Chapter 2

Heritage languages as new dialects

Naomi Nagy

University of Toronto, Department of Linguistics

In order to compare heritage and homeland varieties, to determine whether the heritage varieties constitute new and distinctive dialects, we need innovative methods and a cohesive definition of "new dialect." Toronto's Heritage Language Variation and Change Project provides testing grounds for both: it is designed for intergenerational, cross-linguistic, and diatopic (heritage vs. homeland varieties) analysis of spoken Cantonese, Faetar, Italian, Korean, Russian and Ukrainian. With reference to the heritage varieties examined in this project, I contrast ways of defining new dialects. I then describe methodological innovations that permit variationist analysis of linguistic patterns and the involvement of large numbers of student-researchers who are speakers of the putative new dialects, two elements critical to the success of the project.

1 Introduction

The XVth meeting of *Methods in Dialectology* sought to "bring traditional approaches to dialectology together with the latest advances in data collection technologies, new analysis instruments, and new interpretations of the concept of dialect".[1]

This paper compares interpretations of the concept of 'dialect', and particularly of dialect divergence, contrasting the outcomes of linguistically- and socially-oriented approaches. It then describes some advances in methods applied in a multilingual speech corpus project whose goal is to understand the process of divergence of heritage varieties from their homeland counterparts.

The study of dialect convergence [dc] and divergence [dd] therefore needs to be informed by both subdisciplines [historical linguistics and sociolinguistics]… Research into dc and dd lies at the crossroads between contact linguistics and variationist linguistics, i.e. between the study of language change as a result of

---

[1] http://methodsxv.webhosting.rug.nl/

language contact and the study of language variation as a synchronic manifestation of language change…" (Auer, Hinskens & Kerswill 2004b: 16).

The Heritage Language Variation and Change Project (HLVC, Nagy 2011) is, in fact, motivated by the complications of intersecting contact linguistics and variationist linguistics, as in the above quotation. We develop and use a multilingual corpus for inter-generational, cross-linguistic, and diatopic (heritage vs. homeland varieties) comparisons in order to develop generalizations about the types of variable features, structures or rules that are borrowed earlier and more often in contact contexts.

The ultimate goal is to better understand what happens in contact situations and what the best predictors are of different linguistic outcomes. For this purpose, a set of consistent methods are applied to a set of linguistic variables that are found in a set of HERITAGE LANGUAGES (HLs) spoken in Toronto, Canada. HLs are defined, in the Canadian context, as mother tongues other than Canada's two official languages (French and English), cf. Cummins & Danesi (1990).[2] I will first discuss whether such varieties may be considered new dialects, contrasting definitions based on linguistic factors and attitudes, and then describe some innovations developed in this inquiry.

1.1 When do new varieties constitute new dialects?

While we are all familiar with the maxim that "a language is a dialect with an army and a navy," it is surprisingly difficult to find viable definitions of what constitutes a dialect. From my admittedly outsider's perspective, dialectologists' definitions may be based on structural features and/or community orientation toward the language. Trudgill (1986; 2004) focuses on linguistic effects, that is, types of features or changes in the language, denying the relevance of attitudinal factors to the concept of new dialect formation (see Meyerhoff 2006: 186 for further discussion). Schneider (2003; 2007), in contrast, focuses more on orientation of the community toward the language, while also including linguistic features, in a model designed to describe the trajectory of post-colonial varieties of English. I have also found Auer, Hinskens & Kerswill's (2004) edited book (hereafter AHK) thought-provoking as I consider how HLs may fit into the discussion of new dialect formation. In order to focus on convergence and divergence as particular

---

[2] Mother tongue is "the first language learned at home in childhood and still understood by the person at the time the data was collected. If the person no longer understands the first language learned, the mother tongue is the second language learned. For a person who learned two languages at the same time in early childhood, the mother tongue is the language this person spoke most often at home before starting school…" (Statistics Canada 2014).

aspects of dialect change, they must grapple with the question of what a dialect is. Their basic definition excludes HLs outright:

> We will use the notion of 'dialect' to refer to a language variety which is used in a geographically limited part of a language area in which it is 'roofed' by a structurally related standard variety (Auer, Hinskens & Kerswill 2004b: 1).

This would exclude heritage varieties from being considered dialects of their parent language. Indeed, any diasporic variety cannot be considered a dialect of its homeland language, unless the emigrés land in a country where the same language is spoken. Other aspects of their definitions would seem to be hospitable to the inclusion of HLs as new dialects (discussed in §2).

1.2 How are heritage languages like new dialects?

We begin by situating this study in the Canadian context. Few would question whether Canadian English and Canadian French constitute different dialects from their European counterparts. The varieties of French and English spoken in Canada have been explicitly labeled as distinct varieties for longer than Canada has been a nation. Canadian English has been labeled as a distinct dialect of English since at least 1857, when The Rev. A. Constable Geikie titled a speech he read before the Canadian Institute "*Canadian English.*" Bouchard (1998 [2002]) proposed that a grammatical debate in 1840–41 between Abbé Maguire, Jérôme Demers and Michel Bibaud marked the transition from considering "French spoken in Canada" to the development of the concept of Canadian French. A few years later, in an epistolary novel, Coursen (1846) referred to "the French Canadian dialect," extending the label beyond academic discourse. It is thus possible for people to label as new dialects the varieties of national languages spoken by immigrant groups.

But what of languages that do not enjoy official recognition in Canada? Languages without official status are not named in government documents. I am not aware of academic recognition of these varieties. For example, there is no Cantonese parallel to *The Canadian Oxford Dictionary* or the university course *Canadian English.* To check for less formalized references to HLs as new dialects, I collected online citations paralleling "Canadian French" and "Canadian English" for HLs spoken in Toronto. This is an effort to capture early evidence of the emergence of named status for these varieties, looking for the modern equivalents of Geikie's speech or Coursen's novel. Dialects can be arrayed on a continuum from least to most recognized, as listed in the top of Table 1.

In the first row of Table 1, check marks indicate the existence of this status of recognition, i.e., Named variety, for each language, while blank cells indicate that no such evidence has been found. The next rows of part (a) indicate additional levels of recognition (outside the HLVC project) enjoyed by some of Toronto's HLs. These constitute evidence of new dialects on normative or attitudinal grounds and will be discussed in §2. Table 1 (b) summarizes the linguistic status of differentiation of these varieties from their homeland counterparts. "S" indicates that variationist analysis has found the same pattern of variable usage in homeland and heritage varieties of that language, while "D" marks documented differences between homeland and heritage varieties, that is, evidence of (partial) formation of a new dialect, based on linguistic criteria. Blank cells remain to be filled in by future work, which will also include additional variables. Table 1 (c) shows each community's average Ethnic Orientation score (explained below). A cursory comparison of the top two parts of the table indicates a lack of relationship between these two ways of considering whether a new dialect has emerged. We also see no connection with the community's degree of attitudinal separation from their homeland. These incongruities point up a problem with deterministic approaches to new dialect formation where we might expect similar outcomes across all languages, if social factors weren't relevant.

At this point, we can conduct the same comparative exercise with Schneider's five phases of new dialect formation, with the same unsatisfying lack of convergence in outcome. For this discussion, I refer numerically to the four types of markers laid out in Schneider's (2003: 255) Table 1: (1) History and politics, (2) Identity construction, (3) Use/attitudes, and (4) Linguistic developments/structural effects.

All HLs included in the HLVC project have undoubtedly made it to Phase 1 on all counts. They exhibit Phase 2 markers for (3) Use/attitudes (acceptance of original norm) and (4) Linguistic developments (lexical borrowing, cf. Danesi (1983) for Italian, but not for (1) or (2). (2)has been explicitly probed by the HLVC project, with the results shown in Table 1 (c). Speakers are asked, "Do you think of yourself as Italian, Canadian or Italian-Canadian?" (*mutatis mutandis* for each language). Open-ended responses are quantified on a scale in which a homeland-oriented response (e.g., "Italian") scores two points while "Canadian" scores 0, with mixed responses scoring 1. In the first generation, all language groups average near 1.5, quite homeland oriented. Differences emerge in the second generation, painting the picture in Table 1 (c). Thus the HL communities straddle the IDENTITY construction definitions for Schneider's Phases 1-4.

Toronto's HLs have reached Phase 3 in terms of (2) (3) markers though, against expectation, not (4): our project has uncovered very little structural spreading

Table 1: Comparison of (a) status-related, (b)structural-related, and (c) attitudinal indicators of new dialect formation for a sample of Toronto's heritage languages. Faetar is omitted from section (c) of this table. The trilingual nature of the community eludes my quantification.

	Faetar	Korean	Cantonese	Russian	Italian	Ukrainian
(a) Status of recognition						
Named varieties			√		√	√
Social or demographic attributes ascribed to the variety						√
Linguistic features of variety described				√	√	√
Systematic quantitative analysis of linguistic variation						√
(b) Heritage – Homeland comparison of linguistic features						
Basic vocabulary (Nagy 2011)		S				
Voice Onset Time (Nagy & Kochetov 2013; Kang & Nagy 2013)		D		D	S	
Null vs. pronoun subjects (Nagy 2015; Nagy & Iannozzi 2014)	D			D	S	
(c) Orientation toward heritage nation *vs.* Canada						
0 = "I am Canadian," 1 = mixed, 2 = "I am Korean/etc."	n/a	1.3	1.0	1.6	1.4	1.1

from English to the HLs (see references in Table 1). We find scattered evidence of the "complaint tradition" that constitutes an Attitude marker of Phase 3: e.g., Struk (2000) anticipates the "total extinction" of Ukrainian in (Alberta) Canada due to massive English influence and the negative views of Canadian Cantonese cited in §2.2.

Schneider's model unravels a bit more at Phase 4, where HL speakers in Toronto exhibit markers for (2), as noted above, but none of the other markers of endonormative stabilization. A return to this question when more linguistic features of the HLs have been analyzed will be critical.

2 A little more about Toronto's HLs as new dialects

The following sections describe the status of Toronto's HLs in greater detail, grouping information according to the recognition characteristics in Table 1 (b). Relevant suggestions in AHK are evaluated as they apply to the status of Toronto's HLs. A lack of correspondence between the rankings in Table 1 (b) by linguistic features, à la Trudgill, and by orientation, à la Schneider, in Table 1 (a, c), will be evident. This underscores the inappropriateness of equating one language to one culture, or monolithic descriptions of either (cf. Foley 2005).

2.1 No status as dialects

Searching the web, including academic resources, yielded no hits for "Canadian Korean" or "Canadian Faetar". We are aware of no published descriptions of these varieties, or claims of them as dialects distinct from their homeland varieties. Both have been spoken in Toronto since about the middle of the twentieth century, but have never had large numbers of speakers.

2.2 Named varieties

Speakers of Cantonese outnumber speakers of Korean by more than 10:1 in Toronto (Statistics Canada 2011a), although the Cantonese arrived in the city only about one decade earlier. In the five years between the 2006 and 2011 census, there was an increase of almost 10% in the number of people of Chinese ethnic origin living in Toronto (Statistics Canada 2011b), the majority of whom likely speak Cantonese.[3] While not recognized at the institutional level, "Canadian

---

[3] Imprecise because many respondents indicating that they speak "Chinese" without specifying their variety.

Cantonese" has gained the status of a distinct dialect among (some) community members. The variety is named, as in posts such as these, at http://www.gamefaqs.com/boards/981401-sleeping-dogs/63775307:

> what bothers me, is that it's not authentic cantonese, but *canadian cantonese*. huge difference (#2BloodyBooger, Posted 8/18/2012, emphasis mine)

> Some of the accents are terrible, you can tell they're *Canadian cantonese* speakers. On the other hand, I personally know a lot of people who have both English and Cantonese as their mother tongues, Queen's English accents and all (myself being one of them), and sometimes when we speak, we tend to mix in English words or vice versa to get our point across (ZeroHiei, Posted 8/18/2012, emphasis mine).

Struk (2000: 71) describes Ukish, "a mixture of Ukrainian and English." Italian also exists as a named variety, cf. the article "*Canadian Italian* as a marker of Ethnicity" (Danesi 1983; 1984). Giovanardi, Gualdo & Coco (2003) label the variety as *Italiese*. The varieties that have been named are distinguished by larger numbers of speakers, tentatively a necessary, but not sufficient condition for dialect identification.

2.3 Social or demographic attributions ascribed

Kerswill & Trudgill (2004) propose a characteristic not directly related to linguistic structure: a new dialect is a variety which lacks a "local stable model" and thus cannot be transmitted. This definition does not seem to apply to the HL context because transmission is certainly attested in our corpus of HL speakers of up to five generations since immigration. Many heritage community institutions offer language classes which adhere to what is considered a stable homeland model. Because of, or perhaps in spite of these courses, the heritage variety is transmitted.

Canadian Ukrainian, however, is well-enough established to have a Wikipedia (2014) entry:

> *Canadian Ukrainian* […] is a dialect […] specific to the Ukrainian Canadian community descended from the first two waves of historical Ukrainian emigration to Western Canada. […] Canadian Ukrainian was widely spoken from the beginning of Ukrainian settlement in Canada in 1892 until the mid-20th century. […] cut off from their co-linguists by wars and social changes,

and half the globe [...] exposed to speakers of many other languages in Canada, especially English. [...] introduced to many new technologies and concepts, for which they had no words. Consequently Canadian Ukrainian began to develop in new directions from the language in the "Old Country."

Here demographic information is presented to bolster the status of the new variety: when, where and why the language emerged, and the circumstances that encouraged its divergence.

2.4 Linguistic features described

Although not all have achieved the status of being named varieties, the heritage versions of Russian, Italian and Ukrainian are recognized as valid objects of linguistic study, having been spoken in Toronto for over a century (about twice as long as the others in Table 1). They are the object of descriptions with less negative connotations than are found in the above descriptions of Canadian Cantonese. For example, in a website called "Canadian dialects of European languages," a Canadian Russian dialect is described, but not named (Language Factory 2013):

> Canada's Doukhobor community, especially in Grand Forks and Castlegar, British Columbia, has kept its *distinct dialect of Russian*. It has a lot in common with South Russian dialects, showing some common features with Ukrainian.

This site also mentions Heritage Ukrainian, but no other languages in the HLVC project. The Wikipedia extract about Ukrainian (§2.3) also includes linguistic description. It explicitly mentions linguistic features that distinguish the Canadian dialect from the European dialect of Ukrainian. This variety is well-enough established that there are also published descriptions of phonetic and syntactic variation in the heritage variety, cf. Hudyma (2011), Struk (2000). Danesi (1983; 1984) describes lexical features of Canadian Italian, but claims that it is not grammatically or phonologically distinct from its homeland counterpart:

> From all structural points of view it is essentially Peninsular Italian, *i.e.*, in its phonology [...], morphology [...] and Syntax [...], it is identical to Peninsular Italian, or to any of its regional and dialectal variants. In its lexical repertoire, however, it contains many new words ...

In addition to ascribing specific linguistic features, we could also seek *types* of features in our quest for testable means of identifying new dialects. For example, Auer, Hinskens & Kerswill (2004b: 1), cite Chambers & Trudgill (1998: 5): "a dialect typically displays structural peculiarities in several language components." It goes without saying that "peculiarities" are subjectively defined and that this will be tautologically true if a minority variety is compared to a mainstream variety. This definition is at odds with others offered in AHK. For example, Berruto (1995, cited in Auer, Hinskens & Kerswill 2004b: 11) notes that dialects lose their "oddest features," e.g., loss of certain word order options or prodrop optionality. Similarly, Kerswill & Trudgill (2004: 198) suggest that the leveling process which contributes to new dialect features includes simplification. For example, "invariable word forms, as well as the loss of categories such as gender, the loss of case marking, simplified morphophonemics (paradigmatic leveling), and a decrease in the number of phonemes," stipulating that

> Mixing, leveling, and simplification are the necessary precursors of new-dialect formation. Together, they can be said to constitute *koineisation* (Auer, Hinskens & Kerswill 2004b: 199).

These contradictory definitions may have led to Auer & Hinskens' (2004: 356) summary statement that the connection between variation and change is still unknown. The HLVC project has not yet documented any examples of these types of changes.

2.5 Quantitative analysis of linguistic variation

Establishing the existence of distinct linguistic features of a variety is not sufficient for understanding the diachronic process of new dialect formation. Auer, Hinskens & Kerswill (2004b: 6) suggest, rather, that the patterns of use such features, or the use of "different features more often," is what constitutes dialect distinctions. This is at the heart of the comparative variationist methods (cf. Cacoullos & Travis 2010) applied in the H eritage Language Variation and Change (HLVC) Project. It requires a focus on distributional patterns and conditioning effects, rather than a simpler test of presence vs. absence of certain structures or forms. In a similar vein, Kerswill & Trudgill (2004: 215) note, as part of a series of steps that define new dialect formation, that children of immigrants will have lots of variation. These promising approaches require further quantification – comparing across varieties used by different groups, what does it mean for a group to have "more variation"? Is it simply a larger number of surface forms? How is

that compatible with the processes of simplification that are reported to accompany diffusion of linguistic patterns discussed in §2.4? Can we develop metrics to compare the degree of variation at lexical, phonetic, structural and discourse levels? Until such methods are in place, these definitions also cannot serve as diagnostics of whether a variety constitutes a new dialect. Furthermore, appropriate data will be needed. Beyond the HLVC output,[4] I am aware of no quantified descriptions of variation in Toronto's HLs except Ukrainian (Budzhak-Jones 1994; Chumak-Horbatsch 1987).

2.6 Summary: identifying HLs as new dialects

This survey has illustrated possibilities for recognition of HLs as "diverged" dialects of their homeland variety, ranging from a complete lack of recognition of a distinct dialect (Faetar, Korean) to naming of the transported variety ("Canadian Cantonese," "Italiese," "Ukish"), to attribution of social and linguistic features of the distinct variety (Italian, Russian), and finally to systematic data analysis to substantiate claims of distinct grammars (Ukrainian). As comparable homeland data become available, the HLVC project will be able to investigate both linguistic and attitudinal features for the difference in degrees of recognition of Toronto's heritage varieties as distinct "new Canadian" dialects. The remainder of the paper introduces the project's methods designed to achieve these goals.

3 The HLVC Project

The HLVC project intertwines descriptive and theoretical goals – so that we can answer the question of whether a variety has achieved "new dialect" status on both *linguistic* and *attitudinal* grounds. We document HLs as spoken by immigrants and two generations of their descendants living in the Toronto area. The three-generation model allows for direct application of models such as Trudgill's (1986) model which offer different roles for speakers of each generation. We are building a corpus of transcribed conversational speech, accompanied by relevant information about the speakers' linguistic habits, attitudes, and experiences, available to interested researchers. Our theoretical goals include better understanding of the relationship between language variation and change, to be achieved by pushing variationist research beyond its monolingually-oriented core. A variety of new tools and techniques have been developed to integrate lesser-documented varieties into the variationist tradition.

---

[4] http://projects.chass.utoronto.ca/ngn/HLVC/1_5_publications.php

These developments have stemmed from my being trained in sociolinguistics by a graduate program with a focus on methods useful for investigating well-documented languages (such as English, French and Spanish)[5] by native-speaker, or at least, very fluent fieldworkers and analysts. This approach was at odds with other aspects of my training in formal linguistics which featured data and examples from many lesser-documented languages. This contradiction came to the fore when these two streams of training merged in a dissertation documenting and theorizing variation in Faetar (Nagy 1996), a language that had been subject to little previous description, none quantified or theorized. I was a non-speaker of the variety at the outset of fieldwork. So, some twenty years later, what have I done to modify tools and approaches as I continue in this vein of applying quantitative variationist methods to lesser-studied varieties? How can we best test whether the sociolinguistic generalizations that have emerged from the study of well-documented languages apply more universally?

An important component of the HLVC project is to use the same methods to describe the variable patterns of both homeland and heritage varieties before trying to answer the question of whether the heritage varieties constitute new dialects or not. Innovations developed to allow for parallel analyses of more- and less-documented varieties include:

- integrating transcription, coding and extraction of sociolinguistic variables in ELAN;

- automated forced alignment and formant extraction for languages beyond English;

- a web map with voice clips as examples of the varieties, accessible to non-linguists;

- integration of research and teaching in courses for undergraduate and graduate students, by paid and volunteer research assistants, and by students and professors in nine countries (so far);

- sharing and training for methods, tools, instruments developed in this project and controlled sharing of data.

It is hoped that this project may help predict the future of (these) dialects and advance the study of dialects more generally and that the following brief descriptions of these innovation may prove useful in that endeavor.

---

[5] Nagy & Meyerhoff (2008) found that studies of these three languages constituted some 98% of variationist studies published in two leading sociolinguistic journals.

3.1 HLVC methods of data collection and organization

While generational differences are unquestionably an aspect of new dialect formation, sociolinguists have established that other factors are also necessary for the accurate description of linguistic variation. This necessitates a socially-stratified sample and quantitative analysis that considers the effects of multiple conditioning factors. Addressing this first need, the HLVC project has developed a sampling protocol that uses convenience sampling to recruit and record participants as follows.

For each language, a particular geographic region or city of origin is specified and all speakers in the corpus trace their ancestry to that one locale. This is meant to reduce one parameter of variation in the data, though it allows for variation in both the founder population and successive generations. For Italian, for example, all speakers in the corpus are (descendants of) Calabrese, selected because it is one of the two largest regionally defined groups of Italians in Toronto. Calabria is a region in the south of Italy where 25% of the population currently report speaking either in Italian or in Calabrese (an Italian dialect) and an additional 10% report speaking in Calabrese (ISTAT (Istituto Nazionale di Statistica) 2007: Tavola 10).

Within each language, speakers are selected to fill cells representing all combinations of generation, age, and sex criteria, defined as follows.

Generation:

- GENERATION 1 speakers are born in the home country and moved to Toronto after age 18. They have subsequently been in Toronto 20+ years.

- GENERATION 2 speakers are born in Toronto or came from the home country before age 6. Their parents are in Generation 1.

- GENERATION 3 speakers are born in Toronto. Their parents are in Generation 2.

Age: Four age groups per generation: 60+, 39-59, 21-39, <21.[6]

Sex: Two males and two females represent each age by generation cell.

Our target sample for each language comprises 40 speakers (two speakers of each sex per age group per generation). However, we have only two generations

---

[6] The two youngest groups do not exist for Generation 1, who are older than 38 by definition. Otherwise, age and generation are orthogonal in the design.

for Korean (too recently arrived) and Faetar (population too small to produce a third generation). Additionally, we have representation of Generations 4 and 5 for Ukrainian, and pilot samples for Hungarian and Polish. Currently the corpus includes transcribed recordings for 190 Heritage speakers, across eight languages.

Sociolinguistic research in the variationist paradigm has found that changes in progress are frequently linked to certain patterns of variation, allowing us to use synchronic variation as a tool for understanding change (Bailey et al. 1991; Labov 2001; 2007). In addition to the factors Generation, Age and Sex, we collect Ethnic Orientation information via an oral, open-ended questionnaire which allows us to consider the effects of (self-reports of) speakers' language practices, attitudes and experiences.[7]

The effects of these factors, and, in turn, their ability to help us understand ongoing changes in the variety, are best interpreted through the Comparative Variationist Analysis approach (cf. Labov 1972; Tagliamonte 2006; Walker 2010). Thomason & Kaufman (1988: 111) point out the vexing issue that once contact has occurred, it may not be easy to access the pre-contact variety, yet contrasting these is crucial. Cross-group comparison, an essential component of the approach, allows us to address issues that would ideally be resolved by comparing the pre-contact variety to its post-contact variety. This method involves comparison of rates of forms, as is typical in experimental approaches, but also compares conditioning effects. This approach, with its accumulated knowledge of synchronic patterns that often signal change, augmented by contrasting speakers with greater and lesser contact with English, provides a fast-track view of language change. Rather than contrasting elusive "pure" contact and non-contact varieties, the HLVC project seeks gradually increasing effects on HLs correlating to gradually increasing contact with English, to address these questions sequentially:

1. What aspects of the language vary?

2. How does the variation differ by community? Can we point to specific demographic or attitudinal differences as predictors?

3. Do the patterns of variation suggest that there is change away from the homeland variety? As Thomason (2001) notes, this requires fieldwork and parallel methods in the home countries as well.

---

[7] http://projects.chass.utoronto.ca/ngn/pdf/HLVC/short_questionnaire_English.pdf

Responses to these questions are, so far, based on small samples (see details in publications at http://projects.chass.utoronto.ca/ngn/HLVC/1_5_publications.php). Once the corpus is complete, we will return to the question of whether the quality and/or quantity of change is sufficient to meet a definition of a "new dialect."

To prepare data for this approach, we collect samples of about one hour of conversational speech from each participant, using sociolinguistic interview methodology (Labov 1984). We transcribe the recordings and then code many instances of each variant of phonetic, lexical and structural variables. We have developed an integrated approach for time-aligned orthographic transcribing and coding tokens (instances) of dependent variables as well as the predictors or independent variables in a single file (detailed in Nagy & Meyerhoff 2015). This provides seamless connections between recording, transcript, and coding of the dependent variable (response) and independent variables (predictors), facilitating revision and intercoder reliability testing. In a project that relies on a large and changing team of student researchers, this tight connection between representations of the data at various stages of analysis is imperative. It also allows for the reuse of contextual factor coding (e.g., style, topic, interlocutor) as well as some structural (morphological, syntactic) tags in successive projects. An additional advantage is the archivability of all mark-up related to each data file in a consistent manner in small files, again particularly useful in a large project where different researchers conduct different stages of the work.

Time-aligned transcription also allows us to test the feasibility of using various automated processes which have been developed for better-documented languages, such as forced alignment (of transcription to sound at a segmental level), vowel formant extraction, speech rate calculators which consider amplitude variation, and VOT measures. Preliminary results are promising and suggest that these will be immensely time-saving approaches for analysis of large data sets (Tse & Nagy 2014).

3.2 Integrating research and teaching in HLVC

The inclusion of student-researchers who are speakers of these HLs make it possible to investigate this range of languages. No one researcher can be a native-speaker, let alone expert, in this range of HLs, making the integration of research

and teaching an essential and productive component of the HLVC project.[8] Class-based activities that work with HLVC data encourage the development of critical thinking, writing skills, oral presentation and research methods, affording a more unified focus on research. In turn, the project benefits from insights and innovations from students with differing degrees of familiarity with the communities. I have structured a successful first-year undergraduate course around the premise that the students, as a group, will prepare an article for the journal *Heritage Languages*, about the ethnolinguistic vitality of heritage languages spoken in the Greater Toronto Area and the way the languages are spoken. For this purpose, the course introduces them to definitions of "heritage language"; the concepts of ethnolinguistic vitality, the status of heritage languages, and methods of measuring them; principles of academic writing; field methods and methods for conducting a sociolinguistic analysis. The assignments for this course are posted at http://individual.utoronto.ca/ngn/LIN/courses/TBB199/TBB199.14W_syll.htm. One assignment, collecting and describing resources for heritage language speakers and learners, has developed into an important section of the project's website.[9]

Students' responses to the integration of research in their course were positive, as indicated by their enthusiasm for continued involvement with the project after the course and these excerpts from their course blogs:

> I had never thought before that linguists and researchers might be interested in learning more about heritage languages, but I think it is wonderful that they are doing work related to this area. – Lesia

> Because of this course, I began to realize how you can learn so much about your roots just through language and the importance of heritage languages. It is another thing pushing me to improve my Chinese and hopefully begin to learn Vietnamese. – Ashley

> I decided to take this course because I feel that a heritage language is an integral part of a person, and a part that cannot be ignored, and instead should be embraced. Learning more about other's experiences seem to be very interesting, as is sharing my own encounters and perceptions on heritage languages. I believe I will come out of this class every week with many new ideas and information. – Seiwon

---

[8] Van Herk, De Decker & Thorburn (2015) note the financial benefit that many universities have resources available for developing pedagogical tools, particularly to enable inquiry-based learning and independent research by undergraduates.

[9] http://projects.chass.utoronto.ca/ngn/HLVC/2_1_speakers.php

> After yesterday's class, I'm more interested than ever to learn about heritages languages and how it has been for those who have immigrated to Canada many generations ago! –Siquian

> I am happy to know that Russian is one of the languages that we will be studying, and i am honoured to be able to help with the program/research. Through this course, I am looking forward to learning new academic skills, alongside expanding my knowledge about not only my language but other heritage languages in Toronto. – Evgeny

> My analytical skills have continuously gotten better as has my research and observation skills, which was developed through the multiple assignments that we've had throughout the semester – Claudia

Use of an online data server has made it easier to integrate students into the project. We encourage students to use the audio recordings and time-aligned transcriptions for empirical research as part of their studies, and have integrated a consent process where students acknowledge that they understand the ethical requirements for using the data prior to viewing it. Details are available at https://corpora.chass.utoronto.ca/, a site supported by curriculum development grants. Transcripts and recordings are available for use by scholars at other institutions, through a similar, but offline, consent-granting process.

It is immensely rewarding to tap into the abilities and enthusiasm of students who are members of the communities under investigation. The HLVC project has benefitted from hours of volunteer efforts from students.[10] Students are invaluable for recruiting participants, noticing innovations as potential variables for investigation, transcribing, and keeping channels of communication open between communities and researchers. One example of the latter benefit is the interactive speaker map.[11] A team of students compiled voice clips with time-aligned transcriptions and translations, representing the speech of several members of each generation of the HLs in the project. Speech samples are (roughly) geo-located on a map of Toronto, by residence of the speaker and labeled by language, age, sex and generation. This allows exploration of the possibility that varieties develop differently in different neighborhoods, related to settlement patterns of more and less recent immigration.

---

[10] Recognized at http://projects.chass.utoronto.ca/ngn/HLVC/3_2_active_ra.php and http://projects.chass.utoronto.ca/ngn/HLVC/3_3_former_ra.php.

[11] http://projects.chass.utoronto.ca/ngn/HLVC/4_1_map.php

4 Conclusion

A survey of different ways of describing and defining dialects, presented in the first half of the paper, shows the diversity of approaches, but also suggests a continuum along which varieties progress as they diverge from their parent variety. Dialects may be defined by social and/or linguistic attributes. Using the admittedly limited HLVC data available to date, we are not able to show congruence of outcomes from these different approaches to defining new dialects. However, patterns of relationships between the social and linguistic features may be documented, producing descriptions of the grammars of these varieties which may diverge from their parent varieties. Comparisons of the homeland and heritage (putatively "new") dialects can be made when appropriately organized data is available. The second half of the paper reviewed the methods of the HLVC project, suggesting a productive process for making headway on understanding the relationships between linguistic variation and change in order to answer such questions. I thank the organizers of *Methods XV* for giving me a place to integrate these thoughts.

Acknowledgements

References

Auer, Peter & Frans Hinskens. 2004. The role of interpersonal accommodation in a theory of language change. In Peter Auer, Frans Hinskens & Paul Kerswill (eds.), *Dialect change: Convergence and divergence in European languages*, 335–357. Cambridge, New York: Cambridge University Press.

---

[12] Recognized at http://projects.chass.utoronto.ca/ngn/HLVC/3_1_investigators.php

[13] http://projects.chass.utoronto.ca/ngn/HLVC/3_2_active_ra.php

[14] http://projects.chass.utoronto.ca/ngn/HLVC/3_3_former_ra.php

Auer, Peter, Frans Hinskens & Paul Kerswill (eds.). 2004a. *Dialect change: Convergence and divergence in European languages.* Cambridge, New York: Cambridge University Press.

Auer, Peter, Frans Hinskens & Paul Kerswill. 2004b. The study of dialect convergence and divergence: Conceptual and methodological considerations. In Peter Auer, Frans Hinskens & Paul Kerswill (eds.), *Dialect change: Convergence and divergence in European languages*, 1–51. Cambridge, New York: Cambridge University Press.

Bailey, Guy, Tom Wikle, Jan Tillery & Lori Sand. 1991. The apparent time construct. *Language Variation and Change* 3(3). 241–264.

Berruto, Gaetano. 1995. *Fondamenti di sociolinguistica.* Roma / Bari: Laterza.

Bouchard, Chantal. 1998 [2002]. *La langue et le nombril: Histoire d'une obsession québécoise.* 2nd edn. (Nouvelles études québecoises). Montréal: Fides.

Budzhak-Jones, Svitlana. 1994. Variable rule analysis of V/U alternation constraints in Canadian Ukrainian. *Journal of Quantitative Linguistics* 1(3). 202–211.

Cacoullos, Rena & Catherine Travis. 2010. Testing convergence via code-switching: Priming and the structure of variable subject expression. *International Journal of Bilingualism* 15(3). 241–267.

Chambers, J. K. & Peter Trudgill. 1998. *Dialectology.* 2nd edn. Cambridge, New York: Cambridge University Press.

Chumak-Horbatsch, Roma. 1987. Language use in the Ukrainian home: A Toronto sample. *International Journal of the Sociology of Language* 63. 99–118.

Coursen, Charlotte. 1846. *It blows, it snows: A winter's rambles through Canada.* Dublin: Brady.

Cummins, James & Marcel Danesi. 1990. Lifting the multicultural veil. *Our Schools/ Our Selves* Special Issue: Heritage Languages(Heritage Languages). 13.

Danesi, Marcel. 1983. Canadian Italian as a marker of ethnicity. *NEMLA Italian Studies* 7(8). 99–105.

Danesi, Marcel. 1984. Canadian Italian: A case in point of how language adapts to environment. *Polyphony* 7. 110–113.

Foley, W. 2005. Personhood and linguistic identity, purism and variation. In Peter K. Austin (ed.), *Language documentation and description*, vol. 3, 157–180. London: SOAS.

Giovanardi, Claudio, Riccardo Gualdo & Alessandra Coco. 2003. *Inglese - italiano 1 a 1: Tradurre o non tradurre le parole inglesi?* San Cesario di Lecce (Lecce): Manni.

Hudyma, Khrystyna. 2011. Ukrainian language in Canada: From prosperity to extinction? *Working Papers of the Linguistics Circle of the University of Victoria* 21(1). 181–189.

Kang, Yoonjung & Naomi Nagy. 2013. VOT merger in Heritage Korean in Toronto. *Proceedings of the Canadian Linguistics Association* 2012. 1–13.

Kerswill, Paul & Peter Trudgill. 2004. The birth of new dialects. In Peter Auer, Frans Hinskens & Paul Kerswill (eds.), *Dialect change: Convergence and divergence in European languages*, 196–220. Cambridge, UK; New York: Cambridge University Press.

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

Labov, William. 1984. Field methods of the project on linguistic change and variation. In J. Baugh & J. Scherzer (eds.), *Language in use: Readings in sociolinguistics*, 28–53. Englewood Cliffs, NJ: Prentice Hall.

Labov, William. 2001. *Principles of linguistic change Vol. 2: Social factors*. Vol. 2. Oxford [etc.]: Blackwell.

Labov, William. 2007. Transmission and diffusion. *Language* 83(2). 344–387.

Language Factory. 2013. *Canadian dialects of European languages*.

Meyerhoff, Miriam. 2006. Linguistic change, sociohistorical context, and theory-building in variationist linguistics: New-dialect formation in New Zealand. *English Language and Linguistics* 10(1). 173–194.

Nagy, Naomi. 1996. *Language contact and language change in the Faetar speech community*. University of Pennsylvania PhD dissertation.

Nagy, Naomi. 2011. A multilingual corpus to explore geographic variation. *Rassegna Italiana di Linguistica Applicata* 43(1-2). 65–84.

Nagy, Naomi. 2015. A sociolinguistic view of null subjects and VOT in Toronto heritage languages. *Lingua* 164(2). 309–327.

Nagy, Naomi & Michael Iannozzi. 2014. *Older speakers use more null subjects, but is the variable stable? Accounting for contrasting reports of contact effects*. NWAV44 Chicago.

Nagy, Naomi & Alexei Kochetov. 2013. VOT across the generations: A cross-linguistic study of contact-induced change. In Peter Siemund, Ingrid Gogolin, Monika Schulz & Julia Davydova (eds.), *Multilingualism and language contact in urban areas: Acquisition - Development - Teaching - Communication*, 19–38. Amsterdam: John Benjamins.

Nagy, Naomi & Miriam Meyerhoff. 2008. *The love that dare not speak its name– The fascination with monolingual speech communities in sociolinguistics*. Poster. Houston.

Nagy, Naomi & Miriam Meyerhoff. 2015. Extending ELAN into Variationist Sociolinguistics (description and tutorial).

Schneider, Edgar W. 2003. The dynamics of New Englishes: From identity construction to dialect birth. *Language* 79(2). 233–281.

Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the world*. New York: Cambridge Press.

Struk, Danylo. 2000. Between Ukish and oblivion: The Ukrainian language in Canada today. First wave emigrants: The first fifty years of Ukrainian settlement in Australia. In H. Koscharsky (ed.), *Proceedings of the Third Conference of the Ukrainian Studies Association of Australia, in association with the Shevchenko Scientific Society in Australia, Macquarie University, 3–5 July 1998*, 67–74. Huntington, NY: Nova Science Publishers.

Tagliamonte, Sali. 2006. *Analysing sociolinguistic variation*. Cambridge etc.: Cambridge University Press.

Thomason, Sarah Grey. 2001. *Language contact*. Washington D.C.: Georgetown University Press.

Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Trudgill, Peter. 1986. *Dialects in contact*. Oxford: Blackwell.

Trudgill, Peter. 2004. *New-dialect formation. the inevitability of colonial Englishes*. Edinburgh: Edinburgh University Press.

Tse, Holman & Naomi Nagy. 2014. *Exploring automated formant analysis for variationist study of Heritage Cantonese*. CRC Summer Phonetics/Phonology Workshop, Toronto, June 19, 2014.

Van Herk, Gerard, Paul De Decker & Jennifer Thorburn. 2015. *Undergraduate-conducted surveys: Balancing learning and data collection*. Methods in Dialectology XV, Groningen, August 12, 2015.

Walker, James A. 2010. *Variation in linguistic systems*. New York [etc.]: Routledge.

Statistics Canada. 2014. *Mother tongue of person*.

Statistics Canada. 2011a. *Census profile*.

Statistics Canada. 2011b. *2011 Census Highlights: Factsheet 6*.

Wikipedia. 2014. *Canadian Ukranian*.

ISTAT (Istituto Nazionale di Statistica). 2007. *La lingua italiana, i dialetti e le lingue straniere (2006)*.