METHODS XVIII

Methods in Dialectology XVIII

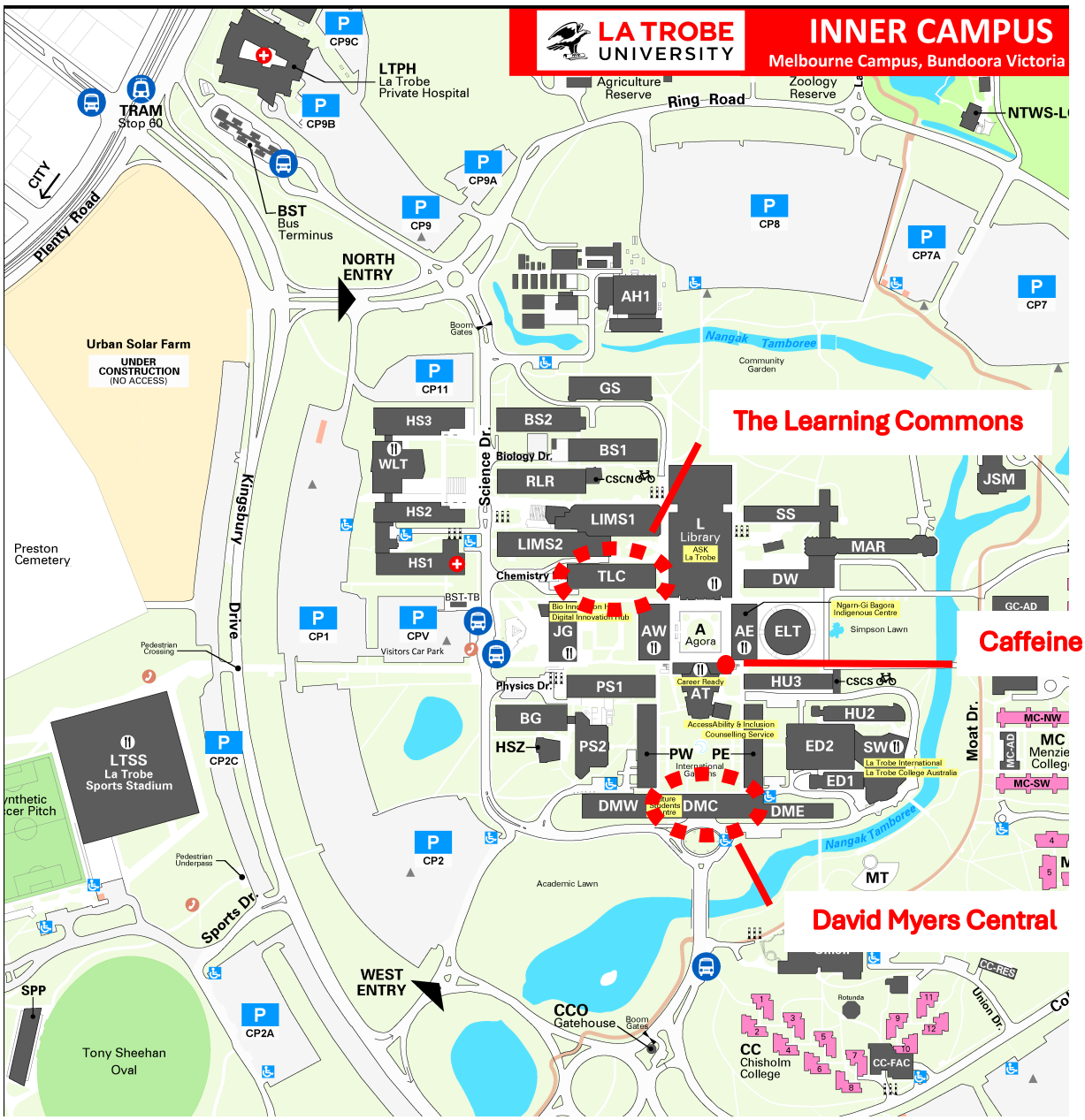
1-5 July 2024
La Trobe University
Melbourne

Conference Handbook

Campus Map

La Trobe University

Bundoora VIC



18th International Conference on Methods in Dialectology

METHODS XVIII

1-4 July 2024

La Trobe University

Melbourne, Australia

Program

	Monday 1 July 2024	
8:30-9:00	Registration (outside David Myers Central DMC-C121)	
9:00-9:30	Introduction and Welcome to Country (David Myers Central DMC-C121)	
9:30-10:30	PLENARY: **The Euphoria and Pitfalls of Working with Archival Language Recordings** Roy Barker, Alison Mount, Jane Simpson (Australian National University) (David Myers Central DMC-C121) *Chair: David Bradley*	
10:30-11:00	**BREAK**	
	Session 1 (Room: TLC-314)	Session 2 (Room: TLC-315)
	ABORIGINAL AUSTRALIAChair: Jane Simpson	*IDENTITYChair: Catherine Travis*
11:00-11:30	Felicity Meakins, Lindell Bromham and Jesse Stewart, "Modelling social salience across multiple variables to show (non)-associations with social categories."	Yoshiyuki Asahi, "A Japanese mayor's ways of using dialect in his political discourse: An analysis of a mayor of Nagoya."
11:30-12:00	Rita Calabrese and Katherine Russo, "Investigating multiword constructions in Indian English and Australian Aboriginal English: From compositionality to grammaticalization and idiomaticity."	Benjamin Purser, "Using variationist sociolinguistic research to inform dialect coaching for performance: addressing a methodological gap."
12:00-12:30	Raymond Kelly, Jean Harkins and Nicholas Watson, "Decolonising dialectology? Viewing language diversity from the inside."	
12:30-1:30	**LUNCH**	

	DIALECT DISTANCEChair: Yoshiyuki Asahi	*LANGUAGE PERCEPTIONChair: Jennifer Thorburn*
1:30-2:00	Hedwig Sekeres, Remco Knooihuizen and Martijn Wieling, “Quantifying dialect distance with acoustic models: A comparison to feature-based distance using single words.”	Heba Bou Orm, “Lebanese Youth Voices: Language attitudes and identity in Sydney.”
2:00-2:30	Carolina Zucchi, “Applying dialectometry to a categorical dataset built from published sources: Discussion of steps and methodological issues based on African Arabic dialects.”	Michol Hoffman and James Walker, “Self-perceptions of ethnic markers in Toronto English.”
2:30-3:00	Wilbert Heeringa and Fumio Inoue, “Relationships among Japanese local dialectal words by Levenshtein distance and multivariate analysis.”	Stefan Dollinger, “What do Canadians think of their English? A national study of language attitudes towards Canadian English.”
3:00-3:30	**BREAK**	
	STUDYING DIALECT DIFFERENCESChair: Stefan Dollinger	*CONTACTChair: David Britain*
3:30-4:00	Jennifer Thorburn, “Mapping attitudes in French-speaking Switzerland.”	Abahan Datta, “Language loss in Bengali: Conflict between standard and regional dialects.”
4:00-4:30	Simon Gonzalez, “Interrogating dialectal disparities in major speech datasets for ASR and linguistics research.”	Kazuko Matsumoto, Akiko Okumura and Kenjiro Matsuda, “Internal and external motivations for koiné formation: Transplanted Brazilian Portuguese in Japan.”
4:30-5:00	Raoul Buurke, Remco Knooihuizen, Wilbert Heeringa and Martijn Wieling, “Measuring sound differences between Standard Dutch and regional languages using automatic quantitative procedures.”	
6:00-8:00	**OPENING RECEPTION: Novotel Preston** **(coach service provided)**	

	Tuesday 2 July 2024	
9:00-9:30	Registration (outside David Myers Central DMC-C121)	
9:30-10:30	PLENARY: **Pursuing Sociolinguistic Diversity in Enregisterment-Rich Settings** Rebecca Starr (National University of Singapore) (David Myers Central DMC-C121) *Chair: Chloé Diskin-Holdaway*	
10:30-11:00	**BREAK**	
	Session 1 (Room: TLC-314)	Session 2 (Room: TLC-315)
	CREOLEChair: Felicity Meakins	*TIME AND SPACEChair: Keiko Hirano*
11:00-11:30	Carina Ahrens and James Grama, "Evaluating traditionally 'stable' variables in a contact language: A study of (th, dh) in Hawai'i Creole."	Takateru Kadoya, Soichiro Harada, Hitoshi Nikaido and Keiko Maeda, "The semantic change of intensifiers observed through geographical variation in the Goto Archipelago, Kyushu, Japan."
11:30-12:00	Kimberley Baxter, "'A who yuh tink dat see Tom?': Syntactic variation in Jamaican Patois."	Madeleine Clews, "The devil in the details: finding the Irish in 19th-century Australian English."
12:00-12:30	Aya Inoue and Amy J. Schafer, "Psycholinguistic evidence for the effect of phonemic orthography: Is it really that hard to process?"	
12:30-1:30	**LUNCH**	
	REGIONAL DIALECTSChair: Kazuko Matsumoto	*SOCIOPHONETICSChair: Michol Hoffman*
1:30-2:00	Catherine Travis, Gerry Docherty and Ksenia Gnevsheva, "Voices of Regional Australia: Assessing local attachment in legacy data."	Carina Ahrens and Anne-Marie Moelders, "Exploring acoustic realisations of Tyneside English: Homophonous realisation of the first-person possessive and object 'me'."
2:00-2:30	Martijn Wieling, Raoul Buurke, Teja Rebernik and Jidde Jacobi, "SPRAAKLAB: A mobile laboratory benefitting dialect data collection."	Lucia Fraiese, Glenys Dale Collard, Celeste Louro, James Walker and Matt Hunt Gardner, "Freakin swimming and everythink: Variable (ING) in youth Aboriginal English."
2:30-3:00	David Britain and Hannah Hedegard, "Keep it in the family: New methods for investigating the role of settler origins in new dialect formation."	Claire Jingyuan Ye, "Morphology and rhotacization in Beijing Chinese: Unravelling complex linguistic dynamics."

3:00-3:30	**BREAK**	
	INDIAChair: David Bradley	*CHANGE IN REAL TIMEChair: Celeste Rodriguez Louro*
3:30-4:00	Saurabh Nath, “Investigating variation of mid-front vowels in Assamese.”	Keiko Hirano, “A longitudinal study of consonant cluster reduction among British and American expatriates in Japan.”
4:00-4:30	Pranesh Bhargava and Krithika Nambiar, “A novel approach for exploring rapid intragenerational accent change through script reform: The case of Malayalam scripts.”	Shoji Takano and Yoshiyuki Asahi, “What constrains the malleability of idiolects over time?: A multivariate analysis of NINJAL”s real-time panel corpora of the standardization of Hokkaido (Furano) Japanese.”
4:30-5:00		James Grama, Isabelle Buchstaller, Anne-Marie Moelders and Lea Bauernfeind, “Disentangling style-shifting and life-span change: A panel study of FACE and first-person possessive ‘me’ in the North-East of England.”

	Wednesday 3 July 2024 **CONFERENCE EXCURSION: An exploration of the Yarra Valley**
8:30	Coach pickup at La Trobe University.
	Puffing Billy steam train tour through the Dandenong Ranges.
	Two-course lunch at Emerald Lake.
	Visit to Healesville Sanctuary to see an extensive range of native Australian animals, such as koalas, kangaroos, platypus, emus and others!
5:00	Coach return to La Trobe University

	Thursday 4 July 2024
	NEW DIALECTS (Room: TLC-314)*Chair: James Walker*
9:30-10:00	Felicity Meakins, Lindell Bromham, Xia Hua and Luis Miguel Rojas Berscia, "Linking micro to macro-level variation in the emergence of new dialects."
10:00-10:30	Fumio Inoue and Yasushi Hanzawa, "New dialect forms over the 250 Years since 'Hamaogi': Geographic and age trends based on multiple correspondence analysis."
10:30-11:00	**BREAK**
	ORGANISED SESSION: LINGUISTIC COVARIATION AND COHERENCE (Room: TLC-314) *Chair: Elena Sheard*
11:00-11:15	Introduction
11:15-11:45	Elena Sheard, Robert Fromont, Josh Wilson Black, Jen Hay, Lynn Clark, and Gia Hurring, "The social meaning and perceptual accessibility of co-varying New Zealand English monophthongs."
11:45-12:15	Elena Sheard and Josh Wilson Black, "Change over the lifespan across covarying New Zealand English monophthongs."
12:15-12:45	Gia Hurring, "Exploring multiplane covariation in New Zealand English."
12:45-1:45	**LUNCH**
1:45-2:15	James Walker, "Co-variation in ethnolects."
2:15-2:45	Catherine Travis and Gan Qiao, "Using covariation to test the intersection of ethnicity and social class."
2:45-3:15	Discussion
3:15-3:45	**BREAK**
3:45-4:45	BUSINESS MEETING (David Myers Central DMC-C121)
4:45-5:45	PLENARY: **/r/ You Listening? Methodological Challenges and Theoretical Insights from Investigating a 'Rhotic' English Dialect** Jane Stuart-Smith (University of Glasgow) (David Myers Central DMC-C121) *Chair: James Walker*
5:45-6:00	CLOSING COMMENTS (David Myers Central DMC-C121)
7:00-10:00	**CONFERENCE DINNER: Welcome to Thornbury (coach service provided)**

METHODS XVIII

Plenary Presentations

Title: **The euphoria and pitfalls of working with archival language recordings**

Authors: **Roy Barker, Alison Mount, Jane Simpson**

We discuss an audio collection of national significance, the 'Barker Collection' of 113 hours of audio recordings, made by Jimmie Barker (1900-1972) between 1968 and 1972. It is one of the largest such collections recorded by an Aboriginal person in Australia. It is also one of the earliest known instances of an Aboriginal person independently recording other speakers for language documentation. The collection is held at the Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS).

Jimmie Barker (1900-1972) was a Muruwari Elder, Cultural Knowledge Holder, linguist, historian, ethnographer, inventor, and sound engineer. The content in his recordings is wide-ranging, from documenting language – including placenames and songs – Muruwari practices and natural history knowledge, to analysing linguistic features and commenting on audio recordings of other speakers and data collected by non-Muruwari researchers. Jimmie also reflects in English on history, including frontier conflict, and current events, including space travel, technological innovations and Halley's comet.

Since 2021, Roy Barker, grandson of Jimmie Barker and Muruwari Cultural Knowledge Holder, has worked to make his grandfather's audio collection culturally accessible, searchable, and useable. He has overseen a team of linguists building an ELAN corpus of the recordings. This first required working with family members to develop a spelling system, and also involved content tagging for building an online cultural portal to the recordings. The project has been an immense source of joy and healing for the extended Barker family.

We describe the process of building the corpus to align with the FAIR principles (findable, accessible, interoperable and reusable) (Wilkinson et al., 2016) and the CARE principles [Collective Benefit, Authority to Control, Responsibility, and Ethics] to support Indigenous data Governance (Carroll et al., 2020). We reflect on unanticipated challenges arising from the intersection of Indigenous cultural and intellectual protocols, Australian copyright law, and the responsibilities of institutions towards depositors and potential users. These have delayed making the corpus publicly available, as well as the distribution of language learning material generated from the recordings. As a result, we encourage researchers working with archival recordings, or indeed creating new recordings, to work closely with community to anticipate and address potential barriers to current and future access.

Pursuing sociolinguistic diversity in enregisterment-rich settings

Rebecca Lurie Starr
National University of Singapore

Prior work on language variation and the enregisterment (Agha 2003) of novel styles views these processes primarily as incidental outcomes of other social phenomena, such as mobility and societal change. In certain circumstances, however, sociolinguistic diversification may be intentionally pursued, with the aim of creating a maximum number of distinctive identities for the purpose of commodification.

In this talk, I consider the circumstances and consequences of what I call *enregisterment-rich settings*, in which individuals and communities actively pursue maximal enregisterment of distinctive styles as an implicit or explicit aim. I highlight the case of Chinese video-sharing platforms as an enregisterment-rich setting using data from three studies by my students and myself involving novel registers: fizzy voice (*qipao yin*), clip voice (*jiazi yin*), and ASMR. The rapid establishment, circulation, and commodification of these registers illustrates both how easily new registers may be constructed in these settings but, at the same time, how difficult it is for novel registers to escape the legacy of existing styles and social types.

Findings from the examples above suggest that the rapid construction of novel registers in these settings results in several common outcomes, including the utilization of these registers in a wide range of commodification processes, contestation of register boundaries, the accretion to new registers of features linked with more established registers, and the use of satirization to display registers without affiliating with their associated personae or in order to problematize the authenticity of their construction. As enregisterment-rich settings become more prominent due to technological and economic shifts, we must consider the construction of sociolinguistic distinction not merely as a part of social life, but, increasingly, as an aim that individuals and institutions may actively pursue as a means of generating commodifiable content. Under these circumstances, we must query assumptions in the field regarding the nature of registers, including how they form, their indexical links with social attributes, and how they may be reinterpreted and repurposed.

/r/ you listening? Methodological challenges and theoretical insights from investigating a 'rhotic' English dialect

Jane Stuart-Smith, University of Glasgow

Some sounds appear to do more social 'work' than others. One such sound is /r/, whose phonetic variants have long been recognized to be socially informative across many languages, including English (Labov 1972). Indeed, the presence/absence of postvocalic /r/ in in words such as *car, card* is a key feature distinguishing rhotic from non-rhotic English dialects (Maguire et al. 2010). At the same time, /r/ is phonetically complex which means that tracking synchronic and diachronic variation in /r/ can present challenges – and opportunities – for the linguist.

This talk will focus on the case of postvocalic /r/ in Scottish English, taking as its basis, results from a series of apparent- and real-time sociophonetic studies, in the field and in the lab, using auditory, articulatory and (different kinds of) acoustic analysis, carried out from 1998 until now. Our early studies confirmed earlier reports of socially-stratified variation in Glaswegian /r/, and specifically loss of coda /r/ in young working-class speakers (Stuart-Smith 2003; Stuart-Smith et al 2007). They also led to the following questions:

- What are the phonetic mechanisms behind the 'loss' of postvocalic /r/? Difficulties with the auditory coding of weak /r/ variants, including ambiguous auditory percepts (Stuart-Smith 2007; Lennon 2024), were resolved by articulatory phonetic analysis, which reveal both tongue shape and timing of tongue gesture as factors in weaker/stronger /r/ variants, and their acoustic correlates (Lawson et al 2011; Lawson et al 2018; Lawson and Stuart-Smith 2021).
- When did /r/-weakening begin, what was the trigger and then the accelerating factors which helped it to spread over the course of the 20th century? Scottish soldiers recorded during the First World War already show weak /r/ (Stuart-Smith and Lawson 2017), which coincides with a change in voice quality around the same period (Sóskuthy and Stuart-Smith 2020); the change took off and was accelerated by sociolinguistic polarisation and TV influence (Stuart-Smith et al 2014).
- How does Glasgow /r/ compare with /r/ more generally? Scaling up the analysis of /r/, over larger numbers of speakers/tokens, allows us to zoom out from local closeups of Glaswegian, to a perspective of postvocalic /r/ within a common acoustic space across Scottish and English dialects. Finally, automated methods also facilitate comparison of Scottish /r/ variation with languages other than English, which are also showing /r/ weakening, specifically Quebec French.

References

Labov, William 1972. *Sociolinguistic Patterns*. Oxford: Blackwell.

Lawson, E., J. Stuart-Smith, and J.M. Scobbie 2018. *The Role of Gesture Delay in Coda /r/ Weakening: An Articulatory, Auditory and Acoustic Study*. *Journal of the Acoustical Society of America* 143 (3): 1646–57. https://doi.org/10.1121/1.5027833.

Lawson, Eleanor, James M. Scobbie, and Jane Stuart-Smith 2011. *The Social Stratification of Tongue Shape for Postvocalic /r/ in Scottish English*. *Journal of Sociolinguistics* 15 (2): 256–68. https://doi.org/10.1111/j.1467-9841.2011.00464.x.

Lawson, Eleanor, and Jane Stuart-Smith 2021. *Lenition and Fortition of /r/ in Utterance-Final Position, an Ultrasound Tongue Imaging Study of Lingual Gesture Timing in Spontaneous Speech*. *Journal of Phonetics* 86 (May): 101053. https://doi.org/10.1016/j.wocn.2021.101053.

Lennon, Robert 2024. *Perception of Ambiguous Rhoticity in Glasgow*. *Journal of Phonetics* 104 (May): 101312. https://doi.org/10.1016/J.WOCN.2024.101312.

Maguire, Warren, April Mcmahon, Paul Heggarty, and Dan Dediu 2010. *The Past, Present, and Future of English Dialects: Quantifying Convergence, Divergence, and Dynamic Equilibrium*. *Language Variation and Change* 22 (1): 69–104. https://doi.org/10.1017/S0954394510000013.

Sóskuthy, M., and J. Stuart-Smith 2020. *Voice Quality and Coda /r/ in Glasgow English in the Early 20th Century*. *Language Variation and Change*. https://doi.org/10.1017/S0954394520000071.

Stuart-Smith, J, E Lawson, and J Scobbie 2014. *Derhoticisation in Scottish-English: A Sociophonetic Journey*. In *Advances in Phonetics*, edited by C Celata and S Calamai, 57–94. John Benjamins. https://doi.org/10.1075/silv.15.03stu.

Stuart-Smith, Jane 2003. *The Phonology of Modern Urban Scots*. In *The Edinburgh Companion to Scots*, edited by John Corbett, Derek J McClure, and Jane Stuart-Smith, 110–37. Edinburgh University Press.

Stuart-Smith, Jane 2007. *A Sociophonetic Investigation of Postvocalic/r/in Glaswegian Adolescents*. In *The 16th International Congress of Phonetic Sciences*, edited by Jurgen Trouvain and William Barry, 1449–52. Saarbrucken.

Stuart-Smith, Jane, and Eleanor Lawson 2017. *Scotland: Glasgow/the Central Belt*. In *Listening to the Past: Audio Records of Accents of English*, edited by Ray Hickey, 171–98. Cambridge: Cambridge University Press. http://eprints.gla.ac.uk/124510/13/124510.pdf.

Stuart-Smith, Jane, Claire Timmins, and Fiona Tweedie 2007. *Talkin' "Jockney"? Variation and Change in Glaswegian Accent*. *Journal of Sociolinguistics* 11 (2): 221–60. https://doi.org/10.1111/j.1467-9841.2007.00319.x.

METHODS XVIII

Methods in Dialectology XVIII

1-5 July 2024
La Trobe University
Melbourne

Regular Papers

Evaluating traditionally "stable" variables in a contact language: A study of (th, dh) in Hawai'i Creole

Carina Ahrens & James Grama
Sociolinguistics Lab, University of Duisburg-Essen

Pidgin (known to linguists as Hawai'i Creole) is an English-lexified creole spoken by some 700,000 speakers across the Hawaiian Islands (and in expatriate communities in the mainland United States; Velupillai 2015). While recent work has identified a number of changes in progress in the vowel system (Grama 2015, 2023) and in discourse-pragmatic *like* (Stabile 2019), little is known about how purportedly diachronically "stable" or traditional variables behave for speakers of Pidgin.

This presentation focuses on variation in the realization of interdental fricatives (here represented as (th) and (dh) for voiceless and voiced varieties). In English (the main lexifier of Pidgin), realizations of these variables (particularly (dh)) is a classic stable variable, contingent on speaker age, gender, and socio-economic status (see, e.g., Labov 2001). Stop realizations of (th, dh) are a typical feature of English-lexified creoles worldwide (Haspelmath 2003), and in Pidgin, these realizations are attested alongside a coronal flap and fricative variants (Sakoda & Siegel 2003). Our study is the first to attempt to circumscribe this variation via a quantitative analysis of the diachronic development of this variable.

Analysis of variation in canonical (th, dh) by local Pidgin speakers from Hawai'i is conducted by drawing on two trend corpora recorded 30 years apart: a subset of the Bickerton collection recorded in the 1970s (Bickerton & Odo 1976) and a comparable corpus recorded in the 2000s (Siegel 2004). Speakers (n=32) are evenly balanced across age (old v. young) and gender (men v. women). In combination, these corpora allow for the investigation of (th, dh) over real- and apparent time. We analyze approximately 8,500 tokens of (th, dh), considering social factors (i.e., gender, age), as well as position of the phone in the word, and grammatical category (i.e. function, content).

Preliminary results suggest a complex pattern. Stop variants are overwhelmingly the most common for canonical (th, dh), alongside fricatives, flaps, and an elided variant. For canonical (dh), two separate changes in progress are evident. In function words, (dh)-stopping shows a recent decrease in the young 2000s speakers, with a concomitant rise in flaps and the elided variant. Fricative forms remain a stable, minority form. In content words, a change in progress towards fricative forms is evident; however, this change is reversed in 2000s speakers, suggesting a shift away from prescribed English norms that coincides with a rise in the overt local prestige of Pidgin (similar to patterns reported in Cajun English; Dubois & Horvath 1998). Canonical (th), by contrast, appears to be a stable variable. In medial and final position, stops are more likely than fricatives, which are more likely than the elided variant. In initial position, women are produce higher rates of fricatives than men. In addition, men show age-graded behavior, where young men in both corpora show greater rates of stops than older men.

These findings indicate that pressures operative in the social hierarchies of lexifiers are not necessarily transferred to creoles, even if those hierarchies are reproduced and maintained by sustained colonizer presence (consider Trask 2000).

References

Bickerton, Derek, and Odo, Carol. 1976. *Change and Variation in Hawaiian English I: General phonology and Pidgin syntax*. Final report on NSF No. GS-39748.

Dubois, Sylvie, and Horvath, Barbara M. 1998. Let's tink about dat: Interdental fricatives in Cajun English. *Language Variation and Change 10*: 245-261.

Grama, James. 2015. Variation and change in Hawai'i Creole vowels. PhD thesis. University of Hawai'i at Mānoa.

Grama, James. 2023. Change over time in [ɛ] and [æ] in Hawai'i Creole. In Skarnitzl, Radek & Volín, Jan (eds.) *Proceedings of the 20th International Congress of Phonetic Science*, 3010-3014. Guarant International.

Haspelmath, Martin and the APiCS Consortium. 2013. Interdental fricatives. In: Michaelis, Susanne Maria & Maurer, Philippe & Haspelmath, Martin & Huber, Magnus (eds.), *The atlas of pidgin and creole language structures*. Oxford: Oxford University Press.

Labov, William. 2001. *Principles of Linguistic Change, Volume 2: Social Factors*. Blackwell.

Sakoda, Kent, and Siegel, Jeff. 2003. *Pidgin Grammar: An Introduction to the Creole Language of Hawai'i*. Bess Press, Inc.

Siegel, Jeff. 2004. External influences and internal variation in current Hawaii Creole English. National Science Foundation (Grant number: 0345959).

Stabile, Claire. 2019. "Like, Local people doing that": Variation in the production and social perception of discourse-pragmatic like in Pidgin and Hawai'i English. PhD thesis. University of Hawai'i at Mānoa.

Task, Haunani-Kay. 2000. Settlers of color and "immigrant" hegemony. *Amerasia Journal 26*(2): 1-24.

Velupillai, Viveka. 2015. *Pidgins, Creoles and Mixed Languages*. Creole Language Library, Vol. 48. John Benjamins Publishing Company: Amsterdam/Philadelphia.

Exploring acoustic realisations of Tyneside English: Homophonous realisation of the First-Person Possessive and Object 'me'?

Carina Ahrens & Anne-Marie Moelders
Sociolinguistics Lab, University of Duisburg-Essen

This paper presents an acoustic analysis of the first possessive (1POS), which can be realized as [maɪ], [mi], [mə], and [mɑ] in the North-East of England (Childs 2013). To date, the research has focused on the change of variants in real and apparent time (Childs 2013, Moelders to appear): Trend data shows that [mɑ] is currently replacing [mi] across time in Tyneside (Moelders to appear). Detailed analysis reveals that proportional differences in the realization of the variable are contingent on social factors: some speakers, especially those working as nurses, draw on the covert prestige of [mi] to index a stance of local belonging as it carries covert prestige as it allows speakers to do local identity work, fostering a joint sense of belonging. Meanwhile, speakers in language-sensitive occupations (such as educators, see Grama et al. 2023) orient away from stigmatized [mi] while they are part of the marché scolaire (Bourdieu & Boltanski 1975) but relax back into vernacularity as they leave the profession. Less attention has been paid to variation in the phonetic realization of 1POS. The present study fills that gap by exploring a panel sample of 12 speakers aged 19 to 72 from Tyneside in the North-East of England. Our analysis follows previous research suggesting that systematic variation is encoded at the level of the lemma (e.g., Drager 2011). Gahl (2008) refuted the assumption that perceived homophones should, as homophones are defined as having the same acoustic realization, be realized in the exact same way regardless of their meaning or function. As a result of phonetic differences in homophone pairs in which one meaning/function is used more frequently than its counterpart, Gahl suggests that acoustic information is stored individually in a speaker's mind for different lemmas. To this end, we investigate the phonetic profile of 1POS and the extent to which lemma-specific realizations of 1POS [mi] are different from the first-person object (1OBJ) [mi]. Auditory coding of approximately 1,500 tokens is combined with the acoustic analysis of F1 and F2 (measurements taken at 5% intervals) over the trajectory of the vowel. Preliminary results indicate that the acoustic realization of [mi] differs depending on its function as 1POS or 1OBJ. When fulfilling the role of 1OBJ, [mi] is longer than when acting as 1POS while the Euclidean Distance stays the same, which results in quicker rate of change over the trajectory, suggesting shortening of the lemma/word of higher frequency. These findings, which are fully in line with Gahl (2008), support exemplar based models of processing, which assume that lemmas/words are stored with acoustic information (e.g., Pierrehumbert 2002). Moreover, our analysis contributes a fine-grained acoustic analysis to the discussion about whether [mi] is either a phonetic realization of 1POS (as argued by Childs 2013) or whether it should be interpreted as "an extension of the object form" for the possessive (Anderwald 2004:177).

References:

Anderwald, Lieselotte. 2004. The varieties of English spoken in the Southeast of England: Morphology and syntax". In Kortmann et al. (eds.), A handbook of varieties of English (Vol. 2), 1367-88.

Bourdieu, Pierre, and Luc Boltanski. 1975. Le Fétichisme de La Langue. Actes de La Recherche En Sciences Sociales 1(4), 2–32.

Childs, Claire. 2013. „I couldn’t really put [mə] finger on it“: Phonetic realisations of the possessive singular ‚my‘ in Tyneside English”. Newcastle Working Papers in Linguistics.

Drager, Katie K. 2011. Sociophonetic variation and the lemma. Journal of Phonetics (Vol.39), 694-707.

Gahl, Susanne. 2008. Time and Thyme Are not Homophones: The Effect of Lemma Frequency on Word Durations in Spontaneous Speech. Language 84(3), 474-496.

Grama, James, Mechler, Johanna, Bauernfeind, Lea, Eiswirth, Mirjam. E., and Buchstaller, Isabelle. (2023). Post-educator relaxation in the U-shaped curve: Evidence from a panel study of Tyneside (ing). Language Variation and Change, 1–26.

Moelders, Anne-Marie. Under Revision. Navigating the vernacular across the life-span: A panel study of the phonetic realization of the first person possessive. Under revision at English Language and Linguistics.

Pierrehumbert, Janet B. 2002. Word-specific phonetics. In Gussenhoven & Wagner (Eds.) Papers in laboratory phonology VII, 101-139.

The role of dialect in the political to construct register in politics
-a case of a major in Nagoya city-

Yoshiyuki Asahi
National Institute for Japanese Language and Linguistics

This paper examines stylistic variation in the speeches by a mayor of Nagoya, Japan, Mr. Takashi Kawamura and discusses how he constructs his identity through the use of one phonological feature in Nagoya dialect. Mr. Kawamura was first elected to be a member of the Japanese Diet from 1992 when he appeared in a public domain. Although at the beginning of his career, he did not use his home dialect features, he gradually employed his Nagoya dialect features in his speeches and discussion at the committee meetings. By the time he became a mayor of Nagoya in 2009, he was a well-known figure who command a large degree of his Nagoya dialect features at national level.

This paper pays a close attention to one variable, monophthongnization of /ai/ and /ae/ into /æ/. This is one of the most distinctive features in Nagoya dialect. I have collected Mr. Kawamura's speech in a number of occasions such as his speech at the general meetings of the Nagoya local Assembly as well as other occasions such as his speech at National Diet when he was one of the members in 1990s, Also, he has shown at TV shows and his speech in various part of Nagoya city as well. The total amount of data sums up to 5.5 hours. I conducted transcriptions, and made quantitative analyses.

The result clearly supports my assumption. He came to use /æ/ in his discourse in accordance with his career. This trend is emphasized when he became a mayor of Nagoya in 2009 although other members of Nagoya city council rarely use this phonological feature. This paper will point out on the basis of the result, that he uses his strategy to perform his character as a major, and he intentionally increase his use of /æ/ in his discourse. This paper discusses this finding in relation to relevant literatures such as Harrington (2006), Mackenize (2017), Sankoff (2019) , and Schilling-Estes (1998).

References:

Harrington, Jonathan. 2006. 'An acoustic analysis of 'happy-tensing'in the Queen's Christmas broadcasts', Journal of Phonetics, 34(4): 439-57.

Mackenzie, Laurel. 2017. 'Frequency effects over the lifespan: A case study of Attenborough's r's', *Linguistics Vanguard*, 3(1).

Sankoff, Gillian. 2019. 'Language change across the lifespan: Three trajectory types', *Language*, 95(2): 197-229.

Schilling-Estes, Natalie. 1998. Investigating 'self-conscious' speech: The performance register in Ocracoke English. *Language in Society* 27.1: 53-83.

"A who yuh tink dat see Tom?": Syntactic Variation in Jamaican Patois

Kimberley Baxter
New York University

Jamaican Patois (JP), has three complementizer-like forms which share meaning with the English complementizer *that*: *seh* (verbal: *say*; complementizer: *that*), which can only be used with thinking, saying, and feeling verbs and adjectives (1); *weh* (relative marker: *what*, *who(m)*, *where*, complementizer: *that/which*) which can only be used for nouns (2); *dat* (complementizer: *that*) a true cognate of the English complementizer *that* (3). This study focuses on *dat* and aims to answer the following questions: Do JP speakers accept subject extraction over the complementizer *dat* from an embedded clause? If so, does acceptability vary across the Jamaican Diaspora?

(1) Mi seh to im seh im nuh
I-1SG say-VERB to him that-COMP he NEG-is not
fi forward ya.
to-PREP come-VERB here.
I said to him that he is not to come here./I told him not to come here.

(2) A Which one a di ting weh yuh did send mi fi get?
FOC-is which one is the thing that-COMP yuh did send me to get?
*Which one of the things (*that) did you send me to get?*

(3) A who yuh tink dat meet Tina?
FOC-is who you think that-COMP meet-PST Tina?
*Who do you think (*that) met Tina?*

The grammatical status of "say" complementizers like *seh* which flout complementizer trace effects in languages like JP has long been the subject of debate (Koopman and Sportiche, 1989; Durrleman, 2008; Veenstra and Besten, 1994; etc.). These studies bypass *dat*, assuming that because subject extraction over an overt complmentizer is ungrammatical in English, it must be the same in JP, an English-lexifier language. To fill this gap, I conduct a judgment survey of JP speakers ($n=35$) from across the Jamaican Diaspora to test the acceptability of subject extraction over *dat* and other complementizers in JP. The survey follows Sobin (1987), which tested the acceptability of subject extraction over a series of complementizers in American English and found variable acceptability of this movement over the complementizer *that*. Similarly, the present study presents participants with 43 utterances including instances of subject extraction, object extraction, and no extraction over the complementizer, which they rate from 1 (least natural) to 5 (most natural).

Most JP participants show a passive acceptance of subject extraction over *dat* (51.4% compared to 47.6% in Sobin (1987)). JP participants show higher rates of strong acceptance

(25% compared to 16.7% in Sobin (1987)) and lower rates of strong rejection (22.8%, compared to 35.7% in Sobin (1987)). Strong acceptance is even higher among participants in Jamaica (36.4%) and the United States (30.8%). Participants in Jamaica have the lowest rate of strong rejection (18.18%), followed by the US (23.1%), and all participants outside Jamaica (25%). This suggests subject extraction over *dat* is more widely accepted in JP than originally thought. These results support previous conclusions in (Baxter, 2023) which show higher acceptance rates among participants in Jamaica and underscores the importance of the viewpoints of language speakers to the formation and analysis of syntactic theory.

References

Baxter, K. (2023). *Complementizer Trace Effects in Jamaican Patois*, poster presented at Linguistic Society of America, 1/7/2023

Besten, H. D., & Veenstra, T. (1994). Fronting. Pidgins and Creoles: An introduction. Amsterdam*: John Benjamins Publishing*, 303-315.

Durrleman, S. (Ed.). (2008). The syntax of Jamaican Creole: A cartographic perspective (Vol. 127). *John Benjamins Publishing*.

Koopman, H., & Sportiche, D. (1989). Pronouns, logical variables, and logophoricity in Abe. *Linguistic Inquiry*, 555-588.

Sobin, N. (1987). The variable status of Comp-trace phenomena. *Natural Language & Linguistic Theory*, 5(1), 33-60.

A Novel Approach for Exploring Rapid Intragenerational Accent Change through Script Reform: The Case of Malayalam Scripts

Dr. Pranesh Bhargava
Assistant Professor, BITS Pilani's Hyderabad Campus
pranesh@hyderabad.bits-pilani.ac.in

Krithika Nambiar
PhD Scholar, BITS Pilani's Hyderabad Campus
krithika.nambiar@wilp.bits-pilani.ac.in

Accent change in younger generations compared with previous generations is driven by external (socio-political) and internal (linguistic) factors [1].

Research on dialects and accents focuses on gradual, long-term changes distinguishing one regional accent from another [2]. However, due to shifts in the linguistic environment, e.g. due to government policies, accents can change within a generation, which these methods may struggle to study [3], [4]. Thus, people who grow up speaking the 'old' accent (OA) might later be exposed to a 'new' accent (NA) as they age [5], leading to interesting questions: e.g. as a consequence of intragenerational accent change, does an individual's spoken accent production adapt or remain stable? How does the comprehension of NA compare to the comprehension of the OA?

We propose a novel approach to studying language variation by examining script reforms as a model system [6]. Just like accents are the various spoken forms expressing the same linguistic content, scripts are different ways of expressing the same linguistic content through written form [7]. Thus, by considering *scripts as visual accents*, accent variations can be studied by proxy.

Malayalam script from Kerala, India, provides a unique opportunity to investigate the impact of rapid, externally-driven language change on individuals within their lifetime. In 1971, the Kerala government in India implemented script reforms, replacing the traditional Malayalam script (TS) with the Reformed Script (RS) [8]. This created two distinct reader groups: those literate in TS before the reform and those literate in RS after the reform.

We investigated whether elderly TS experts would process sentences in TS faster than RS or prefer TS over RS for speaking/writing, considering their childhood/youth exposure to TS, and recent exposure to RS through media and the environment.

With sixty (39 females, range 17-80 years, no cognitive impairment) Malayalam speakers, divided into pre- and post-1971 script reform groups based on age, we used a self-paced reading task for comprehension to measure the reading speed, using 50 sentences in both RS and TS script; and an oral dictation task for the choice of script in the production task.

Results show that the elderly TS experts consistently chose TS for writing [6]. However, surprisingly they read the RS sentences faster than the TS sentences. This might be because the orthographic knowledge of script exists at two levels, i.e. sub lexical, which is more resistant to exposure, and lexical which is prone to be updated due to exposure [9], [10].

On the basis of this finding, we expect that the elderly generation may comprehend NA easily and faster compared to OA. We further expect that when a language production task is given to the elderly generation, they are likely to produce OA rather than NA. We speculate that there might be two levels of phonological accent to a dialect/language similar to orthographic levels in reading and writing [9], [10]. Further studies may find evidence for this.

REFERENCES

[1] F. Cox and S. Palethorpe, "History and accent change," Macquarie University. Accessed: Mar. 07, 2024. [Online]. Available: https://www.mq.edu.au/research/research-centres-groups-and-facilities/healthy-people/centres/centre-for-language-sciences-clas/australian-voices/history-and-accent-change

[2] P. Trudgill, "The Historical Sociolinguistics Of Elite Accent Change: On Why RP Is Not Disappearing," vol. 44, pp. 3–12, Jan. 2008.

[3] O. Mantiri, "Factors Affecting Language Change." Rochester, NY, Mar. 17, 2010. doi: 10.2139/ssrn.2566128.

[4] M. J. Esman, "The State and Language Policy," *International Political Science Review*, vol. 13, no. 4, pp. 381–396, Oct. 1992, doi: 10.1177/019251219201300403.

[5] J. K. Chambers, "Dialect Acquisition," *Language*, vol. 68, no. 4, pp. 673–705, 1992, doi: 10.2307/416850.

[6] K. Nambiar, K. Kishore, and P. Bhargava, "The effect of script reform on levels of orthographic knowledge: Evidence from alphasyllabary Malayalam scripts," *PLOS ONE*, vol. 18, no. 8, p. e0285781, Aug. 2023, doi: 10.1371/journal.pone.0285781.

[7] J. Vaid, "Biscriptality: a neglected construct in the study of bilingualism," *J Cult Cogn Sci*, vol. 6, no. 2, pp. 135–149, Sep. 2022, doi: 10.1007/s41809-022-00101-3.

[8] Government of Kerala, "Malayalam Script-Adoption of New Script for Use-Orders Issued." Government of Kerala, 1971.

[9] K. Apel, "What Is Orthographic Knowledge?," *LSHSS*, vol. 42, no. 4, pp. 592–603, Oct. 2011, doi: 10.1044/0161-1461(2011/10-0085).

[10] K. Apel, V. S. Henbest, and J. Masterson, "Orthographic knowledge: clarifications, challenges, and future directions," *Read Writ*, vol. 32, no. 4, pp. 873–889, Apr. 2019, doi: 10.1007/s11145-018-9895-9.

Keep it in the family: new methods for investigating the role of settler origins in new dialect formation

David Britain and Hannah Hedegard (Bern)

For koineisation and new dialect formation to take place, interactive contact is necessary, and mere propinquity or proximity is not sufficient. Many of the 'classic' examples from the literature of new dialect formation come from sites, therefore, where interaction between speakers of different dialects was intensive, for example, in New Towns (Kerswill and Williams 2000), or in thriving colonial settlements (Trudgill et al 2004). As a result, through levelling, reallocation, simplification and interdialectalisation, new dialects emerge that are distinct from the settler dialects from which they were generated. In some cases, however, where interspeaker contact is limited, new dialect formation is slower, and traces of settler dialects linger longer as the new dialect remains more diffuse and takes more time to focus. One often cited case of this 'slower focussing' is to be found in the English of the Falkland Islands (FIE), a British Overseas Territory in the South Atlantic, Anglophone since 1833, in which only 3500 people live in an area larger than Cyprus and Lebanon. Dominant among the 19th century settlers were migrants from Scotland and the English South-West, although the islands have also seen immigration from south-eastern England and Ireland, and, more recently, other countries. Because of a scattered population and adverse terrain on the islands, inter-settlement communication until very recently was very difficult leading Trudgill to argue that many parts of islands outside of the capital Stanley had 'retained dialects of English which reflect the area of origin of the first settlers…very little mixing has taken place and very little focussing' (1986: 128). In this presentation, we wish to use a family tree methodology to empirically test this claim.

Our data come from a near-million word corpus of informal conversational FIE, collected in 1997 and 2020, transcribed into ELAN. Because of the small population, we were often able to record several members of the same family, and glean important information about their family's heritage and origins. We hypothesise here that if very little mixing and new dialect formation had taken place, we would be able to see evidence of specific settler dialect forms being passed down from one generation to the next within families, something we would not expect to see if there had been considerable mixing (Kerswill and Williams 2000).

We present our family-tree analysis of two linguistic variables: (a) intrusive /r/ - the insertion of /r/ between two vowels (that do not derive from the loss of rhoticity) that straddle a morpheme boundary – e.g. Argentina#and. Intrusive /r/ is known not to be a characteristic of rhotic Englishes, and so should be found less among those with Scottish and South-Western heritage ; (b) the diphthong MOUTH /au/, where we considered the effects of the following phonological environment and diphthong trajectory. This diphthong is known to be sensitive to following phonological environment in Scottish Englishes, which also have shorter trajectories for these diphthongs than varieties found in England to the south. Our family tree analysis shows both that speakers with, for example, Scottish heritage 'pass on' variant preferences to subsequent generations, even in some cases to their grandchildren, but also bear witness to intergenerational change. By exploring the sociolinguistic ecology of the Falklands during the formation phase of FIE, we attempt to explain why this is the case.

Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A. & Trudgill, P. (2004). *New Zealand English: Its Origins and Evolution*. Cambridge: CUP.

Kerswill, P. & Williams, A. (2000). Creating a new town koine. *Language in Society* 29: 65-115.

Trudgill, P. (1986). *Dialects in contact*. Oxford: Blackwell.

Measuring sound differences between Standard Dutch and regional languages using automatic quantitative procedures

Raoul Buurke, Remco Knooihuizen, Wilbert Heeringa, Martijn Wieling
University of Groningen

Abstract

In the presence of a prestigious standard language, regional language forms often become more similar to standard language forms (Hinskens et al., 2005). When analyzing phonetic transcriptions from dialects in the Netherlands in the 1980s (Taeldeman & Goeman, 1996), Wieling et al. (2011) found that Standard Dutch influence on dialects depended on their geographical distance from the economic center of the Netherlands (see Figure 1a). However, the analyzed corpus is known to be subject to transcriber effects due to the large number of transcribers creating the phonetic transcriptions (Buurke & Wieling, 2023; Hinskens & Van Oostendorp, 2006).

Instead of quantifying pronunciation differences using phonetic transcriptions (suffering from transcriber effects), recent large AI speech models can be leveraged to extract feature representations from sound recordings directly, after which a distance between these representations can be computed. For example, this approach has been shown to yield results matching human perceptual judgments (Bartelds et al., 2022) and has been found suitable to detect dialect areas in the Netherlands (Bartelds & Wieling, 2022). We follow the same procedure here to quantify pronunciation differences of dialectal variants compared to Standard Dutch based on a recently collected dialect dataset.

Sound recordings for 98 regional language users of Frisian and Low Saxon were made in 29 locations (see Figure 1b). The participants were presented with a 500- or shortened 150-item list of Standard Dutch target words and asked to translate each word into their local dialect. The 28,589 recorded dialectal translations were automatically segmented using an algorithm for voice detection (Bredin, 2023; Plaquet & Bredin, 2023). Recordings of the word list in Standard Dutch were made by a national news presenter, who was selected (by over 250 voters) out of 20 national and regional new presenters to represent Standard Dutch best. Distances between the acoustic recordings of the Standard Dutch target word and the dialectal translations were computed using the method of Bartelds and Wieling (2022).

We subsequently constructed a generalized additive mixed model (see Wieling, 2018) predicting the distance (between Standard Dutch and dialectal realizations) using a geographical smooth and an appropriate random-effects structure to account for by-speaker and by-word variation. The computed distances do not have a straightforward interpretation and are therefore scaled in the model (i.e., a value of 0 indicates the average distance). The results indicate that the geographical variation of Standard Dutch influence follows a predictable pattern outwards

from the economic center of the Netherlands (see Figure 1), which is in line with earlier findings (Kloeke, 1927; Wieling et al., 2011).

In conclusion, our results suggest that Standard Dutch influence can be measured fully automatically from sound recordings. The findings of Wieling et al. (2011) can be re-evaluated more directly when the 1980s recordings have been segmented, which is ongoing work. Moreover, this method can also be used to investigate local language change in these recording locations between the 1980s and 2020s. These results will be presented fully at the conference, together with details of the automatic procedure, which can be generalized and applied in contexts where single-word recordings need to be compared acoustically.

References

Bartelds, M., De Vries, W., Sanal, F., Richter, C., Liberman, M., & Wieling, M. (2022). Neural representations for modeling variation in speech. *Journal of Phonetics*, *92*, 101137. https://doi.org/10.1016/j.wocn.2022.101137

Bartelds, M., & Wieling, M. (2022). Quantifying Language Variation Acoustically with Few Resources. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 3735–3741. https://doi.org/10.18653/v1/2022.naacl-main.273

Bredin, H. (2023). Pyannote. audio 2.1 speaker diarization pipeline: Principle, benchmark, and recipe. *INTERSPEECH 2023*, 1983–1987.

Buurke, R. S. S. J., & Wieling, M. (2023). Sound Change Estimation in Netherlandic Regional Languages: Reducing Inter-Transcriber Variability in Dialect Corpora. *Taal en Tongval*, *75*(1), 7–28. https://doi.org/10.5117/TET2023.1.002.BUUR

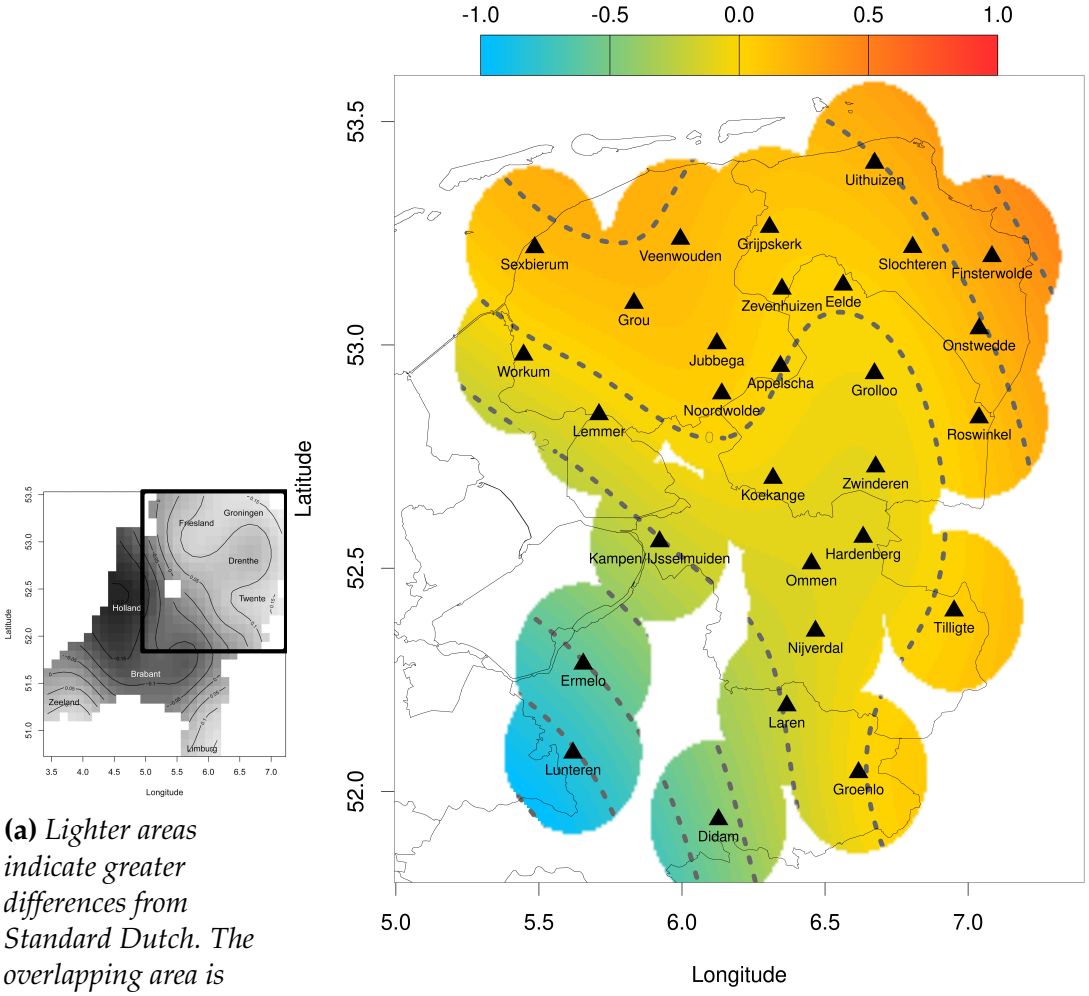
Hinskens, F., Auer, P., & Kerswill, P. (2005). The study of dialect convergence and divergence: Conceptual and methodological considerations. In P. Auer, F. Hinskens, & P. Kerswill (Eds.), *Dialect Change: Convergence and Divergence in European Languages* (pp. 1–48). Cambridge University Press. https://doi.org/10.1017/CBO9780511486623.003

Hinskens, F., & Van Oostendorp, M. (2006). De palatalisering en velarisering van coronale nasaal-plosief clusters in GTR. Talige, dialectgeografische en onderzoekerseffecten. *Taal en Tongval*, *58*, 103–122.

Kloeke, G. (1927). *De Hollandsche Expansie in de zestiende en zeventiende eeuw en haar weerspiegeling in de hedendaagsche Nederlandse dialecten*. Gravenhage: Martinus Nijhoff.

Plaquet, A., & Bredin, H. (2023). Powerset multi-class cross entropy loss for neural speaker diarization. *INTERSPEECH 2023*, 3222–3226. https://doi.org/10.21437/Interspeech.2023-205

Taeldeman, J., & Goeman, A. (1996). Fonologie en morfologie van de Nederlandse dialecten: Een nieuwe materiaalverzameling en twee nieuwe atlasprojecten. *Taal en Tongval*, *48*, 38–59.



(a) *Lighter areas indicate greater differences from Standard Dutch. The overlapping area is marked with a black border. Taken from Wieling et al. (2011).*

(b) *Redder areas indicate greater differences from Standard Dutch. The triangles indicate the recording locations in 2022/2023.*

Figure 1
Plots of geographical smooths of generalized additive mixed models predicting computed distances between Standard Dutch and recorded dialectal translations.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86–116. https://doi.org/10.1016/j.wocn.2018.03.002

Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative Social Dialectology: Explaining Linguistic Variation Geographically and Socially (M. Perc, Ed.). *PLoS ONE*, *6*(9), e23613. https://doi.org/10.1371/journal.pone.0023613

Investigating Multiword Constructions in Indian English and Australian Aboriginal English:

From Compositionality to Grammaticalization and Idiomaticity

***Rita Calabrese,* University of Salerno**

***Katherine E. Russo,* University of Naples "L'Orientale"**

The study analyses verbal MWCs in two varieties of English, namely Indian English and Aboriginal English, with the aim of investigating the development from compositional (e.g. *chase to*) to non-compositional constructions bearing figurative/idiomatic meaning (e.g. *father upon*) in post-colonial varieties of English (Brinton/Closs Traugott 2005; Rodriguez-Puente 2021; Calabrese, Chambers and Leitner 2015). Past literature on this topic has highlighted methodological complexities in the identification and the classification of such constructions especially when automatic annotation tools/procedures are adopted. As Baldwin and Kim (2010) note, explicit lexicalised MWC data contributes to simplify the accurate recognition of MWCs and influences the accuracy of semantic tagging, while lack of MWCs lexical items in a precision grammar may be a significant source of parse errors.

In this study, corpus-based methodology and linguistic diagnostics have been matched to 1) test the efficacy of automatic annotation tools for the detection of 'divergent' (or unmatched) lexico-grammatical and semantic variation with respect to reference corpora; 2) uncover potential differences and/or shared features between Indian and Aboriginal English in multiword verbal constructions. In order to test the two theoretical and methodological issues, the research draws on data from two annotated corpora, the Diachronic Corpus of Indian English (1,000,000 tokens) and the Diachronic Corpus of Aboriginal English (948,000 tokens), specifically compiled to represent different dimensions of linguistic variability over a period of about 150 years (1833–2013). The two corpora have been designed on the model of available multi-genre corpora like ICE-IND and ICE-AUS to provide a comparable configuration to the two corpora. As major changes in a language are assumed to come from spoken language, speech-related written genres such as witness depositions (containing direct speech) were included for the analysis (Culpeper/Kyto 2010) and then compared to selected proceedings from the Old Bailey Corpus dating back to the same time span. Other sections of the corpora were selected and compared to similar samples from the BNC and ARCHER corpora. The two corpora were automatically annotated using Visual Interactive Syntax Learning (VISL) (Bick 2001) and the UCREL Semantic Annotation System (USAS) (Rayson et al. 2004).

The VISL interface proved to be less accurate in identifying archaisms reflecting literal meanings and provided 'fallacious' annotations of unmatched constructions. Therefore, USAS was introduced to refine the semantic analysis of the data. The comparison of annotated data from different time periods suggests that the creation of new combinatory patterns is an overall phenomenon occurring regularly and steadily over time. It is remarkable that the overlap in Indian-specific and Aboriginal-specific occurrences emerges with greater frequency up until the year 1938. A possible explanation may be that in their first phases of evolution the two varieties followed a similar path, still linked to the common lexifier, but after that date they underwent a phase of nativisation (Schneider, 2007).

References

Baldwin, T. & Kim, S.N. (2010), Multiword Expressions. In Indurkhya, N. & Damerau, F. J. (Eds.), *Handbook of Natural Language Processing*, Chapman and Hall/CRC: New York, pp. 267-292.

Bick, Eckhard (2001), The VISL System: Research and applicative aspects of IT-based learning. In *Proceedings of the 13th Nordic Conference of Computational Linguistics (NODALIDA 2001)*. Uppsala University, Sweden.

Brinton, L., and Closs Traugott, E. (2005). *Lexicalization and Language Change*. Cambridge University Press: Cambridge.

Calabrese, R., Chambers J. and G. Leitner (2015). *Language Variation and Change in Post-colonial Contexts*. Cambridge Scholars Publishing, Newcastle Upon Tyne.

Rayson, P., Archer, D., Piao, S. and A. M. McEnery (2004). *The UCREL Semantic Analysis System.* In *Proceedings of the Beyond Named Entity Recognition Semantic Labelling for NLP Tasks Workshop*, Lisbon, Portugal, pp. 7-12.

Rodrìguez-Puente, P., (2021). *The English Phrasal Verb, 1650- Present. History, Stylistic Drift, and Lexicalisation*. Cambridge University Press: Cambridge.

Schneider, E.W. (2007). *Postcolonial English: Varieties Around the World*. Cambridge University Press: Cambridge.

The devil in the details: finding the Irish in 19th-century Australian English

Madeleine Clews
The University of Western Australia

The question of the influence of speakers from Ireland on the early formation of Australian English (AusE) is divisive. On one hand, Horvath (1985), Trudgill (1986; 2004), Troy (1991), and Hickey (2003), among others, assert that such AusE syntactic features as second-person plural pronoun *youse/yez* and utterance-final *but* are of Irish origin; on the other, corpora-based investigations of selected syntactic features by Burridge & Musgrave (2014) and Musgrave & Burridge (2023) find there is insufficient evidence of such features occurring in naturally generated text to enable them to be attributed to IrE. Ransom (1966) and Moore (2008) both comment on an apparent lack of Irish input to the early formation of AusE vocabulary, the former suggesting any such effects would have been mediated through earlier Irish migrations to England (1966: 51) and the latter citing heavy stigmatisation of the variety (2008: 90-91), while historians of Irish ethnicity give voice to frustration at perceived erasure of their heritage from the documentation of AusE—Lonergan (2003) suggesting that an Anglo-centric historical approach has ignored evidence of Irish language in the lexicon, and Whitaker (1998: 65) referring to 'lexicography as genocide'. Meanwhile, Trudgill (1986; 2004), Kiesling (2004) and Mitchell (1993) all refer to a dearth of investigation at the phonological level, Burridge & Musgrave (2014) concluding their analysis with a declaration that they would be 'delighted' to see exploration of possible IrE phonology in AusE. Accordingly, recent research (Clews, under review) using a small written corpus (100,000 words) transcribed from manuscript texts from below—ego documents written by two sisters born in the Colony of Western Australia in the mid-to-late 19th century—has found possible evidence of IrE phonological features.

This paper builds on these recent phonological findings, exploring the texts further to look at examples of possible morphosyntactic fingerprints of IrE such as *it*-clefting (Example 1) and conjunctive *only* (Example 2). These are discussed alongside other non-standard syntax associated with but not exclusive to IrE, such as syncretised past-participle/preterits (Example 3). In the spirit of Mitchell's (2003[1993]) exhortation for a historical study of AusE which engages deeply with cultural, historical, social, political and literary disciplines, these features are then considered alongside other texts, such as a memoir by the great-grandson of the last Fenian convict to remain in Western Australia (Talbot, 2005), located in the same region as the speakers of the original data, who recalls his grandparents using basic vocabulary from the local Indigenous language, Nyungar. This suggests contact between Irish migrant families and Aboriginal communities, possibly also manifest in parallel occurrences of putative IrE features in Australian Aboriginal English (Rodríguez Louro, Collard, & Clews 2024). Drawing a judicial analogy, a case is made here to lower the standard of proof from 'beyond reasonable doubt' to the 'balance of probabilities', where the cumulative weight of evidence readily falls in favour of recognising IrE in Australia's language ecology.

Examples

1) **its** the Road today **links up** with Caves Road (BK, F, b. 1875)
2) everything had gone out of the house **only** one table & one Chair (BK, F, b. 1875)
3) Mother and I **done** a little needlework and a lot of talking (EA, F, b. 1861)

References

Burridge, Kate & Musgrave, Simon (2014). It's Speaking Australian English We Are: Irish Features in Nineteenth Century Australia. *Australian Journal of Linguistics* 34(1): 24-49.

Clews, Madeleine (under review). Dialect formation signs in the Colony of Western Australia: a historical sociolinguistic study.

Hickey, Raymond (2003). Rectifying a standard deficiency: Second-person pronomial distinctions in varieties of English. In Taavitsainen, I. & Jucker, A. H. (Eds.), *Diachronic Perspectives on Address Term Systems.* Amsterdam: John Benjamins Publishing Company. 343-374.

Horvath, Barbara M. (1985). *Variation in Australian English: the Sociolects of Sydney.* Cambridge [Cambridgeshire]: Cambridge University Press.

Lonergan, Dymphna (2003). An Irish-centric View of Australian English. *Australian Journal of Linguistics* 23(2): 151-159.

Mitchell, A.G. (2003[1993]). The Story of Australian English: Users and Environment. *Australian Journal of Linguistics* 23(2): 111-128.

Moore, Bruce (2008). *Speaking our Language: the Story of Australian English.* South Melbourne, Vic: Oxford University Press.

Musgrave, Simon & Burridge, Kate (2023). Irish Influence on Australian English. In Hickey, R. (Ed.), *Oxford Handbook of Irish English.* Oxford: Oxford University Press.

Ransom, W.S. (1966). *Australian English: a Historical Study of the Vocabulary 1788-1898*. Canberra: Australian National University Press.

Rodríguez Louro, Celeste, Collard, Glenys Dale & Clews, Madeleine (2024). Australian Aboriginal English. *The New Cambridge History of the English Language.* Cambridge: Cambridge University Press.

Talbot, Len (2005). *Nannup: a Place to Stop and Rest.* Carlisle, Western Australia: Hesperian Press.

Troy, Jakelin (1991). 'Der mary this is fine cuntry is there is in the wourld': Irish-English and Irish in late eighteenth century Australia. In O'Brien, J. & Travers, P. (Eds.), *The Irish Emigrant Experience in Australia.* Dublin: Poolberg Press. 148-180.

Trudgill, Peter (1986). *Dialects in Contact.* Oxford: Basil Blackwell.

Trudgill, Peter (2004). *New-Dialect Formation The Inevitability of Colonial Englishes.* Edinburgh: Edinburgh University Press.

Whitaker, Ann-Maree (1998). Lexicography as genocide: the Irish influence on the Australian language. *Australian Celtic Journal* 6: 65-72.

Language Loss in Bengali: Conflict between Standard and Regional Dialects

Abahan Datta
abahanjiriya@gmail.com
India
Visva-Bharati

Bengali is the official language and the medium of education in the state (province) of West Bengal in India. However, five recognized dialects of Bengali are in use in the state. But as they have not found a space in the learning arena and the dialect of the state capital region of Kolkata and adjacent areas called Rarhi dialect is the sole medium of education, other dialects conflict with it. It could be an ideal diglossic situation if the government-sponsored standard dialect does not try to appropriate the local ones through official works and school education. This paper will investigate the conflict between a standard and the regional dialects and argue how language loss impacts Bengali in a dynamic sociolinguistic situation.

Research Question
My central question is about language conflict and loss, framed as: how does a powerful dialect obliterate the others from the periphery in the process of a collision? How hegemonic structure and power relations can impact a language system? There is a specific question too. How does a regional language, which is not the official language of a country and is under aggression by the official language called Hindi, is itself becoming an aggressor when promoted to the status of regional official language?

Data
One of my field studies was a village named Vidyadharpur, in the rural district of Birbhum, around 100 km from Kolkata. The survey showed that the elderly inhabitants are trying to preserve their local dialect, while the younger ones are more familiar with the standard one. The elderly people of the village hail from the Austro-Asiatic tribe called Santhal. They traditionally practice agriculture. The younger generation is also into local business, and that is the reason behind gradually shedding their linguistic habits. They have to connect through Lingua Franca Bengali in the marketplace while mixing with the majority ethnic Bengalis. While the society of old people is much more secluded due to their tie to the land. Empirical evidence is shown through phonological, morphological, and lexicon analysis.

A few examples follow:

1. Nasal intonation is prevalent in some words. For example, ʃʊkiɛ > s̃ʊkhiɛ (mean. dried up)
2. Article (as suffixes in Bengali) is used in different forms than the standard ones. For example, tin-ʈe > tin-ʈa (mean. 3 o'clock)
3. Some unknown words are used in the vocabulary, chiefly nouns. Example, pauʈ̪hi (stairs), mokka (corn), akha (oven)

These influences are traced to a tribal language of the region called Santhali. Its influence on the Jharkhandi dialect, mainly spoken in the western forest districts of the state, is very much evident. Jharkhandi dialect has been influenced by the tribal languages of the western region of the state, namely Mundari, Ho, and Kudmali, apart from Santhali. But as this particular village is populated with Santhals, they are influenced chiefly by their mother tongue Santhali, which is used apart from the Bengali dialect. Nasal intonation is one of those, and unknown words for standard Bengali is of Santhali. Though Santhali is now recognized as an additional official state language, its scope in

education and governmental work is very limited as the political structure doesn't provide it, so the Jharkhandi dialect is also not of high status.

Method

The method of this research is a typical dialectology one. This linguistic quantitative research method involves a field survey in the form of an interview, based on a questionnaire provided to the locals, and natural speech recording. Dialect boundaries, dialect areas, pure dialects, transition areas, relic areas, etc, linguistic data can be traced and analyzed from phonological, morphological, and lexical perspectives.

Result

The difference in language use between elders and younger reflects school education. Children educated in standard Bengali nowadays are embracing it as a symbol of status, and the ones going out of the village for business and jobs too, while the uneducated elders are sticking to their habits due to the prestige of identity. Conflict is making a loss for the non-standard dialects, where power makes the ultimate decision through the state.

What do Canadians think of their English? A national study of language attitudes towards Canadian English

Stefan Dollinger (UBC Vancouver)

Canadian English has been studied since the 1940s from a number of angles: lexis (e.g. Avis et al. 1967), phonetics (e.g. Boberg 2008), morphosyntax (Tagliamonte & D'Arcy 2007), pragmatics (e.g. Denis 2020) and perception (e.g. Nagy, Hoffman & Walker 2020). Paradoxically, language attitudes have been studied only to a rather limited degree. Warkentyne (1983), based on Gulden's MA thesis (1979), and Owens & Baker (1984), represent the bulk of language attitude studies on Canadian English so far and are rather dated. The question how Canadian residents think of their own variety of standard English is an important one that has repercussions on linguistic autonomy.

The present paper seeks to address this research gap using a questionnaire survey on language attitudes with closed and open answer questions. Consisting of 25 attitudinal questions and 15 background questions the survey elicited 3143 responses from all 13 provinces and territories in the fall of 2023, which allow an assessment of language attitudes nationally. Preliminary results reveal interesting insights. For instance, 94% consider it "cool" to speak more than one language while 61% consider "multilingualism (speaking more than one language) an important characteristic of the Canadian population". In the sample, 51% are themselves multilingual. Among these, 3% feel "extremely uncomfortable" and 12% "moderately uncomfortable" when speaking "non-English languages in public in Canada".

Besides this multilingual angle, the survey asks a great number of questions on Canadian English. For instance, just 50% (1,153) of respondents have "heard of 'Standard Canadian English'", while 35% have not and 15% are "not sure". Among the 50% that have heard about it, almost three quarters "can't describe it well", while 28% can. Among linguistic features, 24% in the sample consider Canadian spelling as very important, and 34% as important, while 70% answer that "Canadian university departments should encourage Canadian English spelling". Among associations with Canadian English, the attributes "polite" (33%), "(ice) hockey" (14%), "tolerant" (11%), "outdoorsy" (11%) and "sophisticated" (7%) are the most frequently selected ones. Finally, 68% consider "Canadian English a distinct kind of English (e.g. similar to American English being distinct from British English"), while 9% are undecided and 23% opposed. In terms of prestige, "British English" carries for 56% the most prestige, followed, although with distance, by Canadian English (5%) and American English (3%). Pragmatically, 81% of respondents are convinced that "Canadians say 'sorry' more often than Americans".

The data is used to gauge the degree of linguistic autonomy of Canadian English at the beginning of the 2020s. The findings will be compared to an American subsample, with earlier work (Gulden 1979, Dollinger 2019: Fig 39 from 2009), and the historical record (e.g. Hultin 1967, Chambers 1993) for a longitudinal assessment of the status of the variety in the eyes of its speakers. Results suggest that the multilingual component of the population takes on a special role in the maintenance of Canadian English autonomy, as do women and the more highly educated strata of society.

Avis, Walter S., Charles Crate, Patrick Drysdale, Douglas Leechman, Matthew H. Scargill and Charles J. Lovell (eds). 1967. *A Dictionary of Canadianisms on Historical Principles.* Toronto: Gage. Online at www.dchp.ca/dchp2

Boberg, Charles. 2008. Regional phonetic differentiation in Standard Canadian English. *Journal of English Linguistics* 36(2): 129-154.

Chambers, J. K. 1993. 'Lawless and vulgar innovations': Victorian views on Canadian English. In *Focus on Canada,* ed. by. Sandra Clarke, 1-26. Amsterdam: Benjamins.

Denis, Derek. 2020. How Canadian was eh? A baseline investigation of usage and ideology. *Canadian Journal of Linguistics* 65(4): 583-592.

Dollinger, Stefan. 2019. *Creating Canadian English.* Cambridge: Cambridge University Press.

Gulden, Brigitte K. 1979. *Attitudinal factors in Canadian English usage*. Victoria: University of Victoria, Department of Linguistics, MA thesis.

Hultin, Neil C. 1967. Canadian views of American English. *American Speech* 42: 243-260.

Nagy, Naomi, Michol F. Hoffman, James A. Walker. 2020. How do Torontonians hear ethnic identity? *TWPL* 42: 1-18.

Owens, Thompson W. and Paul M. Baker. 1984. Linguistic insecurity in Winnipeg: validation of a Canadian index of linguistic insecurity. *Language in Society* 13: 337-350.

Tagliamonte, Sali A. and Alexandra D'Arcy. 2007. The modals of obligation/necessity in Canadian perspective. *English World-Wide* 28(1):47–87.

Warkentyne, Henry J. 1983. Attitudes and language behavior. *Canadian Journal of Linguistics.* 28: 71-76.

Freakin swimming and everythink: Variable (ING) in youth Aboriginal English

Fraiese, Lucía; Collard, Glenys; Rodríguez Louro, Celeste; Gardner, Matt Hunt; Walker, James

Variable (ING) in English has been recognised for its long-term stability. While most scholarship has focused on mainstream (Cole, 2020; Forrest & Wolfram, 2019; Hazen, 2008; Horvath, 1985) and racialized Englishes (e.g., Kendall & Thomas, 2019; Van Hofwegen & Wolfram, 2010), less in known about its use in Indigenised Englishes (but see Ball & Bernhardt, 2008; Mulder, 1982; Sterzuk, 2003). (ING) variation has been argued to be a vernacular universal (Chambers, 2004), yet how it patterns in Indigenised Englishes is less well understood. In this paper, we consider Australian Aboriginal English (AE), a post-invasion contact-based variety of English spoken by 80% of First Nations people in Australia (Rodríguez Louro & Collard, 2021). While frequent use of the alveolar variant has been reported by Malcolm (2018), quantitative sociolinguistic studies of variable (ING) as used in AE are yet to be offered.

We draw on two original synchronic corpora of naturalistic talk-in-interaction spoken AE data collected in Nyungar *boodja*, southwest Western Australia (Fraiese, 2022-2023; Rodríguez Louro, 2018-2022). In line with the literature (e.g., Horvath, 1985; Kendall & Thomas, 2019; Travis, Grama & Purser, 2023; Walker, 2023), all tokens of variable (ING) (shown in examples 1-3) were extracted and coded impressionistically for their phonetic realisation (alveolar, velar, stopped voiced and stopped voiceless), and for grammatical category, phonological context, and syllable count. The social constraints included speaker year of birth, gender, and proximity of hometown to a major city. A preliminary sample of 13 AE speakers born between 2001 and 2010 (6 women and 7 men) reveals significant interspeaker variation, but the rates of use of the alveolar variant are overall higher than those found in other varieties of English in Australia (as reported in, e.g., Travis et al., 2023; Walker, 2023). Gender is a significant social predictor, with males leading in the use of the alveolar variant, and females showing an overall higher rate of stopping. Place of residence and grammatical category also interact with phonetic realisation, suggesting that those who reside in major cities have higher rates of the alveolar variant overall, and show a distinction between verbal and non-verbal forms. Those who hail from remote areas, however, do not display such a distinction, and show higher rates of velar realisations, which are still significantly lower than those found in other mainstream varieties of English (Walker, 2023). These results confirm observations by Malcolm (2018) of the presence of the alveolar variant in AE. The presence of stopping among females, although also attested elsewhere (Gordon, 1998), makes an interesting case for further exploration.

Examples

1. I got real annoyed when you mob *sayin* that. (Alessa, woman, b. 2010)
2. And then we was just um, like, I don't know, freakin swimming and *everythink.* (Frankie, woman, b. 2007)
3. This fellow was like *coming* close to us. (RWC, man, b. 2001)

References

Ball, Jessica & Bernhardt, B. May (2008). First Nations English dialects in Canada: Implications for speech-language pathology: Indigenous and Colonial Languages: Implications for Speech-Language Pathology Research and Practice. *Clinical linguistics & phonetics* 22(8): 570-588.

Chambers, Jack K. (2004). Dynamic typology and vernacular universals. In Kortmann, B. (Ed.), *Dialectology meets typology dialect grammar from a cross-linguistic perspective.* Berlin: Mouton de Gruyter.

Cole, Amanda (2020). Co-variation, style and social meaning: The implicational relationship between (h) and (ing) in Debden, Essex. *Language variation and change* 32(3): 349-371.

Forrest, Jon & Wolfram, Walt (2019). The Status Of (ING) in African American Language: A Quantitative Analysis of Social Factors and Internal Constraints. *American speech* 94(1): 72-90.

Fraiese, Lucía (2022-2023). The Boarders' Corpus of Australian Aboriginal English. Discipline of Linguistics: The University of Western Australia.

Gordon, Elizabeth (1998). Anythink or Nothink: A Lazy Variant or an Ancient Treasure? *New Zealand English journal* 12(12): 25-33.

Hazen, Kirk (2008). (ING): A vernacular baseline for English in Appalachia. *American speech* 83(2): 116-140.

Horvath, Barbara M. (1985). *Variation in Australian English: the sociolects of Sydney.* Cambridge [Cambridgeshire]: Cambridge University Press.

Kendall, Tyler S. & Thomas, Erik R. (2019). Variable (ING). In Thomas, E. R. (Ed.), *Mexican American English: Substrate Influence and the Birth of an Ethnolect.* Cambridge: Cambridge University Press. 171-197.

Malcolm, Ian (2018). *Australian Aboriginal English: Change and continuity in an adopted language.* Boston: De Gruyter.

Mulder, Jane (1982). The Tsimshian English dialect: the result of language interference. In Bartelt, G. H., Penfield-Jasper, S. & Hoffer, B. (Eds.), *Essays in Native American English.* San Antonio, TX: Trinity University Press. 95–112.

Rodríguez Louro, Celeste 2018-2022. Aboriginal English in the global city: Minorities and language change. DE170100493. Discovery Early Career Researcher Award, Australian Research Council.

Rodríguez Louro, Celeste & Collard, Glenys (2021). Australian Aboriginal English: Linguistic and sociolinguistic perspectives. *Language and Linguistics Compass* 15(5): n/a.

Sterzuk, Andrea (2003). *A study of Indigenous English speakers in the Standard English classroom*. ProQuest Dissertations Publishing.

Travis, Catherine E., Grama, James & Purser, Benjamin (2023). Stability and change in (ing): Ethnic and grammatical variation over time in Australian English. *English world-wide* 44(3): 435-469.

Van Hofwegen, Janneke & Wolfram, Walt (2010). Coming of age in African American English: A longitudinal study. *Journal of sociolinguistics* 14(4): 427-455.

Walker, James A. (2023). Soundin(g(k)) Greek in Melbourne. *Conference of the Australian Linguistic Society 2023, Language Variation and Change Australia VI.* The Universtiy of Sydney.

Identifying Dialectal Disparities in Major Speech Databases used in ASR and Linguistics Research

Simon Gonzalez, Tao Hoang, Jason Littlefield
simon.gonzalez@defence.gov.au; tao.hoang@defence.gov.au; jason.littlefield@defence.gov.au

With the emergence of Machine Learning and use of AI in language research, technologies are being developed to capture dialectal differences (**Gothi & Rao, 2023; Hirayama et al., 2015; Ma et al., 2023**). These technologies tend to use widely established common large databases across the research spectrum. This paper investigates the complex representation of speech dialects in 20 of these major databases widely utilized in Automatic Speech Recognition (ASR) tasks and linguistic research. The databases include Speech Accent Archive (**Weitzman & Major, 2005**), VoxCeleb (**Nagrani, 2017**), LibriSpeech (**Panayotov, 2015**), Common Voice (**Ardila, 2019**), and the Switchboard Corpus (**Godfrey, 1992**). These have emerged as pivotal resources, shaping the advancement of spoken language analysis. Our objective is to provide a nuanced sociolinguistic exploration, focusing on the dialectal distribution within these databases, and to shed light on potential pitfalls that may influence the outcomes of ASR models and linguistic research.

By scrutinizing the sociolinguistic nuances embedded in these datasets, our paper seeks to address critical questions. Are dialects adequately represented, or are there discernible gaps that may skew research outcomes? Are there differences that may favour certain dialects over others? We calculate the Shannon Entropy (**Shannon & Weaver, 1949**) for all databases to examine the balance of distributions of dialects/languages. Our examination also extends beyond the mere quantitative distribution of dialectal samples, delving into the qualitative aspects that contribute to a comprehensive understanding of linguistic diversity. For this, we look at world distribution, dialectal overlap across databases, and speech style used in the data collection.

This paper not only identifies potential pitfalls but also serves as a call to action for academics and researchers engaged in ASR and linguistic studies. As databases play a pivotal role in shaping research outcomes, acknowledging and rectifying flaws is paramount. The sociolinguistic lens through which we approach these databases enhances our ability to discern the subtleties and intricacies of spoken language, fostering a more representative approach in research endeavours. In conclusion, our paper provides valuable insights into the dialectal makeup of major speech databases, emphasizing the need for a more discerning and realistic representation of linguistic diversity.

References

1. Ardila, R., Branson, M., Davis, K., Flandrin, P., Fraser, G., Lee, A., & Weber, J. (2019). CommonVoice: A Massively-Multilingual Speech Corpus. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6530-6534). IEEE.

2. Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). SWITCHBOARD: Telephone Speech Corpus for Research and Development. In *ICASSP-92, 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 1, pp. 517-520). IEEE.

3. Gothi, R., & Rao, P. (2023). Improving Automatic Speech Recognition with Dialect-Specific Language Models. In: Karpov, A., Samudravijaya, K., Deepak, K.T., Hegde, R.M., Agrawal, S.S., Prasanna, S.R.M. (eds) *Speech and Computer. SPECOM 2023*.

4. Hirayama, N., Yoshino, K., Itoyama, K., Mori, S. & Okuno, H. (2015). Automatic Speech Recognition for Mixed Dialect Utterances by Mixing Dialect Language Models. In *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 23. 1-1. 10.1109/TASLP.2014.2387414.

5. Ma, X., Deng, C., Du, D., Pei, Q.: An enhanced method for dialect transcription via error-correcting thesaurus. In *IET Commun*. 17, 1984–1997 (2023). https://doi.org/10.1049/cmu2.12671

6. Nagrani, A., Albanie, S., & Zisserman, A. (2017). VoxCeleb: A Large-scale Speaker Identification Dataset. *arXiv preprint arXiv*:1706.08612.

7. Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). LibriSpeech: An ASR corpus based on public domain audio books. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5206-5210). IEEE.

8. Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois Press.

9. Weitzman, L., & Major, R. C. (2005). *The Speech Accent Archive: Towards a typology of English accents*. College of the Holy Cross, Worcester, MA, USA. Speech Accent Archive.

Disentangling style-shifting and life-span change: A panel study of FACE and 1st person possessive *me* in the North-East of England

James Grama, Isabelle Buchstaller, Anne-Marie Moelders & Lea Bauernfeind
Sociolinguistics Lab, University of Duisburg-Essen

Life-span change and style shifting do not occur independently (Buchstaller 2015; Wagner 2021). Instead, speakers modulate their linguistic behaviour according to considerations that are situated at the intersection of cohort and lifespan-specific marketplaces (see Rickford & Price 2013; Tetreault 2017; Riverain-Coutlee & Harrington to appear). Research suggests that stylistic presentation of self is contingent on a wealth of factors that interact with speaker age. These have been shown to include the conversational topic and the type of speech activity (Labov 1972; Rickford & McNair-Knox 1994), accommodation to interlocutors, cohort and community (e.g., Bell 1984; Tetreault 2017; Wagner 2021), as well as the speaker's orientation to the social meaning of linguistic resources (see papers in Eckert & Rickford 2002). A persistent challenge for panel research is how to disentangle bona fide intra-speaker change from more momentary stylistic shifts, due in large part to a lack of larger panel datasets that allow us to unpack stylistic effects.

This talk addresses the relationship between stylistic shifts and intra-speaker change by drawing on a unique panel corpus consisting of 17 speakers from the North-East of England, recorded first between the ages of 27-53 and again an average of seven years later. We investigate variability in two variables currently undergoing changes in progress, but differentiated by level of linguistic structure and socio-indexical meaning: the FACE vowel and the first-person possessive (1POS). To unpack the effect of style shifting, each recording includes a comparable range of socio-situational contexts. Stylistic variation is conceptualised along two different dimensions: attention paid to speech (an axis of casual-careful-reading; see Labov 2001) and addressee (following Bell's 1984 audience design model). Operationalising these measures allows us to develop a multifaceted model of style shifting, and to hone in on the relationship between synchronic variation and diachronic change.

Analyses reveal a complex relationship between ageing and style. First, the range of stylistic shifts across time is highly contingent on age group. A single cohort—the 30s age group—exhibits the greatest stylistic variation in both variables. FACE undergoes lifespan change; speakers follow along with the broader community change towards monophthongal [eː] between the two recordings, and 1POS undergoes retrograde change, moving away from increasingly common [maɪ] in favor of stigmatized [mi]. By contrast, slightly older speakers (in their 40s at timepoint 1) have developed relatively stable stylistic ranges with more limited shifts across stylistic axes. For the oldest age range (50+), stylistic shifts are only evident in stigmatized 1POS when talking with an unfamiliar interlocutor, the most formal of the contexts considered. These findings are discussed in terms of extant models of variation and change across the lifespan that rely on marketplace pressures (cf. Bourdieu & Boltanski 1975). Our analysis paints a differentiated picture of the relationship between momentary style shifts and more permanent intra-speaker change across the entire adult lifespan. By honing in on the relationship between variation (at one time point) and change (between time points), the present study informs models of intra-speaker linguistic lability (see Rickford 2021; Wagner 2021).

(499 words)

References

Bell, Alan. 1984. “Language Style as Audience Design”. In *Sociolinguistics: A Reader and Coursebook*, N. Coupland and A. Jaworski (eds). New York: St Martin’s Press Inc. 240–250.

Bourdieu, Pierre, & Boltanski, Luc. 1975. “Le fétichisme de la langue.” *Actes de la recherche en sciences sociales* 1(4):2–32.

Buchstaller, Isabelle. 2015. “Exploring linguistic malleability across the life span: Age-specific patterns in quotative use.” *Language in Society* 44(4):457–496.

Eckert, Penelope. 2019. “The individual in the semiotic landscape.” *Glossa: A Journal of General Linguistics* 4(1):14.

Eckert, Penelope, & Rickford, John (eds.). 2002. *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. 2001. The anatomy of style-shifting. In *Style and Sociolinguistic Variation*, P. Eckert and J. Rickford (eds). Cambridge: Cambridge University Press. 85–108.

Rickford, John, & McNair-Knox, Faye. 1994. “Addressee- and Topic-Influenced Style Shift: A Quantitative Socio-Linguistic Study”. In *Perspectives on Register: Situating Register Variation within Sociolinguistics*, D. Biber and E. Finegan (eds). Oxford: Oxford University Press. 235–276.

Rickford, John, & Price, Mackenzie. 2013. “Girlz II women: Age-grading, language change, and stylistic variation.” *Journal of Sociolinguistics* 17(2):143–179.

Rickford, John. 2021. “Stylistic Variation in Panel Studies of Language Change”. *Karen V. Beaman, Isabelle Buchstaller (eds.)*, Language Variation and Language Change Across the Lifespan Theoretical and Empirical Perspectives from Panel Studies. New York and London: Routledge, 77-97.

Tetreault, Chantal. 2017. “Ethnographic Perspectives on Panel Studies and Longitudinal Research.” In *Panel Studies of Variation and Change*, S.E. Wagner and I. Buchstaller (eds). New York: Routledge. 235–255.

Wagner, Suzanne Evans. 2021. “What’s the point of panel studies?” In *Language Variation and Language Change Across the Lifespan: Theoretical and Empirical Perspectives from Panel Studies*, K. Beaman and I. Buchstaller (eds). New York and London: Routledge. 267–280.

Relationships among Japanese local dialectal words by Levenshtein distance and multivariate analysis

Wilbert Heeringa (Fryske Academy) & Fumio Inoue (Tokyo University of Foreign Studies)

The Linguistic Atlas of Japan Database (LAJDB) is a huge data set including transcriptions of no fewer than 2,400 local Japanese dialects and maximally 37 items per location. Using these data we will answer the following questions:

1. What are the relationships among the 2,400 varieties, and what areas do they represent?
2. What are the relationships among the 37 items, and what dimensions do they represent?

Using led-a.org we applied PMI Levenshtein distance (Wieling 2012) to 2400 localities and 37 items of the LAJDB. Ward's clustering and t-distributed stochastic neighbor embedding (t-SNE) were applied to the distances that we obtained among the 2400 localities. With Ward's clustering we found five natural groups which we projected in a map. Using 3-dimensional t-SNE we created a continuum map showing a dialect continuum. The results that we obtained are plausible and consistent with the preceding studies such as those from Fujiwara (1990), Inoue (2001) and others. They reinforce, rather than contradict, the results of previous Japanese dialectologists (Heeringa & Inoue 2023).

Dialectometry usually focuses on geography, i.e. the dimension of local dialects and their inter-relationships. However, it is also possible to analyze the item dimension and their inter-relationships. The two dimensions – dialect variety and item - are two sides of the same coin.

When analyzing the item dimension we followed Nerbonne (2006) who applied factor analysis in order to identify the linguistic structures that are represented by the average Levenshtein distances. We found two main factors. In order to understand the meaning of the factors, we applied Ward's clustering to the loadings of the factors, which divided the 37 items in three sets:

- a set of 14 items that in particular loaded on the first factor;
- a set of 11 items that in particular loaded on the second factor;
- a set of 12 items loading on neither factor:

Subsequently for each of the three sets of items we measured Levenshtein distances on the basis of the items in that set. When applying Ward's clustering to the distances that were obtained for each set, we found the following results:

- items with high loadings on factor 1 represent a tripartite division: north, central and the southern islands;
- items with high loadings on factor 2 represent a dichotomy with a border in the south of the mainland;
- items close to the origin represent a dichotomy with a border in the center of the mainland.

When superimposing the three divisions, a map is obtained that is almost identical to the map with five areas that we initially obtained on the basis of the full set of 37 items.

References

Heeringa, Wilbert & Fumio Inoue (2023), Exploring the Japanese Dialect Geography Dialectometrically: Division and Continuity. *Studies in Geolinguistics* 3, 1-44.

Fujiwara, Yoichi (1990) *Nihongo Hogen Bunpa ron* [Dialect Propagation Theory of Japanese] Tokyo: Musashino Shoin.

Inoue, Fumio (2001) *Keiryoteki Hogen Kukaku* [Quantificational Dialect Classification] Tokyo: Meiji Shoin.

Nerbonne, J. (2006), Identifying Linguistic Structure in Aggregate. In: J. Nerbonne and W. Kretzschmar, Jr. (eds.), Progress in Dialectometry, special issue of Literary and Linguistic Computing, selected proceedings of a workshop at Methods in Dialectology XII, Moncton, Aug. 5, 2005.

Nerbonne, J. (2006). Identifying linguistic structure in aggregate comparison. *Literary and Linguistic Computing*, 21(4), 463-475, https://doi.org/10.1093/llc/fql041.

Wieling, Martijn (2012). A quantitative approach to social and geographical dialect variation. Doctoral dissertation, University of Groningen. URI: https://hdl.handle.net/11370/cd637817-572f-4826-98c1-08272775fb64

A longitudinal study of consonant cluster reduction among British and American expatriates in Japan

Keiko Hirano
The University of Kitakyushu, Japan

This paper reports on linguistic changes among British and American sojourners and expatriates in Japan by comparing three sets of corpus data collected at three time points. The consonant cluster reduction or /-t,d/ deletion (Amos, Kasstan & Johnson 2020; Guy 1980; Guy & Boyd 1990; Labov 1989; Neu 1980; Roberts 1997; Schreier 2005; Tagliamonte and Temple 2005) was chosen as a linguistic token to test the changes in usage by British and American speakers. A gradual shift away from the /-t,d/ deletion is demonstrated in the presentation. This longitudinal study is based on three sets of linguistic data collected from 40 native speakers of English living in Japan: (1) a corpus of English conversations collected in 2000 from young British and American English speakers who had recently arrived in Japan [Data1], (2) a corpus of spontaneous conversations collected in 2001 from the same speakers after one year [Data2], and (3) a corpus of natural conversations collected in 2018–2019 from Britons and Americans who had lived and worked in Japan for an average of 19 years [Data3]. About 8,000 tokens were collected from three datasets.

The analysis of /-t,d/ deletion observed in the three datasets of British English speakers shows that they steadily reduce their use of /-t,d/ deletion from Data1 to Data2 and from Data2 to Data3 (63%–58%–43%). American English speakers, similar to British speakers, decrease their /-t,d/ deletion from Data1 to Data2 to Data3 (67%–65%–56%). Why do these speakers continue to move away from the /-t,d/ deletion during their stays in Japan? Why do these expatriates who have lived in Japan for a long time (Data3) use much lower rates of /-t,d/ deletion than those who just arrived in Japan?

Using Data1 and Data2 mentioned above, Hirano (2013) found systematic decreases in the rates of /-t,d/ deletion in the English spoken by young native speakers from Britain, the United States and New Zealand who arrived in Japan to teach English. Social network theory (Milroy 1980) was used to analyse differences in individual rates of /-t,d/ deletion, and multiple regression analysis confirmed an inverse correlation between deletion and social networks with non-native English speakers. The evidence strongly suggests that this change was induced by frequent contact with Japanese Anglophones.

Frequently encountering non-native English speakers in Japan—mostly Japanese people—and being constantly surrounded by these people in their daily lives for a long time is likeliest to influence how they speak English. Consequently, expatriate speakers tend to reduce the use of /-t,d/ deletion in rapid speech phenomena over time and reflect this change even in conversation with another native speaker of English. The presentation will also report the results of the analysis of /-t,d/ deletion concerning linguistic constraints, such as grammatical status and the following phonological environment.

REFERENCES

Amos, Jenny, Kasstan, Jonathan R., and Johnson, Wyn. (2020). Reconsidering the variable context: A phonological argument for (t) and (d) deletion. *English Today* 36(3):6–13.

Guy, Gregory R. (1980). Variation in the group and the individual: The case of final stop deletion. In W. Labov (ed.), *Locating language in time and space*. New York: Academic. 1–36.

Guy, Gregory R., and Boyd, Sally. (1990). The development of a morphological class. *Language Variation and Change* 2: 1–18.

Hirano, Keiko. (2013). *Dialect contact and social networks: Language change in an Anglophone community in Japan*. Frankfurt: Peter Lang.

Labov, William. (1989). The child as linguistic historian. *Language Variation and Change* 1: 85–97.

Milroy, Lesley. (1980). *Language and social networks*. Oxford: Blackwell.

Neu, Helene. (1980). Ranking of constraints on /t, d/ deletion in American English: A statistical analysis. In W. Labov (ed.), *Locating language in time and space*. New York: Academic. 37–54.

Roberts, Julie. (1997). Acquisition of variable rules: A study of (-t, d) deletion in preschool children. *Journal of Child Language* 24: 351–372.

Schreier, Daniel. (2005). *Consonant change in English worldwide: Synchrony meets diachrony*. Basingstoke: Palgrave Macmillan.

Tagliamonte, Sali, and Temple, Rosalind. (2005). New perspectives on an ol' variable: (t,d) in British English. *Language Variation and Change* 17: 281–302.

Self-perceptions of Ethnic Markers in Toronto English

Michol F. Hoffman, York University and James A. Walker, La Trobe University

Changing patterns of immigration have increased the ethnolinguistic diversity of Canada's largest cities, leading to long-term maintenance of heritage languages and the development of ethnically marked ways of speaking the majority language. Members of ethnolinguistic minority groups who grow up in 'ethnic enclaves' tend to have higher degrees of orientation to their background and may be more likely to use speech features associated with it. Research on the sociolinguistic consequences of urban ethnolinguistic diversity has been conducted (e.g., Authors 2010, Boberg 2014, Denis et al. 2023) but, while anecdotes and public discourse point to awareness of ethnically marked speech, there is little systematic research on the perceptions of speech of individual ethnolinguistic groups. Perceptual studies of ethnolinguistic groups outside of Canada have largely focused on "multiethnolects" (e.g., Kircher and Fox 2019), or on well-established varieties (e.g., Purnell et al.1999).

As part of a larger project on ethnically marked ways of speaking in Toronto, Canada's largest and most ethnolinguistically diverse city, we have conducted perceptual studies drawing on methods from previous work (e.g., Campbell-Kibler 2009, Levon 2014). We tested listener ability to identify the ethnic background of native speakers of Toronto English from five of the largest ethnic groups (British/Irish, Chinese, Italian, Portuguese and Punjabi), stratified by sex and degree of ethnic orientation. Those results confirmed listener awareness of ethnically marked ways of speaking and greater ability to identify speakers who identify more strongly with their ethnicity. Some speakers and ethnic backgrounds were more salient than others. Here, we use direct and indirect methods to explore in more detail the specific linguistic features that may index ethnicity.

We conducted focus groups with second generation Chinese and Italian Torontonians, born and raised in Toronto. The sessions, led by community members, included questions designed to elicit specific features and ways of speaking English associated with Chinese and Italian Torontonians. Consultants listened to recorded speech clips from their own and another ethnic group and were asked to identify linguistic features marking speakers' ethnicities. They also considered potential transfer features from Italian and Cantonese chosen in consultation with community members and based on previous analyses. Finally, they were asked about variables in Toronto English (changes in progress and stable variation) to gauge their awareness of variation and potential associations with particular ethnicities. The sessions also included general discussions of ways and contexts in which Chinese and Italian Torontonians may mark their identities through language.

An analysis of session transcripts reveals that in-group members can identify some specific features marking Cantonese (final consonant devoicing, /θ/and /ð/ stopping and fronting, final /l/ vocalization, and prosodic features including intonation) and Italian (the articulation of /e/ and /o/ with no off-glide). Further, participants associate some variables (e.g., Canadian Raising) with British-Irish Torontonians, some (t/d deletion) with other ethnicities, and report lack of awareness of others (e.g., Canadian Vowel Shift). These findings complement those based on studies of linguistic production to deepen our

understanding of the sociolinguistic consequences of ethnolinguistic diversity on social identity and group interactions.

References

Authors. (2010). Ethnolects and the City: Ethnolinguistic variation in Toronto English. *Language Variation and Change* 22: 37-67.

Author1. (2017) "In the front and in the back": The role of ethnicity in back vowel fronting in Toronto English. presented at the Annual Meeting of the American Dialect Society, Austin Texas.

Boberg, C. (2014). Ethnic divergence in Montreal English. Canadian Journal of Linguistics, 59(1), 55–82.

Campbell-Kibler, K. (2009) The nature of sociolinguistic perception. *Language Variation and Change* 21: 135–56.

Denis, D., Elango, V., Kamal, N. S. N., Prashar, S., & Velasco, M. (2023). Exploring the Vowel Space of Multicultural Toronto English. Journal of English Linguistics, 51(1), 30–65.

Kircher, R., & Fox, S. (2019). Attitudes towards multicultural London English: Implications for attitude theory and language planning. *Journal of Multilingual and Multicultural Development, 40*(10), 847-864. doi:http://dx.doi.org.ezproxy.library.yorku.ca/10.1080/01434632.2019.1577869

Levon, E. (2014) Categories, stereotypes, and the linguistic perception of sexuality. *Language in Society* 43: 539–566.

Purnell, T., Idsardi, W. & Baugh, J. 1999. Perceptual and phonetic experiments on American English dialect identification. Journal of Language and Social Psychology, 18(1), 10-30.

Modelling Time and Space in the Emergence of New Dialects

Xia Hua[1], Luis Miguel Rojas-Berscia[2], Lindell Bromham[1], Felicity Meakins[3]

Australian National University[1], Radboud University[2], University of Queensland[3]

Variation is a fundamental ingredient in all evolutionary systems, including language. All fields of linguistics which deal with language evolution from population-level variation and change to linguistic diversification (e.g., sociolinguistics, creolistics and historical linguistics) have the (relative) use and non-use of language features at their core (e.g., Croft, 2000; Labov, 1994a; Labov, 1994b; Mufwene, 1990).

Modelling the relationship between micro-level processes (i.e., variation among individuals in a population) and macro-level outcomes (language diversity) has proven difficult but is key to understanding how new languages emerge. In order to link the micro and macro, we need to characterise the way that variation between speakers leads to differences between languages. This can be done by mapping clusters of co-varying language features in individuals (idiolects) to regionally distinct varieties of language use (dialects).

Investigating clusters of language variants which pattern according to social categories (i.e., linguistic coherence) across generations of speakers contributes to an understanding of the processes of social differentiation and language change at a local speaker community-level (e.g., Guy, 2013; Labov, 1966, 1972). But there are also potential broader implications for theories of language diversification and evolution. Theories of language evolution are particularly concerned with linguistic variation along the dimensions of time and geography. Variants associated with different generations often are indicative of change over time; and variants associated with different regions are often associated with dialects. The combination of space (region) and time (generations) can contribute to an understanding of how new forms of language emerge (Trudgill, 1983).

In this paper, we introduce a new dataset containing 20 variables (57 variants) used by 168 Shawi speakers of different ages from three regions of Peru. Crucially, this dataset has the necessary time and regional dimensions required to document the emergence of a new dialect of Shawi. We use this dataset to show how the flow of clusters of linguistic variants used by individual speakers has led to the emergence of the Balsapuerto dialect.

References:

Croft, William. (2000). *Explaining language change: An evolutionary approach*. Harlow, England: Longman.

Guy, Gregory. (2013). The cognitive coherence of sociolects: How do speakers handle multiple sociolinguistic variables. *Journal of Pragmatics, 52*, 63-71.

Labov, William. (1966). *Social Stratification of English in New York City*. Washington: Center for Applied Linguistics.

Labov, William. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. (1994a). *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell.

Labov, William. (1994b). *Principles of Linguistic Change: Social factors*. Oxford: Blackwell.

Mufwene, Salikoko. (1990). Transfer and the substrate hypothesis in creolistics. *Studies in Second Language Acquisition, 12*(1), 1-23. doi:10.1017/S0272263100008718

Trudgill, Peter. (1983). *On Dialect: Social and Geographical Perspectives*. Oxford: Blackwell.

New Dialect forms over the 250 Years Since “Hamaogi”
---Geographic and age trends based on multiple correspondence analysis---

Fumio Inoue (Tokyo University of Foreign Studies) & Yasushi Hanzawa (Fukushima University)

Research question: This paper utilizes a lexical tracking study based on a dialect glossary compiled approximately 250 years ago to ask what the trend of dialect is. Specifically, it asks which forms are new dialect forms and what the distributional trend among generations and geographic region was.

Data: Surveys were conducted in 1950 and 2018 on 3 and 4 generations, respectively, covering a total of 7 generations with a maximum age difference of 140 years. The residual rates of the word forms recorded in the *Hamaogi* glossary in the first part of the questionnaire (Yes / No form) have been analyzed in previous papers, where multivariate analysis was applied to analyze age and regional differences (Inoue and Hanzawa 2021). In this paper, we aim to gain an overall understanding of the second part of the questionnaire (choice of actual dialect forms + free entry).

Methods: Multivariate analysis was applied to determine the overall tendency of vocabulary usage. The “Hamaogi word forms” and “standard word forms” were identified by established criteria. For the third category, “other word forms,” the usage rates for each of the seven generations were calculated, and the word forms that were on the rise in the three adult generations of the second survey (generations 4, 5, and 6) were extracted to set up "new dialect" as the fourth category. Line graphs showing usage rates of the four categories for all the generations were useful as clues for analysis.

Results: The results of the multivariate analysis were utilized to make scatter plots to explain the results. As actual examples, we applied the method to new dialect forms, and observed the emergence and spread of new dialects since pre-modern times, their power struggle with other word forms, and their competition with the standard language especially after WWII. The data clearly showed that new dialect forms appeared and spread repeatedly. New dialect is created as a perpetual and perennial process. We also found non-use of new dialect among the seventh generation (middle and high school students). Based on these results, we attempted to develop a historical and theoretical perspective, generalize dialectal change after the modern era, and look at examples of multiple shifts, where new dialect forms take over several times. Applying multivariate analysis was useful for simplifying and summarizing multiple complex phenomena as a whole.

References

Inoue, Fumio and Yasushi Hanzawa (2021) Multivariate analysis of geography and age in dialect vocabulary --- Comprehensive analysis of 250 years of language change --- *Dialectologia: revista electrònica* 27: 97-160.

Psycholinguistic evidence for the effect of phonemic orthography: Is it really that hard to process?

Aya Inoue

Aichi University of the Arts / University of Hawai‘i at Mānoa

Amy J. Schafer

University of Hawai‘i at Mānoa

In writing dialects and creole languages without enforced official orthographies, the spelling conventions of the standard variety or lexifier languages (etymological orthographies) are often used, although these varieties are sometimes very different in terms of the phonological and/or grammatical forms from the standard varieties or lexifiers. Although etymological orthographies look accessible and readable for speakers of standard or lexifier languages, the spelling patterns are often inconsistent depending on the writers' choices and preferences. Furthermore, it is not easy for readers to tell whether those inconsistencies are due to the different writers' choices or simply representing variable phonological and grammatical forms of such varieties themselves.

Yet the other option of orthography, the phonology-based orthographies (phonemic orthographies) are hardly used simply because they look "alien" and "intimidating" (Romaine 2005). It may be true that phonemic orthographies are hard to read if dialect/creole speakers have no previous explanations or practice with them. But what if they received a certain amount of training about the phonemic orthographies?

Such a question is not easy to answer within the traditional methodologies of sociolinguistics because sociolinguistic evidence is typically based on observation and elicitation of naturally occurring language use. Psycholinguistic experiments, however, are effective when we want to see the hypothetical situation: in this case, the readability of phonemic orthographies after readers receive certain training. As a case study, we demonstrate how psycholinguistic evidence can contribute to the discussion of orthographic choice regarding Hawai'i Creole based on the experiment with 87 participants.

Hawai'i Creole (HC) has no official orthography, but two spelling systems have been used in different circles. In creative writing, such as poetry, short stories, and novels, local

voice is often described using Standard American English (SAE) spelling rules (Etymological HC Orthography). A phonemic spelling system for HC also exists (Odo HC Orthography (Odo 1975)), and is mostly used in linguistics.

The experiment tests Etymological vs. Odo Orthography, with bilingual/bidialectal (HC-SAE) speakers of English raised in Hawai'i and monolingual (SAE) speakers of English from the mainland US. The results suggest that expected differences between words that are easier or harder to read emerge in both orthographies. Moreover, in contrast to the popular belief that phonemic orthographies are more difficult to process and learn than etymological orthographies, after merely a short explanation of the phonemic orthography, native speakers of creole from Hawai‘i and newly exposed readers from the U.S. mainland exhibited almost perfect accuracy on comprehension questions for sentences in both creole orthographies. These results suggest that continued practice and exposure to the phonemic Odo Orthography could facilitate reading processing in HC speakers. Arguably, psycholinguistic effects should also be included in the consideration of selecting orthography for creole languages in addition to the context of language policy and language ideology (Schieffelin and Doucet 1998)).

[458 words]

Selected references

Odo, Carol. 1975. Phonological processes in the English dialect of Hawaii. University of Hawai‘i Ph.D. dissertation.

Schieffelin, Bambi B., and Rachelle Charlier Doucet. 1998. The “real” Haitian Creole: Ideology, metalinguistics, and orthographic choice. Language ideologies: Practice and theory, ed. by Bambi B. Schieffelin, Kathryn A. Woolard, and Paul V. Kroskrity, 285–316. New York: Oxford University Press.

Romaine, Suzanne. (2005). Orthographic practices in the standardization of pidgins and creoles: Pidgin in Hawai’i as anti-language and anti-standard. Journal of Pidgin and Creole Languages 20:1. 101-140

Siegel, Jeff. 2005. Literacy in pidgin and creole languages. Current Issues in Language Planning 6 (2): 143–163.

The semantic change of intensifiers observed through geographical variation in the Goto Archipelago, Kyushu, Japan

Takateru Kadoya (Fukuoka Jo Gakuin University)
Soichiro Harada (Nagasaki University)
Hitoshi Nikaido (Fukuoka Jo Gakuin University)
Keiko Maeda (Nagasaki University)

Intensifiers, such as "*very*" in English and "*totemo*" in Japanese, are highly frequently observed in languages throughout the world. Numerous researches have been conducted on the development of intensifiers (De Clerck and Colleman 2013 among many others), and there are cases where clear identifications of their etymologies are included. It is not hard to imagine the existence of intensifiers derived from words meaning "big" or "many," as seen in examples like "*greatly*" in English, and there are studies, such as Zaborski (2006), that point out this development. However, there seems to be a significant gap between content words representing meanings like "big" and "many" and intensifiers that have become grammaticalized with their original meanings bleached. Therefore, this study aims to describe the developmental process of intensifiers by observing geographical variation.

The focus of the investigation is the dialects of several villages on some islands of the Goto Archipelago, located west of Kyushu, Japan. These dialects exhibit significant generational differences, with younger generations strongly influenced by standard Japanese. Consequently, elicitation sessions were conducted targeting individuals aged 60 and above in this study.

The dialects of the Goto Archipelago share intensifying adverbs derived from the root "*zama*" whose original meaning seemingly is "big" or "many." Hence, the study adopts a method of assuming the process of semantic change by observing the geographical variation in these intensifiers.

As a result of the examination of the geographical variants, it was found out that the usage of each intensifying adverb significantly differs among dialects. For example, in one dialect, the original meanings ("big", "many") have been completely bleached, allowing them to modify adjectives meaning "small" and establishing itself as full-fledged intensifier in the dialect. On the other hand, in other dialects, there are constraints based on the original meanings, such as being unable to modify words meaning "small" or words with emotional connotations (e.g., "happy"). This study illustrates a process where an adverb initially indicating "physical magnitude" evolves to express "emotional magnitude," subsequently losing its original meaning and becoming capable of conveying a high degree regardless of the direction of scale

("big" or "small"). This is because of the existence of dialects where the intensifier can modify "emotional magnitude" but cannot modify "physical smallness."

While the evolution of intensifiers' meanings has been studied through classical literatures, corpora, and sociolinguistic generational differences (Parkington 1993, Macaulay 2006, among others), the current study demonstrates that geographical variation observation can contribute to similar research. At least, such studies on dialects in Japan have been limited as far as our knowledge goes, and this research seems to show one approach to the study of dialects where significant variations in intensifier are easily found.

References

De Clerck, Bernard and Timothy Colleman. 2013. From noun to intensifier: *massa* and *massa's* in Flemish varieties of Dutch. Language Sciences 36, 147-160.

Macaulay, Ronald. 2006. Pure grammaticalizaion: The development of a teenage intensifier. *Language Variation and Change* 18-3, 267-283.

Parkington, Allan. 1993. Corpus Evidence of Language Change — The Case of the Intensifier. In M. Baker et al. (eds.) Text and Technology: In honour of John Sinclair, 177-192.

Zaborski, Andrej. 2006. Intensifier 'Very' in Some Afroasiatic Languages. In A. Krasnowlska et al. (eds.) *In the Orient Where the Gracious Light – Satura orientalis in honorem Andrzej Pisowicz*, 229-232.

Decolonising dialectology? Viewing language diversity from the inside

Dr Raymond Kelly
Dr Jean Harkins
Mr Nicholas Watson

Dialectologists have explored the wealth of language variation that lies between and beyond the standardised language constructs of nation-states, discovering realities about language that exist socially over mappable geographic areas, and across more or less identifiable social groups. Decades of descriptive and historical linguistic research into Australia's First Languages have revealed much about probable historical and dialectal relationships among hundreds of language varieties that flourished before the arrival of colonisers (Dixon 1980, 2002; Bowern & Koch 2004; McGregor 2008).

In this paper we are concerned with how this body of knowledge might be enriched by viewing this multilingual, multidialectal landscape through First Nations eyes. The questions about these languages – their history, spread, relatedness, typological features, contact, change, and so on – that seem obvious when examining them from an outside perspective may not be the same questions as their speakers and inheritors have about them. Speakers and inheritors of First Languages today continue to explore and debate socially and culturally vital questions of how linguistic features and choices intersect with geography, kinship, cultural and personal identities, and historical and contemporary intergroup relationships – and how the available documentation of their languages connects with people's direct experiences of language use in families and communities.

As a Dhanggati man who began acquiring his ancestral language in early childhood, Ray Kelly reconnected with it as an adult after many years of English-dominated education and work. The work of descriptive linguists such as Holmer (1966, 1967, and associated audio recordings and notes), Holmer & Holmer (1969) and Lissarrague (2007, 2020) assisted greatly in this. But to fully reclaim an embodied competence commensurate with his obligations as a custodian and transmitter of cultural heritage required more than that. Kelly found that the journey back into his language led into a complex multilingual and multidialectal web of interconnecting language varieties and cultural meanings (Kelly 2015). He continues to seek practical means to understand and disseminate this wealth of culturally embedded language knowledge in ways that are helpful to fellow language custodians and inheritors.

When Australian First Nations people speak of 'Language', often with a capital 'L', they often mean more than just one particular Australian indigenous language or dialect (named or unnamed). The cultural meaning of 'Language' extends to an embodied competence across a linguistic repertoire often encompassing several Indigenous languages and dialects; it may include one or more immigrant languages, as well as a range of language styles or registers required for one's life activities. Harkins (1994) has suggested how this kind of multilingual competence allows speakers to deploy complex variations in response to each communicative context. We will say more about our emerging Indigenous-led understanding of the construct of 'Language', and how it may relate to established concepts of languages, dialects, and what has more recently been termed 'translanguaging' (Wei 2018).

Internal and external motivations for koiné formation:
Transplanted Brazilian Portuguese in Japan

Kazuko Matsumoto (University of Tokyo)
Akiko Okumura (Kanda University of International Studies)
Kenjiro Matsuda (Kobe Shoin Women's University)

This paper explores a newly emerging Brazilian Portuguese (BP) koiné in the Greater Tokyo Area of Japan. In the 1980s, Japan's labour shortage prompted a significant shift in immigration policy, allowing *Nikkei* (Japanese descendants) from Latin America to work in the country. This change led to the emergence of largely Brazilian-origin Latin American communities throughout Japan. Despite extensive research on this population in various fields over the past three decades, there is still a research gap in understanding language contact, dialect contact, and their linguistic consequences, within this immigrant context. To address this gap, we conducted a variationist study on BP in Joso City, Ibaraki Prefecture, within the Greater Tokyo Area.

In order to examine new dialect formation in a Latin American immigrant setting in Japan where language contact is also involved (i.e., of Japanese, the mainstream language in the host society, and Peruvian Spanish, an adstrate language spoken by Peruvian Spanish–BP bilingual community members), this paper incorporates a language contact perspective into models of koineization (Trudgill 1986; Britain 2018) by employing Mufwene's (2001, 2008) concepts of the 'feature pool' and 'xenolectal features'.

We focus on Strong-R (onset /r/) realizations (i.e., [r, x, χ, h, ɾ, ɸ]) by 79 speakers (first- and second-generation immigrants). In Brazil, Strong-R is, diachronically, a feature undergoing in-progress lenition ([r] → [ʀ] → [x]/[χ] → [h]; Rennicke 2015), while synchronically, several different forms ([r, x, h, ɾ]) show different distributions by speakers' geographical upbringing. In the immigrant context in Japan, Strong-R has the potential to be influenced by Japanese [ɾ, ɸ] and Peruvian Spanish [r].

In order to verify an ongoing change in Japanese BP and identify the combined effect of the predictors for this change, we conduct generalized linear mixed-effect models. Generation, gender, dialect background of speakers and their parents, mobility within Japan, frequency of return journeys to South America, position, tonicity and following vowel are entered as fixed effects, with speaker and word as random intercepts.

Generation is used to evaluate ongoing language change within an apparent-time construct (Bailey *et al*., 1991).

The results highlight (a) levelling and focussing towards [h] as a result of koineization and (b) early stages of the adoption of [ɸ], a xenolectal feature resulting from contact with Japanese. The external and internal motivations for the change to [h] are identified to be *local* and *supralocal* levelling and drift. The transition to [ɸ], and its linguistic and social embedding, are discussed in terms of acquisition orders, the structure of the Japanese *kana* syllabary, and speakers' social networks.

This paper demonstrates that in order to account for the formation of diaspora varieties it is crucial to investigate not only which region in the homeland immigrants come from, their number, and the dialect features they bring with them, but also ongoing language change in the homeland, languages in contact in the post-contact society, the social meanings of variants in both *pre*- and *post*-contact societies, and speakers' social networks and mobilities.

References:

- Bailey, G., Wikle, T., Tillery, J., & Sand, L. (1991). The apparent time construct. *Language Variation and Change*, *3*(3), 241–264.
- Britain, D. (2018). Dialect contact and new dialect formation. In C. Boberg *et al.* (eds.), *Handbook of dialectology*. Oxford: Wiley Blackwell.
- Mufwene, S. (2001). *The ecology of language evolution*. Cambridge: Cambridge University Press.
- Mufwene S. (2008). *Language evolution: Contact, competition and change*. London: Continuum.
- Rennicke, I. (2015). *Variation and change in the rhotics of Brazilian Portuguese*. Unpublished doctoral dissertation, University of Helsinki in cotutelle with Universidade Federal de Minas Gerais.
- Trudgill, P. (1986). *Dialects in contact*. Oxford: Blackwell.

Modelling social salience across multiple variables to show (non)-associations with social categories

Felicity Meakins[1], Jesse Stewart[2], Cassandra Algy[3], Lindell Bromham[4], Xia Hua[4]

University of Queensland[1], University of Saskatchewan[2],
Karungkarni Art and Culture Aboriginal Corporation[3], Australian National University[4]

Social salience is the property of a variable which listeners collectively associate with a social category. Variables where there is a high consensus within the community about their association with a social category are considered highly salient. One of the interests in social salience relates to its potential explanatory power in studies of language change. One prediction has been that variables with higher social salience will show higher rates of change (e.g., Baxter, Blythe, Croft, & McKane, 2009; Greenhill et al., 2017, p. E8827; Hinskens, 1996; Kerswill & Williams, 2002; Labov, 1972, pp. 178-179; 1994, p. 78; Rácz, 2013 but see Trudgill (1986) for a different perspective). This prediction has not been empirically tested due to a lack of metrics for measuring social salience in datasets. Perception experiments have been used to determine whether a variable has social salience by assessing the extent to which listeners from a speech community converge on the association of a variant with a social category (e.g., Campbell-Kibler, 2009; Fridland, Bartlett, & Kreuz, 2004; Plichta & Preston, 2005). However, most of these experiments have involved tests of individual variables which do not provide us with enough data to model the relative rates of uptake and loss of different variables based on their social salience in models of language evolution. Experiments which have assessed the social salience of multiple variables only consider the association of the variables with social categories (Llamas, Watt, & MacFarlane, 2016). This paper introduces a new method which provides a measure of salience across multiple variants in the Gurindji Kriol dataset. It also assesses the relative association and non-association of variants with social categories across these variants. This method provides us with the necessary metric required to show how the higher association or non-association of variants with social categories might affect rates of uptake and loss.

References

Baxter, Gareth, Blythe, Richard, Croft, William, & McKane, Allan. (2009). Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change, 21*(2), 257–296.

Campbell-Kibler, Kathryn. (2009). The nature of sociolinguistic perception. *Language Variation and Change, 21*(1), 135-156.

Fridland, Valerie, Bartlett, Kathryn, & Kreuz, Roger. (2004). Do you hear what I hear? Experimental measurement of the perceptual salience of acoustically manipulated vowel variants by Southern

speakers in Memphis. *Language Variation and Change, 16*(1), 1-16.

Greenhill, Simon, Wu, Chieh-Hsi, Hua, Xia, Dunn, Michael, Levinson, Stephen, & Gray, Russell. (2017). Evolutionary dynamics of language systems. *Proceedings of the Royal Society, 114*(42), E8822-E8829. doi:10.1073/pnas.1700388114

Hinskens, Frans. (1996). *Dialect Levelling in Limburg: Structural and Sociolinguistic Aspects*. Tübingen: Niemeyer.

Kerswill, Paul, & Williams, Ann. (2002). 'Salience' as an explanatory factor in language change: Evidence from dialect levelling in urban England. In M. Jones & E. Esch (Eds.), *Language Change: The Interplay of Internal, External and Extra-Linguistic Factors* (pp. 81-110). Berlin: Mouton de Gruyter.

Labov, William. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. (1994). *Principles of Linguistic Change: Internal Factors*. Oxford: Blackwell.

Llamas, Carmen, Watt, Dominc, & MacFarlane, Andrew. (2016). Estimating the relative sociolinguistic salience of segmental variables in a dialect boundary zone. *Frontiers in Psychology, 7*. doi:10.3389/fpsyg.2016.01163

Plichta, Bartek, & Preston, Dennis. (2005). The /ay/s have it: The perception of /ay/ as a North- South stereotype in US English. *Acta Linguistica Hafniensia, 37*(243-285).

Rácz, Péter. (2013). *Salience in Sociolinguistics: A Quantitative Approach*. Berlin: De Gruyter Mouton.

Investigating variation of mid-front vowels in Assamese

Saurabh Kumar Nath
The Australian National University
Saurabh.Nath@anu.edu.au

In the existing literature, it is generally accepted that Assamese has two mid-front vowels, e and ɛ, both represented by the single grapheme এ, however, the relationship between the two is still not quite clear. Some early studies reported that Assamese retains the early Indo-Aryan vowel e in the case of *tatsama* words (that is the words directly borrowed from Sanskrit to Assamese). Otherwise, Assamese generally has the vowel ɛ, and [e] is an allophone of /ɛ/ in certain phonological contexts and there are hardly any minimal pairs to understand their phonemicity. On the other hand, later studies postulate that the relationship between /e/ and /ɛ/ is phonemic as observed in a very limited number of environments, primarily contributed by loan words from English (for example, in b_l environment, bel 'a bell (English)' vs bɛl 'a wood-apple (Assamese)'). Likewise, it was reported earlier that /ɛ/ in Eastern and Central Assamese dialects is realised as /æ/ in Western Assamese, recent studies find a merger between /e/ and /ɛ/ in Western Assamese.

In the light of varied descriptions in early accounts and their respective methods, the study examines the status of present-day Assamese e and ɛ vowels following a speech community approach. Data were drawn from 10 female speakers belonging to two generations (4 adults and 6 children) residing in the district of Barpeta, a Western Assamese dialect region. A list of 75 words (or 225 tokens/speaker) was elicited in a wide variety of phonetic and phonological contexts through reading tasks. Both F1 and F2 values were extracted manually and plotted against each speaker in F1-F2 space normalizing the data.

Based on a series of quantitative analyses, the study suggests that:

a) Overall the relationship between e and ɛ is allophonic, but it is not categorical.

b) The lexical items categorically realising as the vowel e are not always *tatsama* words. In fact, it is phonetically motivated.

c) While age seems an important factor in the acquisition of allophonic relationship, other factors such as educational background and ethnicity also may play a major role.

Keywords: Assamese vowels, Allophonic variation, Language acquisition, Age, Ethnicity

Lebanese Voices: Language attitudes and perceptions of ethnicity in Sydney

Heba Bou Orm, Australian National University

The Lebanese community has a long-standing history of immigration and settlement in Australia, and continues to represent one of the largest ethnic groups in Sydney today (Australian Bureau of Statistics, 2021). The high concentration of Lebanese in Sydney's western suburbs, coupled with a series of sociohistorical events, has led to the stigmatisation of the diaspora, amplified by public discourse on the criminalisation of Lebanese Australian men (e.g., Collins, 2005; Poynting, Noble, & Tabar, 2001). While there has been a surge in research on ethnolectal variation in Australian English, including studies on the production of Lebanese Australians (e.g., see Clothier, 2019 for a review), perception research lags in comparison. Apart from Rieschild's (2007) study on "Lebspeak," which briefly reported on attitudes towards this variety based on survey results, the topic remains largely unexplored.

This paper reports on the use of a mixed methods approach in the study of language attitudes, involving: (1) sociolinguistic interviews with Lebanese Australian teenagers and adults to untangle in-group attitudes towards the "Leb" accent and identity; (2) an online perception experiment using 16 short sound clips of spontaneous speech from Australians of Lebanese (teenage and adult), Italian, Chinese, and Anglo-Celtic background. In the experiment, Sydneysiders from a range of backgrounds were asked to identify the speakers' ethnic background, explain what motivated their choice, indicate where in Sydney they think the speaker is from, and rate them on pre-determined characteristics (Confident, Friendly, Honest, Intelligent, Sophisticated).

Sociolinguistic interviews with Lebanese Australians indicate that in-group attitudes towards the "Leb" accent are overwhelmingly negative, reflecting an underlying sense of linguistic insecurity and highlighting complex ideological links between language, ethnicity, masculinity, and class. Results from mixed-effects models, based on data from 92 responses to the perception experiment, indicate that Sydneysiders are highly attuned to the speech of Lebanese Australian teens, who were the most accurately identified speakers in the study. Listener metalinguistic awareness is further exemplified by comments linking perceived Lebanese ethnicity with specific linguistic features (e.g.,*"th" more like a "d"*) and stereotypes (e.g., *sounds like a guy from western Sydney, particularly the troublemaker ones on the train*). The results also show that evaluations of speakers are more strongly tied to perceived rather than actual ethnicity, reflecting stereotypical characteristics that listeners associate with different groups. For example, compared to speakers perceived as Anglo, those perceived as Lebanese were rated as significantly less intelligent and sophisticated (and Chinese as significantly more intelligent but less confident).

These findings highlight the need for further perception research, to add to our understanding of language attitudes in multicultural settings and to help combat stereotypes that may impede access to social equality.

References

Australian Bureau of Statistics. (2021). *Census of population and housing*.

Clothier, J. (2019). Ethnolectal variability in Australian Englishes. In *Australian English Reimagined* (pp. 155-171). London: Routledge.

Collins, J. (2005). From Beirut to Bankstown: the Lebanese diaspora in multicultural Australia. In *Lebanese Diaspora: history, racism and belonging* (pp. 187-211). Beirut, Lebanon: Lebanese American University.

Poynting, S., Noble, G., & Tabar, P. (2001). Middle Eastern Appearances: "Ethnic Gangs", Moral Panic and Media Framing. *Australian & New Zealand journal of criminology, 34*(1), 67-90. doi:10.1177/000486580103400105

Rieschild, V. R. (2007). Influences of language proficiency, bilingual socialization, and urban youth identities on producing different Arabic-English voices in Australia. *Novitas-ROYAL, 1*(1), 34-52.

Methods in Dialectology XVIII – La Trobe University

Using variationist sociolinguistic research to inform dialect coaching for performance: addressing a methodological gap

Abstract

Dialect coaching for performance, in contrast to naturalistic second dialect acquisition, necessarily relies on explicit instruction (Siegel, 2010: 192-3), though the evidential basis on which that instruction rests remains contested. While linguists have viewed dialect coaching for performance as involving little more than imitation (Siegel, 2010: 65; Watt, 2016: 219), dialect coaches have challenged that assumption, noting that "it is the academic foundation of linguistics which informs the establishment of a character voice in drama" (Gunn 2023: 197). This has given rise to a push for descriptivist and "sociolinguistically aware" approaches to dialect coaching (Thompson et al., 2017: 331-2), and for the inclusion of phonetics instruction in actor training programs (e.g., Armstrong, 2016). The performance industry more widely serves to create and perpetuate representations that influence our social understanding (Brooks & Hébert, 2006; Bell & Gibson, 2011: 558), highlighting the need for informed dialect portrayals in performance media. And although it is recognised that dialect coaches have an important goal of establishing for actors a convincing speech system that is grounded in linguistic reality (Gunn, 2023: 204), there appears to be very little in the way of research to establish methodological bridges between variationist sociolinguistics and dialect coaching.

While there exists piecemeal research that employs sociolinguistic methods to explore aspects of accent or dialect instruction in a broad sense (e.g., Nelson, 2018; Cerreta & Trofimovich, 2018; Knooihuizen, 2019), as an area of applied dialectology, dialect coaches rely specifically on evidence of sociolinguistic variation to inform their instruction (Watt, 2016: 219). I demonstrate here that variationist sociolinguistic work is ideally placed to provide the kind of empirical evidence dialect coaches need, and make some suggestions about how such research can be harnessed by dialect coaches to inform their practice, with a specific focus on vocalic variation in Australian English (cf., Purser et al., 2020).

To do this, I draw on analysis of vowels in Australian English, utilising a spontaneous speech corpus consisting of sociolinguistic interviews with 95 speakers in Sydney, as part of the Sydney Speaks corpus (Travis 2014-2022), which provides a total of 72,700 tokens across 19 vowel types (Purser et al., 2020). I highlight three main ways in which such research is well-placed to inform dialect coaches. First, given the need for an empirical basis for which accent features to teach performers, the use of such spontaneous speech data from a wide range of speakers capturing diversity in the society, across, for example, age, gender, social class, and ethnicity, can better inform acceptable accent portrayals in performance. Second, a dialect coach's work necessarily involves identifying the target speech community and analysing the relevant sociolinguistic features, which may include variation in vowels, as presented here, but also in rhoticity, consonant production, and morphophonological variables, all of which can be extracted from sociolinguistic corpora. And third, recognising that there is no single, precise target point for vowel realisations, and instead focusing on a target range of variation, can consequently provide options to performers to create well-rounded characters, rather than static or stereotyped accent representations. Ultimately, while there remains a methodological gap in adapting variationist research for the performance industry, here I demonstrate ways in which sociolinguistic work can address this gap and inform dialect coaching practice.

References

Armstrong, E. (2016). Efficacy in phonetics training for the actor. *Voice and Speech Review*, *10*(1), 36–52. https://doi.org/10.1080/23268263.2017.1282676

Bell, A., & Gibson, A. (2011). Staging language: An introduction to the sociolinguistics of performance. *Journal of Sociolinguistics*, *15*(5), 555–572.

Brooks, D. E., & Hébert, L. P. (2006). Gender, Race, and Media Representation. In B. J. Dow & J. T. Wood (Eds.), *The Sage Handbook of Gender and Communication* (pp. 297–317). Sage.

Cerreta, S., & Trofimovich, P. (2018). Engaging the senses. *Journal of Second Language Pronunciation*, *4*(1), 46–72. https://doi.org/10.1075/jslp.00003.cer

Gunn, B. (2023). Dialect coaching and linguistics. In *Communicating Linguistics: Language, Community and Public Engagement* (pp. 197–207). Taylor and Francis. https://doi.org/10.4324/9781003096078-20

Knooihuizen, R. (2019). Accuracy and acceptability of second-dialect performance on American television. *English Language and Linguistics*, *23*(2), 229–252. https://doi.org/10.1017/S1360674317000363

Nelson, K. (2018). Accent Modification in the Teleconferencing Industry. *Voice and Speech Review*, *12*(1), 65–76. https://doi.org/10.1080/23268263.2018.1407108

Purser, B., Grama, J., & Travis, C. E. (2020). Australian English over Time: Using Sociolinguistic Analysis to Inform Dialect Coaching. *Voice and Speech Review*, *14*(3), 269–291. https://doi.org/10.1080/23268263.2020.1750791

Siegel, J. (2010). *Second dialect acquisition*. Cambridge University Press.

Thompson, P., Caban, A., & Singer, E. (2017). Vocal traditions: Knight-Thompson Speechwork. In *Voice and Speech Review* (Vol. 11, Issue 3, pp. 329–338). Bellwether Publishing, Ltd. https://doi.org/10.1080/23268263.2017.1402560

Travis, C. E. (2014-2022). *Sydney Speaks*. Australian Research Council Centre of Excellence for the Dynamics of Language, Australian National University. http://www.dynamicsoflanguage.edu.au/sydney-speaks/

Watt, D. (2016). Applied dialectology: Dialect coaching, dialect reduction, and forensic phonetics. In *The Handbook of Dialectology* (pp. 219–232). Wiley Blackwell. https://doi.org/10.1002/9781118827628.ch12

Quantifying dialect distance with acoustic models: A comparison to feature-based distance using single words

Hedwig Sekeres, Remco Knooihuizen & Martijn Wieling
University of Groningen

One of the main challenges in dialectology is the quantification of distance between dialects. Distance can be calculated on the basis of phonetic transcription or the presence of certain features. However, these traditional methods are labor intensive and may lead to an (unintentional) bias for certain features, perhaps excluding other relevant sounds and features (Nerbonne, 2009). Recent developments in natural language processing have made it possible to compare acoustic recordings directly without transcribing or coding.

In this study, a deep acoustic wav2vec 2.0 model that was fine-tuned on Dutch (Bartelds & Wieling, 2022) is used to investigate variation between dialects of Low Saxon (which is closely related to Dutch) in the northern Netherlands. The goal of this study is to test whether acoustic distances as calculated by the model can be used to investigate language variation. In particular, we are interested in the use of this distance measure for comparing individual words, in addition to aggregating all words in the dataset as is standard practice in dialectometry (Heeringa & Nerbonne, 2013).

90 participants from two provinces in the northern Netherlands took part in an online survey in which they were asked to translate 62 words from Dutch into their (Low Saxon) dialect and to record themselves pronouncing them. The words were selected on the basis of previous research into sound change and variation in these dialects (e.g., Reker, 1983; 1986). We then calculated the acoustic distances, but also (for comparison) manually annotated the presence vs. absence of specific features.

For each word, the correlation between acoustic and feature-based distance matrices is assessed using Mantel tests. Additionally, maps based on multidimensional scaling (MDS: similar colors represent similarity regarding the acoustics or features) are visually inspected in order to determine to what degree acoustic distances align with feature-based distances. Both maps are also compared to traditional dialect maps. Preliminary results indicate that acoustic distances align well with feature-based distances. An example of a map generated on the basis of acoustic distances for a single word (*paal*; ‘pole’) as well as a feature-based map (based on two features; so distances are either 0 when both features match, 0.5 when one is different, or 1 when both are different) can be found in Figure 1. In this particular example, the acoustic map captures more variation than the feature-based map, but the overall patterns are similar. Furthermore, both maps appear to capture the gradual differences traditionally observed between the north and south of the region, as well as the more abrupt isogloss separating the easternmost part of the region from the rest.

The use of acoustic models opens up new avenues for explorative research as it allows researchers to investigate language variation or change without focussing on specific sounds or features and allows for larger amounts of data to be processed. This method can be a powerful

tool for identifying areas and words of interest which can then be investigated using more qualitative methods to obtain a deeper understanding of language variation.

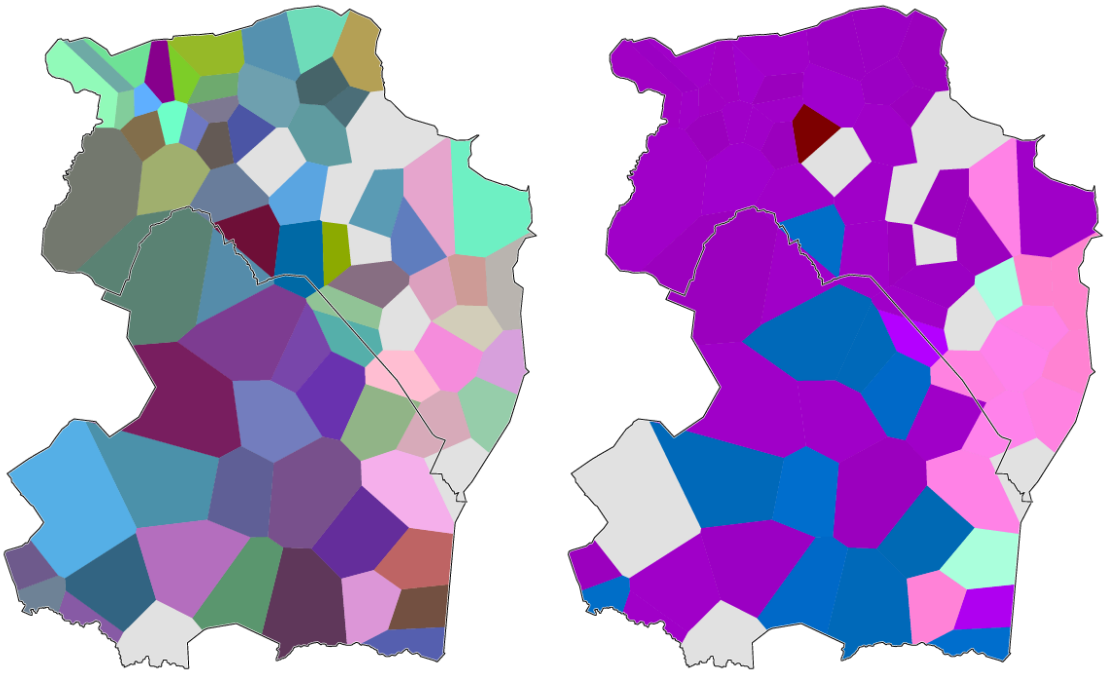


Figure 1: MDS maps for the acoustic distances (left) and feature-based distances (right) of a single word.

Works cited:

Bartelds, M., & Wieling, M. (2022). Quantifying language variation acoustically with few resources. In Carpuat, M., de Marneffe, M., & Meza Ruiz, I. V. (Eds.) *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 3735-3741). Association for Computational Linguistics. 10.18653/v1/2022.naacl-main.273

Heeringa, W., Nerbonne, J. (2013). Dialectometry. In Hinskens, F., & Taeldeman, J. *Language and Space: Dutch* (pp. 624-646). De Gruyter Mouton.

Nerbonne, J. (2009). Data-driven dialectology. *Language and Linguistics Compass, 3*(1), 175-198.

Reker, S. (1983). Tussen Hogelandster Gronings en het standaard-Nederlands. *Driemaandelijksche Bladen, 35*(1), 1-28.

Reker, S. (1986). Schriftelijk dialect-onderzoek in “ronding” in Groningen: Van ‘k wait nait’ tot ‘k woit noit’. *Driemaandelijksche Bladen, 38*(3-4), 178-187.

What constrains the malleability of idiolects over time?:
A multivariate analysis of NINJAL's real-time panel corpora of the standardization of Hokkaido (Furano) Japanese

Shoji Takano
Hokusei Gakuen University

Yoshiyuki Asahi
NINJAL (National Institute for Japanese Language and Linguistics)

A recent surge of panel studies on language variation and change has shed light on the malleability of an individual's grammar throughout their lifetime. These studies have provided valuable insights into how linguistic changes actually occur at the individual level (e.g., Beaman & Buchstaller 2021). However, at the same time, some of the findings have led to conflicting generalizations about the complexities of intra-speaker malleability as the field continues to grow.

It has been observed that while the majority of individuals remain stable in their grammar over time, some adapt their language use to reflect ongoing changes in the community (lifespan change), while others revert back to older usage patterns (retrograde change) (Sankoff 2013). The longitudinal stability of idiolects varies depending on different levels of grammar. As a general rule, vocabulary items are said to be more susceptible to change, while prosodic variables remain largely unaffected (Chambers 2009). Intra-speaker changes over time can also be variable-specific phenomena. The saliency of variables (Oushiro & Guy 2015) as well as their stages of variation at the community level (e.g., "unsteadily advancing changes") (Nahkola & Saanilahti 2004) may be responsible for the activation of lifespan changes.

In Japanese dialect standardization, which is typically a change from above (the level of speech community members' consciousness) (Labov 1966), several findings contradict with the above-mentioned generalizations, primarily based on changes from below. Lifespan changes in Tsuruoka Japanese are consistently observed in lexical accents (Yokoyama & Sanada 2010), whereas retrograde changes are more common in segmental phonology (Yokoyama 2010). In Furano Japanese (Hokkaido), segmental phonology and morphosyntax remain relatively stable, while vocabulary items and lexical accents exhibit versatility over time. In Sapporo Japanese (Hokkaido), however, lexical accents are found to remain almost intact over time. Intra-speaker malleability is constrained by lexical properties such as frequency and familiarity, and it may vary even

across different communities of the same dialect region (Furano vs. Sapporo in Hokkaido), where the rates and stages of advancement of standardization, as well as the members' loyalty to and pride in their local culture, significantly differ (Takano 2021).

This paper represents the first attempt to make sense of such contradictory findings by uncovering various factors determining intra-speaker malleability over time. To achieve this, we utilize the extensive panelist information from the National Institute for Japanese Language and Linguistics (NINJAL)'s large-scale panel corpora, focusing on Japanese dialect standardization in Furuno, Hokkaido. Data were collected in 1959 and 1986 but have remained largely unexplored until now. Our multivariate analyses, conducted through Rbrul, take into consideration both language internal (e.g., lexical items per se, levels of grammar) and sociolinguistic factors (e.g., lexical frequency/familiarity, stages of standardization), along with speaker demographics, including age, gender, and education. Additionally, we consider the regionality index that measures the panelists' integration into the local culture (Chambers & Heisler 1999). Results from our ongoing analyses will reveal the relative effectiveness of these intersecting factors on intra-speaker changes over time and, ultimately, contribute to the development of improved theories of language change (498 words).

References

Beaman, Karen V. & Buchstaller, Isabelle. (eds.) (2021). *Language Variation and Language Change Across the Lifespan: Theoretical and Empirical Perspectives from Panel Studies.* New York: Routledge.

Chambers, Jack K. (2009). *Sociolinguistic Theory (Revised Edition)*. Oxford: Wiley-Blackwell.

Chambers, Jack K. & Heisler, Troy. (1999). Dialect topography of Québec City English. *Canadian Journal of Linguistics* 44(1): 23-48.

Labov, William. (1966). *The Social Stratification of English in New York City.* Washington, D.C.: Center for Applied Linguistics.

Nahkola, Kari & Saanilahti, Marja. (2004). Mapping language changes in real time: A panel study on Finnish. Language Variation and Change 16: 75-921.

Oushiro, Livia & Guy, Gregory R. (2015). The effects of salience on co-variation in Brazilian Portuguese. University of Pennsylvania Working Papers in Linguistics 21(2): 156-166.

Sankoff, Gillian. (2013). Longitudinal studies. In R. Bayley, R. Cameron, & C. Lucas (eds.), The Oxford Handbook of Sociolinguistics, pp. 261-279. Oxford: Oxford University Press.

Takano, Shoji. (2021). Lifespan “changes from above” in the standardization of Japanese regional dialects: Levels of grammar, lexical properties and community characteristics. *Language Variation and Change* 3(3): 297-329.

Yokoyama, Shooichi. (2010). *Onsei kyootsuugoka-no yosoku-to kenshoo* (Predictions and examinations of standardization in speech). Special Lecture 2. The 24th General Meeting of the Phonetic Society of Japan at Kokugakuin University.

Yokoyama, Shooichi, & Sanada, Haruko. (2010). *Gengo-no shoogai shuutoku moderu-ni yoru kyootsuugoka yosoku* (Predictions of dialect standardization by the “life-long assimilation of language change” model). *Nihongo-no Kenkyuu* 6(2): 31-45.

Mapping attitudes in French-speaking Switzerland

Jennifer Thorburn (University of Lausanne)

Although multilingualism is "an essential part of Switzerland's self-conception" (Werlen 2007: 138), the Swiss are not fluent in all national languages; instead, language use is governed by the territoriality principle, which allows each canton to choose its official language(s), resulting in rather rigid linguistic boundaries (Reichenau 1997: 107) between the French-, German-, and Italian-speaking regions. Consequently, there is a "considerable amount" of individual monolingualism in the country, with the Italian-speaking area being the most multilingual and the French-speaking region the most monolingual (Berthele & Wittlin 2013: 184, citing Werlen et al. 2009). (The fourth national language, Romansh, is highly endangered and is an official language only in the canton of Graubünden, alongside German and Italian.)

Studies of language attitudes in Switzerland often focus on different varieties or usage of a single national language (e.g., Singy 1996, L'Éplattenier-Saugy 2002, Prikhodkine 2011, Deboffe 2014, Studler 2017, Schmidlin 2022); attitudes and ideologies toward multiple languages are typically considered in the context of national identity (e.g., Watts 1998, Matser et al. 2010) or language policy (e.g., Berthele 2016, Chevalier 2016), or in relation to English (e.g., Dürmuller 1989, Aloise 2016, Stępkowska 2016).

The present study focuses on the French-speaking area of Switzerland, also known as *Romandie*. Data were gathered by undergraduate students at a local university as part of their "Introduction to Sociolinguistics" module using traditional perceptual dialectology techniques, modelled on Bucholtz et al.'s (2007) study, in which undergraduates in California employed a map task along with a brief demographic questionnaire and two open questions about the best and worst varieties in the area.

In the first year of the project, the map included the entirety of Switzerland. Participants overwhelmingly highlighted the broad language regions of the country, with nearly every respondent noting the *Röstigraben*, the cultural and linguistic border between the French- and German-speaking areas. The Italian-speaking region was consistently but less frequently demarcated. Very few participants highlighted Graubünden, suggesting that Romansh does not factor into most participants' conceptions of Swiss language regions.

In subsequent years, students focused on either *Romandie* as a whole or on a particular canton. Preliminary analyses of the maps of *Romandie* (N= 2,361) show that respondents frequently followed cantonal boundaries to indicate different dialects of French, echoing L'Eplattenier-Saugy's (2002) findings. Geneva was most cited as the place people speak the best (35.1%), with participants referencing its proximity to France, which was also the rationale offered by those who said people from Geneva speak the worst (9.2%). Valais (26.0%) and Jura (24.1%) were the cantons most frequently named as places where people speak the worst. In both the maps and open questions, respondents regularly, and often disfavourably, commented on the French/German bilingualism of these cantons. Finally, a notable number of participants did not provide a best (15.2%) or worst (21.0%) choice, with many of these people questioning the validity and relevance of these prompts, an outcome not discussed in perceptual dialectology studies.

Next steps will consider social predictors such as age, gender, and canton, which have been significant in other studies of attitudes in *Romandie* (Singy 1996, L'Eplattenier-Saugy 2002, Deboffe 2014).

References

Aloise, Adriano. 2016. The study of integrative and instrumental motivation of students learning English as a second language in Lausanne, Switzerland. In Patricia Ronan (ed.), *Perspectives on English in Switzerland*. Lausanne: University of Lausanne, 149-170.

Berthele, Raphael. 2016. Demography vs. legitimacy: Current issues in Swiss language policy. In Patricia Ronan (ed.), *Perspectives on English in Switzerland*. Lausanne: University of Lausanne, 27-51.

Berthele, Raphael, and Gabriele Wittlin. 2013. Receptive multilingualism in the Swiss army. *International Journal of Multilingualism* 10.2, 181-195.

Bucholtz, Mary, Nancy Bermudez, Victor Fung, Lisa Edwards, and Rosalva Vargas. 2007. Hella Nor Cal or totally So Cal?: The perceptual dialectology of California. *Journal of English Linguistics* 35, 325-352.

Chevalier, Sarah. 2016. The value of English in multilingual families. *SPELL: Swiss Papers in English Language and Literature* 33, 97-116. http://doi.org/10.5169/seals-632497

Deboffe, Mathieu. 2014. *Teenagers' attitudes towards Franglais in Amiens and in Lausanne: A comparative sociolinguistic study*. MA thesis, University of Lausanne.

Dürmuller, Urs. 1989. Attitudes towards English as a possible lingua franca in Switzerland. *York Papers in Linguistics* 14, 3-17.

L'Éplattenier-Saugy, Caroline. 2002. A perceptual dialect study of French in Switzerland. In Daniel Long and Dennis R. Preston (eds.), *Handbook of Perceptual Dialectology, Volume* 2. Amsterdam: John Benjamins, 351-365.

Matser, Carine, Jan Pieter van Oudenhoven, Françoise Askevis-Leherpeux, Arnd Florack, Bettina Hannover, and Jerôme Rossier. 2010. Impact of relative size and language on the attitudes between nations and linguistic groups: The case of Switzerland. *Applied Psychology: An International Review* 59.1, 143-158.

Prikhodkine, Alexei. 2011. *Dynamique normative du française en usage en Suisse Romande*. Paris: L'Harmattan.

Reichenau, Christoph. 1997. Eine Debatte, doch kein Dialog. *Swiss Political Science Review* 3.4, 100-108.

Schmidlin, Regula. 2022. Standard variation and linguistic attitudes in German-speaking Switzerland: From the etic to the emic perspective. In Alexandra N. Lenz, Barbara Soukup, and Wolfgang Koppensteiner (eds.), *Standard Languages in Germanic-Speaking Europe: Attitudes and Perception*. Oslo: Novus, 97-120.

Singy, Pascal. 1996. *L'image du française en Suisse romande*. Paris: L'Harmattan.

Stępkowska, Agnieszka. 2016. The Swiss paradigm of multilingualism and English. In Patricia Ronan (ed.), *Perspectives on English in Switzerland*. Lausanne: University of Lausanne, 79-92.

Studler, Rebekka. 2017. Cognitive cultural models at work: The case of German-speaking Switzerland. *Yearbook of the German Cognitive Linguistics Association* 5: 93–106.

Watts, Richard J. 1998. Language, dialect and national identity in Switzerland. *Multilingua* 7.3, 313-334.

Werlen, Iwar. 2007. Receptive multilingualism in Switzerland and the case of Biel/Bienne. In Jan D. ten Thije and Ludgar Zeevaert (eds), *Receptive multilingualism: Linguistic analyses, language policies and didactic concepts*. Amsterdam: John Benjamins, 137-158.

Werlen, Iwar, Lukas Rosenberger, and Jachin Baumgartner. 2009. *Sprachkompetenzen der erwachsenen Bevölkerung in der Schweiz*. Zürich: Seismo.

Voices of Regional Australia: Assessing local attachment in legacy data

Catherine Travis (ANU), Gerry Docherty (Griffith U), Ksenia Gnevsheva (ANU)

The importance of place and belonging in sociolinguistic variation has been acknowledged since the birth of the field (Labov 1963). More recently, there has been much interest in how to quantify measures of local attachment, and meaningful metrics have been developed from questionnaire data, tailored interview questions, as well as from the content of sociolinguistic interviews (e.g., Carmichael 2023; Pabst 2022; Reed 2020). Such methods are built into the research design, and require direct access to the community, options which may not be available when working with recordings collected for another purpose. With such legacy data, information about local attachment must be obtained from content that arises in the existing recordings. In this talk we illustrate and assess a methodology for extracting and quantifying information about local attachment from existing data.

The legacy data used here are a collection of oral history interviews about the 2019-2020 Black Summer bushfires, collected around the township of Braidwood, NSW, Australia (cf., https://braidwoodradio.com.au/podcast/hots/). These oral histories share a lot of characteristics with sociolinguistic interviews, in particular in comprising conversations around a loosely structured set of topics and narratives of personal experience, including 'danger of death' stories (Labov 2006 [1966]).

All interviews contain some information relating to local attachment, though on varying topics. This includes information about length of residence in the area ('we moved here in 2003'); family background ('I come from a farming family, my father's side'); family in the region (one participant talks about his grandmother being evacuated during the fires and 'sleeping in a swag in the courtyard of the pub'); attachment to the area ('we absolutely just love living here'); involvement in local organisations ('I joined up with the X Fire Brigade, and then I decided to join up with Y [fire brigade]'); and importance of community ('when the fires happened, everyone bonded together, it was good').

We systematically extracted all relevant comments, and organised them into topics. Each topic was given a score based on the comments relevant to it (be that just one comment, or several), and then an overall average across all topics was calculated. In this way, it was possible to assign a score to a large number of the participants, despite having different information from them. To assess this methodology, we compare this scoring with information obtained via a demographic questionnaire for a subset of speakers who participated in a follow up interview, allowing us to see what is missed, and what is gained, from the content extraction.

This talk presents a proof-of-concept for the development of a content-based local attachment metric for legacy data, highlighting both the challenge of such an approach — that different information is available for each participant— and its merits— that it allows us to go beyond categories that are pre-determined by the researcher, to identify topics of relevance to the community (Torres Cacoullos & Travis 2018:62; cf., Poplack, Walker & Malcolmson 2006).

In closing, we consider the interplay between community-specific and more broadly applicable themes and the transferability of local attachment metrics.

References

Carmichael, Katie. 2023. Locating place in variationist sociolinguistics: Making the case for ethnographically informed multidimensional place orientation metrics. *Journal of Linguistic Geography*: 1-13. https://doi.org/10.1017/jlg.2023.2

Labov, William. 1963. The social motivation of a sound change. *Word* 19: 273-309. https://doi.org/10.1080/00437956.1963.11659799

Labov, William. 2006 [1966]. *The social stratification of English in New York City* (2nd ed.). Cambridge: Cambridge University Press.

Pabst, Katharina 2022. Putting “the Other Maine” on the map: Language variation, local affiliation, and co-occurrence in Aroostook County English. PhD thesis, Department of Linguistics, University of Toronto.

Poplack, Shana, James A. Walker and Rebecca Malcolmson. 2006. An English "like no other"?: Language contact and change in Quebec. *Canadian Journal of Linguistics / Revue canadienne de Linguistique* 51(2): 185-213.

Reed, Paul E. 2020. Prosodic variation and rootedness in Appalachian English. *University of Pennsylvania Working Papers in Linguistics* 25(2): 107-114. https://repository.upenn.edu/pwpl/vol25/iss2/13/

Torres Cacoullos, Rena and Catherine E. Travis. 2018. *Bilingualism in the community: Code-switching and grammars in contact*. Cambridge: Cambridge University Press.

SPRAAKLAB: A mobile laboratory benefitting dialect data collection

Martijn Wieling, Raoul Buurke, Teja Rebernik & Jidde Jacobi
University of Groningen

One of the research lines in our lab focuses on collecting and analyzing speech data from dialect speakers. However, obtaining a representative sample of dialect speakers may be difficult as most dialect speakers are often older, less mobile individuals who live in rural, localized communities (Auer, 2005; Chambers & Trudgill, 1998). These speakers may be less able and willing to travel to the (urban) university to participate in our studies.

Consequently, researchers involved in large dialect atlas data collection efforts (e.g., Goeman & Taeldeman, 1996) have generally travelled to dialect speakers' homes and recorded their dialectal speech using portable recording equipment. However, due to variable environments, these speech recordings usually also contain undesired noise, for example a ticking clock or traffic noise. While this may not be a problem for dialect data collection when the goal is to first transcribe the data before any analyses are conducted, it is a problem when one wishes to conduct a detailed acoustic analysis. For example, quantifying pronunciation distances on the basis of manual phonetic transcriptions using the Levenshtein distance (e.g., Heeringa, 2004) has long been the gold standard (Wieling et al., 2014), but recently an acoustic-only computational method was found to be superior (Bartelds et al., 2022a; Bartelds et al., 2022b). In addition, collected dialectal speech data may be used for developing speech technology, as long as the recordings are of sufficient quality and background noise is minimized. Consequently, collecting high-quality acoustic data when studying regional dialectal variation has become ever more important.

For this reason, we have developed SPRAAKLAB: a mobile laboratory for collecting speech production data in the field (Rebernik et al., submitted). Our mobile laboratory (see Figure below) is a large van (LxWxH: 7x2.75x3m) with an attractive outside design, custom-built on top of a lowered chassis of a Fiat Ducato. SPRAAKLAB contains two separate rooms: one where the experimenter(s) control all equipment, and a sound-dampened room (-40dB) containing several high-quality microphones. Separating speaker and experimenter, may be helpful in limiting the degree to which a dialect speaker is prone to use more formal speech norms in the presence of a non-dialectal experimenter (the so-called "observer's paradox"; see Cukor-Avila, 2000). Furthermore, the setup of the mobile lab is highly flexible. It allows not only for experiments collecting speech production data from individual speakers, but also for experiments assessing language variation and change involving dyads (e.g., Wieling et al., 2020).

In addition to our regular experiments, we also use SPRAAKLAB to inform the general audience about the research we conduct. We often visit science or music festivals for this purpose, which we combine with experimental data collection. For example, at the *Zwarte Cross* festival (~240,000 visitors) in 2023, we collected dialectal speech pronunciations for 100 words from over 100 people in three days.

During the presentation at Methods XVIII, we will discuss the specifications of SPRAAKLAB and illustrate how we use the mobile laboratory, so that our experiences may be beneficial to others.

References

Auer, P. (2005). Europe's sociolinguistic unity, or: A typology of European dialect/standard constellations. In Delbecque, Van der Auwera, Geeraert (eds.): *Perspectives on variation: Sociolinguistic, historical, comparative*, De Gruyter, pp. 7–42.

Bartelds, M., de Vries, W., Sanal, F., Richter, C., Liberman, M., & Wieling, M. (2022). Neural representations for modeling variation in speech. *Journal of Phonetics*, 92: 101137.

Bartelds, M., & Wieling, M. (2022). Quantifying language variation acoustically with few resources. *Proceedings of NAACL*, 3735-3741.

Chambers, J. K., & Trudgill, P. (1998). *Dialectology*. Cambridge University Press. 2nd edition.

Cukor-Avila, P. (2000). Revisiting the observer's paradox. *American Speech*, *75*(3), 253-254.

Heeringa, W. J. (2004). *Measuring dialect pronunciation differences using Levenshtein distance*. PhD thesis, University of Groningen.

Rebernik, T., Jacobi, J., Buurke, R., Tienkamp, T., Tsiwah, F., Oud, S., Reinders, A., Abur, D., & Wieling, M. (submitted). Mobile Laboratory SPRAAKLAB: technical specifications, practical experiences, and a validation study.

Taeldeman, J., & Goeman, A. (1996). Fonologie en morfologie van de Nederlandse dialecten: Een nieuwe materiaalverzameling en twee nieuwe atlasprojecten. *Taal en Tongval*, 48, 38-59.

Wieling, M., Bloem, J., Mignella, K., Timmermeister, M., & Nerbonne, J. (2014). Measuring foreign accent strength in English: validating Levenshtein distance as a measure. *Language Dynamics and Change*, 4(2), 253 - 269.

Wieling, M., Tiede, M., Rebernik, T., de Jong, L., Braggaar, A., Bartelds, M., Medvedeva, M., Heisterkamp, P., Freire Offrede, T., Sekeres, H., Pot, A., van der Ploeg, M., Volkers, K., & Mills, G. (2020). A novel paradigm to investigate phonetic convergence in interaction. *Proceedings of the 12th International Seminar on Speech Production*, pp. 1 - 4.

Morphology and Rhotacization in Beijing Chinese: Unravelling Complex Linguistic Dynamics

Claire Jingyuan YE
Latrobe University

Rhotacization in the Beijing dialect of Chinese operates on multiple linguistic levels, presenting challenges to variationist analysis. Beyond the phonological process (adding a rhotic sound), rhotacization involves the evolution of the free morpheme 儿 *er* ' descendant' into a bound suffix, signifying a shift from a notional word to a diminutive expression, triggered by semantic change. Because of this complexity, the primary focus of previous studies was on the evolution of rhotacized syllables. Rather than considering other potential variants and the variable context, analysis focused on the distribution of only the rhotacized words.

To address methodological ambiguities, this paper proposes an accountable analysis of rhotacization. Data were collected from participants residing in Beijing's historically central Xicheng and Dongcheng Districts, selected for their relatively stable geographical boundaries and recognized for possessing a highly acknowledged representative Beijing accent. Participants were recorded in reading and conversational contexts.

The initial step used the reading data to identify all potential variants. 58 participants were engaged in reading a short story which included 33 rhotacized syllables, realized by 73 rhotacized morphemes to form 95 potential rhotacized words, thus generating limited and targeted entries (102 in total). Following the Principle of Accountability (Labov 1966, 1972), in addition to the variant of interest (the rhotacized forms), other variants (illustrated in the examples below) were identified, including another diminutive form (the diminutive suffix), the absent form (the zero variant), and the base form.

Example: 石墩 *shítūn* [ʂʴ́ tūn], meaning a block of stone used as a seat;
(1) Rhotacized syllable: 石墩儿 *shítūner* [ʂʴ́tūəɻ]
(2) Zero variant: 石墩 *shítūn* [ʂʴ́ tūn]
(3) Diminutive suffix: 石墩子 *shítūnzi* [ʂʴ́ tūn tsɿ]

Example: 地 *dì* [tì], meaning some place;
(1) Rhotacized syllable: 地儿 *dìer* [tìə˞]
(2) Base form: 地方 *dìfāng* [tìfɑŋ/tìfɑŋ]

These words were classified into four categories (obligatory, recommended, optional and forbidden) based on regulations from the official Putonghua Proficiency Test and the vernacular Beijing rhotacization dictionary (Jia, 2019). After excluding categorical tokens, variation within each category was exhibited as evidence of identification of all potential variants (Table 1).

A further pilot study selected one participant from each gender and age group (totalling six participants) to circumscribe the variable context (Poplack & Tagliamonte 1989). Instead of syllables or words, each morpheme containing a rhotacized syllable was identified and ranked into five levels of frequency (Labov 1982). The following variation analysis started from the highest-frequency morphemes (Table 2) (N=989).

In studying a linguistic variable comprising several linguistic structures, one of the mainstays of variationist analysis is to isolate a specific level, as the departure point, to deliminate the variable context. Here, targeting the intermediate linguistic unit (morphemes), this study elucidates the nuanced variation within rhotacization in Beijing dialect. Only two of the nine morphemes did not exhibit variation. Despite vital internal linguistic constraints, the inherent variation of rhotacization is confirmed. The analysis presented here is intended to serve as a proposal for further research into this intricate linguistic phenomenon.

Key words: rhotacization, Beijing dialect, morpheme, zero variant, variable context;

Table 1 Distribution of variants by regulated categories

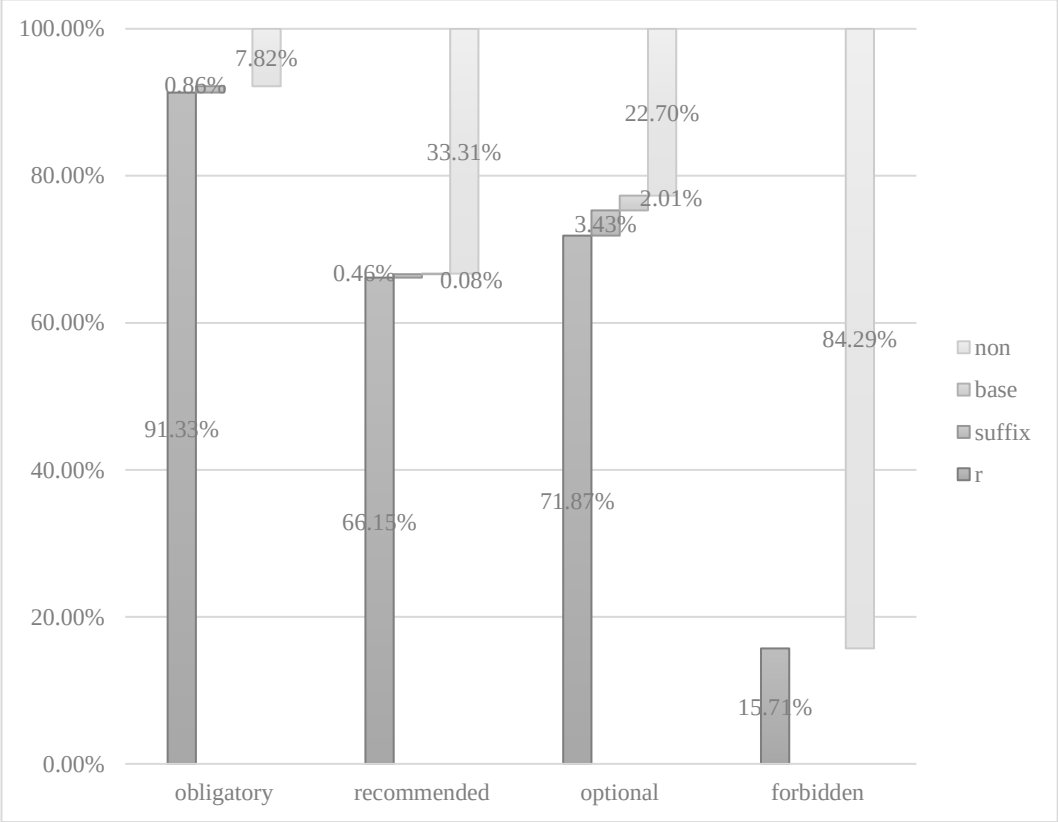
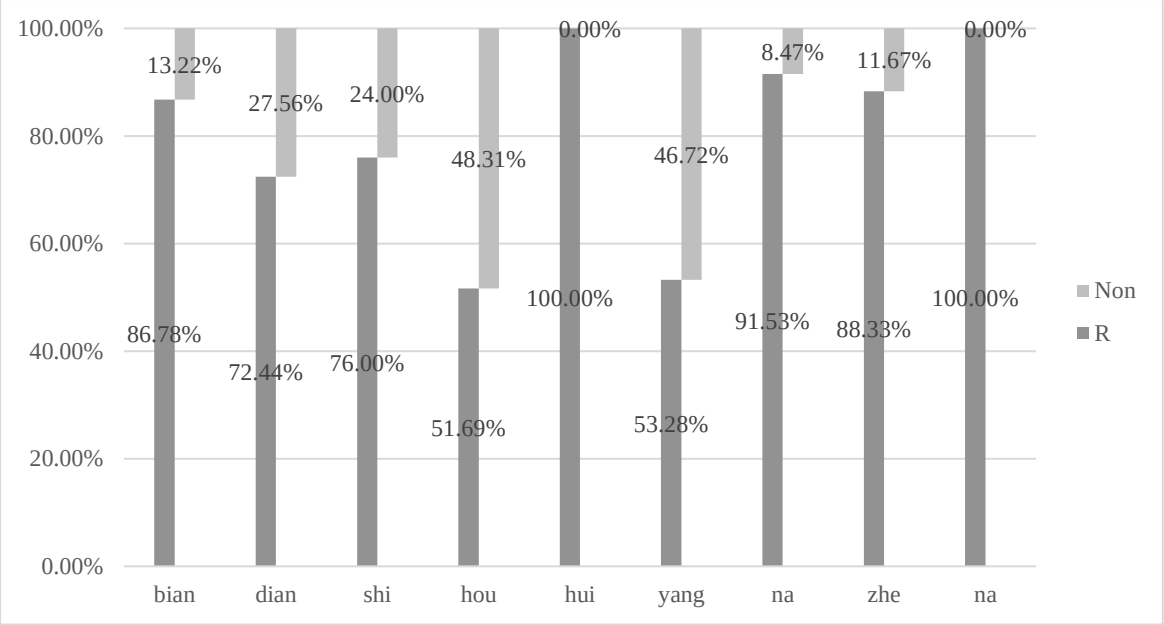


Table 2 Variation of highest-frequency morphemes



Reference

Jia, Caizhu. (2019). *Dictionary of Beijing Rhotacized Words* (2nd edition). Shanghai Education Publisher. (1st edition published in 1990)

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania press.

Labov, W. (1982) Building on empirical foundations, in W.P. Lehmann and Y. Malkiel (eds), Perspectives on Historical Linguistics. Amsterdam and Philadelphia: John Benjamins, pp. 17–92.

National Language Committee of PRC. (2021) *Outline for the Implementation of Putonghua Proficiency Test* (2nd edition). Commercial Press. (1st edition published in 2004)

Poplack, S., & Tagliamonte, S. (1989). There's no tense like the present: Verbal-s inflection in early Black English. *Language Variation and Change*, *1*(1), 47-84.

Applying dialectometry to a categorical dataset built from published sources: discussion of steps and methodological issues based on African Arabic dialects

Carolina Zucchi

BIGSAS, University of Bayreuth, Germany

Commonly in dialectometry comparative word lists are used as input data for the computation of dialect distance matrices, either to quantify phonetic difference employing edit distance string metrics or to determine the degree of lexical similarity between dialects. A more limited number of dialectometric studies extends its focus to morphosyntactic variables (e.g. Jeszenszky et al. 2017, Spruit et al. 2009), typically treated as categorical variables whose values are based on the possible forms attested in dialect atlases.

The present project assumes that, in the context of Arabic studies, while the compilation of new dialect atlases is highly desirable, using available published dialect descriptions as a source of data for dialectometric analysis is also crucial to characterise dialect geography in the African region. This is due to the extensive material available, the still unclear typological classifications of certain Arabic dialects, and practical challenges in conducting new fieldwork in several areas. The existing sources documenting African Arabic dialects allow for comparison of phonological systems, inflectional morphemes, presence/absence of morphosyntactic categories and various lexical elements. Because of diverse transcriptions and methodologies employed by different scholars to record the material, the application of string metrics would in this case be problematic and such variables must be treated as categorical. Choosing variables for the computation of dialect distances and determining their categorical values, however, involves making several methodological choices whose challenging nature was already pointed out by Goebl in his seminal work on dialectometry (Goebl 1984: 33).

In this talk, I will present the current progress made in the compilation of a dataset comprising most of the Arabic dialectal varieties so far documented in Africa. The dataset currently counts 50 dialects geographically spanning from Mauritania to the Sudan and was compiled based on 46 published articles and grammars, predominantly encompassing phonological and morphological features, and to a lesser extent syntactic and lexical aspects. Based on this dataset, I will discuss some key issues related to computing dialect distances with categorical variables, chiefly: how does redundancy of features affect the calculation of dialectal differences? When dialects differ in multiple aspects concerning the same linguistic feature, how many linguistic variables should be defined? Drawing examples from the dataset, I will present alternative ways of coding related linguistic variables (e.g. the presence/absence of a gender distinction in plural subject pronouns and whether, when present, the feminine plural forms are etymologically related) and compare MDS plots and clustering dendrograms produced based on these different methods using the Gabmap web application (Leinonen et al. 2016). I will then discuss how the computation of the distance matrices is affected by these methodological choices, commenting on what seem to be the most and least problematic variable coding systems. This will also prompt a discussion on the benefits and limitations of compiling a dataset from published sources for quantifying dialect distances. Addressing these methodological issues is a step towards establishing criteria for calculating distances based

on categorical variables, relevant not only in dialectometry but also in categorising variants and representing variation on dialect atlas maps.

References:

- Goebl, Hans. 1984. *Dialektometrische Studien: anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*. Tübingen: Niemeyer. Vol. I.
- Jeszenszky, Péter, Philipp Stoeckle, Elvira Glaser and Robert Weibel. 2017. "Exploring global and local patterns in the correlation of geographic distances and morphosyntactic variation in Swiss German". *Journal of Linguistic Geography*, 5(02): 86-108.
- Leinonen, Therese, Çağrı Çöltekin and John Nerbonne. 2016. "Using Gabmap". *Lingua*, 178: 71–83.
- Spruit, Marco René, Wilbert Heeringa and John Nerbonne. 2009. "Associations among linguistic levels". *Lingua*, 119: 1624-1642.

METHODS XVIII

Methods in Dialectology XVIII

1-5 July 2024
La Trobe University
Melbourne

Special Session

Special Session

Linguistic covariation and coherence

The relationships between linguistic variables have received increasing attention in sociolinguistic research in the 21st century. Studies situated within the 'third wave' have consistently explored the inter-related nature of variables in the context of persona construction (cf. Eckert 2012), while non-ethnographic based quantitative analyses have increasingly been applied to covariation beyond vocalic chain-shifts on the community and individual level, in both real and apparent time (e.g., Beaman and Guy 2022).

Central to the quantitative analysis of sociolinguistic covariation are the distinct and multi-faceted methodological decisions required to account for the fact that the object of interest is not only multiple variables, but the relationships between them. Earlier analyses that relied on frequencies or pairwise correlations between individual variables tended to be inconclusive in their findings, often documenting some, but ultimately unsystematic, relationships between different variables (e.g., Oushiro 2016; Guy 2013; Waters and Tagliamonte 2017). With the growing availability of, and capacity to use, more powerful statistical methods, there has been a movement towards multi-dimensional and multi-variate techniques that have proven to be more fruitful in their outputs and insights. For example, Principal Components Analysis (PCA) has been used to reveal systematic covariation between New Zealand English monophthongs (Brand et al. 2021), while a combination of (PCA) and lectal lattices has successfully tracked changes in use of Swabian German dialectal features in trend and panel data (e.g., Beaman 2021; Beaman and Sering 2022).

This session highlights the proliferation and growing potential of such techniques in the analysis of covariation; the range of methodologies includes Principal Components Analysis, Multidimensional Scaling, Constrained Correspondence Analysis, and divisive cluster analysis. Their implementation provides new and exciting insights into comparatively understudied aspects of covariation. Specifically, Multidimensional Scaling is applied in the analysis of the perception of covarying variables (Sheard et al., Abstract 1). Principal Components Analysis is used both to expand research on covariation over the lifespan beyond the work on Swabian German (Sheard and Wilson Black, Abstract 2) and track (a lack of) covariation across vocalic, consonantal, and suprasegmental variables in New Zealand English (Hurring, Abstract 3). Finally, there is substantial scope to further our understanding of patterns of covarying variables in ethnically and linguistically diverse speech communities, and here Constrained Correspondence Analysis (Walker, Abstract 4) and divisive cluster analysis (Travis and Qiao, Abstract 5) are effectively implemented for two different ethnically and linguistically diverse communities in Australia and Canada, with markedly different relationships between variable patterning and ethnic background and identity.

As such, the aim of this special session is to provide a platform for discussion of the conceptual and methodological challenges faced in the analysis of linguistic covariation and coherence, and the opportunities offered by increasingly accessible and powerful statistical techniques. The session provides an opportunity for the sharing of best practice and experiences analysing covariation across different varieties, variables, structural relationships, and time scales. We hope that the panel will strengthen connections between researchers, result in the development and sharing of methodological resources, and ultimately build capacity in the field to employ robust and replicable quantitative analyses of covariation.

References

Beaman, Karen V. 2021. 'Exploring an approach for modelling lectal coherence.' in Hans Van de Velde, Nanna Haug Hilton and Remco Knooihuizen (eds.), *Language Variation–European Perspectives VIII: Selected papers from the Tenth International Conference on Language Variation in Europe (ICLaVE 10), Leeuwarden, June 2019* (John Benjamins Publishing Company: Amsterdam/Philadelphia).

Beaman, Karen V, and Gregory R Guy (eds.). 2022. *The Coherence of Linguistic Communities* (Routledge).

Beaman, Karen V, and Konstantin Sering. 2022. 'Measuring change in lectal coherence across real-and apparent-time.' in Karen V. Beaman and Gregory R. Guy (eds.), *The Coherence of Linguistic Communities* (Routledge).

Brand, James, Jen Hay, Lynn Clark, Kevin Watson, and Márton Sóskuthy. 2021. 'Systematic co-variation of monophthongs across speakers of New Zealand English', *Journal of Phonetics*, 88: 101096.

Eckert, Penelope. 2012. 'Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation', *Annual Review of Anthropology*, 41: 87-100.

Guy, Gregory R. 2013. 'The cognitive coherence of sociolects: How do speakers handle multiple sociolinguistic variables?', *Journal of Pragmatics*, 52: 63-71.

Oushiro, Livia. 2016. 'Social and structural constraints in lectal cohesion', *Lingua*, 172: 116-30.

Waters, Cathleen, and Sali Tagliamonte. 2017. 'Is one innovation enough? Leaders, covariation, and language change', *American Speech*, 92: 23-40.

.

List of contributions

1. Sheard, Elena, Robert Fromont, Josh Wilson Black, Jen Hay, Lynn Clark, and Gia Hurring. *The social meaning and perceptual accessibility of co-varying New Zealand English monophthongs*
2. Sheard, Elena and Josh Wilson Black. *Change over the lifespan across covarying New Zealand English monophthongs.*
3. Hurring, Gia. *Exploring multiplane covariation in New Zealand English.*
4. Walker, James. *Co-variation in Ethnolects*
5. Travis, Catherine, and Gan Qiao. *Using covariation to test the intersection of ethnicity and social class.*

The social meaning and perceptual accessibility of co-varying New Zealand English monophthongs

Elena Sheard[1], Robert Fromont[1], Joshua Wilson Black[1], Jen Hay[12], Lynn Clark[12], Gia Hurring[2]
[1]*New Zealand Institute of Language, Brain and Behaviour, University of Canterbury*
[2]*Department of Linguistics, University of Canterbury*

Previous research on covariation using data from the Origins of New Zealand English (ONZE) corpus has established, by means of Principal Component Analysis (PCA), that monophthongs in New Zealand English are not independent of each other, and that speakers' production of constellations of vowels varies predictably (Brand et al. 2021; Wilson Black et al. 2023). These findings have been replicated using data from the QuakeBox corpus (Hurring et al. Under review). Our project investigates whether this vocalic covariation is accessible to listeners, via a perceptual task implemented using a modified version of the browser-based Audio-Tokens toolbox (Donhauser and Klein 2023).

New Zealand English speakers (n = 117, 75% women) were presented with audio stimuli from 38 Pākehā women between the age of 46 and 55 in the QuakeBox corpus (cf. Walsh et al. 2013) that were randomly distributed across three classification groups (12 stimuli in the first, 13 in the second and third). Following the free classification paradigm, participants were asked to make subgroupings of stimuli that they perceived to sound similar and label them with free text, with no categories or dimensions of contrast specified in advance by the researchers (cf. Clopper 2008). The results of this procedure were converted to a 38 x 38 pairwise dissimilarity matrix of the proportions that each stimuli pair had not been placed in the same group, of the instances where they could have been. We implemented a two-dimensional non-metric multidimensional scaling analysis (MDS) to reduce this matrix to a set of coordinates for each stimulus and dimension (i.e., 38 x 38 to 38 x 2), that can be 'mapped' spatially and from which we can infer underlying structure (i.e., the acoustic cues used to assess similarity). The number of dimensions, which must be specified *a priori*, was determined using a custom function in R that compares the actual reduction in stress (a measure of fit) from adding additional dimensions to the reduction that would be expected based on chance.

We then fit linear regressions with the coordinates for Dimension 1 (D1) and 2 (D2) as the dependent variables. One set of models tested the relationship between D1/D2 and listener labels, with label category (based on manual coding of free text labels) fit as the independent variable. A second set of models tested the relationship between D1/D2 and speaker production. Here, we used two Principal Components (PCs) loadings for the same 38 speakers generated by Hurring et al. (Under review), which capture speakers' positions as 'innovative' or 'conservative' with respect to two sets of co-varying vowels. These PCs were included as independent variables, in interaction with stimuli speech rate and pitch. The combined results from the models show that listeners accurately perceive speaker pitch and speech rate, with slower speed and high pitch production and corresponding label categories both significantly favouring higher D2 ($p < 0.05$). Importantly, D1 is predicted by an interaction between speech rate and PC1 loadings (representing realisations of KIT, FLEECE, TRAP, NURSE, and GOOSE). While the perception of speakers with innovative vowels is less influenced by speech rate, speakers with conservative vowels have higher D1 (predicted by low SES/'rural' labels) when they speak more quickly, and lower D1 (predicted by high SES labels) when they speak slowly. Vowel production does, then, affect the perception of social characteristics, but not in insolation.

These results validate MDS as a data-driven way of approaching the perception of vocalic covariation and show how free classification and labelling can lead to an improved understanding of how listeners perceive social meaning. Speech rate and pitch emerge as primarily accessible to speakers of NZE, but certain co-varying NZE monophthongs also

appear to be secondarily accessible, socially meaningful, and interact with speech rate in affecting perceived social meaning.

References

Borg, Ingwer, and Patrick J.F. Groenen. 2005. *Modern multidimensional scaling : theory and applications* (Springer: New York).

Brand, James, Jen Hay, Lynn Clark, Kevin Watson, and Márton Sóskuthy. 2021. 'Systematic co-variation of monophthongs across speakers of New Zealand English', *Journal of Phonetics*, 88: 101096.

Clopper, Cynthia G. 2008. 'Auditory free classification: Methods and analysis', *Behavior research methods*, 40: 575-81.

Donhauser, Peter W, and Denise Klein. 2023. 'Audio-Tokens: a toolbox for rating, sorting and comparing audio samples in the browser', *Behavior research methods*, 55: 508-15.

Hurring, Gia, Josh Wilson-Black, Jen Hay, and Lynn Clark. Under review. 'How stable are patterns of covariation across time?'.

Walsh, Liam, Jen Hay, Derek Bent, Jeanette King, Paul Millar, Viktoria Papp, and Kevin Watson. 2013. 'The UC QuakeBox Project: Creation of a community-focused research archive', *New Zealand English Journal*, 27: 20-32.

Wilson Black, Joshua, James Brand, Jen Hay, and Lynn Clark. 2023. 'Using principal component analysis to explore co-variation of vowels', *Language and Linguistics Compass*, 17.

Covariation over the lifespan: New Zealand English monophthongs

Elena Sheard & Josh Wilson Black
New Zealand Institute of Language, Brain, and Behaviour, University of Canterbury

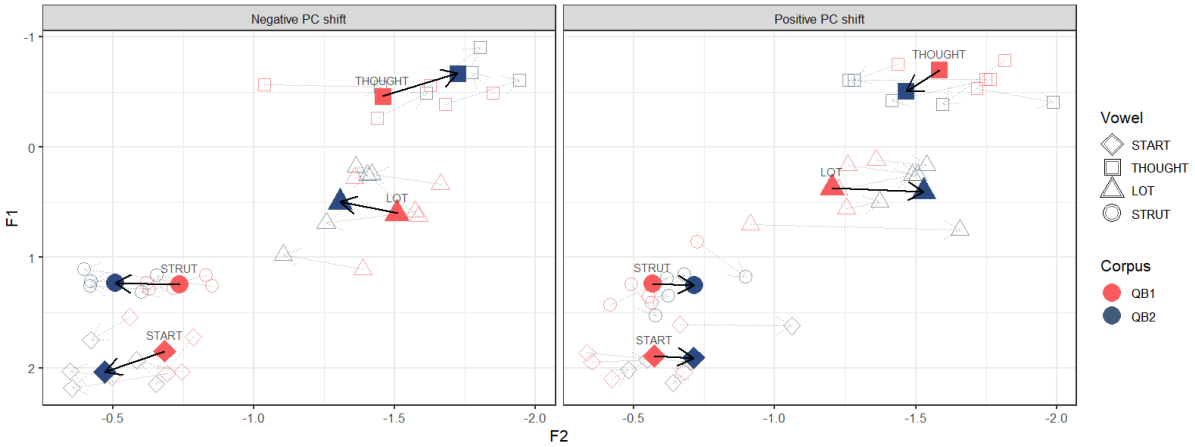
Previous research has uncovered systematic covariation in New Zealand English (NZE) monophthongs by means of Principal Component Analysis (PCA) (Brand et al. 2021; Wilson Black et al. 2023). Work currently in progress has revealed a high level of stability in the 'scores' assigned to speakers by PCA when applied to the same speakers recorded at two time points, around eight years apart (Hurring et al. Under review). This indicates that patterns of vowel covariation in the community have remained stable. In this paper, we explore how the less common instances of *instability* in PC scores over time reveal systematic change across the back vowel spaces of individual speakers (cf. Sankoff 2019).

Our data comes from the QuakeBox project (Clark et al. 2016; Walsh et al. 2013). The original project (QB1) collected recordings of earthquake stories in multiple locations across Christchurch, New Zealand in 2011-2012. The QuakeBox 2 (QB2) project then re-recorded stories from a subset of the QB1 speakers in 2019-2020. Following the workflow established in Brand et al. (2021), we analyse vowel midpoint data from the 51 speakers who satisfy data quantity and quality filtering steps at both recording points. First, Generalised Additive Mixed Models (GAMMs) are fit for the first and second formant of each vowel with gender, age (at QB1), and articulation rate as predictors (n = 22529 (QB1), n = 43403 (QB2)). PCA is then applied to the speaker random intercepts extracted from these GAMMs (51 at each time point) to reveal which vocalic variables pattern together.

At both time points, we see two significant clusters of vocalic variables. PC1 primarily tracks the NZE short front vowel shift and PC2 captures reorganisation of the back vowels. Each speaker is given a 'score' by the PCA representing their position with respect to these clusters. A speaker who has a low score for PC2 will, all things being equal, have a vowel space in which START and STRUT are fronting, and LOT and THOUGHT are fronting and raising. Hurring et al. (Under review) show a significant correlation in the PC scores of speakers when analysed at QB1 and QB2, indicating stable covarying relationships between vowels on the individual level. However, the correlation is much weaker for PC2 than for PC1.

Acoustic analysis of the vowel midpoints for THOUGHT, LOT, STRUT and START for the ten speakers with the largest shifts in PC2 scores (n = 3581) reveals that the five speakers with a negative shift in PC2 scores have raised realisations of THOUGHT over time, while the other three vowels have fronted. The five speakers with a positive shift in PC2 scores have correspondingly lowered realisations of THOUGHT and backed LOT/STRUT/START. These movements are shown in Figure 1, which depicts the average midpoint for each individual speaker between QB1 and QB2. This analysis provides support for a change over the lifespan-based explanation of the weaker correlation between QB1 and QB2 for PC2 relative to PC1. While speakers remain stable with respect to the NZE short front vowel shift (PC1), the relationship between the four back vowels significantly changes for some individuals (PC2). The results validate the use of PCA as a means of tracking both vocalic covariation in the community and revealing changes over the lifespan across covarying vowels. Our application of PCA also suggests that language change over the lifespan can encompasses systematic change in vocalic relationships.

Figure 1 Average back vowel midpoints for the ten individuals with largest shifts in PC2 scores between QB1 and QB2 (hollow shapes). The filled in shapes represent the average shifts for each group overall.



References

Brand, James, Jen Hay, Lynn Clark, Kevin Watson, and Márton Sóskuthy. 2021. 'Systematic co-variation of monophthongs across speakers of New Zealand English', *Journal of Phonetics*, 88: 101096.

Clark, Lynn, Helen MacGougan, Jennifer Hay, and Liam Walsh. 2016. "'Kia ora. This is my earthquake story". Multiple applications of a sociolinguistic corpus', *Ampersand*, 3: 13-20.

Hurring, Gia, Josh Wilson-Black, Jen Hay, and Lynn Clark. Under review. 'How stable are patterns of covariation across time?'.

Sankoff, Gillian. 2019. 'Language change across the lifespan: Three trajectory types', *Language*, 95: 197-229.

Walsh, Liam, Jen Hay, Derek Bent, Jeanette King, Paul Millar, Viktoria Papp, and Kevin Watson. 2013. 'The UC QuakeBox Project: Creation of a community-focused research archive', *New Zealand English Journal*, 27: 20-32.

Wilson Black, Joshua, James Brand, Jen Hay, and Lynn Clark. 2023. 'Using principal component analysis to explore co-variation of vowels', *Language and Linguistics Compass*, 17.

Exploring multiplane covariation in New Zealand English

Gia Hurring, *University of Canterbury*

While recent sociolinguistic research has moved towards exploring covariation in linguistic landscapes, the novelty of this line of research presents gaps in the field (Beaman and Guy 2022; Guy 2013; Nagy and Gadanidis 2022). First, there appears to be inconsistent covariation results among the literature with much of the data either one level of grammar, leaving other variables under-researched (Beaman and Guy 2022; Nagy and Gadanidis 2022). Secondly, previous studies have often lacked sophisticated method for exploring covariation, relying on modelling techniques and pairwise correlations to assume covariation over whole multivariate datasets (Guy 2013; Tamminga 2019).

Considering this second caveat, Brand et al. (2021) developed a 'model-to-PCA' method which successfully explored holistic vowel covariation in NZE. Using generalised additive mixed models (GAMMs) and Principal Component Analysis (PCA), this method allowed for multiple variables to be explored in a lower-dimensional space while being able to hold known sources of variation constant. Thus, covariation can be explored over the whole dataset in a single space. This method was later refined by Wilson Black et al. (2023) and tested by Hurring et al. (Under review) to enhance and support its methodological capabilities. However, this method has yet to be tested on non-vocalic variables, thus, leaving an open door for further investigation.

Giving all above caveats, this portion of my thesis investigates what multiplane covariation may occur in NZE (RQ1), using the more sophisticated model-to-PCA method, and whether this covariation can be sufficiently explained either theoretically and/or in context to NZE (RQ2). The data includes two 'planes' of sound – segmental and suprasegmental. This includes monophthongs, intervocalic /t/, and /s/ for the segmental plane, and rhythm, articulation rate, creaky voice, and modal pitch for the suprasegmental plane. Together, these create a multiplane dataset to both test the model-to-PCA method and explore covariation beyond the vowels in NZE.

The results find that only the vowel variables consistently cluster together with no other clusters across or within the different planes. The remaining variables are either working independently (/s/, creak, modal pitch) or they load as pairwise correlations (between two /t/ variants, and rhythm and articulation rate).

While this answers RQ1, interpreting these results to answer RQ2 requires more investigation. Post-hoc analysis was done on the data at different stages of the model-to-PCA method to see what covariation (and covariation sources) the models control for/strips away. Taken together, these analyses highlight that the PCA may be:

a) representative of uncontrolled social factors; and/or
b) representative of persona-variation (carried by the vowels); or
c) stylistic covariation (variables are too dynamic to be captured in covariation systematically).

Downfalls of the model-to-PCA method itself, and the type of data used, are also considered as possible reasons for the lack of multiplane covariation.

References

Beaman, K. V., & Guy, G. R. (2022). The coherence of linguistic communities: Orderly heterogeneity and social meaning. In *The Coherence of Linguistic Communities* (pp. 1-13). Routledge.

Brand, J., Hay, J., Clark, L., Watson, K., & Sóskuthy, M. (2021). Systematic co-variation of monophthongs across speakers of New Zealand English. *Journal of Phonetics*, *88*, 101096.

Guy, G. R. (2013). The cognitive coherence of sociolects: How do speakers handle multiple sociolinguistic variables? *Journal of pragmatics*, *52*, 63-71.

Hurring, G., Wilson Black, J., Hay, J., & Clark, L. (under review). *How stable are patterns of covariation across time?*

Nagy, N., & Gadanidis, T. (2022). Looking for covariation in heritage Italian in Toronto 1. In *The Coherence of Linguistic Communities* (pp. 106-126). Routledge.

Tamminga, M. (2019). Interspeaker covariation in Philadelphia vowel changes. *Language Variation and Change*, *31*(2), 119-133.

Wilson Black, J., Hay, J., & Clark, L. (2023). Using Principal Component Analysis to explore co-variation of vowels. *Language and linguistics compass*, *17*(1).

Co-variation in ethnolects

James A. Walker, *La Trobe University*

Ethnolects – ethnically marked ways of speaking a majority language (Carlock & Wölck 1981) – are normally identified on the basis of salient linguistic differences between groups differentiated by ethnic background. Although such differences are (implicitly) assumed to reflect transfer from heritage/community languages, tracing individual features across languages is not always possible (Labov 2008). In some cases, ethnolects are reflected in quantitative differences rather than the occurrence of salient features (e.g., Hoffman & Walker 2010). In addition, while most studies of ethnolects have focused on single variables, speakers and listeners undoubtedly rely on the co-occurrence of features to convey and infer linguistic differences for socio-symbolic purposes (Eckert 2008a, 2008b). In this paper, we consider whether studying the co-variation of multiple linguistic features can help to define an ethnolect.

Our focus is Toronto, Canada's largest city and one of the most multicultural cities in the world. Although heritage languages are maintained to different degrees, all groups are shifting to English and there are anecdotal reports of ethnolects, especially in in-group contexts. While individual ethnolectal features have been investigated (e.g. Hoffman 2010; Hoffman & Walker 2010), we do not know how multiple co-varying features act together to constitute a 'lect'.

An earlier study (Walker et al. 2022) examined co-variation of vowel features, but this paper carries the analysis forward by investigating co-variation among vowel and consonant variables in a corpus of spoken English in Toronto. From sociolinguistic interviews conducted by in-group community members with 23 speakers of different backgrounds (British/Irish, Chinese, Italian, Portuguese), we extracted a representative sample of tokens for six variables representing stable variation, ongoing changes and potentially ethnolectal features: Canadian Raising of the onset of the diphthongs /aw/ and /ay/ before voiceless consonants; the Canadian Vowel Shift (CVS), the retraction and lowering of the front lax vowels /ɪ/, /ɛ/ and /æ/; the fronting of /o/, /u/ and /aw/; the place of articulation of word-final unstressed *-ing* (ING); stopping of word-final velar nasals (NK); and deletion of word-final /t, d/ in consonant clusters (TD). Co-variation was investigated using Constrained Correspondence Analysis (CCA; Meyerhoff & Klaere 2017), a multivariate method for measuring the relative contribution of each variable to a smaller number of underlying factors. Preliminary analysis of the data shows that individual speakers can be positioned within a space defined by these factors, correlating to some extent with responses to the Ethnic Orientation (EO) questionnaire administered as part of the interview. These results suggest that ethnolects are not well- defined but rather reflect considerations of identity and context.

References

Carlock, E. & Wölck,W. 1981. A method for isolating diagnostic linguistic variables: The Buffalo ethnolects experiment. In D. Sankoff & H. Cedergren (eds.), *Variation omnibus*. Edmonton: Linguistic Research Inc. 17–24.

Clyne, M., Eisikovits, E. & Tollfree, L. 2001. Ethnic varieties of Australian English. In D. Blair & P. Collins (eds.), *English in Australia*. Amsterdam: John Benjamins. 223–238.

Eckert, P. 2008a. Variation and the indexical field. *Journal of Sociolinguistics* 12: 453–476.

Eckert, P. 2008b. Where do ethnolects stop? *International Journal of Bilingualism* 12: 25-42.

Hoffman, M.F. 2010. The role of social factors in the Canadian Vowel Shift: Evidence from Toronto. *American Speech* 85:121–140.

Hoffman, M.F. & Walker, J.A. 2010. Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language Variation and Change* 22: 37-67.

Labov, W. 2008. Mysteries of the substrate. In M. Meyerhoff & N. Nagy (eds.), *Social lives in language — Sociolinguistics and multilingual speech communities*. Amsterdam: John Benjamins. 315–326.

Meyerhoff, M. & Klaere, S. 2017. A case for clustering speakers and linguistic variables: Big issues with smaller samples in language variation. In I. Buchstaller & B. Siebenhaar (eds.), *Language Variation – European Perspectives VI: Selected Papers from the Eighth International Conference on Language Variation in Europe (ICLaVE 8)*. Amsterdam: John Benjamins, 22–46.

Walker, J.A., Hoffman, M.F. & Meyerhoff, M. 2022. What's in a lect? Coherence in phonetic and grammatical variation. In K. Beaman & G.R. Guy (eds.), *The coherence of linguistic communities: Orderly heterogeneity and social meaning*. London/New York: Routledge, 71-86.

Using covariation to test the intersection of ethnicity and social class

Catherine Travis & Gan Qiao, *Australian National University*

Ethnicity has long captured the attention of sociolinguists, with extensive examination of the ways in which ethnic minorities respond to patterns of variation and change in the wider community, as well as how they use language to mark ethnic orientation (e.g., Hoffman & Walker 2010; Labov 2001). Such work tends to foreground ethnicity as a defining feature and consider it independently of other social factors. In this talk, we explore patterns of crossspeaker covariation in a set of variables in Australian English to demonstrate a fundamental intersection between ethnicity and social class.

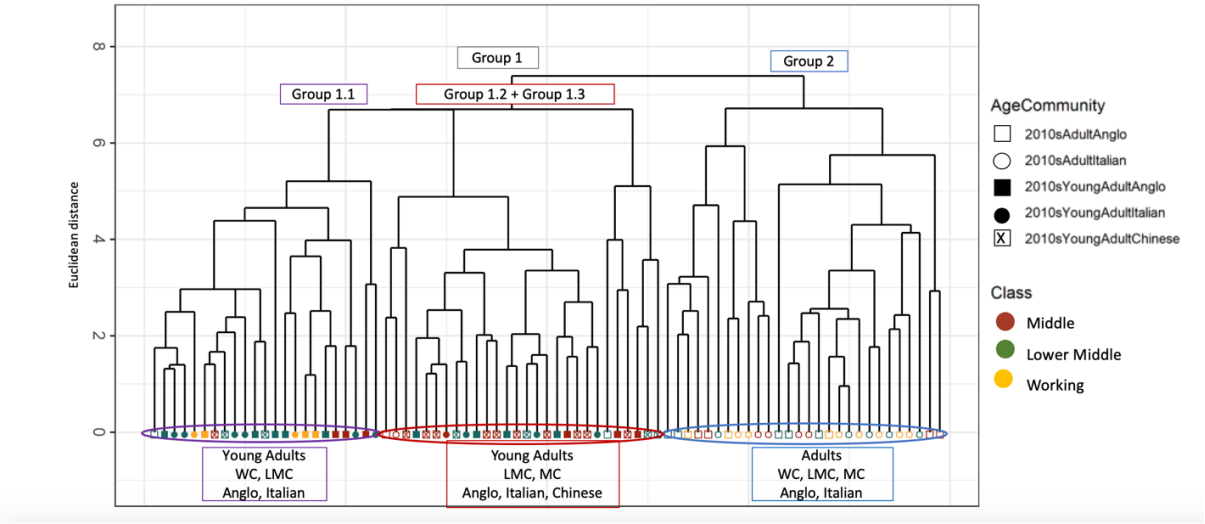
The data come from a large-scale study that brings together sociolinguistic interviews with 230 Australians recorded in Sydney at two time points, in the 1970s (Horvath 1985) and the 2010s. The speakers represent some of Australia's largest ethnic groups—Anglo-Celtic, Italian, Greek and Chinese Australians—stratified according to age, gender, and social class. Transcriptions totalling some 1.25 million words provide large amounts of data for linguistic analysis, in which to explore patterns of cross-speaker covariation.

Previous work on this dataset has found that apparent differences across ethnic groups rarely correlate with measures of ties to ethnic heritage, such as social network, language proficiency, or ethnic orientation. However, there is evidence of a strong intersection between ethnicity and social class. For instance, Chinese Australian young adults in the 2010s tend to orient towards overtly prestigious norms, with patterning most closely resembling that of middle-class Australians (e.g., Grama & Travis & Gonzalez 2021; Qiao & Travis 2022; Travis & Grama & Purser 2023). Here, we test the broader applicability of this intersection by asking not how different social groups pattern linguistically, but how speakers with similar linguistic behaviour are grouped socially (cf., Horvath & Sankoff 1987).

We conduct independent regression analyses of a set of variables (FLEECE and FACE vowels, prevocalic *the*, *-ing*, word-final *-er*, existential *there's*, modals of obligation, quotatives), analysing linguistic predictors only, without any pre-assigned social categories. We extract the speaker random intercepts from each model, as a metric against which to meaningfully compare across speakers (cf., Drager & Hay 2012), and then conduct Divisive Cluster Analysis based on these random intercepts to produce clusters of speakers who pattern similarly linguistically (cf., Haddican et al. 2021). Finally, we identify the social correlates of each cluster. We find that there is no clustering based solely on ethnicity, as ethnic minorities cluster first and foremost with their age cohort, and then according to social class. This can be seen in Figure 1 for the 2010s data, where Chinese Australians cluster with Middle- and Lower Middle-Class Anglo and Italian Australians, while Working Class Australians of these latter groups cluster separately.

By considering the co-variation across speakers to identify emerging social clusters we are able to empirically demonstrate the intersectionality of age, ethnicity, and social class not just for single variables, but across sets of variables, allowing us to highlight that the best way to understand ethnicity is to look at its intersection with other social factors.

Figure 1 Dendrogram of a DIANA cluster analysis (age, ethnicity, social class); 2010s (79 speakers) 9 variables: FACE, FLEECE, *-er, (ing), there's, the+Vowel, have got to, need to, quotative be like*



References

Drager, Katie and Jennifer Hay. 2012. Exploiting random intercepts: Two case studies in sociophonetics. *Language Variation and Change* 24(1): 59-78. https://doi.org/10.1017/S0954394512000014

Grama, James, Catherine E. Travis and Simon Gonzalez. 2021. Ethnic variation in real time: Change in Australian English diphthongs. In Hans Van de Velde, Nanna Haug Hilton, and Remco Knooihuizen (eds), *Studies in Language Variation*, 292-314. Amsterdam: John Benjamins. https://www.jbe-platform.com/content/books/9789027259820-silv.25.13gra

Haddican, Bill, Michael Newman, Cecelia Cutler and Christina Tortora. 2021. Aspects of change in New York City English short-a. *Language Variation and Change* 33(2): 135-163. https://doi.org/10.1017/S0954394521000120

Hoffman, Michol F. and James A. Walker. 2010. Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language Variation and Change* 22(1): 37-67. https://doi.org/10.1017/S0954394509990238

Horvath, Barbara. 1985. *Variation in Australian English: The sociolects of Sydney*. Cambridge: Cambridge University Press.

Horvath, Barbara and David Sankoff. 1987. Delimiting the Sydney speech community. *Language in Society* 16(2): 179-204. https://www.jstor.org/stable/4167835

Labov, William. 2001. *Principles of linguistic change: Social factors*, vol. 2, 3 vols. Oxford: Blackwell.

Qiao, Gan and Catherine E. Travis. 2022. Ethnicity and social class in pre-vocalic *the* in Australian English. In Rosey Billington (Ed.), *Proceedings of the Eighteenth Australasian International Conference on Speech Science and Technology* (pp. 56-60): Australasian Speech Science and Technology Association. https://sst2022.files.wordpress.com/2022/12/qiao-travis-2022-ethnicityand-social-class-in-pre-vocalic-the-in-australian-english.pdf

Travis, Catherine E., James Grama and Benjamin Purser. 2023. Stability and change in (ing): Ethnic and grammatical variation over time in Australian English. *English World-Wide* 44(3): 435-469. https://doi.org/10.1075/eww.22043.tra