features. We take the former to be features whose frequencies are significantly higher in speech than in writing, and the latter to be those with a reversed distributional pattern. In order to highlight the most important statistical relationships in the data, we restricted ourselves to features with r^2 values

---

[5] The three 'features' grouped under lexical complexity are not linguistic features in the usual sense of the word; they characterise properties of texts, rather than properties of items within the grammar of English. The term was retained for convenience.

[6] The number of words in each text was calculated with the WordList function in *WordSmith Tools* (version 5.0). Type/token ratios and mean word lengths were also calculated with this function.

[7] r^2 indicates the percentage of variance in the normalised frequencies that is accounted for by knowing the register. It is computed by dividing the sum-of-squares between groups by the total sum-of-squares.

Table 2. *Results of ANOVAs for the linguistic features*

Feature	Spoken mean	Written mean	F	p-value	r^2	Type
A. The noun phrase						
Articles	64.8	87.4	269.1	.000	.21	Anti-colloquial
Demonstratives	28.7	18.2	398.9	.000	.29	Colloquial
Possessive determiners	14.3	15.4	3.4	.065		
Quantity determiners	6.9	7.3	3.5	.063		
Semi-determiners	1.7	1.8	2.0	.158		
First-person pronouns	37.5	11.4	496.6	.000	.33	Colloquial
Second-person pronouns	20.9	3.6	704.8	.000	.41	Colloquial
Third-person pronouns	18.8	16.2	5.5	.019	.01	
Nonpersonal pronoun *it*	15.8	7.6	341.8	.000	.25	Colloquial
Indefinite pronouns	3.3	1.5	146.5	.000	.13	
Numerals	20.1	20.9	0.8	.378		
Singular common nouns	127.3	161.1	233.1	.000	.19	
Plural common nouns	32.6	60.6	438.5	.000	.31	Anti-colloquial
Common nouns neutral for number	2.4	3.1	10.3	.001	.01	
Proper nouns	28.5	32.6	7.0	.008	.01	
Temporal nouns	9.5	8.8	2.4	.122		
Place nouns	0.2	0.3	2.9	.091		
Total other nouns[a]	83.7	76.6	3.26	.071		
Nominalisations	17.2	30.7	151.7	.000	.13	
Adjective+noun sequences	28.3	53.9	535.5	.000	.35	Anti-colloquial
Noun+noun sequences	21.6	29.0	42.9	.000	.04	
S-genitives	1.8	4.2	107.4	.000	.10	
Of-genitives	14.1	25.2	239.3	.000	.19	
B. The verb phrase						
Common activity verbs	24.8	21.8	21.0	.000	.02	
Common mental verbs	23.2	11.8	328.3	.000	.25	Colloquial
Common communication verbs	11.2	8.9	33.8	.000	.03	
Common existence verbs	1.6	3.5	240.3	.000	.19	
Be as main verb	42.3	25.4	564.4	.000	.36	Colloquial
Have as main verb	8.3	4.0	359.0	.000	.26	Colloquial
Do as main verb	9.9	2.4	487.9	.000	.33	Colloquial
Past tense	27.9	31.6	5.0	.025	.01	
Present tense	74.8	42.8	482.3	.000	.33	Colloquial
Perfect aspect	4.7	6.5	53.5	.000	.05	
Progressive aspect	6.4	2.9	296.0	.000	.23	Colloquial
Agentless passives	4.2	7.5	114.8	.000	.10	Anti-colloquial
By-passives	0.4	1.0	79.5	.000	.07	
Get-passives	0.2	0.1	26.7	.000	.03	
Possibility modals	5.2	5.6	2.3	.127		
Necessity modals	1.5	1.9	13.6	.000	.01	

Table 2. *Continued.*

Feature	Spoken mean	Written mean	F	p-value	r^2	Type
Prediction modals	8.2	5.6	90.4	.000	.08	
Possibility semi-modals	0.5	0.3	22.9	.000	.02	
Necessity semi-modals	2.3	1.0	189.0	.000	.16	
Prediction semi-modals	3.0	0.6	287.2	.000	.22	Colloquial
C. Adjectival, adverbial and prepositional phrases						
Predicative adjectives	3.6	3.1	11.8	.001	.01	
Common attributive adjectives: descriptors	4.4	6.3	79.7	.000	.07	
Common attributive adjectives: classifiers	3.9	6.7	83.1	.000	.08	
Phrasal comparatives and superlatives	0.7	1.4	81.4	.000	.08	
Inflectional comparatives and superlatives	1.9	2.9	63.7	.000	.06	
Comparative clauses and other degree complements	0.3	0.6	48.0	.000	.05	
-ly adverbs[b]	9.7	12.4	81.6	.000	.08	
Non-*ly* adverbs	68.4	39.6	483.7	.000	.33	Colloquial
Time adverbs	5.8	3.9	104.4	.000	.09	
Place adverbs	2.4	3.5	40.9	.000	.04	
Degree adverbs: amplifiers	6.1	3.0	257.5	.000	.20	
Degree adverbs: downtoners	1.8	2.5	53.3	.000	.05	
Stance adverbs	1.4	1.4	0.7	.405		
Linking adverbs	1.2	1.9	51.6	.000	.05	
Total prepositions	84.6	119.4	505.2	.000	.34	Anti-colloquial
D. Subordination						
That-relative clauses	3.5	3.5	0.1	.795		
Wh-relative clauses	3.2	3.2	0.2	.623		
Pied piping	0.3	0.9	110.9	.000	.10	
Stranded prepositions	1.4	0.7	46.1	.000	.04	
That-verb complement clauses	3.1	2.6	13.3	.000	.01	
That-adjective complement clauses	0.4	0.4	0.0	.886		
Wh-clauses as complements	2.3	1.2	148.4	.000	.13	
Infinitival complements	5.8	6.4	9.0	.003	.01	
Ing-participial complements	1.2	1.2	0.6	.442		

Table 2.   *Continued.*

Feature	Spoken mean	Written mean	F	p-value	r^2	Type
Past participial whiz-deletion relatives	1.4	3.6	287.4	.000	.22	Anti-colloquial
Present participial whiz-deletion relatives	2.2	3.0	43.3	.000	.04	
E. Other phrasal and clausal features						
Phrasal coordination	5.1	10.3	242.8	.000	.20	
Clausal coordination	27.9	25.2	25.5	.000	.03	
Clausal subordination	13.0	8.2	232.2	.000	.19	
Conjuncts	1.8	3.2	107.2	.000	.10	
Existential *there*	2.9	1.8	93.8	.000	.09	
No-negation	4.5	1.7	189.9	.000	.16	
Not-negation	12.0	6.3	266.9	.000	.21	Colloquial
Comment clauses	7.7	0.5	558.6	.000	.36	Colloquial
Hedges	1.3	0.7	87.0	.000	.08	
Emphatics	9.9	5.1	270.7	.000	.21	Colloquial
F. Reduced forms and dispreferences						
Contraction	26.2	3.9	643.5	.000	.39	Colloquial
Subordinator *that*-deletion	7.4	2.2	357.6	.000	.26	Colloquial
Split infinitives	0.2	0.1	0.9	.345		
Split auxiliaries	2.4	2.6	3.2	.074		
G. Lexical complexity						
Type/token ratio	25.0	35.0	811.5	.000	.45	Anti-colloquial
Mean word length	4.2	4.8	574.1	.000	.36	Anti-colloquial
Lexical density	0.6	0.6	49.7	.000	.05	

[a]This includes all nouns other than those counted as the six types of nouns listed above. It provides a further characterisation of the 'nouniness' of a text. In Biber (1988) this is defined differently, and includes all nouns other than nominalisations and gerunds.
[b]This feature replaces Biber's 'total adverbs', which includes any adverb that is longer than five letters and ends in *-ly*.

higher than 0.20. Table 2 shows the results of the ANOVAs, including mean normalised frequencies for speech and writing.[8]

The majority of the selected linguistic features (66 out of 86) exhibit highly significant differences ($p < 0.001$) in their distributions across speech and writing. However, only 25 features, 17 colloquial and 8 anti-colloquial, are strong indicators of the speech-and-writing divide, with r^2 values of higher than 0.20.

Some of the most remarkable register differences can be seen for features indicating a general concern with the 'here' and 'now' of the communicator. The present tense,

---

[8] Normalisation does not apply to the three features in the category of lexical complexity.

for example, can be used to describe events and states that are ongoing or existing at the time of utterance. Demonstratives and first- and second-person pronouns establish reference to the communicators and in relation to aspects of the immediate situational context. Moreover, speech contains more expressions with emotional and attitudinal meanings: common mental verbs (e.g. *think*, *know*, *want*), which denote psychological states experienced by animate subjects; prediction semi-modals (*be going to*, *be about to*, *want to*), which express the communicator's prediction and volition; comment clauses (e.g. *I mean*, *you know*, *you see*), which indicate various kinds of attitude to the proposition expressed; emphatics, which signal a strong degree of certainty (e.g. *for sure*, *such a*). Speech also differs significantly from writing in having a higher frequency of structurally simpler forms such as non-*ly* adverbs (as opposed to -*ly* adverbs typically derived from adjectives), words with shorter lengths and reduced forms such as contractions (e.g. *'ll*, *'m*), and subordinator *that*-deletion (e.g. *I think (that) this is the best solution*). As for *not*-negation and the progressive, their popularity in speech can be ascribed to the functional prominence of the verb. Biber *et al.* (1999: 65–6) have shown that verbs are more common in conversations than in academic writing. Since both negation and aspectual marking are tied to verbs, it is not surprising that these features appear more often in speech.

The anti-colloquial features identified in this study are generally indicative of an informative, compact style. Among such features are plural common nouns, adjective+noun sequences and past participial whiz-deletion relatives (e.g. *the chemical (which is) produced by this process*), reliance on which indicates more compact encoding of information. Many studies have revealed that written English has become increasingly oriented towards nominal discourse (e.g. Biber & Finegan 1997; Biber & Gray 2011). This orientation may explain our finding that articles and prepositions, which co-occur with nouns, also emerge as anti-colloquial features. The hallmark of informativeness is perhaps type/token ratio, which represents the number of different lexical words in texts. A higher mean score for writing on this measure suggests that writing is more varied in lexical choice, hence informationally richer. Correspondingly, speech is marked by features which express a lower level of semantic specificity, and which are therefore less informative: *be*, *have* and *do* as main verbs, and the nonpersonal pronoun *it* (e.g. *It's warm today*; *It's getting dark*). Finally, the fact that the agentless passive is also selected as anti-colloquial reflects that, in addition to packing a large amount of information, writing tends to be more detached and abstract (Biber & Finegan 1989; Wanner 2009).

3.4   *Measuring colloquiality*

We are now in a position to calculate colloquiality scores for texts to be used for the diachronic analysis of PhilE and AmE. The colloquiality score of a text summarises its preference for colloquial features and its dispreference for anti-colloquial features. Since some of the features have remarkably higher normalised frequencies (e.g. the present tense) than others (e.g. prediction semi-modals), we calculated the standardised

Table 3. *Corpus texts used in the diachronic comparisons*

		PhilE		AmE	
		1960s	1990s	1960s	1990s
Press editorials	Source	B in Phil-Brown	W2E in ICE-Phil	B in Brown	W2E in ICE-US
	No. of texts	15	10	27	10
Learned	Source	J in Phil-Brown	W2A in ICE-Phil	J in Brown	W2A in ICE-US
	No. of texts	60	40	80	40
Fiction	Source	K in Phil-Brown	W2F in ICE-Phil	K in Brown	W2F in ICE-US
	No. of texts	45	20	20	20

scores (z-scores) of the normalised frequencies, so that the contributions of high- and low-frequency features are transformed to a common scale and can be compared. The colloquiality score of a text is defined as the sum of the standardised frequencies of all colloquial features minus the standardised frequencies of all anti-colloquial features in this text.[9]

Parallel texts were then chosen from our dataset to build a diachronic corpus of PhilE and AmE. Specifically, this involved matching text categories in Phil-Brown and Brown to those in ICE-Phil and ICE-US. Unfortunately, the lack of representations of earlier PhilE speech prevents us from studying diachronic trends in the spoken language. Nevertheless, it is still possible to make comparisons with regard to three written registers, press editorials, academic writing and fiction, which represent distinct social functions and levels of formality. Table 3 shows the text selection scheme for our diachronic corpus. Colloquiality scores were calculated for all texts contained therein.

4 Variation in colloquiality scores

In this section we report quantitative findings gained from analysing the colloquiality scores of texts in the diachronic corpus. We begin by outlining general patterns of register variation in the data and proceed to determine whether and how far colloquialisation has progressed in the two varieties during the thirty-year period.

4.1 Register variation

Figure 1 presents the mean colloquiality scores for the diachronic corpus. At first glance, regional and diachronic variation in colloquiality scores is less noteworthy than overarching patterns of register variation. The results largely confirm our expectations

---

[9] Type/token ratios and mean word length were also standardised.

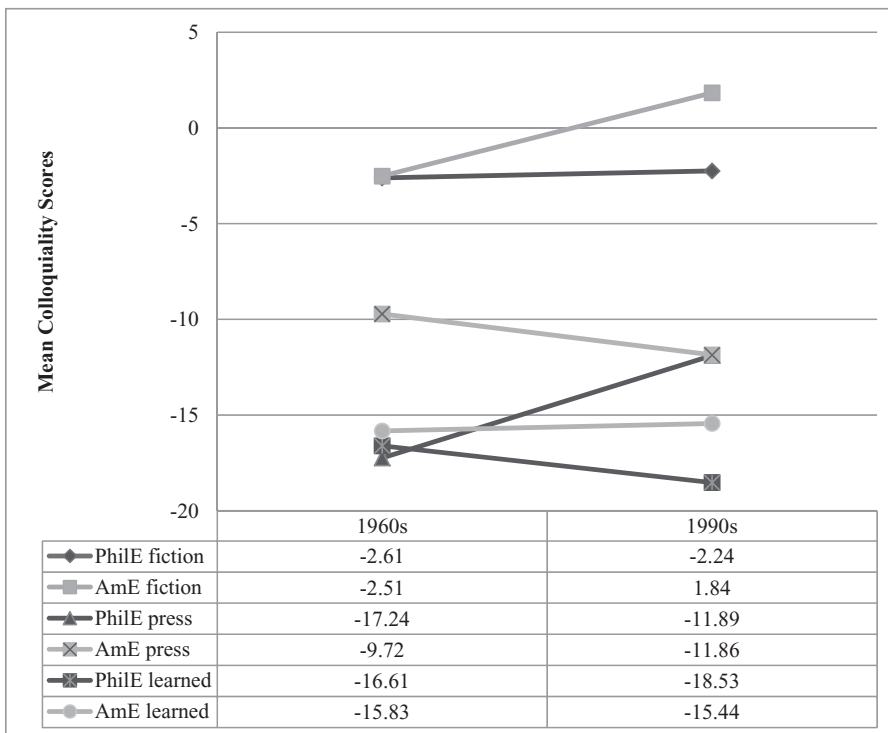|	1960s	1990s
◆ PhilE fiction	-2.61	-2.24
■ AmE fiction	-2.51	1.84
▲ PhilE press	-17.24	-11.89
✕ AmE press	-9.72	-11.86
⊠ PhilE learned	-16.61	-18.53
● AmE learned	-15.83	-15.44

Figure 1. Mean colloquiality scores of texts in the diachronic corpus

about the colloquiality levels of the three registers. Despite variation over time, the most colloquial register is fiction, with PhilE and AmE having respective mean scores of -2.43 and 0.34.[10] This is not surprising as writers create fictional worlds often by producing linguistic features that are imitative of those of face-to-face conversation. Sometimes narration is developed mainly through the conversation of characters. In comparison, the two informative registers are clearly anti-colloquial, with press editorials having mean scores of -14.57 (PhilE) and -10.79 (AmE), and learned writing -17.57 (PhilE) and -15.64 (AmE), all of which are far below zero.

Figure 2 presents a more elaborate picture of general patterns of register variation by showing the mean standardised frequencies of the 25 features regardless of variety and time period. As in figure 1, we find in figure 2 a notable divide between fiction on the one hand, and press editorials and learned writing on the other. Fiction has higher frequencies of all colloquial features except demonstratives and present tense verbs. The frequency gap with demonstratives reflects a combination of two factors: first, a greater reliance in fiction on explicit reference to construct imagined situational contexts, and second, a stronger need in press and learned writing to

---

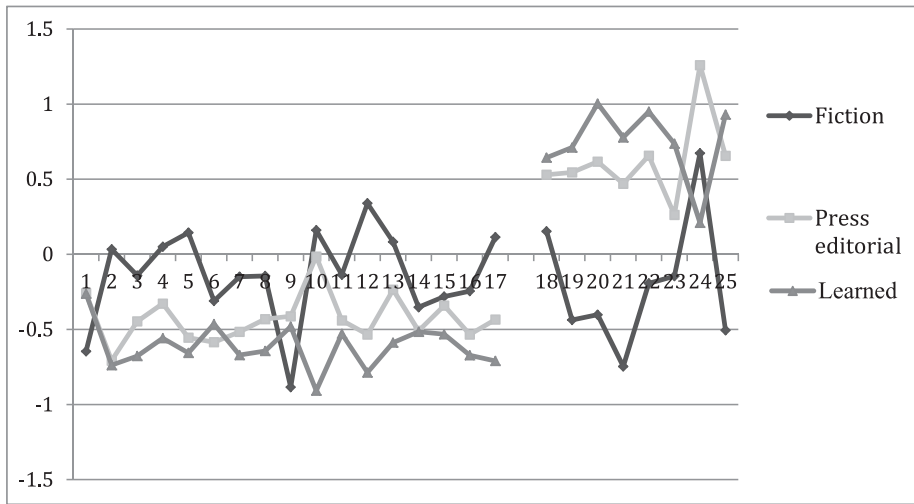[10] These are the averages of the 1960 and 1990 scores in figure 1.

Figure 2. Mean standardised frequencies of colloquial and anti-colloquial features
in the three registers

Note: Colloquial features: 1 = demonstratives; 2 = first-person pronouns; 3 = second-person
pronouns; 4 = nonpersonal pronoun *it*; 5 = common mental verbs; 6 = *be* as main verb; 7 =
have as main verb; 8 = *do* as main verb; 9 = present tense; 10 = progressive aspect; 11 =
prediction semi-modals; 12 = non-*ly* adverbs; 13 = *not*-negation; 14 = comment clauses;
15 = emphatics; 16 = contractions; 17 = subordinator *that*-deletion. Anti-colloquial features:
18 = articles; 19 = plural common nouns; 20 = adjective+noun sequences; 21 = agentless
passives; 22 = total prepositional phrases; 23 = past participial whiz-deletion relatives; 24 =
type/token ratio; 25 = mean word length.

establish anaphoric links to events and concepts introduced in the immediate text. The
relative shortage of present tense verbs in fiction is apparently due to its concern with
narration, for which the past tense tends to be a preferred choice. As a mirror image of
the findings for colloquial features, most anti-colloquial features are less common in
fiction than in the other two registers. The only exception to this pattern is type/token
ratio. Its lowest score in learned writing reflects a smaller range of vocabulary as
determined by this register's specialised content and narrow focus. What is especially
interesting about the findings in figure 2 is that the gap between fiction and the two
informative registers is much greater for colloquial than for anti-colloquial features.
This means that the higher colloquiality level of fiction is defined not so much by its
preference for colloquial features as by its dispreference for anti-colloquial features.

4.2   Diachronic variation

Having outlined general patterns of register variation, we move on to examine
diachronic variation in colloquiality scores. A comparison of the 1960s and 1990s
data reveals a striking change in PhilE press editorials (mean score -17.24→-11.89),

suggesting that this register has become substantially more colloquial over the thirty-year period. The outcome of this change, together with a mild decline in colloquiality in the corresponding American data (-9.72→-11.86), is that PhilE and AmE press editorials of the 1990s are far more similar in their colloquiality level than they were in the 1960s. The second notable change concerns AmE fiction, whose mean score rises from -2.51 to 1.84. By contrast, PhilE fiction, which starts at roughly the same point in the 1960s, remains stable over time (-2.61→-2.24). As for learned writing, its colloquiality level has not changed much in AmE (-15.83→-15.44) and has even decreased slightly in PhilE (-16.61→-18.53). In general, what these patterns illustrate is that the three registers differ in their degree of susceptibility to colloquialisation during the second half of the twentieth century. It is useful here to consider the distinction drawn by Hundt & Mair (1999) between 'agile' and 'uptight' styles. On the whole, fiction and press editorials are typical 'agile' registers in their receptiveness to linguistic innovations, whereas learned writing – a notably 'uptight' register – is more resistant to change. The distinction in question can be traced to different readership types. Fiction and press writing are typically produced subject to the economic pressure to attract bigger audiences and hence to conform to emerging and 'fashionable' styles. By contrast, learned writing targets a relatively stable group of readers and its stylistic conventions are thus more entrenched.

Before we turn our attention to diachronic changes in PhilE, a comment should be made concerning the findings for AmE. While the AmE fiction data clearly support the colloquialisation hypothesis, the mild decline in the colloquiality score for AmE press editorials appears somewhat perplexing at first sight. Upon closer examination, two factors emerge as responsible. First, compared with the 1960s press editorials, the 1990s texts employ far fewer first-person pronouns (mean standardised frequencies: -0.85 < -0.52) as well as fewer mental verbs (-0.69 < -0.36). Similar patterns were identified by Westin (2002) in her diachronic analysis of editorials in British 'upmarket' newspapers published over the past century. The implication is that in these texts there has been a move away from the explicit marking of author stance. Another remarkable difference between the 1960s and 1990s AmE editorials is that the latter greatly outstrip the former not only in the use of plural common nouns (0.57>0.02) and adjective+noun sequences (0.78>0.45), but also in mean standardised word length (0.80>0.38). Increases in the scores of these three anti-colloquial features are indicative of DENSIFICATION, a well-documented discourse process via which information is compacted into a smaller number of words (e.g. Leech *et al.* 2009; Biber & Gray 2011). As economy of expression is highly valued in newspaper language, it is not surprising that American press editorials have moved towards informational density and become less colloquial over time.

The evolution of PhilE press editorials follows an entirely different path, showing a strong colloquialising tendency which calls for further investigation into their linguistic and situational properties. As we have seen above, Philippine press editorials of the 1960s are highly formal and closely resemble learned writing. Figure 3
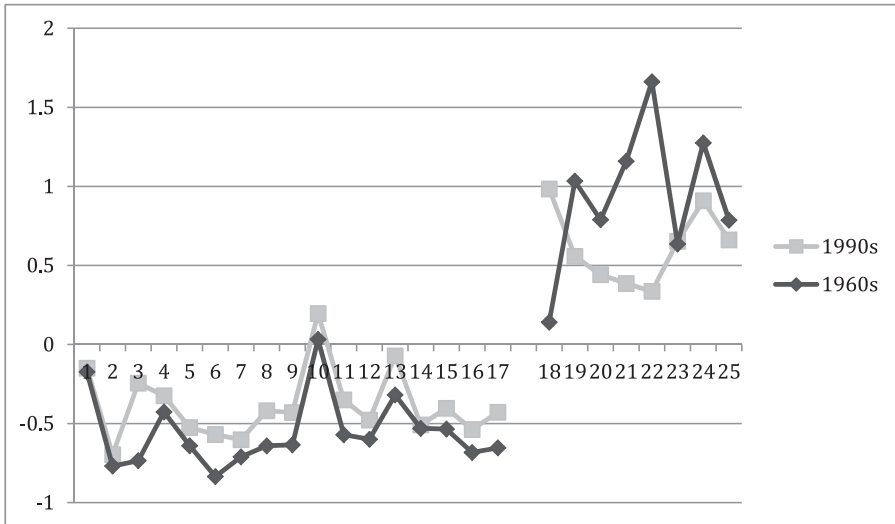
Figure 3. Mean standardised frequencies of colloquial and anti-colloquial features
in PhilE press editorials

Note: Colloquial features: 1 = demonstratives; 2 = first-person pronouns; 3 = second-person
pronouns; 4 = nonpersonal pronoun *it*; 5 = common mental verbs; 6 = *be* as main verb; 7 =
have as main verb; 8 = *do* as main verb; 9 = present tense; 10 = progressive aspect; 11 =
prediction semi-modals; 12 = non-*ly* adverbs; 13 = *not*-negation; 14 = comment clauses;
15 = emphatics; 16 = contractions; 17 = subordinator *that*-deletion. Anti-colloquial features:
18 = articles; 19 = plural common nouns; 20 = adjective+noun sequences; 21 = agentless
passives; 22 = total prepositional phrases; 23 = past participial whiz-deletion relatives; 24 =
type/token ratio; 25 = mean word length.

indicates that compared with their 1990s PhilE counterparts they are considerably
less reliant on second-person pronouns (-0.73 < -0.24), *be* as main verb (-0.84
< -0.57), *not*-negation (-0.32 < -0.07) and subordinator *that*-deletion (-0.65 <
-0.43). In line with these patterns is a higher type/token ratio (1.28>0.91) and a
much stronger preference for the majority of the anti-colloquial features, including
plural common nouns (1.03>0.56), adjective+noun sequences (0.79>0.44), agentless
passives (1.16>0.39) and prepositional phrases (1.66>0.34). Taken together, these
findings reflect a tendency for 1960s PhilE press language to be more compact in
content and more complex in lexical choice and syntactic structure.

    The following excerpts, taken from Phil-Brown and ICE-Phil, illustrate the stylistic
preferences of 1960s and 1990s PhilE press editorials respectively:

(1) There is one project of the present administration the significance of which appears to
    have been lost in the welter of more sensational albeit transient scandals, both petty and
    monstrous in the government. This is the extension of the railroad into the Bicol region's
    southernmost areas and into the Cagayan valley. Studies have been made on this project
    and although millions of pesos are involved in its realization, the administration has
    repeatedly signified determination to push it through as a part of ambitious program to

link Luzon to Mindanao through the Visayas. The major phase of the project, of course, is the railroad line's extension into the Cagayan valley up to Aparri, including not only the laying of rail lines north of the Carballo mountain range but also the construction of a tunnel through the mountain. A huge engineering job, from the erection of durable railbeds that can withstand the periodic overflow from the Cagayan River to tunneling through a vast mountain complex, will involve considerable work and expenditure which could discourage the weak of heart. (Phil-Brown B03)

(2) While lending rates have skyrocketed, interest rates on deposits continue at 2 to 3 percent. For the banks, it seems it is all take, take, take and never any give. In the present crisis, bankers continue to laugh all the way to the bank. It is as if the banks had a God-given right always to make money and never to lose any, no matter what the situation for the rest of the country. A time-honored principle of human relations is that those who shaft you should be shafted back. The banks are not afraid, for they believe in the Imeldific 'rule' that those who have the gold make the rules. That's true, but those who hold the rod can change the rules. The banks have money but the people have power. They can tell an administration seeking vindication at the polls to bring the banks to heel or under state control before they ruin the country and hurt the people some more. (ICE-Phil W2E010)

Excerpt (1) differs from (2) in the level of concentration of nouns (0.30 vs 0.22 token per word). More importantly, noun phrases in excerpt (1) often contain one or more modifiers or complements which serve to form complex structures containing detailed and precise information. These elements include attributive adjectives (e.g. *sensational albeit transient scandals*, *ambitious program*), nouns (e.g. *railroad line*, *engineering job*), *s*-genitives (e.g. *Bicol region's southernmost areas*, *railroad line's extension*) and prepositional phrases (e.g. *the major phase of the project*, *the erection of durable railbeds*). By comparison, in excerpt (2) the nominal constructions are less complex, have shorter lengths and feature fixed lexical combinations (e.g. *interest rates*, *human relations*). As for the verb phrase, excerpt (1) features agentless passives (*which appears to have been lost*, *studies have been made*). These constructions enable inanimate entities to become grammatical subjects, thereby allowing information to be structured in a coherent way. Excerpt (2), on the other hand, is characterised by a heavy reliance on verbal constructions in the active voice, as well as the use of *be* and *have* as main verbs (*it is all take, take, take and never any give*; *The banks have money but the people have power*) and high-frequency verbs (e.g. *continue to laugh all the way*; *bring the banks to heel*). The overall effect of these linguistic differences is that excerpt (1) is more informationally compact and requires a considerable amount of processing effort on the part of the reader, while excerpt (2) is more speech-like and less structurally complex.

The question that arises from the above discussion is what accounts for, first, the strong colloquialising tendency in PhilE editorials, and second, the absence of such a tendency in PhilE fiction and learned writing of the same period. This question can be addressed by further examining the sociohistorical backdrop against which PhilE has progressed. Unlike in Inner Circle countries/regions, English was 'transplanted' to the Philippines as a result of US occupation and colonisation, via the establishment of a regular system of English teaching at the beginning of the twentieth century.

In the following decades, which saw the formation of PhilE, English teaching in the classroom was conducted mainly via grammatical analysis and imitation exercises. Teachers presented Anglo-American literary canons including those of Matthew Arnold, Washington Irving, Henry Longfellow and Ralph Waldo Emerson, as well as those of Shakespeare, as examples of 'good English' (Gonzalez 2008; Martin 2008). Exposure to such texts and sustained writing practice had direct effects on students' writing. Gonzalez (1991) noted that Filipino students of the colonial period tended to write compositions in an antiquated, lofty Victorian style featuring archaic expressions and florid sentences, what he called the 'Philippine classroom composition style'. In 1925, the Board of Educational Survey conducted a comprehensive study of the Philippine public system and reported that 'children in upper grades seem to have a "reciting" knowledge of more technical English grammar than most children in corresponding grades in American schools' (cited in Martin 2008: 251). The continuing influence of the colonial pedagogic tradition is reflected in the penchant for complex and formal grammatical features in the earlier PhilE press. It is not surprising to see this penchant gradually weaken as teaching practices in the Philippines evolved over the years to incorporate functional approaches, such as communicative language teaching, content-based instruction and English for specific purposes (Gonzalez 2008). With such approaches, the focus on formal 'correctness' is often replaced by an emphasis on the appropriateness of language use in meaningful, communicative contexts.

There are several other social and demographic forces that are likely contributors to the colloquialisation of Philippine press language. First, increased international communication may have prompted Philippine journalists to adhere to the conventions of international journalism as promoted by the more colloquial American newspapers. Importantly, census figures show that the proportion of English speakers in the Philippines rose from around 39 per cent in 1960 to 56 per cent in 1990 (Gonzalez 2004). This substantial rise in English literacy resulted in the expansion of the general reading public, whose needs were better catered for by less complex and abstract styles. Furthermore, while television and radio stations have witnessed a rapid growth in Filipino programmes in recent years, the print media in the Philippines have long been dominated by English (Gonzalez & Bautista 1986; Dayag 2004). All of the important newspapers that enjoy national circulation are in English, including the *Philippine Daily Inquirer*, *Manila Bulletin* and *Philippine Star*. Philippine newspapers also tend to have a wide readership and a strong social impact. Citing figures from the 2000 *Philippine Media Factbook*, Dayag (2004) reported that 29 per cent of the entire Philippine population read newspapers and as high a proportion as 48 per cent of Metro Manila residents did so. It was in response to the wide influence of Philippine print media on the Philippine community that press freedom was severely curtailed during the 1972–81 martial law period. At that time, a number of newspapers critical of Marcos's military rule were terminated and replaced by those in favour. As a result, 'the intellectual lights went out, along with the other "inalienable" rights of the Filipino people', and the Philippines 'went into a deathly journalistic silence' (Mijares 1976: 325). In summary, improved English literacy and a popular audience are among

the factors responsible for the marked shift of PhilE press editorials towards colloquial styles.

By contrast, the diachronic stability of PhilE fiction and learned writing can be traced to the lack of a general readership and the limited social impact of these two registers. As discussed earlier, the readers of learned writing tend to be members of a highly educated group comprised of academics, professionals, students and the like. This register can therefore afford to remain 'uptight', i.e. unaffected by the interests and demands of the general public. On the other hand, the social context that has fostered English fiction writing and reading in the Philippines is drastically different from that in the United States. The Philippines has not had a large community of English fiction readers. An important reason why fiction continues to be written and published in English in the Philippines is the provision of institutional support from the government, which regularly sponsors literary contests such as the Commonwealth Literary Awards and the National Artists Awards, in an effort to promote culture and the arts in the country. Despite these efforts, PhilE fiction remains largely unread by the majority of the local people. Lamenting the unpopularity of literature in the Philippines, Gonzalez (1988: 36) notes the general impression that Filipinos are not book readers. According to writer Charles Ong, 'a novel in English that sells a thousand copies in three or four years, itself a rarity, is deemed as a best-seller by Philippine standards' (cited in Hau 2008: 320). As for the demographic makeup of the intended audience of PhilE fiction, Hau (2008: 321) writes:

> [I]ts production and reception are restricted to a minority of cultural workers in the publishing, journalism, and educational sectors and to a small percentage of the student and professional population in the Philippines, a fact that accounts for the seemingly 'incestuous' nature of literary production and consumption, and the preeminence of a 'personal' politics of authorship in the country.

The inability of PhilE fiction to reach a bigger audience places it in sharp contrast with Philippine newspapers, with their wide circulation. It is perhaps not surprising that the martial law period saw the leading personalities of the print media arrested and detained for their comments in newspapers, but not in fiction, poetry and drama. On this point Casper (1995: 5) asks:

> Was it because established novelists, poets and playwrights were assumed to be beneficiaries of 'capitalist imperialism'…whose ambitions coincided therefore with Marcos's own? Or had the authors turned to trivia as a safeguard, abandoning a long tradition of polemicism in literature? Or could it be that Marcos considered such works irrelevant inventions, temporary entertainments…?

Wherever the explanation may lie, it can be argued that limited readership and social impact lie behind the observed 'uptightness' of PhilE fiction and, in the same vein, that of PhilE learned writing.

5 Conclusion

In this article we have reported the findings of a corpus-based study of stylistic change in a non-native variety of English, PhilE, alongside its 'parent variety', AmE. Our investigation has focused on possible signs of colloquialisation, a well-noted diachronic trend for previously formal, literary writing to shift to informal, speech-like styles. Adopting a bottom-up approach, we have derived a comprehensive measure of colloquiality based on a total of 86 grammatical features. This measure, itself a summary of 25 strong indicators of the speech–writing divide, allows us to determine and compare the extent to which different texts favour linguistic features that are typical of speech, or colloquial features, and disfavour those typical of writing, or anti-colloquial features. According to our approach, colloquialisation is interpreted as a dual process involving not only the shift of writing towards a speech-like style, but also the shift away from a writing-like style. We believe a combination of the two opposing trends better encapsulates the process of 'writing becoming more like speech' than any single trend.

We have employed the measure of colloquiality to analyse texts in a diachronic, parallel corpus of PhilE and AmE. Drawn in part from the Brown corpora and the *International Corpus of English* family, the diachronic corpus consists of three written registers with distinct situational characteristics and spans around thirty years from the 1960s to 1990s. Our analysis of colloquiality scores has revealed several noteworthy patterns. Regarding register variation, we have seen that when transformed to a common scale it is anti-colloquial features, not colloquial features, that most clearly signal the difference between creative and informative writing in contemporary English. Texts representing these two registers are not drastically different in their use of colloquial features. Rather, the most remarkable differences on a global level lie in the frequencies of anti-colloquial features, which indicate overall lexical diversity and informational richness.

Regarding diachronic variation, evidence for colloquialisation varies across registers. There are considerable increases in the colloquiality scores of Philippine press editorials and American fiction over the time span under investigation. On the other hand, learned writing has not shown remarkable changes irrespective of variety. We have argued that the distinction drawn by Hundt & Mair (1999) between 'agile' and 'uptight' registers is particularly useful for interpreting the diachronic findings. Popular registers which are driven by the need to cater to a large readership tend to be open towards stylistic innovations, whereas specialised registers are less receptive to change with their small and stable audiences. Differences in the nature of the intended audience account not only for the different findings for the two popular registers and learned writing, but also for the contrast between the rapid colloquialisation of PhilE press editorials and the stability of PhilE fiction during the same time period.

Importantly, we have seen that the evolution of PhilE registers cannot be explained by a simple process involving emulation of AmE. This is inevitable given the unique sociohistorical circumstances in which PhilE has evolved. PhilE's colonial history

imparts to it an elitist character, placing it in a hierarchical relationship with the local languages. The patterns described in this study lend support to the general observation made in previous research that PhilE is less colloquial than AmE. However, there is no convincing evidence for monostylisticism since stylistic differentiation is on the whole fairly marked in PhilE (despite an affinity between press and learned writing of the 1960s). The conclusion is that PhilE speakers are no less sensitive to the stylistic conventions of the three registers than native English speakers. However, future research is required to see whether the same can be said about spoken and other informal registers.

Authors' addresses:

School of Foreign Languages
Renmin University of China
59 Zhongguancundajie, Haidian
Beijing 100872
China
xinyue.yao@ruc.edu.cn

School of Humanities and Languages
University of New South Wales
Sydney NSW 2052
Australia
p.collins@unsw.edu.au

References

Aarts, Bas, Joanne Close, Geoffrey Leech & Sean Wallis (eds.). 2013. *The verb phrase in English: Investigating recent language change with corpora*. Cambridge: Cambridge University Press.

Alberca, Wilfredo. 1978. The distinctive features of Philippine English in the mass media. Unpublished PhD thesis, University of Santo Tomas, Manila.

Bautista, Ma. Lourdes. 2000. *Defining standard Philippine English: Its status and grammatical features*. Manila: De La Salle University Press.

Bautista, Ma. Lourdes. 2004. The verb in Philippine English: A preliminary analysis of modal *would. World Englishes* 23(1), 113–28.

Bautista, Ma. Lourdes (ed.). 2011. *Studies of Philippine English: Exploring the Philippine component of the International Corpus of English*. Manila: Anvil Publishing.

Bautista, Ma. Lourdes & Kingsley Bolton (eds.). 2008. *Philippine English: Linguistic and literary perspectives*. Hong Kong: Hong Kong University Press.

Biber, Douglas. 1986. Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language* 62(2), 384–414.

Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, Douglas. 2003. Compressed noun-phrase structures in newspaper discourse: The competing demands of popularization vs. economy. In Jean Aitchison & Diana Lewis (eds.), *New media language*, 169–81. London: Routledge.

Biber, Douglas & Susan Conrad. 2009. *Register, genre and style*. Cambridge: Cambridge University Press.

Biber, Douglas & Edward Finegan. 1989. Drift and the evolution of English style: A history of three genres. *Language* 65(3), 487–517.

Biber, Douglas & Edward Finegan. 1997. Diachronic relations among speech-based and written registers in English. In Terttu Nevalainen & Leena Kahlas-Tarkka (eds.), *To explain the present: Studies in the changing English language in honour of Matti Rissanen*, 253–75. Helsinki: Société Néophilologique.

Biber, Douglas & Bethany Gray. 2011. Grammatical change in the noun phrase: The influence of written language use. *English Language and Linguistics* 15(2), 223–50.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.

Borlongan, Ariane Macalinga. 2009. A survey on language use, attitudes, and identity in relation to Philippine English among young generation Filipinos: An initial sample from a private university. *The Philippine ESL Journal* 3(1), 74–107.

Borlongan, Ariane Macalinga. In progress. The Phil-Brown Corpus project (accessed November 2011).

Bureau of Census and Statistics, Department of Commerce and Industry, Republic of the Philippines. 1954. *Summary and general report on the 1948 census of population and agriculture*, part I: *Population* Manila: Bureau of Printing.

Casper, Leonard. 1995. *The opposing thumb: Decoding literature of the Marcos regime*. Quezon City: Giraffe Books.

Collins, Peter. 2009. Modals and quasi-modals in world Englishes. *World Englishes* 28(3), 281–92.

Collins, Peter (ed.) 2015a. *Grammatical change in English world-wide*. Amsterdam: John Benjamins.

Collins, Peter. 2015b. Recent diachronic change in the progressive in Philippine English. In Collins (ed.), 271–96.

Collins, Peter, Ariane Macalinga Borlongan & Xinyue Yao. 2014. Modality in Philippine English: A diachronic study. *Journal of English Linguistics* 42(1), 68–88.

Collins, Peter & Xinyue Yao. 2013. Colloquial features in World Englishes. *International Journal of Corpus Linguistics* 18(4), 479–505.

Dayag, Danilo. 2004. The English-language media in the Philippines. *World Englishes* 23(1), 33–45.

Dayag, Danilo. 2008. English-language media in the Philippines: Description and research. In Bautista & Bolton (eds.), 49–66.

Dayag, Danilo. 2012. Philippine English. In Ee-Ling Low & Azirah Hashim (eds.), *English in Southeast Asia: Features, policy and language in use*, 91–100. Amsterdam: John Benjamins.

Dita, Shirley. 2011. The grammar and semantics of adverbial disjuncts in Philippine English. In Bautista (ed.), 33–50.

Du Bois, John W., Wallace L. Chafe, Charles Meyer, Sandra A. Thompson, Robert Englebretson & Nii Martey. 2000–5. *Santa Barbara Corpus of Spoken American English*, parts 1-4. www.linguistics.ucsb.edu/research/santa-barbara-corpus (accessed December 2015).

Francis, W. Nelson & Henry Kučera. 1964, 1971, 1979. *A Standard Corpus of Present-Day Edited American English, for Use with Digital Computers (Brown)*. www.helsinki.fi/varieng/CoRD/corpora/BROWN/ (accessed December 2015).

Gonzalez, Andrew (ed.). 1988. *The role of English and its maintenance in the Philippines*. Manila: Solidarity.

Gonzalez, Andrew. 1991. Stylistic shifts in the English of the Philippine print media. In Jenny Cheshire (ed.), *English around the world: Sociolinguistic perspectives*, 333–63. Cambridge: Cambridge University Press.

Gonzalez, Andrew. 1992. Philippine English. In Tom McArthur (ed.), *The Oxford companion to the English language*, 765–7. Oxford: Oxford University Press.

Gonzalez, Andrew. 1997. The history of English in the Philippines. In Ma. Lourdes Bautista (ed.), *English is an Asian language: The Philippine context*, 7–16. Sydney: Macquarie Library Ltd.

Gonzalez, Andrew. 2004. The social dimensions of Philippine English. *World Englishes* 23(1), 7–16.

Gonzalez, Andrew. 2008. A favorable climate and soil: A transplanted language and literature. In Bautista & Bolton (eds.), 13–27.

Gonzalez, Andrew & Ma, Lourdes Bautista. 1986. *Language surveys in the Philippines (1966-1984)*. Manila: De La Salle University Press.

Hackert, Stephanie & Dagmar Deuber. 2015. American influence on written Caribbean English: A diachronic analysis of newspaper reportage in the Bahamas and in Trinidad and Tobago. In Collins (ed.), 389–410.

Hau, S. Caroline. 2008. The Filipino novel in English. In Bautista & Bolton (eds.), 317–36.

Hundt, Marianne. 1997. Has BrE been catching up with AmE over the past thirty years? In Magnus Ljung (ed.), *Corpus-based studies in English*, 135–51. Amsterdam: Rodopi.

Hundt, Marianne & Christian Mair. 1999. 'Agile' and 'uptight' genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics* 4(2), 221–42.

ICE: *International Corpus of English*. http://ice-corpora.net/ice/ (accessed December 2015).

Kortmann, Bernd & Kerstin Lunkenheimer (eds.) 2013. *The electronic world atlas of varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in contemporary English: A grammatical study*. Cambridge: Cambridge University Press.

Lim, JooHyuk & Ariane Macalinga Borlongan. 2011. Tagalog particles in Philippine English: The case of *ba*, *na*, *'no* and *pa*. *Philippine Journal of Linguistics* 42, 59–74.

Llamzon, Teodoro. 1969. *Standard Filipino English*. Quezon City: Ateneo de Manila University Press.

Mair, Christian & Marianne Hundt. 1995. Why is the progressive becoming more frequent in English? A corpus-based investigation of language change in progress. *Zeitschrift für Anglistik und Amerikanistik* 43(2), 111–22.

Mair, Christian & Claudia Winkle. 2012. Change from *to*-infinitive to bare infinitive in specificational cleft sentences: Data from World Englishes. In Marianne Hundt & Ulrike Gut (eds.), *Mapping unity and diversity world-wide: Corpus-based studies of New Englishes*, 243–62. Amsterdam: John Benjamins.

Martin, Isabel Pefianco, 2008. Colonial tradition and the shaping of Philippine literature in English. In Bautista & Bolton (eds.), 245–60.

Martin, Isabel Pefianco. 2010. Periphery ELT: The politics and practice of teaching English in the Philippines. In Andy Kirkpatrick (ed.), *The Routledge handbook of world Englishes*, 247–64. London: Routledge.

Martin, Isabel Pefianco. 2014a. Philippine English revisited. *World Englishes* 33(1), 50–9.

Martin, Isabel Pefianco. 2014b. Beyond nativisation? Philippine English in Schneider's Dynamic Model. In Sarah Buschfeld, Thomas Hoffmann, Magnus Huber & Alexander Kautzsch (eds.), *The evolution of Englishes: The Dynamic Model and beyond*, 70–85. Amsterdam: John Benjamins.

Mijares, Primitivo. 1976. *The conjugal dictatorship of Ferdinand and Imelda Marcos I*. San Francisco: Union Square Publications.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik 1985. *A comprehensive grammar of the English language*. London: Longman.

Schneider, Edgar. 2007. *Postcolonial English: Varieties of English around the world*. Cambridge: Cambridge University Press.

Sibayan, Bonifacio. 1994. Philippine language problems. In Jasmin E. Acuña (ed.), *The language issue in education*, 47–86. Manila & Quezon City: Congress of the Philippines,

Social Weather Stations, 2006. *March 2006 Social Weather Survey: National proficiency in English declines*. Social Weather Stations report. www.sws.org.ph/pr060418.htm (accessed 10 September 2015).

Szmrecsanyi, Benedikt. 2013. *Grammatical variation in British English dialects*. Cambridge: Cambridge University Press.

Thompson, Roger. 2003. *Filipino English and Taglish: Language switching from multiple perspectives*. Amsterdam: John Benjamins.

Van Rooy, Bertus & Ronel Wasserman. 2014. Do the modals of Black and White South African English converge? *Journal of English Linguistics* 42(1), 51–67.

Wanner, Anja. 2009. *Deconstructing the English passive*. Berlin: Mouton de Gruyter.

Westin, Ingrid. 2002. *Language change in English newspaper editorials*. Amsterdam: Rodopi.

Xiao, Richard. 2009. Multidimensional analysis and the study of world Englishes. *World Englishes* 28(4), 421–50.