Sociolinguistics and Minimalist Syntax

David Adger, E Jamieson and Jennifer Smith

DRAFT 7th December 2020

1. Separation

One of the most famous passages from Chomsky's (1965) *Aspects of the Theory of Syntax* is the following:

"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, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance. (Chomsky 1965:3)."

This passage was written in a context where many linguists of the time would have assumed that linguistic theory should be concerned with external linguistic behaviour: linguists should classify and analyse utterances into phonemes, morphemes, words, phrases, sentences etc. The behaviour was the focus of the analysis. The passage was, like many of Chomsky's writings at the time and since, an attempt to justify a different viewpoint, to shift the perspective of linguistics from one which is concerned with language as observable behaviour, to one which is concerned with the internal cognitive systems that are (partly) responsible for that behaviour. Beyond this, the goal was to highlight the importance of understanding what the general properties of those cognitive systems are that go beyond the idiosyncratic features of an individual's grammar, and so could be said to be part of the linguistic endowment of humans in general. Chomsky's point here is that linguistic behaviour is evidence for linguistic theory, not its object.

Chomsky makes this point by proposing a number of idealizations. One is an idealization away from memory limitations, shifts of attention, etc., allowing one to leave aside such properties of performance when attempting to characterize a system of linguistic knowledge. Another is an idealization about the speaker-listener: they know the language of the speech community they belong to perfectly, allowing the theorist to leave aside issues of acquisition, second language status, and attrition. A final idealization is about the speech community itself, where by speech community, Chomsky means something like "speakers of English/Mandarin/Yoruba/etc.": this idealization takes the behaviour of the members of that community to be essentially alike, so idiolectal and dialectal variation can be ignored by the theorist.

Chomsky's proposition is that, if one builds a theory adopting these idealizations, that theory will then more closely approach the nature of the cognitive system he takes to be central to accounting for human language: the grammatical system (both (morpho)-syntax and phonology).

These idealizations are not, however, injunctions, and of course linguistic theory has developed in depth of understanding and breadth of empirical coverage over the decades. Questions about first and second language acquisition, attrition, processing and idiolectal/dialectal variation require us to relax these idealizations in particular ways. If one is interested in the other systems of cognition that enter, with the grammatical system, into linguistic behaviour, then such relaxation is necessary, even adopting the perspective Chomsky endorses.

The issue of whether Chomsky's (1965) idealizations were relevant in linguistics more broadly became quickly important in the development of sociolinguistic work, where these idealizations were thought to be problematic. For example, Weinreich, Labov & Herzog (1968:100) argue "the generative model for the description of a language as a homogenous object is itself needlessly unrealistic and represents a backward step from structural theories capable of accommodating the facts of orderly heterogeneity". Weinreich et al. point here to the issue of how the generative models of the time generally idealized away from the issue of variation within language.

If one wanted to analyse variation within a language and at the same time to adopt the *Aspects* model of grammars as constituted of optional and obligatory rules, two linked empirical questions arise for the theory: how do obligatory rules vary across speakers (idiolectal/dialectal variation), and how do optional rules vary across speakers and how might they be "orderly"? In other words, what is the nature of variation within an individual's grammar and how does that variation itself vary across individuals?

To tackle these questions, it's important to have a clear conception of what we mean by variation in linguistics. Variation between grammars is fairly clear: are the basic components of the grammars set up identically or not? Variation within a grammar is a slightly different concept. It is, of course, obvious that no two utterances are identical, but two utterances can be, in some sense, linguistically the same. This is clearest with respect to the notions of phoneme and allophone, but holds for other levels of linguistic structure too, entailing a certain kind of variation: a single unit at one level corresponds to multiple units at another. Some of this variation is determined categorically by linguistic context, as in the classical notions of phoneme and allophone. Some of this variation is not. The Structuralist tradition in North America usually took variation that was not deterministic to be "free": essentially arbitrary and outside of the linguistic system (Bloch 1948). Early generative grammar took the same perspective.

Labov's early work, brought together in his 1966 volume *The Social Stratification of English in New York*, showed that the relationship between phonemes and their realizations was neither categorical, nor arbitrary. A single phoneme may have multiple realizations, and those realizations were, Labov argued, influenced by, but not determined by, linguistic and social factors, so the variation was not "free" but structured into the "ordered heterogeneity" that Weinrich et al. discuss.

Key to uncovering this structure was the notion of "linguistic variable" (Labov 1966), loosely described as "variable ways of saying the same thing" (Labov 1978). Labov (1972a), in his study of negative concord, also demonstrated that "orderly heterogeneity" existed not only

in phonetic form, but also in morphosyntax. In that paper he adopted a transformational rule-based account of negation involving the interaction of a categorical rule of negative attraction (which handled the syntax of negative quantifiers in subject position) with a variable rule of negative concord (which accounted for when a negative quantifier is preceded by negation). The negative concord rule was constrained to operate in certain linguistic contexts. Labov showed that the syntax of negation across dialects and idiolects of English could be understood by weighting the linguistic contexts on the negative concord rule. The rule applied across a wider set of contexts in some dialects compared to others. For example, in some dialects it was restricted to only apply within clauses, while in others it could apply cross-clausally. He argued that these contexts could be understood as probabilistic conditions on the application of the rule, and different conditions applied in different dialects and idiolects. In addition, there was a crucial social aspect to these linguistic conditions. For example, young Black speakers who were members of tight-knit Black peer groups in New York had close to categorical use of the negative concord rule, while those who were on the fringes of those peer groups used it only a proportion of the time. Later work expanded the social contexts beyond dialectal and idiolectal differences to class, ethnicity, gender and age. Further, a growing range of linguistic phenomena were shown to be variable in this way, which opened up a new understanding of how languages vary and change.

This new way of thinking meant that the methods and the theoretical models for understanding language had to be quite different from those that Chomsky laid out for generative grammar. Rather than idealizing away from the behaviour of members of the speech community, one needed to gather rich empirical data from that community. In other words, a bottom up approach was developed, based on observed linguistic behaviour. Similar to Chomsky's idealization, the sociolinguistic speech community "share rules of grammar" but crucially these are "in the form of variable rules" (Labov 1982:19), which accounted for the observed orderliness within the variation. Further, given that the grammar was partly constituted by not just grammatical forms, but also by the probabilistic influence of the linguistic and social contexts in which those forms existed, it seemed clear that such data was best understood through statistical models. Computer programs were developed that could model the multidimentional factors influencing variant form choice and how those factors were weighted with respect to each other. The theoretical explanation of the patterns uncovered was best made, it was argued, in terms of linguistic and social categories that were constituted at a supra-individual level, as opposed to the individual internal nature of grammars in the generative tradition.

These two approaches to language focused on different research questions, but beyond this, they took quite different positions on the nature of language. Because of this, from the 1970s to the end of the 20th century, the two approaches continued to develop in different ways, looking at different aspects of linguistic data, using different methodologies, developing quite different theoretical perspectives, and paying only scant attention to each other's results.

One of the clearest differences was in data, and methods for collecting and analyzing data, which emerged from the very different views of what language is. Work in generative grammar throughout the 1970s and 1980s focused on theory development and extending

the cross-linguistic base of the theory. This would have been close to impossible without the use of, primarily introspective, acceptability judgments as evidence for or against particular theoretical ideas. At the same time sociolinguistic approaches argued against the use of acceptability judgments except as an exploratory tool. A major part of the reason for this was the focus on language as a supra-individual object. Labov showed that speakers might "agree that a certain form is completely unacceptable, yet use it themselves freely in every-day speech" (Labov 1996:78). He suggested that this arose from the non-standard, often highly stigmatized nature of particular dialect forms, where such social conditions may intervene to obscure what is acceptable or not in a speaker's grammar, leading to the conclusion that that "if the social pressures on a minority or local community are well recognized, we can employ considerable caution in interpreting intuitive judgments" (Labov 1996: 102). Sociolinguists instead advocated the notion that the core data should be derived from spontaneous, naturalistic speech (Labov 1972b) as gleaned in, for example, the sociolinguistic interview (Labov 1982). Such sociolinguistic data was seen as a sample of the object of study, and a careful set of methodological principles was developed around handling it.

At the theoretical level too, the perspectives diverged, especially when attention was paid to morphosyntactic variation. Labov's early work on syntactic variation in negative concord discussed above involved a variable rule. However, many syntactic phenomena, even when they apparently involve a linguistic variable, are not related by rule. For example, Weiner and Labov (1982) argue that agentless passives and impersonal constructions form a linguistic variable, but these constructions are clearly not related by syntactic rule. At the same time, Lavandera (1978) showed that attempts to extend the classical notion of sociolinguistic variable (multiple forms with the same meaning) to morphosyntax were problematic, and later work by Romaine (1984) and Cheshire (1987) pressed home the point that, without a clear notion of what sameness in meaning is, the notion of linguistic variable was simply not apposite for many cases of syntactic variation.

These theoretical and methodological developments in sociolinguistics took place during a period when generative grammar was itself moving away from using phrase structure and transformational rules towards the Principles and Parameters approach. In that approach, grammatical rules were replaced by the idea that all languages shared broad structural principles of Universal Grammar, which could vary along limited parameters, so that a grammar for a language was a setting of these parameters, rather than a system of rules. It was unclear at the time how the probabilistic information that was so central to the sociolinguistic project could be incorporated into such a theory.

The key differences between syntactic theory and variationist sociolinguistics at the time can be elegantly captured by Wolfram's observation that "Most studies of linguistic variation are more preoccupied with working out the descriptive details of fluctuation than they are with the specification of parameters that bind fluctuating form' (Wolfram 1991:22)."

It became, then, over time, less and less obvious how to reconcile the philosophical, methodological and theoretical differences between these two approaches to understanding language (Labov 1987).

2. Rapprochement

Government and Binding Theory, a particular theoretical system developed during the 1980s within the broader Principles and Parameters framework, was an attempt to attack the question of how to account for the wide grammatical diversity of the world's languages, while developing a theory which was highly constrained, so that the inductive generalizations of the child acquiring her language were restricted. The architecture of this approach posited a number of principles, modular in nature but highly interactive, which could vary in certain respects (via parameters) across grammars. This allowed the theory to both tackle the way that human languages were acquired and still account for variation across languages.

The parameters can be thought of as the positions of metaphorical binary switches (Chomsky 1986:146): for example, one grammar might be thought of as having the specification …1101001… with a minimally different grammar being …1101000…, where only a single parametric switch is in a different position.

This idea, that differences between languages could be dealt with by specifying parameters in their grammars was originally focussed on fairly large differences between languages, such as whether they allowed extraction from the same domains (Rizzi 1982), or what their basic word order was (Travis 1984). However, even early on in the development of Government and Binding Theory, Borer (1984:29) hypothesized that all differences between languages could be tied down to idiosyncratic properties of lexical items, and, further, that the relevant lexical items were not the major categories (Noun, Verb, Adjective), but rather what we now call functional categories such as tense, agreement, aspect, definiteness etc. Borer had taken these to be elements of the inflectional system of a language, while Fukui (1988) adapted her proposal to the emerging view of clause structure during the 1980s, expanding its scope to include categories such as complementizers, determiners, etc.

Borer's hypothesis was, at the time, not obviously true, given that many basic word order properties seemed to be best stated as properties of major categories (such as the Head Parameter). However, further developments in the theory of phrase structure and movement in the early 1990s allowed one to see how small changes in the properties of single functional heads in a grammar could give rise to large changes in how the language that that grammar underpinned looked (Kayne 1996). By the time of Chomsky's (1995) *The Minimalist Program*, this idea was central to the generative theory of cross-linguistic variation.

As Kayne (2000) pointed out, this view of parameters not only allowed one to capture differences between languages, but should also extend to dialects and even idiolects. Kayne distinguished parametric variation between languages from microparametric variation between individuals who spoke the same language (though technically the two kinds of variation are accomplished by the same theoretical mechanisms, so the distinction is on a cline).

This set of developments meant that as early Minimalist syntax developed through the mid 1990s, there was actually a theory of how to begin to model at least some aspects of the behaviour of a non-homogeneous speech community. Given that parameters, including microparameters, are a fundamental theoretical concept in need of investigation, it followed that evidence from non-homogenous speech communities and, indeed, non-ideal speaker-listeners was necessary, so the idealizations that had held sway within theoretical syntax since *Aspects* needed to be modified somewhat. Generative grammar had always included some work on dialectal and idiolectal variation (Kimball & Aissen 1971, Carden 1973), but early Minimalism, via the notion of microparameters as minimally different properties of functional elements, provided a theoretical system that allowed researchers to tackle the linguistic knowledge of non-standard speakers and the syntax of varieties which minimally differed from each other and from the standard.

One of the earliest attempts to use Kayne's microparametric ideas to model the variation found in a non-homogeneous speech community was Henry (1996), who examined the morphosyntax of Belfast English via a number of small parametric differences from other varieties of English.

One case that Henry addressed was variation in word order in imperatives in Belfast English of the following sort

(1) Sit you down!

(2) You sit down!

Henry showed that there is variation within the Belfast speech community in who allows the inverted form in (1), but that all speakers who do allow it, also allow the uninverted form in (2). This variation appeared to be changing across the generations: Henry showed that, for younger speakers, only unaccusative verbs allow the inversion, so for these speakers, but not for the older speakers, there is a contrast between (3) and (4):

(3) Go you away!

(4) *Read you that!

Henry modelled this as a microparameter that was very similar in nature to parameters which distinguish the overall word order of languages. For the older speakers, she proposed that the main verb raises to C in these imperatives, while for the younger speakers it remains in situ. Further, for both sets of speakers, the subject can remain overt and in situ when there is no Tense node (which is absent in imperatives), distinguishing the variety from other varieties of English. The parameter that causes the verb to raise is very clearly a microparameter: it involves a single property of a particular C (imperative C – main verbs do not raise to declarative or interrogative C in Belfast English). Further, to explain the variability between (1) and (2), the parameter that causes verb movement must be optionally set in a way reminiscent of the optional rules that were used to handle free variation in early generative grammar.

Similar work on dialectal variation further enriched the phenomena that generative syntax investigated (Barbiers et al. 2002, Barbiers et al. 2008). Some of this research was given impetus by Kayne's remark that closely related varieties that differed in just small ways were a kind of natural controlled experiment allowing linguists to determine the varied effects and properties of single parameters (Kayne 2005:8-10). However, such work, while explaining the differences between idiolects, does not immediately extend to the kinds of variation where different variants could be used by the same speaker even in the same sentence. A striking example of such individual level variation comes from Smith's (2000) work on variation in the Buckie speech community of North-East Scotland:

(5) A lot of families does na get what that cats get.
"A lot of families don't get what those cats get."

Here we see variability in verbal agreement, with the plural subjects triggering both singular agreement (*does*) and plural (*get*).

There are a number of ways of tackling such variation. Most syntacticians' immediate instincts are to look at the particular grammatical properties of the sentences (do you get singular in the first clause because the subject has the singular noun *lot* as a potential head? Does the apparent singular demonstrative not have the same effect in the second clause because it is a non-head? Is the distinction because of *do*-support in the first clause but a main verb in the second? Etc.). The intuition is that the variation will actually end up being categorical, once enough contextual information is supplied. Sociolinguistic work uses similar linguistic categories, usually informed by fairly observable surface properties of grammar to examine this kind of variation (what are the grammatical properties of the subject in terms of case, agreement features etc? What are the grammatical properties of the verb, such as main vs. auxiliary status, tense, strong/weak verb type, etc.?). However, the sociolinguists expectation is not that the variability will eventually turn out to vanish, but rather that it will be structured by these linguistic (and potentially other social) contexts in a probabilistic way. We return to this difference in expectations in the next section.

Though these developments brought sociolinguistic and generative research closer together philosophically, and in terms of acknowledging that an idealization away from variation was no longer necessary for theory development, the microparametric approach had no space within the model for structured individual level variation of the kind that is captured by variable rules or Labov's notion of linguistic variable. The two approaches were becoming more compatible, but were still distant.

3. Reconciliation

Over the same period as the development of work on microparameters, an alternative viewpoint was being developed from generative syntactic work in language change that provided a way to think about the question of individual level variation.

Kroch (1989) investigated a number of historical changes in a variety of languages, arguing that the changes took place at a constant rate across time, irrespective of the context that the changing elements found themselves in (main clause vs embedded, etc.), a result he

called the Constant Rate Effect. This paper drew upon both sociolinguistic methodology and Government and Binding theory analyses of clausal and phrasal structure to argue that the Constant Rate Effect entailed that, when grammars changed, they did so at an abstract level, entailing parametric change. Kroch (1994:180) took these findings further and argued that syntactic change was triggered by "competition between grammatically incompatible options which substitute for one another in usage." Kroch argued that these results entailed that an individual speaker's cognitive system contains multiple grammars in competition with each other, each with a distinct parametric setting, and that the choice of which grammar to use in any particular setting may be dependent on both grammatical and extra-grammatical factors.

Yang's (2000) dissertation enriched this basic concept with a novel model of language acquisition, which proposed that all possible grammars, as specifications of parameters, were available to the child initially, and that these were pruned away probabilistically by the input, so that the child was left with a small number of grammars, probabilistically organised, in adulthood.

However, more Minimalist conceptions of grammar had moved away from the approach to conceptualising parameters seen in Kroch and Yang's work. Rather than the locus of parameters being the Principles of Universal Grammar, Minimalism took it to be functional lexical items, following the direction of Borer's hypothesis discussed in section 2. As Borer herself noted, if cross-linguistic variation is attributed to properties of lexical items, "the burden of learning is placed exactly on that component of grammar for which there is strong evidence of learning: the vocabulary and its idiosyncratic properties." (Borer 1984:29).

For example, Chomsky in his 1995 *The Minimalist Program* distinguishes the grammars responsible for the verb-raising differences between English and French discussed by Emonds (1978) and Pollock (1989) by proposing that particular lexical items, functional heads, are "strong" or "weak", leading to different syntactic behaviours. For example, T in French will attract the finite verb, while T in English will not. Chomsky suggested that this was because in both languages T is specified with a verbal category feature V, but in French V is strong, and forces the finite verb, which matches with it, to be local to it. The V feature in English T is weak, and does not. The relevant locality is enforced by the verb raising to T in French. The principles that build syntactic structure are identical across languages, and indeed in both languages even the featural content of the functional category T is the same, but the syntactic properties of these features differ, giving rise to different syntactic structures.

With this move, Chomsky introduced the idea that functional heads, and indeed particular features of functional heads, could have what Adger and Svenonius (2011) would later term "second order features": strength, interpretability, etc. These second order features are treated as instructions to the computational system, and they are Minimalism's way of encoding grammatical dependency. The property of interpretability, for example, is used to create a dependency between matching features at some distance from each other in a clause via the operation Agree, while strength essentially forces matching features to be local to each other via the operation Move (later, Internal Merge). Variation in these second

order features is what gives variation in the grammatical properties of languages. In Rizzi's (2017) terms, there is a "format for parameters," where a parameter in this case is a functional head, bearing a feature that triggers an operation of some sort: Agree and/or Merge (including internal Merge) and/or Spellout.

Adger and Smith (2005) point out that this kind of perspective on parametric variation offers a different way of thinking about individual level variation. Rather than assuming that a speaker has distinct grammars in their head, which was Kroch's model, one can simply say that the range of (functional) lexical items available to a speaker is richer in cases where there is individual level variation. For example, if there is variation in agreement with a subject, this can be thought of as involving not two distinct grammars, with different parametric settings, but the existence of two semantically equivalent T categories, one of which bears an agreement feature that the other lacks. The grammar of the speaker exhibiting variable agreement has an extra option, but there is no sense in which that speaker has multiple grammars.

This perspective unifies the explanation of idiolectal variation between individuals with individual variation within a speaker: it's all just a matter of what lexical items are available to a speaker. The grammatical rules or principles are invariant across speakers (indeed broadly invariant across human beings), but the operation of those principles is dependent on the featural properties of functional heads.

To see how this might work concretely, take an example of variable agreement from Buckie:

(6) a. Buckie boats were a' bonny graint.
'Buckie boats were all nicely grained.'
b. The mothers was roaring at ye comin' in.
'The mothers were shouting at you to come in.'

Here we see a plural subject appearing with a plural verb (*were*) in (6a) and with a singular verb (*was*) in (6b). Under an Adger and Smith type proposal, the idea would be that there are two functional lexical items of category T that can be Merged into clause structure. One bears an uninterpretable number feature, and the other does not. In (6a), T[*u*Num:] Agrees with the subject *Buckie boats*, and is valued as T[*u*Num:pl], which is eventually spelled out as *were*. In (6b), T has the DP *the mothers* Merged as its specifier, but no Agree operation takes place, and the verb is spelled out in a default form insensitive to number, as *is*. Schematically we have:

(7) a. T_1[*u*Num:] … DP[Num:pl] → T_1[~~*u*Num:pl~~] … DP[Num:pl]

b. T_2 … DP[Num:pl] → T_2 … DP[Num:pl]

Agree in (7a) values the number feature of T as plural, but either Agree doesn't apply in 7(b), or its application is vacuous. In terms of the morphology, the valued number feature in (7a) leads to T being spelled out as *were*, while the lack of number feature in (7b) leads to T being spelled out as *was*. Crucially, both T heads have an identical interpretation in the semantics, because the number feature on T is uninterpretable and so is ignored by the

semantic systems (this is the import of the strikethrough of the feature). This provides us with a way of having two distinct phonological forms with the same semantic interpretation in a way that depends on which lexical items the system has and how they combine. Of course, two distinct forms with the same interpretation is the classical definition of the linguistic variable in variationist sociolinguistics.

Not only is this view distinct from the multiple grammars approach, it is also a different way of thinking about variation than the classical view described in Section 1 above, where variation is about how an abstract element at one linguistic level is realized as multiple elements at another (what Adger (2006) calls "Variation in Exponence"). Although it leads to the same effect (different forms at one level associated with the same representation at another), it does so not through rules of exponence, but rather through the possible combinations the system allows. Adger (2006) proposes the term "Combinatorial Variability" for this idea.

Combinatorial Variability also gives us a possible model for understanding non-uniform frequencies in syntactic variants. An example of this is given in Cheshire, Adger and Fox (2013). This is a study of adolescent speech in inner city London. It shows that the teenagers use both *who* and *that* to introduce subject relative clauses with animate heads, just as in Standard English:

(8) apparently a chav is like someone that wears like big gold chains

(9) I'm the only one who's gone to college

However, Cheshire et al. show that, for adolescent speakers in this community, the variation between *who* and *that* is not free (as it appears to be for older members of the community). It is structured by whether the referent of the head noun of the relative clause is informationally important: *who* is used more often than *that* when the referent of the head noun of the relative is an ongoing topic of the discourse. The variation between *who* and *that* is controlled by topicality.

Cheshire et al. propose to capture this behaviour via three lexical items that serve to introduce subject relative clauses. These three items are differentiated featurally as follows:

(10) a. C_1[relativiser:+]

b. C_2[relativiser:+, *u*Animate:+]

c. C_3[relativiser:+, *u*Animate:+, *u*Topic:+]

Consider what happens when a subject relative clause is built with an animate topical antecedent. The relative C will enter into an Agree dependency with that DP. Any of these feature bundles would be able occur with an animate topical antecedent.

Schematically, we have the following situation:

(11) a. NP[Animate:+, Topic:+] C_1[relativiser:+]

b. NP[Animate:+, Topic:+] C_2[relativiser:+, ~~*u*Animate:+~~]

c. NP[Animate:+, Topic:+] C_2[relativiser:+, ~~*u*Animate:+~~, ~~*u*Topic:+~~]

All three possibilities have identical semantic interpretation, since the features checked under Agree and annotated by a strikethrough are ignored by the semantic interface. Now assume that C_1 has the phonological form *that*, while both C_2 and C_3 have the phonological form *who*. On the assumption that the lexical items are chosen randomly, we model a 2/3rds proportion of *who* variants: there are more ways, in this context, for the system to provide *who* than *that*.

Contrast this with when the antecedent is non-topical. In this case, Cheshire et al. assume that it is syntactically marked as [Topic:-]. Such a non-topic DP will be unable to combine with C_3, as the feature values clash.

(12) a. NP[Animate:+, Topic:-] C_1[relativiser:+]

b. NP[Animate:+, Topic:-] C_2[relativiser:+, ~~*u*Animate:+~~]

c. *NP[Animate:+, Topic:-] C_2[relativiser:+, ~~*u*Animate:+~~, ~~*u*Topic:+~~]

Given that C_1 is pronounced *that* and C_2 is pronounced *who*, this models the study's finding that a higher proportion of relativizers pronounced as *who* occur when the antecedent is topical.

This is a different way of modelling the probabilistic nature of variation than the standard approach in sociolinguistics: rather than attaching probabilities to rules, it models different frequencies found in corpora by a system that gives different routes to a particular phonological form. The variants are really the syntactic categories (unlike in a Variation in Exponence model, where they are the exponents). This model of variation then interacts with statements or rules of exponence. If each variant has a distinct exponent, then the number of usable syntactic categories will provide the proportion of expected variants. However, if some variants are homonymous, then this will give rise to a non-random frequency distribution of exponents, just as we've seen in the Cheshire et al. study. Adger (2006) provides an algorithm that derives the distribution of homonymy in the variants and argues that that algorithm gives the correct results in a range of cases.

A major advantage of the Combinatorial Variability approach is that it says nothing new. Minimalist syntactic theory, as generally understood, contains all the mechanisms that are necessary to understand parametric variation, individual level variation, and indeed frequency effects in the latter. The theory does not need to be extended because it already provides a model for the kinds of variation that are of interest to sociolinguists. The Combinatorial Variability approach has also been used to provide a way of theoretically constraining how linguistic contexts can impact on variants (Adger 2014), and how dialectal

variation and combinatorial variation interact in a single system. For example, Adger & Smith (2010) model dialectal variation as a difference in feature inventory, and show how this can account for patterns of variability and categoricity in agreement.

While Combinatorial Variability allows one to model both dialectal variation and intra-speaker variability, thus bringing together key aspects of Minimalist syntax and sociolinguistics, Combinatorial Variability, unlike variable rules, doesn't provide a space for integrating social categories or the impact of processing on variants. To model this, Adger (2007) proposes a more general model of how the Combinatorial Variability system integrates into the systems of use.

Adger terms the collection of lexical items that can combine in the syntax the Pool of Variants (so C_1, C_2, C_3 above would be a Pool of Variants). How homonyms are distributed within the Pool of Variants impacts on the frequency distribution of the surface forms in just the way we've seen here. Adger proposes that the Pool of Variants may also be externally structured by the processing systems that effect lexical access: some items are generally highly routinized and easily accessible, so that when they appear within the Pool of Variants this property itself has an effect on whether that item is chosen; recency may have a similar effect.

Within this model any particular use of a variant can be seen as a selection of an element of the Pool of Variants (PoV) in a particular context. The selection is done via a choice function, U. U itself is sensitive to all sorts of properties of the context of utterance (C) and the variants in the Pool. Schematically, we have

(13) $U(PoV, C) \rightarrow v_i \in PoV$

The model takes the function U to be extremely complex and to be:

"sensitive to all sorts of properties of the elements of PoV: their phonology, their sociolinguistic connotations, whether they have been encountered recently, their frequency of occurrence in the life of the language user who is speaking, whether the language user likes that particular word, etc. It is also sensitive to many aspects of the context of utterance: the information structure of the discourse, pragmatic expectations about the interlocutor's knowledge, social expectations about appropriateness etc." (Adger 2007:696).

The overall system, then, attempts to provide an architecture in which the technology of Minimalist syntax can be combined with the insights of sociolinguistics into a single model. Syntax itself is "socio-free" in the same way that it is phonology-free (Zwicky and Pullum 1983), which is very different to the way that variable rules were used to model variation. Further, (13) is about cognitive capacities, as opposed to being about supra-individual categories, but nevertheless, the overall architecture allows generative syntax and variationist sociolinguistics to integrate in a way that combines the insights of both.

Other approaches to incorporating variability within minimalism have suggested applying the Variation in Exponence idea but within the formal model of Distributed Morphology.

Nevins and Parrott (2010) argue that the morphological operation of impoverishment can be made probabilistic and can be used to explain a wide range of levelling patterns discovered in dialectological work. Henry (2005) has argued that, to deal with individual level variation, variable rules need to be incorporated into minimalist syntax. Following on from an influential volume edited by Cornips & Corrigan which was a major step in reconciling the generative and sociolinguistic approaches to language variation (Cornips & Corrigan 2005), there has been an explosion of work which, while not focusing on the architectural questions we have discussed here, expands minimalist syntactic theory to non-standard data (e.g. Poletto 2006, Tortora 2006, Corver & van Koppen 2010, Hasty 2012, Myer 2013, Biggs 2015, Thoms & Sailor 2018, Wood & Zanuttini 2018, Gallego 2019, Willis 2019, van Alem to appear, Weir 2020).

Just as the theoretical perspectives that once diverged are beginning to be integrated, the methodological divergences that began in the 1970s (outlined in section 1) have also started to narrow. Generative syntax has seen an expansion of methods, including moves away from informal acceptability judgments and towards incorporating formal judgments of gradient acceptability using tasks such as Likert scales (e.g. Schütze 1996). These gradient measurements have opened up the possibility of exploring variation in the grammar of an individual, while still maintaining the judgment task as a primary method.

Furthermore, work in microvariation has designed approaches to judgment tasks specifically to tap into variability, addressing the concerns raised in sociolinguistics (e.g. Labov 1996) that judgment tasks are unreliable sources of data. Cornips & Poletto (2005) detail ways to conduct oral interviews with community insiders, allowing for ad hoc adaptations to materials to avoid judgments based on factors irrelevant to the question at hand, as well as elicitation of alternative grammatical structures. Incorporating qualitative metalinguistic data from these interviews into the analysis of the judgments also works to improve their reliability (Jamieson 2020).

Some of these methodological innovations grew out of large projects that were developed to examine varieties of Dutch, Italian, and the Scandinavian languages at a much more fine-grained scale than traditional dialectology and to build Atlases of the geographical variation (Barbiers et al. 2006, Benincà & Poletto 2007, Johannesson et al. 2014). More recent work in this vein has directly embraced methods from sociolinguistics, building large-scale corpora of both spoken and informal written data and using formal theoretical tools to analyse these (e.g. Tortora et al. 2017, Willis 2020).

While the earlier large scale Atlas projects just mentioned focussed on judgment tasks, the Scots Syntax Atlas (Smith et al. 2019) gathered both extensive judgment data and a large spoken corpus from the same set of speakers. This combination of methods captures the natural variation found in the speech community that has been central to sociolinguistic work, as well as the in-depth judgments from speakers about their own grammar that have been so valuable to syntactic theory development, allowing for deeper investigation of individual variation alongside the broader picture of community use (e.g. Thoms et al. 2019 on auxiliary contraction). This work combines the methods of generative grammar and those of sociolinguistics to provide a detailed picture of how specifically syntactic variation

works across geographical domains, and feeds directly into theoretical questions that are important for both disciplines.

Beyond this enrichment of the kinds of data generative grammar is now incorporating, there have also been recent developments in methods for analysis. In particular, there is also potential to combine judgment data with statistical mapping models in order to evaluate hypotheses about the underlying sources of variation in a grammar.

For example, van Craenenbroeck et al. (2019) use Correspondence Analysis (Greenacre 2007) in order to assess how well different theoretical accounts of verb clusters can account for the considerable levels of variation in these structures found across Dutch dialects. This study looks at how different verb cluster orders distribute geographically and determines statistically how well that corresponds with possible analyses of these orders via 64 parameters (conceived of as properties of functional categories). It builds on the assumption that the variation can be accounted for at least in part by the grammar and argues that "cluster orders that are close together… should be the result of the same or highly similar parameter setting, while orders that are further apart should have fewer parameter values in common." (van Craenenbroeck et al. 2019:348). The relevant parameters for the variation can then be established according to how much variance they can explain in the data. An important result is that incorporating these kinds of statistical models into syntactic theories can allow us to see what proportions of variation can be accounted for by the grammatical system itself, and how much is due to external properties contributed by other cognitive systems, such pragmatic or social factors.

4. Conclusion

This article has sketched out in very broad terms the reasons why generative grammar and sociolinguistics, which were closely aligned in the early years, diverged, and how theoretical developments in generative grammar, especially since the advent of the Minimalist approach, has enabled a rapprochement between the two fields. Certainly, at the level of providing analytical and theoretical tools, and in terms of an interest in the details of variation within and across grammars and varieties, the two fields are now more closely aligned than for many years. There is however a good way to go. The issue of how to understand linguistic contexts for variants within minimalism is still wide open (though see Adger 2014), and a syntactic theory of this may well be important for understanding what the constraints on linguistic contexts are. This is an area where a joint effort could yield considerable results. There is also the question of how to more deeply understand the interaction between systems of use and systems of linguistic structure, both empirically and architecturally. Nevertheless, we contend that the advent of ideas from Minimalist Syntax have helped bring the insights of sociolinguistics together with those of generative grammar.

5. References

Adger, D. 2006. Combinatorial variability. Journal of Linguistics. 42(3) 503-530.

Adger, D. 2007. Variability and modularity: A response to Hudson. Journal of Linguistics. 43(3) 695-700.

Adger, D. 2014. Variability and grammatical architecture. In C. Picallo (ed.) Variation in a minimalist perspective. New York: Oxford University Press. 179-196.

Adger, D. & Smith, J. 2005. Variation and the Minimalist Program. In L. Cornips & K. Corrigan (eds.) Syntax and variation: Reconciling the biological and the social. Amsterdam: Benjamins. 149-178.

Adger, D. & Smith, J. 2010. Variation in Agreement: a lexical feature-based approach. Lingua. 120(5) 1109-1134.

Adger, D. & Svenonius, P. 2011. Features in minimalist syntax. In C. Boeckx (ed.) The handbook of linguistic minimalism. Oxford: Blackwell. 27-51.

Barbiers, S., Cornips, L. & van der Kleij, S. 2002. Syntactic microvariation. Electronic publication of the Meertens Instituut. http://www.meertens.knaw.nl/projecten/sand/synmic

Barbiers, S. et al. 2006. Dynamic Syntactic Atlas of the Dutch Dialects (DynaSAND). Amsterdam, Meertens Institute. http://www.meertens.knaw.nl/sand/

Barbiers, S., Koeneman, O., Lekakou, M. & van der Ham, M. H. 2008. Microvariation in syntactic doubling. Bingley: Emerald.

Benincà, P & Poletto, C. 2007. The ASIS enterprise: a view on the construction of a syntactic atlas for the Northern Italian dialects. Nordlyd 34 35-52.

Biggs, A. 2015. A new case for A-movement in Northwest British English. In Proceedings of the 32nd West Coast Conference on Formal Linguistics. Cascadilla Proceedings Project. 218-227.

Bloch, B. 1948. A set of postulates for phonemic analysis. Language 24(1) 3-46.

Borer, H. 1984. Parametric syntax: Case studies in Semitic and Romance languages. Dordrecht: Foris.

Carden, G. 1973. Dialect Variation and Abstract Syntax. In R. Shuy (ed.) Some new directions in linguistics. Washington: Georgetown University Press. 1-34.

Cheshire, J. 1987. Syntactic variation, the linguistic variable, and sociolinguistic theory. Linguistics 25 257-282.

Cheshire, J., Adger, D. and Fox, S. 2013. Relative *who* and the Actuation Problem. Lingua 126. 51-77.

Chomsky, N. 1965. Aspects of the theory of syntax. Cambridge, MA: MIT Press

Chomsky, N. 1986. Knowledge of language: Its nature, origin and use. New York: Praeger Publications.

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

Cornips, L. & Corrigan, K. 2005. Syntax and variation: Reconciling the biological and the social. Amsterdam: John Benjamins.

Cornips, L. & Poletto, C. 2005. On standardising syntactic elicitation techniques. Lingua 115 939-957.

Corver, N., & van Koppen, M. 2010. Ellipsis in Dutch possessive noun phrases: A micro-comparative approach. Journal of Comparative Germanic Linguistics 13 99–140.

Emonds, J. 1978. The verb complex V'-V in French. Linguistic Inquiry 9(2) 151-175

Fukui, N. 1988. Deriving the differences between English and Japanese: A case study in parametric syntax. English Linguistics 5 249–270.

Gallego, A. 2019. The syntactic variation of Spanish dialects. Oxford: Oxford University Press.

Greenacre, M. 2007. Correspondence analysis in practice. London: Chapman & Hall.

Hasty, J. D. 2012. We might should oughta take a second look at this: A syntactic re-analysis of double modals in Southern United States English. Lingua 122 1716-1738.

Henry, A. 1995. Belfast English and Standard English: dialect variation and parameter setting. Oxford: Oxford University Press.

Henry, A. 2005. Non-standard dialects and linguistic data. Lingua 115(11) 1599–1617.

Jamieson, E. 2020. Viewing dialect change through acceptability judgments: A case study in Shetland dialect. Glossa 5(1):19 1-28

Johannessen, J. B., Vangsnes, Ø, Priestley, J., & Hagen, K. 2014. A multilingual speech corpus of North-Germanic languages. In Spoken corpora and linguistic studies. Amsterdam: John Benjamins. 69-83.

Kayne, R. 1996. Microparametric syntax: some introductory remarks. In J. Black & V. Motapanyane (eds.) Microparametric syntax and dialect variation. Amsterdam: Benjamins. ix – xviii.

Kayne, R. 2000. Parameters and universals. Oxford: Oxford University Press.

Kayne, R. 2005. Some notes on comparative syntax, with special reference to English and French. In G. Cinque & R. Kayne (eds.) The Oxford handbook of comparative syntax. Oxford: Oxford University Press. 3-69.

Kimball, J. & Aissen, J. 1971. I think, you think, he think. Linguistic Inquiry 2 241-246.

Kroch, A. 1989. Reflexes of grammar in patterns of language change. Language Variation and Change 1 199-214.

Kroch, A. 1994. Morphosyntactic variation. In Beals, K., J. Denton, R. Knippen, L. Melnar, H. Suzuki and E. Zeinfeld (eds.) Papers from the 30th regional meeting of the Chicago Linguistic Society, vol. 2: The parasession on variation in linguistic theory. Chicago: Chicago Linguistic Society. 180-201.

Labov, W. 1966. The social stratification of English in New York City. Washington, D.C.: Center for Applied Linguistics.

Labov, W. 1972a. Negative attraction and negative concord. Language 48(4) 773–818.

Labov W. 1972b. Some principles of linguistic methodology. Language in Society 1(1) 97–120.

Labov W. 1978. Where does the sociolinguistic variable stop? A response to Lavandera. Working Papers in Sociolinguistics 44. Austin: Southwest Educational Development Laboratory.

Labov, W. 1982. Building on empirical foundations. In W.P. Lehmann & Y. Malkiel (eds.). Perspectives on historical linguistics. Amsterdam: Benjamins. 17-92.

Labov, W. 1987. Some observations on the foundation of linguistics. http://www.ling.upenn.edu/~wlabov/Papers/Foundations.html

Labov, W. 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: Chicago Linguistic Society. 77-106.

Lavandera, B. 1978. Where does the sociolinguistic variable stop? Language in Society 7. 171-182.

Myler, N. 2013. On 'coming the pub' in the North West of England: Accusative unaccusatives, dependent case and preposition incorporation. Journal of Comparative Germanic Linguistics 16(2–3) 189–207.

Nevins, A. & Parrott, J. K. 2010. Variable rules meet Impoverishment theory: Patterns of agreement leveling in English varieties. Lingua 120 (5) 1135-1159.

Poletto, C. 2006. Asymmetrical pro drop in the Northern Italian Dialects. In P. Ackema, P. Brandt, M. Schoorlemmer & F. Weerman (eds.) Arguments and agreement. Oxford: Oxford University Press. 159–194.

Pollock, J.-Y. 1989. Verb movement, Universal Grammar and the structure of IP. Linguistic Inquiry 20(3) 365–424.

Rizzi, L. 1982. Issues in Italian syntax. Dordrecht: Foris.

Rizzi, L. 2017. On the format and locus of parameters: The role of morphosyntactic features. In M. Piattelli Palamarini & S. Karimi (eds.) The Linguistic Review (41) Special issue on parameters. 159-192.

Romaine, S. 1984. On the problem of syntactic variation and pragmatic meaning in sociolinguistic theory. Folia Linguistica 18 409-439.

Schütze, C. 1996. The empirical base of linguistics: Grammaticality judgments and linguistic methodology. Chicago: University of Chicago Press

Smith, J. 2000. 'You Ø na hear o' that kind o' things': Negative do in Buckie Scots. English World-Wide 21(2) 231–259.

Smith, J., Adger, D., Aitken, B., Heycock, C., Jamieson, E. & Thoms, G. 2019. The Scots Syntax Atlas. University of Glasgow. http://scotssyntaxatlas.ac.uk.

Thoms, G., Adger, D., Heycock, C. & Smith, J. 2019. Syntactic variation and auxiliary contraction: The surprising case of Scots. Language 95(3) 421-455

Thoms, G. & Sailor, C. 2018. When silence gets in the way: extraction from do-ellipsis in British Dialects. In S. Hucklebridge & M. Nelson (eds.) NELS 48: Proceedings of the Forty-Eighth Annual Meeting of the North East Linguistic Society (vol 3). 145-154.

Tortora, C. 2006. The case of Appalachian expletive *they*. American Speech 81(3) 266–296.

Tortora, C., Santorini, B., Blanchette, F. & Diertani, C. 2017. The Audio-Aligned and Parsed Corpus of Appalachian English (AAPCAppE), version 0.1. www.aapcappe.org

Travis, L. 1984. Parameters and effects of word order Variation. PhD Thesis. MIT.

van Alem, A. to appear. Licensing imperative subjects without an imperative operator: Evidence from word order in West-Germanic imperatives. Journal of Comparative Germanic Linguistics.

van Craenenbroeck, J., van Koppen, M. & van den Bosch, A. 2019. A quantitative-theoretical analysis of syntactic microvariation: Word order in Dutch verb clusters. Language 95(2) 333-370

Weiner, E.J. & Labov, W.1983. Constraints on the agentless passive. Journal of Linguistics 19 29-58.

Weinreich, U., Labov, W., & Herzog, M. 1968. Empirical foundations for a theory of language change. In W. Lehmann & Y. Malkiel (Eds.) Directions for historical linguistics. Austin: University of Texas Press. 95-195.

Weir, A. 2020. Resultatives, goal PPs, and postverbal subjects: From Scotland to Belfast. Journal of Linguistics 58. doi:10.1017/S0022226720000407

Willis, D. 2019. Dialect syntax as a testbed for models of innovation and change: Modals and negative concord in the Syntactic Atlas of Welsh Dialects. Glossa 4(1) 1-30.

Willis, D. 2020. Using social-media data to investigate morphosyntactic variation and dialect syntax in a lesser-used language: Two case studies from Welsh. Glossa 5(1) 1-33.

Wolfram, W. 1991. The linguistic variable: Fact and fantasy. American Speech 66(1) 33-32.

Wood, J., & Zanuttini, R. 2018. Datives, data and dialect syntax in American English. Glossa 3(1) 1-22.

Yang, C. 2000. Knowledge and learning in natural language. PhD thesis. MIT.

Zwicky, A. & Pullum, G. 1983. Cliticization vs. inflection: English n't. Language, 59(3) 502–513.

.