Pre-publication version:

Childs, Claire. Forthcoming. Sociosyntax and the weight of the linguistic within sociolinguistics. In Daniel Duncan and Mary Robinson (eds) *English Sociosyntax*. Berlin: De Gruyter.

2. Sociosyntax and the weight of the linguistic within sociolinguistics

Claire Childs, University of York

1 Introduction

One of the most well-known contrasts made in debates about linguistic theory and methodology is the distinction between generative syntax and sociolinguistics. Generative syntax as a discipline was founded by Noam Chomsky in the 1950s and 1960s, with the goal of identifying and understanding the universal properties of language – a Universal Grammar – through the study of competence, i.e., innate linguistic knowledge (Chomsky 1966: 9). In the 1960s and 1970s, William Labov's work in sociolinguistics – focusing instead on "language in use within the speech community" (Labov 1972a: 183) – stood in stark contrast to the internally-oriented generativist framework. From a sociolinguistic perspective, language is intended for communication (Labov 2001: 3), so performance, i.e., language use, is the main object of study. The aim is to understand the "orderly heterogeneity" of language (Weinreich, Labov, and Herzog 1968), which typically involves quantifying linguistic data and examining the linguistic and social factors that contribute to the patterns of variation that are observed.

Given this contrast between generative syntax and sociolinguistics as they were first founded, it is no surprise that these disciplines are often depicted as representing two ends of a dichotomy – see, e.g., claims that "Labov and Chomsky hold completely opposite views" (Viereck 1986: 418). However, this opinion largely reflects how the two subfields were initially conceived, rather than the reality today. Although generative syntax and sociolinguistics clearly still differ in their theory of language and methodologies, the two subfields have shifted towards each other over the past few decades (see also Adger, Jamieson, and Smith Forthcoming), particularly with how they view intuition data from grammaticality/acceptability judgements versus production data in a corpus of recordings or texts.[1] While intuitions were initially seen as granting a window into linguistic competence, and production data granting an insight into performance in its social context, the boundaries are now much more blurred. Both kinds of data can be used to analyse both performance and competence. Generative syntax and sociolinguistics can be brought together better than ever through a sociosyntactic approach of the kind at the centre of this volume, to develop a more well-rounded understanding of grammatical variation and change in terms of the possible rules of the internal grammar and the social realisation of language in the community (Wilson and Henry 1998; Cornips and Corrigan 2005a, 2005b).

[1] I use the term *corpus* to refer to "a collection of texts assumed to be representative of a given language put together so that it can be used for linguistic analysis […] [where] there is a consensus that a corpus deals with natural, authentic language" (Tognini-Bonelli 2001: 2). I use the term *text* in a broad sense where it can refer to language that is spoken (recordings or transcripts), written or signed.

Within sociolinguistics, grammatical variables are often considered less likely to mark social differences compared to phonetic variables (Labov 1993; Hudson 1996; Cheshire 1999), but this may partly reflect the fact that the majority of sociolinguistic work has focused on phonetic variables. Often, the same kinds of grammatical variables are researched time and time again, perhaps because of an interest in particular kinds of alternation (e.g. standard versus nonstandard forms) or a need to choose features that are frequent enough in a corpus to successfully quantify (Cheshire 2005). As such, we know relatively little about the social meaning of grammatical forms (Moore 2021; Moore and Spencer 2021). We know more about the social factors that might be implicated in grammatical variation, through corpus-based work in a range of different communities, but it is important not to equate social meaning with social correlations in usage.

This chapter proceeds firstly with an overview of the main positions taken by generative syntax and sociolinguistics, including some of the criticisms that Chomsky and Labov have directed at each other's approach (Section 2). Section 3 focuses on acceptability judgements and corpora as sources of linguistic data and how they shed light on language as an internal and an external object. Section 4 looks at the relevance of social meaning and social factors for grammatical variation, and why linguistic factors may emerge as more significant for these variables. This is followed by the conclusion in Section 5.

2 The line drawn between generative syntax and sociolinguistics

The line often drawn between generative syntax and sociolinguistics has its origin in some of Chomsky and Labov's earliest works. Chomsky (1965: 4) argued that linguistics is not about language use: language *is* the internal grammar. He criticised disciplines that collect data without explanatory principles (for him, sociolinguistics included) with this famous allusion to "butterfly collecting": "You can also collect butterflies and make many observations. If you like butterflies, that's fine; but such work must not be confounded with research, which is concerned to discover explanatory principles of some depth and fails if it does not do so" (Chomsky 1979: 57). The same "butterfly collecting" metaphor was used ten years prior by DeCamp (1969), who called upon sociolinguists to move from collecting data – which he nevertheless flagged as important – to developing an overarching linguistic theory.[2] Although Chomsky (2000: 156) later acknowledges sociolinguistics as "a perfectly legitimate inquiry, externalist by definition", he notes that it "borrows from internalist inquiry into humans, but suggests no alternative to it, to my knowledge". Such remarks set out a dichotomy between sociolinguistics as more descriptive and generative syntax as more explanatory. Some sociolinguists have actually made similar criticisms. Cheshire (1987: 278) remarks upon "the lack of a coherent theory of language as it is used in its social context", while Chambers (2005: 216) muses:

> Most sociolinguists, no less than most theoreticians, go about their business as if they are engaged in self-contained, hermetically sealed research with no implications beyond the immediate results that, say, women in Amman tend to use glottal stops talking to other women but uvulars talking to men, or that, say, ergatives and transitives subcategorize periphrastic modals but unaccusatives do not.

[2] The "butterfly collecting" metaphor did not originate with DeCamp, as it has been used in other scientific fields, but he appears to be the first linguist to use it (Souag 2017).

In other words, collecting data and making observations about language use is clearly important, but this can run the risk of ignoring potential interactions between competence and performance that could contribute to an explanation of why particular patterns of language variation and change emerge. The "why" question is also at the centre of other criticisms that sociolinguistic analyses of grammatical variation can sometimes lack sufficient justification for why particular predictors are included in statistical models (Henry 2002) or that "the methodology does not require a detailed analysis of syntax" (Cheshire 2005: 85). As Tagliamonte and Rupp (2023: 17) argue in their account of the zero indefinite article, an approach that combines syntactic, sociolinguistic, dialectological and/or historical insights can help provide "the best possible means of coming to a fulsome explanation" of the variation. Potential shortcomings of sociolinguistic analyses of grammatical variation can be addressed through sociosyntactic investigations that take a more hypothesis-driven, theory-oriented approach, integrating more detailed syntactic analysis alongside quantitative evidence to test particular accounts of the variation.

On the sociolinguistic side of the debate, Labov has expressed resistance to the term *sociolinguistics* because "it implies that there can be a successful linguistic theory of practice which is not social" (Labov 1972a: xiii). This stance clearly stands in opposition to the generative perspective, which is not concerned with language as a social phenomenon. Labov's main criticisms of generative syntax are that it relies on introspective judgements of sentences as a basis for research. He argues that the variation in judgements between different researchers may make their conclusions unreliable (Labov 1972a, 1972b, 1996). He also expresses concern that acceptability judgements do not necessarily match what we observe in language use (Labov 1996). While this is most certainly true, it seems that it would be an irrelevant argument to generative linguists, since language use is just not within their remit.

Despite the disagreements between Chomsky and Labov in their approach to language, they have also acknowledged some common ground. In a 1978 interview later published in Chomsky (2004), Chomsky argues that his assumption of homogeneity in speakers' language capacities, speech communities, or how they acquire language, as required by generativist theory (see Section 3), does not preclude an interest in variation. Chomsky (1980: 224) also acknowledges that we may gain more understanding of our innate linguistic knowledge if we also consider pragmatic competence, i.e. "knowledge of conditions and manner of appropriate use, in conformity with various purposes". However, this point seems to have flown largely under the radar for both generativists and sociolinguists (Chambers 2005: 216). A further concession from Chomsky is that, in principle, evidence used to gain insight into the nature of the internal grammar could be derived from data other than acceptability judgements (including experimental work, the study of language change, and data from language acquisition) and that these kinds of data should be embraced if they prove to be as useful as acceptability judgements (Chomsky 1986: 37). Labov (1972a: 199) also recognises the role that intuition data can play in developing our knowledge of language variation but suggests that it needs to be interpreted in light of observations from natural language use.

Generative syntax as a subfield has also come to orient more towards variation than it did in the 1960s and 1970s. Principles and Parameters theory (Chomsky 1981), for example, put variation more at the centre of generativist enquiry. Parameters represent binary settings that could account for variation between languages, distinguishing, for example, pro-drop and non-pro-drop languages. Although Principles and Parameters theory was criticised by some (Newmeyer 2004; Boeckx 2014) and it fell out of favour among many syntacticians with the development of the Minimalist Program (Chomsky 1995), the two frameworks are not incompatible, as both are consistent with the idea that variation emerges from properties of functional heads (Roberts and Holmberg 2005; see also Section 4.1). Furthermore, Principles

and Parameters theory has continued to be used by some researchers, particularly in the area of micro-parametric comparative syntax where one compares closely related varieties of a language to better understand the properties of the language faculty and the parameters underlying linguistic variation (Kayne 2000, 2013).

The tenets of micro-parametric syntax – to understand the locus of variation within the grammar and what types of variation are possible/impossible (Baker 2008) – are also central to sociosyntax, the approach at the centre of this volume. However, sociosyntax is not as clearly associated (at least not yet) with one particular theory or perspective within generative linguistics. Furthermore, while the comparative method is central to micro-parametric syntax (Baker 2008), sociosyntax can involve either comparative cross-dialectal work or research within a single language variety. Cornips and Corrigan (2005: 7) note in their edited volume on sociosyntax that "[c]ommon ground for all the papers [in that volume] is that each attempts to achieve an adequate understanding of the mechanisms determining syntactic variation and change by combining insights from both paradigms" (referring to generative linguistics and sociolinguistics). Thus, there is flexibility in how insights/methods from each field are integrated, which will depend on the research questions. For example, research into grammatical variation across geographical space adopts a similar approach to micro-parametric syntax research in that it is inherently comparative. However, such dialectological research eschews the traditional generative method of introspection in favour of syntactic elicitation tasks or other kinds of acceptability judgement surveys, since much larger geographical coverage and scale is needed. These tasks are often coupled with some recordings of spontaneous conversation – a data source that generativists traditionally do not use – to understand the structure of syntactic variation in dialects of a language. See, for instance, work on The Syntactic Atlas of the Dutch Dialects (Cornips and Jongenburger 2001; Barbiers, Cornips, and Kunst 2007) and The Scots Syntax Atlas (Smith et al. 2019), which have both formal syntactic and sociolinguistic angles to their research. Projects like these embrace the intuition data that has long been the cornerstone of generativist methodology but adapt the data collection process to target particular speech communities and demographics, elicit judgements from a larger sample of participants, and compare intuitions and production. Doing so helps us to understand language as both an internal and an external object, as discussed in the next section.

3 Language as an internal and an external object

Linguistic competence clearly contributes to linguistic performance, but performance is also affected by a multitude of other factors that are, according to the generativist perspective, outside linguistic theory: "Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shift of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance" (Chomsky 1965: 3).

This famous quotation has been criticised outside the generativist enterprise for detaching language from its social context (Hymes 1971). However, Chomsky's argument is that to understand the fundamental principles of language, there is a need to abstract away from some of the complexity one might find in reality. He argues that abstractions of this kind are commonplace in the natural sciences and any potential oversights can be acknowledged and taken into account (Chomsky 1979: 57). Indeed, every field of research – in and outside of linguistics – involves some degree of abstraction. Sociolinguists, for instance, regularly adopt the apparent-time hypothesis to investigate language change in

synchronic data, which relies on an assumption that speakers' grammars do not fundamentally change after they have acquired their language (Bailey et al. 1991). Some research indicates that speakers' use of variants can vary across the lifespan. Although this has been reported most often for phonological variables (Sankoff and Blondeau 2007; MacKenzie 2017), the frequency of use of syntactic variants over an individual's lifespan can also change, and changes to their grammar, albeit rare, are possible (Fonteyn and Petré 2022). Regardless, the apparent-time hypothesis works as exactly that – a hypothesis – that is supported in most cases but can be re-evaluated when counter-evidence indicates that the hypothesis needs to be revised in a given scenario.

We need linguistic data that allows us some degree of insight into the internal grammar and/or linguistic performance, while understanding that all data is necessarily an abstraction, because of our inability to tap directly into speaker competence and the issues surrounding data collection such as sampling and representativeness. The generativist tradition relies predominantly on native speaker intuitions of sentences, which can inform us as to what is linguistically possible or impossible in someone's grammar. The sociolinguistic enterprise has traditionally been more focused on the analysis of language use – most often, in spontaneous speech. However, as discussed in the remainder of this section, judgements can tell us about performance as well as competence, and language use – performance – can lead to discoveries about competence.

3.1 Acceptability judgements

Chomsky (1965: 21) has long emphasised the importance of intuitions in building linguistic theories, but notes that speakers may not be aware of the extent of their linguistic knowledge: "Although there is no way to avoid the traditional assumption that the speaker-hearer's linguistic intuition is the ultimate standard that determines the accuracy of any proposed grammar, linguistic theory, or operational test, it must be emphasized, once again, that this tacit knowledge may very well not be immediately available to the user of the language". Chomsky notes that this seems to be a somewhat paradoxical position: speakers' intuitions are fundamental to developing a theory, but the speaker cannot necessarily access them. However, he explains that, given "an appropriately constructed context", a person will attach one particular interpretation to a given sentence, despite it having structural ambiguity (Chomsky 1965: 21). The person is therefore able to access their language intuitions with a bit of coaxing. For example, Chomsky discusses the structurally-ambiguous example *Flying planes can be dangerous*, which has two interpretations – either (i) planes that are flying can be dangerous, or (ii) the act of flying a plane can be dangerous. Someone will most likely interpret the sentence in one of the two ways and may not realise the ambiguity until it is pointed out to them, at which point they will likely acknowledge the other interpretation also (Chomsky 1965: 21).

Sentences are therefore abstract entities that, when meanings become attached to them in context, become tokens of sentences, essentially instantiated in performance (see Postal 2009) – which, incidentally, is what sociolinguists have traditionally been most interested in. From a generative perspective, sentences are generated by the grammar, and if a sentence violates rules of the grammar, then it is less grammatical; acceptability meanwhile relates to a sentence being "suitable, appropriate, adequate to the purpose at hand, etc." (Chomsky 1975: 8). A grammaticality judgement is therefore a direct reflection of competence whereas an acceptability judgement is affected by performance (Chomsky 1975: 7). However, the consensus is that speakers cannot access their grammars directly, so the notion of a grammaticality judgement is a misnomer (Bader and Häussler 2010; Sprouse, Schütze, and

Almeida 2013). In other words, speakers can only make acceptability judgements – which are "a characteristic of the stimulus as perceived by a speaker" (Bard, Robertson, and Sorace 1996: 33). Competence feeds into acceptability, but acceptability is also affected by numerous other factors (Chomsky 1965; Schütze 1996; Sprouse, Schütze, and Almeida 2013; Lau, Clark, and Lappin 2016). These include task-based factors such as the mode of presentation – written versus aural (Murphy 1997) – or the number of items and amount of repetition. In the case of syntactic satiation, for instance, the acceptability of a construction increases over time as a participant encounters more examples of it (Snyder 2000). Pragmatic and processing-related factors such as the frequency of the words in the sentence or the plausibility of the proposition expressed can also affect judgements (Cornips and Poletto 2005). For example, the sentence *There's three people coming to the party* was rejected by a speaker of Belfast English not because of the nonstandard singular agreement, but because they did not consider a party with three people to be much of a party (Henry 2005a: 1603).

Given this array of factors that may influence how people judge sentence acceptability, plus our inability to measure these directly (only indirectly), the exact mechanisms involved in making judgements remain unclear and open to debate. A usage-based perspective suggests that someone making a judgement may rely on stored knowledge of past experiences with similar sentences (Gerasimova and Lyutikova 2020: 135). It could also involve imagining a scenario in which the sentence might be said. In this vein, acceptability judgements are reflections on *perceived usage*. They are "a type of metalinguistic performance" (Francis 2022: 103). It is perhaps not so surprising, then, that they are subject to external influences, especially when it comes to the standard language ideology. A native speaker of English, for example, when presented with (1) and (2) in an acceptability judgement task, would be able to recognise that (1) is found in Standard English while (2) is not, even if the speaker themselves finds (2) acceptable and natural in their vernacular:

(1) *He saw her yesterday.*
(2) *He seen her yesterday.*

The extent to which a speaker is influenced by the nonstandard status of (2) when judging its acceptability will vary from person to person, but there is always a risk that someone will misreport a sentence as unacceptable because of its nonstandard status even though it is part of their vernacular grammar (Labov 1996; Henry 2005a; Cornips 2006). On the other hand, speakers might be accepting of sentences that are not possible in their own grammar, but they (erroneously) think that those sentences are used by others in the same community (Jamieson 2020).

The fact that individuals vary in their acceptability judgements has led some to question the validity of findings that are drawn from one researcher's intuitions, which is common practice in generativist research (Labov 1972c; Wasow and Arnold 2005). Despite acknowledgements from formal syntacticians that speakers of the same language do not have the same judgements (Kayne 2000: 8), information about how syntactic judgements are collected, and from whom, is often absent or lacking in generative syntax publications. That said, when acceptability judgements collected by informal methods and presented in syntax publications have been re-tested among a larger pool of participants, there is a high level of convergence in the judgements between the two (Sprouse and Almeida 2012; Sprouse, Schütze, and Almeida 2013). It is worth noting, however, that the replicability of judgements may depend on the construction involved and could vary for languages other than English (Linzen and Oseki 2018).

If individuals have different judgements, which may reflect different grammars, the term "language" or "dialect" in regular parlance must represent an average of the variation

found at an individual speaker level (DeCamp 1969; Lieberman 1984: 14). Within sociolinguistics, it is the overarching unit of the *speech community* in which structure emerges, which reflects speakers' "participation in a set of shared norms" (Labov 1972a: 121). Within the speech community, we find systematic variation according to style and social factors. Changes to the speech community, such as migration and language contact, can lead to linguistic change and the development of new varieties, as with the development of the koine Fiji Hindi (Siegel 2001: 176) and the emergence of Multicultural London English (Cheshire et al. 2011). Although acceptability judgements are essentially an individual-level phenomenon, it is assumed that speakers are broadly aware of the language of the speech community; their judgements appear to be influenced by their exposure to variation (Cornips 2006). In this vein, the notion of the speech community is an abstraction in a similar way that the notion of an ideal speaker-listener is, but Chomsky (1979: 54) argues that this is a sensible approach to take: "[y]ou study ideal systems, then afterwards you can ask yourself in what manner these ideal systems are represented and interact in real individuals".

The fact that acceptability judgements do not necessarily reflect patterns of language use, as discussed by Labov (1996), is not a fundamental problem to linguistic enquiry either, if we assume that perception and production are not equivalent. This seems to be an uncontroversial position. It is the cases where we find a contrast between intuitions and performance that are particularly interesting, and the fact that there is a contrast does not entail that we should disregard all intuition data (Newmeyer 1983: 66). By that logic, we could equally decide to dispense with spoken data, just because it does not tally with intuition data. Sometimes a contrast between production and perception data might arise simply because of the way the data was collected. For example, Labov (1996) reports how an informant had claimed to have never heard positive *anymore* (which means roughly "nowadays", e.g. *it's difficult anymore*) but she did use it in her speech. Her intuition therefore did not tally with her actual language use. However, as Labov and his team adapted their elicitation techniques as they collected their data, they subsequently gained more consistent answers from participants. Specifically, more consistent judgements were elicited when the positive *anymore* was used as part of a complaint – e.g. *farmers are pretty scarce around here anymore* (Labov 1996). Such examples highlight the importance of social context and pragmatics in designing elicitation tasks. Judgements from a sample of participants might include some ratings that relate to elements of the sentence that the researcher had not intended to be the feature of interest. This noise in the data might be averaged out overall, but giving participants an opportunity to comment on their judgements as part of the task can help to illuminate what is driving their ratings (Henry 2005a, 2005b; Jamieson 2020). I would argue that the reason why introspection of one's own judgements seems to work relatively successfully for generative syntacticians (as discussed earlier) is that the researcher can "ask" themselves why they have certain judgements. The syntactician can also "filter out" some of the factors that they know are irrelevant to their object of study and focus on what linguistic properties are affecting their judgements. Reflecting upon their judgements in this way leads the syntactician to evaluate and refine their analysis, and construct additional sentences to test further hypotheses, in an iterative process.

Despite criticisms that linguists should not rely on their own acceptability judgements, all linguists do this to some extent, even those using corpus-based methods (see Section 3.2). Felix (1987) points out that although Labov has criticised introspection, he nevertheless relies on judgements in his description of the Negative Attraction rule: Labov (1972a: 235) notes that the rule is "obligatory for all dialects, but sentences where it has not applied, such as **anybody doesn't sit there* are un-English in a very striking way". However, Labov has never denounced the use of introspection entirely, but has said that it can be useful if integrated

with insights from spontaneous speech (Labov 1972a: 199). A more general point is that preparing linguistic data for quantitative analysis necessarily requires intuitions from the author(s) for defining the variable and variants, delineating the contexts in which the variants can alternate with no change in meaning, deciding which factors may be relevant for inclusion in statistical models, and coding tokens according to these factors of interest. Some qualitative analysis is needed to describe grammatical properties and successfully quantify variation (Langacker 2016). Documenting these processes is essential to improve replicability and transparency; this may include the use of decision trees, explicit coding schema, and/or inter-rater reliability checks (see, e.g., Childs 2021; Eiswirth 2022; Wagner et al. 2015).

Acceptability judgement and other kinds of perceptual data, e.g. from experiments, is now a regular part of the sociolinguist's toolkit. It allows for systematic comparison of what is linguistically possible or not in different dialects, while also paying due attention to the individual and social differences that arise, and positing explanations for the variation that emerges (see, e.g., Barbiers 2005). Alternatively – or perhaps at the same time – one can also make use of corpus data to investigate what speakers actually do with language in interaction, as discussed in the next subsection.

3.2 Corpus data

Recordings of vernacular speech have long been at the centre of sociolinguistic research, because of the importance placed on the vernacular as the style "in which the minimum attention is paid to speech, [and which] provides the most systematic data for linguistic analysis" (Labov 1984: 29). Sociolinguists are also interested in language change, and while it is not possible to go back in time to ask people for their intuitions about language, we can analyse data in a corpus of recordings or texts from different timepoints to track linguistic change. Similar methods are used in historical linguistics, with written texts and longer time periods (see, e.g., Kroch 1989). The focus of sociolinguistics on shorter time periods likely reflects (i) an interest among sociolinguists in changes in progress rather than changes that are complete or near completion; (ii) the fact that audio recordings only started to be made from the late 19th century onwards and many of the oldest recordings have been lost, damaged or are not publicly available, so we must rely on newer recordings; (iii) an interest in social factors affecting linguistic change, which may be easier to investigate using contemporary corpora that have more detailed metadata.

Data availability is a key issue in corpus-based research more broadly. Certain language varieties are better represented in corpora than others – most obviously, English. This can have negative consequences, as corpus-based methods that are developed on languages such as English may not apply as readily to languages with different syntax, e.g., those with more morphological marking (Szmrecsanyi and Rosseel 2021), such as Arabic. The infrequency of many grammatical variables compared to phonological variables also poses a problem for any analysis (Henry 2005a; Labov 1972c: 190–191). We also run into the issue of negative evidence: no corpus can tell us where linguistic forms *cannot* be used. They can only provide positive evidence of where forms can be (and are) used. The absence of a form from a corpus could mean that it is not possible in the language variety the corpus represents, or it might exist in the variety but just happen not to occur in the dataset. Although some corpus linguists have argued that negative evidence can be obtained from a corpus (Stefanowitsch 2006), through a form of collostructural analysis that involves comparing the frequency of a form in and outside a construction, and the frequency of that construction with and without the form (Stefanowitsch and Gries 2003), others have criticised this approach. Some have argued that the underlying statistical assumptions are problematic and have proposed alternative calculations (Schmid and Küchenhoff 2013; Küchenhoff and Schmid 2015), while others

have argued that the approach does not pay enough attention to the semantics of the linguistic forms (Bybee 2010). Although Gries (2012, 2015) has defended the original method, it is clearly not without its controversies. Such methods have also not been adopted within variationist sociolinguistics, where the relevant unit of linguistic analysis is the linguistic variable. The negative evidence problem therefore provides a compelling argument as to why we sometimes still require intuition data as well as corpus data.

Just as corpora can be used for the study of language use and language change, they can be used to better understand aspects of competence, i.e., the linguistic structures underpinning variation. MacKenzie (2013), for example, uses Switchboard – a corpus of telephone conversations between American English speakers, recorded in the early 1990s – to study the phonological form of contracted auxiliary verbs. She shows how these forms fundamentally arise from two underlying structures (long and short allomorphs), but are subject to further phonological/phonetic processes. Her account can explain the distribution of intermediate contracted forms – tokens where full contraction is not possible due to phonotactic rules and a vowel is required (compare *it'[əl] be nice* to *they'[l] be nice*). Childs (2025) also uses corpus evidence – from the Spoken British National Corpus 2014 (Love et al. 2017) – to shed light on contraction as it interacts with negation and stative possession, demonstrating that Subject Type, a constraint that had previously been identified as relevant to variation between stative possessive *have* and *have got* (see Tagliamonte 2003), arises because of the different propensities for pronouns and Noun Phrases to trigger contraction.

Corpus data, if available, is therefore a useful tool for the analysis of linguistic variation and change, particularly for tracking linguistic change, testing different accounts of variation, or drawing connections between different linguistic variables. In this way, corpora can be used for the study of both competence and performance. Although the negative evidence problem is a limitation of corpus-based methods, intuition data can fill these gaps. Of these two methods, "neither appears to be wholly privileged as a window on a speaker's internal grammar" (Henry 2005a: 1616), but they can inform and complement each other.

4 The weight of linguistic factors in sociolinguistics

Generative syntax remains predominantly focused on understanding the properties that are shared by all languages, but it has been criticised for its tendency to focus on standard languages as opposed to nonstandard dialects (Hinskens, Auer, and Kerswill 2005; Adger and Trousdale 2007), which may result in relevant linguistic facts being missed because of biased sampling. Sociolinguistics meanwhile has an intrinsic focus on variation and regularly, though not exclusively, studies less codified varieties of a language. This orientation towards dialectal variation brings to the fore differences that may be explained by both linguistic and social factors, though the weighting of these may vary. This section focuses on this issue, with discussion of the social evaluation of linguistic forms (Section 4.1) and the social correlates of variants in use (Section 4.2), before considering how linguistic factors appear to be more significant and operate more consistently for grammatical variables across different communities compared to social factors (Section 4.3).

4.1 Social evaluation

Phonetic variants often carry social meaning for speakers (Eckert 1989, 2000; Campbell-Kibler 2010; D'Onofrio 2018), as do other aspects of the voice such as intonation and voice quality (Podesva 2007, 2011). Grammatical variation, on the other hand, is regularly characterised as less of a social marker than phonetic variation (Labov 1993;

Hudson 1996; Cheshire 1999). This contrast between phonetic and grammatical features is captured by The Interface Principle (Labov 1993) which refers to the fact that speakers evaluate surface-level forms. Phonetic variation is more "surface"-level in nature, whereas grammatical variation is "deep" (Labov 1993). However, the situation is complicated: some grammatical variables appear to be more surface-like and are socially evaluated, while others are less likely to receive attention or social comment. The social meaning of grammatical variants is vastly understudied, so there is much yet to be discovered in this domain (Moore and Spencer 2021).

Romaine (1981) proposed that there are four types of variables: phonological, morphophonemic, morphosyntactic or morpholexical, and "pure" syntactic. Phonological variables are the most widely studied, classic sociolinguistic variables, that involve variation between two or more sounds and are phonologically conditioned, such as post-vocalic /r/ or vowels such as FACE or GOAT. Morphophonemic variables are somewhat similar but are affected by both phonological and morphological/grammatical factors. For example, *t/d*-deletion is influenced by the phonological context but is also sensitive to morphological class (Romaine 1981; Baranowski and Turton 2020). Morphosyntactic or morpholexical variables can also be influenced by phonological and grammatical factors, but they involve the presence of particular lexical items that mark a syntactic relation of some kind (Romaine 1981). These comprise most of the grammatical variables regularly studied by sociolinguists. Romaine (1981) gives the examples of negative concord (Labov 1972d), negation in Montreal French (Sankoff and Vincent 1977)[3] and the deletion of complementiser *that* (Kroch and Small 1978). Finally, with "pure" syntactic variables, "a whole construction or arrangement of items which alternates is required" (Romaine 1981: 17). Romaine's (1981) example of a "pure" syntactic variable is the agentless passive (Weiner and Labov 1983), e.g. the distinction between *John saw Sandy* and *Sandy was seen by John.* Weiner and Labov (1983) found that no social factors were implicated in the agentless passive variation, and Romaine (1981) suggests that this is potentially characteristic of "pure" syntactic variables. Levon and Buchstaller (2015) warn about the potential circularity of such arguments, however. Is a linguistic item syntactic *because* it does not attract social evaluation, or is it syntactic regardless of that? Can more surface-level variants be less socially evaluated? There is no clear answer to these questions yet.

Trying to classify variables that are traditionally thought of as "grammatical" into one of Romaine's categories may be tricky because the boundaries between them are fuzzy (Cheshire 1987). There have been debates, for example, as to whether negative concord is fundamentally lexical as opposed to morphosyntactic/grammatical in nature, because it attracts such strong overt stigma (see Labov 1993; Meyerhoff and Walker 2013). Smith and Holmes-Elliott (2022) find that Buckie English speakers use negative concord at different rates depending on the lexical items involved in the construction, leading them to agree that the variation at least in part is lexical in nature. Negative concord is also sensitive to verb type (Childs 2017). That said, the fact that negative concord is not found in certain syntactic constructions where it is possible, namely existentials, suggests that there is a syntactic component as well (Smith and Holmes-Elliott 2022). Incidentally, the proposal that some grammatical variation has a lexical quality is entirely in-keeping with a popular school of thought within contemporary generative syntax in which variation is ultimately "attributable to differences in the features of particular items (e.g., the functional heads) in the lexicon" (Baker 2008: 353). According to this view – often called the Borer-Chomsky Conjecture (based on Borer 1984; Chomsky 1995) – variation emerges "in the mapping from the syntactic module to PF [phonetic form], not in the syntactic module itself" (Barbiers 2014:

[3] This is a French publication that was subsequently published in English as Sankoff and Vincent (1980).

199). The actual frequencies of variants in language use can then vary further depending on external factors (Adger 2006). Therefore, the sociolinguistic view of (at least some) grammatical variation is not so far removed from a generative view in this respect.

Addressing some of the "fuzziness" between Romaine's (1981) categories of variable, MacKenzie and Robinson (This Volume) make a different distinction between two types of grammatical variables: "variation in realization" variables and "variation in order" variables.[4] Under their account, variation in realization refers to phonological realizations of morphemes (e.g. verbal *-s* or past tense morphology) while variation in order refers to word order (e.g. the dative alternation). Through a systematic review of papers published in the journals *Language Variation and Change* and *Journal of Sociolinguistics*, they set out to test whether "variation in realization" and "variation in order" variables are different in terms of their likelihood to have social significance. They found that both types of variables could pattern according to social factors. However, social factors were less likely to be investigated in the first place for "variation in order" variables than "variation in realization" variables. In part, this might reflect a tendency for word order variation to be studied in historical texts, where detailed social information is often lacking (MacKenzie and Robinson This Volume). Speakers might be less aware of variation in word order than variation in realization – for instance, Corrigan and Robinson (This Volume) suggest that embedded inverted questions like *I asked what is it*, which involves word order variation, may be below the level of consciousness for native speakers of Mid-Ulster English. However, as MacKenzie and Robinson (This Volume) discuss, researchers have rarely studied perceptions of word order variation, so there is much more to find out in terms of people's awareness of these types of variables and any social meaning attached to the variants.

It is fair to say that grammatical features that involve an obvious deviation from the standard variety in *form* have been most likely to attract attention and stigma (as also reflected in the findings of MacKenzie and Robinson, This Volume). It is perhaps no accident that stigmatised forms, which have a high degree of salience and are subject to extensive social commentary, and will vary on both stylistic and social dimensions, are the same forms that sociolinguists reach for most often when carrying out grammatical analysis (Cheshire 2005). Nonstandard agreement and negative concord, for example, have "quite fixed social meanings associated with class and ethnicity" (Eckert 2019: 758–759). Even the most regular users of negative concord can shift more towards standard negation depending on the setting, which shows speakers' awareness of the construction and the negative attitudes it can garner (Labov 1972d; Smith and Holmes-Elliott 2022). Nonstandard verbal *-s* similarly elicits negative attitudes – users are perceived as less professional (Levon and Buchstaller 2015; Levon, Buchstaller, and Mearns 2020). Not only that, but it requires just one token of nonstandard verbal *-s* for such judgements to be made, whereas phonetic variants may require many more instances before the same judgements arise (Levon, Buchstaller, and Mearns 2020). The fact that there is a written standard that we can compare grammatical variants to makes identifying non-standard grammatical variants relatively easy, whereas this is more difficult for phonetic variation (Moore 2023: 212). Therefore, the suggestion that grammatical features do not carry social meaning as readily as phonetic variation is not always true (see also MacKenzie and Robinson This Volume). This will of course depend on the nature of the phonetic variants too. For instance, Levon, Buchstaller, and Mearns (2020) compared perceptions of nonstandard verbal *-s* to other variables such as (ing) and the FACE vowel, which were all produced by speakers from the North East of England. These phonetic

[4] MacKenzie and Robinson (This Volume) put aside some variables that do not readily fit into one of these categories (e.g. discourse-pragmatic variables).

features might not be as salient or obviously nonstandard to the listeners – who were from Scotland and the North East and North West of England – compared to nonstandard verbal *-s*.

Deviations from the standard in terms of *context* might, in some cases, not be as noticeable to speakers as deviations in form. While Buckie English speakers used negative concord less frequently in conversation with a community outsider compared to a community insider, no such shift occurred with their use of non-quantificational *never*, despite both forms of negation being nonstandard (Smith and Holmes-Elliott 2022). The fact that non-quantificational *never* involves the use of an otherwise Standard English item–*never*–but in a nonstandard syntactic/semantic context, might explain why speakers are not as aware of it and do not style shift with different interlocutors. The specific syntactic or discourse context might also make a difference, however. Non-quantificational *never* in a more obviously nonstandard syntactic environment, e.g. with an elided verb and for a particularly emphatic pragmatic purpose like an outright denial, might be particularly salient (see Childs 2021). For example, an accusation of *You broke the glass!* could be countered with a denial *No I never* – compare the pragmatic force of this instance of *never* with a more neutral case such as *I never organised the party*. At face value, the latter could be a Standard English sentence referring to a work party that used to happen every year but the speaker did not organise; alternatively, it could be a nonstandard sentence referring to one single party (e.g. last week) that the speaker is denying having organised. The context in which nonstandard forms are embedded therefore might modulate how well speakers accept a construction (in an acceptability judgement task or experiment) or how they style shift (if at all) in different conversational settings.

The social evaluation of a linguistic form is also not necessarily consistent between social groups, or from person to person, even within the same speech community. Women are often more likely to adhere to the standard language ideology and be less accepting of constructions that appear to deviate from those norms (Allen 1986; Labov et al. 2011). Listeners judging nonstandard features that are more local to them, and that they are more familiar with, may judge them more negatively than listeners who are less familiar with them (Levon and Buchstaller 2015); however, in other cases, people are more accepting of nonstandard features that are potential markers of local identity compared to more supralocal stigmatised forms (Jamieson et al. 2024). People's pragmatic language abilities might also correlate with how much they attend to the social meaning of variants. For example, people from the south of England who had higher scores on a Pragmatic Language scale[5] – which equated to lower pragmatic language ability – were (perhaps somewhat counter-intuitively) more sensitive to the frequency of the Northern Subject Rule (NSR, nonstandard verbal *-s*) in speech when judging a speaker's professionalism, whereas people from the same area with higher pragmatic language ability did not appear to attend to the frequency of the feature when making their judgements (Levon and Buchstaller 2015). Native and non-native speakers of a language also do not necessarily notice the same linguistic features or have the same social connotations for those features (Davydova and Hazen 2021).

Overall, although grammatical variation may be less susceptible to social evaluation than phonetic variation, nonstandard grammar is likely to attract particular attention for its deviation from the standard language, which may invoke stereotypes associated with low education levels or lower socioeconomic status. Features such as negative concord that are a vernacular universal of English (Chambers 2012) and are consistently the subject of negative attitudes may have more stable social meanings across separate communities, while less

[5] The Pragmatic Language scale "aims to tap into an individual's ability to engage in the social aspects of language, including their relative ease of effective communication and their ease/difficulty in holding fluent, reciprocal conversations" (Levon and Buchstaller 2015: 330). A higher score indicates a lower ability in this regard, while a lower score reflects a higher ability.

widespread grammatical variants may have more locally-nuanced social meanings. The social evaluation of forms likely depends on the level of linguistic structure in which they reside, but there is no set way of determining this; it will necessitate combining data from a variety of sources, such as acceptability judgements, experiments and style-shifting in language use.

4.2 Social patterning

The previous subsection focused on the social evaluation of grammatical forms, which needs to be differentiated from the social *patterning* of features. By social *patterning,* I am referring to correlations between the use of a particular variant with a particular social group. Although the social meaning of a form may stem from the use of that form by certain users, the reverse is not necessarily true – i.e., a social group A may use a variant more than social group B, but that does not mean that the variant indexes (socially) group A.

To give an example, Cheshire (1999: 72) suggests that nonstandard singular agreement with existential *there* might not be used "for social evaluation and the consequent marking of social groups". Walker (2020) studies this construction in the speech of different ethnic groups in Toronto – British/Irish, Chinese and Italian – and finds variation in how often each group uses it. For example, second/third generation speakers with an Italian background use standard plural agreement only 9% of the time, compared to 45% among older speakers of British/Irish descent. These and other patterns of variation according to ethnicity might reflect differences in where and how speakers acquired English (Walker 2020). What we do not know, however, is whether the speakers are *aware* of how often they use standard or nonstandard agreement, or whether they act as markers of those ethnic groups. Similarly, in their study of *was*/*were* variation in London, Cheshire and Fox (2009: 20) find that Black Caribbean speakers use nonstandard *was* the most, while Bangladeshi speakers use it the least. Ethnicity is a significant factor in their analysis, but the authors note that they treat it as "a broad indicator of language background, rather than necessarily being a salient variable in its own right" (Cheshire and Fox 2009: 20). In other words, the fact that a speaker is Bangladeshi is not in itself a direct explanation for their low frequency of nonstandard *was*, but may reflect some other linguistic or social fact. Cheshire and Fox (2009: 19) explain that the Bangladeshi ethnic group who migrated to London in the 1970s and 1980s were able to form very tight-knit networks with each other, with "very little contact with the white community". In this case, then, the Bangladeshi group's lower frequency of nonstandard *was*, which is used more often in London more broadly, may reflect aspects of their social network structure and the acquisition of variation rather than the form itself acting as an identity marker for speakers in other ethnic groups. Evidence from acceptability judgements, experiments or style-shifting in production (as discussed in Section 4.1) would be able to elucidate whether the variants do have social meaning for the social groups who are using the forms or (potentially) avoiding them.

Just as ethnicity is a social factor that can also reflect linguistic facts pertaining to language background, region is a social factor that can represent different underlying grammatical systems. Returning to *was*/*were* variation, even within a country as relatively small as England, there is extensive geographical variation for this variable which goes beyond variation in frequency. The system of constraints differs across the regions. Some areas of England, for instance, have a *was*-levelling system, e.g., the North East (Beal 1993) and parts of the Midlands (Anderwald 2002; Pietsch 2005). Other areas pattern towards *were*-levelling, e.g., Greater Manchester (Moore 2011) and parts of Lancashire, Yorkshire and Derbyshire (Pietsch 2005). These patterns suggest that person/number is implicated in *was*/*were* variation, or at least was historically, as these levelled systems developed. However, the picture becomes more complicated as certain areas also have a polarity-based

distinction where *was*-levelling is present in affirmative contexts but *weren't* appears in negative contexts, as found in areas including York (Tagliamonte 1998), East Anglia (Britain 2002) and the South East (Cheshire 1982; Cheshire and Fox 2009). To add yet more layers of linguistic complexity, some regions have the NSR, which permits singular agreement with Noun Phrase subjects but not pronouns, unless the pronoun subject is non-adjacent to the verb (Montgomery 1994). As the NSR originated in Northern England and Scotland (Cole 2014), it is perhaps most pertinent to those areas, but a similar Subject Type Constraint seems to operate even outside NSR areas (Schilling-Estes and Wolfram 1994; Cheshire and Fox 2009), and adjacency effects may be present in non-NSR varieties also (Pietsch 2005). Certain constructions like existential *there*, interrogatives and tag questions are also susceptible to non-agreement more generally (Cheshire and Fox 2009; Moore 2011; Tagliamonte 1998). We cannot, at least in this case, reduce regional distributions down to frequency differences alone, as this would mask differences in the internal rules of the dialects involved.

Age is perhaps the most important external predictor of all within sociolinguistics, as it grants insight into ongoing language change. One variant may increase in frequency over time at the expense of another, or we may see age-grading, where speakers shift away from nonstandard forms as they enter the workforce and develop their career, due to pressures from the linguistic marketplace (Sankoff and Laberge 1978), whereas later in life, they may revert to using nonstandard variants more often. Such changes in the frequency of variants across the lifespan are clearly sensitive to social identity and external pressures. Changes to grammars are meanwhile often most observable when comparing younger and older generations in a community where a traditional construction or system is gradually becoming lost. Younger speakers may shift away from a more traditional, local set of constraints and move towards a supralocal tendency (Durham 2013). These changes involve an internal component as well as an external component. The linguistic input that children hear around them has a role in shaping the development of their internal language, which then has consequences for their linguistic production; future generations of children are then exposed to this as input which similarly has consequences for their own internal language, and so on, resulting in language change over time (Lightfoot 1989, 2010).

To summarise, ethnicity, region and age are all social factors on the surface but they may also fundamentally correspond to linguistic facts. Ethnicity might represent facts about language background and language contact. Region might represent inherent structural differences in the dialects of the areas under study. Age effects might represent a change in the underlying grammar between older and younger generations. Other external factors may have similar linguistic ramifications, while others may be more wholly social and external in nature. Although we could treat social factors at face value as social correlates of language variation, which may be descriptively interesting, we can further explore the linguistic and social facts underpinning our choice of social predictors to try and establish whether we are dealing with more superficial frequency differences or deeper differences in the speakers' grammars.

4.3 The robustness of internal constraints in grammatical variation

Grammatical variables are typically primarily conditioned by internal constraints, in that the main constraint or constraints on the variation are linguistic and they are the most robust factors affecting the variation across different communities. Why is this the case? Firstly, if grammatical variables are less susceptible to social evaluation, as discussed in Section 4.1, this might be a major contributing factor, as researchers would either avoid looking at social factors because they expect that they do not exist, or they might look at

social factors but find no significant results (see discussion in MacKenzie and Robinson This Volume). Secondly, grammatical variables are notoriously difficult to define using the notion of the linguistic variable, as discussed at length in the sociolinguistic literature (Sankoff 1973; Lavandera 1978; Dines 1980; Romaine 1981; Cheshire 1987; Buchstaller 2009). While variants of a phonetic variable have no semantic meaning of their own and can be easily equated as alternating forms within a given set of linguistic contexts, as required by the definition of the linguistic variable as "alternative ways of 'saying the same thing'" (Labov 1972a: 94), grammatical variants necessarily contribute to the semantics of a phrase or sentence. Some have argued that any change in grammatical form necessarily results in a change in meaning (Bolinger 1977). Although this can largely be overcome by considering grammatical variants to mean the same thing if they are broadly functionally equivalent and refer to the same thing, grammatical variants do not exist in isolation and are inevitably affected by other sentence elements (as Childs 2025 discusses with regard to negation). As such, it is perhaps no surprise that internal factors come out strongest in analyses of the variation, because the variants themselves may have subtly different meanings or linguistic consequences for the rest of the sentence.

The greater significance of internal factors for grammatical variation has been demonstrated for variables including the agentless passive as discussed earlier (Weiner and Labov 1983), embedded inverted questions (Corrigan and Robinson This Volume), and the alternation between *not*-negation (e.g. *I didn't see anybody*) and *no*-negation (e.g. *I saw nobody*). The latter varies according to verb type across a range of British and Canadian Englishes: BE/HAVE favour *no*-negation and lexical verbs favour *not*-negation (Tottie 1991a; Tottie 1991b; Varela Pérez 2014; Childs 2017; Wallage 2017; Burnett, Koopman, and Tagliamonte 2018; Childs et al. 2020). Variation between *not*- and *no*-negation according to social factors is, in contrast, much less consistent across locations even just within Britain (Childs 2017). Furthermore, while sex/gender and age are significant in Toronto for Burnett, Koopman, and Tagliamonte (2018), they are not significant in the same community in Childs et al. (2020).

Particle verb variation between the joined variant (e.g. *phoned up the bank*) and the split variant (e.g. *phoned the bank up*) is also overwhelmingly conditioned by linguistic factors, including the length of the direct object (Kroch and Small 1978; Gries 2003; Lohse, Hawkins, and Wasow 2004; Szmrecsanyi 2005; Rodríguez-Puente 2016; Grafmiller and Szmrecsanyi 2018; Röthlisberger and Tagliamonte 2020; Lee and MacKenzie 2023), the syntactic weight or complexity of the object (Gries 2003; Cappelle 2009; Haddican and Johnson 2012; Rodríguez-Puente 2016; Haddican et al. 2020), and semantic transparency (Kroch and Small 1978; Szmrecsanyi 2005; Cappelle 2009; Grafmiller and Szmrecsanyi 2018; Lee and MacKenzie 2023). Other relevant linguistic factors include the frequency/idiomaticy of the phrase/meaning or its collocational strength, discourse status/accessibility or focus, concreteness or definiteness, presence of a post-modifying directional Prepositional Phrase/adverbial, rhythm or stress placement, and structural persistence or priming (Gries 2003; Szmrecsanyi 2005; Cappelle 2009; Rodríguez-Puente 2016; Grafmiller and Szmrecsanyi 2018; Lee and MacKenzie 2023). Some of these factors may be interrelated and may work together to aid the processing of information (Gries 2003; Lohse, Hawkins, and Wasow 2004) or to mark information structure (Dehé 2002).

External factors that may be relevant to particle verb variation are less widely studied than internal factors. The two that are most investigated are register/genre (Gries 2003; Cappelle 2009; Grafmiller and Szmrecsanyi 2018; Haddican et al. 2020; Lee and MacKenzie 2023) and language variety (Szmrecsanyi 2005; Haddican and Johnson 2012; Grafmiller and Szmrecsanyi 2018; Haddican et al. 2020). Such external factors can interact with more significant linguistic factors (Giles 2003) or apply less consistently across different datasets:

for example, Kroch and Small's (1978) suggestion that the particle verb alternation varies according to the social status of radio hosts versus callers was not borne out in Lee and MacKenzie (2023) where they tested this in more contemporary data alongside other factors.

As Röthlisberger and Tagliamonte (2020) note, though, the lack of social patterning for particle verb variation may in part emerge because of the unavailability of relevant social background information for the speakers/texts in the corpora that have been used in previous studies. As such, they set out to investigate the potential impact of social factors on particle verb variation in six communities in Ontario, Canada. In doing so, they find significant effects of birth year, community and occupation, but the linguistic factor of the length of the direct object is again most significant (Röthlisberger and Tagliamonte 2020). However, Röthlisberger and Tagliamonte (2020) included only two linguistic factors – length of direct object and idiomaticity of the verb semantics – alongside five social factors – birth year, gender, education, occupation and community. As many more linguistic factors have been found to be significant in prior research (as discussed above), a future model might instead consider as many of those same linguistic factors as feasibly possible before adding social factors to it, because some of the variation could potentially be explained by the inclusion of the missing linguistic factors. Szmrecsanyi (2005) takes this kind of approach to his analysis of persistence/priming effects for this variable, noting that he is considering these in addition to linguistic factors that have previously been identified as relevant, to see whether persistence/priming makes a difference on top of existing constraints. The inclusion of known factors alongside his new additions "minimises the likelihood that what appears to be a relevant factor is, in fact, a statistically spurious artefact of some other, not included factor" (Szmrecsanyi 2005: 120).

Although the discussion so far indicates that internal factors tend to be observed more consistently across different studies of the same variable, in cases where a feature has been transported from one geographically disparate language variety to another, e.g. via diffusion, internal factors may be less geographically consistent. For example, although quotative BE LIKE might be better described as a lexical feature than a grammatical feature (see Trudgill 2014), it has syntactic implications as the verb marks agreement with the subject (e.g. *I was like*; *they were like*). Quotative *be like* is thought to have its origins in American English, but it has since been transported to other varieties of English overseas. Despite the United States, England and New Zealand being located far from each other, the varieties spoken there have all adopted the same three major linguistic constraints on the use of *be like* with the same direction of effects: it is more common to express thought as opposed to speech, to express 1st person as opposed to 3rd person, and is more frequently used for mimetic reenactment (Buchstaller and D'Arcy 2009). The linguistic factor of tense/temporal reference is an exception that was not adopted in the same way in all three locations – it likely represents "high context" information that "is created in and through the local routinization of forms in the respective local variety" (Buchstaller and D'Arcy 2009: 322). BE LIKE meanwhile has vastly inconsistent social patterning between the three varieties for speaker sex and socioeconomic status; social information is therefore not as readily transported along with the linguistic form (Buchstaller and D'Arcy 2009). This is despite the fact that social factors are often particularly important in driving linguistic changes more generally, especially at the outset or the middle stages (Eckert 2000: 226).

In terms of measuring the weight of linguistic and social factors in sociolinguistic analyses, we are fortunate to be in an age of sociolinguistics where statistical tools and methods are becoming increasingly advanced. The availability of "big data" – large corpora or other databases – lends itself to techniques like mixed-effects regression modelling that can take into account multiple predictors – linguistic, social, stylistic, etc. – that may simultaneously affect linguistic variation and change. We can also critically evaluate the

results from older studies that used statistical methods that have since been superseded. Variationist sociolinguistics in the 1970s developed the notion of the "variable rule" which was central to the VARBRUL software package that could assess the relative impact of numerous factors contributing to variation (Cedergren and Sankoff 1974; Rousseau and Sankoff 1978; Sankoff and Labov 1979; Guy 1988). VARBRUL and its successor Goldvarb (Sankoff, Tagliamonte, and Smith 2005) were common packages used in sociolinguistic work, but fell out of favour – and out of use – around the late 2000s to early 2010s. Johnson (2009) demonstrated that the statistical assumptions underlying the models were not appropriate for linguistic data. Goldvarb assumes that each token of a variable is an independent observation, which is clearly not the case given that the same speaker can produce more than one token, leading to nesting effects which caused the software to overestimate the significance of social factors (Johnson 2009). Carrying out mixed-effects regression modelling, typically in the software R (R Core Team 2024), instead allows for the inclusion of random effects for factors such as "speaker" which cannot be easily replicated across studies but are important to take into account in modelling the variation, especially as speakers vary in how often they use each variant. Including random effects alongside fixed factors leads to more accurate results and can increase the researcher's confidence in the results for the other factors in the model (Johnson 2009; Gorman and Johnson 2013; Walker 2013). It is possible that social factors that were flagged as significant in models from Goldvarb might not be so relevant if the data were re-analysed with the contemporary statistical alternatives. Those older observations are nevertheless useful as they act as hypotheses that can be tested further using modern methods, if they have not already been replicated.

Overall, in any study of a grammatical variable, we expect linguistic factors to be significant. Social factors may be less important or, if they are relevant, less likely to be replicated across communities. The social correlates of a variant in one community do not necessarily emerge in another community, whether that be a case of recent linguistic diffusion or where the variable has been used in the communities for hundreds of years. Statistical models can include previously attested linguistic factors as a means of verifying their applicability to a new dataset and to ensure that the main constraints of the grammar are accounted for. Additional linguistic and social factors then add more layers to our understanding of the variation.

5 Conclusion

Generative syntax and sociolinguistics have often been characterised as opposite approaches within linguistics, with the former focusing on what is universal to language, innate, and linguistic, and the latter focusing on what varies within language, the external, and the social. Since the subfields were first developed in the 1960s and 1970s, however, they have gradually shifted closer towards one another. Intuition data that was once associated solely with generative approaches is now fully embraced within sociolinguistics. The use of such methods within sociolinguistics is not at all controversial. Extending introspection methods to larger scale acceptability judgement surveys in speech communities, analysed using quantitative methods, helps to overcome the criticisms that intuitions from one researcher may be unreliable. Although generative syntacticians have not necessarily embraced corpora to the same extent as sociolinguists have embraced intuition data, there is a greater synergy between formal and sociolinguistic methods as part of sociosyntactic research investigating the structures that underpin linguistic variation in its social context.

Acceptability judgements and production data both "add distortion of different kinds" to our understanding of linguistic competence (Gerasimova and Lyutikova 2020: 135–6). Both can be seen as essentially types of language performance, but both can be used to glean insights into competence, especially if we use a corpus to test hypotheses about linguistic structure. The combination of both intuition and usage data – either within or across studies – is the best of both worlds: "[d]ata from a variety of distinct sources and methods, properly interpreted, can be used to converge on right answers to hard questions" (Labov 1972c: 119).

As sociolinguistics has traditionally focused more on phonetic variation, we know relatively little about the social evaluation of grammatical features. However, we know a lot more about the social correlates of grammatical variation in language use: linguistic constraints appear to be the primary and most consistent constraints for grammatical variables, across different communities, whereas social patterns are much more diverse. Where social correlations do arise, delving deeper into possible reasons behind those correlations, through perceptual or judgement tasks, will help us ascertain which grammatical choices are more automatic and below the social radar, and which choices might be used with a greater sense of agency for identity marking or to achieve specific goals in interaction.

References

Adger, David. 2006. Combinatorial variability. *Journal of Linguistics* 42(3). 503–530. https://doi.org/10.1017/S002222670600418X.

Adger, David, E Jamieson & Jennifer Smith. Forthcoming. Sociolinguistics and Minimalist syntax. In Evelina Leivada & Kleanthes K. Grohmann (eds.), *The Cambridge Handbook of Minimalism and its Applications*. Cambridge: Cambridge University Press.

Adger, David & Graeme Trousdale. 2007. Variation in English syntax: Theoretical implications. *English Language and Linguistics* 11(2). 261–278. https://doi.org/10.1017/S1360674307002250.

Allen, Harold B. 1986. Sex-linked variation in the responses of dialect informants: Part III, grammar. *Journal of English Linguistics* 19(2). 149–176. https://doi.org/10.1177/007542428601900202.

Anderwald, Lieselotte. 2002. *Negation in Non-Standard British English: Gaps, Regularizations and Asymmetries*. London: Routledge.

Bader, Markus & Jana Häussler. 2010. Toward a model of grammaticality judgments. *Journal of Linguistics* 46(2). 273–330. https://doi.org/10.1017/S0022226709990260.

Bailey, Guy, Tom Wikle, Jan Tillery & Lori Sand. 1991. The apparent time construct. *Language Variation and Change* 3(3). 241–264. https://doi.org/10.1017/S0954394500000569.

Baker, Mark C. 2008. The macroparameter in a microparametric world. In Theresa Biberauer (ed.), *The Limits of Syntactic Variation*, 351–372. Amsterdam: John Benjamins.

Baranowski, Maciej & Danielle Turton. 2020. TD-deletion in British English: New evidence for the long-lost morphological effect. *Language Variation and Change* 32(1). 1–23. https://doi.org/10.1017/S0954394520000034.

Barbiers, Sjef, Leonie Cornips & Jan P. Kunst. 2007. The syntactic atlas of the Dutch dialects: A corpus of elicited speech and text as an online dynamic atlas. In Joan C. Beal, Karen P. Corrigan & Hermann L. Moisl (eds.), *Creating and Digitizing Language Corpora, Volume 1: Synchronic Databases*, 54–90. New York: Palgrave Macmillan.

Barbiers, Sjef. 2005. Word order variation in three-verb clusters and the division of labour between generative linguistics and sociolinguistics. In Leonie Cornips & Karen P. Corrigan (eds.), *Syntax and Variation: Reconciling the Biological and the Social*, 233–264. Amsterdam: John Benjamins.

Barbiers, Sjef. 2014. Syntactic doubling and deletion as a source of variation. In M. Carme Picallo (ed.), *Linguistic Variation in the Minimalist Framework*, 197–223. Oxford: Oxford University Press.

Bard, Ellen Gurman, Dan Robertson & Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language* 72(1). 32–68. https://doi.org/10.2307/416793.

Beal, Joan C. 1993. The grammar of Tyneside and Northumbrian English. In James Milroy & Lesley Milroy (eds.), *Real English: The Grammar of English Dialects in the British Isles*, 187–213. London: Longman.

Boeckx, Cedric. 2014. What Principles and Parameters got wrong. In M. Carme Picallo (ed.), *Linguistic Variation in the Minimalist Framework*, 155–178. Oxford: Oxford University Press.

Bolinger, Dwight. 1977. *Meaning and Form*. London: Longman.

Borer, Hagit. 1984. *Parametric Syntax: Case Studies in Semitic and Romance Languages*. Dordrecht: Foris Publications.

Britain, David. 2002. Diffusion, levelling, simplification and reallocation in past tense BE in the English Fens. *Journal of Sociolinguistics* 6(1). 16–43. https://doi.org/10.1111/1467-9481.00175.

Buchstaller, Isabelle. 2009. The quantitative analysis of morphosyntactic variation: Constructing and quantifying the denominator. *Language and Linguistics Compass* 3(4). 1010–1033. https://doi.org/10.1111/j.1749-818X.2009.00142.x.

Buchstaller, Isabelle & Alexandra D'Arcy. 2009. Localized globalization: A multi-local, multivariate investigation of quotative *be like*. *Journal of Sociolinguistics* 13(3). 291–331. https://doi.org/10.1111/j.1467-9841.2009.00412.x.

Burnett, Heather, Hilda Koopman & Sali A. Tagliamonte. 2018. Structural explanations in syntactic variation: The evolution of English negative and polarity indefinites. *Language Variation and Change* 30(1). 83–107. https://doi.org/10.1017/S0954394517000266.

Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Campbell-Kibler, Kathryn. 2010. The sociolinguistic variant as a carrier of social meaning. *Language Variation and Change* 22(3). 423–441. https://doi.org/10.1017/S0954394510000177.

Cappelle, Bert. 2009. Contextual clues for particle placement: Multiplicity, motivation, modeling. In Alexander Bergs & Gabriele Diewald (eds.), *Contexts and Constructions*, 145–192. Amsterdam: John Benjamins.

Cedergren, Henrietta J. & David Sankoff. 1974. Variable rules: Performance as a statistical reflection of competence. *Language* 50(2). 333–355. https://doi.org/10.2307/412441.

Chambers, J. K. 2005. Sociolinguistics and the language faculty. *The Canadian Journal of Linguistics* 50(1-4). 215–239. https://doi.org/10.1017/S0008413100003716.

Chambers, J. K. 2012. Global features of English vernaculars. In Raymond Hickey (ed.), *Areal Features of the Anglophone World*, 261–276. Berlin: De Gruyter Mouton.

Cheshire, Jenny. 1982. *Variation in an English Dialect: A Sociolinguistic Study*. Cambridge: Cambridge University Press.

Cheshire, Jenny. 1987. Syntactic variation, the linguistic variable, and sociolinguistic theory. *Linguistics* 25(2). 257–282. https://doi.org/10.1515/ling.1987.25.2.257.

Cheshire, Jenny. 1999. Taming the vernacular: Some repercussions for the study of syntactic variation and spoken grammar. *Cuadernos de Filologia Inglesa* 8. 59–80. https://revistas.um.es/cfi/article/view/65681.

Cheshire, Jenny. 2005. Syntactic variation and spoken language. In Leonie Cornips & Karen P. Corrigan (eds.), *Syntax and Variation: Reconciling the Biological and the Social*, 81–106. Amsterdam: John Benjamins.

Cheshire, Jenny & Sue Fox. 2009. *Was*/*were* variation: A perspective from London. *Language Variation and Change* 21(1). 1–38. https://doi.org/10.1017/S0954394509000015.

Cheshire, Jenny, Paul Kerswill, Sue Fox & Eivind N. Torgersen. 2011. Contact, the feature pool and the community: The emergence of Multicultural London English. *Journal of Sociolinguistics* 15(2). 151–196. https://doi.org/10.1111/j.1467-9841.2011.00478.x.

Childs, Claire. 2017. Integrating syntactic theory and variationist analysis: The structure of negative indefinites in regional dialects of British English. *Glossa: A Journal of General Linguistics* 2(1). Article 106. https://doi.org/10.5334/gjgl.287.

Childs, Claire. 2021. The grammaticalisation of *never* in British English dialects: Quantifying syntactic and functional change. *Journal of Linguistics* 57(3). 531–568. https://doi.org/10.1017/S0022226720000092.

Childs, Claire. 2025. What counts as "saying the same thing"? Variables and variants in English negation. In Frances Blanchette & Cynthia Lukyanenko (eds.), *Perspectives on Negation: Views from Across the Language Sciences*. 253–288. Berlin: De Gruyter Mouton.

Childs, Claire, Chris Harvey, Karen P. Corrigan & Sali A. Tagliamonte. 2020. Transatlantic perspectives on variation in negative expressions. *English Language and Linguistics* 24(1). 23–47. https://doi.org/10.1017/S1360674318000199.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1966. *Topics in the Theory of Generative Grammar*. Berlin: De Gruyter Mouton.

Chomsky, Noam. 1975. *The Logical Structure of Linguistic Theory*. New York: Plenum Press.

Chomsky, Noam. 1979. *Language and Responsibility: Based on Conversations with Mitsou Ronat*. Sussex: The Harvester Press.

Chomsky, Noam. 1980. *Rules and Representations*. New York: Columbia University Press.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris Publications.

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. Westport, CT: Praeger.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Chomsky, Noam. 2004. Language and the human mind (29 September 1978). In Carlos-Peregrín Otero (ed.). *Noam Chomsky: Language and Politics*, 221–242. 2nd edn. Edinburgh: AK Press.

Cole, Marcelle. 2014. *Old Northumbrian Verbal Morphosyntax and the (Northern) Subject Rule*. Amsterdam: John Benjamins.

Cornips, Leonie. 2006. Intermediate syntactic variants in a dialect-standard speech repertoire and relative acceptability. In Gisbert Fanselow, Caroline Féry, Ralf Vogel & Matthias Schlesewsky (eds.), *Gradience in Grammar: Generative Perspectives*, 85–105. Oxford: Oxford University Press.

Cornips, Leonie & Karen P. Corrigan. 2005a. Convergence and divergence in grammar. In Peter Auer, Frans Hinskens & Paul Kerswill (eds.), *Dialect Change: Convergence and Divergence in European Languages*, 96–134. Cambridge: Cambridge University Press.

Cornips, Leonie & Karen P. Corrigan. 2005b. Toward an integrated approach to syntactic variation: A retrospective and prospective synopsis. In Leonie Cornips & Karen P. Corrigan (eds.), *Syntax and Variation: Reconciling the Biological and the Social*, 1–27. Amsterdam: John Benjamins.

Cornips, Leonie & Willy Jongenburger. 2001. Elicitation techniques in a Dutch syntactic dialect atlas project. *Linguistics in the Netherlands* 18(1). 53–63. https://doi.org/10.1075/avt.18.08cor.

Cornips, Leonie & Cecilia Poletto. 2005. On standardising syntactic elicitation techniques (part 1). *Lingua* 115(7). 939–957. https://doi.org/10.1016/j.lingua.2003.11.004.

Corrigan, Karen P. & Mary Robinson. This Volume. Polish newcomers acquiring questions and questioning in a local dialect.

Davydova, Julia & Kirk Hazen. 2021. The role of linguistic structure in the perceptions of vernacular speech: Evidence from L1 English and English as a foreign language. *English World-Wide* 42(3). 273–298. https://doi.org/10.1075/eww.00073.dav.

DeCamp, David. 1969. Is a sociolinguistic theory possible? In James A. Alatis (ed.), *Report of the 20th Annual Round Table: Linguistics and the Teaching of Standard English to Speakers of Other Languages or Dialects*, 157–173. Washington, DC: Georgetown University Press.

Dehé, Nicole. 2002. *Particle Verbs in English: Syntax, Information Structure and Intonation*. Amsterdam: John Benjamins.

Dines, Elizabeth R. 1980. Variation in discourse – "and stuff like that". *Language in Society* 9(1). 13–31. https://doi.org/10.1017/S0047404500007764.

D'Onofrio, Annette. 2018. Personae and phonetic detail in sociolinguistic signs. *Language in Society* 47(4). 513–539. https://doi.org/10.1017/S0047404518000581.

Durham, Mercedes. 2013. *Was/were* alternation in Shetland English. *World Englishes* 32(1). 108–128. https://doi.org/10.1111/weng.12009.

Eckert, Penelope. 1989. *Jocks and Burnouts: Social Categories and Identity in the High School*. New York: Teachers College Press.

Eckert, Penelope. 2000. *Linguistic Variation as Social Practice*. Oxford: Blackwell.

Eckert, Penelope. 2019. The limits of meaning: Social indexicality, variation, and the cline of interiority. *Language* 95(4). 751–776. https://doi.org/10.1353/lan.2019.0072.

Eiswirth, Mirjam E. 2022. Developing and testing interaction-based coding schemes for the analysis of sociolinguistic variation. *Language & Communication* 87. 11–28. https://doi.org/10.1016/j.langcom.2022.05.001.

Felix, Sascha W. 1987. *Cognition and Language Growth*. Dordrecht: Foris Publications.

Fonteyn, Lauren & Peter Petré. 2022. On the probability and direction of morphosyntactic lifespan change. *Language Variation and Change* 34(1). 79–105. https://doi.org/10.1017/S0954394522000011.

Francis, Elaine J. 2022. *Gradient Acceptability and Linguistic Theory*. Oxford: Oxford University Press.

Gerasimova, Anastasia & Ekaterina Lyutikova. 2020. Intralingual variation in acceptability judgments and production: Three case studies in Russian grammar. *Frontiers in Psychology* 11. 120–138. https://doi.org/10.3389/fpsyg.2020.00348.

Gorman, Kyle & Daniel Ezra Johnson. 2013. Quantitative analysis. In Robert Bayley, Richard Cameron & Ceil Lucas (eds.), *The Oxford Handbook of Sociolinguistics*, 214–240. Oxford: Oxford University Press.

Grafmiller, Jason & Benedikt Szmrecsanyi. 2018. Mapping out particle placement in varieties of English: A study in comparative sociolinguistic analysis. *Language Variation and Change* 30(3). 385–412. https://doi.org/10.1017/S0954394518000170.

Gries, Stefan Th. 2003. *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement*. New York: Continuum.

Gries, Stefan Th. 2012. Frequencies, probabilities, and association measures in usage-/exemplar-based linguistics: Some necessary clarifications. *Studies in Language* 36(3). 477–510. https://doi.org/10.1075/sl.36.3.02gri.

Gries, Stefan Th. 2015. More (old and new) misunderstandings of collostructional analysis: On Schmid and Küchenhoff (2013). *Cognitive Linguistics* 26(3). 505–536. https://doi.org/10.1515/cog-2014-0092.

Guy, Gregory R. 1988. Advanced Varbrul analysis. In Kathleen Ferrara, Becky Brown, Keith Walters & John Baugh (eds.), *Linguistic Change and Contact*, 124–136. Austin, TX: Department of Linguistics, University of Texas.

Haddican, Bill & Daniel Ezra Johnson. 2012. Effects on the particle verb alternation across English dialects. *University of Pennsylvania Working Papers in Linguistics* 18(2). 31–40. https://repository.upenn.edu/handle/20.500.14332/44878.

Haddican, Bill, Daniel Ezra Johnson, Joel Wallenberg & Anders Holmberg. 2020. Variation and change in the particle verb alternation across English dialects. In Karen V. Beaman, Isabelle Buchstaller, Sue Fox & James A. Walker (eds.), *Advancing Socio-*

Grammatical Variation and Change: In Honour of Jenny Cheshire, 205–228. London: Routledge.

Henry, Alison. 2002. Variation and syntactic theory. In J. K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, 267–282. Oxford: Blackwell.

Henry, Alison. 2005a. Non-standard dialects and linguistic data. *Lingua* 115(11). 1599–1617. https://doi.org/10.1016/j.lingua.2004.07.006.

Henry, Alison. 2005b. Idiolectal variation and syntactic theory. In Leonie Cornips & Karen P. Corrigan (eds.), *Syntax and Variation: Reconciling the Biological and the Social*, 109–122. Amsterdam: John Benjamins.

Hinskens, Frans, Peter Auer & Paul Kerswill. 2005. The study of dialect convergence and divergence: Conceptual and methodological considerations. In Peter Auer, Frans Hinskens & Paul Kerswill (eds.), *Dialect Change: Convergence and Divergence in European Languages*, 1–48. Cambridge: Cambridge University Press.

Hudson, Richard A. 1996. *Sociolinguistics*. 2nd edn. Cambridge: Cambridge University Press.

Hymes, Dell. 1971. Competence and performance in linguistic theory. In Renira Huxley & Elisabeth Ingram (eds.), *Language Acquisition: Models and Methods*. New York: Academic Press.

Jamieson, E. 2020. Viewing dialect change through acceptability judgments: A case study in Shetland dialect. *Glossa: A Journal of General Linguistics* 5(1). Article 19. https://doi.org/10.5334/gjgl.979.

Jamieson, E., Jennifer Smith, David Adger, Caroline Heycock & Gary Thoms. 2024 (First View). 'When intuitions (don't) fail': Combining syntax and sociolinguistics in the analysis of Scots. *English Language and Linguistics*. https://doi.org/10.1017/S1360674323000679

Johnson, Daniel Ezra. 2009. Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3(1). 359–383. https://doi.org/10.1111/j.1749-818X.2008.00108.x.

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

Kayne, Richard S. 2013. Comparative syntax. *Lingua* 130. 132–151. https://doi.org/10.1016/j.lingua.2012.10.008.

Kroch, Anthony S. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1(3). 199–244. https://doi.org/10.1017/S0954394500000168.

Kroch, Anthony & Cathy Small. 1978. Grammatical ideology and its effect on speech. In David Sankoff (ed.), *Linguistic Variation: Models and Methods*, 45–55. New York: Academic Press.

Küchenhoff, Helmut & Hans-Jörg Schmid. 2015. Reply to "More (old and new) misunderstandings of collostructional analysis: On Schmid & Küchenhoff" by Stefan Th. Gries. *Cognitive Linguistics* 26(3). 537–547. https://doi.org/10.1515/cog-2015-0053.

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

Labov, William. 1972b. *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphia: University of Philadelphia Press.

Labov, William. 1972c. Some principles of linguistic methodology. *Language in Society* 1(1). 97–120. https://doi.org/10.1017/S0047404500006576.

Labov, William. 1972d. Negative attraction and negative concord in English grammar. *Language* 48(4). 773–818. https://doi.org/10.2307/411989.

Labov, William. 1993. The unobservability of structure and its linguistic consequences. Paper presented at New Ways of Analyzing Variation in English (NWAVE) XXII, Ottawa, 16 October, 1993.

Labov, William. 1996. When intuitions fail. In L. McNair, K. Singer, L. Dolbrin & M. Aucon (eds.), *Papers from the Parasession on Theory and Data in Linguistics, Chicago Linguistic Society 32*, 77–106.

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

Labov, William, Sharon Ash, Maya Ravindranath, Tracey Weldon, Maciej Baranowski & Naomi Nagy. 2011. Properties of the sociolinguistic monitor. *Journal of Sociolinguistics* 15(4). 431–463. https://doi.org/10.1111/j.1467-9841.2011.00504.x.

Langacker, Ronald W. 2016. Working toward a synthesis. *Cognitive Linguistics* 27(4). 465–477. https://doi.org/10.1515/cog-2016-0004.

Lau, Jey Han, Alexander Clark & Shalom Lappin. 2016. Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive Linguistics* 41(5). 1202–1241. https://doi.org/10.1111/cogs.12414.

Lavandera, Beatriz R. 1978. Where does the sociolinguistic variable stop? *Language in Society* 7(2). 171–182. https://doi.org/10.1017/S0047404500005510.

Lee, Naomi & Laurel MacKenzie. 2023. Social role effects on English particle verb variation fail to replicate. *Canadian Journal of Linguistics* 68(2).329–343. https://doi.org/10.1017/cnj.2023.13.

Levon, Erez & Isabelle Buchstaller. 2015. Perception, cognition, and linguistic structure: The effect of linguistic modularity and cognitive style on sociolinguistic processing. *Language Variation and Change* 27(3). 319–348. https://doi.org/10.1017/S0954394515000149.

Levon, Erez, Isabelle Buchstaller & Adam Mearns. 2020. Towards an integrated model of perception: Linguistic architecture and the dynamics of sociolinguistic cognition. In Karen V. Beaman, Isabelle Buchstaller, Sue Fox & James A. Walker (eds.), *Advancing Socio-Grammatical Variation and Change: In Honour of Jenny Cheshire*, 32–54. London: Routledge.

Lieberman, Philip. 1984. *The Biology and Evolution of Language*. Cambridge, MA: Harvard University Press.

Lightfoot, David. 1989. The child's trigger experience: Degree-0 learnability. *Behavioral and Brain Sciences* 12(2). 321–334. https://doi.org/10.1017/S0140525X00048883.

Lightfoot, David. 2010. Language acquisition and language change. *Wiley Interdisciplinary Reviews: Cognitive Science* 1(5). 677–684. https://doi.org/10.1002/wcs.39.

Linzen, Tal & Yohei Oseki. 2018. The reliability of acceptability judgments across languages. *Glossa: A Journal of General Linguistics* 3(1). Article 100. https://doi.org/10.5334/gjgl.528.

Lohse, Barbara, John A. Hawkins & Thomas Wasow. 2004. Domain minimization in English verb-particle constructions. *Language* 80(2). 238–261. https://doi.org/10.1353/lan.2004.0089.

Love, Robbie, Claire Dembrie, Andrew Hardie, Vaclav Brezina & Tony McEnery. 2017. The Spoken BNC2014: Designing and building a corpus of everyday conversations. *International Journal of Corpus Linguistics* 22(3). 319–344. https://doi.org/10.1075/ijcl.22.3.02lov

MacKenzie, Laurel. 2013. Variation in English auxiliary realization: A new take on contraction. *Language Variation and Change* 25(1). 17–41. https://doi.org/10.1017/S0954394512000257.

MacKenzie, Laurel. 2017. Frequency effects over the lifespan: A case study of Attenborough's r's. *Linguistics Vanguard* 3(1). Article 20170005. https://doi.org/10.1515/lingvan-2017-0005.

MacKenzie, Laurel & Mary Robinson. This Volume. Spelling out syntactic variation.

Meyerhoff, Miriam & James A. Walker. 2013. An existential problem: The sociolinguistic monitor and variation in existential constructions on Bequia (St. Vincent and the Grenadines). *Language in Society* 42(4). 407–428. https://doi.org/10.1017/S0047404513000456.

Montgomery, Michael. 1994. The evolution of verb concord in Scots. In Alexander Fenton & Donald A. McDonald (eds.), *Studies in Scots and Gaelic: Proceedings of the Third International Conferences on the Languages of Scotland*, 81–95. Edinburgh: Canongate.

Moore, Emma. 2011. Interaction between social category and social practice: Explaining *was/were* variation. *Language Variation and Change* 22(3). 347–371. https://doi.org/10.1017/S095439451000013X.

Moore, Emma. 2021. The social meaning of syntax. In Lauren Hall-Lew, Emma Moore & Robert J. Podesva (eds.), *Social Meaning and Linguistic Variation: Theorizing the Third Wave*, 54–79. Cambridge: Cambridge University Press.

Moore, Emma. 2023. *Socio-Syntax: Exploring the Social Life of Grammar*. Cambridge: Cambridge University Press.

Moore, Emma & Sarah Spencer. 2021. "It just sounds proper common": Exploring the social meanings expressed by nonstandard grammar. *Linguistics and Education* 63. Article 100933. https://doi.org/10.1016/j.linged.2021.100933.

Murphy, Victoria A. 1997. The effect of modality on a grammaticality judgement task. *Second Language Research* 13(1). 34–65. https://doi.org/10.1191/026765897671676818.

Newmeyer, Frederick J. 1983. *Grammatical Theory: Its Limits and Its Possibilities*. Chicago: University of Chicago Press.

Newmeyer, Frederick J. 2004. Against a parameter-setting approach to typological variation. *Linguistic Variation Yearbook* 4. 181–234. https://doi.org/10.1075/livy.4.06new.

Pietsch, Lukas. 2005. *Variable Grammars: Verbal Agreement in Northern Dialects of English*. Tübingen: Max Niemeyer Verlag.

Podesva, Robert J. 2007. Phonation type as a stylistic variable: The use of falsetto in constructing persona. *Journal of Sociolinguistics* 11(4). 478–504. https://doi.org/10.1111/j.1467-9841.2007.00334.x.

Podesva, Robert J. 2011. The salience and social meaning of declarative contours: Three case studies of gay professionals. *Journal of English Linguistics* 39(3). 233–264. https://doi.org/10.1177/0075424211405161.

Postal, Paul M. 2009. The incoherence of Chomsky's "biolinguistic" ontology. *Biolinguistics* 3(1). 104–123. https://doi.org/10.5964/bioling.8689.

R Core Team. 2024. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Roberts, Ian & Anders Holmberg. 2005. On the role of parameters in Universal Grammar: A reply to Newmeyer. In Hans Broekhuis, Norbert Corver, Riny Huijbregts, Ursula Kleinhenz & Jan Koster (eds.), *Organizing Grammar: Linguistic Studies in Honor of Henk Van Riemsdijk*, 538–553. Berlin: De Gruyter Mouton.

Rodríguez-Puente, Paula. 2016. Particle placement in Late Modern English and twentieth-century English: Morpho-syntactic variables. *Folia Linguistica Historica* 37(1). 145–175. https://doi.org/10.1515/flih-2016-0005.

Romaine, Suzanne. 1981. On the problem of syntactic variation: A reply to Beatriz Lavandera and William Labov. *Sociolinguistic Working Paper 82*. 3–40. Austin, TX: Southwest Educational Development Laboratory.

Röthlisberger, Melanie & Sali A. Tagliamonte. 2020. The social embedding of a syntactic alternation: Variable particle placement in Ontario English. *Language Variation and Change* 32(3). 317–348. https://doi.org/10.1017/S0954394520000174.

Rousseau, Pascale & David Sankoff. 1978. Advances in variable rule methodology. In David Sankoff (ed.), *Linguistic Variation: Models and Methods*, 55–69. New York: Academic Press.

Sankoff, David & Suzanne Laberge. 1978. The linguistic market and the statistical explanation of variability. In David Sankoff (ed.), *Linguistic Variation: Models and Methods*, 239–250. New York: Academic Press.

Sankoff, David & William Labov. 1979. On the uses of variable rules. *Language in Society* 8(2-3). 189–222. https://doi.org/10.1017/S0047404500007430.

Sankoff, David, Sali A. Tagliamonte & Eric Smith. 2005. Goldvarb X: A Variable Rule Application for Macintosh and Windows. University of Toronto and University of Ottawa.

Sankoff, Gillian. 1973. Above and beyond phonology in variable rules. In Charles-James N. Bailey & Roger W. Shuy (eds.), *New Ways of Analyzing Variation in English*, 44–61. Washington, DC: Georgetown University Press.

Sankoff, Gillian & Hélène Blondeau. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83(3). 560–588. https://doi.org/10.1353/lan.2007.0106.

Sankoff, Gillian & Diane Vincent. 1977. L'emploi productif du *ne* dans le français parlé à Montréal. *Le français moderne* 45(3). 243–256.

Sankoff, Gillian & Diane Vincent. 1980. The productive use of *ne* in spoken Montreal French. In Gillian Sankoff (ed.), *The Social Life of Language*, 295–310. Philadelphia: University of Philadelphia Press.

Schilling-Estes, Natalie & Walt Wolfram. 1994. Convergent explanation and alternative regularization patterns: *Were/weren't* levelling in a vernacular English variety. *Language Variation and Change* 6(3). 273–302. https://doi.org/10.1017/S0954394500001691.

Schmid, Hans-Jörg & Helmut Küchenhoff. 2013. Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics* 24(3). 531–577. https://doi.org/10.1515/cog-2013-0018.

Schütze, Carson T. 1996. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: University of Chicago Press.

Siegel, Jeff. 2001. Koine formation and creole genesis. In Norval Smith & Tonjes Veenstra (eds.), *Creolization and Contact*, 175–197. Amsterdam: John Benjamins.

Smith, Jennifer, David Adger, Brian Aitken, Caroline Heycock, E Jamieson & Gary Thoms. 2019. The Scots Syntax Atlas. https://scotssyntaxatlas.ac.uk (accessed 5 August 2024).

Smith, Jennifer & Sophie Holmes-Elliott. 2022. Mapping syntax and the sociolinguistic monitor. In Tanya Karoli Christensen & Torben Juel Jensen (eds.), *Explanations in Sociosyntactic Variation*, 58–89. Cambridge: Cambridge University Press.

Snyder, William. 2000. An experimental investigation of syntactic satiation effects. *Linguistic Inquiry* 31(3). 575–582. https://doi.org/10.1162/002438900554479.

Souag, Lameen. 2017. October 29. "Butterfly-collecting: The History of an Insult." *Jabal al-Lughat: Climbing the Mountain of Languages*. http://lughat.blogspot.com/2017/10/ (accessed 5 August 2024).

Sprouse, Jon & Diogo Almeida. 2012. Assessing the reliability of textbook data in syntax: Adger's *Core Syntax*. *Journal of Linguistics* 48(3). 609–652. https://doi.org/10.1017/S0022226712000011.

Sprouse, Jon, Carson T. Schütze & Diogo Almeida. 2013. A comparison of informal and formal acceptability judgements using a random sample from *Linguistic Inquiry* 2001–2010. *Lingua* 134. 219–248. https://doi.org/10.1016/j.lingua.2013.07.002.

Stefanowitsch, Anatol. 2006. Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistic Theory* 2(1). 61–77. https://doi.org/10.1515/CLLT.2006.003.

Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243. https://doi.org/10.1075/ijcl.8.2.03ste.

Szmrecsanyi, Benedikt. 2005. Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory* 1(1). 113–150. https://doi.org/10.1515/cllt.2005.1.1.113.

Szmrecsanyi, Benedikt & Laura Rosseel. 2021. English corpus linguistics. In Bas Aarts, April McMahon & Lars Hinrichs (eds.), *The Handbook of English Linguistics*, 29–44. 2nd edn. Hoboken, NJ: Wiley Blackwell.

Tagliamonte, Sali A. 1998. *Was/were* variation across the generations: View from the city of York. *Language Variation and Change* 10(2). 153–191. https://doi.org/10.1017/S0954394500001277.

Tagliamonte, Sali A. 2003. "Every place has a different toll": Determinants of grammatical variation in cross-variety perspective. In Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of Grammatical Variation in English*, 531–554. Berlin: De Gruyter Mouton.

Tagliamonte, Sali A. 2012. *Variationist Sociolinguistics: Change, Observation, Interpretation*. Hoboken, NJ: Wiley-Blackwell.

Tagliamonte, Sali A. & Laura Rupp. 2023. Legacies of syntactic change in a conservative dialect, York, England. *Journal of Historical Syntax* 7(6-19). 1–22. https://doi.org/10.18148/hs/2023.v7i6-19.161.

Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Tottie, Gunnel. 1991a. *Negation in English Speech and Writing: A Study in Variation*. London: Academic Press.

Tottie, Gunnel. 1991b. Lexical diffusion in syntactic change: Frequency as a determinant of linguistic conservatism in the development of negation in English. In Dieter Kastovsky (ed.), *Historical English Syntax*, 439–468. Berlin: De Gruyter Mouton.

Trudgill, Peter. 2014. Diffusion, drift and the irrelevance of media influence. *Journal of Sociolinguistics* 18(2). 214–222. https://doi.org/10.1111/josl.12070.

Varela Pérez, José Ramón. 2014. *Variation and change in negative constructions in contemporary British English: A corpus-based approach*. Santiago de Compostela: Universidade de Santiago de Compostela dissertation.

Viereck, Wolfgang. 1986. Social dialectology: A plea for more data. In Harold B. Allen & Michael D. Linn (eds.), *Dialect and Language Variation*, 415–425. Orlando, FL: Academic Press.

Wagner, Suzanne Evans, Ashley Hesson, Kali Bybel & Heidi Little. 2015. Quantifying the referential function of general extenders in North American English. *Language in Society* 44(5). 705–731. https://doi.org/10.1017/S0047404515000603.

Walker, James A. 2013. Variation analysis. In Robert J. Podesva & Devyani Sharma (eds.), *Research Methods in Linguistics*, 440–459. Cambridge: Cambridge University Press.

Walker, James A. 2020. Modelling socio-grammatical variation: Plural existentials in Toronto English. In Karen V. Beaman, Isabelle Buchstaller, Sue Fox & James A. Walker (eds.), *Advancing Socio-Grammatical Variation and Change: In Honour of Jenny Cheshire*, 165–184. New York: Routledge.

Wallage, Phillip. 2017. *Negation in Early English*. Cambridge: Cambridge University Press.

Wasow, Thomas & Jennifer Arnold. 2005. Intuitions in linguistic argumentation. *Lingua* 115(11). 1481–1496. https://doi.org/10.1016/j.lingua.2004.07.001.

Weiner, E. Judith & William Labov. 1983. Constraints on the agentless passive. *Journal of Linguistics* 19(1). 29–58. https://doi.org/10.1017/S0022226700007441.

Weinreich, Uriel, William Labov & Marvin I. Herzog. 1968. Empirical foundations for a theory of language change. In Winfred P. Lehmann & Yakov Malkiel (eds.), *Directions for Historical Linguistics*, 97–195. Austin, TX: University of Texas Press.

Wilson, John & Alison Henry. 1998. Parameter setting within a socially realistic linguistics. *Language in Society* 27(1). 1–21. https://doi.org/10.1017/S0047404500019709.