# Why Migrant Languages in the Australian National Corpus?

**Michael Clyne**
**Monash University and University of Melbourne**

## 1. Introduction

For a good many years, Australia enjoyed an outstanding reputation internationally for its pluralistic language policies. They were largely informed by research and by advocacy resulting from research, including research into the actual languages spoken in Australia. More recently there has been a decline in commitment not only by governments to multiculturalism but also by university departments to teaching and research on immigrant multilingualism and language contact despite indications of continuing interest in the wider community in multilingualism and among students in learning and researching about it. This decline is evident from the failure of universities to appoint or replace experts in the field and from the abolition of undergraduate courses in the field. The former has led to the unavailability of adequate supervision in some places and/or a discouragement of graduate research in the field. Yet there is both anecdotal and quantitative evidence of student demand where a course is or was available, whether at a university or at an Australian Linguistic Institute. The University of Western Sydney, for example, has over 200 enrolments for an introduction bilingualism subject. Australia still has an unfinished agenda in research into community languages and the pooling of data is one way of continuing achievements in the field.

In this paper I will begin with a brief overview of Australia's language demography, focusing on migrant (community) languages. I will then discuss some of the corpora available and how they were collected and utilized in research. Comparable corpora of German, Dutch, Croatian, Vietnamese, Mandarin, German, German/Dutch, German/Hungarian, and Italian/Spanish in contact with English from bi- and trilinguals in different generations and vintages will be used to demonstrate what can be gleaned from the corpora to show the typological and sociolinguistic factors in contact-induced language change. It will be argued that as migration to Australia comes from an ever increasing number of source countries, there will be more opportunities for the expansion of the corpus and further research with the promise of more insights into linguistics, sociolinguistics, psycholinguistics and language and languages-in-education policies. Past experience shows that there is an intersection of data between migrant languages and ethnolects of Australian English, as I shall discuss. It will also be shown that if migrant languages data is not incorporated in the Australian National Corpus, it is likely to be lost forever. The only other way to ensure its safety is for it to be held overseas and this would mean segmenting it according to the language.

## 2. Demography

According to the 2006 Census, 16.8% of the Australian population then used a language other than English at home. It was 31.7% in Sydney and 27.9% in Melbourne but much lower outside the metropolitan areas (4.9% in NSW outside Sydney and 4.7% in Victoria other than Melbourne). These are underestimates as they do not include young people who have moved out of the parental home and use the community language in the homes of parents or other relatives and in community groups, or elderly people living on their own who respond "no" to the question "Does this person speak a language other than English in the home?" Over the past 15 years, most European languages in Australia have undergone a considerable decline in number of home users while languages of Asian, Middle Eastern, or Latin American origin have experienced substantial increases. Table 1 shows the top 20 languages, nationwide, in 2006, with comparative statistics from 1991 and 2001.

 Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus, ed. Michael Haugh et al., 44-54. Somerville, MA: Cascadilla Proceedings Project.

Table 1.
*Top 20 LOTES Spoken at Home in Australia in 2006*

| | Top20 LOTEs in 2006 | Speakers in 1991 | Speakers in 2001 | Speakers in 2006 | %Change since 1991 | % Change since 2001 |
|---|---|---|---|---|---|---|
| **1** | Italian | 418801 | 353605 | 316893 | -24.3 | -10.4 |
| **2** | Greek | 285702 | 263717 | 252222 | -11.7 | -4.4 |
| **3** | Cantonese | 163266 | 225307 | 244554 | +49.8 | +8.5 |
| **4** | Arabic | 162855 | 209372 | 243662 | +49.6 | +16.4 |
| **5** | Mandarin | 54430 | 139288 | 220596 | +305.3 | +58.4 |
| **6** | Vietnamese | 110185 | 174236 | 194858 | +76.8 | +11.8 |
| **7** | Spanish | 90477 | 93593 | 97998 | +8.3 | +4.7 |
| **8** | Tagalog /Filipino | 59109 | 78878 | 92330 | +56.2 | +17.1 |
| **9** | German | 113335 | 76443 | 75634 | -33.3 | -1.1 |
| **10** | Hindi | 22727 | 47817 | 70013 | +208.1 | +46.4 |
| **11** | Macedonian | 64428 | 71994 | 67831 | +5.3 | -5.8 |
| **12** | Croatian | 63081 | 69851 | 63615 | +0.8 | -8.9 |
| **13** | Korean | 19756 | 39529 | 54619 | +176.5 | +38.2 |
| **14** | Turkish | 41966 | 50693 | 53858 | +28.3 | +6.2 |
| **15** | Polish | 66933 | 59056 | 53390 | -20.2 | -9.6 |
| **16** | Serbian | 24336 | 49203 | 52534 | +115.9 | +6.8 |
| **17** | French | 45496 | 39643 | 43219 | -5.0 | +9.0 |
| **18** | Indonesian | 29803 | 38724 | 42038 | +41.1 | +8.6 |
| **19** | Persian | | 25238 | 37155 | | +47.2 |
| **20** | Maltese | 52997 | 41393 | 36517 | -31.1 | -11.8 |

Current rates of increase suggest that the number of languages with more than 100,000 home users is likely to rise by three – Hindi, Filipino, and Spanish – in the 2011 Census. By then, Mandarin will have overtaken the other community languages and Arabic will probably be in second place. In the meantime, a steady development of African languages in Australia is taking place. (Clyne, Hajek, & Kipp, 2008)

Sociolinguistic research based on macrodata from the Census and/or smaller scale studies (e.g., Garner, 1988; Clyne & Kipp, 1999, 2006; Clyne & Fernandez, 2005; Fernandez & Clyne, 2007; Martin, 1996) have shown differential rates and patterns of language shift to English as the only home language. A Dutch-born Australian's chance of speaking only English at home is about 26 times as great as that of a Vietnamese-born Australian. Apart from Vietnamese, the languages best maintained are those from the eastern Mediterranean (Macedonian, Turkish, Arabic, Greek) and those resulting from recent migration from Asia and Africa (such as Dinka, Somali, Tigrinya, Mandarin, Khmer). Those least maintained are languages of the postwar assimilation era such as Dutch, German, Maltese, French, Hungarian, Lithuanian, and Latvian. Although these differentials are conditioned by many factors related to pre- and postwar experiences, including the sociolinguistic status of the language in the homeland and the attitudes to other languages in the migration country (Clyne, 2006, 2007), such variables as cultural distance and the importance of language – or rather bilingualism? – as a cultural core value are especially salient. In the second generation, the shift rate is consistently higher and the continuum of shift differential similar to that of the first generation except that the intergenerational shift is overproportionate for Mandarin and Cantonese. As recent censuses do not collect data on parents' country of birth, the most recent second generation estimates are from 1996 and the high shift for Chinese languages may have decreased due to a larger population and smaller proportion of exogamous marriages. Differences in language-use patterns influence the rate of language change but not necessarily transference or switching patterns.

## 3. Availability of Corpora

Bettoni (1991) lists 23 community languages on which studies have been carried out in Australian universities: Italian, German, Dutch, Polish, Russian, Japanese, Yiddish, Greek, Vietnamese, Maltese, Swedish, 'Chinese', French, Latvian, Macedonian, Serbo-Croatian, Czech, Estonian, Hungarian, Punjabi, Scots Gaelic, Sorbian, and Spanish. The last seven languages at the time had been the subject of only one study each.

Since then, the number of languages investigated has increased by at least nine: Arabic, Danish, Filipino, Finnish, Hebrew, Korean, Somali, Tamil, and Turkish. There have been new and different studies of some of the other languages, including tri- and multilinguals as well as bilinguals. Many studies are the result of doctoral theses, largely from Monash or Melbourne. There are also some studies (e.g., Kovács, 2001, on Finnish and Hungarian) by visiting scholars. The majority of the studies accounted for are based on actual speech samples. While some university ethics committees these days require researchers to preserve data for 5 years and then destroy it, the corpora are and remain of great significance for as long as they are in working order.

The rest of this paper centres around some Monash corpora through which I will give some examples of what has been possible to glean, both on individual languages and comparatively. This is the data that I am proposing should initially be part of the Australian National Corpus (ANC) and can be added to later from other studies of language contact between English and community languages.

All the corpora under consideration have a number of common features facilitating comparison. Apart from some of the German settlement tapes, they were recorded in Victoria, principally in Melbourne. They all include conversations between one researcher (or a team of researchers) and a respondent. The topics (in the community language [CL]) include a typical day, first impressions of Australia (or growing up bilingually in Australia), and the story of a book read or a film seen. In order to compare lexical items and grammatical features, respondents were also asked to talk about some pictures – a beach scene (in the CL), and a city scene and a scene from the country of origin (in English). For trilingual speech, different community language material was sought in each of the languages. Some corpora also include other data including conversations in the family taped by the participant(s).

The original corpus of 200 (first and second generation) German-English postwar bilinguals recorded in 1962-64 is unfortunately largely no longer extant (though most transcriptions are). However, a similar corpus from 50 prewar German-speaking refugees (recorded in 1968) is fairly intact. It should be stressed that the data has been in numerous locations in four buildings at two universities, often under sub-optimal conditions. As will be evident from Table 2, some recordings and transcriptions of about 200 interviews with second, third, and fourth generation German-English bilinguals in former German enclaves in the Victorian Western District, the Wimmera, and parts of South Australia have been lost (for details of the settlements, which date back to the 19th century, see Clyne, 1968, 1972; Kipp, 2003). The same applies to the Dutch-English bilinguals recorded in Melbourne, the Dandenong Ranges, and the Latrobe Valley in 1970-71. It should be noted that the recordings not on CD are still extant on reel-to-reel tape, so that one priority would be to have them recopied to equipment that is readily available today. We have described this study as 'longitudinal' because in 1987, Kees de Bot (then Nijmegen, now Groningen) spent his study leave at Monash and recorded new interviews with 42 of the informants. In addition to the regular interview schedule, he asked the respondents about their language-use patterns and views on possible attrition or improvement in both languages. As those selected for the 1987 interviews were those whose English and Dutch were strong and clearly differentiated in 1970-71, another five who had experienced dense switching and much convergence in the earlier interview were subsequently (1991) recorded again by a research assistant. Of those rerecorded in the 1980s or 1990s, 11 who could be contacted and were still able to be interviewed were recorded again in 2005. This study can claim to be one of the few and probably the most extensive longitudinal study of bilinguals. It is able to test the validity of some of the recent assumptions about second-language attrition and first-language reversion (De Bot & Clyne, 1994; see also Section 4.13).

Of the two groups of trilinguals – Dutch-German-English and Hungarian-German-English – mentioned, the former will be used to explore whether the respondents behave like Dutch-English

bilinguals in Dutch and like German-English bilinguals in German. The Hungarian-German-English trilinguals enable us to establish whether the effects of English on German are different if the third language is more distant typologically from English than German. This group also facilitates comparisons in the integration of English lexical transfers into an isolating and an agglutinative language.

Table 2.
*The Corpora*

**1. German**
**a. Western District of Victoria:**
19 CDs, 53 participants, transcriptions of 47

**b. Wimmera:**
16 CDs, 74 participants, transcriptions of 71
*No CD recording found for 7 participants transcribed.

**c. South Australia:**
15 CDs, 75 participants, transcriptions of 42 available
Barossa Valley: 53 participants, transcriptions of 29
Hahndorf and district: 12 participants, all transcribed
Lobethal and district: 3 CDs, 10 participants, 1 transcription found

**d. Pre-war:**
28 CDs, covering 49 of 55 participants, transcriptions of 51

**2. Dutch Longitudinal**
**1971:** 22 CDs covering 76 of 200 participants. Remaining recordings on 15 tape reels.
Transcripts on CD cover 175 participants.
**1986-1991/92:** 12 cassettes, 47 participants
Transcripts on CD cover 29 participants.
**2005:** 1 CD, 11 participants, all transcribed.

**3. Hungarian-German-English trilinguals**
26 CDs, 36 participants, transcriptions of 30

**4.Dutch-German-English trilinguals**
38 CDs, 38 participants, transcriptions of 36

In addition, there are corpora of particular subgroups, such as 20 third generation German-English bilinguals in Melbourne (which show how some grammatical features are lost when the third generation takes its grammar from the second; Clyne, 1997). Parallel to the other groups of trilinguals, there is a similar corpus from Italian-Spanish-English trilinguals. From two of my PhD graduates, Jim Hlavac and Tuc Ho-Dac, I obtained part of their corpora on second generation Croatian and first and second generation Vietnamese in Melbourne, respectively, with an analysis of certain issues of switching, which I wanted to compare between languages. Both of the students had published revised versions of their theses (Hlavac, 2003; Tuc Ho-Dac, 2003). From another PhD graduate, Lin Zheng, I received some relevant fragmentary data on second generation and young first generation Mandarin speakers in junior secondary school (see article, Zheng, 1997).

## 4. What we have Gleaned from these Data[1]

Among other things, comparative studies of the corpora have facilitated the following research findings. The more radical change in definite article, plural allomorph, and, in some cases even personal endings is due to Dutch having less choice than German. The radical change in syntax can be attributed to less morphological marking in Dutch than in German. In sum, findings suggest, in relation to the discussion on Dutch being the language between German and English (Van Haeringen, n.d.; Hüning et al., 2006), that Dutch is really typologically somewhat closer to English (see Sections 4.1 to 4.5 and 4.7).

### *4.1. Gender Assignment*

While German in Australia integrates lexical transfers from English by assigning the gender of the definite article of a German synonym (e.g., die high school <die Schule) or German homophone (e.g., der roof < der Ruf, 'reputation'), Dutch tends towards a uniform non-neuter definite article *de*. This is particularly evident in lexical transfers from English even where the semantic equivalent is neuter (e.g., de fence, cf Dutch equivalent het hek), but also in Dutch words that have some homophony with English equivalents (e.g., de boek, Dut. het boek 'book';de café, Dut. het café, 'pub'). This convention even affects other Dutch words but usually only in individual cases (e.g., de paard, Dut. het paard, 'horse'; de gezin, Dut. het gezin, 'family'). This effect can be attributed to *de* being the most prevalent article and also resembling English *the*. Dutch, like other languages in contact, also assigns gender to nouns according to the gender of the nearest semantic equivalent (de library < de bibliotheek).

### *4.2. Plurals*

While the most common plural allomorph chosen for English nouns transferred into German is –s (the marker of non-native origin), ffor example, beaches, groceries, referees, some informants employ zero (Plumber), -en (Beachen), -n (Refereen). Dutch in Australia strongly tends towards a uniform plural form –s. This tendency can be attributed to -s being the most common plural allomorph in Dutch, compared to the variety in German (e.g., -e, -en, -s).

The –s allomorph is particularly evident in lexical transfers from English in Dutch (e.g., shops, groceries, tools) but also in Dutch nouns that bear some homophony with English equivalents (e.g. zondags 'Sundays', Dut. zondagen; koes 'cows', Dut. koeien). It also occurs in non-homophone Dutch words but usually only in individual cases (e.g., hoofdleidings, Dut. hoofdleidingen 'mains', lantaarnpaals, Dut. lantaarnpalen 'lamp-posts').

### *4.3. Word Order*

Dutch speakers are far more likely to generalize the fixed subject-verb-object ( SVO ) word order from English than are German speakers and do so in the first generation, whereas this occurs in the second generation German speakers if at all. SVO contrasts with V2 word order (including fronting) in main clauses in both Dutch and German, and with verb final in subordinate clauses. An example of SVO generalization in Dutch is:

(1) Maar als wij praten in het Hollands, zij verstaat drommels goed.
But if we talk Dutch, she understands damned well.
Dut: Maar als wij in het Hollands praten, versataat zij drommels goed.

[1] For more details, see Clyne (2003).

### *4.4. Inflections*

There is a slight tendency towards analytic (bare, uninflected) personal forms of verbs among second generation Dutch-English bilinguals (e.g., kan, moet, kijk in the first and third person plural; homeland Dutch kunnen, moeten, kijken). This tendency does not have a parallel in any of the other corpora from European languages and can be explained by the smaller range of personal endings in Dutch in most European languages other than English.

### *4.5. Morphosyntactic Change*

The more rapid syntactic change among Dutch-English bilinguals may be attributed to less rich morphology in Dutch than in German, where case endings indicate the relation between parts of the sentence and there is no need for word order to do this. *Such changes are accelerated in contact situations* and it is only recently that research on colloquial speech in the Netherlands and Flanders has started to pick up the beginnings of a similar tendency. However, Dutch-German-English trilinguals are more conservative in their syntactic change in Dutch than are bilinguals and more radical in German. This tendency shows the effect of trilingualism on all the languages.

### *4.6. Grammatical Integration*

The comparison between the integration of English lexical transfers in the German and Hungarian of Hungarian-German-English trilinguals also shows up a typological feature – the morphology of Hungarian requires a full integration into the recipient language. According to one's view of grammar, Hungarian has between 12 and 27 cases, many expressed through postpositions, for example:

szép swimmingpool*al* 'a nice swimming-pool, egy opshop*ba* 'into an op-shop', Acland Street*re* 'to Acland Street' as opposed to the uninflected: swimming pool, shop, Flinders Street in German.

### *4.7. Convergence*

Because many lexical items are very similar in Dutch and English, and also because of the adult Dutch speakers' phonological transference from Dutch into English, there is a considerable amount of convergence. This means that it is very difficult to identify which items are Dutch and which are English, for example:

*(2*) Ik hebt *een kop of tea* or something.
'I have a cup of tea or something'
where *kop of* are bilingual homophones in the speech of the respondent's parents, who are the source of his Dutch, and

*(3) Dat's* one of *de nieuwer* plaatsen *in Holland.*
'That's one of the newer places in Holland'
where only *one of* and *plaatsen* are clearly in English and Dutch respectively and the other items could be either language.

### *4.8. Triggering*

In all the corpora, overlapping items facilitate (trigger) switching between languages. In combinations of languages in bi- or trilinguals other than Dutch-English bilinguals and Dutch-German-English trilinguals,[2] the items that are most likely to facilitate (or trigger) an interlingual switch (Clyne, 2003) are lexical transfers from English used in one or more other languages and proper nouns common to the languages, for example:

[2] German-English, Croatian-English, Vietnamese-English, Italian-Spanish-English, Hungarian-German-English

(4) Das ist ein Foto gemacht an der *beach* could be kann be kann sein in Mount Martha
That's a photo taken on the beach Could be (Eng.) can (Ger.) be (Eng.) can be (Ger.) in Mount Martha.

Lexical facilitation of switching by *beach* (which the participant uses in both languages) is promoted by syntactic convergence between the two languages. Languages requiring a high level of integration of lexical transfers have less potential for switching facilitation. Hungarian-German-English trilinguals switch as the result of a trigger-word a total of 46 times from German to English and 6 times from Hungarian to English.

Trigger-words do have strong switching facilitation function in Dutch-English bilinguals, for example:

(5) Ik heb gelezen '*Snow White, come home'*. It's about a winter-pet.

The switching is facilitated by the title of the book.

However, among Dutch-English bilinguals alone, the most common source of lexical facilitation is bilingual homophones, where the homophony is the result of convergence by the informants (see comparisons, Clyne, 2003, Table 5.5, p. 170). For example:

(6) I don't know *what/wat* ze doen.

Items such as *wat, is, was*, and *de/the* are identical in the speech of many Dutch-English bilinguals.

It was partly our data that led Muysken (2000) to characterize his third type of code-mixing/switching ('congruent lexicalization'), which refers to interlingual sharing of grammatical structures through overlap or convergence and lexical items originating from two or more languages. The triggering theory (Clyne, 1967, 1980, 2003) has been discussed and refined in recent articles by overseas scholars (Broersma & de Bot, 2006; Broersma, Isurin, Bultena, & de Bot, 2009; De Bot, Broersma, & Isurin 2009; Broersma, in press) and was one of the foci of a conference in late 2007 at Ohio State University.

### *4.9. Typological Variation*

Each additional combination of languages from which we have data adds to our knowledge of contact phenomena and plurilingual processes. Thus, while the Dutch-English data give us better insights into the effects of convergence, Vietnamese-English and Mandarin-English data indicated that tonal factors can facilitate switching. Ho-Dac (1996, 2003) found that 85% of switches occurred where the Vietnamese items immediately before the switch were in a mid to high pitch tone, these being the tones that Vietnamese speakers are most likely to equate with English pitch and stress – unstressed syllables with mid tones and stressed syllables with high tones. In Mandarin-English language contact, it is falling and neutral tones that facilitate switching, with 97% of switches in Zheng's corpus following such a tone, corresponding to English pitch and stress (Zheng, 1997).

### *4.10. Trilinguals*

I have already briefly referred to trilinguals and the way in which all their languages interact to produce tendencies that are unlike those of bilinguals. This tendency also applies to the facilitation of switching. A trigger word may be common to two or three languages, for example:

(7) Dan ga ik naar de *shops* einkaufen
Then go I to the (Dut) shops (Eng) to shop (Ger).

In Example 7, *shops*, although English in origin, is used by the participant in all three languages – English, German, and Dutch – and facilitates a switch from Dutch to German.

In addition, if a feature is shared by two of the languages of a trilingual, they tend to transfer that feature to the third languages as well, for example, in Hungarian-German-English trilinguals:

(8) Hungarian szép napos nap van
Nice sunny day is

The construction in Example 8 is due to both German and English collocating *sunny* and *day*. In Homeland Hungarian, *nap* means both 'sun' and 'day'. It would be customary to say:

Szép napos idö van
Nice sunny weather is.

On the other hand, some trilinguals will avoid lexical items or structures that acceptable in two or more languages but which they themselves perceive to be too close, for example:

(9) Voordat ik studeerde…ben gaan studeren.
Before I studied… took up studying.

The inchoative construction *ben gaan studeren* is peculiar to Dutch among the three languages.

### *4.11. Universal Constraints*

Data from the German-English and Dutch-English bilingual corpora were among those which demonstrated some weaknesses with constraints (syntactic integrity, free morpheme, and especially government constraints) (Clyne, 1987). Comparative data from all the corpora mentioned (Croatian-English, Dutch-English, Dutch-German-English, German-English, Hungarian-German-English, Italian-Spanish-English) drew attention to problems in Myers-Scotton's (1993) attribution of Embedded Language Island to lack of congruity. On the other hand, Hlavac (2003) worked largely within Myers-Scotton's model and Hlavac's work is frequently cited by Myers-Scotton as supporting her model. The comparative analysis also made possible a reappraisal of current psycholinguistic models in relation to language contact behaviour (Clyne, 2003).

### *4.12. Modal Particles and Discourse Markers*

A comparison of the use of L1 modal particles by bi- and trilinguals with some combination of English, German, Dutch, and Hungarian shows that consensus imposing particles are the ones most likely to be dropped first because they do not comply with the 'mainstream Australian way of behaving'. On the other hand, those discourse markers which reflect an adaptation of mainstream Australian behaviour (such as *well, you know, sort of, anyway*) are most likely to be transferred into the other languages.

### *4.13. L2 Attrition and L1 Reversion*

The Dutch longitudinal study facilitated a critical assessment of the myth that as (healthy) people get older, they will lose skills in their L2 (in this case English) and revert to their L1. The study shows a considerable stability in the Dutch, English, and switching patterns of the informants over time, with changes in either language (positive or negative, perceived or based on linguistic analysis) largely determined by lifestyle changes. Another important factor is the threshold of Dutch and English of the speaker in middle age (e.g., in the early 1970s) (De Bot & Clyne, 1989, 1994). In her study of the role of bilingualism in Dutch old people's villages around Melbourne, Seebus (2008) shows that older Dutch migrants need both their languages to express their identity and share their life's experiences.

Many of Seebus's informants fear second-language attrition as a lack of English would cut them off from their children and grandchildren. The Dutch study demonstrates the usefulness of rerecording informants over a number of decades. A National Corpus could facilitate this.

## 5. Ethnolects

Ethnolects (ethnic varieties) of Australian English used for in-group communication within a family or ethnic community are a link between our community language corpora and the Monash Corpus of Australian English. Often ethnolects are used where language shift is in progress or complete and the identity functions of a community language are transferred to the ethnolect of Australian English. This is so particularly where there is ethno-religious geographical concentration and the group has its own schools (Clyne, Eisikovits, & Tollfree, 2003).

We have in the Monash Corpus of Australian English recordings of interviews with three generations of descendants of German settlers in former German enclaves in the Victorian Western District as well as self-taped conversations between three generations of the extended family and with same-age friends. The recordings show a gradual disappearance of ethnolectal features, whether lexical, syntactic, or phonological. The German settlement recordings give us insights into the ethnolect of the same localities (sometimes within the same family) one and two generations earlier and the German used by those informants which could be the basis of the ethnolectal features. The data from family and friends of Melbourne schoolchildren in the Monash Corpus of Australian English provide us with information on in-group communication based on Chinese, Greek, Yiddish, and other substrata.

## 6. Data on Migration and Settlements

On the tapes from old German settlements we have stories from local history and folklore and from family histories. On most of the other recordings there are accounts of migration history and impressions of Australia at different times. These may be of interest to colleagues in history, sociology, anthropology, and other fields. I would argue that more data should be collected collaboratively across disciplines. Such a process would take the pressure off migrants and their descendants, who are among the most overused informants.

## 7. Where can Data be Found?

The corpora listed in Table 2 are held in the Language and Society Centre, School of Languages, Monash University (Clayton). I would suggest using them as a starting-point. This would involve the completion of copying to CDs the Dutch-English corpus and any other data extant on other types of recording and retranscribing those transcriptions that have been lost. Then it would be useful to try to obtain additional corpora for other languages and groups, and from other parts of Australia from colleagues and students, where they are still available. The additional Croatian and Vietnamese samples are part of the corpora collected by Jim Hajek and Tuc Ho-Dac. Most PhD graduates retain their corpora. It would be good to contact them as soon as the facilities for digitalization and/or storage are available. Colleagues and students embarking on data collection could be encouraged to contribute corpora and so could winners of the immigrant bilingualism/language contact prize awarded by ALS and ALAA.

It has been noted that some of the data (recordings and transcriptions) have been lost. This has occurred with other corpora too. In the late 1960s and 1970s a colleague at a university in another state and his graduate students gathered and analyzed data in old German settlements in that state. Like the former German enclaves whose data is included in Table 2, the German settlements now contain virtually no German-English bilinguals. When the colleague died, the data were dumped in a storeroom and eventually disappeared. The inclusion of community languages in the Australian National Corpus will be an important ecological exercise. I will refrain from telling any stories about our corpora, their journeys, and their possible fate.

## 8. Concluding Remarks

The above discussion has shown the value of comparative studies of bi- and trilinguals whose overarching contact language is English in an Australian context – their value for typological insights into language change and for sociolinguistic, linguistic, and psychological understandings of bi- and trilingualism. Different languages, migration vintages, and generations can be compared if corpora are available based on a similar methodology. Such corpora can also be utilized by other disciplines. In this way more information will be available to harness our language resources in education, business, and many other fields and Australia's outstanding record on typological studies based on indigenous languages could be extended to ones on community languages.

## References


Bettoni, Camilla. (1991). Other community languages in Australia. In M. Clyne (Ed.), *Linguistics in Australia* (pp. 75-90). Canberra: Academy of Social Sciences in Australia.

Broersma, Miriam. (In press). Triggered codeswitching between cognate languages. *Bilingualism: Language and Cognition*.

Broersma, Miriam, & Kees de Bot. (2006). Triggered codeswitching: A corpus-based evaluation of the original triggering hypothesis and a new alternative. *Bilingualism: Language and Cognition, 9,* 1-13.

Broersma, Miriam, Ludmila Isurin, Sybrema Bultena, & Kees de Bot. (2009). Triggered code-switching: Evidence from Dutch-English and Russian-English bilinguals. In Ludmila Isurin, Don Winford, & Kees de Bot (Eds.), *Multidisciplinary approaches to code switching*. Amsterdam: Benjamins.

Clyne, Michael. (1967). *Transference and triggering*. The Hague: Nijhoff.

Clyne, Michael. (1968). Deutscher Idiolekt und deutscher Dialekt in einer zweisprachigen Siedlung in Australien. *Wirkendes Wort, 18,* 84-95.

Clyne, Michael (1972) *Perspectives on language contact.* Melbourne: Hawthorn Press.

Clyne, Michael. (1973). Decay, preservation and renewal: Notes on some Southern Australian German settlements. *AUMLA, 29,* 33-43.

Clyne, Michael. (1980). Triggering and language processing. *Canadian Journal of Psychology, 43,* 400-406.

Clyne, Michael (1987). Constraints on code-switching – how universal are they? *Linguistics* 25, 739-64,

Clyne, Michael, (1997). The speech of third generation German-English bilinguals in Australia. In Wolfgang Wölck & Aniek de Houwer (Eds.), *Recent studies in contact linguistics* (pp. 36-43). Bonn: Dümmler.

Clyne, Michael. (2003). *Dynamics of language contact*. Cambridge: Cambridge University Press.

Clyne, Michael. (2005). *Australia's language potential.* Sydney: University of New South Wales Press.

Clyne, Michael. (2006). Some exploratory comments relating sociolinguistic typology to language shift. In Janet Fuller & Linda Thornburg (Eds.), *Studies in contact linguistics* (pp. 207-230). Frankfurt: Lang.

Clyne, Michael. (2007). Sociolinguistic continuity from old to new homeland: Factors in language maintenance and shift seen from the Australian situation. In Jeroen Darquennes (Ed.), *Contact linguistics and language minorities* (= Plurilingua 30) (pp. 91-102). Hamburg: Asgard.

Clyne, Michael, Edina Eisikovits, & Laura Tollfree. (2003). Ethnolects as in-group markers. In Anna Duszak (Ed.), *Us and others* (pp. 133-157). Amsterdam: Benjamins.

Clyne, Michael, & Sue Fernandez. (2005). Period of residence as a factor in language maintenance: Hungarian-English bilinguals in Australia as a case study. *ITL International Journal of Applied Linguistics, 149/150,* 1-20.

Clyne, Michael, John Hajek, & Sandra Kipp. (2008). A tale of two multilingual cities in a multicultural continent. *People and Place, 16* (3), 1-8.

Clyne, Michael, & Sandra Kipp. (1999). *Pluricentric languages in an immigrant context: Spanish, Arabic and Chinese*. Berlin: Mouton de Gruyter.

Clyne, Michael, & Sandra Kipp. (2006). *Tiles in a multilingual mosaic: Macedonian, Somali and Filipino in Melbourne.* Canberra: Pacific Linguistics.

De Bot, Kees, & Michael Clyne. (1989). Language attrition revisited. *Studies in Second Language Acquisition, 9,* 167-177.

De Bot, Kees, & Michael Clyne. (1994). A 16-year longitudinal study of language attrition in Dutch migrants in Australia. *Journal of Multilingual and Multicultural Development, 15,* 17-28.

De Bot, Kees, Miriam Broersma, & Ludmila Isurin. (2009). Sources of triggering in code-switching. In Ludmila Isurin, Don Winford, & Kees De Bot (Eds.), *Multidisciplinary approaches to code switching*. Amsterdam: Benjamins.

Fernandez, Sue, & Michael Clyne. (2007). Tamil in Melbourne. *Journal of Multilingual and Multicultural Development, 28,* 169-187.

Garner, Mark. (1988). Ethnic languages in two small communities: Swedish and Russian. *International Journal of the Sociology of Language, 72,* 37-50.

Hlavac, Jim. (2003). *Second generation speech.* Frankfurt: Lang.

Hüning, M., et al. (Eds.). (2006). *Nederlands tussen Duits en Engels: Handelingen van de workshop op 30 september en 1 october 2005*. Leiden: SNL.

Kipp, Sandra. (2003). A language community from a historical perspective: Homogeneity and variation. *International Journal of the Sociology of Language, 180,*

Kovács, Magdolna. (2001). *Code-switching and language shift in Australian Finnish in comparison with Australian Hungarian*. Abo: Abo Akademi University Press.

Martín, Mario. (1996). Spanish language maintenance and shift in Australia. Unpublished PhD thesis, Australian National University, Canberra.

Muysken, Pieter. (2006). *Bilingual speech*. Cambridge: Cambridge University Press.

Myers-Scotton, Carol. (1993). *Duelling languages*. Oxford: Oxford University Press.

Seebus, Ingrid. (2008). "Dinkum Dutch Aussies": Language and identity among elderly Dutch-Australians. Unpublished PhD thesis, Monash University, Clayton.

Tuc Ho-Dac (1996). Languages in Contact: Vietnamese-English code-switching in Melbourne. Unpublished PhD thesis. Monash University.

Tuc Ho-Dac (2003). *Vietnamese-English bilingualism. Patterns of code-switching.* Lodon: Routledge-Curzon.

Van Haeringen, C.B. (n.d.). *Nederlands tussen Duits en Engels.* The Hague: Servire.

Zheng, Lin. (1997). Tonal aspects of code-switching. *Monash University Linguistics Papers, 1* (1), 53-63.

# 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



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



## 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:

Clyne, Michael. 2009. Why Migrant Languages in the Australian National Corpus? In *Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus*, ed. Michael Haugh et al., 44-54. Somerville, MA: Cascadilla Proceedings Project.

or:

Clyne, Michael. 2009. Why Migrant Languages in the Australian National Corpus? In *Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus*, ed. Michael Haugh et al., 44-54. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2287.