

Multicultural Australia: Accounting for Our Diversity in the Australian National Corpus

Loy Lising
Macquarie University

1. Introduction

Various authors on corpus linguistics emphasize the significance of representativeness (Atkins, Clear, & Ostler, 1992; Biber, 1990; Biber, 1993). Representativeness takes into consideration both the speakers of the language and the various domains where language is used. The idea of representativeness in relation to the target population and representativeness of the data within that population are discussed in Atkins, Clear & Ostler (1992), Biber (1993) and Summers (1993). However, the Australian multicultural situation makes the notion slightly more complicated and necessitates clarification of a number of issues including the scope of the national corpus, and the language or languages represented in the corpus. Since a corpus is aimed at collecting and making available a body of material that can be analyzed to characterize a particular language, representativeness suggests a fit between the target population or users of a language and the sample data that make up the corpus that is meant to represent these very users of the language. If the corpus is to be used as the basis for making descriptions of a particular language (Biber, 1990) then it is imperative that the users of the language who utilize and inevitably effect/initiate change to the form of that language need to be clearly identified. Biber (1990) posits that the most fundamental consideration in corpus building is not necessarily sample size but a “thorough definition of the target population” (p. 1).

Representativeness is also used to refer to the various registers, text categories and ultimately the sample data included in the corpus and how they can fully and sufficiently provide a comprehensive sampling of the communicative experiences of the target population. Below, these conceptions of representativeness - that of target population and of text types - will be explored from two angles. First, options for representing the linguistically and culturally diverse group of migrants in Australia within the Australian National Corpus (AusNC) will be examined. Second the various communicative events and the range of registers for which data need to be collected have to be considered.

In particular, this paper will argue for a representation in the AusNC of Australians who were born overseas and whose dominant language may not necessarily be English. Moreover, it will also suggest that it includes Australians whose dominant language may be English but come from a migrant family. It will then outline the Migrant Language Corpus (MLC) within the AusNC.

2. Migrants in Australia

2.1. *Representativeness: Target Population*

It is widely acknowledged that Australia is a multicultural and multilingual society. Although from 1788 to 1996 British immigrants were the largest, single national group of migrants, the trend slowly changed especially after the Whitlam government officially ended the White Australia Policy in the early 70s. The arrival of around 170,000 Displaced Persons from Eastern and Southern Europe in the mid 70s laid the foundation for a multicultural Australia (Jupp, 2002). Since then, Australia has seen some significant growth in its migrant population not just from Europe but also from Asia and the Middle East. Currently, statistics indicate that around a quarter of the total population was born overseas (ABS, 2008).

As the AusNC plans for the building of the national corpus, it is important that it considers the breadth of its representation and also takes into account how various migrant groups can feature in the

corpus. The AusNC can now take the unique position of having in the national corpus samples of English varieties in Australia. More importantly, AusNC needs to address the issue of language: what language or languages will be represented in the national corpus? This needs to be clarified since a number of migrant groups come from various linguistic experiences. A large number of migrants come from countries where English is an official second language or a lingua franca such as in India, the Philippines, Malaysia and Sri Lanka, and educated users of English from these countries often consider themselves as “functionally” native in English (Kachru, 1996). There are also those who come from monolingual societies where English is not used at all and have moved into Australia and immediately found a strong and vibrant ethnic and linguistic community which may have allowed them to operate mostly in their first language and have found no need to learn and use English extensively. For the corpus to be representative of the range of languages migrants use in their new migrant community, the framework for the AusNC needs to capture the various domains in which these languages are used.

Based on the *Report on Migration Program* for 2007-2008 (DIAC, 2008), the top ten migrant groups in Australia are: UK, China, India, Philippines, South Africa, Malaysia, Sri Lanka, USA, Korea and Indonesia. This is excluding New Zealand as it is not counted as part of the Migration Program because of its special relation with Australia. Altogether, migrants to Australia come from 100 source countries. Including all 100 countries in the discussion here would be too long and as the intention is to give an indication of migrants’ employment participation, qualification and languages used at home, doing so using the top ten migrant countries should be sufficient to begin with.

Although participation of people in society varies and is dependent upon various factors including but not limited to employment, social network and education, an indication of where these top ten migrant groups are in relation to employment, education and languages at home will be indicative of some of the location of their participation and the languages used. It is acknowledged that some members of the migrant community are not in the workforce for various reasons even if they are highly educated. It is also equally important to point out that some migrants with low educational qualification or less workforce participation can be just as engaged in their participation in society. Information used in the subsequent discussion is derived from *Community Information Summaries* (DIAC, 2008).

2.1.1. Employment

The Hawke government in the mid 80s became more concerned with the economic viability of migrants (Jupp, 2002) and put an emphasis on a higher intake of migrants under the *Skills* category. This resulted in the steady rise of skilled migrants. Of the top ten source countries of migrants, there is quite a high employment participation rate compared to 68.6% of the total Australian population. Of the UK-born migrants there is 58.8% labour force participation, of the China-born 56.3%, India-born has 72.3%, Philippine-born 73.1%, South Africa-born 75.1%, Malaysia-born 67.3%, Sri Lanka-born 70.9%, USA-born 70.3%, Korea-born 51.8% and Indonesia-born 62.1%. While employment alone is not a sole indicator of migrants’ participation in society, it is nonetheless revealing of the fact that they are highly involved in the local workforce and this can suggest possible participations through their interaction with colleagues at work and the various local clients their companies may engage with. This supports the view that it is crucial to capture their conversations in the workforce as part of the data in the AusNC.

2.1.2. Educational qualification

The educational qualifications of the overseas born migrants is also one factor that is interesting to examine as it can certainly provide an indication of their level of literacy be it in English or their first language (L1). This level of educational qualification may also be relative to their employment participation. Such information allows for a consideration of the kinds of text types whether in English or L1 which the Migrant Language Corpus within the AusNC can consider. Of the UK-born migrants, 55% had some form of higher non-school qualification, 26.8% had a diploma or higher qualification and 20.3% had certificate level qualification. The other two English speaking source countries have the following figures: South Africa-born with 68.1%, 47.5% and 13.8%, respectively and the USA-born

with 70%, 55.8% and 8.4%. The figures are not very different in comparison to the Asian group which are: China-born (55%, 42.2% and 4.8%), India-born (76.1%, 60.2% and 7.3%), Philippine-born (64.9%, 42.8% and 1.9%), Malaysia-born (66.8%, 54.3% and 6.6%), Sri Lanka-born (64.8%, 45.3% and 10.2%), Korea-born (56.1%, 41.7% and 5%) and Indonesia-born (62.7%, 48.1% and 6.9%). These figures indicate a relatively high level of differentiated qualifications across the top ten migrant groups. These various levels of educational qualification are insightful as the AusNC considers the various text types it can include in its collection. Although these qualifications do not certainly indicate specific genres migrants engage with, they can be indicative of possible text types that can be included in the corpus.

2.1.3. *Languages spoken at home*

The other information derived from the report which is insightful to examine in considering a Migrant Language Corpus in the AusNC is the migrants' reported language/s at home. They are as follows: English for UK, USA and South Africa with a low percentage of Afrikaans spoken at home for the South Africa-born. With the Asian migrants, there are some who speak English at home and some who speak their L1. The China-born migrants reported Mandarin as the main home language with 64.7% of the total also claiming to speak English well and 34.6% not able to speak English well. The India-born reported English with 30% of the total speaking Hindi or Punjabi. Of the Philippine-born, there was an equal spread of Filipino, Tagalog and English reported spoken at home with 96% of the Filipino or Tagalog speakers reporting speaking English well. The Malaysia-born reported three languages spoken at home: English, Cantonese and Mandarin with 94% of those who speak non-English reported speaking English well. The Sri Lanka-born reported 35% using English at home with the rest using Sinhalese and Tamil with 94% speaking English well. The South Korea-born reported Korean as the main home language with 62% also speaking English well. Finally, the Indonesia-born reported Indonesian as the main home language with 90% claiming they speak English well. Except for native English countries, the majority of these figures indicate high instances of bilingualism among the rest of the top ten migrant groups. These reports indicate the richness and complexity of languages used by migrants in various domains. Capturing these in a migrant language corpus in the AusNC will make for very rich data.

Migrants' participation in society is naturally not just limited to their employment participation nor is it dependent upon their educational qualification. However, having an understanding of their active involvement in the Australian workforce provides an indication that they may also be highly engaged in other domains in society. Given all the other possible participation they have in other areas in the community, how is the AusNC going to capture their interactions which can possibly be a rich data source for reflecting language contact, and possibly language change, not just on the migrants' languages but also on Australian English? The challenge will be to represent these instances of interactions in the sample.

2.2. *Representativeness: Target Text Types*

Another facet of the issue of representativeness that needs to be clarified is the question of the sample data from the target population that represent their communicative activities in English and/or other languages. The various communicative events can suggest the different registers that need to be represented in the corpus can be further explored. The choice of text types can then be planned and predetermined once the target population is clarified. As Biber (1993) pointed out "representativeness refers to the extent to which a sample includes the full range of variability in a population" (p. 1). After accounting for some of the members of the multicultural part of the Australian community, appropriate sampling needs to take place to reflect the uses of English and other languages within these groups. How is their variability going to be accounted for in the corpus?

It may also be beneficial for the AusNC to survey pre-existing corpora that have already begun the process of corpus building among overseas-born Australians. How is the AusNC going to incorporate these existing corpora so as to capitalise on the rich data that already exists? Identifying what the possible registers are where overseas-born Australians participate and use English or their local

languages will be an enriching part of the AusNC. As Biber (1990) pointed out once you have identified who is going to be part of your target population, you can “identify an adequate sampling frame... stratified samples are almost always more representative than non-stratified samples (and they are never less representative)” (p. 2).

As language is constantly evolving it is impossible to fully represent all communicative events that happen within the lives of the target population. To be clear and consistent with the data collection, however, feasible boundaries and parameters for determining which registers and genres should be incorporated in the corpus needs to be identified and defined.

3. Migrant Language Corpus

The Migrant Language Corpus (MLC) is envisaged to be a sub-corpus within the AusNC that will include text types produced by Australian migrants. The main objective of this sub-corpus is to have a representation of migrants’ use of English or other languages in their participation in various domains in society. This allows the AusNC to be consistent with and representative of the fact that Australia is a multicultural and multilingual society. Having the MLC within the AusNC provides data for some of the research agendas outlined for the corpus. Peters (2008) identifies these various research agendas for building a national corpus.

NLP uses, e.g. data mining; humanistic/sociological/historical inquiry; linguistic variation: social, regional, diachronic; discourse analysis: nonfiction and fiction, written v. spoken usage, online v. print documentation; lexical/lexicographic profiling and phraseology; grammatical research: morphology and syntax terminology/terminography; phonetic and prosodic research and gestural research, including sign language (p. 4).

The use of sample data which includes text produced by overseas-born Australians would allow for analyses that would inform linguistic variation and discourse analysis as outlined above. Moreover, it would also speak to issues of language and identity: how migrants retain their language in general or way of speaking/acquiring English in particular or adapt to local forms to portray their transnational and globalized identities and express their multicultural selves. One of the most urgent linguistic needs that Australia has as a multicultural nation is to be able to build the relevant resources to build on information about the various users of English in our society and the rich diverse contexts in which they use English and other languages.

The AusNC has the potential to place itself in a unique position to provide rich data that address various research issues relative to overseas-born Australians and Englishes in the context of migration. The following considerations can be made in thinking of MLC within the AusNC.

The target population for sampling should include both first and second generation migrants to provide as much range in the data as possible.

Given that one of the prerequisites for entry into Australia for most migrants is English proficiency, it can be expected that most overseas-born migrants, particularly those who came under the *Skills* category program, will come with English that differs in terms of lexical, phonological, grammatical and pragmatic features in varying degrees from standard Australian English. Moreover, if they come from countries where English is used as the official second language and/or the lingua franca like in the case of Malaysia, India and the Philippines, then it is likely that they will come with a very strong sense of ownership of their variety of English (Kachru, 1996). There is significant potential for the national corpus to feature data on English varieties in the context of migration. Given their strong sense of ownership of the kind of English which, for most of them, is representative of their identity, how will the corpus capture change over time in the kind of English they speak and the impact of Australian English on their variety and vice versa?

During their initial stage of settlement, newly arrived migrants to Australia engage to varying degrees in various domains in society. Multiple factors such as familiarity with the host country due to previous visits, presence of extended families, and other such factors can help facilitate immediate mobility in society.

While it is ideal to have a collection of both written and spoken texts like those in similar corpora such as the International Corpus of English (ICE), newly arrived migrants may not necessarily have immediate opportunities to engage in all such text types as those outlined in the ICE (Bautista, Lising, & Dayag, 2004). For the written text types, for example, they may only have limited opportunities to engage in non-printed correspondence text types like social and business letters which can include letters to home and job application letters that migrants are most likely to write a lot of as they look for work. Non-printed academic writing text types may also be possible especially for those who seek to upskill themselves if employment does not present itself easily. It may also be possible to collect academic writing and reportage text types but only if these can be provided by a representative number of migrants, otherwise, it raises the question of representativeness. Therefore, the large part of the migrant language corpus would be of spoken text types.

The possible text types in the migrant language will largely include dialogic data in English such as interviews, social interactions, transactional interactions, and further study and workplace interactions.

3.1. Interviews

Periodic individual interviews can be arranged throughout the course of data collection to capture language samples of newly arrived migrants and to document their views of themselves, their own variety of English and their reflections on settlement. The interviews will not only capture insightful views they hold of their language and their migration but also will provide a record of the kind of English they bring to Australia.

3.2. Periodic Family (Home) Interactions

It can be assumed that much of the interactions of most migrants especially during their initial years of settlement as they endeavour to build their social network will be naturally at home with their family. It will be interesting and insightful to capture language/s in the home to get a sense of whether the first language is retained or English is adopted for use. Moreover, if parents come with their own brand of English and the children acquire the new variety, how do the different varieties interact? For consistency, families can be asked to nominate any family time where discussion among family members is richest.

3.3. Social Interactions

New migrants to Australia, especially those without extended families, may not necessarily have immediate opportunities for social interactions unless there is an established local ethnic community that they can readily join upon arrival and/or if they belong to some religious, social or sporting organizations. Social interactions that take place, be it with extended families, ethnic community members or local community members, can be part of the social interaction data. How do migrants interact and in what language in their new community? Do they communicate in English or their first language?

3.4. Transactional Interactions

A large part of many migrants' initial settlement processes include familiarization with the local way of life in general and with social services such as Centrelink and the Roads and Traffic Authority. These kinds of interactions can provide insights into their (un) successful settlement as they negotiate with various agencies. The AusNC's ability to capture these interactions will provide evidence not just of the migrants' language but also of communication strategies they employ and issues that they regularly confront. This information could also easily form part of an orientation package for new migrants.

These transactional interactions can also include further study interactions which some migrants might engage in if they opt to upskill or do further study to have their current qualifications recognized.

3.5. Workplace Interactions

As discussed above, there is quite a high employment participation rate among skilled migrants. Incorporating in the corpus recordings of their various interactions at work would greatly enhance the corpus. Since a predominant motivation for migrants is economic and social mobility The AusNC can place itself in a significant position to provide the necessary data that would allow investigation into various aspects of intercultural communication at work.

There are many aspects to communicative practices at work that can be examined in greater depth if the corpus is to assist in creating a workplace environment that facilitates understanding of intercultural communication (Clyne, 1994). The corpus can provide data that reveals communicative instances that may be challenging for both migrants and locals. It can also allow researchers to explore sites of difficulty in communication where varied cultural practices and assumptions can impede successful communicative experiences. How will workplace negotiations between employees from diverse cultural background feature in the corpus?

Issues on intelligibility are also in the forefront in discussions on intercultural communication (Munro & Derwing, 2002). Local Australians are sometimes unfamiliar with the variety of accents of migrants from diverse linguistic background, and conversely, migrants, in some instances, may have difficulty with understanding the Australian accent. These and other issues are significant causes of miscommunication and possibly disempowerment (on both sides) and can be a very revealing site for investigation.

4. Summary & Implications

Australia's ethnically and linguistically diverse migrant community plays an active role in and is a vital part of the greater Australian society and needs to be considered as planning for the infrastructure of the AusNC takes place.

Once data which include various migrant groups are incorporated in the AusNC, the issue of language in the context of migration can be added to the research agenda outlined by Peters (2008). Much has been written about different regional varieties of English as they exist in the endogenous context. Very few studies, however, delve into the forms, usage, function and acceptability as they are used in the context of migration. How do functionally native speakers of English from countries where it is used as a lingua franca use their English or their first languages and negotiate in and/or adapt to the local English variety as they begin to settle in Australia?

An inclusion of the MLC provides significant potential for the AusNC to be strategic and unique in its scope and purpose.

References

- Atkins, Sue, Jeremy Clear, & Nicholas Ostler. (1992). Corpus design criteria. *Literary and Linguistic Computing*, 7(1), 1-16.
- Australian Bureau of Statistics (2008). <http://www.abs.gov.au/ausstats/abs@.NSF/7d12b0f6763c78caca257061001cc588/F1C38FAE9E5F2B82CA2573D200110333?opendocument>
- Bautista, Maria Lourdes S., Loy V. Lising, & Danilo T. Dayag. (2004). *ICE-Philippines Lexical Corpus* – CD-ROM. London: International Corpus of English.
- Biber, Douglas. (1990). Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing*, 5(4), 257-269.
- Biber, Douglas. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243-257.

- Clyne, Michael. (1994). *Inter-cultural communication at work: cultural values in discourse*. Cambridge: Cambridge University Press.
- Community Information Summaries*. Available at <http://www.immi.gov.au/media/publications/statistics/comm-summ/textversion/>
- Department of Immigration and Citizenship (DIAC) (2008). *Report on migration program for 2007-2008*. Available at <http://www.immi.gov.au/media/statistics/pdf/report-on-migration-program-2007-08.pdf>
- Fact Sheet 4 – More than 60 years of post-war migration*. Available at <http://www.immi.gov.au/media/fact-sheets/04fifty.htm>
- Jupp, James. (2002). *From White Australia to Woomera: The Story of Australian Immigration*. Cambridge: Cambridge University Press.
- Kachru, Braj. (1992). Teaching World Englishes. In Braj Kachru (Ed.), *The other tongue: English across cultures* (2nd ed.). Urbana: University of Illinois Press.
- Kachru, Braj. (1996). English as an Asian Language. In Bautista, Maria Lourdes (Ed.), *English is an Asian language: The Philippine context*. NSW: The Macquarie Library Pty. Ltd.
- Kachru, Yamuna. (2008). Language variation and corpus linguistics. *World Englishes*, 27(1), 1-8.
- Peters, Pam. (2008). Architecture for a multipurpose Australian National Corpus. Keynote paper presented at *Designing the Australian National Corpus Workshop*, 4-5 December, UNSW, Australia.
- Summers, Della. (1993). Longman/Lancaster English Language Corpus - Criteria and design. *International Journal of Lexicography* 6(3), 181-208.
- Munro, Murray J., & Tracey M. Derwing. (2002). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning*, 49(1), 285-310.

Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus: Mustering Languages

edited by Michael Haugh, Kate Burridge,
Jean Mulder, and Pam Peters

Cascadilla Proceedings Project Somerville, MA 2009

Copyright information

Selected Proceedings of the 2008 HCSNet Workshop on
Designing the Australian National Corpus: Mustering Languages
© 2009 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 978-1-57473-435-5 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Lising, Loy. 2009. Multicultural Australia: Accounting for Our Diversity in the Australian National Corpus. In *Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus*, ed. Michael Haugh et al., 37-43. Somerville, MA: Cascadilla Proceedings Project.

or:

Lising, Loy. 2009. Multicultural Australia: Accounting for Our Diversity in the Australian National Corpus. In *Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus*, ed. Michael Haugh et al., 37-43. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2286.