Linguistic Typology at the Crossroads 5-2 (2025): 161-187 ISSN 2785-0943

Sociolinguistics meets typology: Insight from vernacular speech to account for cross-linguistic patterns

SALI TAGLIAMONTE
UNIVERSITY OF TORONTO

Submitted: 05/06/2024 Revised version: 02/08/2025
Accepted: 07/08/2025 Published: 25/02/2026

Abstract

This study explores how two English grammatical systems with alternative grammatical options expose typological tendencies. A study of negation, *no* vs. negative quantifiers (*nothing*) vs. negative polarity items (*any/anybody*) reveals that the syntactic patterning of English indefinites and negative objects lines up with the hard syntax of the same expressions in closely related languages. A study of variable possession (*the* vs. *my/our*) reveals that while much of the system is stable, but nuances of contrast between UK and Canadian patterns expose differences in conceptions of the personal domain that relate to the cultures in which the varieties are spoken. Taken together these results make several important contributions: 1) the value of natural speech data in providing insights into linguistic and sociolinguistic typology operating in tandem; 2) the importance of statistical tools to uncover relevant patterns and contrasts in variable data. The findings affirm that patterns in single variety, when viewed across communities, dialects and languages reveal that linguistic systems are a continuum of interlocking structural and tractile contrasts in the broader context of human language.

Keywords: sociolinguistics, typology, negative quantifier, negative polarity item, possessive 'the', possessive pronoun.

1. Introduction

In this paper, I explore how vernacular speech data exposes cross-dialectal contrasts, which can, in turn, be used for insights into typology and vice versa. Comparative sociolinguistics (Tagliamonte 2002) and variationist analysis (Tagliamonte 2006;

2025) provide the theoretical approach and methodological toolkit, offering a quantitative, empirical means to understand the systems underlying language variation.

Using two different grammatical systems as case studies, I demonstrate how cross-dialectal regularities expose typological tendencies. A study of negation, (1), reveals that the soft syntax of English (eng) indefinites and negative objects aligns with the hard syntax of the same expressions in Scandinavian languages with object raising (Burnett et al. 2018: 103). (2) a study of possession of prototypical possessums (Gardner & Tagliamonte 2020) reveals a pattern found in other Indo European languages whereby nouns that are inherently possessed ('legs' and 'arms') do not require overt possessive marking (i.e., *my/our*) but can be marked with a definite pronoun, i.e., in English with *the.* Further, the analysis exposes a typological nuance between Canadian and British patterns based on varying cultural patterns of communal possession.[1]

(1)
a. There were ***no*** jobs to be had.
b. The were***n't any*** great places to eat.

(2)
a. 'Cause ***the*** house, I mean ***our*** house, it's not that big.
b. Had a picnic and went out on ***the*** bike.

1.1. The data

The data come from several sources: the Ontario Dialects Project (ODP), a 13-million-word archive of conversations with individuals from 21 communities across the largest province of Canada, Ontario. This is a long-term documentation project that has been underway since the early 2000's (Tagliamonte 2003–2006 et seq.).
Figure 1 shows a map of the communities in Ontario.

[1] All the examples come from the spoken language corpora under investigation. Note the local terminology and other linguistic features of interest in the examples, e.g., subject dislocation in (3).

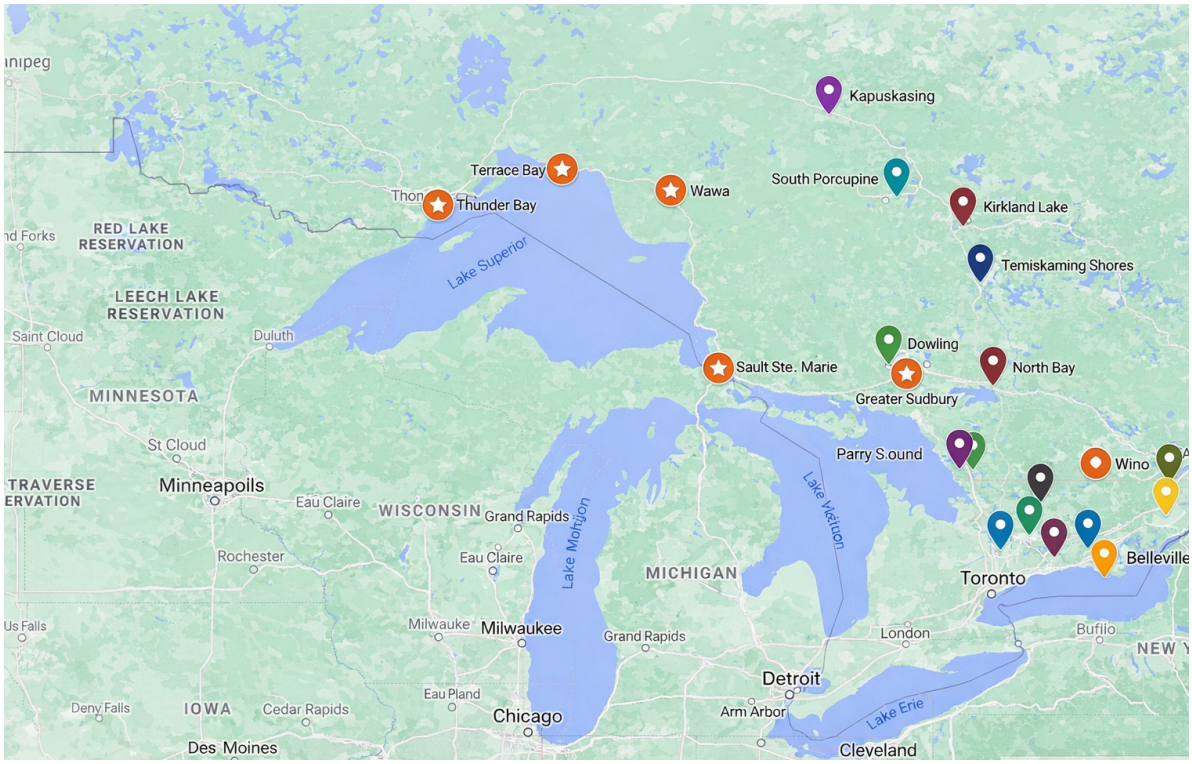

Figure 1: Map of Ontario showing the Ontario Dialects Project communities (Google maps).

In the UK, the data come from several different sources. The first, is the York English Corpus (YEC), a 1.2-million word archive of conversations with individuals born and raised in York collected in 1997 (Tagliamonte 1996–1998). The other corpora include a 3 million word archive of UK dialects comprising elderly people in small communities across the UK (ROOTS), (Tagliamonte 2001–2003; 2013) and a multitude of small studies conducted by students at the University of York from communities across the UK between 1997-2003 (Tagliamonte 2000–2001). In collaborative research, these materials have been supplemented from corpora collected in other universities, in this work the NECTE corpus, a compendium of vernacular speech collected in Newcastle, UK (Allen et al. 2007; Beal et al. 2007). Figure 2 shows a map of the communities in the UK.

All the materials in the synchronic corpora were collected through on the ground fieldwork and lengthy informal recordings of conversations with people born and raised in the communities. Due to legacy recordings in some locales, the time depth of these materials is substantive, with individuals born from late 1800's (from legacy recordings) to individuals born in the early 2001's. Further, because the corpora comprise geographic, social, and linguistic contrasts, the analyst can probe language features in these materials from varying perspectives. In this paper, I focus in on the

value of these materials for typology highlighting findings that have broad relevance; however, the scope of the data has to potential to explore many types of contrast.

Figure 2: Map of the United Kingdom showing locations of fieldwork (Google maps).

1.1.1. Can sociolinguistic data be useful for the study of typology?

Sociolinguistic data is spoken interaction in informal settings. It is often not ideally constituted, regular across places or data sets, or equally 'good' on many dimensions. When there is more than one person in the interview, they often talk over each other and the sound quality of a recording can be hampered by unexpected local conditions, e.g., grandfather clocks, aquariums, dogs, etc. The data are not balanced, as in corpus linguistics where there is a data extraction plan and cut-off point so that the number of words by category (e.g., register) is equal. This type of methodology is not the case in sociolinguistic data. The fieldworker aims for approximately an hour of conversation with each community member; however, it may be less or more. Another attribute of sociolinguistic data is that it is unpredictable. The fieldworker cannot anticipate what someone will talk about or for how long or what extraordinary words

and expression may come out of their mouth, or not. Finally, sociolinguistic data is crosscut by innumerable regional, social, and cultural factors.

Nevertheless, the Canadian and UK repositories are among the largest corpora of vernacular speech in the world and cover at least 120 years in birth dates of the individuals from the late 1800's up to the early 2000's. The data are sufficient for finding and studying linguistic variables that are notoriously rare in spoken corpora and may never appear in written sources. Comparative studies offer the ultimate check on interpretation and for discovering nuances both linguistically but also historically and culturally. Moreover, these materials offer a comparative perspective that is a must when considering issues of typology.

1.2. The value of social information and linguistic theory

I will argue that sociolinguistic metadata such as date of birth, gender, education, and job type have the potential to offer important insights into typology. Further, the different types of communities across Ontario and the UK afford the ideal opportunity to contrast varieties and communities where local features may expose unique forms and patterns. While a sociolinguistic approach to typology is not new, its application has mostly been applied to historical texts which have a strong bias towards the written language, formal settings and literate individuals (Walkden et al. 2023: 550). Most corpus-based studies come from sources other than the spoken language, including dialect surveys, atlases etc. Even when spoken language is investigated the corpora under investigation often do not provide representative coverage of community level social contrasts nor metadata about the characteristics of the individuals (e.g., Szmrecsanyi & Grafmiller 2023).

Building on the premise of the journal "Linguistic Typology at the Crossroads" whereby cross-linguistic tendencies are typically found in discourse use as pathways of language change (Keller 1994; Hawkins 2004; 2014; Haspelmath 2021), I suggest that social information, e.g. gender, occupation, education alongside geographic and regional differences, e.g. population size, distance from urban centers, can add depth and insights to these trajectories.

I also build on theoretical linguistic research that has demonstrated grammatical contrasts that are portrayed as 'hard' vs. 'soft'. Hard contrasts are the structural properties that create grammaticality differences between languages, the mainstay of typology, whereas soft contrasts are preferential tendencies between languages (e.g.

Bresnan et al. 2001). For example, in syntax languages with more flexible word order may allow for different types of alternations compared to languages with rigid word order. Similarly, in semantics, the meanings of individual words and how they interact within the meaning system of a language in conjunction with the cultural ecology in which they are used can also influence the types of alternations that occur. The same is true of phonological, morphological, and discourse-pragmatic patterns. I will argue that nuances in the constraints that operate on variable systems provide an ideal mirror of these soft contrasts and offer evidence for typological patterns. Further, trajectories of change can be mapped to geography (e.g. Poplack & Tagliamonte 1999) and sociolinguistic typology (Trudgill 2011; Szmrecsanyi & Wälchli 2014; Tagliamonte & Brook 2016; Torres Cacoullos & Travis 2019), although depending on the variable, these may or may not apply.

1.3. The value of the spoken vernacular

The spoken vernacular, what has been referred to as "naturally occurring speech" or "what language users do", is argued to be the "gold standard" for typology (Du Bois & Troiani 2023). Moreover, Miestamo et al. (2023) suggest that by "tapping into the rich vein of functional diversity found in naturally occurring conversation, typology will be better positioned to discover new forms of structural diversity". In the same way, dialect data are ideally positioned to add to the knowledge base of typology. Because of their low status dialects are often regarded as degenerate versions of the prevailing linguistic norms, therefore, they tend to preserve linguistic phenomena that are not recorded in descriptive accounts and so many of their features remain unattested in the published literature. The linguistic phenomena that endure in dialects offer insight into earlier typological structures in the same language as well as their history and culture that led to their evolution (Du Bois & Troiani 2023).

The deep knowledge that lives within community members as they use their vernaculars emanates the "imprint of ages emblazoned in vocabulary and expressions" (Tagliamonte 2016). Indeed, the words and expressions people use provide information far greater than their conscious knowledge of their language. From the broader contemporary context (currently in the early decades of the 21st century), it is also important to emphasize that dialects in rural locales all over the world are dying rapidly. The type of sociolinguistic fieldwork and documentation of research programs focusing on regional dialects aims to bring the linguistic treasures

held within dialect data into the open and into academic study as an enduring legacy for future generations. For the purposes of typology, this information has the potential to offer linguistic details of form and structure that are not represented in more formally gathered language materials.

1.3.1. A gold mine of linguistic data

Spoken vernacular dialect data collected in community context (i.e., representing age, gender and other socially relevant information) provides a cornucopia of sociolinguistic contrasts. Moreover, due to the wide range of birthdates of the individuals who live in geographically diverse locations, the Canadian and UK archives are replete with remnants of receding and emerging features. Conservative features from earlier times include zero articles, (3a), double demonstratives, (3b), old preterit forms, (3c), etc. and discourse-pragmatic markers, (4). The most infamous discourse-pragmatic marker in contemporary English is *like* (4a) (e.g., D'Arcy 2017), but other markers are developing (4b) and there are many old ones, (4c-d).

(3)

a. My mother's father was ***Ø relatively prosperous*** *farmer*. (mschiff, woman, 78, Parry Sound)
b. One day, we had ***this here snowstorm*** down here. (dhinds, man, 77, Christie)
c. I ***come*** home to Tatlock and I stayed there and ***run*** the mail. Looked after the post office. (mcarson, woman, 81, Ottawa Valley)

(4)

a. I think ***like*** I don't know ***like*** I guess.
b. No ***wait***, we did. ***Wait***, did we? .
c. ***See***, it would start about eight o-clock in the evening and go 'til about four o-clock in the morning, ***you see.***
d. If you did get onto the roads, ***why***, you never knew when you were going to get home ***or anything like that.*** (gbillings, man, 78, Ottawa Valley, 0101199_5)

Global mobility has led to a myriad of new language contact situations that — at the turn of the 21st century —can transcend geography offering new uncharted territory

for investigation, e.g. booming mega trends (e.g., Tagliamonte & Smith 2021) and so called "Black Swans" (Taleb 2007; Tagliamonte et al. 2016), that defy historical predictions about change.

In summary, spoken language data, messy and unpredictable as it is, offer the potential for deep insights into typology by seeking out rare phenomena in the dialects of closely related languages and deviations within types across varieties and dialects.

2. Crosslinguistic variation

As introduced earlier, recent research in linguistic theory has discovered that grammatical contrasts that distinguish languages can present as preferential tendencies *within* languages, suggesting that there is a critical connection between the syntactic structure and semantic organization of complex expressions and how they are used by speakers. A phenomenon that is categorical in one language, may be optional or variable in another.

Setting the scene for this kind of an approach, Bresnan et al. (2001) report on a comparison of person hierarchy effects and grammatical voice in Lummi, a Salish language compared to English. In Lummi, transitive predicates with third-person actors and first-person patients must be in passive voice; however, first or second person actors and third person patients must be in the active (Jelinek & Demers 1983). I illustrate with English equivalents in (5).

(5)

a. I am known by the man
   1st person 3rd person

b. The man knows me
   3rd person 1st person

In English, a third person actor with a first person patient is grammatical with either active or passive voice; however, in a study of these alternates in spoken language in the Switchboard corpus (Godfrey et al. 1992), notable contrasts were discovered. First and second person actors with a third person patient occur only in the passive

(n = 0/6246 cases) and third person actors with first or second person patients occur overwhelmingly (97%) in the active (n = 472/486). Although both options are grammatical, an active and passive contrast is operational in usage patterns. The critical facet of this result is that it demonstrates that hard contrasts in some languages may manifest as soft contrasts in others, a finding that has been repeated for other features of English, such as the dative alternation (e.g. Bresnan et al. 2007; Bresnan & Ford 2010) and the animacy effects in the genitive alternation (e.g. Zaenen et al. 2004; Wolk, et al. 2013; Szmrecsanyi et al. 2017; Szmrecsanyi & Grafmiller 2023). If we turn these observations into a workable hypothesis then an important question to ask is: "to what extent do the apparently variable English patterns have correspondences with invariant syntactic patterns in other languages?" (Burnett et al. 2018: 91).

3. Two case studies of variation

In the next section, I highlight the results of two case studies with the goal of offering insights from large archives of sociolinguistic data to the study of typology. The two case studies are excerpted from research from the corpora described in section 1.1: 1) grammatical choices in negative patterns with negative quantifiers and polarity indefinites (Childs et al. 2015; Burnett et al. 2018) and 2) the alternation between definite article and possessive pronouns (Gardner & Tagliamonte 2020). In both case studies, there are two standard options that mean more or less the same thing, i.e. ***I don't know anything/I know nothing*** and ***my bike/the bike,*** thus constituting a 'linguistic variable' (Labov 1972 et seq.). Further details on situating these linguistic phenomena in time and space, details of methodology and results can be found in the research papers.

3.1. Negation patterns

In English negative constructions, negative quantifiers (e.g., *no, none, nothing, nobody,* etc.) and negative polarity indefinites (NPI's) (e.g., *any*, *ever, at all*) alternate in constructions where either one is grammatical, (6a-b).

(6)

a. *There were* ***no*** jobs to be had. (There *were****n't/not any*** jobs to be had)

b. *The were****n't any*** great places to eat. (There were ***no*** great places to eat)

These alternates occur with near equal frequency in naturally occurring speech. The question is what determines whether an individual will use one alternate vs. the other?

Many studies — most based on historical written materials — report a consistent pattern. A hierarchy of verb constructions conditioning the options: existential *be* occurs most often with negative quantifiers, followed by stative *have*, then copula *be* and finally lexical verbs, which tend to prefer negative polarity indefinites, also referred to as NPI's. Notice however, that both options are grammatical in each category, as in (7) with NPI's and in (8) with *no*.

(7) Existential *be* → stative *have* → Copula *be* → Lexical verbs

a. There *wasn't* ***anything*** else to do. (prowlett, man, b. 1943, Toronto, Ontario)[2]
b. They *haven't* acquired ***anything*** yet. (pronenen, woman, b. 1955, Toronto, Ontario)
c. She *was no* good for ***anybody***, that woman. (D/131, man, b. 1931, Newcastle, UK)
d. They *don't look* ***any*** different. (echapman, woman, b. 1942, York, UK)

(8) **Existential *be* → stative *have* → Copula *be* → Lexical verbs**

a. There *was* ***no*** comaradship. (pmarshall, man, b. 1949, Wheatley Hill, UK)
b. My tractor *had* ***no*** brakes, ***no*** nowt. (crabeale, man, b. 1967, York, UK)
c. I *'m* ***no*** different now. (rburkett, female, b. 1968, Belleville, Ontario).
d. My grandmother *spoke* ***no*** English. (oholtby, b. 1951, Toronto, Ontario)

In Childs et al. (2015), we examined the patterns of use of these alternates in individuals born and raised in two Canadian communities, a large city (Toronto) and a small town (Belleville) along with and several northeastern communities in the UK, York and Newcastle (two cities), Durham (a small city) and Wheatley Hill (a small town).

Figure 3 provides a more nuanced perspective, distinguishing Canada vs. the UK and showing the overall rates of negation in each community.

[2] Each example indicates the individual's pseudonym, their perceived gender, date of the birth and place of residence.

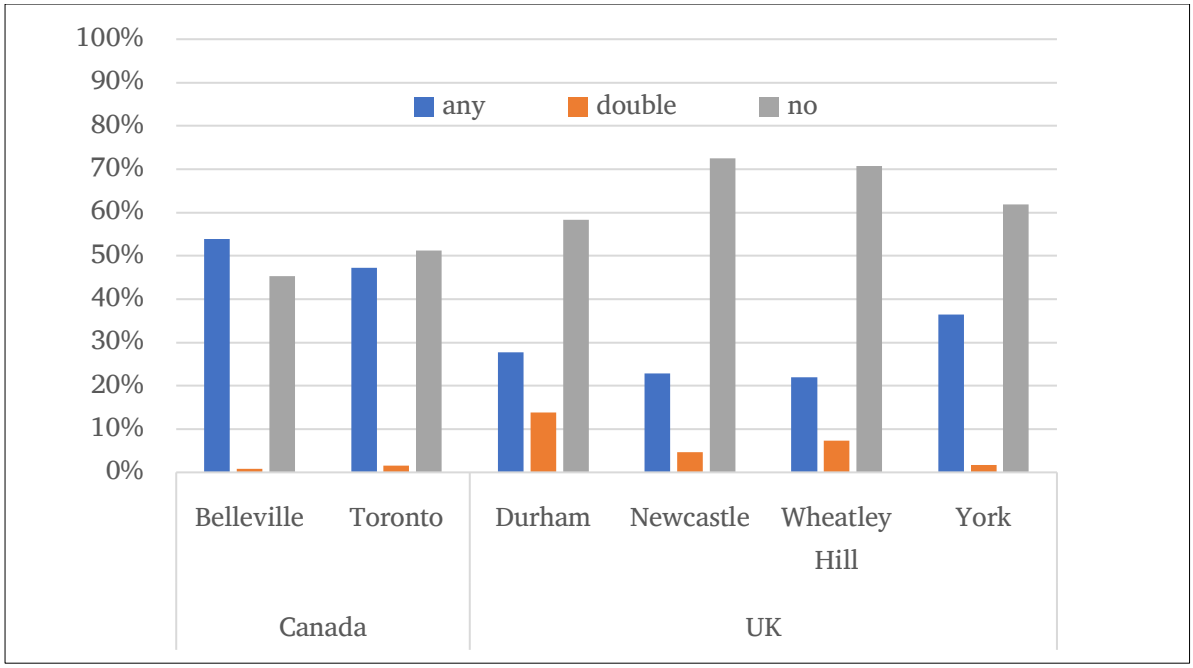

Figure 3: Rate of negative quantifiers across communities in Canada and the UK.

Double negation (i.e., negative concord) is virtually absent in Toronto and Belleville as well as in York. It also occurs rarely (6.6% of the time) in Newcastle, Wheatley Hill, and Durham. In contrast, variation between *no*- and *any*-negation is present in all varieties, but the distribution is markedly different by variety. In Canada, the two constructions have near-equal frequency, with a slight preference for *no*-negation in Toronto. In England, *no*-negation dominates at 63% in York and ranges from a high of 73% in Newcastle to 58% in Durham. Given that *no*-negation is the older form, it appears that UK dialects, at least in the northeast are more conservative than Canadian dialects in Ontario. However, there is notable inter-variety coherence in overall rates of variants by community regardless of population size or location.

In the next step, we replicated the coding schema of earlier studies and tested for the effect of the hierarchy of verb/contruction types that had been reported in studies of historical written English. Figure 4 shows the rate of *no*-negation according to these construction types, amalgamating the data in the UK as "Modern Spoken" and in Canada as 'Toronto, Canada' and comparing it with the rates reported in earlier studies of written data from the UK (Early Modern and Modern English).

The construction hierarchy in the UK and CDA data mirrors, in large part, the one found in Early Modern and Modern Written English. Each data set exhibits the same overall hierarchy: higher rates of negative quantifiers, i.e. *I know nothing*, with existential *be*, then copula *be*, *have* and lexical verbs the least, i.e., lexical verbs tend

to occur with *any* negation, i.e. *I don't **like** anything*. However, notice that rates of negative quantifiers are lower overall in spoken data.

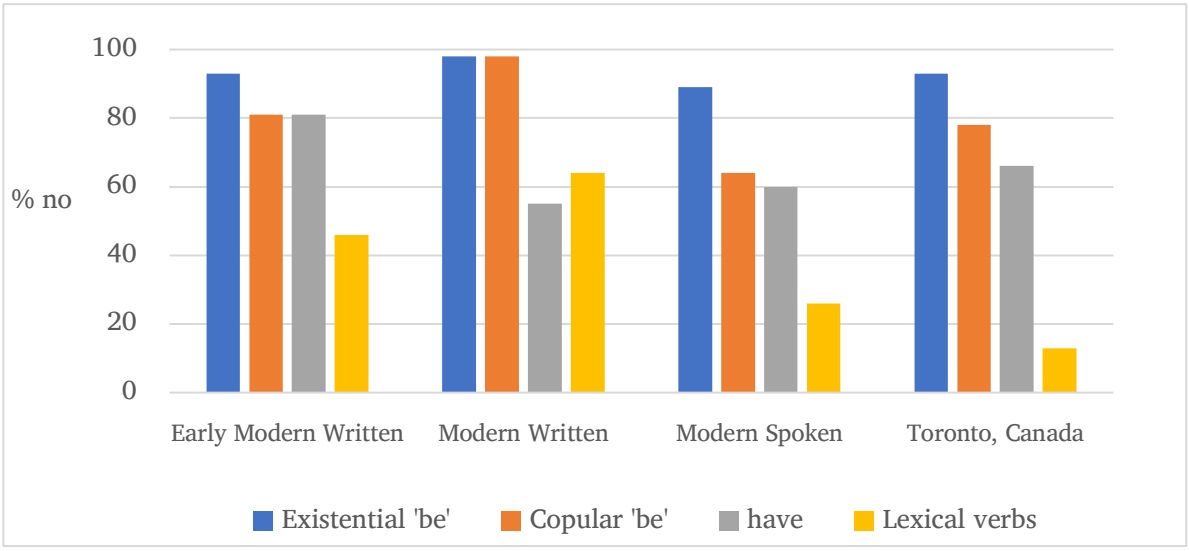

Figure 4: Rate of negative quantifiers in written compared to modern spoken English.

Moreover, Early Modern English differs from Modern English in having greater use of negative quantifiers with *have* and 2) the modern spoken data from Canada has far greater use of negative polarity items with lexical verbs. Note too that the spoken British and Canadian English data are much the same, except for the nuance that the UK data has higher use of negative quantifiers with lexical verbs, (9) than the Canadian data where it is rare.

(9) I *see **no*** point in it. (JS/169, man, b. 1982, Newcastle)

These differences mean that a frequency-based explanation does not entirely explain the variation and that deeper consideration of these differences was warranted.

In a later study focusing on the same negative options, we augmented the analysis by collaborating across disciplines, adding a socio-semanticist with a specialization in mathematical tools (Burnett) and a formal syntactician (Koopman) to the research team. Then we went back into the data to conduct a more in-depth analysis. The question we aimed to address was: *what other linguistic patterns may underlie the linear order of the constraint hierarchy?*

3.1.1. Expanding the perspective to typology

To re-focus attention on underlying structural patterns, we scrutinized languages related to English for insight, for example, Scandinavian languages. Kayne (1998; 2000) demonstrated the key syntactic contrast of object raising that distinguishes negative quantifiers and negative polarity items based on the nature of the verb. In English the syntactic position occupied by object negative quantifiers differs according to the syntactic properties of the other morphosyntactic material that the indefinite combines with. For example, in some verbal constructions, such as an existential (10a), the direct object *nothing* has undergone a negative quantifier shift to a higher syntactic position than it would occupy if it had occurred in a structure with a lexical verb (10b) or a participle (10c).

(10)

a. There'*s* *nothing* to do here. (katrina, woman, b. 1967, York, UK)
b. He eats *nothing*. (AA/613, man, b. 1985, Newcastle, UK)
c. She got her degree from Ryerson. She*'s* *doing* *nothing* with it. (akaran, woman, b. 1979, Toronto, Cda)

In other words, only the verb *be* and the object can raise out of a verb phrase. In contrast, other verbs cannot raise, so the object does not raise either, (11).

(11)

a. There are *no* → Neg-Q
   There are *no* absolute rules now. (ekempt, woman, b. 1928, Toronto, Ontario)

b. There *aren't* ***any*** … → NPI
   There *are****n't*** a lot of mosquitos. (asinkic, woman, b. 1962, Toronto, Ontario)

Based on Kayne's analysis, we predicted that when it is possible for the object to move upwards in the syntactic structure, negation is more likely to be realized with negative quantifiers.

Figure 5 shows a notable contrast: where the negative is higher than VP, a negative quantifier is used; where it is lower than VP it appears as negative polarity item. There is only a small amount of variation.

3.1.2. Results of typology on variation — negation

We coded the Ontario and UK data (see section 1.1) according to whether there was object raising above the verb phrase or not (see Burnett et al. 2018 for coding and other information). The results are illustrated in Figure 5.

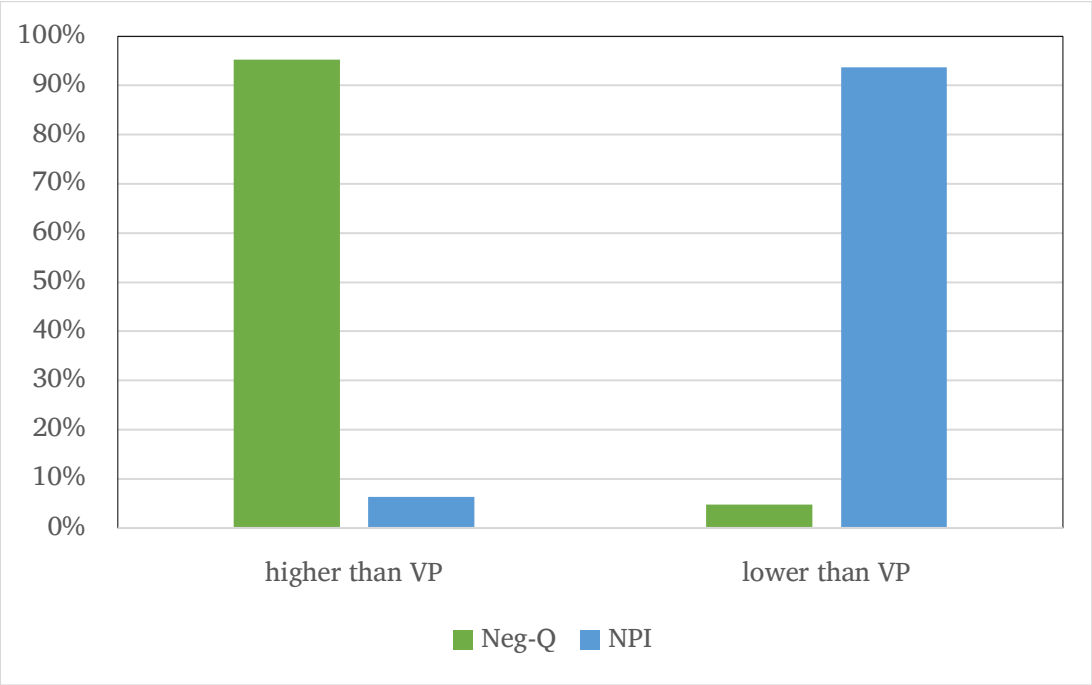

Figure 5: Rate of negation constructions by syntactic domain.

3.1.3. Summary

The key finding is that when the negation constructions are coded according to syntactic domain rather than construction type, the variants are nearly completely distributed according to syntactic position: *no/not* appears in the higher syntactic domain, 95% and negative quantifiers (e.g., anything) appear in the lower syntactic domain, 95%. The variation is highly circumscribed even though both options are grammatical in the language.

The cross-linguistic regularities from the comparative syntactic study show that the structural position that the negative quantifier or negative polarity it occupies plays and important role in negative patterns in English. A hard contrast found in some Scandinavian languages is manifested as a soft contrast in English.

In section 3.2. I turn to the second case study, personal domain possessums.

3.2. Personal domain possessums

English offers two possibilities for marking possession in the personal domain, a definite article, (12a-b) and a possessive determiner, (13a-b), and both options are grammatical.

(12)

a. ***The*** *house*, I mean ***our*** *house*; it's got- it's not that big. (mlondry, woman, b. 1959,York, UK)

b. I was at home and ***my kids*** couldn't eat the ice-dream fast enough out of ***my freezer***. … … ***Our*** cottage is like a cottage from years ago. … I mean we light fire and ***the kids*** still do their hot-dogs and stuff … (rsinardo, woman, b. 1964, Toronto, Cda)

Where does an English speaker use *the* and when *my/our?* Upon reflection, even native speakers, wonder where one or the other form should be used; yet in unreflectingly vernacular conversation both options occur regularly. Grannis (1972) lamented that "the major modern theoretical approaches to language description [have] not yet developed [the] means of adequately accounting for the referential complexities involved in the use of the definite article." He called this the "Definite Article Conspiracy". To illustrate, Grannis (1972) elaborated that a man with a missing cat might say to his wife, "have you seen *the* cat?" If she hasn't, he might then ask his neighbour, "Have you seen *our* cat?" To a complete stranger, he might then say, "Have you see *a* cat?" But to any one of these people the same man could say, "I have to feed *the* cat." Likewise, to your wife, you could say, instead of "*the* cat", "where's *our* cat?" Furthermore, you could also call your wife or husband, *"the* wife/*the* husband*"*. This description underscores the robust variability of forms.

In Gardner and Tagliamonte (2020), we studied the patterning of these options in a comparative study of spoken vernacular discourse in the UK and Canada using the York English Corpus and the Toronto English Archive (section 1.1). We based our data extraction using the concept of personal domain (Bally 1926) in which membership in the personal sphere of the speaker is the relevant dimension of contrast (McWhorter 2002: 224). We focussed on nouns "conventionally understood to be within the personal domain" (Bally 1926: 233). Common or prototypical words from several personal domain-type semantic categories (e.g. body parts, intimates,

conveyances, etc.) were used as search keywords, including *house, back garden/yard, car, boyfriend/girlfriend,* etc.), (13a-b).

(13)

a. The only thing I've ever gotten stolen was – when I– the day I got married, someone stole a case of beer from ***our*** *backyard*, someone stole gas out of ***the*** *car*. Siphoned it out of ***the*** *car*. (njalna, woman, b. 1964, Toronto, Cda)
b. We had to take ***the*** car for some work doing and we use a chap in Elvington and … so I drove ***our*** car and Wayne went on his motorbike. (cbiggs, woman, b. 1964, York, UK)

Dictionaries and usage guides consistently report that when possessive *the* is used the context is informal, (14a-c).

(14)

a. (*informal*) used instead of a possessive pronoun to refer to someone with whom the speaker or person addressed is associated: *I'm meeting the boss | How's the family?* (Oxford Dictionary of English, 2nd edition revised)
b. British informal *my, our* (Oxford Dictionary of Canadian English)
c. Colloquial: *my:our, the dog, the car* (Oxford Dictionary of Current English)

Possessive *the* is also said to be used only to refer to someone who the speaker or interlocutor is associated with such as "*the* boss" or "*the* family". It may also used for body parts, (15a-b). Notice here that the possessum — the thing possessed — is restricted to possessions of humans.

(15)

a. *how's the arm today?* used as a function word before names of some parts of the body or of the clothing as an equivalent of a possessive adjective (Merriam-Webster)
b. used to refer to a part of the body, e.g. *Lieutenant Taylor was wounded in* ***the*** *knee; How's* ***the*** *ankle? Is it still hurting?* (Longman Dictionary of Contemporary English, Advanced Learner's Dictionary)

Possessive *the* is also said to be related to the habit of men referring to their wives, or

children with *the*, as in "How's *the* wife" or "How are *the* kids?" It is also a familiar way of referring to a husband or father, as in "*the* old man". However, such use is restricted to "some men", perhaps suggesting social stratification as well as stylistic correlates of "informality" and "familiarity". It may also be a regional dialect difference.

In summary, both linguistic and social factors are reported to be involved in the usage patterns of possessive *the* vs. posessive pronouns *my/our* in English with personal domain possessums.

3.2.1. Expanding the perspective to typology

To focus on underlying structural patterns, we scrutinized languages related to English for insight. In this case the relevant comparison is the Indo-European language family where a special relationship between definiteness and possession is documented. First, possessed nominals that are marked with a possessive pronoun do not require overt definite marking because they are inherently definite. Second, objects that are considered "inherently possessed" — like body parts — do not need overt possessive marking and usually only appear with definite marking. This contrast is a hard contrast in several languages like French (fra) and Italian (ita), as in (16a-b). However, across languages the patterns differ slightly and in languages such as Spanish and English, there is alternation.

(16)

a. French
Je me lave ***les*** *mains*
I to.myself wash ***the*** *hands*
'I wash my hands'

b. Italian
mi lavo ***le*** *mani*
to.myself wash ***the*** hands
'I wash my hands'

Our goal, therefore, was to uncover the patterns underlying the choice between *the* vs. *my/our* in Canada and the UK as represented by Toronto and York in a set of objects within the personal domain across semantic categories.

Because alienability is basically the distinction between body parts and everything else, we adopted Bally's concept of "personal domain" in order to extend the coverage to a more fine-grained categorization schema.

Bally considered personal domain to be entirely subjective:

> nothing prevents the collective imagination from attributing to the self objects that normally have their own independent existence, or, conversely, of detaching those things which in reality cannot be. The extent of the domain is determined by the cultural outlook of each linguistic group. Its limits may vary from language to language and vary within the same language during the course of its evolution.

Therefore, we left open the possibility that the choice may not only be linguistic but also determined by the varietal differences and cultural context in time and space.

3.2.2. Results of typology on variation — possession

First, we discovered that the rates of possessive *the* and possessive pronouns *my/our* were remarkably similar in Toronto and York. Overall possessive *the* was used in about 30% of the possessive personal domain contexts we included in the study. Then we performed a mixed-effects logistic regression model in R (R 2022) to test the statistical validity of the many potential social and linguistic patterns that had been mentioned in the literature, while treating individual as random.

We discovered that most of the predictors that had been attested in the literature were not significant. Grammatical person was not significant, nor was syntactic position of the possessum, its grammatical number, whether it was human or not, animate or not or alienable or not. Notably, alienable possessums, e.g. *dogs, kayaks,* were more likely to be used with possessive *the* than inalienable possessums, e.g. *arms, bellies,* despite the claim in Quirk et al. (1985) that possessive *the* is idiomatically preferable to a possessive pronoun in inalienable contexts.

Instead, alienable possessumes had a higher rate of possessive *the* than inalienable possessums, the reverse of what was predicted. With respect to the social factors, men used slightly more possessive *the* in Toronto and women use slightly more possessive *the* in York, but these differences were not significant. We found no trend in apparent time: younger speakers did not use less possessive *the*. Comparing speaker education and occupation did not reveal any significant patterning either.

The best fit of the model to our data categorized the possessums into semantically related objects, rather than an animacy contrast between human vs non-humans, or animates vs non-animates. Table 2 shows the rate of possessive *the* in Toronto and York across possessum categories in the study.

	Bikes	**Body parts**	**Family**	**Homes**	**Pets**	**Romantic partners**	**Vehicles**	**Yards**
				Toronto				
the	12%	9.9%	26.1%	51.1%	75%	4.5%	57.2%	49.7%
my/our	88%	90.1%	73.9%	48.9%	25%	95.5%	42.8%	50.3%
				York				
the	28.8%	5.4%	40.5%	47.3%	85.5%	3.92%	63.8%	76.58
my/our	74.4%	71.2%	94.6%	59.5%	52.7%	14.5%	96.1%	23.4%

Table 2: Rate of possessive *the* and possessive pronouns *my/our* by semantic category.

In both places, pets, and vehicles have the highest rates of possessive *the* and bikes, body parts. Homes, (e.g. cottages, cabins, and chalets) are robustly variable with both forms in both. Non-romantic family members (e.g. children, kids, and in-laws) are positioned in the middle between these groups, but note the striking difference in use of possessive *the* — Toronto (26.1%) and York (40.5%).

Another difference between York and Toronto is with 'yards': this category has one of the highest rates of possessive *the* in York (76.6%), but in Toronto yards have 49.7% possessive *the*, patterning with vehicles and homes.

3.2.3. Summary

The key finding is that there is remarkable similarity between Toronto and York in the rate and conditioning of marking possession with *the* vs. *my/our*. This parallelism extends to the lack of significant linguistic and social conditoning on the variation (see Gardner & Tagliamonte 2020: 246, Table 6).

Taken together these results suggests that the choice of possessive *the* and possessive *pronoun my/our* is a stable, soft contrast in English. However, rather than marking a hard contrast based in inalienability as in some Indo-European languages, in English it encodes communal possession. For example, pets, which occur with possessive *the* 75% and 85% of the time in this cross-variety comparison, are generally co-owned by

a group of people, such as a family in both cultures.

A similar personal domain interpretation can be expected for the family car or the family home. Body parts on the other hand only belong to one person and these are overwhelmingly encoded with *my/our*. Likewise, romantic partners like husbands and girlfriends are the least communally owned, and the ownership is more reciprocal, at least in contemporary western culture.

However, the interesting nuance to this linguistic system is that in two areas of personal sphere, the UK and Canada diverge considerably. In York 'yard' and other green spaces are very often referred to with possessive *the*. In Toronto, non-romantic family members (e.g. 'children', 'sister') are referred to overwhelmingly with possessive *my/our*.

4. Speculating on the impact of culture

The difference between 'cars' and 'bikes' is probably the clearest illustration of the striking regularity across dialects. Figure 6 shows the rate of possessive *the* in 'cars' vs. 'bikes' in Toronto (Canada) and York (United Kingdom).

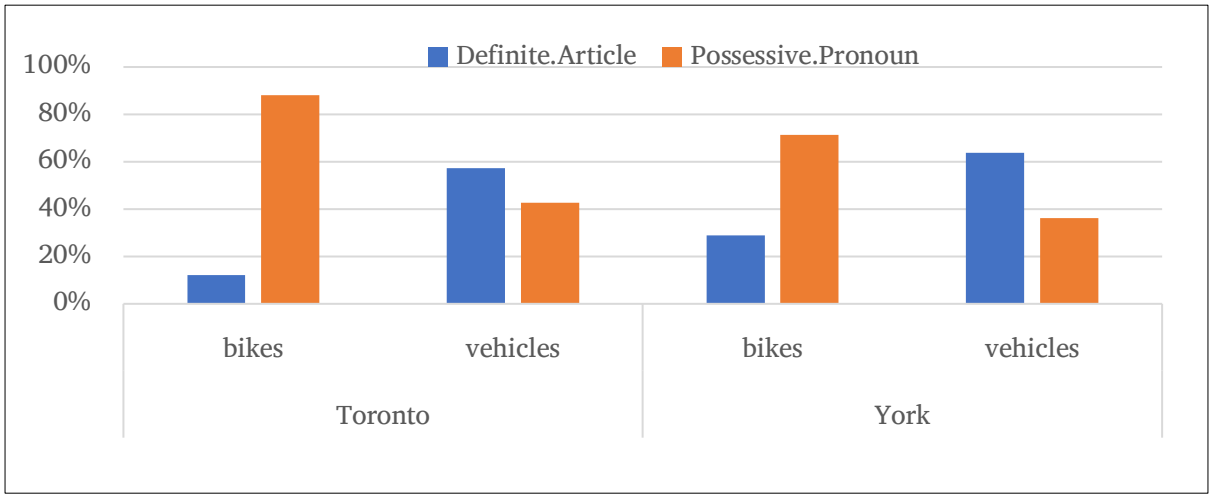

Figure 6: Rate of possessive *the* vs. possessive pronouns *my/our* with bikes vs. vehicles.

Cars (vehicles) which are generally thought of as belonging to a family occur with possessive *the* about 50% of the time, while bikes, which generally belong to only one person, have more *my* and the contrasts across varieties is parallel. In contrast, Figure 7 also highlights a striking deviation between the two dialects, namely the rate of possessive *the* with yards and children in Toronto compared to York.

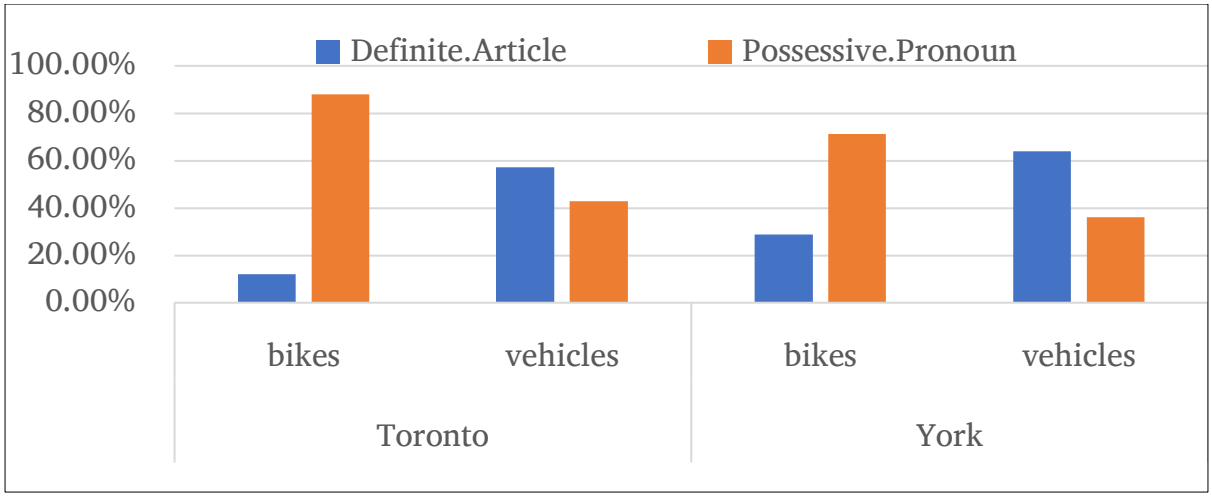

Figure 7: Rate of definite article *the* vs. possessive pronouns *my/our* with bikes vs. vehicles.

The patterns here are antithetic. In Gardner & Tagliamonte (2020:249), we speculated that the higher rate of possessive *the* for yards (76% compared to 54%) among York speakers is likely due to the fact that the green spaces in York are spaces like the garden and the family allotment are spaces that are conceived as being more communal than the often expansive front and back yards of Torontonians. It is not clear however, why children, i.e. *the children* not *my children,* would by the use of *the* be construed as conceptually more shared in York than in Toronto is beyond the scope of the study, but is also suggestive of cultural differences. We can conclude that where cultures treat items in the same way, we can expect similar even universal patterns in the constraints on alternating options (i.e. *the* vs. *my/the*), but when cultures differ in their conception of items in terms of the personal sphere, we can expect differences.

This discovery aligns with earlier research in demonstrating that culture is a key attribute of typological variance (e.g., Evans 2010; Song 2010) and highlights a novel direction for study. Possession is subjective, varying across cultures and time (Bally 1926). These results suggest this malleability is susceptible to the ways space is conceived as well as in how human relationships are construed. If so, then there is also the possibility that the conception of personal domain is reflected in linguistic variation as one semantic category or another can shift over time and that these can be culturally constructed and locally distinctive. I suggest therefore that targeting linguistics variation in systems such as the marking of possession that are implicated in changing cultural norms and practises are ripe for diachronic study and will provide interesting fodder for future research.

The key finding is that when possessive constructions are coded according to personal domain, there was no correlation between the use of possessive *the* and the traditional users of non-standard speech, e.g., men or working-class speakers, as predicted in much of the related literature on this topic. Instead, we discovered that possessive *the* is typically used instead of a possessive pronoun (i.e., *my/our)* with possessums that are possessed communally (i.e., *the dog, the car*) rather than individually (*my bike, my boyfriend, my back*). In this case study, cross-linguistic comparison exposes a typological contrast in the personal domain, with nuances that reflect differences between two regional varieties. A striking finding is that there seems to be a critical link with culture in the variable systems of certain categories that warrant further investigation.

5. Discussion

Examination of two variable linguistic systems in spoken English language data have revealed insight into typology, in particular the relevance of hard and soft contrasts. In both linguistic systems, cross-linguistic regularities common in other languages were profitably used to understand why there is a choice, in one case between negative quantifiers and negative polarity items and in the other between possessive *the* and possessive pronouns (*my/our)*. In the former case, we identified a structural contrast that underlies an often-reported construction hierarchy in related languages. In the latter case, the marking of possession found a largely stable system. However, we also uncovered a dialect difference between the UK and Canada that exposed an area of 'softness' in conceptions of personal domain that are different. Importantly, what we know about variable systems is enhanced by expanding the knowledge base to include a typological perspective thereby increasing understanding and explanation more generally.

The two case studies also highlight the value of natural spoken language interactions collected using sociolinguistic methods to provide insights into typology. They also show that the tools of quantitative analysis and statistics applied to corpus data keyed to time and place, culture and context offer an important means to detect the details of typological patterns and how these play out in dialectal and cross-linguistic variation, all leading to more integrated explanations. Patterns in single variety, when viewed across communities and across dialects and languages confirm

that linguistic systems are a continuum of interlocking patterns in the broader context of human language.

Acknowledgements

I gratefully acknowledge the support of the funding agencies that enabled the research to be conducted, in Canada spanning 2001-present from the Social Science Research Council (SSHRC), in the UK spanning 1995-2003 from the Economic and Social Science Research Council (ESRC). I express my sincere gratitude to the individuals in the communities in the UK and Canada who shared their stories and reminiscences with me and my research team. The corpora were constructed by a skilled team of research assistants in the sociolinguistic laboratories in York and Toronto. Data extraction, coding and analysis for the two studies was made possible by collaboration with *Heather Burnett, Karen Corrigan, Claire Childs, Christopher Harvey, Hilda Koopman,* and *Matt Hunt Gardner*.

Abbreviations

CDA = Canada
NECTE = Newcastle Corpus of Tyneside English
NPI = Negative Polarity Item
YEC = York English Corpus
ODP = Ontario Dialects Project
UK = United Kingdom

References:

Allen, William & Joan Beal & Karen Corrigan & Warren Maguire & Hermann Moisl. 2007. A linguistic 'time-capsule': The Newcastle Electronic Corpus of Tyneside English. *Creating and Digitizing Language Corpora: Diachronic Databases*, 16–48. Basingstoke: Palgrave Macmillan.

Bally, Charles. 1926. L'expression des idées de sphère personnelle et de solidarité dans les langues indoeuropéenes. In Franz Frankhauser & Jakob Jud (eds.), *Festschrift Louis Gauchat*, 68–78. Aarau: Sauerländer.

Beal, Joan & Karen Corrigan & Hermann Moisl (eds.) 2007. *Using Unconventional Digital Language Corpora: Volume 1: Synchronic Corpora*. Basingstoke, Hampshire: Palgrave Macmillan Limited.

Bresnan, Joan & Shipra Dingare & Chris D. Manning 2001. Soft constraints mirror hard constraints: Voice and person in English and Lummi. In M. Butt & T. Holloway King (eds.), *Proceedings of the LFG01 Conference*, 13–32. Stanford: CSLI Publications.

Bresnan, Joan & Anna Cueni & Tatiana Nikitina & Harald R. Baayen. 2007. Predicting the dative alternation. In Gerlof Boume & Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Arts and Sciences.

Bresnan, Joan & Marilyn Ford 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86(1). 168–213.

Burnett, Heather & Sali A. Tagliamonte & Hilda Koopman. 2018. Soft Syntax and the Evolution of Negative and Polarity Indefinites in the History of English. *Language Variation and Change* 30(1). 83–107.

Childs, Claire & Christopher Harvey & Karen Corrigan & Sali A. Tagliamonte. 2015. Comparative sociolinguistics insights in the evolution of negation. *Selected Papers from New Ways of Analyzing Variation (NWAV) 43*, Philadelphia, PA: University of Pennsylvania Working Papers in Linguistics, 21, Article 4.

D'Arcy, Alexandra. 2017. *Discourse-Pragmatic Variation in Context: 800 years of like*. Amsterdam and New York: John Benjamins.

Du Bois, John W. & Giorgia Troiani. 2023. *Typology and its data: Functional monoculture or structural diversity?* Naturally occurring data in and beyond linguistic typology. Bologna. 18–19 May 2023, Bologna.

Evans, Nicholas. 2010. 504 Semantic Typology. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, 0. Oxford University Press.

Gardner, Matt Hunt & Sali A. Tagliamonte. 2020. The bike, the back, and the boyfriend: Confronting the “definite article conspiracy” in Canadian and British English. *English World-Wide* 41(2). 226–255.

Godfrey, John J. & Edward C. Holliman & Jane McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. *Acoustics, Speech, and Signal Processing*, 517–520. New York: Institute of Electrical and Electronics Engineers.

Grannis, Oliver C. 1972. The Definite Article Conspiracy in English”. *Language Learning* 22. 275–289.

Haspelmath, Martin. 2021. Explaining grammatical coding asymmetries: Formfrequency correspondences and predictability. *Journal of Linguistics* 57(3). 605–633.

Hawkins, John. 2004. *Efficiency and complexity in gramamrs*. Oxford: Oxford University Press.

Hawkins, John A. 2014. *Cross-linguistic variation and efficiency*. New York: Oxford University Press.

Jelinek, Eloise & Richard A. Demers. 1983. The agent hierarchy and voice in some Coast Salish languages. *International Journal of American Linguistics* 49. 167–185.

Kayne, Richard S. 1998. Overt vs. covert movement. *Syntax* 1. 128–191.

Kayne, Richard S. 2000. *Parameters and Universals*. Oxford: Oxford University Press.

Keller, Rudi. 1994. *On language change: The invisible hand in language*. London and New York: Routledge.

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

McWhorter, John. 2002. What happened to English? *Diachronica* 19(2). 217–272.

Miestamo, Matti & Ksenia Shagal & Olli Silvennoinen & Chingduang Yurayong. 2023. *Typology and usage: Explanatory perspectives with special reference to negation*. Naturally occurring data in and beyond linguistic typology. Bologna. 18–19 May 2023, Bologna.

Poplack, Shana & Sali A. Tagliamonte. 1999. The grammaticalization of *going to* in (African American) English. *Language Variation and Change* 11(3). 315–342.

Quirk, Randolph, & Sidney Greenbaum & Geoffry Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. New York: Longman.

Song, Jae Jung. 2010. Setting the Stage. In Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*, 0. Oxford University Press.

Szmrecsanyi, Benedikt & Jason Grafmiller. 2023. *Comparative variation analysis: syntactic variation in World Englishes*. Cambridge: Cambridge University Press.

Szmrecsanyi, Benedikt, & Jason Grafmiller & Joan Bresnan & Anette Rosenbach & Sali A. Tagliamonte & Simon Todd. 2017. Spoken syntax in a comparative perspective: the dative and genitive alternation in varieties of English. *Glossa* 86. 1–27.

Szmrecsanyi, Benedikt & Bernhard Wälchli (eds.) 2014. *Aggregating dialectology, typology, and register analysis: Linguistic variation in text and speech*. Berlin: de Gruyter.

Tagliamonte, Sali A. 2000–2001. Vernacular Roots: A database of British dialects. Research Grant B/RG/AN 6093/APN11081. Arts and Humanities Research Board of the United Kingdom (AHRB).

Tagliamonte, Sali A. 2001–2003. Back to the roots: The legacy of British dialects. Research Grant. Economic and Social Research Council of the United Kingdom (ESRC). #R000239097.

Tagliamonte, Sali A. 2002. Comparative sociolinguistics. In J.K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *Handbook of language variation and change*, 729–763. Malden and Oxford: Blackwell Publishers.

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

Tagliamonte, Sali A. 2013. *Roots of English: Exploring the history of dialects.* Cambridge: Cambridge University Press.

Tagliamonte, Sali A. 2016. *Making Waves: The History of Variationist Sociolinguistics*. Malden and New York: Wiley-Blackwell Publishers.

Tagliamonte, Sali A. 2025. *Analysing sociolinguistic variation, 2nd edition*. Cambridge: Cambridge University Press.

Tagliamonte, Sali A. & Marisa Brook. 2016. *Adaptive change in sociolinguistic typology: The case of relative 'who'*. LSA Annual Meeting 2016, January 7-10, 2016. Washington, D.C., USA.

Tagliamonte, Sali A., Alexandra D'Arcy & Celeste Rodrigues-Louro. 2016. Outliers, impact and rationalization in linguistic change. *Language* 92(4). 824–849.

Tagliamonte, Sali A & Jennifer Smith. 2021. Obviously undergoing change: Adverbs of evidentiality in the UK and Canada over 100 years. *Language Variation and Change* 33(1). 1–25.

Taleb, Nassim Nicholas. 2007. *The Black Swan: Impact of the highly improbable*. New York: Random House.

Torres Cacoullos, Rena & Catherine E. Travis. 2019. Variationist typology: Shared probabilistic constraints across (non-)null subject languages. *Linguistics* 57(3). 653–692.

Trudgill, Peter J. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.

Walkden, George & Gemma Hunter McCarley & Raquel Montero & Molly Rolf & Sarah Einhaus & Henri Kauhanen. 2023. Sociolinguistic Typology Meets Historical Corpus Linguistics. *Transactions of the Philological Society* 121(3). 546–567.

Wolk, Christoph, Bresnan, Joan, Rosenbach, Anette & Szmrecsanyi, Benedikt. 2013. Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change. *Diachronica* 30(3). 382–419.

Zaenen, Annie & Jean Carlette & Gregory Garretson & Joan Bresnan & Andrew Koontz-Garboden & Tatiana Nikitina, Catherine O'Connor & Tom Wasow. 2004. *Animacy encoding in English: why and how*. Proceedings of the 2004 ACL workshop on discourse annotation. Barcelona. July 2004.

CONTACT

sali.tagliamonte@utoronto.ca