

## **AusBrown: A new diachronic corpus of Australian English**

*Peter Collins, University of New South Wales and  
Xinyue Yao, Renmin University of China*

### ***Abstract***

*This paper presents a newly-compiled diachronic corpus of Australian English (AusBrown). With four sampling time points (1931, 1961, 1991 and 2006), AusBrown is designed to match the current suite of British and American ‘Brown-family’ corpora in both sampling year and design. We provide details of the composition and compilation of AusBrown, and explore the broader context of its ‘Brown-family background’ and of complementary Australian corpora. We also overview research based on the Australian corpora presented, including several AusBrown-based papers.*

### ***1 Introduction***

In this paper we introduce a new Brown-family corpus of Australian English (AusE), ‘AusBrown’, compiled by the authors of this paper. In addition to providing information on the design and compilation of AusBrown, we identify other corpora of AusE, discuss AusBrown’s ‘family background’, and comment on related research.

What is the rationale for the compilation of AusBrown? The current suite of British and American ‘Brown-family’ corpora, spanning the period 1931–2006, address the need for parallel corpora that can be used for comparative diachronic research. AusBrown is not the only Australian corpus in existence (see further below), but it is the only one that addresses the same needs as its British and American Brown-family counterparts, thereby paving the way for comparative research on changing patterns in AusE, British English (BrE) and American English (AmE) from the early 20<sup>th</sup> century to the early 21<sup>st</sup> century. More specifically, AusBrown provides a resource for scholars to explore the development of AusE as it progressed towards linguistic independence – or ‘endonormativised’ (Schneider 2007) – over the course of the 20<sup>th</sup> century (see also Moore

2008), and to compare developments in AusE with those in the AusE's 'parent variety', BrE, and in the most transnationally influential variety of English, AmE.

The structure of the remainder of the paper is as follows. The next two sections deal in turn with the dual contextual backgrounds – “Aus” and “Brown” – of our new corpus, as reflected in its name. Section 2 surveys previous corpora of AusE, and the research based on them, and Section 3 discusses the extended Brown family of corpora. In Section 4 we focus on the design and compilation details of AusBrown. Section 5 overviews research that has been based on AusBrown and other diachronic corpora of AusE. Section 6 is reserved for our concluding remarks.

## **2 Corpora of Australian English**

The development of computer corpora in Australia dates back to the 1980s. The first modest foray was the Melbourne-Surrey Corpus, a collection of newspaper editorials from the *Melbourne Age* 1980–81, which was designed to match a small (100,000-word) British corpus of editorials from the *Times* (Ahmad and Corbett 1987).

In 1986 Pam Peters, Peter Collins and David Blair undertook compilation of the Australian Corpus of English (ACE). ACE was designed to match the Brown Standard Corpus of Present-day Edited American English (Brown) and the Lancaster-Oslo/Bergen Corpus (LOB). The design of ACE and its parallels is detailed in the next Section. Suffice it to add here that the sampling year for ACE, 1986 (symbolically a quarter-century later than 1961, the sampling year for Brown and LOB), was chosen for two main reasons: a number of the genres in Brown and LOB were not published in Australia in 1961, and at the same time we wanted to provide a contemporary reference corpus. Another point of difference with Brown and LOB involved the fiction samples: the relative unavailability of some fiction genres required adjustments, but the overall 3:1 ratio of non-fiction:fiction was maintained (see Table 1 below). For more information on the make-up of ACE see <http://www.hit.uib.no/icame/ace/ace-man.htm>, and Peters 1987; Peters et al. 1988; Collins and Peters 1988; Green and Peters 1991. ACE has been used in a large number of studies, mainly by Peters (on grammatical/morphological usage and style: 1993, 1994, 1998, 2001), and Collins (on modality 1988, 1991a, 1991b; existentials 1992; extraposition 1994a; comparative clauses 1994b; indirect objects 1995; passives 1996; *let*-imperatives 2004; and exclamatives 2005a), and Collins and Peters (2004/8).

The next Australian corpus to be compiled (by Peters) was ICE-AUS, the Australian component of the International Corpus of English, a large project initiated in 1988 by the late Sidney Greenbaum, the then Director of the Survey of English Usage. The aim was to prepare a set of one million word corpora representing the English of countries where English is either a first or second language. ICE corpora compiled thus far represent Australia, Cameroon, Canada, East Africa (Kenya, Malawi, Tanzania), Fiji, Great Britain, Hong Kong, India, Ireland, Jamaica, Kenya, Malta, Malaysia, New Zealand, Nigeria, Pakistan, Philippines, Sierra Leone, Singapore, South Africa, Sri Lanka, Trinidad and Tobago, USA. Each corpus was to comprise a predetermined set of both spoken texts (60%) and written texts (40%), with samples collected between 1990 and 1994 (a temporal window which had to be moved forward for some of the more recent corpora): see <http://www.ucl.ac.uk/english-usage/projects/ice-gb/index.htm>. ICE-AUS has been used in a wide range of studies. Most of the chapter authors in Peters, Collins and Smith's (2009) volume *Comparative studies in Australian and New Zealand English: Grammar and beyond* use data from ICE-AUS (along with other corpora), as they were enjoined to do by the editors: Allan and Burridge; Collins; Elsness; Holmes, Sigley and Terraschke; Hundt; Mair; Miller; Mulder, Thompson and Williams; Peters; Peters and Funk; Peterson; Quinn; and Smith. ICE-AUS is also used in Peters' (2007) *Cambridge guide to Australian English Usage*, and Collins's (2009a) book *Modals and quasi-modals in English*. Other studies include Peters (on comparison 1996; subjunctive 1998; negation 2008); Collins (on *get*-passives 1996; modality 2005b, 2007, 2009b; progressives 2008, information-packaging 2010; and *there*-existentials 2011); Collins and Yao (on modality 2011a; the verb phrase 2011b; and the perfect aspect 2012).

A more specialised Australian corpus is the Australian Radio Talkback (ART) corpus, comprising 257,000 words of unscripted spoken language, some from public radio (the ABC) and some from commercial radio, recorded in 2004–06. ART was used by a number of the authors in Peters et al. (2009) (Allan and Burridge; Elsness; Holmes, Sigley and Terraschke; Miller; Peters; and Smith), and by Peters and Collins (2012) in their study of colloquialism.

All of the Australian corpora discussed thus far have been synchronic in design. More recent corpora have been diachronic, like AusBrown. The first of these to be compiled was the two million word Corpus of Oz Early English (COOEE), compiled by Clemens Fritz, which comprises texts of various sizes from more than a hundred different sources, including books, letters, diaries, proclamations, and newspapers (for corpus information see Fritz 2007b). The time span covered, 1788–1900, is divided into four periods which “not only

mark important changes in the history of white Australia, but also signify transitions in the linguistic history of early AusE” (Fritz 2007b: 69): 1788–1825 (“the penal colony”), 1826–1850 (“settlers and emancipists”), 1851–1875 (“the impact of the gold rushes”), and 1876–1900 (“nativisation and urbanisation”), each represented by 500,000 words of text. In every period there is also a similar number of words from each of the four text registers, as follows (along with their total number of words and percentage of the corpus): speech-based (75,000/15%), private written (175,000/35%), public written (200,000/40%), and government English (50,000/10%). In addition to the COOEE-based publications by Fritz listed in Section 5, there are four others, as listed below for AusCorp and discussed in Section 5 (Collins 2014, 2015b; Collins and Yao 2014; and Yao 2015).

AusBrown is not the first diachronic corpus of AusE prepared by the present authors. In 2012–13, as part of an Australian Research Grant-funded project, we compiled AusCorp, a 20<sup>th</sup> century corpus whose text samples are organised decade-by-decade and which comprises c.340,000 words of news and fiction samples. The news section contains around 170,000 words sourced from major Australian newspapers such as the *Sydney Morning Herald* and *The Australian*. The fiction section also comprises around 170,000 words, with some bias towards the late 20th century. Text-extracts are taken from two sources: anthologies of Australian literature such as *The Macquarie PEN anthology of Australian literature* (Jose 2009); and the *Digital Archive of Colonial Australian Popular Fiction*, an online collection of Australian fiction produced during the period spanning the 19th century and early 20<sup>th</sup> century (<http://www.apfa.esrc.unimelb.edu.au/>). For each writer, only about 2,000 words were selected. Four of the studies discussed in Section 5 below make use of AusCorp data: Collins (2014, 2015b), Collins and Yao (2014), and Yao (2015).

Finally, the Australian Diachronic Hansard Corpus (ADHC), a collection of parliamentary debates from the House of Representatives, is still under development (by Haidee Kruger and Adam Smith of Macquarie University). In its current form, the corpus consists of 2.3 million words, sampled across five equally-spaced periods corresponding to major social changes in Australia: 1901–1905; 1931–35; 1961–5; 1991–5; and 2011–15. ADHC is projected to consist of 15–20 million words in its final form (see Kruger and Smith 2018).

### **3 Survey of the Brown family of corpora**

The Brown family of corpora, whose ‘core’ members represent AmE, and BrE, also includes as ‘extended family’ members AusE, Indian English (IndE), New

Zealand English (NZE) and Philippine English (Phile). The make-up of the Brown-family corpora is displayed in Table 1 below (with that of AusBrown discussed in Section 4).

The four ‘core’ members of the Brown family are the Brown, LOB, Frown and FLOB corpora. The first to be compiled – in 1963/4 by W. Nelson Francis and Henry Kučera at Brown University was Brown (“Standard Corpus of Present-day Edited American English”).<sup>1</sup> Brown contains over 1 million words (500 samples of c.2,000+ words each) of running text of edited English prose printed in the United States during the calendar year 1961, and distributed across fifteen text categories – nine informative and six imaginative (see Table 1). We are indebted to Francis and Kučera for recognising the value of what is now referred to as a ‘reference corpus’: a corpus designed as a general linguistic research resource encompassing a wide range of genres.

The next member of the ‘Brown family’ to appear, a British parallel to Brown, LOB (the Lancaster-Oslo/Bergen Corpus, originally the Lancaster Corpus), was prepared during the 1970s by Geoffrey Leech (Lancaster) and his Norwegian colleagues Stig Johansson (Oslo), and Knut Hofland (Bergen). The preparation of two further ‘core’ Brown family members commenced in 1991, when a team headed by Christian Mair (Albert-Ludwigs-Universität Freiburg) undertook the compilation of a Freiburg-based update of LOB (F-LOB). A sampling year of 1991, three decades on from that for Brown and LOB, was selected in order to provide linguists with an empirical basis for the study of language change in progress. Shortly after work began on F-LOB, a second project was initiated by the same Freiburg team with a 1992 Brown-update called Frown.

Four more recent British and American additions to the Brown family have further extended its chronological span. Two of these have 1931 as their target sampling year: the Before-LOB Corpus (B-LOB), prepared by Geoffrey Leech and Paul Rayson (Lancaster), and the B-Brown corpus, prepared by Marianne Hundt (Zurich). The other two, with a median sampling point of 2006 – British English 2006 (BE06) and American English 2006 (AE06) – have been compiled by Paul Baker (Lancaster).<sup>2</sup> Also, reportedly at an advanced stage of compilation, is a comparable LOB corpus for 1901.<sup>3</sup>

Projects in countries other than England and The United States, have produced further matching corpora, albeit with some adjustments to the original generic design necessitated by local differences.

During the early 1980s S. V. Shastri (Shivaji University) and colleagues compiled the Kolhapur Corpus of Indian English. 1978, symbolically thirty years after Independence, was set as the sampling year. Modifications to the composition of the fiction section of the corpus – see Table 1 – were necessi-

tated by the fact that “the amount and kind of imaginative writings in a second language situation such as in India is very different from that in a first language one such as the American or the British situation” (Shastri 1986 n.p.).

Shortly after the commencement of the ACE project (discussed in the previous Section), a team headed by Laurie Bauer (Victoria University) prepared the Wellington Corpus of Written New Zealand English (WCWNZE). The selection of 1986 as the sampling year, as for ACE, acknowledges the importance of comparisons between Australian and New Zealand data.

In 2011, Ariane Borlongan (De La Salle University, Manila) compiled ‘Phil-Brown’, with 1961 as median sampling year. Even though text category sizes are not parallel to Brown, some Brown categories are not represented, and the million-word target was not achieved (see Table 1), Phil-Brown has proven to be – in conjunction with (the written component of) ICE-Phil – a valuable resource for studies of diachronic variation in contemporary (written) PhilE: Collins, Borlongan, Lim and Yao (2014); Collins, Borlongan and Yao (2014); Collins, Yao and Borlongan (2014); Borlongan and Dita (2015); Collins (2015c; 2016); Collins and Borlongan (2017). Borlongan is currently compiling a Philippine Before-Brown corpus (PBB) at the Tokyo University of Foreign Studies (pc).

*Table 1:* The composition of the Brown-family corpora: target sampling years are indicated in brackets and figures represent numbers of 2,000-word text samples

	B-LOB (1931)	B-Brown (1931)	Brown (1961)	LOB (1961)	Phil-Brown (1961)	Kolhapur (1978)	ACE (1986)	WCWNZE (1986)	FLOB (1991)	Frown (1992)	BE06 (2006)	AE06 (2006)
A Press reportage	44	44	44	44	14	44	44	44	44	44	44	44
B Press editorials	27	27	27	27	15	27	27	27	27	27	27	27
C Press reviews	17	17	17	17	29	17	17	17	17	17	17	17
D Religion	17	17	17	17	65	17	17	17	17	17	17	17
E Skills, trades and hobbies	38	36	36	38	-	38	38	38	38	36	38	36
F Popular lore	44	48	48	44	-	44	44	44	44	48	44	48
G Belles lettres, biographies, etc.	77	75	75	77	30	70	77	77	77	75	77	75
H Miscellaneous	30	30	30	30	59	37	30	30	30	30	30	30

J Learned	80	80	80	80	41	80	80	80	80	80	80	80
K General fiction	29	29	29	29	45	58	29	126	29	29	29	29
L Mystery and detective fiction	24	24	24	24	-	24	15	-	24	24	24	24
M Science fiction	6	6	6	6	-	2	7	-	6	6	6	6
N Adventure and western fiction	29	29	29	29	30	15	8	-	29	29	29	29
P Romance and love story	29	29	29	29	7	18	15	-	29	29	29	29
R Humour	9	9	9	9	-	9	15	-	9	9	9	9
S Historical fiction	-	-	-	-	-	-	22	-	-	-	-	-
W Women's fiction	-	-	-	-	-	-	15	-	-	-	-	-
Total	500	500	500	500	335	500	500	500	500	500	500	500

#### 4 *AusBrown*

The most recent addition to the Brown family is a suite of four Australian ‘mini-’ or ‘sub-’ corpora compiled by the present authors, which we refer to collectively as ‘AusBrown’ (see Table 2 below for the composition of AusBrown).<sup>4</sup> The target sampling years for the four subcorpora match those of their British and American counterparts: 1931 (B-LOB and B-Brown), 1961 (LOB and Brown), 1991 (FLOB and Frown), 2006 (BE06 and AE06), the only exception being Frown, which was sampled in 1992. The epithet ‘target’ reflects the fact that, in the case of the first three sampling dates the occasional paucity of available texts required extending the sampling period several years before and after these specific years. Currently AusBrown comprises c.720,000 words, with each of the four subcorpora comprising c.180,000 words (90 texts, each c.2,000 words) divided evenly among the three much-studied and arguably most important Brown genre categories: fiction, learned and press reportage. When AusBrown is completed it will comprise c.960,000 words (each subcorpus containing 120 texts, or c.240,000 words). The choice of genres selected had both a practical motivation (AusE materials are considerably easier to assemble for fiction, learned and press than for the other Brown categories) and a theoretical one (these three influential genres cover much of the range of variation in printed English and thus, not surprisingly, have featured prominently in corpus-based register-oriented studies, such as Hundt and Mair 1991, and Biber et al. 1999).

All AusBrown fiction and learned authors were subjected to biographical investigation to check for ‘Australianness’ (i.e. they had to have been born in Australia, or to have migrated to Australia when less than ten years of age). A further basis for exclusion from AusBrown was evidence of extended periods of time spent overseas, a not uncommon phenomenon for a country of immigrants, as Hundt (2012) has noted for New Zealand.

Each of the three genres presented its own set of sampling challenges. For fiction, given that each subcorpus contains only 30 texts (by comparison with 126 texts in their one-million-word British and American counterparts), we could do little more than attempt to represent as many of the fictional subcategories (as listed in Table 1) as was feasible in view of the comparatively smaller Australian market, especially in the 1930s and 1960s. Our fiction texts were collected from novels, short story collections and literature anthologies. They were mainly scanned from library copies, and then converted from .pdf to .txt files. Most of the literary extracts were taken from Australian publishers, so as to reduce the impact of regional variation with regard to editorial preferences, but occasionally we had to resort to non-Australian publishers to fill our quota. For the learned category, following the disciplinary composition of Category J of the Brown-family corpora, we sought to sample a roughly equal proportion of texts from the humanities (history, philosophy, literary criticism, etc.), social sciences (politics, education, psychology, etc.), and hard sciences (biology, veterinary science, engineering, etc.). Our learned texts were collected primarily from academic journals and books published in Australia, many of which were available in digital form. Occasionally, in order to meet quota requirements, it was necessary to include a non-Australian publisher, and – in the absence of relevant biographical information – to accept affiliation with an Australian institution as evidence of an author’s “Australianness”.

For press reportage, texts were mainly downloaded from the Australian National Library’s Trove online search engine (<https://trove.nla.gov.au/>), which provides a digitalised archive of historical Australian newspapers dating back to the 19<sup>th</sup> century. We chose news reports published during the years 1931, 1961 and 1991, with months and dates randomised. Together they cover thirteen newspapers with relatively long and continuous publication records, including *The Sydney Morning Herald*, *The Newcastle Sun*, *The Canberra Times* and so on. For 2006, which is not covered in Trove, texts were selected from the ProQuest database ([www.proquest.com/libraries/academic/news-newspapers](http://www.proquest.com/libraries/academic/news-newspapers)) using the same random sampling as for the three earlier periods.

A special note has to be added regarding the 1991 AusBrown subcorpus. Part of the data for this subcorpus was taken from the ICE-AUS corpus, which was sampled for the same time but does not offer enough texts (20 for press



reportage, 20 for fiction and 40 for academic writing, equally partitioned into humanities, social sciences, natural sciences, and technology).

After all texts were catalogued, a random sample of around 2,000 words from each text was selected and converted into .txt format through OCR software. The output was manually proofread to eliminate possible errors arising from the automatic recognition process. Further subsequent processing of the data was required, in preparing the two grammar-focused AusBrown publications prepared thus far (Collins and Yao 2018; Yao and Collins forthcoming): see Section 5 below. To facilitate the search for grammatical features, raw texts were part-of-speech tagged with CLAWS, and the C7 tagset was chosen to maintain consistency with the existing Brown corpora (<http://ucrel.lancs.ac.uk/claws/>).

Table 2: Composition of AusBrown as of March 2018: figures represent numbers of 2000-word text samples

	Target Sampling Year			
	1931	1961	1991	2006
A Press reportage	30	30	30	30
J Learned	30	30	30	30
K General fiction	30	30	30	30
Total	90	90	90	90

## 5 Corpus-based diachronic studies of AusE grammar

This sections surveys corpus-based diachronic studies of AusE grammar – including those based on AusBrown – which began to appear a little over a decade ago, and which provided the rationale for the preparation of AusBrown. We can identify two publications that have provided a stimulus for many of these studies. The first of these is the pioneering work of Geoffrey Leech and his colleagues on short-term change in BrE and AmE, in their landmark 2009 volume, *Change in Contemporary English*. This study demonstrated the power of parallel Brown-family corpora to shed light on short-term change in a range of grammatical categories, including the modal auxiliaries, progressive, subjunctive, passive, genitive and relative clauses. Further studies of the grammar of BrE and AmE, using both spoken and written diachronic corpora, were reported in Aarts et al. (2013). More recently linguists have begun to undertake diachro-

nic studies of a number of postcolonial varieties of English, including AusE. One catalyst for such studies – the second of the two stimuli to which we have referred above – was the special issue of the *Journal of English Linguistics* (Volume 42, 2014, “Diachronic Approaches to Modality in World Englishes”) in which the editors, Dirk Noël, Johan Van der Auwera and Bertus Van Rooy, argued the case cogently and forcefully for addressing this neglected area. Collins’s study of modality in the 19<sup>th</sup> and 20<sup>th</sup> centuries in AusE, BrE and AmE in this issue (Collins 2014), based on data from COOEE, AusCorp and ARCHER (A Representative Corpus of Historical English Registers), shows AusE to have been evolving in the same general direction as the two longer-established varieties, albeit with its endonormative independence being asserted in its users’ more extreme dispreference for the modals and their reluctance to embrace the quasi-modals as enthusiastically as users of AmE.

Pioneering forays into 19<sup>th</sup> century AusE morphosyntax, using COOEE data, were made by Fritz (2006, 2007a, 2007b). More recently Collins and Yao (2014) have used data from COOEE and AusCorp to examine changes in various VP categories in AusE over the same period. Collins and Yao find a sufficient number of divergences in frequency and rate of change between AusE and the BrE and AmE reference varieties to suggest a degree of linguistic independence for the Antipodean variety.

Collins’s (2015a) volume – the inspiration for which he acknowledges to have been Dirk Noël et al.’s special issue of *JEngL* – applies and extends the techniques of corpus linguistics and diachronic linguistics to the task of describing and explaining grammatical change in English varieties (or sub-varieties in some cases) other than the two supervarieties. AusE is the focus of four chapters: Collins’s chapter (Collins 2015b) uses data extracted from COOEE and AusCorp to examine developments in ten morphosyntactic variables: *-t/-ed* past verb forms, *’s*-genitives, the mandative and *were*-subjunctive, concord with collective nouns, light verbs, non-finite complementation with *help* and *prevent*, *do*-support, and *be*-passives). Australian usage is found to have diverged from that of its British colonial parent towards that of AmE. Peters’s (2015) chapter studies developments with ‘dual adverbs’ such as *slow/slowly* using data from ICE-AUS and ICE-GB representing the late 20<sup>th</sup> century, and two subcorpora representing the second half of the 19<sup>th</sup> century: one from COOEE and the other opportunely assembled from British novels and news reports of this period. In Rodriguez Louro’s chapter analyses of oral history data derived from 39 speakers of AusE born between 1874 and 1983 lead her to suggest that the grammaticalisation of *think* was essentially a late 20<sup>th</sup> century phenomenon. Yao (2015)

uses data from COOEE and AusCorp to examine shifts in the use of the present perfect and the preterite in AusE.

Two papers report independent studies of colloquialisation in AusE. Collins and Yao (2018), the first AusBrown-based study, investigates the influence of (grammatical) colloquialisation on developments in the grammar of AusE, a variety noted for its penchant for (lexical) colloquialism, against the historical backdrop of the progressive decline of Britishness in Australia. Kruger and Smith (2018) use a spoken rather than written corpus – ADHC (1901–2015; see Section 2 above) – to examine the role of colloquialisation (and densification) in AusE over this period.

Yao and Collins (forthcoming) is the first study to investigate the overall development of AusE grammar in the 20<sup>th</sup> century using AusBrown and the full suite of its British and American counterparts. The study uses a pioneering aggregate approach to investigate grammatical variability in the three varieties at four time points. Over the period investigated, BrE is found to follow the lead of AmE, moving in the same direction but at a faster rate, so that by the early 20<sup>th</sup> century the gap between the two varieties is considerably smaller than in the 1930s. The rapidity of change in AusE is noted to have been particularly strong in the three-decade period from the 1960s to the 1990s (unsurprisingly, Yao and Collins argue, citing the work of historians who have observed these years to have been a time of rapid social progress in Australia).

Finally, Peter Collins, Xinyue Yao, Haidee Kruger, Adam Smith and Minna Korhonen (ms) address the question of why registers change differently over time in different regional varieties. Multidimensional analyses are used to explore grammatical change in four major registers of English: three written registers (press reportage, fiction, learned writing) in AusE, BrE and AmE, and one speech-based register (edited parliamentary speech, or ‘Hansard’) in AusE and BrE. Data sources are AusBrown, ADHC, and a collection of British Hansard texts compiled by the authors.

## **6 Conclusion**

The primary purpose of this paper has been to introduce AusBrown as a new resource for the diachronic investigation of AusE which, despite its relatively modest size by comparison with the other matching members of the Brown-family of corpora, has already demonstrated its capacity to shed light on the grammatical development of AusE in the 20<sup>th</sup> century. A secondary aim has been to contextualise AusBrown, by providing details of its Brown-family background and of complementary Australian corpora. Finally, we have also

summarised research based on the Australian corpora presented, including three AusBrown-based papers. Anyone seeking access to AusBrown is welcome to contact us.

### Notes

1. An important caveat regarding the use of the word ‘standard’ in the title of the Corpus is that it “does not in any way mean that it is put forward as ‘standard English’; it merely expresses the hope that this corpus will be used for comparative studies where it is important to use the same body of data” (Francis and Kučera 1964; available at <http://clu.uni.no/icame/manuals/BROWN/INDEX.HTM>).
2. B-LOB, B-Brown, BE06 and AE06 are all accessible through <https://cqp-web.lancs.ac.uk/> (the Corpus Query Processor maintained by Andrew Hardie at Lancaster University).
3. See <http://www.helsinki.fi/varieng/CoRD/corpora/BLOB-1931/> and Baker (2009).
4. This terminological practice is comparable in some ways to the use of the term International Corpus of English (ICE) to refer to a collection of individual national variety corpora (ICE-GB, ICE-HK, ICE-AUS, etc.).

### References

- Aarts, Bas, Joanne Close, Geoffrey Leech and Sean Wallis (eds.). 2013. *The verb phrase in English: Investigating recent language change with corpora*. Cambridge: Cambridge University Press.
- Ahmad, Kurshid and Greville Corbett. 1987. The Melbourne-Surrey Corpus. *ICAME Journal* 11: 39–43.
- Baker, Paul. 2009. The BE06 Corpus of British English and recent language change. *International Journal of Corpus Linguistics* 14: 312–337.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.
- Collins, Peter. 1988. The semantics of some modals in contemporary Australian English. *Australian Journal of Linguistics* 8: 261–286.
- Collins, Peter. 1991a. *Will and shall* in Australian English. In S. Johansson and A-B. Stenström (eds.). *English computer corpora*, 181–200. Berlin: Mouton de Gruyter.

- Collins, Peter. 1991b. The modals of obligation and necessity: Some Australian findings. In K. Aijmer and B. Altenberg (eds.), *English corpus linguistics: Studies in honour of Jan Svartvik*, 145–165. London: Longman.
- Collins, Peter. 1992. Cleft existentials in English. *Language Sciences* 14: 419–433.
- Collins, Peter. 1994a. Extraposition in English. *Functions of Language* 1: 1–18.
- Collins, Peter. 1994b. The structure of English comparative clauses. *English Studies* 75: 157–165.
- Collins, Peter. 1995. The indirect object construction in English: An informational approach. *Linguistics* 33: 35–49.
- Collins, Peter. 1996. *Get*-passives in English. *World Englishes* 15: 43–56.
- Collins, Peter. 2004. *Let*-imperatives in English. *International Journal of Corpus Linguistics* 9: 299–320.
- Collins, Peter. 2005a. Exclamative clauses in English. *Word* 56: 1–18.
- Collins, Peter. 2005b. The modals and quasi-modals of obligation and necessity in Australian English and other Englishes. *English World Wide* 26: 249–273.
- Collins, Peter. 2007. *Can/could* and *may/might* in British, American and Australian English: A corpus-based account. *World Englishes* 26: 471–488.
- Collins, Peter. 2008. The progressive in World Englishes: A corpus-based study. *Australian Journal of Linguistics* 28: 225–249.
- Collins, Peter. 2009a. *Modals and quasi-modals in English*. Amsterdam: Rodopi.
- Collins, Peter. 2009b. Modals and quasi-modals in world Englishes. *World Englishes* 28: 281–292.
- Collins, Peter. 2010. Information-packaging constructions in some new Englishes. In S. Slembrouck, M. van Herreweghe and M. Taverniers (eds.), *Studies in honour of Anne-Marie Simon-Vandenberghe*, 105–116. Ghent: Academia Press.
- Collins, Peter. 2011. Singular agreement in *there*-existentials: An intervarectal corpus-based study. *English World-Wide* 33(1): 53–68.
- Collins, Peter. 2014. Quasi-modals and modals in Australian English fiction 1800–1999, with comparisons across British and American English. *Journal of English Language and Linguistics* 42: 7–30.
- Collins, Peter. 2015a. *Grammatical change in English world-wide*. Amsterdam: John Benjamins.

- Collins, Peter. 2015b. Diachronic variation in the grammar of Australian English: Corpus-based explorations. In P. Collins (ed.) (2015a), 15–42.
- Collins, Peter. 2015c. Recent diachronic change in the progressive in Philippine English. In P. Collins (ed.) (2015a), 271–296.
- Collins, Peter. 2016. Grammatical change in the verb phrase in contemporary Philippine English. *Asiatic* 10: 50–67.
- Collins, Peter and Ariane Borlongan. 2017. Has Philippine English attained linguistic independence? The grammatical evidence. *The Philippine ESL Journal* 19: 10–24.
- Collins, Peter, Ariane Borlongan, Joo-Hyuk Lim and Xinyue Yao. 2014. The subjunctive mood in Philippine English: A diachronic perspective. In S. Pfenniger, O. Timofeeva, A. Gardner, A. Honkapohja, M. Hundt and D. Schreier (eds.). *Contact, variation, and change in the history of English*, 259–280. Amsterdam: John Benjamins.
- Collins, Peter, Ariane Borlongan and Xinyue Yao. 2014. Modality in Philippine English: A diachronic study. *Journal of English Linguistics* 42: 68–88.
- Collins, Peter and Pam Peters. 1988. The Australian Corpus Project. In M. Kytö, O. Ihalainen and M. Rissanen (eds.). *Corpus linguistics, hard and soft*, 103–120. Amsterdam: Rodopi.
- Collins, Peter and Pam Peters. 2004. Australian English morphology and syntax. In B. Kortmann (ed.). *Varieties of English*. 2 vols. Berlin: Mouton de Gruyter.
- Collins, Peter and Pam Peters. 2008. Australian English morphology and syntax. In K. Burridge and B. Kortmann (eds.). *Varieties of English 3: The Pacific and Australia*, 341–361. Berlin: Mouton de Gruyter.
- Collins, Peter and Xinyue Yao. 2011a. Modals and quasi-modals in New Englishes. In M. Hundt and U. Güt (eds.). *Mapping unity and diversity in New Englishes*, 35–53. Amsterdam: John Benjamins.
- Collins, Peter and Xinyue Yao. 2011b. Developments in the verbal systems of the “old” and “new” Englishes. In C. Gitsaki and R. Baldauf (eds.). *The future of applied linguistics: Local and global perspectives*, 328–349. Cambridge Scholars Publishing: Newcastle, UK.
- Collins Peter and Xinyue Yao. 2014. Grammatical change in the verb phrase in Australian English: A corpus-based study. *Australian Journal of Linguistics* 34: 496–514.

- Collins Peter and Xinyue Yao. 2018. Colloquialisation and the evolution of Australian English: A cross-varietal and cross-generic study of Australian, British and American English from 1931 to 2006. *English World-Wide* 39: 253–277.
- Collins, Peter, Xinyue Yao and Ariane Borlongan. 2014. Relative clauses in Philippine English: A diachronic perspective. In L. Vandelandotte, K. Davidse, C. Gentens and D. Kimps (eds.). *Corpus interrogation and grammatical patterns*, 123–146. Amsterdam: Rodopi.
- Collins, Peter, Xinyue Yao, Haidee Kruger, Adam Smith and Minna Korhonen. N.d. (Ms). Diachronic register change: A corpus-based study of change in four registers in Australian English from 1931 to 2006, with comparisons across British and American English.
- Francis, W.N. and H. Kučera. 1964. *Manual of information to accompany A Standard Corpus of Present-Day Edited American English, for use with digital computers*. Providence: Rhode Island Department of Linguistics, Brown University. Available at <http://clu.uni.no/icame/manuals/BROWN/INDEX.HTM>.
- Fritz, Clemens. 2006. Resilient or yielding? Features of IrE syntax and aspect in early Australia. In T. Nevalainen, J. Klemola and M. Laitinen (eds.). *Types of variation: Diachronic, dialectal and typological*, 280–301. Amsterdam/Philadelphia: Benjamins.
- Fritz, Clemens. 2007a. From Plato to Aristotle – investigating early Australian English. *Australian Journal of Linguistics* 24: 57–97.
- Fritz, Clemens. 2007b. *From English in Australia to Australian English: 1788–1900*. Frankfurt: Peter Lang.
- Green, Elizabeth and Pam Peters. 1991. The Australian corpus project and Australian English. *ICAME Journal* 15: 37–53.
- Hundt, Marianne and Christian Mair. 1991. ‘Agile’ and ‘uptight’ genres: The corpus-based approach to language in progress. *International Journal of Corpus Linguistics* 4: 221–242.
- Hundt, Marianne. 2012. Towards a corpus of early written New Zealand English – news from Erewhon? *Te Reo* 55: 51–74.
- Johansson, Stig, in collaboration with Helen Goodluck. Nd. *Manual of information to accompany the Lancaster-Oslo/Bergen Corpus of British English, for use with digital computers*. Oslo: Department of English, University of Oslo.

- Jose, Nicholas. 2009. *Macquarie PEN anthology of Australian literature*. Sydney: Angus and Robertson.
- Kruger, Haidee and Adam Smith. 2018. Colloquialisation versus densification in Australian English: A multidimensional analysis of the Australian Diachronic Hansard Corpus (ADHC). *Australian Journal of Linguistics* 38(3): 293–328.
- Leech Geoffrey and Stig Johansson. 2009. The coming of ICAME. *ICAME Journal* 33: 5–20.
- Leech, Geoffrey, Marianne Hundt, Christian Mair and Nicholas Smith. 2009. *Change in contemporary English: A grammatical study*. Cambridge: Cambridge University Press.
- Moore, Bruce. 2008. *Speaking our language: The story of Australian English*. Sydney: Oxford University Press.
- Noël, Dirk, Johan Van der Auwera and Bertus Van Rooy (eds.). 2014. Special issue of *Journal of English Linguistics* (Volume 42, “Diachronic Approaches to Modality in World Englishes”).
- Peters, Pam. 1987. Towards a corpus of Australian English. *ICAME Journal* 11: 1–13. Peters, Pam. 1998. In quest of international English: Mapping the levels of regional divergence. In A. Renouf (ed.). *Explorations in corpus linguistics*, 281–292. Amsterdam: Rodopi.
- Peters, Pam. 1993. Corpus evidence on some points of usage. In J. Aarts, P. de Haan and N. Oostdijk (eds.). *English language corpora: Design, analysis and exploitation*, 247–256. Amsterdam: Rodopi.
- Peters, Pam. 1994. American and British influence in Australian verb morphology. In U. Fries, G. Tottie and P. Schneider (eds.). *Creating and using English language corpora*, 149–158. Amsterdam: Rodopi.
- Peters, Pam. 1996. Comparative insights into comparison. *World Englishes* 15: 57–68.
- Peters, Pam. 1998. The survival of the subjunctive: Evidence of its use in Australia and elsewhere. *English World-Wide* 19: 87–102.
- Peters, Pam. 2001. Corpus evidence on Australian style and usage. In D. Blair and P. Collins (eds.). *English in Australia*, 163–178. Amsterdam: John Benjamins.
- Peters, Pam. 2007. *Cambridge guide to Australian English usage*. Cambridge: Cambridge University Press.



- Peters, Pam. 2008. Patterns of negation: The relationship between NO and NOT in regional varieties of English. In T. Nevalainen, I. Taavitsainen, P. Pahta and M. Korhonen (eds.). *The dynamics of linguistic variation: Corpus evidence on English past and present*, 147–162. Amsterdam, John Benjamins.
- Peters, Pam. 2015. Dual adverbs in Australian English. In P. Collins (ed.) (2015a), 179–204.
- Peters, Pam and Peter Collins. 2012. Colloquial Australian English. In B. Kortmann and K. Lunkenheimer (eds.). *The Mouton world atlas of variation in English*, 585– 595. De Gruyter Mouton.
- Peters, Pam, Peter Collins, and Adam Smith. 2009. *Comparative studies in Australian and New Zealand English: Grammar and beyond*. Amsterdam: John Benjamins.
- Peters, Pam, Peter Collins, David Blair and Alison Brierley. 1988. The Australian Corpus Project: Findings on some functional variants in the Australian press. *Australian Review of Applied Linguistics* 11: 20–30.
- Rodriguez-Louro, Celeste. 2015. The evolution of epistemic marking in West Australian English. In P. Collins (ed.) (2015a), 205–219.
- Schneider, Edgar. 2007. *Post-colonial English: Varieties around the World*. Cambridge: Cambridge University Press.
- Shastri, S.V. in collaboration with C.T. Patilkulkarni and G.S. Shastri. 1986. *Manual of information to accompany the Kolhapur Corpus of Indian English, for use with digital computers*. Kolhapur: Department of English, Shivaji University.
- Yao, Xinyue. 2015. The present perfect and the preterite in Australian English: A diachronic perspective. In P. Collins (ed.) (2015a), 247–268.
- Yao, Xinyue and Peter Collins. 2012. The present perfect in world Englishes. *World Englishes* 31: 386–403.
- Yao, Xinyue and Peter Collins. 2018. Exploring grammatical colloquialisation in non-native English: A case study of Philippine English. *English Language and Linguistics* 22/3:457–482.
- Yao, Xinyue and Peter Collins. Forthcoming. Developments in Australian English grammar from 1931 to 2006: An aggregate approach to dialectal variation and change. *Journal of English Linguistics*.