

Workshop on community language corpora in Australia

9-10 November 2023

Engma Room (3.165 HC Coombs Building, ANU)

Organisers: Li Nguyen & Catherine Travis

Thursday 9 November	Session 1	
11.20	Li Nguyen & Catherine Travis	Intro and Welcome
11.30	Invited speaker: Adrian Vickers	Creating a Multi-lingual On-line Archive
12.10	Simon Musgrave	Government documents as a parallel multilingual corpus: An opportunity which may be disappearing
12.35	Van Tran, Sharynne McLeod, Sarah Verdon, Kate Margetson	Vietnamese language use and maintenance in Australia
1.00	Lunch	
	Session 2	
2.00	Antonia Rubino	Exploring the trilingualism of Italian migrants in Australia: Evolving data, methodologies and insights
2.25	Li Nguyen	CanVEC: The Canberra Vietnamese English corpus
2.50	Paola Escudero, Gloria Pino Escobar, Milena Hernandez Gallego, Chloé Diskin-Holdaway, John Hajek	The Little Multilingual Minds corpus: Educators' and children's speech in heritage languages
3.15	Afternoon Tea	
	Session 3	
3.45	Invited speaker: John Hajek	The many corpora of Michael Clyne, his students and colleagues
4.25	Eleni Elefterias-Kostakidis	The Greek-Australian Oral Archive Project
4.50	Jaime Hunt and Sacha Davis	'Oh, wunderbar, ein bisschen Deutsch!' German-speakers in Newcastle and the Hunter Valley through the lenses of oral history and linguistics
5.15	Close Day 1	
Friday 10 November	Session 4	
9.20	Li Nguyen & Catherine Travis	Welcome Day 2
9.30	Invited speaker: Ingrid Piller	Life in a new language: Data-sharing and re-use across multiple sociolinguistic ethnographies
10.10	James Walker, Stavroula Nikoloudis	Corpora of Greek in Melbourne
10.35	Ksenia Gnevsheva, Catherine Travis	A spoken corpus of Australian English by Russian and Chinese speakers
11.00	Chloé Diskin-Holdaway	The Superdiversity corpus
11.25	Morning tea	
	Session 5: Panel discussion	
11.55	Panel discussion	Francesco Ricatti, Adrian Vickers, John Hajek, Ingrid Piller
12.45	Lunch	
2.00	Close	

Workshop on community language corpora in Australia

9-10 November 2023

Engma Room,

Organisers: Li Nguyen & Catherine Travis

Australia is a highly multilingual and multicultural society, with more than 490 languages coming from around 300 ancestries and cultural traditions (ABS, 2021, 2022). For decades, the language and cultural maintenance of various immigrant groups have been under investigation by many scholars, not only in linguistics but also in history, sociology, anthropology, and many other disciplines. This work has amassed a large body of data reflecting the languages of these communities, providing information about how Australia's immigration history has contributed to the country today.

The purpose of this workshop is to bring together scholars working with language corpora from across different disciplines. The workshop is being run as part of the [Language Data Commons of Australia \(LDaCA\)](#), which is working to build national research infrastructure for the Humanities and Social Sciences, facilitating sustainable access to and controlled use of digital language corpora for linguists, scholars across the Humanities and Social Sciences, and non-academics.

The workshop will consist of presentations on language data collected from Australian immigrant communities for different research purposes, and will close with a panel discussion on needs and challenges around managing and archiving community language data in a way that is ethical, legal and culturally sensitive, and how LDaCA can help support that.

Invited speakers

Life in a new language:

Data-sharing and re-use across multiple sociolinguistic ethnographies

Ingrid Piller (Macquarie University)

This presentation introduces a data-sharing and re-use project involving six separate sociolinguistic ethnographies, which were brought together to examine the language learning and settlement experiences of 130 migrants to Australia from 34 different countries in Africa, Asia, Europe, and Latin America over a period of 20 years. In addition to a brief overview of project findings related to participants' lived experience of navigating the dual challenge of language learning and settlement, I will focus on epistemological issues related to researching "migrant languages" in Australia.

Creating a Multi-Lingual On-line Archive

Adrian Vickers (The University of Sydney)

As part of the Opening the Multi-Lingual Archive of Australia project, our research team has drawn on public and private archives and libraries to create an on-line resource that presents primary sources on Australian history in languages other than English (but not including Indigenous languages). I will discuss the issues of selection and translation in the creation of the OMAA archive <https://omaa-arts.sydney.edu.au/> and the technical solutions that we have come up with. This discussion will also include an outline of further developments relating to potential integration with other Digital Humanities projects, such as LDaCA.

The many corpora of Michael Clyne, his students and colleagues

John Hajek (The University of Melbourne)

Michael Clyne (1939-2010) was undoubtedly Australia's greatest sociolinguist during his lifetime. He was a pioneer in many different areas of linguistic research, making invaluable contributions over many decades to our understanding of bilingualism, multilingualism, language contact, code-switching, intercultural communication, language maintenance and shift as well as the presence and fate of migrant languages in Australia. He was also a pioneer in the recording of non-indigenous languages on this continent, and is responsible - both directly and indirectly - for the creation of a number of different language corpora. Many of these were recorded by his graduate students (a number of whom went on to become his colleagues) as well as by other researchers working with Michael on a range of projects. His interests (and their interests) extended from individual migrant languages to their influence on spoken English, as well as to the linguistic effects of trilingualism (e.g. Dutch-German-English, Hungarian-German-English and Spanish-Italian-English). In the last part of his life, he was particularly preoccupied by the long-term preservation of his own corpus of German speakers recorded in a range of Australian rural settlements – a matter which at least in theory - was ultimately resolved through permanent archiving in Germany.

In this contribution, I give an overview of Michael's many corpora, their content, examples of their use in research (e.g. Clyne et al 2015), as well as their current fate. I also provide some initial information about the many other corpora collected by his former students and colleagues, their content and purposes – as best as we know about them for the moment. Together they provide a remarkable panorama of language diversity and multilingual speech in Australia, and in Melbourne in particular.

References

- Hajek and Slaughter (2014) *Challenging the monolingual mindset*. Clevedon: Multilingual Matters.
- Clyne, M. (2009) Why migrant languages in the Australian National Corpus? In Michael Haugh et al. (eds), *Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus*. Somerville, MA: Cascadia Proceedings Project, pp.44-54.
- Clyne, M., Slaughter, Y., Hajek, J. and D. Schüpbach. (2015). On the relation between linguistic and social factors in migrant language contact. In R De Busser and R. J. LaPolla (eds). *Language, structure and environment. Social, cultural and natural factors*. Amsterdam: John Benjamins, pp.149-176.

Presentations

The Superdiversity corpus

Chloé Diskin-Holdaway (The University of Melbourne)

The 'Superdiversity' project aims to investigate the acquisition of Australian English (AusE) by recent migrants to Melbourne. The main corpus, created in 2017, consists of data from three groups of adult speakers: 12 native AusE speakers (7F, 5M), 10 Irish migrants (5F, 5M) and 14 Chinese migrants (5F, 5M). Gender was self-reported. The Chinese participants came from Mainland China and Hong Kong. All of them reported "Chinese", "standard Chinese" or "Mandarin" as their first language. The Irish came from across the island of Ireland. The years of birth ranged from 1976 to 1995. The average length of residence (LoR) for the Irish migrants was 5;8 ($SD=3;9$) and for the Chinese was 2;5 ($SD=1;10$). The AusE and Irish group consisted mostly of professionals; and the Chinese group consisted mostly of students, with some recent graduates.

All participants completed a series of tasks recorded on the University of Melbourne campus in a sound-attenuated studio or a perception lab. The tasks included a wordlist, sociolinguistic interview of approximately one hour with two interviewers, mid-sagittal ultrasound imaging of words read aloud, a written background questionnaire including a social network task, and a twenty-minute vowel categorization task. In 2018, the 'Paddy's Day' sub-corpus was added, consisting of 17 Irish (7F, 10M) recorded completing the same wordlist at an outdoor festival. This sub-corpus includes participants who have been in Australia for over 40 years.

Preliminary analyses have shown LoR (Diskin et al. 2018a, 2019) and social network (Diskin-Holdaway 2021; Diskin-Holdaway et al. under review) to be correlated with adoption of Australian-like patterns of phonetics and discourse-pragmatics, as well as perception (see also Diskin et al. 2018b). English proficiency has had weaker effects in the acquisition of Australian-like perception of vowel contrasts (Diskin-Holdaway et al. under review). Further work investigating the ultrasound data is underway.

References

- Diskin, C., Loakes, D., & Clothier, J. (2018a). The phonetics and sociophonetics of laterals: a study of Irish migrants in Australia. Paper presented at *SocioPhonAus*, Brisbane, Australia.
- Diskin, C., Loakes, D., & Clothier, J. (2018b). Varietal differences in categorisation of /l e æ/ : A case study of Irish and Australian English listeners in Melbourne. *Proceedings of the 17th SST*, Coogee, Sydney.
- Diskin, C., Loakes, D., Clothier, J., & Volchok, B. (2019). A sociophonetic analysis of vowels produced by female Irish migrants: Investigating second dialect contact in Melbourne. *Proceedings of the International Congress for the Phonetic Sciences*, Melbourne, Australia, 1873-1877.
- Diskin-Holdaway, C. (2021). *You know and like* among migrants in Ireland and Australia. *World Englishes*, 40(4), 562-577.
- Diskin-Holdaway, C., Loakes, D., Clothier, J. (under review). Variability in cross-language and cross-dialect perception. How Irish and Chinese migrants process Australian English vowels. *Phonetica*.

The Greek-Australian Oral Archive Project

Eleni Elefterias-Kostakidis (University of New South Wales)

The Greek Australian Archive Project is a collaboration involving the University of NSW, the State Library of NSW and the Greek Orthodox Community of NSW. It is a collection of oral interviews and photos dealing with the experiences of Greek migrants who arrived in Australia from Greece, Cyprus, Egypt, Romania and Asia Minor during the post-war period. The project's original focus was on neighbourhoods and how the interviewees made a place for themselves and their families within the broader community. Interviewees were asked questions about their life before arrival in Australia and why they migrated, how they migrated and where they settled. Not every interviewee answered the same questions though there were a few general questions answered by all. Interviews were conducted in Greek or English or a mixture of both. There were challenges getting interviews from elderly people who spoke only Greek or a mixture of both 'broken' English and 'kitchen' Greek or who had heavy accents. There are many issues raised in the transcribing process because of these challenges, especially when dealing with different dialects. Other challenges included making the interviewee feel comfortable enough to open about their personal lives. The data that has been collected will be used as part of The Greek Australian Archive Website, currently being created by the State Library of NSW for free public access.

I will be presenting a couple of short extracts from live interviews and showing some photos of the subjects and their stories which may include the topics of travel to Australia, settlement, communal living on arrival, marriage, work, aspirations, challenges or possibly some of the more sensitive topics, such as that of racism, mental illness, domestic violence and/or depression.

The Little Multilingual Minds corpus: Educators' and children's speech in heritage languages

Paola Escudero, Gloria Pino Escobar (Western Sydney University),

Chloé Diskin-Holdaway and John Hajek (The University of Melbourne)

We introduce a corpus of child language in Spanish, Mandarin, and Vietnamese collected within the Little Multilingual Minds (LMM) program, which provides education in heritage languages (HL) for children of three major immigrant communities. We aim to document the development of languages other than English in Australia and assess the quality and quantity of input children receive when exposed to educational content in their HL.

In 2020, we conducted an observational study of a bilingual early childhood setting where children's language production amounted to only 5% of the recorded speech in 8 weeks, with approximately 3 hours of recordings per working day. The remaining 95% of speech was spoken by educators. We also found a decrease in Spanish language input in the transition from ages 2-3 to 3-5, when children are preparing for formal entry into primary school. Thus, our first LMM program was co-designed with the centre to address the language shift to English and enable children's Spanish development throughout their preschool years (ages 3 to 6) and beyond.

The current LMM corpus includes over 1300 LMM educational sessions of 45-90 mins each conducted entirely in Spanish for children from 2 to 8 years, featuring both facilitator/teacher input and children's voices. We also have over 300 one-on-one sessions assessing linguistic and cognitive skills in both English and Spanish, each lasting 40-60 mins. Additionally, our corpus includes over 100 LMM sessions conducted in Vietnamese and 35 in Mandarin for children aged 3 to 5, lasting between 30 to 45 mins each. We will present the corpus and the analysis conducted so far, as well as planned projects from HDR students and the research team. We will also present the current data management plan and our plans to collaborate with the wider community via a digital platform.

[A spoken corpus of Australian English by Russian and Chinese speakers](#)

Ksenia Gnevshva and Catherine Travis (The Australian National University)

Despite much recent interest in the impact of ethnic diversity on Australian English, linguistic analyses have primarily focused on native speakers of Australian English, or second-generation migrants. The Accented Australian English corpus aims to capture Australian English as spoken by a group not captured in such studies, namely second-language speakers of English who are first-generation migrants to Australia.

The corpus currently comprises recordings with 30 Russian and 30 Mandarin speakers who migrated to Australia as adults from Russian-speaking countries and mainland China, and have been in Australia for between 2 and 22 years. The recordings come from sociolinguistic interviews of around 60 minutes, conducted in English by community members, and covering topics such as the participants' experiences growing up in their home countries; their migration history; their linguistic background; encounters with discrimination; and so on. While all participants can converse with ease in English, they vary in their proficiency, allowing us to ask questions about language acquisition, with our main interest being in the acquisition of sociolinguistic variation, or the degree to which second-language speakers pick up variation patterns of the majority community. The corpus can also serve for perception work. One such project currently underway is exploring linguistic discrimination by using excerpts from the recordings as stimuli to assess Australian listeners' judgements about the employability of these second-language speakers in comparison with Australians with English as a first language. And the content of the interviews offers insights into the lived experiences of members of these migrant communities in contemporary Australia, contributing to the record of multilingual Australia.

The corpus was collected in 2019 in Sydney and Canberra, under funding from the Australian government. It will be onboarded to LDaCA, where it will be made accessible for approved projects.

'Oh, wunderbar, ein bisschen Deutsch!' German speakers in Newcastle and the Hunter Valley through the lenses of oral history and linguistics

Jaime Hunt and Sacha Davis (*The University of Newcastle*)

This talk presents a spoken corpus collected for the ongoing multidisciplinary linguistics/oral history project German as a heritage language and culture in Newcastle and the Hunter Valley. So far, the partially transcribed corpus contains over 40 hours of semi-structured interviews. The 38 participants, aged between 16 and 87 years (27 female and 13 male), have ancestry in the German-speaking areas in Europe and belong to either the first, second, or third generation of such migrants to the greater Newcastle region, who form the largest group in the region with non-Anglo-Celtic ancestry (ABS, 2019). Participants are from three migratory waves of German speakers: (1) those with German-Jewish ancestry arriving in 1938-1939, (2) post-war working-class refugees and economic migrants in the 1950s and 1960s, and (3) middle class/educated professionals arriving from the 1970s onwards.

Interviews with first-generation speakers are predominantly in German, while those with later generations are in both English and (where possible) German, depending on the interviewee's language skills. Thus, the data collected acts as a source for not only analysing the potential impact of English upon the first language of German-speaking migrants, but also contributes to a contemporary corpus of heritage German as spoken today. Furthermore, the data allows for the exploration of historical, social, and cultural trends in the German-speaking community as the topics covered in each interview include family origin and migration history, the settlement process in Newcastle, the formation and decline of German-speaking communities, the knowledge and use of German (both present and past), language education, family language policy, and ongoing connections with family abroad.

Our multidisciplinary approach has elucidated particularly rich descriptions of the migrant experience, including how German language practices have changed over time, its fluctuating utility and significance at the local and national levels, and the complex interplay between societal, familial, and individual attitudes to language use, as well as state-sanctioned immigrant language policies (Hunt & Davis, 2019, 2022). This allows for a deeper understanding of the social factors around community formation and maintenance, everyday practices of culture and language, familial and societal attitudes to linguistic and cultural identity, the rise and decline of a surprisingly complex patchwork of only partially overlapping German communities, and the role of familial networks and ties to Europe, which frequently take Novocastrians of German-speaking descent to different destinations to both places of family origin and current centres of German culture in Europe.

References

- ABS, A. B. o. S. (2019, 29.3.2019). Census. Commonwealth of Australia. Retrieved May 9, 2019 from <https://www.abs.gov.au/websitedbs/D3310114.nsf/Home/Census?OpenDocument&ref=top Bar>
- Hunt, J. W., & Davis, S. E. (2019). Social and historical factors contributing to language shift among German heritage-language migrants in Australia: An overview. *Linguistik online*, 7, 159-180. <https://doi.org/10.13092/lo.100.6025>
- Hunt, J. W., & Davis, S. E. (2022). 'So, mein Deutsch ist schlecht ... ': echoes of societal attitudes and education language policies within the family language policies of second- and third- generation German speakers in Newcastle, Australia. *International Journal of Multilingualism*, 1-18. <https://doi.org/10.1080/14790718.2022.2037609>

Government documents as a parallel multilingual corpus:

An opportunity which may be disappearing

Simon Musgrave (Language Data Commons of Australia)

Different levels of government in Australia have accepted for several decades that residents and citizens from culturally and linguistically diverse backgrounds benefit from having information provided to them in their primary languages. The federal government publishes *Australian Government Language Services Guidelines: Supporting access and equity for people with limited English* (most recent [version](#) 2019). State governments may have similar documents; for example, the Victorian government publishes [Using Translation Services - Victorian Government Guidelines](#) and [Using Interpreting Services - Victorian Government Guidelines](#). The result of such thinking is that many government documents are translated into multiple languages. How decisions are made about which languages to support varies, but it is certainly not uncommon for material to be provided in more than ten languages; a 2007 survey of Victorian Government websites found translated material in about 60 languages in one case. These relatively easily available materials could be collected to construct a multilingual parallel corpus which would be of great benefit to researchers in various fields.

Several factors contribute to the value of such a resource. Firstly, the material is consequential and therefore translations are prepared to high standards. Translated versions of government information may not have the full legal force of the English original, but great care is taken to make the versions in various languages as accurate a reflection of the original as possible. This also means that although a Cantonese version and a Turkish version of a document are not directly related by translation, each is a reliable translation of the same original and therefore it is possible to treat them as parallel texts. Therefore a collection of government documents in various languages would be both a series of highly parallel corpora, each consisting of material in English and one other language, but also at least to some extent a multilingual parallel corpus. Much of the material falls in areas such as health care and would support research on delivery of information to culturally and linguistically diverse populations on important topics. As mentioned, there is variation in which languages are used for which purposes, as well as variation over time which reflects demographic changes, but in a pilot project looking at one federal agency and one state agency, it was simple to find documents which had translations to a consistent group of five languages other than English.

Advances in machine translation technology are impacting on language services everywhere, including Australian government services. Several local councils use services such as Google Translate to make all of their online presence available in tens of languages. While this potentially opens another line of research, comparison of human translations and machine translations, moves in this direction suggest that if assembling government documents in multiple languages is a valuable project, undertaking it sooner rather than later is advisable. One would hope that documents are archived by government agencies, but the value of this material to researchers may not be understood and considered when making decisions about archiving and preserving material.

CanVEC – the Canberra Vietnamese English corpus

Li Nguyen (The Australian National University)

This talk introduces the Canberra Vietnamese-English corpus (CanVEC), an original corpus of natural speech produced by 45 Vietnamese-English bilingual speakers in Canberra, Australia. The corpus consists of 23 natural conversations (~90,000 words, 10 hours of speech) across two generations of speakers, ranging in age from 12–67. First generation is the first members of a speaker’s family who emigrated to Australia and have been living in Canberra for at least ten years, and second generation are those born in Australia or who arrived before the age of five, and whose parents qualify as the first generation. Generation thus can, but need not, correlate with age in the corpus.

We also collected other demographic information such as occupation or highest level of education, as well as participants’ self-rated language proficiency, language attitudes, ethnic orientation, and language use within their social network. To maximise its future use, we also semi-automatically annotated CanVEC with language-marking, Part-of-Speech (POS) tags and English translations (Nguyen & Bryant, 2020). This is first annotated corpus of natural speech of the Canberra Vietnamese community.

Data from CanVEC has allowed us to examine Vietnamese heritage language, Australian English, and code-switching as spoken in the community (Nguyen, 2021). We were also able to probe some little-understood variables such as Vietnamese subject and object expression, and consider the extent to which they appear vary across different generations. Excitingly, several computational works have also used the code-switching portion from CanVEC to build multilingual models and evaluate how natural code-switching production affects machine performance in various domains (Nguyen et al 2023a, Nguyen et al 2023b, Chan et al 2023). Ultimately, CanVEC offers a useful resource for researchers from various disciplines who are interested in the daily interactions and natural language use within the Vietnamese community in Australia.

Exploring the trilingualism of Italian migrants in Australia:

Evolving data, methodologies and insights

Antonia Rubino (The University of Sydney)

Studies investigating the sociolinguistic situation of the Italian migrants in Australia since mass migration date back to the late 1960s and span across different migrant generations and migration waves. Since the majority of post-war migrants arrived from Italy as dialect-Italian bilinguals, a major analytical focus has concerned the dynamics amongst the three main languages within the Italian Australian migrant community, namely, Italian, dialect(s) and English.

Research has been conducted through a wide range of approaches, at different sociolinguistic levels. Analyses of Census data have demonstrated that the use of Italian decreases across generations, albeit at an ‘intermediate’ rate compared with other migrant groups (Clyne 2005). Studies in the sociology of language framework (Bettoni & Rubino 1996) have shown that post-war migrants and their children self-report dialect as the dominant language in informal contexts, and Italian for more formal contexts; and that dialect is lost faster than Italian. The translingual practices of migrants and their children have been explored in interview settings (Bettoni 1991; Kinder 1994), in spontaneous conversations (Cavallaro 2010, Rubino 2014) and in the Italian speaking media (Rubino 2016), also accounting for ideologies and attitudes related to such practices.

Against the background of this rich tradition of studies, this paper will focus on a project conducted with young bi- (English, Italian) and trilingual (English, Italian and dialect) third generation Australian Italians, which explores their linguistic competence, practices, preferences and identity claims (Rubino 2021). The study was conducted through a mixed method approach. A detailed sociolinguistic survey was distributed online to obtain data from a relatively large sample of participants. This was accompanied by a set of in-depth interviews that delved further into the issues investigated through the survey. As will be discussed, this project yields important insights into long-term maintenance of both heritage languages amongst Australian Italians.

References

- Bettoni, Camilla. 1991. "Language variety among Italians." In *Language in Australia*, edited by Suzanne Romaine, 263-269. Cambridge: Cambridge University Press.
- Bettoni, Camilla, and Antonia Rubino. 1996. *Emigrazione e comportamento linguistico. Un'indagine sul trilinguismo dei siciliani e dei veneti in Australia*, Galatina: Congedo.
- Cavallaro, Francesco. 2010. *Transgenerational Language Shift: From Sicilian and Italian to Australian English*. Melbourne: Italian Australian Institute/La Trobe University.
- Clyne, Michael. 2005. *Australia's Language Potential*. Sydney: UNSW Press.
- Kinder, John. 1994. "Il recupero della sintassi nell'italiano della seconda generazione in Australia." In *Italiano lingua seconda/lingua straniera*, edited by Anna Giacalone Ramat, and Massimo Vedovelli, 343-361. Rome: Bulzoni.
- Rubino, Antonia. 2014. *Trilingual Talk in Sicilian-Australian Families. Playing out Identities through Language Alternation*. Palgrave Macmillan
- Rubino, Antonia. 2016. Constructing pseudo-intimacy in an Italo-Australian phone-in radio program, *Journal of Pragmatics*, 103, 33-48.
- Rubino, Antonia. 2021. Language competence, choice and attitudes amongst Italo-Australian youth, *Journal of Multilingual and Multicultural Development*, 42(3), 275-290.

Vietnamese language use and maintenance in Australia

Van Tran, Sharynne McLeod, Sarah Verdon, Kate Margetson (Charles Sturt University)

Vietnamese is one of the top three languages spoken in Australia, other than English, with its speakers comprising of approximately 1.2% of the Australian population. The Australian Research Council-funded Discovery Grant project, VietSpeech, is the first large-scale project studying language competence and home language maintenance of Vietnamese-Australian families. The project collected data from over 300 families with Vietnamese heritage living in Australia. Main studies include: 1) Vietnamese-Australian families' linguistic multi-competence and language maintenance ($n = 271$); 2) Australian Vietnamese-English-speaking children's speech acquisition in Vietnamese and English ($n = 154$); Development of a multilingual speech program ($n = 15$); and 4) Feasibility and efficacy of a Vietnamese-Australian children's speech and language maintenance program ($n = 30$). Key findings include: 1) Maintaining Vietnamese as a home language does not have a negative impact on English language competence; 2) Successful home language maintenance is associated with parents consistently speaking with children in Vietnamese from an early age and having positive attitudes towards home language and culture; 3) Both children and adults have a relatively high percentage of consonants correct (PCC) in both English and Vietnamese; and 4) Parents express a desire to support their children in home language maintenance and bilingual development but may lack knowledge on how to do so effectively. The study also emphasises the importance of collaboration from various stakeholders including parents, schools, community, and government to support Vietnamese language maintenance.

Corpora of Greek in Melbourne

James Walker, Stavroula Nikoloudis (La Trobe University)

This paper documents ongoing projects to develop corpora of spoken Greek in Melbourne, based on two sources: legacy archival materials and new sociolinguistic interviews.

The Dardalis Archives, a collection curated by the La Trobe University Library, consists of thousands of objects (photographs, books, correspondence, newspapers, films, cassette recordings) donated by members of the local Greek community, reflecting the migration of Greeks to Australia, especially during the post-WWII period. From 458 cassettes recorded between 1980 and 2005, we have identified 97 of at least average sound quality that contain interviews with over 200 migrants and prominent community members about migration, descriptions of their houses and the community and histories of Greeks in Australia. These recordings, which are in the process of being digitised, will be prioritised for transcription.

A newer project involves conducting sociolinguistic interviews with Melbourne residents of Greek background. In the first phase, conducted in late 2022 and early 2023, Greek Studies students at La Trobe University recruited 33 members of their extended social networks and engaged them in discussions of their migration and family history as well as administering a questionnaire about self-identification, social networks and language acquisition and use. Participants representing three generations and ranging in age from 18 to 82, were interviewed in Greek (15 preferred to speak English but all interviews contain some mixture of the two languages). These digitally recorded interviews, lasting between 34 minutes and over two hours, are being transcribed and the questionnaire answers extracted.

The purpose of these corpora is twofold: to provide an understanding of the Greek experience in Melbourne, in terms of migration and community history and patterns of language maintenance and shift, and to document the input varieties of Greek and the extent to which linguistic variation is being maintained in subsequent generations.