

Matt Hunt Gardner* and Viktorija Kostadinova

Gettin' sociolinguistic data remotely: comparing vernacularity during online remote versus in-person sociolinguistic interviews

<https://doi.org/10.1515/lingvan-2022-0069>

Received June 7, 2022; accepted September 5, 2023; published online January 9, 2024

Abstract: The following paper examines the use of the stable sociolinguistic variable (*-ing*) across two different interview modalities: “classic” in-person sociolinguistic interviews and identical interviews conducted remotely over online video chat. The goal of this research was to test whether a change in modality results in style-shifting, as quantified by different rates of formal/standard [-ɪŋ] versus informal/non-standard [-ɪn]. Results show that when the internal linguistic constraints governing (*-ing*) variation are taken into account, there is not a significant difference between modalities, suggesting both modalities are equally formal (or informal). This suggests that remote online video chats are a viable method for collecting sociolinguistic data.

Keywords: sociolinguistic interviews; (ING); COVID-19 pandemic; language variation; style-shifting

1 Introduction

Adjusting data collection as a consequence of the safety measures precipitated by the outbreak of COVID-19 raised urgent questions about data quality, validity, and comparability, along with technical dilemmas for researchers across the social sciences. For sociolinguists, the central and highly valued types of data have traditionally been those collected in naturalistic settings, through face-to-face conversations, when speakers (ostensibly) pay the least amount of attention to speech (e.g., Labov 1984) – a very difficult type of data to collect amid travel restrictions, masking, and mandatory social distancing. As Poplack (1993: 252) succinctly lays out, the “relatively homogeneous, spontaneous speech reserved for intimate or casual situations [is] taken to reflect the most systematic form of the language acquired by the speaker, prior to any subsequent efforts at (hyper-)correction or style shifting”. Great care is thus taken to capture straightforwardly casual, vernacular speech by, among other things, minimizing the observer’s paradox (Labov 1972a: 209). Transitioning to remote data collection requires replacing face-to-face sociolinguistic interviews with a similar strategy adapted for videoconferencing software. But does communicating through Zoom, Skype, FaceTime, and so on, alter how much attention is paid to speech, the influence of the observer’s paradox, or the ability to capture speech that is “naturalistic”? The present study addresses these questions by quantifying English vernacularity or degree of self-monitoring as a function of formal/standard velar (*-ing*) use (see Labov 2001b, among many other references), and comparing the same participants under two different speaking conditions: face-to-face sociolinguistic interviews and sociolinguistic interviews conducted via video chat. Our driving hypothesis is that if video chatting as a modality is a fundamentally less “naturalistic” way of communicating, the overall rate of [-ɪŋ] use for (*-ing*) should be higher (as [-ɪn] generally decreases and [-ɪŋ] increases in use as speakers pay more attention to their speech).

*Corresponding author: Matt Hunt Gardner, University of Oxford, Oxford, UK, E-mail: matt.gardner@ling-phil.ox.ac.uk. <https://orcid.org/0000-0002-1878-4232>

Viktorija Kostadinova, University of Amsterdam, Amsterdam, Netherlands, E-mail: v.kostadinova@uva.nl. <https://orcid.org/0009-0005-1911-8329>

2 Sociolinguistic interviews and style shifting

The sociolinguistic interview, at its most basic, involves a researcher sitting down with a person face-to-face, posing questions, and recording the answers using good quality equipment (Meyerhoff et al. 2015: 47). This method is the “bread-and-butter data collection method of variationist sociolinguistics” (Kendall 2010: 351). It is the “premier data collection tool of the variationist sociolinguist” (Schilling 2013: 92). It is the “only means of obtaining the volume and quantity of recorded speech that is needed for quantitative analysis” (Labov 1984: 29). In other words, the sociolinguistic interview is integral to *doing* sociolinguistics. Thus a crisis, like a global pandemic, which disrupts the ability to do sociolinguistic interviews also disrupts the ability to do variationist sociolinguistics.

When Labov (1984) describes the optimal sociolinguistic interview, he envisions a discourse event, led by the researcher, with the participant(s) eventually engaging in speech with the least amount of self-monitoring. This speech style, according to Labov, comes closest to speech that is not influenced by being overtly observed by a researcher. As Labov states (1972a: 113), “to obtain the data most important for linguistic theory, we have to observe how people speak when they are not being observed”. Overcoming this observer’s paradox requires putting interview participants at ease (McPherson and Smoke 2019: 60), which relies both on the skills of the interviewer, who aims to build rapport, and the structure of the interview questions, which should be community-specific and carefully designed to lead interview participants from general, impersonal, non-specific topics or questions to more specific, personal ones that (usually) involve nostalgia or heightened emotions (see also Tagliamonte 2006). These latter types of questions elicit personal narratives, in which “community norms and styles of personal interaction are most plainly revealed, and where style is regularly shifted towards the vernacular” (Labov 1984: 32).

Differences in elements of a sociolinguistic interview have been shown to correlate with differences in speech elicited from the same participants. For example, in Fischer’s (1958) investigation on (-*ing*) variation among New England schoolchildren different interview types elicited very different rates of formal/standard [-*m*] use. This same pattern of style-shifting for different linguistic tasks was corroborated by Labov (1966), Wolfram (1969), Trudgill (1974a, 1974b), and others. Labov (1972b) argues that speakers decrease the rate at which they use non-prestigious speech features as a function of the degree to which they are self-monitoring, and that more formal situations or situations in which participants are made to focus on how they are speaking beget more self-monitoring.

Similarly, different interviewers have been found to elicit differing rates of non-standard features from the same participant (e.g., Rickford and Price 2013), and, complicating matters, interviewers themselves have been shown to use different rates of non-standard features (specifically non-standard [-*m*]) with different interview participants (e.g., Kendall 2010). The relationship between speech style and audience is well documented (Bell 1984; Coupland 1980), as is that between speech style and conversational topic (e.g., Bell 2001; Grieser 2019; Kiesling 2009; Labov 1966). Ervin-Tripp (2001) argues that a change in the circumstances of a speech event – like, for example, moving from an in-person to an online video chat – can have equally if not more profound effects on the linguistic features used:

Circumstantial shifts can change [linguistic] features not because of addressee behaviour or stereotypes about the addressee, but because the psycholinguistics of production and feedback are altered [...] These changes do not necessarily bear on, for instance, dialectal features, but they can so deeply alter the possibilities to edit or monitor speech that the role of dominant norms or stereotypes in production can be affected. (Ervin-Tripp 2001: 48)

Writing a decade ago, Sindoni (2011: 220) concluded that “videochats are used by young adults for everyday, spontaneous conversations, much alike to what happens in face-to-face interactions. However, web-based interactions differ in significant ways from face-to-face verbal exchanges and, more generally, from traditional social interactions”. This would suggest that conversational data elicited from videoconferencing tools represents a fundamentally different type of speech event and is thus not comparable to in-person interviews. Kiesling (1998), for example, found that among 11 fraternity men, rates of (-*ing*) variants were not consistent across

different speech events, such as sociolinguistic interviews, fraternity meetings, and casual socializing. In summary, while the exigences of a pandemic make online remote interviewing tempting, without empirical evidence to show otherwise, there is the danger that the modality of videoconferencing itself could trigger more monitored, formal, and potentially (hyper-)corrected speech.

In the context of the COVID-19 pandemic, the use of videoconferencing software to collect sociolinguistic data *has* been evaluated, but only (to our knowledge) insofar as it is appropriate for (socio)phonetic study (e.g., Calder and Wheeler 2022; Calder et al. 2022; Freeman and De Decker 2021; Sanker et al. 2021; Zhang et al. 2021). Though not an evaluation of *sociolinguistic* interviews, Bleaman et al.'s (2022) evaluation of interview modalities from publicly available talk show recordings is germane to the present study. They compare articulation rate, vowel space size, and use of (-ing) variants among guests of *The Late Show with Stephen Colbert* who were interviewed in-studio in the months prior to the pandemic and then again during the pandemic via Zoom. They find significant differences in articulation rate and vowel space size across interview modalities, with speakers slowing down their speech and using more of the vowel space during Zoom interviews. The authors label this “medium shift”, which they argue is motivated by speakers' desire to maximize intelligibility. They also argue medium shift is distinct from style-shifting as there was no significant difference in non-standard (-ing) use between modalities. Bleaman et al. (2022) posit that medium shift is likely non-salient and that it may be short-lived without lasting consequences. As people become more familiar with videoconferencing software the instinct to medium shift will wane.

So is speech captured using videoconferencing software as informal as speech captured in person? To answer this question, in the following sections, we use the diachronically stable, yet style-stratified variable (-ing) to assess the relative formality of each modality. We use (-ing) as a marker of vernacularity because of its stability through time and across varieties of English. We hypothesize that if videoconferencing conversations are perceived by speakers to be more formal they will pay more attention to what they are saying, and, in turn, use more of the prestigious [-ɪŋ] variant. Though we acknowledge the limitations of the attention-to-speech approach to style (see discussion in Eckert and Rickford 2001), we consider this approach to be reasonable and reliable in operationalizing style in the specific context of comparing the same individuals across two modalities.

3 (-ing)

The alternation in English production between an alveolar [-ɪn] and a velar [-ɪŋ] in words like *feeling*, *something*, and *morning*, as in (1),¹ “has been a staple of sociolinguistic research since the advent of the modern field” (Hazen 2006: 581).

- (1) a. *Well, she was since dead. And I remember the first night I slept in there, I woke up a couple of times with the feel[-ɪn] of someth[-ɪn] go[-ɪn] like this on the bed. And I told Nana the next morn[-ɪn]. I wasn't scared, because at that age, you're not scared. Like, you know? And I told her and she said “Oh,” she said, “that was just Grandma com[-ɪn] back, upset you were sleep[-ɪn] in her bed.” But, I don't know if that's what it was, or if she was com[-ɪn] in to check on me and I didn't wake up. I don't know. Like, you know, but that used to be the old battleaxe's room anyway.* (OF, in-person interview)
[associated audio-1a-GardnerKostadinova.mp3 with example (1a)]
- b. *They didn't teach music or anyth[-ɪn]. You had gym and then you had your regular classes. Ok? So anyth[-ɪn] extra was done by the teachers after the fact. Ok? So, uh, they'd say, “Alright, we're go[-ɪn] to have like a study group”, or “a read[-ɪn] group” or, you know. But now the ball and the curl[-ɪn] and stuff like that was all well-organized through the school.* (OF, online remote interview)
[associated audio-1b-GardnerKostadinova.mp3 with example (1b)]

¹ Here and throughout we include extended examples of speech from our sample of Cape Breton English (see Gardner 2017, and Section 4) to showcase the comparable informal nature of both the in-person and online remote interviews. Speaker labels indicate age (older, boomer, or millennial), sex (male or female) and speaker ID (for millennials only).

The general finding across L1 adult varieties of English has been that the alveolar variant of *[-ing]*, *[-m]*, is more common among speakers with lower social status than among speakers with higher social status. The variable is salient, and thus when speakers are using formal styles of speech, their rates of the formal variant, *[-m]*, increase. This has been tested by researchers by controlling the relative formality of linguistic tasks (e.g., Labov 1966; Trudgill 1974b). The variable is considered to be stable diachronically in Modern English (e.g., Labov 2001a).

Across studies, several internal linguistic constraints underlie the variation between *[-m]* and *[-n]*, including preceding and following phonological context, number of syllables in the word, and word-type (Tagliamonte 2012: 190). Function words like indefinite pronouns, as in (2a), and prepositions, (2b), favour *[-m]* relative to lexical words. The *[-m]* form also consistently shows higher frequencies for the progressive, as in (2c), lower frequencies for participles, (2d), and adjectives, (2e), and the lowest frequencies for gerunds and nouns, as in (2f) and (2g) (Labov 2001b: 87). These constraints likely reflect the separate origin of the two variants (see Houston 1985 and the analysis therein).

- (2) a. *I think **something**[-m]’s been going on.* (MF3, online remote interview)
[associated audio-2a-GardnerKostadinova.mp3 with example (2a)]
- b. *We used to go on hikes all the time and, uh, camping one or twice **dur**[-m] the summer. Uh, and it was a good group of guys we had there.* (OM, online remote interview)
[associated audio-2b-GardnerKostadinova.mp3 with example (2b)]
- c. *Oh. Oh they’re **tak**[-m] down the wall, right?* (MF1, online remote interview)
[associated audio-2c-GardnerKostadinova.mp3 with example (2c)]
- d. *I walk most days. I’m going to start **gett**[-m] the bus, I think, when it’s cold.* (MF2, in-person interview)
[associated audio-2d-GardnerKostadinova.mp3 with example (2d)]
- e. *That’s an **interest**[-m] question since this year we did nothing.* (BF, in-person interview)
[associated audio-2e-GardnerKostadinova.mp3 with example (2e)]
- f. *I would have a hard time not **hav**[-m] a dishwasher.* (MF3, online remote interview)
[associated audio-2f-GardnerKostadinova.mp3 with example (2f)]
- g. *Yeah, that’s quite a **ceil**[-m].* (BM, online remote interview)
[associated audio-2g-GardnerKostadinova.mp3 with example (2g)]

4 Methods

To compare in-person and online remote sociolinguistic interviews, we revisit a set of sociolinguistic interviews collected by the first author in Cape Breton, Nova Scotia, Canada, in 2009–2011 (Gardner 2017). These interviews were guided by Labov (1984) and Tagliamonte (2006). Interview questions were community-specific, but based on the interview schedule included as an appendix to Tagliamonte (2006),² itself adapted from Labov (1966, 1984). The *relative* vernacularity of this data has not been assessed; however, it has been used to demonstrate systematic phonological and morphosyntactic vernacular patterns (e.g., Gardner 2017; Roeder and Gardner 2013). Additionally, the data contains many hallmarks of casual speech, including non-standard pronunciation, word forms, and sentence structure; high rates of vowel and function word reduction or deletion; frequent use of discourse markers; use of expletives; and much laughter, joking, and sharing of personal stories, as in (1).

New interviews were conducted in 2021 with nine of the original participants in the 2009–2011 data. The number of secondary interviews is small, as it was incredibly challenging to reconnect with former participants. The nine participants who were interviewed a second time include one man and one woman born between 1935 and 1945 (labelled “older”), one man and one woman born between 1955 and 1965 (labelled “boomers”) and two men and three women born in the 1980s (labelled “millennials”). In 2009–2011 the older speakers were retired (and were formerly working class). In 2021 they were still retired. In 2009–2011 the two boomers were white-collar professionals. In 2021, the boomer man was retired and the boomer woman had changed careers but was working

2 Available at https://www.cambridge.org/pk/files/5213/6689/9619/2846_APPENDIX_B.pdf (accessed 8 December 2023).

as a manager in the financial industry. In 2009–2011 three of the millennials were university students, while both millennial man 1 and millennial woman 2 had recently graduated and were working as language teachers (though not as linguists). In 2021, all millennials were working in white-collar jobs. Aside from the older speakers, the returning participants are all university-educated. All but millennial woman 2 continue to live in the region.

The online remote interviews followed the same interview protocol as the original data collection. Participants were invited to join a Zoom meeting set up by the first author. Participants joined the online meeting using the device of their choosing (mobile phone, tablet, or computer) and from the setting of their choosing. Interviews lasted approximately 40 min and were recorded using Zoom's built-in audio and video recording function. Participants were aware that they were being recorded, and consented to such. An advantage of using Zoom is that it allows for the independent recording of each person in a meeting, which facilitated the automatic transcription of just the participants' audio feed. Participants were not given any instruction for capturing their voice while taking part in the Zoom meeting. Some participants used the built-in microphone of their mobile device, some used the built-in microphone of their computer, others used Bluetooth wireless or wired headphones with a built-in microphone. As the acoustic quality of the recorded audio, aside from being clear enough to distinguish *(-ing)* variants, was not a priority, we decided that having participants interact with their mobile devices or computers in whatever way they chose and not overtly commenting on their recording set-up promoted a more casual interaction.

Transcription of both the first and second round of interviews, as well as extraction and coding of the relevant tokens, was facilitated by a number of automated tasks preceding a final manual coding and verification of the data set. The interviews were transcribed with automatic speech recognition in Python, using AssemblyAI's speech-to-text API (<https://www.assemblyai.com>). Each *(-ing)* token in the transcribed text was extracted, together with its timestamp in the interview, and the broader context in which it occurred. To aid coding for grammatical function, we used the Python NLP library spaCy (Honnibal and Montani 2017) to tag the part of speech for each *(-ing)* token. This automatic coding was manually verified. The data was then imported into ELAN for impressionistic coding of the dependent variable (either [-m] or [-ɪ]), and phonological context.

4.1 Exclusions

As *(-ing)* variation occurs only in multisyllabic words, monosyllabic word-final *-ing* are not included in the analysis. Further, we exclude tokens of the verb *going*, either as a lexical verb, as in (3a), as the first element of a coordinated verb construction (Stefanowitsch 2000), as in (3b), or as a pseudo- or semi-modal future marker, as in (3c). In the variety, all forms of *going* are nearly always reduced in speech (cf. Labov et al. 1968: 250–253; Pullum 1997: 87).

We also chose to set *trying to* aside ($N = 50$), as in (3d), because it is very frequently (88 %, $n = 44$) pronounced as some version of *tryna* in the data. The reduction of *going (to/and)* and *trying to* are independent processes of grammaticalization or reduction adjacent to, but not linked to, variation between [-ɪ] and [-m] for *(-ing)*. This is consistent with other analyses of *(-ing)* (e.g., Hazen 2008; Houston 1985; Labov et al. 1968; Tagliamonte 2004, etc.). The prevalence of *gonna* and *tryna* in our data is indicative of vernacular speech (e.g., Gonzalez 2020) – a fact that further suggests our in-person and Zoom interviews are comparable.

- (3) a. *I have a gut feel[-ɪ] someth[-ɪ]'s been **going** ['gowən] on for a while.* (MF3, online remote interview)
[associated audio-3a-GardnerKostadinova.mp3 with example (3a)]
- b. *And then we ended up **going and** ['gowən.ən] buy[-ɪ] a lot of stuff.* (BF, online remote interview)
[associated audio-3b-GardnerKostadinova.mp3 with example (3b)]
- c. *And the thing is, when Boston Pizza first put their, like, chalk line out for where they were **going to** ['gə.ñə] put their build[-ɪ], everyone was like, "Ok, they're just tak[-ɪ] up space from the movie theatre. No one's **going to** ['gə.ñə] go there. It's not **going to** ['gə.ñə] be popular". But it's quite popular.* (MM1, in-person interview)
[associated audio-3c-GardnerKostadinova.mp3 with example (3c)]

- d. *I wanted to do someth[-ɪŋ] that I- seemed straightforward, um. Some classes looked more complicated than others, um. And I wasn't comfortable, like, at first, to start, like, **trying to** ['tʁaɪn.tə] like, do, like, different, like, role-play[-ɪŋ] stuff. So I was like, "What if I just played, like, a meek, nice lady?"* (MF1, online remote interview)
[associated audio-3d-GardnerKostadinova.mp3 with example (3d)]

5 Results

After the above exclusions, a total of 1,028 (-ing) tokens were extracted from the in-person interviews and 811 tokens were extracted from the online remote interviews. Figure 1 shows the overall distribution of (-ing) variants by speaker across the two interview modalities and 10 years of real time. Compared to younger speakers, both older speakers show much lower rates of the formal/standard [-ɪŋ] variant. Millennial speakers, on the other hand, use very high rates of standard [-ɪŋ]. Across modalities/real time, each speaker has remarkably similar, but not identical rates of [-ɪŋ]. For example, the four oldest speakers and millennial woman 3 use less [-ɪŋ] in the online interviews, while millennial man 1 and millennial woman 1 use slightly more [-ɪŋ] during the online interviews. Whether these small differences across modalities and real time are statistically significant, or whether the differences in rates may be due to differences in the internal linguistic constraints governing the variation between [-ɪŋ] and [-ɪŋ] is analysed below.

In order to test whether a change in interview modality significantly affected our participants' use of (-ing), we controlled for the following internal linguistic constraints, based on the extant literature on (-ing) variation: phonological context and grammatical category.

5.1 Phonological context

Not all studies find phonological environment to be a significant constraining factor for (-ing). Labov (2006: 255) claims that (-ing) variation "takes place at the morphological rather than the phonological level"; however, Tagliamonte (2004) does find number of syllables, preceding phonological context, and following phonological context to condition specifically nominal (-ing) words in York, in the UK. Hazen (2008) also found an effect for

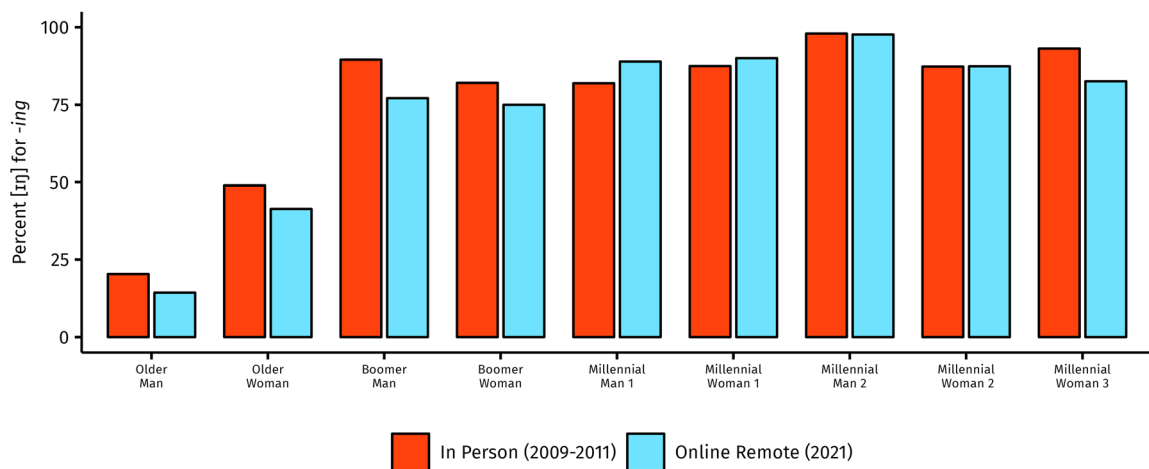


Figure 1: Percentage of formal/standard [-ɪŋ] for (-ing) by speaker and interview modality/time.

phonological context in his sample of US Appalachian speech, though this effect interacted heavily with grammatical part of speech (see also Hazen 2014: 49–51). In our data, the rate of standard [-ɪŋ] is lower in two-syllable words (74 %, 1,075 of 1,446) compared to longer words (95 %, 490 of 517). This is consistent with Tagliamonte (2004). As with Hazen (2014), there is substantial collinearity between number of syllables and grammatical part of speech. For example, 99 % (82 of 84) of noun tokens have two syllables. For this reason we chose not to include number of syllables as a potential control constraint in our inferential statistical modelling (which assumes little to no collinearity between predictors) in favour of including grammatical category (the more commonly attested predictor).

Table 1 shows the rate of [-ɪŋ] by preceding sound. The relatively higher rate of [-ɪŋ] after dorsals and [-ɪŋ] after coronals coincides with findings reported by Tagliamonte (2004) and Hazen (2008); conversely, Hazen (2008) reports higher rates of [-ɪŋ] with bilabials compared to labiodentals, while we report the inverse.

As with syllable type, preceding sound is collinear with grammatical category. For example, all indefinite pronouns have preceding coronals, all prepositions have preceding liquids, and 92 % (77 of 84) of nouns have either preceding coronals or liquids. Again, in order to prioritize grammatical category, preceding sound was not included in the inferential statistical analysis.

Our findings for following sound do not align with those of Tagliamonte (2004) and Kendall (2010), who found that [-ɪŋ] was most likely before a dorsal consonant and less likely before other consonants, vowels, and pauses. Importantly, Tagliamonte (2004) excludes some following contexts that we choose to include. The first is (-ing) followed by /g/. While neutralization is possible here, we did find variation: of the 15 tokens, two (13 %) occur with [-ɪŋ], as in (4). If these highly-[-ɪŋ]-favouring pre-/g/ contexts are removed, our data gets less like Tagliamonte (2004), rather than more like it with respect to the frequency of [-ɪŋ] before dorsal consonants.

- (4) a. *The basement was full of parts and doodads and the frigg[-ɪŋ] **garage** was filled, and ...* (BF, online remote interview)
[associated audio-4a-GardnerKostadinova.mp3 with example (4a)]
- b. *Honestly, I- I can tell when a guy is a good-look[-ɪŋ] **guy**.* (MM1, in-person interview)
[associated audio-4b-GardnerKostadinova.mp3 with example (4b)]

The second key difference in how we coded following phonological context was the differentiation between oral and nasalized vowels. Words that began with vowels immediately followed by a nasal sound were categorized as nasalized vowels. These consisted mostly of words like *on*, *in*, and *and*, and the reduced forms of *him* [ɪm] and *them* [ɛm] – both of which can also be realized as [ɪm]. The [-ɪŋ] variant occurred more frequently before these nasalized vowels (83 %) compared to before non-nasalized vowels (71 %); see Table 2.

- (5) *I love bring[-ɪŋ] **him** [ɪm] out around people and like introduc[-ɪŋ] **him** [ɪm] and show[-ɪŋ] **him** [ɪm] off.* (MM1, in person interview)
[associated audio-5-GardnerKostadinova.mp3 with example (5)]

Table 1: Number of occurrences of [-ɪŋ] (*n*), relative proportion of [-ɪŋ] (%), and total number of (-ing) tokens (Total *N*) by preceding sound for all speakers and both interview modalities/times.

Preceding sound	<i>n</i>	%	Total <i>N</i>
Coronal	869	86	1,007
Labiodental	96	83	115
Liquid	114	76	115
Vowel	200	74	270
Bilabial	104	73	142
Dorsal	182	65	279

Table 2: Number of occurrences of [-ɪŋ] (*n*), relative proportion of [-ɪŋ] (%), and total number of (-ing) tokens (Total *N*) by following sound for all speakers and both interview modalities/times.

Following sound	<i>n</i>	%	Total <i>N</i>
Pause	197	94	209
Nasal consonant	114	86	168
Nasalized vowel	113	83	136
Dorsal consonant	96	83	115
Labial consonant	209	82	254
Coronal consonant	454	78	583
Oral vowel	352	71	498

5.2 Grammatical category

Labov (1989: 87) reports an implicational hierarchy of grammatical conditioning for (-ing) with [-ɪŋ] occurring “most in progressives and participles, less in adjectives, even less in gerunds and least of all in nouns”. This hierarchy reflects the history of the two variants, which derived from an alveolar morpheme marking participles and a velar morpheme marking verbal nouns (Houston 1985; Labov 2001a: 88). In Modern English, words that are more noun-like favour [-ɪŋ] relative to words that are more verb-like. Although the exact classification of words as either noun-like or verb-like has been variable from study to study, the general finding that [-ɪŋ] is favoured by participles and [-ɪŋ] is favoured by nouns is consistent (e.g., Hazen 2008; Tagliamonte 2004). In our analysis, we collapse progressives and participles. We further do not distinguish gerundial nouns and gerundial participles (Huddleston and Pullum 2002).

Table 3 shows that nouns in our data do occur most frequently with [-ɪŋ] relative to other grammatical parts of speech while participles occur with [-ɪŋ] the least. Given the small number of proper noun tokens in our data, we merge these with other nouns in our subsequent analyses.³

Prior to regression modelling we subjected our data to a conditional inference tree (CIT) analysis⁴ to determine if our elaborated coding scheme for each linguistic control predictor could be simplified. A CIT analysis can “expose the quantitative structure of a data set, pinpointing fine-grained distinctions among predictors” (Tagliamonte et al. 2016: 832). The partitioned trees indicate where there are statistically significant differences between predictor levels if all other input predictors are simultaneously considered (Hothorn et al. 2006; Schweinberger 2023). Figure 2 shows that for grammatical function, the significant distinction is gerunds/

Table 3: Number of occurrences of [-ɪŋ] (*n*), relative proportion of [-ɪŋ] (%), and total number of (-ing) tokens (Total *N*) by grammatical category for all speakers and both interview modalities/times.

Grammatical category	<i>n</i>	%	Total <i>N</i>
Proper noun	7	100	7
Noun	75	97	77
Adjective	173	90	192
Indefinite (pronoun)	294	87	339
Preposition	11	85	13
Gerund	261	76	342
Participle	626	72	869

³ Other analyses of (-ing) exclude proper nouns; however, our seven tokens were all *Thanksgiving*, which does alternate with *Thanksgivin'* in the broader Cape Breton English corpus, so we chose to include them.

⁴ The CIT analysis was performed in *R* (R Core Team 2021) using the partykit package (Hothorn and Zeileis 2015).

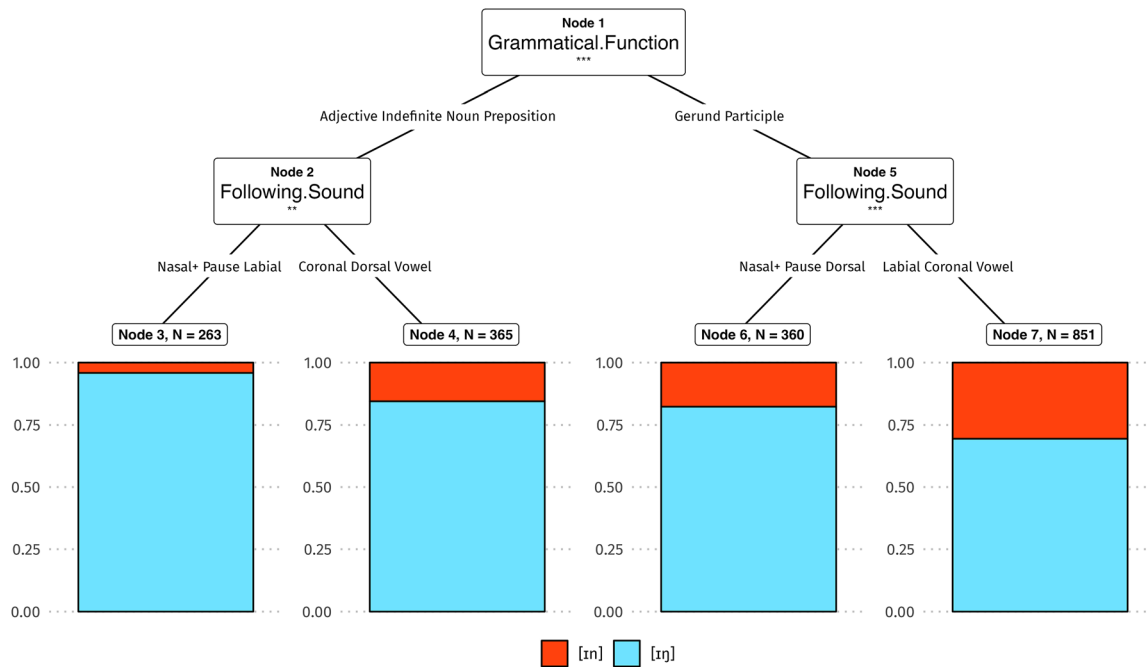


Figure 2: Conditional inference recursive partitioning tree for the realization of *(-ing)* as *[-ɪŋ]* or *[-ɪn]* with grammatical function, following sound, and interview modality as input parameters ($N = 1839$). Nasal+ includes nasal consonants and nasalized vowels. $***p < 0.01$. $***p < 0.001$.

participles versus all other grammatical categories. It also shows that while following nasals+ (nasal consonants and nasalized vowels) and pauses favour *[-ɪn]* relative to following coronals and vowels, whether following dorsals and labials favour *[-ɪn]* depends on whether *(-ing)* is part of a gerund/participle or another grammatical category. Given the findings of the CIT analysis we choose to collapse grammatical function into a two-way predictor (gerund/participle vs. nouns, etc.) and (initially) keep the six-way distinction for following sound. Subsequent analyses (not shown), however, indicated that the relevant distinction once the additional random effect of speaker was considered is between pause and all other following sounds. With respect to our research question, the CIT analysis indicates that interview modality, which was included as an input parameter, adds no explanatory value to describing the variation between *[-ɪn]* and *[-ɪŋ]*. This would justify our concluding that there is little difference in relative vernacularity between data collected in person or remotely. Nevertheless, using the linguistic predictors as controls, we further conducted mixed-effects logistic regression modelling to determine whether the small differences that do exist between the interview modalities observed in Figure 1 are statistically significant.

Table 4 presents a mixed-effects logistic regression model of *[-ɪn]* for *(-ing)* among the nine participants in the data.⁵ There are 1,839 *(-ing)* tokens analysed, of which 79 % are *[-ɪn]*. The model includes the fixed effects of grammatical function, following sound, and interview modality, fixed interactions between interview modality and the other two predictors, as well as the random effect of individual speaker. The Akaike information criterion (AIC) is an estimator of the prediction error of this model. A lower AIC is a better fit (more explanatory) than a higher AIC. The AIC of a model built with just the random effect of speaker is 1,579, which is significantly higher than that of the model presented ($\chi^2 = 130.46$, $df = 7$, $p < 0.001$).⁶ The marginal and conditional R^2 values show the proportion of the data explained by the fixed effects predictors and the fixed effects predictors plus the random effect respectively. For this model, the fixed effect predictors are assumed to have a constant relationship with the

5 Models were built in *R* (R Core Team 2021) using the *lme4* package (Bates et al. 2015, 2023). Additional calculations were made using functions from the following packages: *multcomp* (Hothorn et al. 2008); *MuMIn* (Bartón 2019); *JGmermod* (Graffmiller 2018); and *performance* (Lüdtke et al. 2021).

Table 4: Mixed-effects logistic regression testing the fixed effects of Grammatical Function, Following Sound, and Interview Modality, a fixed interaction between Grammatical Function and Interview Modality, a fixed interaction between Interview Modality and Following Sound, and a random intercept of Speaker on the realization of [-ɪŋ] for (-ing) among nine speakers of Cape Breton English. Treatment contrast coding. Model fit by maximum likelihood (Laplace approximation). Model converges with BOBYQA optimizer with <20,000 iterations. Coefficients reported in log odds.

Observations: 1,839 (overall frequency of [-ɪŋ] for (-ing) 79 %, Total N = 1,447)							
AIC: 1,459.2, Marginal R ² : 0.14, Conditional R ² : 0.50							
Fixed effects	Coef.	SE	z	Sig.	Observations		
					Total N	[-ɪŋ]	%
Intercept (nouns, etc. + pause + in-person)	4.60	0.79	5.87	***	1,335	1,077	81
Grammatical function (vs. nouns, etc.)					66	64	97
Participle & gerund	-1.56	0.24	-6.62	***	1,211	887	73
Following sound (vs. pause)					193	182	94
All other sounds	-2.06	0.58	-3.57	***	1,646	1,265	77
Interview modality (vs. in-person)					1,028	810	79
Online remote	-0.85	0.78	-1.09		811	637	79
Interaction: Interview modality (online remote) and grammatical function (participle & gerund)	0.22	0.34	0.65		45	39	87
Interaction: Interview modality (online remote) and following sound (all other sounds)	0.49	0.75	0.65		224	191	85
Random effects	SD		Group N				
Speaker (intercept) ^a	1.54		9				

Notes. *** $p < 0.001$. ^aMean by speaker = 79 ± 9.2 %.

response variable (choice of [-ɪŋ] or [-ɪn]) across all observations, but, by setting speaker as a random effect, we allow for that fixed relationship to vary from person to person (i.e., accounting for participants having a differing overall likelihood of [-ɪŋ]). The mean rate of [-ɪŋ] use per speaker is 75 %, with a standard deviation (SD) of ± 9.2 %.

In Table 4, the model is built with treatment contrast coding, whereby the coefficient of the intercept represents the likelihood (expressed in log odds) of [-ɪŋ] when each of the fixed effect predictors are set to a reference value. These reference values each have the highest frequency of [-ɪŋ] (as shown by the distributions under “observations”), though this choice is arbitrary. The coefficients for each non-reference value represent the change in likelihood of [-ɪŋ] if that parameter is switched. Coefficients are reported in log odds, which range from $-\infty$ to $+\infty$ and are centred around zero. Positive coefficients indicate an increased likelihood of [-ɪŋ] (relative to the likelihood of the reference value), negative coefficients indicate a decreased likelihood of [-ɪŋ]. Based on the standard error (SE) and z-score, whether this difference in likelihood is significant is calculated and represented by asterisks. For example, the negative coefficient for participle and gerund indicates [-ɪŋ] is less likely for this group compared to nouns, etc., and the three asterisks indicate this difference is significant ($p < 0.001$). Conversely, even though the likelihood of [-ɪŋ] is lower for the online remote data, given its negative coefficient, compared to the in-person data, this difference is not significant.

The relevant finding in Table 4 for our present investigation is that when linguistic factors are accounted for, the difference between interview modalities is not significant. Consequently, we must accept the null hypothesis that there is no difference between the in-person and online remote data with respect to (-ing) variation. The first corollary of this result is that both modalities elicit equally vernacular data. The second corollary of this result, given that our two modalities represent 10 years of real time, is that the (-ing) variable is, in fact, stable among these speakers.

6 Alternative models using number of syllables (AIC = 1,425.8) and preceding context (AIC = 1,425.9) instead of grammatical function, with which both are collinear, were built. In no model was there a significant effect of modality or a significant interaction of modality with another predictor.

Table 4 also includes fixed interactions between interview modality and grammatical category and between interview modality and following sound. These interaction terms permit testing whether the effects of grammatical category and following sound are significantly different when the level of interview modality changes from in-person to online remote. The coefficient for the interaction terms represents any additional change in likelihood of [-ɪŋ] when both predictors are set to non-reference values. As neither interaction term's coefficient is significantly different from zero, we must accept the null hypothesis that a difference in interview modality does not correspond to a difference in the effect of grammatical category or following sound. In other words, the effects of grammatical category and following sound are consistent across 10 years of real time, and in data collected via in-person interviews and remotely online.

6 Discussion and conclusion

Our analysis confirms previously established grammatical and phonological constraints on stable (-ing) and finds no effect of interview modality on its variation or on its governing constraints. This suggests that video chat may be a highly viable option for collecting sociolinguistic interview data.

The advantages of collecting sociolinguistic data using remote online sociolinguistic interviews are numerous. Using videoconferencing software is cheaper and faster and its built-in functions facilitate data storage, transcription, coding, and analysis. These advantages, however, should be weighed in light of a number of caveats.

First, we chose a stable socially and stylistically stratified variable, whose similar patterns of variation have been confirmed by a number of studies in different contexts. This might make it less likely to be subject to interview-modality constraints (see also Bleaman et al. 2022). Other variables, especially those related to clarity, articulation, or discourse organization, may be more susceptible to the effect of interview modality.

Second, the two sets of interviews were conducted by the same interviewer using the same interview protocol and were with the same speakers. This design was essential for our analysis; however, it did mean that during the second round of interviews the participants were already familiar with the interviewer. This relationship may have moderated the artificial nature of the video chat interviews and led to speech more informal than might have otherwise been captured.

Third, our results may reflect the reality that for our nine speakers video chat is comfortable, familiar, and regular. The two older speakers and the retired boomer man note in their interviews that the only time they ever use videoconferencing software (FaceTime, Zoom, etc.) is when chatting with their friends and relatives. This may not be the case for all older speakers in all communities. The millennials and the female boomer speaker reported using videoconferencing software to communicate with friends and family too, but also for their jobs during (and even after) the restrictions imposed by the COVID-19 pandemic. Those who work in blue-collar jobs or who are from peripheral speech communities may have differing familiarity with videoconferencing software, so the likelihood of eliciting casual speech from these speakers via video chat may vary.

In sum, then, our results suggest that videoconferencing software is a viable option for collecting sociolinguistic data. If (-ing) is typical, sociolinguistic variables that are salient, socially marked, and prevalent should not be affected by interview modality. These results, however, should be interpreted in light of the above caveats. These caveats do point to additional empirical research that will help improve and refine the use of video chat in data collection in the future. Our small-scale study also contributes to the growing research on innovative data collection and lends credibility to researchers using videoconferencing software for studying language variation and change.

Acknowledgment: We thank our participants for agreeing to be re-interviewed after 10 years. We also thank our research assistant Jona van der Schelde, our audience at NWAV 49, and the Universitas 21 Researcher Relief Fund. The in-person data collection for this project took place on Cape Breton Island, known as Unama'ki (the Land of Fog), which is part of the ancestral and unceded territories of the Mi'kmaq people, who the authors acknowledge as the past, present, and future caretakers of the island.

References

- Bartón, Kamil. 2019. Mumin: Multi-modal inference. R package. Version 1.46.5. <https://CRAN.R-project.org/package=MumIn> (accessed 8 December 2023).
- Bates, Douglas M., Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software* 67(1). 1–48.
- Bates, Douglas M., Martin Maechler, Ben Bolker & Steven Walker. 2023. lme4: Linear mixed-effects models using ‘Eigen’ and S4. R package. Version 1.1-35.1. <https://CRAN.R-project.org/package=lme4> (accessed 8 December 2023).
- Bell, Allan. 1984. Language style as audience design. *Language and Society* 13. 145–204.
- Bell, Allan. 2001. Back in style: Reworking audience design. In Penelope Eckert & John R. Rickford (eds.), *Back in style*, 145–204. Cambridge: Cambridge University Press.
- Bleaman, Isaac L., Katie Cugno & Annie Helms. 2022. Medium-shifting and intraspeaker variation in conversational interviews. *Language Variation and Change* 34(3). 305–329.
- Calder, Jeremy & Rebecca Wheeler. 2022. Is Zoom viable for sociophonetic research? A comparison of in-person and online recordings for sibilant analysis. *Linguistics Vanguard*. 20210014. <https://doi.org/10.1515/lingvan-2021-0014> (Epub ahead of print).
- Calder, Jeremy, Rebecca Wheeler, Sarah Adams, Daniel Amarello, Katherine Arnold-Murray, Justin Bai, Meredith Church, Josh Daniels, Sarah Gomez, Jacob Henry, Yunan Jia, Brienna Johnson-Morris, Kyo Lee, Kit Miller, Derrek Powell, Caitlin Ramsey-Smith, Sydney Rayl, Sara Rosenau & Nadine Salvador. 2022. Is Zoom viable for sociophonetic research? A comparison of in-person and online recordings for vocalic analysis. *Linguistics Vanguard*. 20200148. <https://doi.org/10.1515/lingvan-2020-0148> (Epub ahead of print).
- Coupland, Nikolas. 1980. Style-shifting in a Cardiff work-setting. *Language in Society* 9(1). 1–12.
- Eckert, Penelope & John R. Rickford (eds.). 2001. *Style and sociolinguistic variation*. Cambridge: Cambridge University Press.
- Ervin-Tripp, Susan. 2001. Variety, style-shifting, and ideology. In Penelope Eckert & John R. Rickford (eds.), *Style and sociolinguistic variation*, 44–56. Cambridge: Cambridge University Press.
- Fischer, John L. 1958. Social influences on the choice of a linguistic variant. *WORD* 14(1). 47–56.
- Freeman, Valerie & Paul De Decker. 2021. Remote sociophonetic data collection: Vowels and nasalization over video conferencing apps. *The Journal of the Acoustical Society of America* 149(2). 1211–1223.
- Gardner, Matt Hunt. 2017. *Grammatical variation and change in Industrial Cape Breton*. Toronto: University of Toronto PhD dissertation. <https://hdl.handle.net/1807/80940> (accessed 8 December 2023).
- Gonzalez, Chloe. 2020. *Tryna*. Yale grammatical diversity project: English in North America. <http://ygdp.yale.edu/phenomena/tryna> (accessed 29 March 2022).
- Grafmiller, Jason. 2018. Jgmermod: Custom functions for mixed-effects regression models. R package. Version 0.2.0. <https://github.com/jasongraf1/JGmermod> (accessed 8 December 2023).
- Grieser, Jessica A. 2019. Investigating topic-based style shifting in the classic sociolinguistic interview. *American Speech* 94(1). 54–71.
- Hazen, Kirk. 2006. In/ing. In Keith Brown (ed.), *Encyclopedia of language & linguistics*, 2nd edn., vol. 5, 581–582. Oxford: Elsevier.
- Hazen, Kirk. 2008. (ing): A vernacular baseline for English in Appalachia. *American Speech* 83(2). 116–140.
- Hazen, Kirk. 2014. Methodological choices in language variation analysis. In Eugene Green & Charles F. Meyer (eds.), *The variability of current World Englishes*, 41–64. Berlin: De Gruyter.
- Honnibal, Matthew & Ines Montani. 2017. Spacy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. Available at: <https://spacy.io/>.
- Hothorn, Torsten, Frank Bretz & Peter Westfall. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50(3). 346–363.
- Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15(3). 651–674.
- Hothorn, Torsten & Achim Zeileis. 2015. Partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research* 16(118). 3905–3909.
- Houston, Ann Celeste. 1985. *Continuity and change in English morphology: The variable (ING)*. Philadelphia: University of Pennsylvania PhD dissertation. <http://repository.upenn.edu/edissertations/1183> (accessed 8 December 2022).
- Huddleston, Rodney D. & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Kendall, Tyler. 2010. Accommodating (ing): Individual variation in mixed-ethnicity interviews. In Barry Heselwood & Clive Upton (eds.), *Proceedings of methods XIII: Papers from the thirteenth international conference on methods in dialectology, 2008*, 351–361. Frankfurt am Main: Peter Lang.
- Kiesling, Scott F. 1998. Men’s identities and sociolinguistic variation: The case of fraternity men. *Journal of Sociolinguistics* 2(1). 69–99.
- Kiesling, Scott F. 2009. Style as stance. In Alexandra Jaffe (ed.), *Stance: Sociolinguistic perspectives*, 171–194. Oxford: Oxford University Press.
- Labov, William. 1966. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Labov, William. 1972a. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, William. 1972b. Some principles of linguistic methodology. *Language in Society* 1(1). 97–120.
- Labov, William. 1984. Field methods of the project of linguistic change and variation. In John Baugh & Joel Sherzer (eds.), *Language in use: Readings in sociolinguistics*, 28–53. Englewood Cliffs, NJ: Prentice Hall.

- Labov, William. 1989. The child as linguistic historian. *Language Variation and Change* 1(1). 85–97.
- Labov, William. 2001a. *Principles of linguistic change. Vol. 2: Social factors*. Oxford: Blackwell.
- Labov, William. 2001b. The anatomy of style-shifting. In Penelope Eckert & John R. Rickford (eds.), *Style and sociolinguistic variation*, 85–108. Cambridge: Cambridge University Press.
- Labov, William. 2006. *The social stratification of English in New York City*, 2nd edn. Cambridge: Cambridge University Press.
- Labov, William, Paul Cohen, Clarence Robins & John Lewis. 1968. *A study of the non-standard English of Negro and Puerto Rican speakers in New York City. Cooperative research project 3288. Vol. 1: Phonological and grammatical analysis*. Philadelphia: U.S. Regional Survey.
- Lüdecke, Daniel, Mattan S. Ben-Shachar, Indrajeet Patil, Philip Waggoner & Dominique Makowski. 2021. Performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software* 6(60). 3139.
- McPherson, Paul & Trudy Smoke. 2019. *Thinking sociolinguistically: How to plan, conduct, and present your research project*. London: Macmillan.
- Meyerhoff, Miriam, Erik Schlee & Laurel MacKenzie. 2015. *Doing sociolinguistics: A practical guide to data collection and analysis*. New York: Routledge.
- Poplack, Shana. 1993. Variation theory and language contact. In Dennis R. Preston (ed.), *American dialect research: Celebrating the 100th anniversary of the American Dialect Society, 1889–1989*, 251–286. Amsterdam: John Benjamins.
- Pullum, Geoffrey K. 1997. The morpholexical nature of English *to*-contraction. *Language* 73(1). 79–102.
- R Core Team. 2021. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Rickford, John & Mackenzie Price. 2013. Girlz II women: Age-grading, language change and stylistic variation. *Journal of Sociolinguistics* 17(2). 143–179.
- Roeder, Rebecca & Matt Hunt Gardner. 2013. The phonology of the Canadian Shift revisited: Thunder Bay and Cape Breton. *U. Penn Working Papers in Linguistics* 19(2). 18.
- Sanker, Chelsea, Sarah Babinski, Roslyn Burns, Marisha Evans, Jeremy Johns, Juhay Kim, Slater Smith, Natalie Weber & Claire Bower. 2021. (Don't) try this at home!: The effects of recording devices and software on phonetic analysis. *Language* 97(4). e360–e382.
- Schilling, Natalie. 2013. *Sociolinguistic fieldwork*. Cambridge: Cambridge University Press.
- Schweinberger, Martin. 2023. *Tree-based models in R*. Brisbane: The University of Queensland, Australia: School of Languages and Cultures. <https://slcladal.github.io/tree.html> (accessed 8 December 2023).
- Sindoni, Maria Grazia. 2011. Online conversations: A sociolinguistic investigation into young adults' use of videochats. *Classroom Discourse* 2(2). 219–235.
- Stefanowitsch, Anatol. 2000. The English *go*-(PRT)-and-*VERB* construction. In Lisa J. Conathan, Jeff Good, Darya Kavitskaya, Alyssa B. Wulf & Alan C. L. Yu (eds.), *Proceedings of the 26th annual meeting of the Berkeley Linguistic Society. General session and parasession on aspect*, 259–270. Berkeley, CA: Berkeley Linguistics Society.
- Tagliamonte, Sali A. 2004. Someth[in]'s go[ing] on!: Variable *ing* at ground zero. In Britt-Louise Gunnarsson, Lena Bergström, Gerd Eklund, Staffan Fidell, Lise H. Hansen, Angela Karstadt, Bengt Nordberg, Eva Sundergren & Mats Thelander (eds.), *Language variation in Europe: Papers from the second international conference on language variation in Europe, ICLAVE 2*. Upsala, Sweden: Department of Scandinavian Languages, Uppsala University.
- Tagliamonte, Sali A. 2006. *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.
- Tagliamonte, Sali A. 2012. *Variationist sociolinguistics: Change, observation, interpretation*. Malden, MA: Wiley-Blackwell.
- Tagliamonte, Sali A., Alexandra D'Arcy & Celeste Rodríguez-Louro. 2016. Outliers, impact, and rationalization in linguistic change. *Language* 92(4). 824–849.
- Trudgill, Peter. 1974a. Linguistic change and diffusion. *Language in Society* 3(2). 215–246.
- Trudgill, Peter. 1974b. *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.
- Wolfram, Walt. 1969. *A sociolinguistic description of Detroit Negro speech*. Washington, DC: Center for Applied Linguistics.
- Zhang, Cong, Kathleen Jepson, Georg Lohfink & Amalia Arvaniti. 2021. Comparing acoustic analyses of speech data collected remotely. *The Journal of the Acoustical Society of America* 149(6). 3910–3916.