John Benjamins Publishing Company

This is a contribution from *Questioning Theoretical Primitives in Linguistic Inquiry*
Papers in honor of Ricardo Otheguy.
Edited by Naomi L. Shin and Daniel Erker.
© 2018. John Benjamins Publishing Company

This electronic file may not be altered in any way.
The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.
Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.
For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).
Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

Variable grammars

Competence as a statistical abstraction from performance. Constructing theories from data

Gregory R. Guy
New York University

Linguists generally postulate a mental grammar which children infer from the speech they encounter, and then use to generate their own speech productions. This grammar is often assumed to be invariant and categorical. Language in use, however, is massively variable: the child encounters diversity at the level of dialect, sociolect, and idiolect. Furthermore, all units of language have multiple realizations and fuzzy boundaries. This raises a fundamental question: if the data is variable, even continuous, how does the child arrive at a grammar that is categorical and discrete? I argue that the system that a learner infers is not invariant and discrete, but rather one that recognizes, incorporates, manipulates, and generates variability.

Keywords: child language acquisition, variable grammar, variable rules, variationist sociolinguistics, linguistic variation

1. Introduction

The pioneering work by Cedergren and Sankoff (1974) on the place of variation in linguistic theory begins with the statement "Speech performances are here considered as statistical samples drawn from a probabilistic language competence." The paper articulates the foundational 'variable rule' model for generating a quantitative representation of the distribution of a variable in speech from a set of linguistic (and possibly extra-linguistic) predictors. Hence, it focuses mainly on the question of production: how an underlying competence can generate "orderly heterogeneity" (Weinreich, Labov, & Herzog, 1968): the regular patterns of variability that are evident in speech. As the above quotation indicates, the authors affirm an abstract competence distinct from performance, but one that is probabilistic rather than deterministic. Implicitly, this model also raises questions about acquisition: how

https://doi.org/10.1075/sfsl.76.04guy
© 2018 John Benjamins Publishing Company

the child develops probabilistic linguistic competence through observation and interaction.

The prevailing view of acquisition among linguists postulates a mental linguistic system – a grammar – which children infer from the speech they encounter (perhaps aided by a prewired universal grammar), and then use to generate their own speech productions. We emulate this process in our own work, by inferring theories of grammar from the speech we observe (perhaps aided by 'intuitions' which are presumed to provide fairly direct access to the grammar). In many theoretical traditions, this grammar is assumed to be invariant and categorical. Language in use, however, is massively variable: the child learner and the linguist encounter diversity at the level of dialect, sociolect, and idiolect, as well as stylistic variation within the usage of each individual. Language usage also exhibits what Weinreich et al. (1968) call 'inherent variability': all units of language (articulatory gestures, phonemes, syntactic structures, semantic interpretations, etc.) have multiple realizations and fuzzy boundaries. This raises a fundamental question: if the data is variable, and in some cases even continuous, how does the child arrive at a grammar that is categorical and discrete?

There are two logical approaches to this question. The one that prevails in most formal theoretical linguistics postulates that language acquisition is an abstracting, generalization-seeking process, which overrides observed diversity in the pursuit of categorical rules and units of grammar. This projects the theory-constructing predilections of the linguist onto the mental processes of the child. This approach relegates the diversity of speech to grammar-external processes in production – the deformations that Chomsky famously attributed to "memory limitations, distractions, shifts of attention and interest, and errors" (Chomsky, 1965: p. 3). But there is a second alternative that obviates the question: acknowledge that the linguistic system that the child learns is not invariant and discrete, but rather is one that recognizes, incorporates, manipulates, and generates variability. In other words, the mental grammar reflects the "probabilistic language competence" that Cedergren and Sankoff postulate.

This is the approach that is argued for here. Probabilistic variability is not just an inherent property of language use, but of linguistic competence itself. The mental linguistic system that comprehends 'orderly heterogeneity' in the input and produces it in the output incorporates probabilistic processing in its internal structure. For the linguist, recognizing the probabilistic capacity of linguistic competence provides the basis not only for a theory of language use, but also for an adequate theory of language.

Evidence for this probabilistic competence is apparent in all facets of language. It is abundantly clear that speakers have an exquisitely subtle capacity for understanding social variability. Simply by hearing the voice of a person speaking a

© 2018. John Benjamins Publishing Company
All rights reserved

language that we know, we are able to make good to excellent estimates of whether they are a native speaker or not, what dialect they speak, the speaker's sex, approximate age, and physical or emotional states like sleepiness, illness, anger, tension, humor, level of formality or politeness, familiarity with the interlocutor, etc. In some communities, we may also be able to identify social class or educational level. Besides recognizing and correctly interpreting the linguistic cues associated with these social variables, we have a variable productive capacity – more limited in scope, but universally present in all human beings. We all vary stylistically, adapting our speech to interlocutors, social settings, topics, locations, and purposes; we also manipulate sociolinguistic variables to construct our own social identities, to conduct our relationships with others, and to communicate stances and emotions. The variable elements of language involved in these processes permeate the linguistic system, including phonetic, phonological, morphological, syntactic, and lexical elements.

It is also abundantly clear that variability of this sort is not random and disorganized, but orderly. Linguistic variation is orderly in terms of its systematic relationships with the social dimensions described above, and also in terms of its systematic conditioning by linguistic structure. A half-century of research on linguistic variation shows how variables of all kinds are constrained by the linguistic contexts in which they occur, appearing more frequently under one condition and less under another. The variant realizations of such a variable are analogous to allophones or allomorphs, but instead of occurring categorically in one context and never in another, they are probabilistically distributed in regular patterns across the spectrum of relevant environments.

This linguistic conditioning of variables, the relative favorability or unfavorability of particular contexts to the occurrence of a given variant, provides essential clues to the structure of the linguistic system, just as much as the categorical distributions privileged in much of formal linguistic theory. It is certainly true that categorical conditions abound, giving evidence about language structure: in English, /ŋ/ only occurs in syllable-final position, never initial, and articles only occur pre-head; in German, verbs are always final in subordinate clauses. But probabilistic phenomena also provide significant evidence about grammatical structure. In English, /θ/ is usually, but not always, word-initial or final, while /ð/ is usually intervocalic. This distribution reflects the fact that these were historical allophones of a single phoneme, voiced intervocalically, voiceless elsewhere. The occurrence of final /ð/ in verbs reflects the historical presence of verbal inflections, now lost: *breathe, teethe, bathe*. In syntax, 'heavy,' more complex NPs are more likely but not certain to occur last in double object constructions, perhaps reflecting processing constraints. Overt pronominal subjects are disfavored, but not impossible, in the second clause of conjoined VPs with the same subject referent (e.g., *When it was*

© 2018. John Benjamins Publishing Company
All rights reserved

Harry's turn to speak, he stood up and (he) launched into a diatribe about health insurance.)

Although often treated as a dichotomy, these two kinds of conditioning – the categorical and the probabilistic – are in quantitative terms, simply different points on a continuum. Categorical constraints are at the extreme ends of the probability distribution: under a given condition, the phenomenon of interest always occurs – i.e., has a probability of 1, or never occurs – a probability of 0. A variable occurrence means a probability between 0 and 1. Specific rates of occurrence of a phenomenon can be obtained empirically, but theoretical claims about categorical behavior are not often empirically validated, so some supposedly categorical properties represent the untested hypothesis of the analyst (Bresnan, Cueni, Nikitina, & Baayen, 2005; Bresnan & Ford, 2010).

The non-categorical probabilistic distributions of linguistic variables that occur in natural language are clearly perceptible to speakers. The psychological literature amply documents the human cognitive capacity to track and accurately match probabilities (e.g., Estes, 1950; Gallistel, Krishan, Liu, & Miller, 2014). Thus, probabilistic information is available and accurately processed in both perception and production. It requires something of a leap of faith to postulate that such information is nevertheless excluded from linguistic cognition – i.e., from competence and grammar.

Finally, a probabilistic component in the grammar is essential for an adequate account of language change. Linguistic change is always associated with synchronic variability; a linguistic variable marks a point in the grammar where change is possible. Like biological evolution, linguistic change depends on the occurrence of variety, and consists of the gradual expansion of one variant at the expense of another. All historical changes for which the time-course can be tracked with sufficient resolution show a gradual quantitative shift from old to new forms. An attempt to model such facts in terms of a succession of discretely but infinitesimally different categorical grammars is more faith-based than realistic. A probabilistic grammar provides an integrated account of synchronic variability and diachronic change: what changes across time is the same parameter that governs the choice between the alternant forms at a given point in time.

2. Variation in linguistic theory

Formal theories designed around categorical conditions and discrete distributions have mostly dealt with the facts of variability in three ways. One is to relegate variation to performance, and hence declare it to be irrelevant to or outside of the grammar. We will present data contradicting this position below. The second common

© 2018. John Benjamins Publishing Company
All rights reserved

approach is to postulate that apparent variation is actually categorically constrained by as yet undiscovered conditions. Hence, any apparent 'variation' actually reflects a mixture of discretely different contexts that the analyst has not yet been able to identify, some that are postulated to categorically require variant A, and others that categorically require variant B. This tactic is inspired by historical examples of apparent indeterminacy resolved by subsequent discoveries; the prototype is Karl Verner's (1877) discovery of a stress constraint on the 'Grimm's law' sound changes in Proto-Germanic. Grimm (and others) showed that Proto-Indo-European voiceless stops mostly became voiceless fricatives in Germanic, but in some words the Germanic reflex was a voiced stop. Verner demonstrated that the latter outcome occurred word internally after unstressed syllables. The Neogrammarian school took this as evidence in support of the hypothesized 'exceptionlessness' of sound change; any apparent exceptions would eventually be explained in terms of categorical conditions (Jankowsky, 1972). The Grimm-Verner case is compelling, but as a systematic account of variation, this model is sorely lacking. As Otheguy (2008) dryly notes, "The supposition that, when a successful linguistic analysis is produced, variation will disappear is not well founded." In fact, such a hypothesis is unfalsifiable, and hence it is undeserving of scientific attention.

Formal theories that engage more seriously with questions of variation take a third approach: they seek to delimit the grammatical elements involved in alternation, for example by means of parameters that specify contrasting grammatical structures. These capture typological differences between languages (e.g., Null Subject and Non-null Subject languages), but they can also be used to model language-internal variation in terms of alternating parameter settings. This is essentially the approach taken by Kroch's grammar competition model (Kroch, 1989, 1994), which treats variation during the course of syntactic change in terms of two competing grammars, present simultaneously in the speech community and in the minds of individual speakers. One grammar generates variant A, the other generates variant B (perhaps reflecting contrasting parameter settings), and the mixture in speech reflects speakers' varying use of each grammar. Diachronically, this model explains change in terms of a changing frequency of choosing one grammar over the other.

This approach has also been taken by scholars working in the Optimality Theory (OT) framework. The discrete but competing constraints in such models can, in principle, generate many possible surface forms, depending on their relative rankings or strengths. If rank order is fixed, surface realizations are invariant, but with partial ordering (cf., Anttila, 1997) or weighted stochastic ordering (cf., Boersma & Hayes, 2001), surface variation can be generated, and specific frequencies of occurrence can be modeled. For example, a variable deletion process could be modeled by competition between a constraint favoring deletion and a faithfulness constraint favoring retention. If the FAITH constraint has the highest ranking

© 2018. John Benjamins Publishing Company
All rights reserved

say 75% of the time (or in Anttila's model, in 75% of possible orders), the surface output will show 25% deletion.

A problem that these approaches have faced is the interaction of multiple constraints on a variable process. To take one example, coda –s deletion in Spanish is consistently and significantly constrained by the preceding context, following context, syllabic stress, position in the word, speech rate, and morphological status (cf., Hoffman, 2004, among others). These are orthogonal, and quantitatively cumulative: i.e., a lot of weaker contexts favoring one outcome can collectively outweigh a very strong context favoring a different outcome. An account of such facts in an OT approach will attribute each of these effects to a different constraint set, and generate the surface variability by the many possible orders they can all take. It is not clear that such a grammar can adequately predict the observed patterns of surface frequencies, as it would be very difficult to compute, much less learn. Crucially, a theory that maintains the standard OT convention that the highest ranked constraint prevails, regardless of what outcomes are favored by lower-ranked constraints, cannot replicate the probabilistically cumulative distribution of variants documented in 50 years worth of studies of linguistic variation.

In response to this problem there have appeared variants of OT that question that convention and permit cumulative weighting. A recent study by Zuraw and Hayes (2017) explores approaches in Harmonic Grammar and Maximum Entropy that effectively render these models probabilistic, and, consequently, make possible a more straight-forward computation of multiple constraint effects, what they call "intersecting constraint families." Their favored mathematical model uses a logistic function, just like the variable rule model. At the extremes, categorical effects are accommodated in the model by assigning weights to such contexts that are so large as to permit only one outcome, which is a standard property of logistic models.

Such approaches mark an important step towards a linguistic theory that is empirically grounded and capable of modeling natural language. An intellectually honest appraisal of what linguistic theories need to account for must include the variable properties of language as well as the categorical ones. A probabilistic grammar has room for both, and the advantage of being learnable, computable, and accurate.

3. Probabilistic competence

Leaving aside the substantial work that speakers do in perceiving, acquiring, and producing the socially meaningful elements of linguistic variation, let us consider how the linguistic distribution of variables illuminates the structure of language and facilitates the acquisition of linguistic competence.

© 2018. John Benjamins Publishing Company
All rights reserved

We begin with a phonological example. English has a general process of simplifying coda clusters by deleting coronal stops, so that utterances like *'east side'* or *'send me'* will routinely be produced without the final /t/ or /d/ in the first word. This process is sociolinguistically variable, showing the social and stylistic distribution typical of a stigmatized variant: less deletion by higher status speakers and in more careful speech styles. But, at the same time, it is subject to linguistic conditioning that shows this process to be intimately embedded in the phonology, morphology, and lexicon of English.

One constraint on coronal stop deletion (CSD) that is evident in many varieties of English is the place and manner of the preceding segment. As Guy and Boberg (1997) show (see Table 1), deletion is favored when the preceding segment is coronal (e.g., s,z > f,v; n > m,ŋ), noncontinuant (p,b > f,v; n > l), and nonsonorant (s,z > l; p,b,k,g > m,ŋ). These are the features that characterize the targets of deletion, /t/ and /d/, which are [+ cor, -cont, -son]. Hence, the deletion rates are highest after sibilants, which are coronal and nonsonorant (e.g., *most*), and stops, which are noncontinuant and nonsonorant (e.g., *act*, *apt*), and /n/, which is coronal and noncontinuant (e.g., *hand*). These segments all share two points of phonological similarity (i.e., two features) with the deletion target. By comparison, deletion rates are lowest following segments that share only one point of similarity with the deletion target, such as liquids and noncoronal fricatives (e.g., *cold*, *left*). The obvious generalization is that the probability of deletion is a function of the similarity between the context and the target of deletion: there is more deletion as the preceding segment is more similar to the final /t,d/.

The overall effect of this constraint on CSD is to avoid surface sequences of adjacent similar segments, and to enhance phonological contouring. This thus reflects the Contour Principle (Leben, 1973; Yip, 1988) – the general phonological

Table 1. Preceding segment constraint on English coronal stop deletion (from Guy & Boberg, 1997)

	N	% deleted	Factor weight[a]
Two features shared with target:			
/s,z,ʃ,ʒ/ + cor, -son	276	49	.69
/p,b,k,g/ -cont, -son	136	37	.69
/n/ + cor, -cont	337	46	.73
One feature shared with target:			
/f,v/ -son	45	29	.55
/l/ + cor	182	32	.45
/m,ŋ/ -cont	9	11	.33

[a] Probability estimate of the effect of a predictor, from a multivariate logistic regression with Varbrul2; 1 = categorical deletion, and 0 = categorical retention.

© 2018. John Benjamins Publishing Company
All rights reserved

tendency to prefer alternations of phonological units, which can be observed in constraints on tone sequences, syllable structure (CV), metrics, etc. It operates as a categorical constraint on some processes in some languages (hence the name 'Obligatory Contour Principle,' or OCP). The categorical prohibition against sequences of identical segments in English is one such 'obligatory' case; thus the –s and –ed suffixes of English are obligatorily realized as syllabic in words with a root ending in –s or –t/d (*glasses, baited* vs. *cats, tapped*). The epenthetic vowel in the suffix creates a CVC contour, and avoids a prohibited *ss or *tt sequence.

Phonologists such as those cited above have treated the Contour Principle as part of competence, a constraint on the grammar. The probabilistic grammar argued for here incorporates both the obligatory constraints against adjacent identical elements and the probabilistic constraint against adjacent similar elements that is evident in coronal stop deletion as instantiations of one generalized constraint on the phonology. In the obligatory cases, the constraint effects have a probability of 1, while in the coronal stop deletion case, the effect has a high probability when adjacent segments are very similar and contour violations are most marked, and a declining probability as the segments are less similar and contour violations are minimal.

Coronal stop deletion is also sensitive to morphological structure. A much-replicated finding is that inflected words like *missed* and *packed* undergo significantly less deletion than uninflected words where the final cluster is part of the root, like *mist* and *pact* (cf., Guy, 1980, 1991, 1996; Santa Ana, 1992; Fruehwald, 2012). For many speakers, a third morphological category is also distinctive, the irregular past tense forms where the final stop is an affix, but other alterations to the root are also present, especially root vowel changes; e.g., *keep-kept, tell-told, lose-lost, leave-left*. These words undergo deletion at an intermediate rate. Table 2 gives the relevant deletion rates from three representative studies of the variable in different corpora of sociolinguistic interviews.

Table 2. Morphological class constraint on English coronal stop deletion in three studies (% deleted)

	Monomorphemes *mist, pact*	**Irregular past** *lost, kept*	**Regular past** *missed, packed*
Guy (1991)	38	34	16
Santa Ana (1992)	58	41	25
Fruehwald (2012)	49	37	22

Various explanations of these facts have been proposed. Functionalist accounts appeal to the different functional loads carried by the final stop in each class (Guy, 1996). Accounts that appeal to differences among the classes in internal boundaries (Guy, 1980) or structural organization (Fruehwald, 2012) argue that the

© 2018. John Benjamins Publishing Company
All rights reserved

phonological process interacts with and is constrained by elements of morphological structure. Guy (1991) and Santa Ana (1992) propose a derivational account in the lexical phonology framework where the deletion process applies both within the lexicon and postlexically, such that the three classes differ in the number of points in their respective derivation where they are exposed to the possibility of deletion. This model predicts an exponential order among the rates of stop retention (i.e., non-deletion) of the three classes, which, it will be noted, provides an excellent fit to the data in all the studies cited in Table 2. But all of these theoretical treatments begin from the same understanding of the place of variation with respect to competence: the sensitivity of the process to the morphology of the words it affects is *prima facie* evidence that it is part of the grammar. And since the effect is quantitative – deletion is inhibited in inflections, but not categorically prohibited – this further entails that the grammar is probabilistic.

Probabilistic grammatical constraints on variation are not confined to phonological variables. Similar patterns of non-categorical but grammatically regular distribution of variants occur in syntax, involving relationships at a distance, making reference to constituency structure, sequential ordering, movement, etc. Two examples from Brazilian Portuguese illustrate the kinds of phenomena at issue. Portuguese, like many Indo-European languages, has number agreement between subject and verb, and within the noun phrase. Historically, and prescriptively in the contemporary standard variety, both of these agreement relations are presumed to be categorical: plural subjects are expected to always occur with plural verbs, and all items other than adverbs in a plural NP/DP are expected to be plural-marked. However, in popular speech number agreement is variable in both verbs and NPs. Crucially, this variation is not random, but highly constrained by syntactic structure.

Consider first the constraints on subject-verb agreement in popular Brazilian Portuguese (PBP). Prominent among them is a linear position effect. Subjects located immediately before the verb trigger the highest rates of verbal plural marking. Preverbal subjects that are distant from the verb (i.e., separated by intervening material such as adverbs or modifying phrases) are somewhat less likely to trigger plural marking, while post-verbal subjects are associated with very low rates of plural marking on the verb.

Table 3. Subject position constraint on verbal number agreement in popular Brazilian Portuguese (from Guy, 1981)

Subject position	N	% plural marked	Factor weight
Immediate preverbal	1861	74	.54
Distant preverbal	597	56	.31
Post-verbal	199	27	.18

© 2018. John Benjamins Publishing Company
All rights reserved

Again, these facts are susceptible to various theoretical explanations. Coelho (2000) accounts for the low rates of agreement with post-verbal subjects in unaccusative verbs (e.g., *Sumiu os taxi* 'The taxis disappeared$_{\text{-sg}}$' as opposed to ***sumiram** os taxis* '…disappeared$_{\text{-pl}}$') as a consequence of VP-internal subjects failing to raise to a preverbal position, and hence not undergoing feature checking. Given that post-verbal subjects are now rare in Brazilian Portuguese (as the Ns in Table 3 illustrate), it is possible that the grammar now disqualifies some of these from subjecthood. Feature matching, spreading, or percolation accounts are also possible. But any adequate account of these facts must explain why these particular syntactic structures are associated with these specific rates of occurrence of the syntactic operation that generates number agreement. Agreement is not a randomly selected option in an otherwise categorical grammar; rather it is variable operation embedded in a probabilistic syntax.

Another position constraint on agreement in PBP appears in variable nominal agreement. The observation there, systematically replicated in many studies and obvious from the most minimal inspection of colloquial speech, is that plural markers abound early in an NP, and become rarer towards the end. Relevant data from two studies with robust Ns are given in Table 4: Guy (1981) studied illiterate working class speakers in Rio de Janeiro, while Oushiro (2014) investigated a social cross section of speakers in São Paulo. Numerous studies in other regions of Brazil have found similar results (e.g., Braga, 1977; Scherre, 1988).

Table 4. Position constraint on nominal number agreement in popular Brazilian Portuguese in two studies.

Position in NP	Guy (1980) – Rio de Janeiro		Oushiro (2014) – São Paulo	
	N	% plural marked	N	% plural marked
First	5,247	95	14,246	99.7
Second	3,947	28	16,934	86
Third	552	21	2,725	86
Fourth and fifth	42	11	225	87

Plural marking in PBP is nearly categorical in the first word in an NP. In the Rio de Janeiro data, drawn from maximally vernacular speakers with no schooling and low social status, there is a huge drop (67%) in the second position, followed by progressively lower rates of plural marking in later positions. In the São Paulo data, drawn from a corpus that includes many speakers with secondary or tertiary education and middle to high social status, the decline in plural marking between the first and second positions is a more modest 15%, and there is no significant decline thereafter. In both varieties, it is almost always the case that marking never skips a

© 2018. John Benjamins Publishing Company
All rights reserved

word; any unmarked word will never be followed by a marked word[1] (hence *as casas branca, as casa branca* 'the white houses' are possible, but *a casas branca, as casa brancas, a casa brancas* are unattested, and typically rejected by native speakers.)

These data are presented in terms of the linear position of words in the DP/NP, but other studies have pursued a more detailed breakdown by the structure of the phrase. By far the most common phrase structure in the data is a Determiner + Noun sequence, in which the determiner will almost always be marked and the noun will often not be marked (*os amigo, umas casa* 'the friends, some houses'). But the initial position may also be occupied by an adjective (*velhos amigo* 'old friends'), possessive (*meus amigo* 'my friends'), quantifier (*todos amigo* 'all friends'), or even a noun (*amigos velho* 'old friends'), and in almost all of these structures, the first word will bear a plural marker. Scherre (1988) argues for a structural analysis that contrasts items occurring before the nominal head, the nominal head itself, and items occurring after it. However, for most of the data, this kind of analysis is coincident with one relying on the linear position.

What grammar of agreement is adequate for generating this regular pattern in PBP? Simply declaring agreement to be optional rather than obligatory is clearly inadequate. A satisfactory account needs to generate essentially obligatory marking in initial or pre-head position, more or less regardless of the word class occupying that structural slot, and declining probabilities of plural marking in subsequent positions. It must also predict that marking never skips an eligible word (i.e., there is no marked word with an eligible unmarked word to its left in the same phrase).

One model that satisfies these requirements relies on probabilistic feature spreading: the plural phi-feature is attributed to a node dominating the entire phrase, and then percolates or spreads down the tree from left to right, with a probability of recursion that is less than 1. The spreading model explains the absence of skipping: if the plural feature has not spread to a given node, it is not available for spreading to a subsequent node.

Alternative models to explain these facts that are not internally probabilistic are difficult to sustain. A grammar competition model (Kroch, 1994) might work for the São Paulo data, if it postulated alternation between a grammar with obligatory agreement and one with only initial plural marking, with the former selected about 86% of the time. But for such a model to generate the pattern found in the Rio de Janeiro data (and elsewhere in other studies) – a progressive decline in plural

1. Some exceptions to this occur in phrases with double determiners, such as *as minhas casas* 'the my houses' and *todas as casas* 'all the houses.' Such cases are occasionally produced without a plural marker on the first word: *a minhas casas, toda as casas*. This may indicate that the first element projects a higher node outside the core DP in which agreement operates.

© 2018. John Benjamins Publishing Company
All rights reserved

marking in later positions – would require either a very complicated inventory of competing grammars, or some statement about which grammar is selected at what rate for NPs of different lengths. It would also require some stipulation to prohibit skipping. None of these models would offer a compelling alternative to a grammar with probabilistic generalizations, rules, and constraints.

4. The fuzzy lexicon

Probabilistically orderly variability is also evident in the lexicon. The lexicon is commonly conceived as the repository of lexically arbitrary information – the phonological content and semantic significance of words and morphemes, and information about derivational and inflectional class membership. In the traditional generative conception, the phonological representation of the word or morpheme is cast in terms of a unique string of phonemes, which serves as the underlying representation for any occurrence of the word in an actual utterance. When variant realizations of morphemes are found that cannot be reduced to a single underlying representation, allomorphs – alternative representations – may be recognized, but categorical theories define these as bound by categorical constraints (e.g., the *-en* participial suffix in English occurs only with a specific subset of verbs: *taken, frozen, eaten, fallen*, etc.). In this view, the lexical entry has no place for probabilistic information such as frequency of occurrence, or likelihood of variant realizations of segments or inflections in various contexts.

Quantitative studies of variation, however, provide substantial evidence of probabilistic lexical properties, consistent with a non-deterministic, 'fuzzy' model of representation. We will consider two kinds of such evidence: probabilistic representations of lexical frequency, and probabilistic indications of lexical exceptionality. We also discuss usage-based or exemplar theories of the lexicon which directly incorporate information about realizations, frequencies, and contexts.

4.1 Lexical frequency

Lexical frequency effects in variation have been much studied in recent decades. A number of variable processes have been shown to be systematically conditioned by the frequency of occurrence of a lexical item. One example is the *–ing*/*–in* alternation in English. The occurrence of the *–in* alternant is positively correlated with the frequency of the root with which it occurs. Figure 1 shows the results of a study by Laturnus, de Vilchez, Chaves, and Guy (2016).

© 2018. John Benjamins Publishing Company
All rights reserved

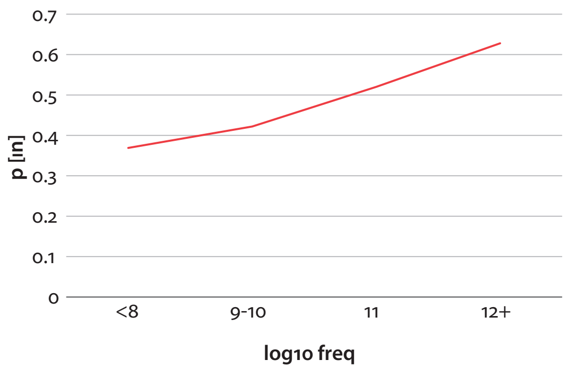

Figure 1. [ɪn] by lexical frequency in American English (from Laturnus et al. 2016)

Similar facts obtain in the case of English coronal stop deletion. Figure 2 is taken from Guy, Hay, and Walker (2008), showing the positive correlation between lexical frequency and deletion.

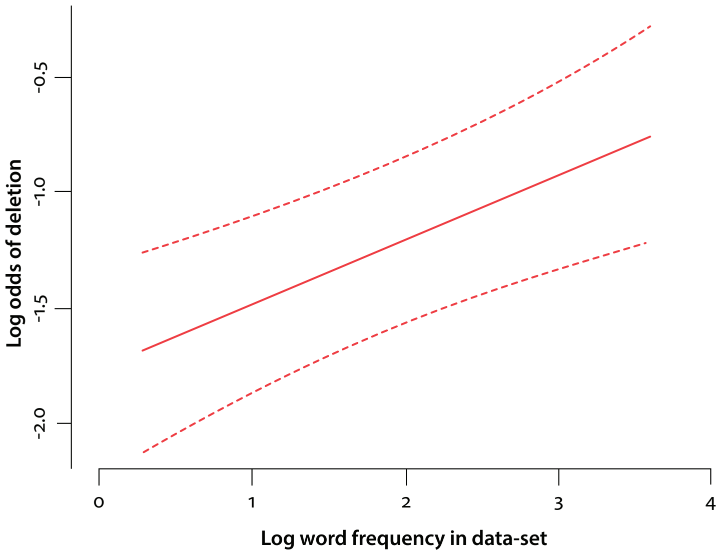

Figure 2. Coronal stop deletion increases with lexical frequency (from Guy et al., 2008)

Such findings imply that the lexicon contains information about lexical frequency, which is available to condition selection of variants or interaction with phonological processes. The quantitative evidence also illuminates the question of what lexical elements are associated with this frequency information – roots or fully derived forms?

© 2018. John Benjamins Publishing Company
All rights reserved

The frequency constraint on coronal stop deletion illustrated in Figure 2 happens to interact with the morphological constraint shown in Table 2. The nature of this interaction is that the frequency effect – increasing deletion with increasing frequency – is apparent only for uninflected root forms, whereas the regular past tense forms are unaffected by lexical frequency. Figure 3, from Myers and Guy (1997), illustrates this point. The implication is that frequency information is associated only with roots, which are stored in the lexicon, and not with regularly derived forms.[2] This is consistent with Pinker's (1999) 'words and rules' model of the lexicon.

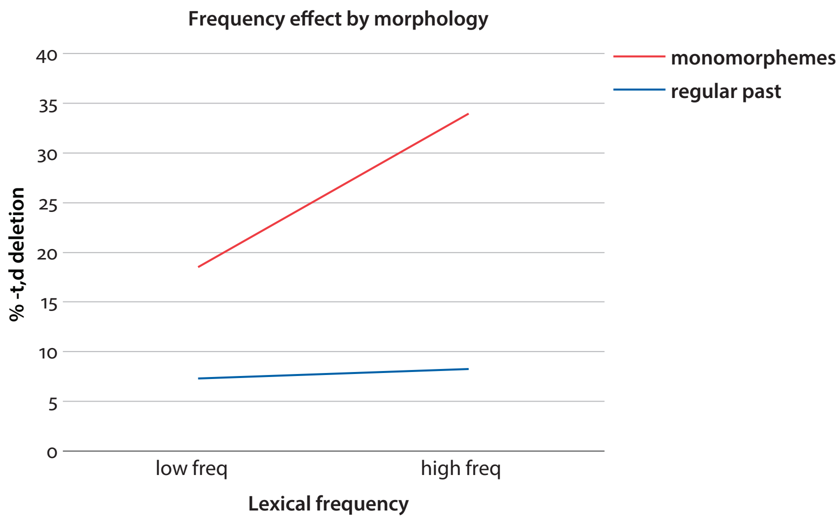

Figure 3. Coronal stop deletion: frequency interacts with morphological structure

What models of lexical representation can account for these facts? There are two contenders that appear to achieve at least observational and descriptive adequacy. One is the usage-based phonology of Joan Bybee (2001, *inter alia*), in which probabilistic properties of the lexicon are a central focus. In this approach, the lexicon is postulated to include remembered exemplars of potentially every utterance of every word the speaker has ever produced or perceived, tagged with information about the linguistic and social contexts in which the forms occurred. Hence, memory alone provides complete information about the quantitative distribution of words and variable processes, including lexical frequency, and idiosyncrasy. Bybee proposes a gradual process of reduction and assimilation through repetition to explain positive correlations with frequency such as those demonstrated above.

2. Irregular past tense forms (*left, kept*, etc.) also show increased deletion with lexical frequency, suggesting they are also stored as lexical entries, as Pinker argues.

© 2018. John Benjamins Publishing Company
All rights reserved

A grammar that operated on these lines can accurately generate many aspects of variation and change in considerable quantitative detail, and in addition to capturing frequency effects, it is capable of accounting for entrenchment of morphologically irregular forms, lenition and assimilation processes, and even socio-stylistic variation (because exemplars are tagged with social information). This is all accomplished by incorporating probabilistic information directly into linguistic representations. However, these virtues are achieved at a considerable modeling cost: the theory requires an essentially unlimited number of predictor variables (each exemplar), and implies great demands on the processing and storage capacity of the language faculty. The theory also appears to over-predict the effect of lexical frequency and the presence of lexical idiosyncrasy (every word has its own distinctive exemplar cloud), and to under-predict productivity, as in the production and perception of novel items for which the speaker has no exemplars.

The second alternative is the 'variable rule' model first articulated in Cedergren and Sankoff (1974). As we have noted, this paper makes the case that grammar is itself internally probabilistic. The authors also lay out a formalism to model such probabilistic grammatical operations. They adapt a conventional generative model by allowing rules, processes, and constraints to be associated with probability weightings.

This model straightforwardly permits frequency effects and other lexically specific constraints: lexical items that are involved in variable processes may be associated with individual probabilities of undergoing those processes. As with an exemplar model, these probabilities will be updated based on experience, but the variable rule model requires saving only the probability value in memory, not the entire exemplar cloud. It is thus more cognitively economical. The variable rule model also avoids the prediction issues mentioned above. Productivity is a central design feature, as in any generative grammar. Lexical idiosyncrasies and frequency effects are not intrinsic to the model. Constraint values are generalized across the entire lexicon, and will prevail wherever significant lexically specific quantitative patterns are absent, such as in most lower-frequency words (cf., Erker & Guy, 2012).

4.2 Lexical exceptions

A conventional assumption about the nature of grammar is that phonological processes apply uniformly across the lexicon, to all words that satisfy the relevant structural description. This is the synchronic version of the Neogrammarian hypothesis of exceptionless sound change. Nevertheless, it has long been recognized that some words behave idiosyncratically. This creates a problem for theoretical models that provide no internal mechanism for lexical exceptions to sound change

© 2018. John Benjamins Publishing Company
All rights reserved

or phonological alternations; it also presents a challenge for the language learner. How are lexical exceptions represented in the grammar?

In studies of variation, it is not uncommon to discover that certain words systematically fail to conform to the general distribution or rates of occurrence of variants that prevail for most of the lexicon. Such words are quantitatively exceptional. A case in point is the word *and* in English, which is uttered without its final /d/ at an extremely high rate, well in excess of phonologically comparable words like *hand, band*. This suggests that it is a quantitative exception to the process of CSD. In the Guy *et al.* study (2008) of early New Zealand English, *and* showed 80% final stop deletion vs. 29% deletion for all other words in the corpus.

How might such a pattern be represented in the grammar? For lexical frequency, we have just argued that high frequency lexical items are associated with individual probabilities of undergoing particular phonological processes. But this strategy is not adequate for the lexical exception cases, because of the way they interact with other phonological constraints. As shown in Table 5, following context ordinarily has a powerful effect on CSD: for the general lexicon, deletion is over five times as likely in pre-consonantal contexts than pre-vocalic. But this constraint is much weaker for *and*, where pre-consonantal contexts are associated with only a 17% increase in deletion compared with pre-vocalic contexts.

Table 5. *and* in New Zealand English: constraint effects are weaker for lexical exceptions

Following	**Other words**			***and***	
	N	**% del**		**N**	**% del**
Context:					
__C	1339	58.3		315	87.9
__V	1477	10.4		182	75.3
Range:		47.9	>		12.6

These facts are difficult to generate using a model in which *and* is simply associated with a higher probability of undergoing CSD than other words. If the following context is an independent predictor that applies equally to all words, then its effect should be constant, regardless of whether a given word is more favorable to deletion than most other words. A more straightforward model is that exceptional words have multiple lexical entries – allomorphs – in the lexicon. In addition to its full lexical entry /ænd/, English *and* must have an additional allomorph that underlyingly lacks a final stop: /æn/ or /n/. In usage, speakers vary stochastically, or perhaps stylistically, between these allomorphs. When the latter allomorph is selected, the surface realization lacks a final /d/ regardless of what the following context is. The surface corpus will therefore contain a mixture of forms, some of which were subject to the deletion process and affected by its constraints, and others for which

© 2018. John Benjamins Publishing Company
All rights reserved

deletion and its constraints were irrelevant. This produces the surface attenuation of the following context effect that we observe for exceptional *and*.

Similar cases of lexical exceptions have been studied in a number of variable processes in English, Spanish, and Portuguese. Thus, Hoffman (2004) finds that the discourse markers *entonces* 'so' and *pues* 'well' in Salvadoran Spanish have exceptionally high rates of final –s deletion, and also show weaker or nonsignificant effects of the constraints on deletion that are evident in the general vocabulary, such as syllabic stress and following context. Woods (2008) shows that the pronouns *I* and *my* are realized as monophthongal at an exceptionally high rate in Southern American English, and are weakly or insignificantly affected by constraints such as speech rate and following context. All such cases involve function words that undergo some phonological process at an exceptionally high rate. They are high frequency items, but frequency alone does not explain them; they differ significantly from other high frequency forms, notably by the attenuated effect of constraints on the process. The evidence suggests that such cases reflect lexicalization: the general phonological process has been encoded in the lexical entry as an allomorph. This would certainly be a reasonable conjecture for a language learner, faced with the kind of evidence we see in Table 5, and it is especially likely for function words, which are known diachronically to behave in exceptional ways.

5. Acquisition

Linguistic competence in a language must be acquired. Knowledge of a language is not an 'app' downloaded from a single source and hence uniform across all users. Rather, it is painstakingly constructed anew by each speaker, through observation, interaction, and use, by means of processes of association, deduction, inference, and generalization. So our questions are: how do these processes work in the face of variable input, and what kind of grammar do they yield? Are the cognitive skills that enable language acquisition tuned to respond to probabilistic input? Does normal engagement in language use lead the learner to a deterministic or a probabilistic grammar?

The data every child encounters will always display variability on a considerable scale. This presents the child with an inferential conundrum: does the variability reflect idiosyncrasy everywhere – words have lexically specific phonological, morphological, and syntactic properties, and speakers are unique and inconsistent? Or alternatively, are there generalizations to be made, and if so, what are the appropriate generalizations? In particular, is it even possible for learners to perceive and acquire quantitative patterns?

© 2018. John Benjamins Publishing Company
All rights reserved

Yang (2016) elaborates a 'variational model' of language acquisition that gives a probabilistic formulation to the learner's grammatical development:

> The child has a statistical distribution over the space of possible grammars … and it is this distribution that changes in response to linguistic data. As learning proceeds, the child will access the target grammar with increasing probability, while the non-target but linguistically possible grammars will still be used, albeit with decreasing probabilities. (Yang, 2016: p. 6)

In Yang's view, children derive generalizations in accord with a function that evaluates productivity and processing cost in the pursuit of an efficient grammar. Initially, every word or structure may be treated as idiosyncratic and simply memorized, but patterns that are sufficiently general, with exceptions that do not exceed a certain limit, will be formulated as rules, permitting a more compact representation and reducing the inventory of forms that require idiosyncratic, 'exceptional' treatments. They thus pursue generalizations, regular patterns, unmarked forms, and default settings, but do so quantitatively not categorically, tracking probabilities not asterisks.

With this work Yang makes an important contribution to recognizing probabilistic processes in the mind of the child learner in the course of language acquisition. However, in his model, the child ultimately seeks deterministic rules, albeit ones that have exceptions, in modest, memorizable numbers. Language learning is probabilistic, but for Yang, the grammars to be learned are not. We argue that the probabilistic cognitive processing Yang recognizes in acquisition is operative in the grammar itself – in the child's linguistic competence.

The learnability of probabilistic distributions is, as we have noted, well-established in psychological research. Linguistic studies of variation in early language acquisition demonstrate that children converge their productions of linguistic variables on the quantitative patterns that they hear in the input. For example, Smith, Durham, and Fortune (2007, 2009) showed children acquiring phonological and syntactic variables and their associated probabilistic constraints at a very early age – by 3;0 – and acquiring sociolinguistic constraints soon thereafter in ways that reflect caregiver and community behavior. But this convergence is limited by the developmental state of the child's mental grammar. One notable example occurs in the acquisition of the morphological constraint on English coronal stop deletion discussed above in Section 3. The pattern seen in Table 2 is replicated by children at a fairly early age for the monomorphemic and regular past tense words, with higher deletion in the former than in the latter. However, Guy and Boyd (1990) have shown that the intermediate deletion levels found for the irregular past tense forms are not acquired until late adolescence or young adulthood.

© 2018. John Benjamins Publishing Company
All rights reserved

Consequently, the productions of young children match their parents' deletion rates very closely for monomorphemes and regular past verbs, but not for irregulars in the *lost, kept, told* class. This is illustrated in Figure 4 from Roberts's (1994) study of children and parents in Philadelphia. Guy and Boyd argue that this represents a developmental stage in the child's grammatical analysis of verbal morphology. In the early grammar, these irregular forms are treated as unanalyzed wholes, suppletive past tense forms analogous to *were* and *thought*. Hence, for CSD, they are treated as monomorphemes. But with increasing linguistic maturity speakers come to a morphological analysis in which such words are derivationally related to their roots, so the final alveolar stop acquires morphemic status, and consequently, a lower rate of deletion.

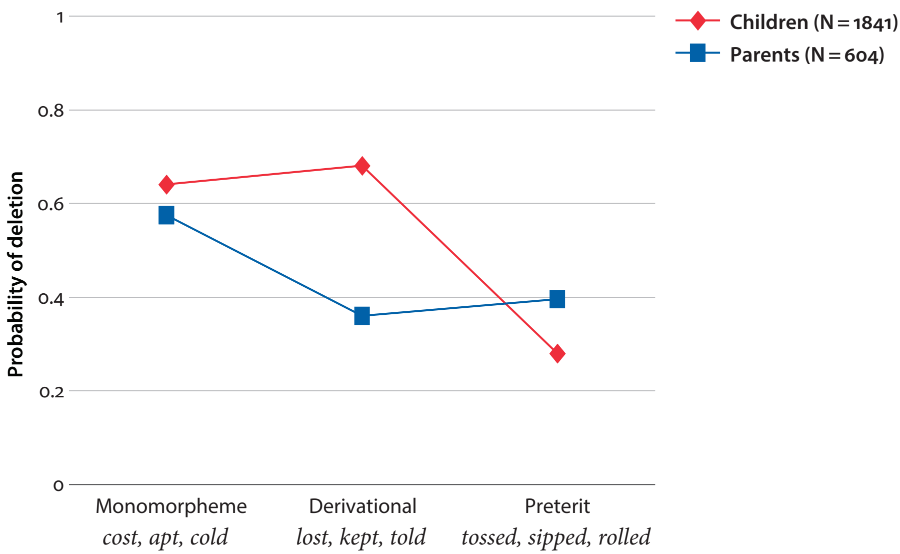

Figure 4. Probability matching of 16 children, 3–5 years old, Philadelphia. From Roberts (1994), Figure 7.4

These facts indicate that probabilistic information is intimately embedded in the grammar. Acquiring the grammar depends on attending to and acquiring the probability distributions of variants, and inducing a grammatical structure that allows the learner to reproduce them. The probability limits are 0 and 1, and these are the values that will be acquired for categorical phenomena, but learners attend to, and learn, phenomena with probabilities that lie between 0 and 1.

© 2018. John Benjamins Publishing Company
All rights reserved

6. Conclusions

Variation is manifestly present in language use. Since the grammars that linguists seek to describe are the mental objects that permit speakers to produce language, they must operate so as to generate variation. A model of grammar that fails to do so also fails to achieve the most elementary level of observational adequacy. So the task for linguistics is to identify the grammatical architecture that produces linguistic variability.

This task may be approached initially from the standpoint of the language learner. Faced with great diversity in the input, the learner observes quantitative patterns and formulates statistical inferences about the shape of the grammar that can generate those patterns. The grammar is constrained: it has to be learnable, it has to be productive – capable of perceiving and generating novel utterances, and it has to be human – i.e., consistent with human cognitive capacities and with the possibilities permitted by the language faculty or Universal Grammar. The easiest patterns to learn are those that are maximally or minimally likely – occurring with probabilities of 1 or 0. But the processes of learning, and the need to arrive at a grammar that achieves convergence with the input, require the learner to attend to the whole observable range of probabilities. Indeed, all the aspects of language incorporate probabilistic properties: in the input, in perception, in inference and acquisition, in the form of the grammar and lexicon, and in production.

The construction of linguistic theories necessarily parallels what the language user does. Linguists have always encountered variable, messy data, and always sought patterns, rules, generalizations. Where theory has gone wrong is when it privileges the regularities to the neglect of the messiness. We cannot discover adequate theories of language if we ignore quantitative patterns in the input, restrict the internal workings of the grammar to categorical, deterministic processes, and generate only categorical patterns in the output. Rather, we must continue to emulate the children, by embracing linguistic variation, and incorporating it into our theories.

References

Anttila, A. (1997). Variation in Finnish phonology and morphology (Doctoral dissertation, Stanford University).

Boersma, P., & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry*, 32, 45–86. https://doi.org/10.1162/002438901554586

Braga, M. L. (1977). A concordância de número no sintagma nominal no triângulo mineiro (Master's thesis). Pontifícia Universidade Católica do Rio de Janeiro.

© 2018. John Benjamins Publishing Company
All rights reserved

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, H. (2005). Predicting the dative alternation. In G. Bouma, I. Krämer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–94). Amsterdam: Royal Netherlands Academy of Science.

Bresnan, J., & Ford, M. (2010). Predicting syntax: processing dative constructions in American and Australian varieties of English. *Language*, 86, 168–213. https://doi.org/10.1353/lan.0.0189

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511612886

Cedergren, H., & Sankoff, D. (1974). Variable rules: performance as a statistical reflection of competence. *Language*, 50, 333–355. https://doi.org/10.2307/412441

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

Coelho, I. L. (2000). A ordem V DP em construções monoargumentais: uma restrição sintático-semântica (Doctoral dissertation, Universidade Federal de Santa Catarina, Brazil).

Fruehwald, J. (2012). Redevelopment of a morphological class. *Pennsylvania Working Papers in Linguistics*, 18(1), 77–86.

Erker, D., & Guy, G. R. (2012). The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language*, 88, 526–557. https://doi.org/10.1353/lan.2012.0050

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, 57(2), 94. https://doi.org/10.1037/h0058559

Gallistel, C. R., Krishan, M., Liu, Y., & Miller, R. (2014). The perception of probability. *Psychological Review*, 121(1), 96–123. https://doi.org/10.1037/a0035232

Guy, G. R. (1980). Variation in the group and the individual: The case of final stop deletion. In W. Labov (Ed.), *Locating language in time and space* (pp. 1–36). New York, NY: Academic Press.

Guy, G. R. (1981). Linguistic variation in Brazilian Portuguese: Aspects of the phonology, syntax, and language history (Doctoral dissertation, University of Pennsylvania). Ann Arbor, MI: University Microfilms International.

Guy, G. R. (1991). Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change*, 3, 1–22. https://doi.org/10.1017/S0954394500000429

Guy, G. R. (1996). Form and function in linguistic variation. In G. Guy, C. Feagin, J. Baugh, & D. Schiffrin (Eds.), *Towards a social science of language: Papers in honour of William Labov* (Vol. 1, pp. 221–252). Amsterdam & Philadelphia, PA: John Benjamins. https://doi.org/10.1075/cilt.127.15guy

Guy, G. R., & Boberg, C. (1997). Inherent variability and the obligatory contour principle. *Language Variation and Change*, 9, 149–164. https://doi.org/10.1017/S095439450000185X

Guy, G. R., & Boyd, S. (1990). The development of a morphological class. *Language Variation and Change*, 2, 1–18. https://doi.org/10.1017/S0954394500000235

Guy, G. R., Hay, J., & Walker, A. (2008). Phonological, lexical, and frequency factors in coronal stop deletion in Early New Zealand English. Paper presented at LabPhon 11, Victoria University of Wellington, Wellington.

Hoffman, M. (2004). Sounding Salvadorean, phonological variables in the Spanish of Salvadorean youth in Toronto (Doctoral dissertation, University of Toronto).

Jankowsky, K. (1972). *The neogrammarians: A re-evaluation of their place in the development of linguistic science*. The Hague: Mouton.

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

© 2018. John Benjamins Publishing Company
All rights reserved

Kroch, A. (1994). Morphosyntactic variation. In K. Beals, J. Denton, R. Knippen, L. Melnar, H. Suzuki, & E. Zeinfeld (Eds), *Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parasession on Variation and Linguistic Theory*. (Vol. 2, pp. 180–201).

Laturnus, R., de Vilchez, N., Chaves, R., & Guy, G. R. (2016). *Dialect, priming, and frequency effects on (-ING) variation in English*. Paper presented at 45th Annual Conference on New Ways of Analyzing Variation (NWAV-45), Simon Fraser University, Vancouver, Canada.

Leben, W. (1973). *Suprasegmental phonology* (Doctoral dissertation, MIT).

Myers, J. & Guy, G. R. (1997). Frequency effects in variable lexical phonology. In C. Boberg, M. Meyerhoff, & S. Strassel (Eds.), *A Selection of Papers from NWAV* 25 (University of Pennsylvania Working Papers in Linguistics 4:1), pp. 215–228.

Otheguy, R. (2008). Affirming differences, valuing variation, and dismissing dialects in modern linguistics. *Studies in Hispanic and Lusophone Linguistics*, 1, 223–233. https://doi.org/10.1515/shll-2008-1013

Oushiro, L. (2014). *Identidade na pluralidade: Avaliação, produção e percepção linguística na cidade de São Paulo*. Doctoral dissertation, Universidade de São Paulo, Departamento de Linguística.

Pinker, S. (1999). *Words and rules*. New York, NY: Harper Perennial.

Roberts, J. (1994). Acquisition of variable rules: (-t, d) deletion and (ing) production in preschool children (Doctoral dissertation, University of Pennsylvania).

Santa Ana, O. (1992). Chicano English evidence for the exponential hypothesis: A variable rule pervades lexical phonology. *Language Variation and Change*, 4, 275–288. https://doi.org/10.1017/S0954394500000818

Scherre, M. M. P. (1988). Reanálise da concordância nominal em português (Doctoral dissertation, Universidade Federal do Rio de Janeiro).

Smith, J., Durham, M., & Fortune, L. (2007). "Mam, my trousers is fa'in doon!": Community, caregiver and child in the acquisition of a Scottish dialect. *Language Variation and Change*, 19, 63–99. https://doi.org/10.1017/S0954394507070044

Smith, J., Durham, M., & Fortune, L. (2009). Universal and dialect-specific pathways of acquisition: Caregivers, children, and t/d deletion. *Language Variation and Change* 21:69–95. https://doi.org/10.1017/S0954394509000039

Verner, K. A. (1877). Eine Ausnahme der ersten Lautverschiebung. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen*, 23, 97–130.

Weinreich, U., Labov, W. & Herzog, M. (1968). Empirical foundations for a theory of language change. In W. P. Lehmann, & Y. Malkiel (eds.), *Directions for Historical Linguistics* (pp. 95–195). Austin: University of Texas Press.

Woods, L. (2008). Monopthongization of /ay/ in Southern American English (Doctoral dissertation, New York University).

Yang, C. (2016). *The price of linguistic productivity; How children learn to break the rules of language*. Cambridge, MA: MIT Press.

Yip, M. (1988). The obligatory contour principle and phonological rules: A loss of identity. *Linguistic Inquiry*, 19, 65–100.

Zuraw, K., & Hayes, B. (2017). Intersecting constraint families: An argument for harmonic grammar. *Language*, 93, 497–548. https://doi.org/10.1353/lan.2017.0035

© 2018. John Benjamins Publishing Company
All rights reserved