

The grammar of use and the use of grammar

Gregory R. Guy (New York University)

Abstract: Language is a uniquely social phenomenon – it is acquired exclusively through interaction with other users. The use of language is further characterized by orderly heterogeneity: regular patterns of social and linguistic conditioning. Since a speaker's knowledge of language derives from and includes knowledge of usage, it incorporates knowledge about variability, frequency and social significance.

An adequate linguistic theory must account for this social, interactive nature of language. Dichotomous oppositions between system and usage, competence and performance, i-language and e-language, etc., are misleading, if they incline us to favor puristic concepts of the mental grammar as an idealized, invariant, categorical system. But an entirely usage-based model, devoid of abstract operations, is also inadequate. Rather, the empirical evidence indicates that speakers do construct mental representations, abstractions, and operations to guide their production, but these are probabilistic and variable, rather than deterministic and discrete. From probabilistically patterned input, speakers infer inherently variable grammars, from which they generate productions that display orderly diversity.

1. Language as a social phenomenon

Language is a social phenomenon, seated in society, and acquired by the individual only through social interaction – i.e., usage. This means that the primary information people possess about their languages is information about how they are used by in communities by speakers. Such information about social usage provides the base for learning language, and for using it productively and competently; ultimately this is what we KNOW when we know a language. Consequently, knowledge of language necessarily incorporates some level of familiarity with all the linguistic diversity, complexity, and variability that we encounter in the world around us. Therefore I will argue that the linguistic systems we construct during the course of language acquisition and use are necessarily formed by, sensitive to, and generative of, variability and use. This has the consequence that the traditional distinctions in linguistics that oppose system vs. usage, competence vs. performance, etc. are false dichotomies – false in the sense that they are not very helpful in constructing coherent and adequate theories of language. This does not mean, however, that there is only usage, or that there is no linguistic system or grammar; rather, I conclude that linguistic systems – in the sense of abstract representations and processes – clearly do exist, but they emerge from use and experience, are constructed out of it and

dependent upon it. The empirical evidence will belie both kinds of theoretical purism – that which denies the relevance of usage and that which denies abstraction and system.

Recognizing the fundamentally social nature of language – its status as both the product and medium of social interaction – is essential to achieving an adequate account of language, and hence an adequate linguistic theory. If linguists conceptualize language as primarily or exclusively the property of the individual mind, we are bound to go wrong. Consider, first, the origins of language, both phylogenetically in the species, and ontogenetically in the individual. In the individual, it is clear that social interaction is the *sine qua non* of language development: without it, there is no language. This is evident in several ways. Humans who have no social interaction as children, who are not part of a speech community, do not develop any language at all. Child language development occurs exclusively through social interaction, so that the fortunately rare cases of isolated individuals – so-called ‘wild’ children, and children isolated due to abuse – do not develop language (cf. Fromkin 1974). But given a human community and social interaction, SOME communicative system will emerge, even if no prior language is available, or if there are other hindrances to communication. Thus we have the emergence of contact languages, pidgins, home signing systems, and the like, appearing in situations where people interact but cannot do so on the basis of normal language acquisition. On this evidence we may further conclude that the phylogenetic situation was the same: the *raison d’être*, the ultimate motivation for language evolution in the first place, was social interaction and communication, not as some would have it, as an aid to the solitary activity of thought. The evolutionary competitive advantage of *homo loquens* ‘the talking hominid’ was social communication (Mufwene 2011). The appropriate conclusion is that societies have languages, individuals do not. Indeed, specific ‘languages’ such as English, Cantonese, and Kikuyu only persist and continue to have regular patterns and continuities because they are the ongoing means of communication in complete human societies. When a language ceases to be used by a speech community, it dies. But when an individual dies, the languages he or she spoke survive, so long as they are used by a community. Thus the structures and system of language do not exist in isolation from human interaction, from usage.

Usage and interaction therefore provide both the data and the motivation for language acquisition: for developing the mental capacity to speak ourselves. And let us be clear, usage provides not just some of the data, but ALL of the data: we have in fact no other kinds of evidence about how our languages work or how to be a speaker except

what we hear and perceive from those around us.¹ This is true of the child language learner, and it is also true of the linguist; even a linguist's intuition about grammaticality is a PRODUCT of the system. The idea that intuition and introspection give us a usage-free pathway to inspect the inner workings of language directly is misguided. But given that our information comes only from usage, the next questions for the linguist are, what do we do with that information, and what does the mental capacity to be a speaker consist of?

These are the questions that point us to the 'system' part of the title of this volume. But we must be cautious in how we use the terminological distinction between 'system' and 'usage', lest we fall into the essentialist trap of believing that two different labels must refer to two essentially distinct things – that the 'system' is something separate from all the evidence obtained from usage. I take this pair of terms to be a contemporary update of the familiar dichotomy that has beguiled and bedeviled linguistic theory since Saussure famously distinguished *langue* from *parole*. In my view, the grand Swiss scholar, widely considered to be the father of modern linguistics, did the discipline a disservice with his love of dichotomies, particularly those opposing synchronic and diachronic linguistics, and *langue* and *parole*. Linguistics has been mesmerized ever since by the seductive metaphorical opposition between a system and its products, subsequently restated as competence vs. performance, I-language vs. E-language, grammar vs. usage.

Before analyzing the substance of this dichotomy, let us consider how these concepts have worked in terms of the sociology of the field – what have they meant for the people and practices in linguistics? From this perspective, I believe the dichotomy has flourished for two reasons. First, it's a simplifying assumption: it designates all variability, heterogeneity, idiosyncrasy, and other messy stuff as belonging to something other than 'real' language, and allows us to set it aside while we figure out the general patterns. This is a typical step in the early stages of a scientific field: it is easier to work out generalizations and models if we can start by ignoring some of the complexity of reality. Thus a simple model of mechanical motion might start by ignoring friction, and a simple model of gravity might start by ignoring relativity. These were in fact the way those theories developed in physics, so we can surely forgive our predecessors in linguistics for doing the same thing when they postulated a categorical, homogeneous, abstract mental grammar, ignoring diversity in society and variability in the individual.

¹ I neglect here the still-debated issue of innateness; even if individuals possess an innate mental faculty that aids in language acquisition, it does not aid them in acquiring any particular language in the absence of linguistic interaction with other speakers.

But when a science achieves greater maturity and self-confidence, it needs to revisit the simplifications and incorporate the inconvenient facts previously ignored, if it is to continue to progress. I think linguistics is clearly at that point: we need to marry our models of language with linguistic reality – focusing on questions like: how does language work, how can we communicate and understand each other by means of speech, and what do we do when we are doing that?

However, it is obvious that not all linguists agree on this point. There is a tendency in the field to reify these simplifying assumptions as if they were a fundamental truth about the nature of reality; this approach leads to Chomsky's surprising position (hardly altered in the last half-century) that "observed use of language ... surely cannot constitute the actual subject matter of linguistics, if this is to be a serious discipline" (*Aspects of the Theory of Syntax*, 1965, p.4). This remarkable, but once widely accepted position deserves serious reflection. Why would linguists adopt such a stance towards the very substance of language? To me, it seems to contradict reality, basic scientific empiricism, and even common sense, to assert that a serious science cannot be based on the study of what people do when they communicate with language. So why has such a position been so influential in our discipline? One possible answer that merits consideration is that the Chomskyan position – opposing competence/system to performance/usage while simultaneously defining the study of competence as the 'serious' science – is a self-justifying ideology in the service of the interests of a particular school and a particular methodology. It licenses the practitioner to ignore messy data from production, and avoids tiresome empirical testing of one's theories. It saves one the trouble of doing fieldwork, and allows one to work while introspecting comfortably in an armchair. Even better, it exalts theory constructors over data analyzers and fieldworkers. This is all very attractive – Chomsky offers linguists a way to make our problems more tractable, to make our working conditions more pleasant, and to feel superior, all at once. Of course, even within generative syntax there has been increasing discomfort with the shakiness of badly gathered introspection (cf. Schütze 1996), but I-language, not E-language, continues to constitute the fundamental focus of Chomskyan linguistics.

The problem that constantly confronts such an approach is that the diversity of linguistic reality keeps undermining the dichotomy. Looking at the world, it is clear that system and usage interpenetrate: within the linguistic system there is variability, reflecting patterns of usage, and within variable usage there is system and structure. There are social, usage-based constraints on the grammar, and systematic, grammar-based constraints on social usage. So at the present state of our knowledge, the dichotomy

has little explanatory or interpretive value. It doesn't even have much practical value, if it is now doing more to obstruct progress in the field than to facilitate it. It is time to abandon this conceptualization of our problems, and move to dealing with language as it is, not as we would imagine it to be.

2. Orderly heterogeneity: the systematic nature of usage

A more illuminating approach begins with the two observations enunciated by Weinreich, Labov and Herzog (1968) as elements of their "Empirical Foundations for a Theory of Language Change". These are "inherent variability" and "orderly heterogeneity". Orderly heterogeneity is the observation that variability in language is not random and arbitrary, but structured and systematic. It is linguistically structured, quantitatively constrained by the linguistic system. It is also socially structured: speakers in a community are not randomly different from one another; rather, linguistic diversity systematically reflects social organization: people's usage follows regular patterning by age, sex, class, ethnicity, linguistic experience, interlocutor, purpose, context, acts of identity, and so on. Our data about language, and our knowledge of it as users, look like the following examples.

In New York City, where I live and work, users of English hear the social patterning in the use of coda /r/ that Labov famously reported in 1966, reproduced here as Figure 1. The rates of /r/ production show simultaneous class and stylistic differentiation, such that higher status speakers use more /r/, and everybody uses more /r/ in their more careful or formal styles. Crucially, these patterns are regular, systematic, and pervasive, and they are regularly produced, perceived, and interpreted by NYC English speakers.

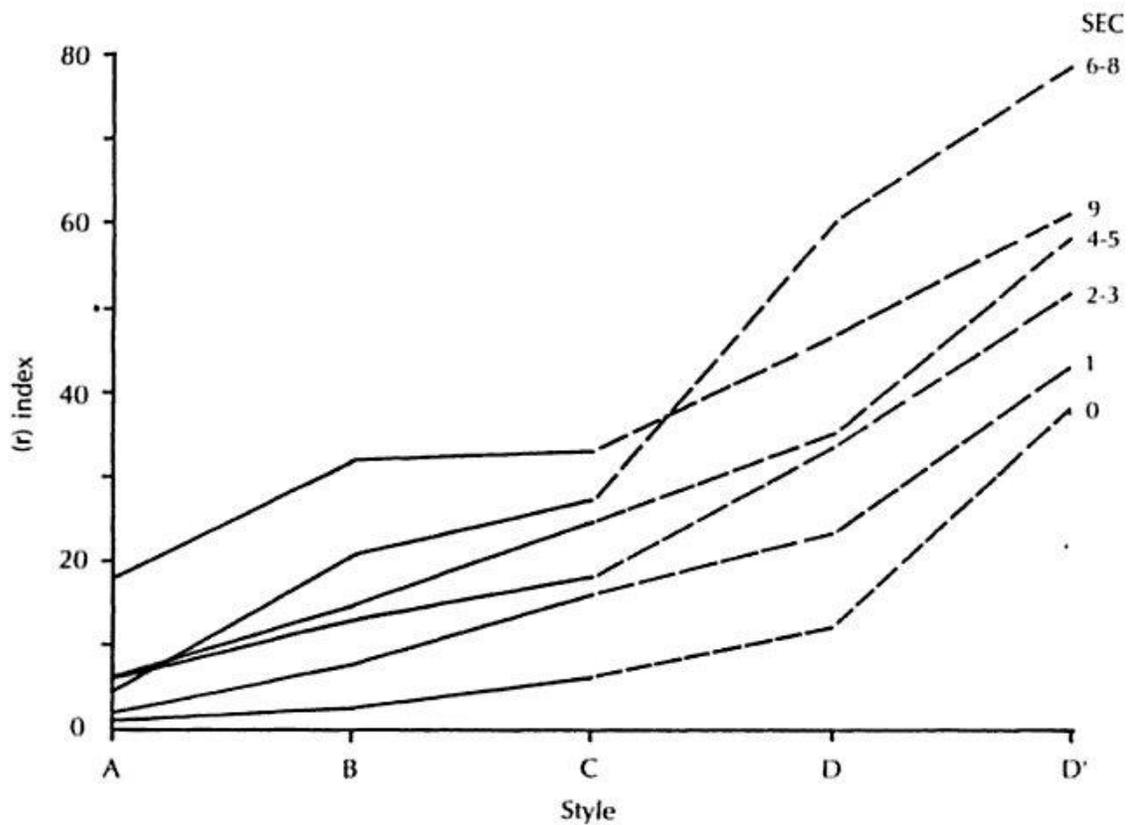


Figure 1. Coda /r/ production in New York City English: Stratification by socioeconomic class and speech style (from Labov 1966)

Another pattern of class differentiation is regularly encountered when something about the language is changing – specifically, when the change is a spontaneous innovation emerging within the speech community. In such situations we typically encounter a curvilinear distribution, with a peak in the lower middle class or upper working class, as illustrated in Figure 2 from Labov’s (1980) study of two ongoing vowel changes in Philadelphia English, the fronting of the nucleus of (aw) (e.g., *ounce*, *house*) and the raising of the nucleus of (ey) in closed syllables (e.g., *made*, *take*)

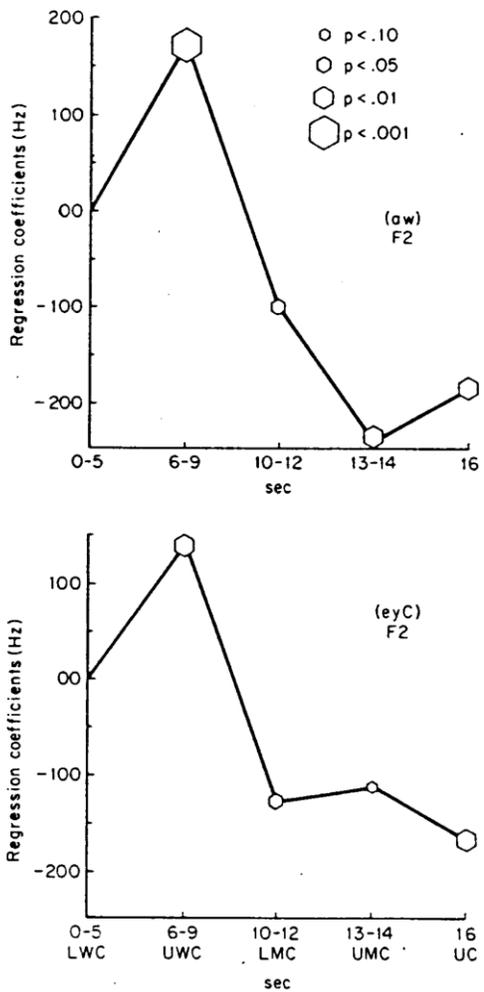


Figure 2. Curvilinear social stratification of two vowel changes in Philadelphia English (from Labov 1980: 261)

Another social dimension that is regularly reflected in orderly linguistic heterogeneity is gender. Thus women typically lead in language changes of several kinds, as illustrated in Figure 3, showing the class and gender distribution from Guy et al.'s (1986) study of an intonational change in progress in Australian English, namely the innovative use of a rising intonation in declarative clauses.

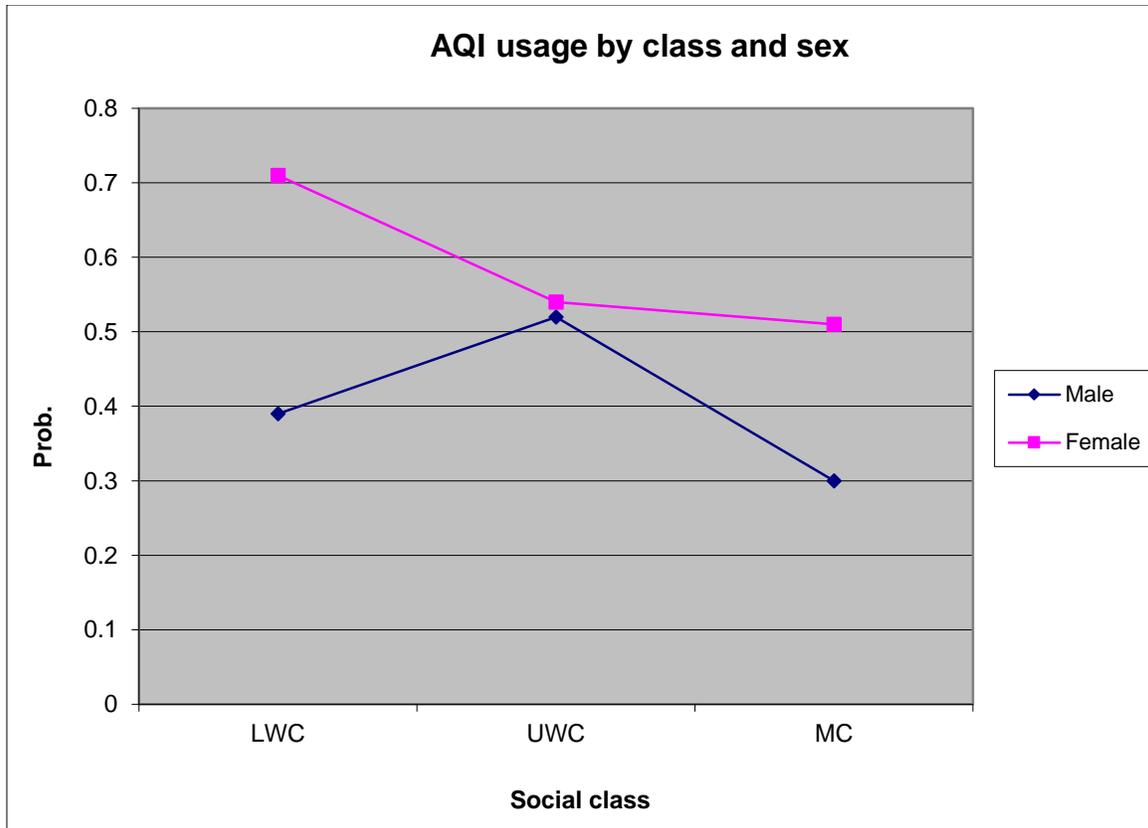


Figure 3. Gender and class distribution of an intonational change in progress in Australian English (data from Guy et al. 1986, p. 37)

Another regularly patterned feature of every human being's experience of language use is age differentiation, and this also turns out to be crucially associated with ongoing language change. The classic time course of language change is the s-shaped curve. This is illustrated in Figure 4, showing Chambers' (2002) data for the loss of /h/ before /w/ in words like *which*, *whine* in Canadian English. The age distributions are similar in four different regions (Montreal, Southern Ontario (GH), the Ottawa Valley and Quebec City). At a given moment in time, only 30-40% of the oldest speakers were using deletion, while the youngest people were approaching 100%.

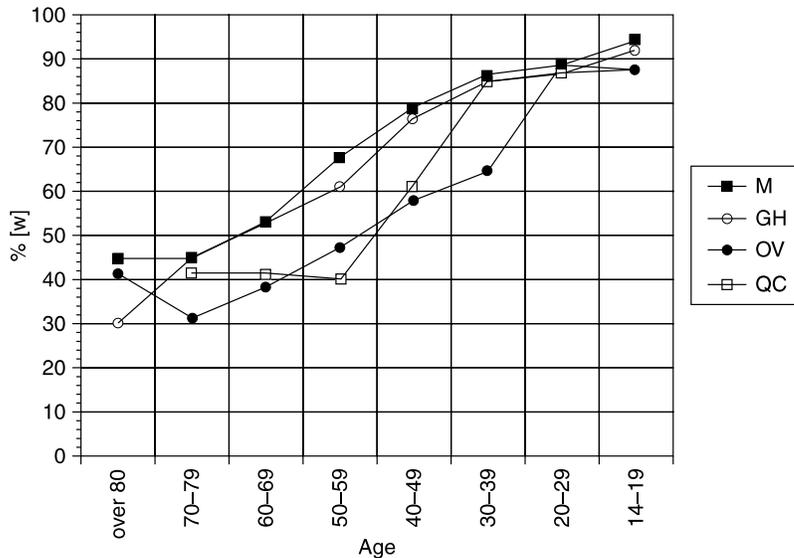


Figure 4. Loss of /h/ before /w/ in Canadian English (from Chambers 2002)

The generality of this pattern of age differentiation extends beyond segmental phonology. In Figure 5 the same s-shaped curve appears in Chamber and Heisler’s data on a morpholexical change, the substitution of *snuck* for historical *sneaked* as the past tense of *sneak*.

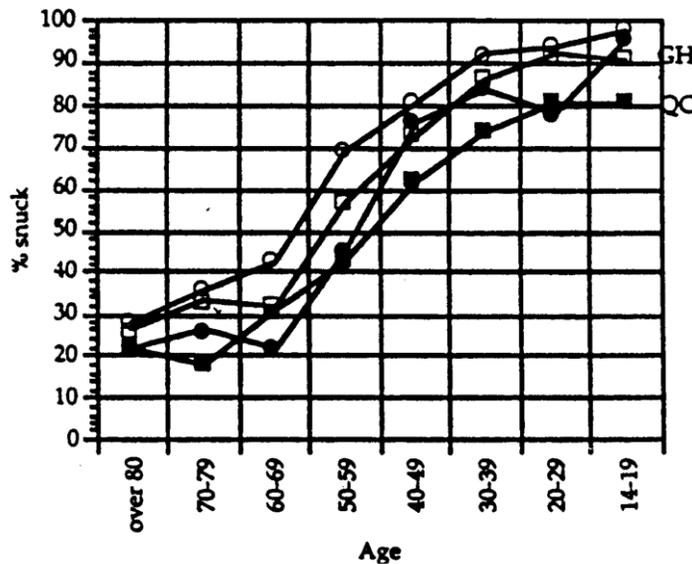


Figure 5. Replacement of *sneaked* by *snuck* in two dialects of Canadian English (from Chambers & Heisler 1999). (QC= Quebec City, GH=Golden Horseshoe, i.e. Southeastern Ontario)

Strikingly, the s-shaped curve of linguistic change reproduces across real time, as seen in Figure 6, from Kroch’s (1989) study of the syntactic change in English that introduced *do* as an auxiliary verb in questions and negations. The pervasiveness of such patterns

implies that during language change, speakers systematically encounter in usage information about the direction of the change, and engage with older forms and newer forms at the same time. Grandparents, parents, and children are all speaking differently, and since these generational differences are an intimate part of everyone's linguistic experience, speakers effectively know which way the change is heading – they can hear what is new and what is old in the voices of their own speech community. More generally, they are regularly exposed to systematic differences in frequency of occurrence of linguistic variables, and through the social distribution of these frequency patterns, are aware of the social significance of quantitative information. This must have implications for their construction and operation of their linguistic systems.

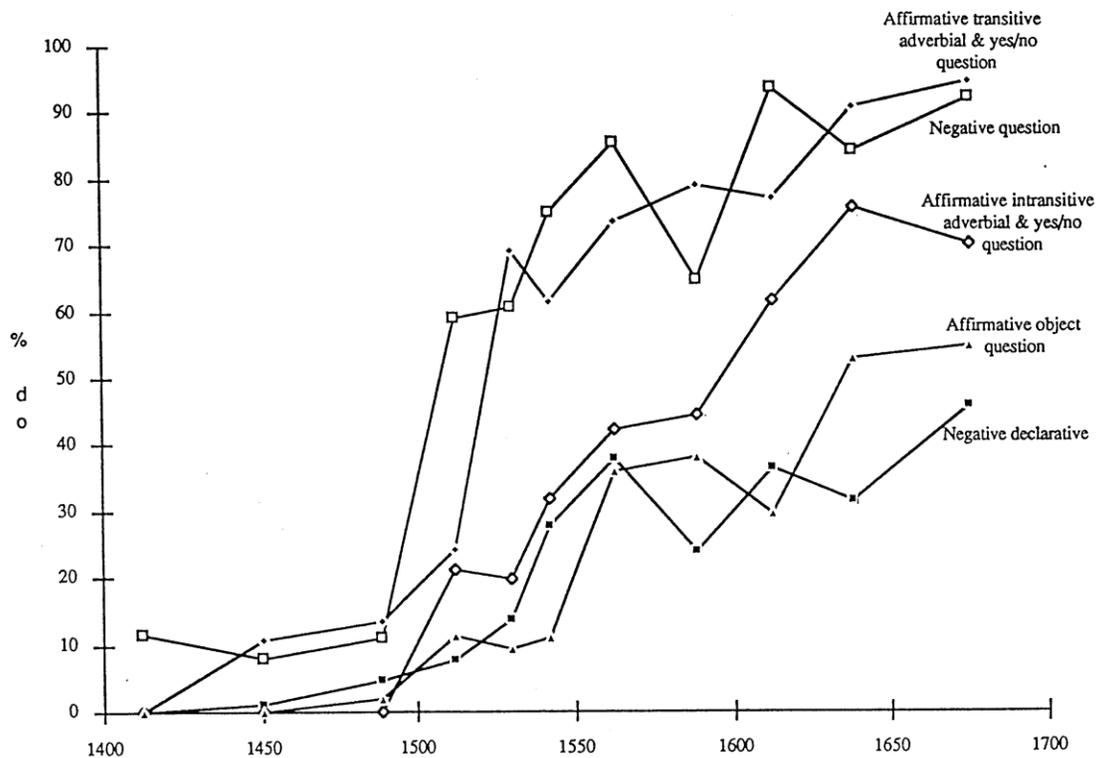


Figure 6. The rise of periphrastic *do* in English (Kroch 1989, p. 223, based on data from Ellegård 1953)

All the above examples illustrate the point that the diversity of language in use is not chaotic; rather it is ORDERLY, along social dimensions. But it is also orderly in the linguistic sense, that is, the linguistic constraints on variability systematically reflect the organizing principles of language. Note that Kroch's data in Figure 6 show systematic conditioning by syntactic construction that persists across three centuries. A further example comes from my research on Brazilian Portuguese (Guy 1981). Plural marking in

popular Brazilian Portuguese is highly variable. Unlike other dialects of Portuguese, and standard varieties of Spanish, Italian, etc., where number agreement is obligatory across the NP, as in (1), Popular Brazilian Portuguese (PBP) shows optional or variable plural marking, as in examples like (2).

(1) Standard Portuguese: categorical number agreement in NP

- a. *os velhos amigos* ‘the old friends’ (cf. sg. *o velho amigo*)
- b. *as casas brancas* ‘the white houses’ (cf. sg. *a casa branca*)

(2) Popular Brazilian Portuguese: variable number marking in NP

- a. *os amigo, os velho amigo, os velhos amigo*
- b. *as casa, as casa branca, as casas branca*

Instead of obligatory agreement, PBP shows high, nearly categorical rates of plural marking in the first position in an NP, with marked declines in subsequent positions, as shown in Table 1 (from Guy 1981).

Table 1. Plural marking in popular Brazilian Portuguese, by position in NP

<i>Position in NP:</i>	<i>% plural marked</i>	<i>N</i>
first	95	5247
second	28	3947
third	21	552
fourth and fifth	11	42

This is a regular, systematic constraint on plural marking in PBP, replicated in many studies of the phenomenon. Plural marking is variable, but not chaotic or random (nor idiolectally differentiated). The constraint is syntactically defined: it is not a function of word-class – first position words get plural markers whether they are heads (*amigos*), determiners (*os amigo*), adjectives (*velhos amigo, bons homem* ‘good men’), etc. This is highly systematic – it is extremely rare to hear a Brazilian speaker say something like *o amigos*, with initial zero followed by second position plural marker, or *as casa novas*, with first position marker, second position zero, and third position marked. Instead, it looks like a rule that marks the first word, followed by a low probability copying rule, from left to right: very orderly, but heterogeneous and probabilistic. This is grammar in usage, orderly heterogeneity. The simplistic dichotomy that postulates a categorical invariant grammar, and consigns frequencies and variability to usage, obscures this order.

3. Inherent variability: the diversity of the linguistic system

The cases presented above are but a small sample of an overwhelming set of empirical observations demonstrating that variability is a pervasive but orderly feature of language use. On the basis of such evidence, Weinreich, Labov and Herzog (1968) further argue that variability is ‘inherent’ in language – i.e., it permeates the linguistic system. The classic dichotomous approach, which distinguishes the linguistic system from usage, postulates that the object of study of linguistics is a homogeneous monostylistic idiolect, which is internally invariant (cf. Chomsky 1965). Alas this imaginary object does not exist in the world; indeed, no observation of language even approximates it. Every speech community includes diverse speakers, every individual commands different styles, every utterance includes variable elements. Speakers clearly perceive, process, produce, comprehend, and manipulate variability in all aspects of language. Thus variability is, in a word, an INHERENT feature of the linguistic system, and no adequate account of the linguistic system can fail to accommodate variability. Consequently, any grammar of a language, or theory of grammar, that fails to account for variability is inadequate on its face – it does not even reach Chomsky’s most elementary level of ‘observational’ adequacy. Furthermore, if variability is an intrinsic part of the grammar, then we lose one of the motivations for the grammar/usage distinction.

Now, elsewhere in this volume Newmeyer has objected that quantitative properties like those illustrated above are not ‘in’ the grammar, but lie outside it, in a ‘user’s manual’, or derive from interactions between the grammar and various grammar-external factors. Nevertheless, the evidence shows that this restatement of the system/usage dichotomy continues to be inadequate, as well as theoretically profligate rather than parsimonious. A cogent case in point is the deletion of final coronal stops in English, which shows an exponential relation among retention rates in three morphological categories.

The relevant categories reflect three derivational levels of words in English: underived or monomorphemic forms like *best*, *old*, which have the full consonant cluster in their dictionary entries, have the highest rate of deletion and lowest rates of retention. Irregular ‘semiweak’ verbs like *left*, *told* are derived early in the phonology, and have intermediate rates of retention. And finally, the highest rates of retention are found in regular past tense forms like *missed*, *rolled*, which are derived late in the phonology. Strikingly, the retention rates in the three categories are exponentially related. I have argued in previous work (Guy 1991, 1992) that this shows an iterated application of a single variable rule, with a constant ‘base’ probability of applying, which operates throughout the several stages of a derivation in a multilevel phonology. The highest retention rates are found in past tense forms that are exposed once to the deletion process

at the final level of the phonology; forms that are exposed at two levels have the square of the basic retention rate, and the underived forms that are exposed at three levels show the cube of the base rate. This relationship was demonstrated in my original work, and it has subsequently been confirmed in a large number of other studies. The data from three such studies appear in Table 2, which shows in the first three columns, the number of tokens analyzed, the percentage of those forms which had a retained /t/ or /d/, and the probability of retention predicted by my model, using the ‘best fit’ for the base rate of retention shown in the fourth column. Thus Santa Ana’s speakers appear to have an underlying base probability of retention of .75, predicting retention rates of 75% in regular pasts, the square of this value – 56.3% – in irregular pasts, and the cube of this value – 42.2% – in underived forms. These numbers fit very closely to the observed percentages of deletion (74.3%, 59.3%, and 42.1%). A chi-square test shows that the differences between the predictions of the model and the observations shows they are not significant ($p=.57$, where the criterion for significance is typically taken to be $p<.05$).

Table 2. The exponential relationship: coronal stop retention in three data sets

	N	Observed % retained	Predicted exponential progression	Model-fitting
Corpus 1 (Guy 1991, 7 speakers)				
Underived	658	61.9	.614 (n=3)	Best-fit $pr = .85$ Chi-square = 1.28, $p=.55$
Irregular Past	56	66.1	.723 (n=2)	
Regular Past	181	84.0	.850 (n=1)	
Corpus 2 (Santa Ana 1992, 45 speakers)				
Underived	3724	42.1	.422	Best-fit $pr = .75$ Chi-square = 1.17, $p=.57$
Irregular Past	297	59.3	.563	
Regular Past	836	74.3	.750	
Corpus 3 (Bayley 1994, 16 speakers)				
Underived	2065	43.3	.439	Best-fit $pr = .76$
Regular Past	541	76.2	.760	

Five further tests of the model by Labov and his students appear in Table 3. In no case is the exponential model statistically rejected (i.e., the predictions of the model are never significantly different, with $p<.05$, from the observed data), and in the studies that have compared it to alternative models, it fits as well or better than the alternatives.

Table 3. Five tests of the exponential relationship in coronal stop deletion, 1991–1997 [W. Labov, p.c.]

Year		Regular Past	Irregular Past	Underived Words	Best fit p_r	Chi-square, sig

		<i>missed, rolled</i>	<i>lost, told</i>	<i>best, old</i>		
1991	N ret/tot	79/100	29/53	221/539		
	p _r	.79	.73	.74	.74	.37, p>.85
1992	N ret/tot	93/116	32/64	250/583		
	p _r	.80	.71	.75	.75	.93, p>.65
1995	N ret/tot	323/404	149/229	496/922		
	p _r	.80	.81	.81	.81	.67, p>.75
1996	N ret/tot	85/96	62/82	219/374		
	p _r	.88	.87	.84	.84	.56, p>.75
1997	N ret/tot	209/258	71/90	491/906		
	p _r	.81	.89	.82	.82	1.99, p>.40

This is thus a robust and systematic quantitative feature of English phonology. It is eminently rule-governed; it is not sporadic or random; and it shows a highly specific mathematical relationship. English speakers do not simply delete final stops more in underived words and less when the stop represents an affix; rather, they delete these categories in a specific ratio. It is hard to imagine any other process than iterated application of a single probabilistic operation that can generate these numbers. For instance, it can't be adequately modeled by just assigning separate probabilities to the three categories. Functional and usage-based accounts that differentiate these classes by their functional load (e.g., avoid deleting tense markers because of their communicative content) fail to predict any specific quantitative relation among them. The only model that explains the exponential relationship is one in which a single operation (stochastic deletion) is recursively applied in the derivation of forms, with the mathematical result that the associated probability is multiplied by itself one, two, or three times. Recursion and derivation in language are ordinarily understood as grammatical operations. In this case, those operations are quantified.

Strong confirmation of this model is found in the way the process interacts with other constraints. Those that are external to the word, such as the favoring effect of a following consonant, are not multiplied during the derivation; rather, they apply only once in the postlexical phonology, after words are inserted into a phrase marker. Consequently they are constant in magnitude across the different derivational classes. However, internal constraints such as the effect of a preceding consonant are indeed present throughout a derivation; consequently they appear magnified in underived words,

which experience them repeatedly, compared to regular past tense forms, which experience their effects only once. These predictions are quantitatively confirmed, as shown in Table 4, which expresses the contextual effects on the process as partial probabilities of deletion occurring in a context (a context with a value of 1 implies categorical deletion, while a value of 0 implies categorical retention; intermediate values above .5 indicate deletion favoring contexts, and those below .5 indicate deletion-inhibiting contexts). The word-internal constraint shows a much larger range between favoring and disfavoring contexts in the underived words than in the regular past tense forms – the predicted amplification through iterated applications of the deletion process, whereas the word-external constraint is uniform in magnitude for both morphological classes.

Table 4. Internal vs. external constraints on coronal stop deletion: interaction with derivational class (Guy 1992) (partial probabilities from separate analyses of morphological classes)

4a. Internal constraint: Preceding segment effect on coronal stop deletion

- - - Morphological Class - - -		
<i>Preceding Segment</i>	<i>Underived</i>	<i>Regular Past</i>
Sibilants	.66	.67
Obstruents (stops, other fricatives)	.49	.46
Nasals	.59	.41
Liquids	.27	.44
Range:	.39	> .26

4b. External constraint: Following segment effect on coronal stop deletion

- - - Morphological Class - - -		
<i>Following Segment</i>	<i>Underived</i>	<i>Regular Past</i>
Consonants	.73	.65
Vowels	.31	.24
Pause	.45	.63
Range:	.42	= .41

What this implies is that variability and quantitative properties are found in the system, inside the grammar. And as we saw in the previous section, systematic, regular ‘grammatical’ properties are also found within the use of language. So the dichotomy that opposes system and usage, assigning invariant and categorical properties to the system/grammar, and variable and probabilistic properties to usage, is turning into an obstacle to explanation, rather than a facilitator.

4. Towards an integrated theory: grammar emerges from experience

What then are the elements of a more coherent vision that eschews the facile system vs. usage dichotomy in pursuit of a model of the fundamental unity of grammar and use? We can start where everyone starts, as a child encountering the language in use in the community around us. Usage constitutes our entire input. We have an intelligent mind, perhaps even endowed with specialized neural networks that facilitate language processing. But whether or not language is cerebrally special, we face the general problem of identifying units, collocations, and productive principles that will allow us not simply to reproduce the specific utterances that we have heard, but to form our own novel utterances that will be correctly interpreted by others. We must make our output well-formed, so we have to figure out what ‘well-formedness’ consists of.

Basically, we have to find PATTERNS. The patterns are the grammar, the system. Where do they come from? The usage-based perspective on this issue, associated with linguists like Bybee (2001, 2002) and Pierrehumbert (2001, 2006), argues that system is emergent, consisting of generalizations across observed usage. Let me present an example from my research with Daniel Erker on Spanish pro-drop (Erker & Guy 2012).

Spanish has optional use of subject personal pronouns (SPPs). They can occur overtly or be omitted, as in (3), where both full and omitted forms communicate the same meaning.

(3)	Overt subject personal pronoun	<i>Yo quiero</i>	‘I want’
	Omitted subject personal pronoun	<i>Quiero</i>	‘I want’

So how does a speaker know or decide where to use one or the other? Much previous research on this topic has turned up some systematic, widely general patterns of use that are governed by morphosyntactic and discursive structures of Spanish. For example, SPP occurrence is regularly constrained by verbal morphology, verb semantics, and discourse reference (cf., *inter alia*, Otheguy & Zentella 2012). The morphological constraint contrasts different tense/mood/aspect forms, with the result that TMA categories with more distinctive verbal inflection (e.g., the preterit, where every person/number category has a distinctive inflected form) are associated with lower probabilities of pronoun occurrence than those with less distinctive inflection (e.g., the imperfect, where first and third person singular forms are systematically identical). The verbal semantics constraint favors SPP use for verbs of mental activity, while those of external activity show less SPP occurrence. And the discourse level constraint considers the flow of reference in a text: a subject which makes reference to a different person than the subject of a preceding

sentence (switch reference) is more likely to be expressed by an overt pronoun. These patterns are confirmed in our data, as shown in Table 5.

Table 5. Three constraints on Spanish SPP occurrence (from Erker & Guy 2012, pp. 540-541)

<i>Tense-Mood-Aspect form of verb</i>	<i>N</i>	<i>% overt SPP</i>
Imperfect Indicative	708	43%
Present Indicative	2695	36%
Preterit Indicative	877	29%
F=9, p<.001		
<i>Semantic content of verb</i>		
Mental Activity	840	45%
Stative Verb	1438	36%
External Activity	2601	31%
F=27.8, p<.001		
<i>Switch Reference</i>		
Switch in reference from previous clause	2653	40%
No switch in reference	2233	29%
T=8.1, p<.001		

These effects are quite regular and systematic, recurring in many studies. They constitute valid generalizations about Spanish syntax. But when Erker and I looked at their distribution with respect to the lexical frequency of the verb, we found that these patterns are primarily associated with high frequency words. First consider TMA. Dividing the verbs into low frequency and high frequency forms, we find that the main effect of TMA is primarily a phenomenon of the frequent words. The differences among the TMA categories is modest (although significant) in infrequent verbs, but dramatically greater among high frequency words. This is graphically illustrated in Figure 7; in this and subsequent figures the diverging lines indicate stronger constraint effects in the high frequency words, and significant differences between contexts are indicated by red stars.

TMA and Frequency

% Pronouns Present

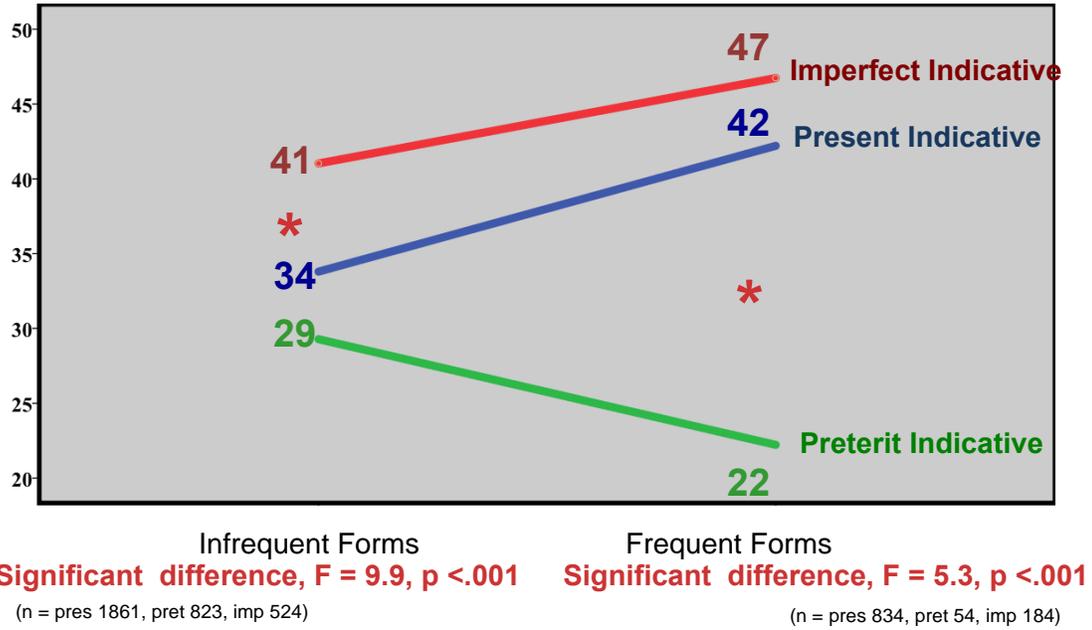
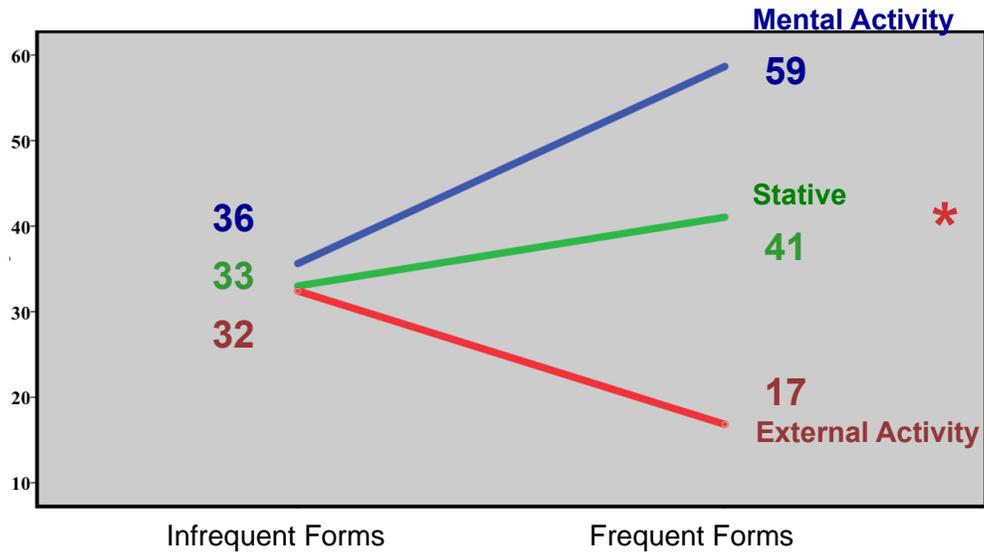


Figure 7. TMA and frequency in Spanish pro-drop (from Erker & Guy 2012, p. 544)

Similar results emerge for the other constraints on Spanish pro-drop. In the case of semantic content, the picture is even clearer: no significant differences among semantic categories in the infrequent forms, but among the frequent forms, the contrast is forcefully evident (Figure 8).

Semantic Content and Frequency

% Pronouns Present



No difference $F = .94, p = .38$ $(n = 494, 924, 2393)$ **Significant difference, $F = 51.2, p < .001$** $(n = 346, 514, 208)$

Figure 8. Semantic content and frequency in Spanish pro-drop (Erker & Guy 2012, p. 543)

And for switch reference (Figure 9), clause sequences involving a switch in reference favor more pronoun expression than those with no switch for both low and high frequency verbs, but again the effect is significantly greater for high frequency words.

Switch Reference and Frequency

% Pronouns Present

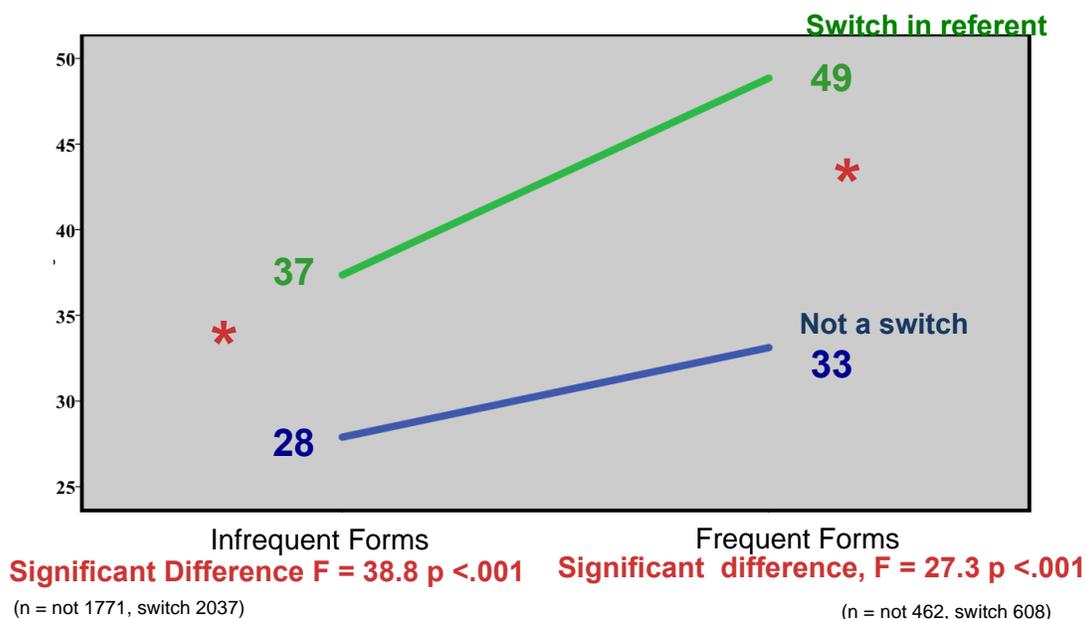


Figure 9. Switch reference and frequency in Spanish pro-drop (Erker & Guy 2012, p. 544)

The robust generalization that emerges from these results is that the various constraints on Spanish pro-drop that have been well documented in many previous studies are regularly much stronger for high frequency verbs. The standard interpretation of these constraints has been to suppose that a speaker's grammatical representation of verbal properties and discourse structure governs their probabilistic choice between using a null or overt pronoun. But these results indicate that the grammatical properties such as tense-mood form or semantic category are actually activated among the words that speakers most often encounter and use in speech: this is a classic characteristic of usage, not a property of grammar in the traditional sense.

Importantly, our data show that this pattern is consistent across two different nationality groups – Dominicans and Mexicans – that have substantially different overall rates of pro-drop. Dominican Spanish is well-known to use a much higher rate of overt SPPs than Mexican Spanish, but both dialects show the same pattern of constraint effects and interaction with lexical frequency, as is illustrated for semantic content in Figure 10.

Frequency and Semantic Content by nationality

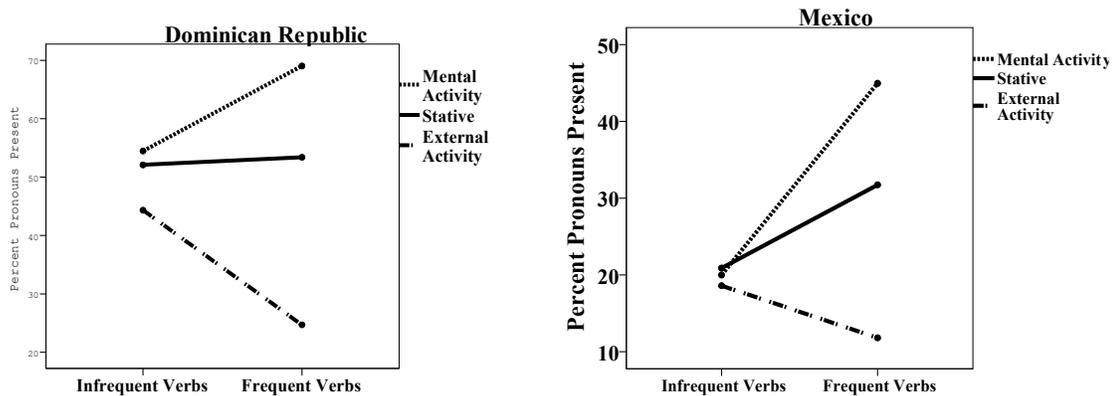


Figure 10. Frequency and semantic content effects on pro-drop in Dominican and Mexican Spanish (Erker & Guy 2012, p. 547)

Erker and Guy argue that these results show that the grammatical analysis, in terms of TMA classes, semantic classes, and even the discourse patterning, is emergent, rather than primary. Speakers presumably need some minimal level of experience with words and structures to begin to discern patterns and formulate hypotheses. Consequently they display robust patterns in high frequency forms, but weak to nonexistent patterns in the low frequency forms. In fact, the low frequency forms basically default to the dialect-specific overall rate of pronoun occurrence. But for verbs that are frequently encountered in usage, distinct generalizations begin to emerge.

5. Towards an integrated theory: grammar governs production

The evidence just presented suggests that speakers infer grammatical properties and 'rules' from experience, as shown by the fact that they do a 'better' – or at least more robust – job of inferring them about words that they hear and use more often. Grammar is thus emergent from and derivative of experience, rather than *a priori* or primary. But we should not leap from such evidence to the conclusion that there is no mental grammar at

all, that speakers simply replicate the quantitative data that they encounter in their linguistic input, without constructing any mental apparatus of abstract representations, patterns, and operations. On this point I think I part company with positions taken by Bybee (2001) and some other usage-based theorists: the evidence suggests that speakers do in fact construct mental grammars – abstract analyses, categories, and operations – to enable and govern their own productions.

An illustration of this point comes from my work with Sally Boyd (Guy & Boyd 1990) on the acquisition of the morphological constraint on English coronal stop deletion that was discussed above. There is a distinctive developmental pattern in the treatment of the irregular, semiweak past tense forms with respect to stop deletion: deletion rates in this category decline with age, as shown in Figure 11.

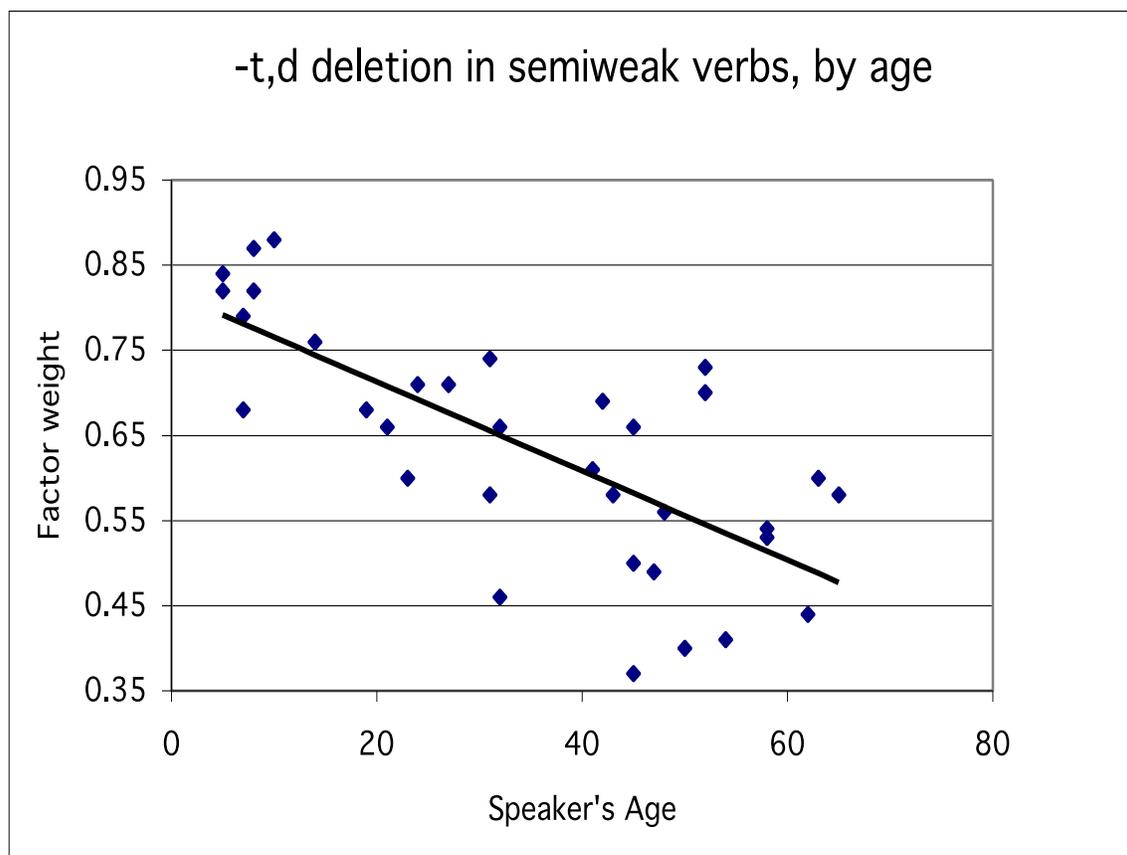


Figure 11. Age grading of English final stop deletion in irregular semiweak past tense verbs (from Guy & Boyd 1990, p.8)

Boyd and Guy interpret this pattern in terms of the mental representation of these categories. The youngest children (aged 4-5 in this study) evidently have just two form classes for English past tenses: strong and weak. The semiweak forms, having salient root

vowel changes, are perforce assigned to the strong class, with mental representations like *tell~toll*, *keep~kep*, so they have essentially categorical absence of the final stop. In our study this appears as very high rates of t,d deletion for the youngest children. In the next developmental stage, speakers set up distinct representations for these words that incorporate the final coronal stop, but treat these as suppletive allomorphs, not derived forms. At this stage, reached in adolescence for our speakers, these words are deleted at approximately the same rate as underived words. The lowest rate of deletion is not reached for most people until adult life, and represents the development of a derived representation in which the final -t,d is identified as an affix, and attached at the first level of the lexicon. This then generates the lowered deletion rate in these forms.

One consequence of this acquisitional pattern is that children do not simply reproduce the deletion rates that they hear from their parents in semiweak verbs, because they do not have the same mental representations. Although they are capable of matching parental input with extraordinary precision in those morphological categories where they share the same representations, in the semiweak verbs their deletion rates reflect their own mental grammar, not the productions they hear from their parents, as illustrated in Labov's figures for a family from a Philadelphia suburb shown in Figure 12. The seven-year old son converges closely on his parent's deletion rates for underived and regular past tense forms, but for the *tell-told*, *keep-kept* class, he is maximally distinct, using the high deletion rates associated with the grammar typical of his stage of language development.

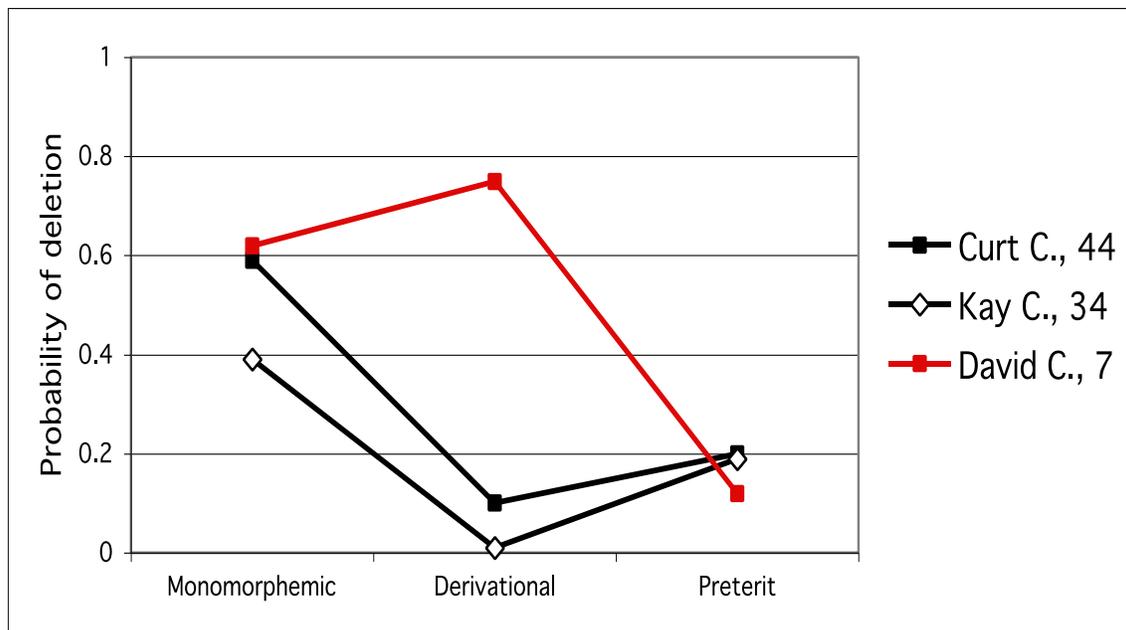


Figure 12. Generational differences in coronal stop deletion in one family (Labov, p.c.)

Consider the cognitive task confronted by the language learner who seeks to discern an optimal representation for this fragment of English morphophonology. They encounter massive evidence that English distinguishes present and past tense verb forms, and that it has two form classes – strong verbs with root vowel changes, and weak verbs with the coronal stop affix. But there are only about 14 lexical items that have **both** of these alternations, like *tell-told*, *leave-left*, *keep-kept*, etc. Hence in order to set up a distinct representation for such forms, the child must first pick the relevant words out of the crowd, recognize that they have variably occurring final stops, and then further recognize that the rate of stop deletion is subtly different in the holistic and derived representations. This takes appreciable amounts of time and data, so in childhood English speakers are unable to replicate the adult treatment of the semiweak verbs. Roberts (1993) study of Philadelphia children demonstrates this with large Ns, as shown in Figure 13.

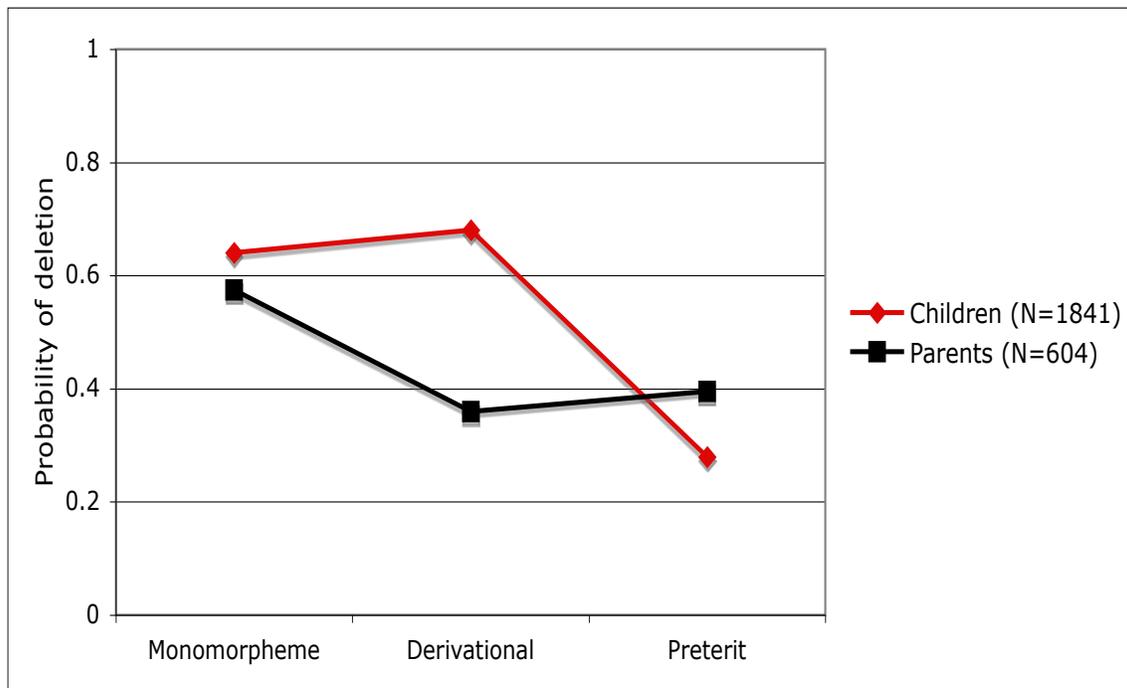


Figure 13. Generational differences in coronal stop deletion in a large corpus (Roberts 1993)

The appropriate conclusion is that, here again, the grammatical representation emerges from – or is inferred from – usage and experience, but developing this representation is a process that each individual must go through, and that may not even lead them to exactly the same analysis in the end. Nevertheless, it is the grammatical representation that the

speaker has inferred that governs their production, rather than a simple repetition of what they have heard.

6. Conclusion

The evidence drawn from the study of actual human language is incompatible with an idealized model that seeks to characterize linguistic systems in isolation from use; it is also incompatible with a model that seeks to characterize usage in isolation from any kind of abstract system. Hence I conclude that puristic models at both extremes of the theoretical spectrum on these issues are destined to fail: a puristic GENERATIVE model, which keeps a strict separation between grammar and usage, fails to give an adequate account of the interpenetration of structure, variability, and probabilistic properties in both grammar and usage, while a puristic USAGE-BASED model, which denies abstraction, fails to account for grammatically governed divergences between experience and production, like the results in Figures 12 and 13. An adequate model requires an integrated approach: usage supplies language acquirers with all of their data, including a vastly enriched fountain of information about social diversity, directions of change, and the orderly linguistic structure of inherent variability. From this input, they construct the set of inferences, representations, and operations that we call the grammar. Crucially, the grammar incorporates and encompasses variability and quantification, enabling speakers to do the fine quantitative tuning of their productions that is so fundamental to situating one's speech appropriately in the social universe, and conveying appropriate messages about interactive stance, speech style, and identity construction. The mental grammar thus mediates between experience and production: production does not derive directly from input, but is governed by the internalized grammar.

In conclusion, doing linguistics without usage is like doing zoology without animals – you can try it, but it yields a very limited theory that suffers from a gaping, self-inflicted wound, a theory which is liable one day to be trampled by elephants. But at the same time, studying usage without addressing systematicity and generalization is not doing linguistics at all: the language user does in fact construct patterns, generalizations, and processes, and uses them to speak. Linguistics must therefore ultimately reject the Saussurean dichotomy, and embrace all the faces that language presents to us: the usage, the systems that emerge from it, and the society that has created and maintained it. Rather than policing the hypothetical frontier between system and use, linguists can move on to more productive and interesting pursuits, investigating precisely how these facets of human linguistic experience interact.

References

- Bayley, Robert (1994): Consonant cluster reduction in Tejano English. *Language Variation and Change* 6: 303–326.
- Bybee, Joan (2001): *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, Joan (2002): Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14: 261–290.
- Chambers, Jack K. (2002): Patterns of variation including change. In: Jack K. Chambers, Peter Trudgill and Natalie Schilling-Estes (eds.), *The handbook of language variation and change*, 349–372. Malden, MA: Blackwell.
- Chambers, Jack K. and Troy Heisler (1999): Dialect topography of Quebec City English. *Canadian Journal of Linguistics* 44: 23–48.
- Chomsky, Noam (1965): *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
- Ellegård, Alvar (1953): *The Auxiliary Do: The Establishment and Regulation of its Use in English*. Gothenburg: Acta Universitatis Gothoburgensis.
- Erker, Daniel and Gregory R. Guy (2012): The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language* 88: 526–557.
- Fromkin, Victoria (1974): The linguistic development of Genie. *Language* 50: 528–54.
- Guy, Gregory R. (1981): *Linguistic Variation in Brazilian Portuguese: Aspects of the phonology, syntax, and language history*. Sydney: Sydney University, Department of Linguistics.
- Guy, Gregory R. (1991): Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change* 3: 1–22.
- Guy, Gregory R. (1992): Contextual conditioning in variable lexical phonology. *Language Variation and Change* 3: 223–239.
- Guy, Gregory R. and Sally Boyd (1990): The development of a morphological class. *Language Variation and Change* 2–1: 1–18.
- Guy, Gregory R., Barbara Horvath, Julia Vonwiller, Elaine Daisley and Inge Rogers (1986): An intonational change in progress in Australian English. *Language in Society* 15: 23–52.

Kroch, Anthony (1989): Reflexes of grammar in patterns of language change. *Language Variation and Change* 1–3: 199–244.

Labov, William (1966): *The social stratification of English in New York City*. Washington: Center for Applied Linguistics.

Labov, William (1980): The social origins of sound change. In: William Labov (ed.), *Locating language in time and space*, pp. 251–265. New York: Academic Press.

Mufwene, Salikoko (2011): Language evolution: an ecological perspective. *Perspectives* 4. Réseau français des Institutes d'études avancées (www.rfiea.fr).

Otheguy, Ricardo and Ana Celia Zentella (2012): *Spanish in New York*. New York: Oxford University Press.

Pierrehumbert, Janet (2001): Exemplar dynamics: Word frequency, lenition, and contrast. In: Joan Bybee and Paul Hopper (eds.), *Frequency effects and the emergence of linguistic structure*, 137–157. Amsterdam: John Benjamins.

Pierrehumbert, Janet (2006): The next toolkit. *Journal of Phonetics* 34: 516–530.

Roberts, Julie (1993): *The acquisition of variable rules: t,d deletion and –ing production in pre-school children*. Ph.D. dissertation, University of Pennsylvania.

Santa Ana, Otto (1992): Chicano English evidence for the exponential hypothesis: A variable rule pervades lexical phonology. *Language Variation and Change* 4: 275–289.

Schütze, Carson (1996): *The empirical basis of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.

Weinreich, Uriel, William Labov and Marvin Herzog (1968): Empirical foundations for a theory of language change. In: Winfred Philipp Lehmann and Yakov Malkiel (eds.), *Directions for historical linguistics*, 95–195. Austin: University of Texas Press.