

Evidence of Australian cultural identity through the analysis of Australian and British corpora

Rachel Muntz

Department of Linguistics, University of Wales, Bangor

1. Abstract

This paper reports on findings of a keywords analysis comparing the ACE corpus of written Australian English (AusE) and the Flob corpus of written British English (BrE). The main aims of the study are to provide evidence of the cultural representativeness of ACE and to identify significant features of lexical frequency associated with written AusE which are distinct from written BrE. This paper shows how the physical environment has helped shape the Australian lexicon. It also demonstrates clear differences in the use of colour and personal reference terms. Although this study has discovered many lexical features which confirm that Australia's identity is very separate from that of Britain's, it is demonstrated that the British influence on Australian English and culture is still pervasive.

2. Background

2.1 Setting the context of this study: Australia's cultural and linguistic identity

On November 6, 1999 the people of Australia voted narrowly against creating an Australian republic. They voted to retain Britain's queen as its head of state, and not to replace her with an Australian citizen. However, the referendum was highly politically charged and the no vote was carried more due to factors such as fears that the proposed constitutional changes might induce political and economic instability, and rather less due to Australians feeling culturally tied to Britain (Stephens, 1999). The referendum was a defining point in several decades of debate on Australia's cultural identity. During the late 1980s and early 1990s (the time period ACE and Flob represent) former prime minister Paul Keating vigorously promoted such debate, encouraging Australians to reflect on the question "What does it mean to be Australian?" (Snow, 1999). Elements of this coming-of-age discourse included the growing awareness of Australia's economic role and geographical position within the Asia-Pacific region; greater acknowledgement of the breadth of the population's ethnicity due to worldwide immigration; and increased recognition of indigenous communities. In discussing who and what it was, Australia was also coming to terms with who and what it *wasn't* anymore: Britain.

Obviously, one of the greatest, enduring influences the British had on Australia was linguistic (Gorlach, 1991). In the 200 years plus since English was transported to Australia, first with the convicts and administration, and then with the free settlers, AusE and BrE have developed in their own unique directions. Of course, the most striking difference now is phonological (Trudgill & Hannah, 1985). An Australian accent, indeed any accent, is a marker not only of linguistic identity, but of cultural identity too. It marks the speaker as a member of a speech *community*. That is, we can tell from this accent that the person shares common ground in the form of knowledge, beliefs and assumptions with the other members of that community - the community here being at the level of the nation (Clark, 1996). But what about subtler linguistic features, at syntactic and lexical levels? Do they also provide clear signals as to the cultural identity of the speaker? This paper seeks to answer this question by analysing a very specific characteristic at the lexical level: frequency of word use in written AusE, as compared to written BrE. Such research is valuable, not only because it adds breadth and depth to the literature on linguistics and corpus linguistics, but also because it can contribute to a young country's search for self-knowledge and sense of identity.

2.2 Previous studies and theories of culture and linguistics

For many years now linguists have been investigating links between culture and language. Any discussion of this subject inevitably refers to Benjamin Lee Whorf and the theories of linguistic relativity and linguistic determinism (Gumperz & Levinson, 1996). Although these theories are focused on the binary factor of the presence versus absence of a word for a particular concept in different languages, here they are applied to information on the frequency of use of a word within different varieties of the same language. The basic premise of linguistic relativity is that our thoughts and points of view are influenced by the language we speak due to the way the language is syntactically and semantically structured. Linguistic determinism is a more radical and less popular theory which claims that the language we speak dictates the way we think about our physical and social world (Clark, 1996). Clark expands the linguistic relativity theory to include not merely major language

communities, but “any cultural community that corresponds to people’s social identities” (1996: 353). So applying the theory of linguistic relativity to a cross-cultural study of two varieties of English which exist in very different parts of the physical world seems a good way to test this theory. Could the English that the convicts and settlers took with them to Australia cope with the demands of a new physical environment? Did it need to change as some words became less useful and others more useful? Undoubtedly it did, and this paper seeks to uncover the nature of these changes.

On the subject of Australian culture, work by Wierzbicka (1992) has provided interesting linguistic insights into its nature by examining Australians’ fondness for using diminutive forms of nouns and proper nouns. Of course, this is a not just an academic observation. The Lonely Planet guide to Australia recommends that “if you want to pass for a native try speaking slightly nasally, shortening any word of more than two syllables and then adding a vowel to the end of it, making anything you can into a diminutive” (1998: 56). Seriously, though, introspection and consultation of Wierzbicka’s work and cultural guidelines found in travel guides provided a starting point for identifying semantic categories which might be present in ACE, such as struggle/adversity and mateship.

However, the true beginning point for this paper was a study by Leech and Fallon entitled *Computer corpora – What do they tell us about culture?* (1992). This paper was the first systematic attempt to use computer corpora as a source of cultural information. They investigated significant differences in word frequencies between American and British English, based on Hofland and Johanssen’s (1982) published frequency tables, and attempted to identify which differences were attributable to cultural difference. A particular strength of their paper was its frank discussion of the limitations of corpus research and clear description of the goals of their study, which excluded analysing what they termed ‘linguistic contrasts’ which included differences in the two language varieties due to spelling conventions or ‘lexical’ differences where the two varieties simply, through convention, use different lexemes to denote the same thing. Examples of such a lexical linguistic difference between Australian and BrE would be the British *crisps* and *lorry* whose Australian counterpart terms are *chips* and *truck*. Whilst interesting to note, and whilst unique to a particular community’s lexicon, these are superficial differences in convention not indications of one community’s need to refer to a something more often than another’s. “Words evolve in a community in direct response to their usefulness and usability in that community” (Clark, 1996: 341). And corpus frequency data can tell us whether the same word is more useable in a certain community and culture. These are Fallon and Leech’s ‘non-linguistic contrasts’, those which cannot be explained by linguistic code or variety. Excluding proper nouns, Leech and Fallon grouped their results into the following 15 domain categories which occurred naturally within their dataset: sport; transport and travel; administration and politics; social hierarchy; military and violence; law and crime; business; mass media; science and technology; education; arts; religion; personal reference; abstract concepts; and ifs, buts and modality. Their generalised conclusion on the cultural differences evidenced in American and British English were that American culture was...

...masculine to the point of machismo, militaristic, dynamic and actuated by high ideals, driven by technology, activity and enterprise – contrasting with one of British culture as more given to temporizing and talking, to benefiting from wealth rather than creating it, and to family and emotional life, less actuated by matters of substance than by considerations of outward status. (Leech & Fallon, 1992: 44-45).

2.3 The present study: cultural differences in AusE and BrE

To a certain extent this paper is a replication of Leech and Fallon’s study, but with the following important differences:

1. This study contrasts different language varieties - British and Australian English
2. This study analyses language from a different time period - the late 1980s/early 1990s
3. The minimum significance level used in this study was $p=0.01$, meaning its results are of a higher statistical significance than those of Leech and Fallon’s study which used $p=0.05$
4. This study has developed its own more suitable domain categories
5. This study includes proper nouns referring to place names, as it was felt they could provide information on a country’s place within the world, which, in turn, contributes to its identity
6. This study includes spelling conventions as they provide evidence of cultural choices and language variation in progress
7. Increased effort has been made in the present study to verify concordances of results, making it less likely that polysemous words and biased dispersion have compromised results
8. This study is more asymmetrical– it focuses on observations about the Australian variety of English as compared with BrE, rather than the other way round.

2.4 About the corpora

The study is a near-synchronic study of contemporary Australian and British Englishes. The instruments used are the Australian Corpus of (written) English (ACE), and the Freiburg Lancaster-Oslo/Bergen (Flob) corpus of British written English. The two corpora are comparable in size, date, composition and parentage (Peters & Smith, online; Hundt et al, online). Both contain around 1 million words from 500 texts (around 2000 words per text). ACE consists of texts published in 1986, whilst Flob's texts come from 1991. Both were modelled on the LOB (Lancaster-Oslo/Bergen) corpus which represented the BrE of 1961. The compilers of both Flob and ACE used random selection methods to select their component texts. These commonalities make the two corpora highly comparable and this contributes to the robustness of results. However, perfect replication of content was not possible due to shortages of Australian publications in certain fiction genres, namely mystery/detective fiction (category L), science fiction (category M), romance (category P) and western and adventure fiction (category N). The shortfall in these areas was compensated for by the inclusion of human relationship and fantasy genres in the romance category, bush fiction in the western and adventure fiction genre and through the addition of two additional sections: historical fiction (S) and women's fiction (W), for which there are no corresponding categories in Flob. It is noted throughout this paper where it is felt these differences may have skewed the study's results. The categories and corresponding number of texts of each corpus are described in the table below.

Table 1: Text composition of ACE and Flob

Category	Number of texts	
	ACE	Flob
A: Press: Reportage	44	44
B: Press: Editorial	27	27
C: Press: Review	17	17
D: Religion	17	17
E: Skills, trades and hobbies	38	38
F: Popular Lore	44	44
G: Belles lettres, biographies, essays	77	77
H: Miscellaneous (e.g. government and industry reports)	30	30
J: Science	80	80
K: General Fiction	29	29
L: Mystery and detective fiction	15	24
M: Science fiction	7	6
N: Adventure and western	8 (including bush fiction)	29
P: Romance and love story	15 (including fantasy and human relationship fiction)	29
R: Humour	15	9
S: Historical fiction	22	-
W: Women's fiction	15	-
Total texts:	500	500

3. Objectives

The main objectives of this study of frequency data are:

1. To assess ACE's representativeness of AusE and Australian culture
2. To identify some key semantic domains and lexemes within these domains which occur more in Australian written English than in British written English
3. To provide evidence that Australia's cultural comparator is still Britain

4. Method

Using the computer software program Wordsmith, a keywords or keyness analysis was performed by comparing the entire frequency wordlists of both the ACE and Flob corpora. A log-likelihood test was applied to the data, using a p-value of ≤ 0.01 . The default on the number of results returned by the program was suspended (minimum frequency = 1; database minimum frequency = 1; associate minimum frequency = 5). This yielded a total number of 5323 results (2486 results for the Flob corpus; 2837 results for the ACE corpus). These were the words which occurred significantly more often in one

corpus than in the other. The fact that more results occurred in the ACE corpus is partly due to the fact that due to different encoding of the two corpora - parts of hyphenated words were counted in the ACE corpus as separate entries. Although this could have been avoided by changing the WordSmith settings at the outset, these results were filtered out manually in the next stage of analysis. This was a mistake in the methodology and may have compromised a few of the results, but in the overall analysis, it is not expected to have a significant impact.

The next stage of the analysis was to break down the long list of results into categories. These categories were at first based on those that Fallon and Leech found useful in their analysis. However, it soon became apparent that a list more appropriate to this data was necessary. Some categories were retained, some were dropped and some new categories were added. The 11 categories focused on in this study are as follows:

1. place names: proper nouns referring to other countries
2. origin/ethnicity adjectives*
3. multiculturalism, prestige forms and borrowings into English*
4. American spelling conventions*
5. colour terms*
6. geography *
7. housing and communities*
8. fauna and flora*
9. weather and clothing *
10. personal references *
11. abstract concepts: challenge and adversity

More categories were created in the analysis, but as the focus of the project narrowed, only these categories were chosen for further analysis. For each of the categories marked with an asterisk, concordances and dispersion patterns were checked. In instances where the word occurred more than 3 times in the same text, only 3 occurrences were attributed. This prevented skewing of results by sifting out data which was not normally distributed. Unfortunately, due to time constraints, it was not possible to check the distribution of the other categories and, as such, caution should be taken in interpreting these results. At this stage of the analysis each entry is checked for instances of polysemy. Any meanings which fall outside the target semantic domain were also discounted. This type of analysis is more rigid (and more time consuming!) than that used by Fallon and Leech and ensures more robust results. By eliminating unsuitable results, the numerical data given by the computer in the first instance on the keyness of each word was altered and recalculated manually using Johanssen and Hofman's frequency tables as a guide to determine the significance of the new result.

5. Results

5.1 Place names: proper nouns referring to other countries

ACE and Flob mention certain countries significantly more often than the other. Each writes about itself the most and then about its near geographical neighbours. ACE mentions Asia-Pacific countries such as *New Zealand*, *Papua New Guinea*, *Timor*, *Malaya* and *Vietnam* and Flob mentions European countries such as *Ireland*, *France*, *Portugal*, and *Germany*. Political events at the times the corpora were compiled have impacted on these results. For example, *Iran*, *Iraq* and *Bosnia* are due to this.

5.2 Origin/ethnicity adjectives

This trend also holds true for adjectives such as *British* and *Australian*. *Aboriginal* people are represented in ACE's results. Also present is the diminutive *Aussie*, evidence that the custom of abbreviating words is so entrenched that it is strong enough to occur in the written language, not just the spoken language. Interestingly, like the proper nouns results, Flob contains many more origin/ethnicity adjectives than ACE. These observations on place names and origin, while fairly straightforward, provide evidence of the representativeness of the two corpora - results are as expected, with each country writing about itself more than all others.

5.3 Multiculturalism, prestige forms and borrowings into English

Other related terms which occurred significantly more often in ACE included *multicultural*, *migrants*, *migration*. These seem relevant to Australia's increasing awareness, in the media at least, of itself as a multicultural country made up of migrants. However, as the media input into ACE is substantial, there is potential overrepresentation of the word *multiculturalism* in particular, as it has been something of a buzzword since the Keating years (late 80s, early 90s). Evidence in ACE of multiculturalism's influence on AusE is negligible. There is no trend of the Italian, Greek or Chinese

languages, which came to Australia with large numbers of speakers, making any inroads into written English. This is in contrast to the British corpus, which shows the influence of the French language on its English. Several French words are significant in the Flob corpus: *le, dans, qui, tout, elle*. But it appears that this tendency to drop in *quelques mots du français* is a habit that Australians do not practice. The reason for their inclusion in BrE is probably due to French words being considered prestige forms in Britain. If this is not the case in Australia, then what are Australia's equivalent prestige forms? It is important to note that French is not an immigrant language in Britain. Following this reasoning, perhaps it can be predicted that Italian, Greek and Chinese are not likely to be sources of prestige borrowing in AusE in the future.

5.4 American spelling conventions

One form of borrowing which Australians seem to be incorporating into their writing is American spelling of words that the British spell with containing 'ou', such as *color, humor, harbor, behavior, neighborhood*. The ACE results contain 11 such lexemes, whilst Flob yielded one American spelling of a different kind: *center*. Whilst the majority of Australians prefer the British spelling, it will be interesting to observe if this is in fact evidence of language variation in progress and whether in 50 years time the old-world 'u' in such words will be obsolete in AusE, as has happened in the new-world English of America.

5.5 Colour terms

One of the most unexpected findings of this study was the discovery of the significance of colour terms in ACE. In descending order of frequency, the following colours occurred significantly more often in ACE: *white, red, blue, green, yellow, brown, pink, orange, beige*. Intriguingly, there were no colours of particular British significance. This list from ACE includes eight of Berlin and Kay's (1991) 11 basic colour groups, plus a supplementary colour – *beige*. The absence of the other three groups (black, purple and grey) means these are used with broadly equal frequency in both Australian written English and British written English. In addition to this, the word *colours* itself was a significant result in ACE. How can these convincing results be explained? Berlin and Kay's work was used to disprove the strong form of the linguistic relativity theory and to support the existence of linguistic universals. But they were only concerned with the existence, not the frequency of terms. Here we have clear evidence of one linguistic community finding use for colour terms far more often than another. From this data, we can say with some certainty that Australians feel the need to classify things in terms of colour to a much greater extent than their British counterparts. One possible explanation is geophysical, due to the difference in the quantity or quality of natural light in Australia. But this is only a theory based on introspection and a similar theory by Van Wijk (cited in Berlin & Kay, 1991) was rejected by Berlin and Kay. Further analysis of the concordances of these colour terms would reveal the reasons behind this result.

5.6 Geography

This section centres around physical and geographical features. Obviously, there are great differences between Australia and Britain in this respect and this is borne out in the data. However, the results show that Australians write about these features much more, evidence that such things are more central to the Australian psyche, e.g. *land, landscape, rocks, hill(s), mountains, soil, parks, ground*. This compares to Flob's few entries such as *meadow, woodlands, moor*, terms, which, in any case, border on being linguistic rather than non-linguistic contrasts. Here it must be noted that the differences in the component texts of the two corpora may have had an effect on the results. The inclusion of bush fiction in ACE must be taken into account, but equally, Flob contains western fiction which ACE does not. Words pertaining to the coast were present in roughly equal numbers in both corpora, though ACE's results were related to the *beach* and *surf*, whilst Flob's included items of a maritime theme: *buoy, moorings*. In a largely arid country such as Australia, the importance of sources of water is also evident. *Creek, drainage, swamp, waterhole, lagoon* and *dam* are all present in the ACE results, whilst Flob yields only *reservoir*. An interesting avenue to investigate would be to see if there is any link between ACE's plethora of geographical terms to the colour terms which were so prevalent.

5.7 Housing and communities

We are also able to see how the two countries arrange their territory. Whereas Australia has *blocks* of land, with the ubiquitous *shed* in the *backyard*, in Britain there exist *estates* (both of the housing estate and country estate variety), *manors, castles* and *mews*. People live in *villages* in Britain, whereas in Australia there are *towns* and remote *townships*. However, in both countries, most people live in *urban* areas and this is a big theme with *cities, suburb, suburbs* and *suburban* all present in

ACE. Difference in housing design is also evident, with Australian dwellings tending to have a *veranda (h)*.

5.8 Fauna and flora

In a similar vein to the geography category, large categories for flora and fauna appeared in ACE, but not in Flob. Surprisingly, there was only one entry in the Flob column (*dolphin*) against 59 in ACE. The ACE results consisted mainly of native animals and plants, such as *dingo, kangaroo, gum, wattle*, which of course would not be expected to be found in Flob. But Britain has its own native animals and plants which are not found in Australia, e.g. *squirrel*, yet these are absent from the results. Both countries have many kinds of animals in common, such as *foxes, rats, the bird and cat*. But these are more significant to AusE and there are no corresponding entries for Flob. Generic terms such as *tree, leaf, plants, ferns, grass* are all significant in ACE, with no counterparts in Flob. Another interesting result is *croc* which, as an abbreviated form, shows the common Australian practice of shortening words. As with the term *Aussie*, the fact that *croc* appears in the written form means it is firmly entrenched in the Australian lexicon.

5.9 Weather and clothing

Not surprisingly, these domains show a trend of rain and cold weather words in Flob e.g. *precipitation, freezing*, against Australia's heat and high winds, e.g. *sun, hot, cyclone, wind*. Corresponding clothing for each kind of weather was also noted: *shorts and bikini* against *suits, boot, and cloak*.

5.10 Personal reference

This category yielded some interesting results. A preference for third person singular personal pronouns and possessives is evident in Flob with *he, him, himself, his, she, her, herself, it, itself* all more significant. In ACE, first person singular and first and third person plural pronouns are present: *I, my, myself, us, our* and *they*. A complementary result was found for personal reference nouns. It seems where BrE prefers the third person pronoun, Australians choose from a number of nouns: *men, bloke(s), boys, father(s), husband*, or the idiomatic *mate(s)*. In the same category, Flob yields just *gentlemen* and *sir*. Nouns referring to women in ACE include *woman, women, women's, girls, mum, daughters*. *Feminists* and *wimmin* are certainly due to the presence of women's fiction texts in ACE and it is also a possibility that the results referring to men may be influenced by this, as well as those for children which include *kid(s), teenagers, and toddlers*. Reflecting the phonological changes in AusE, Australians have begun to write *meself* and *ya*, mainly in reported speech.

5.11 Abstract concepts: challenge and adversity

One small, uniquely Australian category emerged from the data - that of struggle and adversity. If you come from the wrong side of the tracks, you "grew up in Struggletown". The figure of the "Aussie battler" still looms large in the national psyche. Even if you come from the right side of the tracks, "life's not meant to be easy". *Struggles, tough, battler, conflict, sweat, feat, overcoming* are results from ACE which reflect this.

6. Discussion and conclusion

6.1 Representativeness of ACE

This study has found that ACE is highly representative of Australian written English, as anticipated results, such as certain proper nouns, were present. The fact that the text make-up of ACE is different to that of Flob needed to be borne in mind, but was not found to have a great influence on results overall.

6.2 Significant semantic domains

Overall then, what lexical frequency evidence have these corpora provided us with which prove Australia's cultural identity is significantly different from Britain's? Results from 11 significant semantic domains were identified and discussed. Because of the focus of this paper, most of these domains deal with lexemes which are more significant to AusE than to BrE. Several of the domains are related to geography or the physical environment (categories, 1, 2, 6, 7, 8 & 9) and it seems that members of the Australian speech community have more use for such terms than the members of the British speech community. This shows how the language, as part of the people, has adapted to a new physical environment. Evidence of multiculturalism was found in ACE, although for a country which talks up this aspect of its identity, there are not the corresponding number of origin/ethnicity adjectives (category 2) which would support this image. Category 1 confirmed that Australia is very well aware of

its location in the Asia Pacific region. Certain BrE prestige forms were shown not to have any AusE equivalents – potential verification of what is known as the ‘tall poppy syndrome’ in Australia, where anyone acting above their station is quickly mown down. This notion goes hand in hand with the ethos of struggle and adversity. The best finds of this study, though, are those relating to the colour and personal reference terms. Clear trends were found in the different cultural uses for these terms and further study of concordances would reveal the reasons behind these trends.

6.3 Cultural comparator

Despite all this linguistic evidence of Australia’s independent identity, this study provides evidence that Britain may still be Australia’s everyday cultural comparator. The results closely reflect some of the most commonly observed facts and stereotypes about Australia and its culture: the size of its coastline, barren interior, native animals and hot weather. Whilst this assures us of the validity of the corpora, it also tells us that Australians have inherited a way of thinking about such things from Britain. Australia does have a massive coastline, but only in comparison with Britain, not in comparison with the United States. It is a dry country in comparison with Britain, but not compared with Middle Eastern countries. Australians and Britons think of it as a hot country, but the Aborigines probably didn’t think in such terms, etc. If the ACE corpus had been compared with its American counterpart, Frown, for instance, then presumably, another set of significant words would have been yielded. But I suspect that these results would not have matched so closely with observations by social commentators about characteristics of Australian culture as the Flob results, partly because Australia has shared physical similarities (size of country, melting pot population, indigenous inhabitants). So although AusE has many of its own lexical frequency characteristics, at a deeper level – in the way Australians think – perhaps they are not so different from the British. In Whorfian terms then, perhaps the fact that Australians speak a type of English dictates the way Australians think about their physical environment, if not the frequency with which they use those words to describe it.

7. References

- Berlin B, Kay P 1991 *Basic color terms: their universality and evolution*. Berkeley, University of California Press.
- Clark HH 1996 Communities, commonalities and communication. In Gumperz JJ, Levinson SC (eds), *Rethinking linguistic relativity*. Cambridge, Cambridge University Press: 324-355.
- Gorlach M 1991 *Englishes: studies in varieties of English 1984 -1988*. Amsterdam, John Benjamins.
- Gumperz JJ, Levinson SC (eds) 1996 *Rethinking linguistic relativity*. Cambridge, Cambridge University Press.
- Hofland K, Johansson S 1982 *Word frequencies in British and American English*. Harlow, Longman.
- Hundt M, Sand A, Siemund R *Manual of information to accompany the Freiburg - LOB corpus of British English ('FLOB')*. Accessed 25/1/01 at:
<http://khnt.hit.uib.no/icame/manuals/flob/INDEX.HTM>
- Lonely Planet 1998 *Australia*. Melbourne, Lonely Planet Publications.
- Leech G, Fallon R 1992 Computer corpora – what do they tell us about culture? *ICAME Journal* 16: 29-50.
- Peters P, Smith A *Manual of information to accompany the Australian corpus of English (ACE)*. Accessed 25/1/01 at: <http://www.hd.uib.no/icame/ace/aceman.htm>
- Snow, D 13 November 1999 When no means yes. *Sydney Morning Herald* Accessed 20/01/01 online at: <http://www.smh.com.au/news/review/9911/13/review4.html>
- Stephens, T 13 November 1999 The great divide. *Sydney Morning Herald* Accessed 20/01/01 online at: <http://www.smh.com.au/news/review/9911/13/review6.html>
- Trudgill P, Hannah J 1985 *International English: a guide to varieties of standard English* (second edition) London, Edward Arnold.
- Wierzbicka A 1992 *Semantics, culture, and cognition: universal human concepts in culture-specific configuration*. Oxford, Oxford University Press.