University of Pennsylvania Working Papers in Linguistics

Volume 3
Issue 1 *(N)WAVES and MEANS: A selection of papers from NWAVE 24*

Article 2

1-1-1996

# Post-Saussurean Linguistics: Toward an integrated theory of language

Gregory R. Guy

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/pwpl/vol3/iss1/2
For more information, please contact libraryrepository@pobox.upenn.edu.

# Post-Saussurean Linguistics: Toward an integrated theory of language

This working paper is available in University of Pennsylvania Working Papers in Linguistics: http://repository.upenn.edu/pwpl/vol3/iss1/2

# Post-Saussurean Linguistics:
# Toward an integrated theory of language

Gregory R. Guy
York University

## 1 Introduction: The Saussurean basis of modern linguistics

Five years from now the temporal odometer is going to roll over into the 2000s and there is going to be a veritable orgy of retrospective analyses being produced on every conceivable subject. In this paper I want to jump the gun a bit and get in early with an end-of-the-century stock-staking on the subject of linguistics.

The discipline of linguistics has clearly had a good 20th century. Actually, it had a pretty good 19th century, too; one might say linguistics has been on a 200-year roll ever since Sir William Jones's speech to the Royal Asiatick Society in 1786 which first postulated the Indo-European language family. But the 20th has probably been the hottest century for the discipline since Panini. From fairly humble beginnings with a primarily historical focus, it has exploded into a substantial enterprise dealing with many different aspects of human language. Numerous important discoveries have been made about the nature, operations and development of language. And with this burgeoning knowledge has come many new models and theories about how the various aspects of language are organized. In the last 30 years especially there has been a huge production of named theories of various aspects of linguistic structure, such as the Chomsky's 'Standard Theory' and its extensions, Principles and Parameters, Lexical Functional Grammar, Generalizaed Phrase Structure Grammar and its successors, the minimalist program, Autosegmental Phonology, Lexical Phonology, Optimality Theory, etc. So if progress in the field were measured in terms of theory-production, we would have to conclude that linguistics is in great shape and that its practitioners were displaying great energy and creativity.

However, from another point of view, there is some cause for concern. Merely producing a new theory does not constitute progress unless the new theory represents some kind of advance over the one it replaces: either it accounts for the same things in a more compelling way, or it accounts for other things that the old model leaves unexplained. Now when new theories are advanced in linguistics, such claims are always made about them. Of course, no-one would ever advocate a new theory on the grounds that it was WORSE than its predecessors. And some, perhaps most, of these claims are valid, so, in a micro-sense, linguistic theory may still be making progress. But it is also obvious that when conflicting theories compete for the same turf they can't all be true. One could look at the diversity of theoretical approaches in current linguistics and see not energy and creativity, but fragmentation, confusion and incoherence. The range of facts about language is partitioned and re-partitioned and each theory defines a small area over which it operates and constructs its models without regards to any other facts about language. As a result, the models proposed in one theory are usually incompatible with those of other theories, especially those dealing with other aspects of language and not much more than lipservice is paid to the problem of how the various models might be integrated.

Because of this fragmentation and lack of cohesion, it is not clear that the field is making general progress in a more macro-sense. Language, it seems obvious to me, is by its nature an INTEGRATED system. All human beings use all of its aspects in their everyday lives with astonishing facility: they use the phonology and syntax and lexicon all at once to produce utterances; they switch effortlessly between production and perception; they all acquire language in childhood and use it fluently throughout acquisition; they all have a communicative competence encompassing the norms of usage in their spech communities; they can all use language in many different ways: expressively, purposefully, analytically, persuasively and playfully; and they all have a sense of change: of what is new and what is old and where things are heading. And most of the time, human speakers do all this smoothly, without any glitches occurring at the 'interfaces' between the different parts of linguistic structure and without any subjective experience that they are switching between, say, synchronic and diachronic operations, or between a discourse module, a syntax module, a phonological module, etc. So in real life, language appears to be integrated, seamless and smooth. Linguistic theory, on the other hand, is anything BUT integrated. So far from being a seamless garment, it is more like scattered bits of wildly clashing fabric, many of which haven't even been assembled in the same room yet. If one were to try to put together a model of the human language faculty out of the different theories now extant of the various parts of language, one would get little more than a compendium of glitches.

Now to defend the discipline against this complaint, one could argue that the problem is reflective of the immaturity of the discipline, that eventually greater integration will be achieved. One could point to the recent interest in the 'interfaces' between various structural levels as evidence of a move towards theoretical integration. (The Linguistic Institute at Ohio State in 1993 took 'Interfaces' as its organizing theme, for example.) Such unifying trends are encouraging, but they are not going to take us where we need to go. They have limited aims, mainly that of unifying formal synchronic theories of the various levels of linguistic structure and therefore they ignore certain important aspects of the problem. In fact, I will argue that there are some directions that linguistics CANNOT make progress in, because of certain aspects of how the enterprise is currently conceived. In several important ways, the dominant conceptual framework of 20th century linguistics makes it a virtue to be theoretically fragmented, to partition the data and address it piecemeal and to explicitly and deliberately ignore other aspects of language. What I want to do in this paper is to sketch some aspects of how this works and show that this fragmentation is unjustified.

I will address two dimensions of opposition, two conceptual dichotomies that have been organizing principles for much recent work in linguistics. Despite the theoretical fragmentation that I have referred to, there is a common conceptual framework about the nature of language and the proper organization of the discipline that underlies most of modern linguistics. Much of this framework derives from the man who is sometimes referred to as the founder of modern linguistics, Ferdinand de Saussure. Saussure was a great dichotomizer and both of the dichotomies I'm going to consider were most influentially articulated by him. They are the distinction between **synchrony and diachrony** and the opposition between **langue and parole**. The synchrony/diachrony dichotomy divides language temporally, seeing the system at any moment as an ahistoric time-slice and similarly divides the discipline into two different perspectives that are explicitly conceptualized as having little or nothing to do with each other. The langue/parole dichotomy divorces the abstract systematic structure of language from its actual use and for many purposes privileges the structure as the principal focus of the discipline. Significantly, Saussurean linguistics explicitly denies the possibility of integrating across these two divides. There can be a linguistics of synchrony and another of diachrony and a linguistics of langue and at least hypothetically a linguistics of parole (although Saussure and most of his succeşsors make it clear that *langue*-istics is the real linguistics), but they will all have distinct goals and methods and never the four shall meet. Each member of these oppositions has an antithetical status to the other member and theoretical integration across them is not seen as possible or even desirable. This position continues to be reflected in much current work and as long as it prevails, an integrated account of the phenomenon of language, one which reflects the integrated experience of its users, will contiinue to elude us. So what I wish to do is to attack these two dichotomies and argue for a Post-Saussurean approach to linguistcs, in which we seek a theoretical synthesis that supersedes these oppositions.

## 2 Integrating synchrony and diachrony

I will begin with the opposition between synchony and diachrony. In the Saussurean view, synchronic linguistics describes the structure of language at a given point in time and accounts for the linguistic competence of the speaker, who has no knowledge of the history of his or her language. A synchronic grammar must eschew all historical devices and treat the language as if it were static and immutable. Historical, or more properly, diachronic linguistics, on the other hand, is free to use a completely independent set of models and explanatory principles to account for language change, but it cannot rely on the synchronic grammars of the speakers of a language to generate change, because these are intrinsically static and without directionality.

In Saussure's words:

> The opposition between the two points of view — the synchronic and the diachronic — *is absolute and admits no compromise.* (Harris 1983: 83 translation of the *Cours de linguistique generale*; emphasis mine)

His rationale for this position can be seen in another quote from the *Cours:*

> The first thing which strikes one on studying linguistic facts is that *the language user is unaware of their succession in time:* he is dealing with a state. Hence the linguist who wishes to understand this state must rule out of consideration everything which brought that state about and *pay no attention to diachrony.* (1983: 81; emphasis mine)

In my view, these arguments are highly debatable. In fact, this seems a nonsensical strategy for developing a theory, given that we know that all languages are always changing. If we took a snapshot of a ball dropping, we would see it apparently suspended in mid-air. But would we construct a theory of physics to explain this synchronic time-slice? If we did, would it be very useful or satisfying? I think it is far more sensible to base our discipline on all the facts at our disposal, which includes the fact of continuous linguistic change. Of course, this position is hardly original with me; 20th century linguistics has been blessed with a sustained anti-Saussurean tradition on this question, which is strongly exemplified by some of my own teachers, such as William Labov. But the mainstream of synchonic linguistics has followed Saussure on these points, so the arguments against him bear repeating. What I present here is my own take on the matter.

We should begin by noting that a static synchronic theory of language structure makes all change appear impossible. If we conceive of language as a static articulated structure, like a building, how can it change into some other structure? Buildings do not evolve into other buildings — say, starting out as a house and then growing up to become a skyscraper or a parking garage. Rather, tampering with such a structure is more likely to lead to catastrophic collapse. But when we look at the history of a language like English, we find it changing drastically, from inflecting to isolating morphology, from verb-final/

verb-second to SVO syntax, undergoing a Great Vowel Shift and numerous other phonological alterations, in the course of 1000 years and all of this happens without its speakers, as far as we know, experiencing anything like a catastrophe.

There are two main lines of argument against the Saussurean position. First, on the historical side, note that Saussure's position seems to rule out any STRUCTURAL explanation or causation of linguistic change. If there was anything in the structure or operations of a language at a particular point in time that was unstable, that inclined it toward change in some particular direction, then the "opposition" between synchrony and diachrony would not be absolute and furthermore, such structural facts could be available to the speakers as part of their knowledge of language. So if Saussure's position is to be maintained, the only possible causes of change are social and historical events that are external to language structure. This would be an unhappy and unlikely, state of affairs for most linguists. Indeed, extensive work on the causes of change have turned up a variety of kinds of structural causes. Hence the post-Saussurean approach prefers a theory of language that incorporates vectors and identifies fundamental tensions in the linguistic system that drive change; in other words a theory that is DYNAMIC and DIRECTIONAL, one that shows the apple dropping, as it were. If we can demonstrate general principles of change and inherent directionality in change processes, then the Saussurean division is not a necessary or adequate basis for organizing linguistic theory.

Second, we should question the assumption that speakers know nothing about language change, history, directionality. I will argue that speakers have a lot of information at their disposal about what is old and what is new in language and about which way the language is headed. If this is true, then the basic rationale for S's position, cited above, is falsified and the whole dichotomy is suspect.

## 2.1 Evidence that speakers have about history

I will deal with this second argument first. What evidence is available to speakers about ongoing change in their language? It seems clear that there are several kinds of sociolinguistic facts available to speakers that suggest change. First, there are the obvious cases of major sociohistorical events triggering language change. Thus, the history of English cannot be understood without reference to the Norman conquest and subsequent contact with French; and pidgin and creole languages would not even exist were it not for the dramatic social events associated with slavery. The speakers in such situations are surely acutely aware of the facts and causes of the changes going on in their languages. So in these extreme cases, the social situation — the presence in the community of different language varieties with different statuses — facilitates or causes language change.

But the same thing is true in less dramatic ways. When we look at sociolinguistic variation in speech communities around us, we find that they contain a diversity of linguistic variants or varieties, used by different groups and associated with different statuses and evaluations. These diverse linguistic forms are, in a sense, in competition with each other and over time, some of them expand and other are lost. As far as we know, this situation is true of ALL human speech communities, which means that the same kind of process that causes linguistic change in the extreme cases — the process by which sociolinguistic diversity leads to language change — this process is also operating in all communities, albeit on a smaller scale. In an important sense, therefore, variation and change are two different faces of the same thing: variation is the synchronic face of change and change is a diachronic outcome of variation.

This allows the possibility that speakers are aware of ongoing change in the ordinary, relatively homogeneous speech communities for the same reasons they are in the more extreme cases of contact and conquest. In other words, the social distribution and evaluation of competing variants provides evidence about what is new and expanding in a language vs. what is older and in decline. Over the last three decades, numerous studies of change in progress have shown that there are characteristic social distributions of innovative forms. Let us briefly consider this kind of evidence.

The most obvious social dimension of an ongoing change is AGE. Simply put, incoming, expanding linguistic variants are used more by the young and the language of older speakers is 'archaic' in the sense that it reflects the usage that predominated when *they* were young. The typical age-distribution of an innovation is the S-shaped curve: the peak usage of the innovation is found among teenagers and young adults. Of course, in looking at age-grading we must always take care to distinguish ontogenetic processes of individual maturation from phylogenetic processes like language change. (Thus, the fact that all babies born this year do not yet speak any language is not evidence that the human species is losing the power of speech.) But there are ways to sort out the two situations and when we subtract maturational processes, residual age-grading is suggestive of change in progress.

Another distinctive dimension of the social distribution of change is GENDER. In western industrial societies, most known changes-in-progress are lead by women, who are often as much as a generation ahead of their male counterparts. The counterexamples in the literature to this generalization are often the kind of exceptions that prove the rule. For example, in Labov's Philadelphia studies, men have been found to lead in the centralization of (ay) before voiceless consonants (a.k.a. 'Canadian raising'), as in *fight, right, pipe.* What makes this change interesting is that it is a retrograde movement, reversing the historical direction of change. This is a Great Vowel Shift vowel, starting in Old English as /i/ and then developing a centralized nucleus which was subsequently lowered along a non-peripheral track. So if we believe Labov, Yaeger and Steiner, ALL English speakers once said [rʌyt]. Philadelphia males may therefore have been historical laggards in the lowering and lead now only by virtually of having been behind until everyone turned around.

Finally, the CLASS distribution of innovative forms is also distinctive, at least for the type that Labov calls 'change from below'. Labov has identified a 'curvilinear' pattern of class distribution whereby the peak usage of the innovation is found in the 'interior' social groups at the middle of the class scale: roughly speaking, the lower middle and upper working class speakers. Kroch (1978) identifies an alternative, in which there is a simple inverse linear correlation between class and innovative usage. But in both models, the highest status groups use the innovations less than the second and third-highest groups.

All of these characteristic distributions are illustrated in the findings of Guy, Horvath, Vonwiller, Daisley and Rogers (1986), dealing with the use of high-rising intonations in declarative clauses in Australian English. In that article, we called this phenomenon "Australian questioning intonation", but the march of language change in the last decade has rendered this designation obsolete, because the phenomenon is now widespread in North America and other parts of the English speaking world. But terminology apart, the facts clearly show that it is an expanding innovation in English and was an active change in progress in Australia in the early 1980s. The age distribution of AQI, given in Table 1, shows a clear S-shaped curve, peaking in the older teens. The sex distribution in Table 2 shows women in the lead and the class distribution in Figure 1 shows a curvilinear pattern for men and a Krochian linear pattern for women, with in both cases less usage by the highest status group than the second-highest.

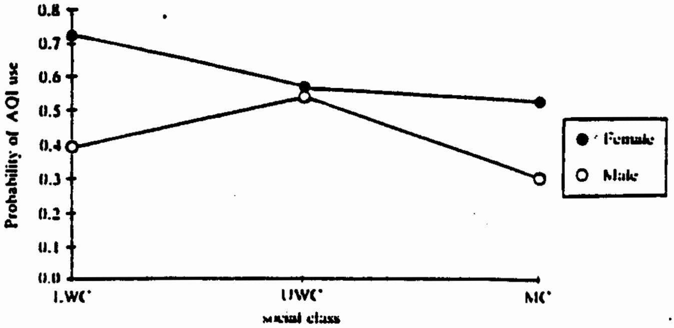
| Age Group | No. of tone groups | %AQI | Factor Weight |
|---|---|---|---|
| 11-14 | 5,032 | 1.6 | .64 |
| 15-19 | 15,067 | 2.0 | .67 |
| 20-39 | 4,386 | .5 | .51 |
| 40+ | 19,642 | .2 | .21 |

**Table 1. Age distribution of AQI use.** (from Guy et al. 1986, Table 2)

| Speaker sex | No. of tone groups | %AQI | Factor Weight |
|---|---|---|---|
| Male | 53,769 | 1.0 | .41 |
| Female | 53,916 | 2.2 | .59 |

**Table 2. Sex distribution of AQI use.** (from Guy et al. 1986, Table 4)

**Figure 1: Class and sex distribution of AQI use.** (from Guy et al. 1986, Fig.2)



As we were doing this research on AQI, it became clear that our consultants were all at least subjectively aware that it was an innovation and that it had this kind of distribution. Everybody knew that children did it more than adults and women more than men; most people associated it with a lower class status and many adults explicitly identified it as an innovation that had occurred within their lifetimes: 'We didn't do that when we were kids.' All of this goes to undermine the Saussurean position quoted above. It is *not true* that these "language users [were] unaware of [the] succession in time" of the facts about this form. This kind of awareness may have been exceptionally conscious in the AQI case, but if the social distributions are systematic and have some regular association with the distribution of innovations in general, then speakers everywhere should have some access to this kind of information and Saussure was simply wrong to conclude that the linguist "must pay no attention to diachrony." Rather, people DO know about diachrony and directionality from personal experience. This experience may be limited to the lifespan of an individual; there's no point in claiming that people have knowledge of remote historical events; but on the micro-level, where linguistic changes are actually instantiated and advanced, speakers often are aware that change is underway and participate in it or resist it for clear social motives.

## 2.2 Linguistic factors in change

Now let us consider the other argument against the Saussurean position, the issue of structural causes of change. If linguistic systems of a certain type tend to change in one direction and not another, if there are identifiable general principles governing change processes, then arguably these dynamic vectors should inhere within the grammar, like other linguistic universals and once again we cannot rule out of our descriptions and theories any kind of historical evidence. Looking at the historical literature, one finds numerous proposals that there ARE structural principles governing sound change. Lenition, deletion, assimilation, pattern pressure causing analogic changes, etc. are all considered unmarked, natural directions of change and their converses are all marked and unusual. Such proposals may be most evident in phonological change, but there are also theories of linguistic factors driving syntactic, morphological and lexical change. I will consider two examples: one dealing with sound change and another with morphosyntax.

**Vocalic chain-shifts.** One of the most precise and testable formulations of a general principle of phonological change is found in Labov, Yaeger and Steiner's (LY&S, 1972) study of chain shifting in vowels. A chain-shift is a linked series of sound changes such as the Great Vowel Shift in English, which have the form a -> b -> c; in other words a series of phonemic units each shift one position along a sequence of adjacent phonetic values. LY&S find that three general principles appear to govern vocalic chain-shifts:

I. In chain shifts, tense vowels rise.
II. In chain shifts, lax vowels usually fall, especially the lax nuclei of upgliding diphthongs.
III. In chain shifts, back vowels move to the front. (1972: 106)

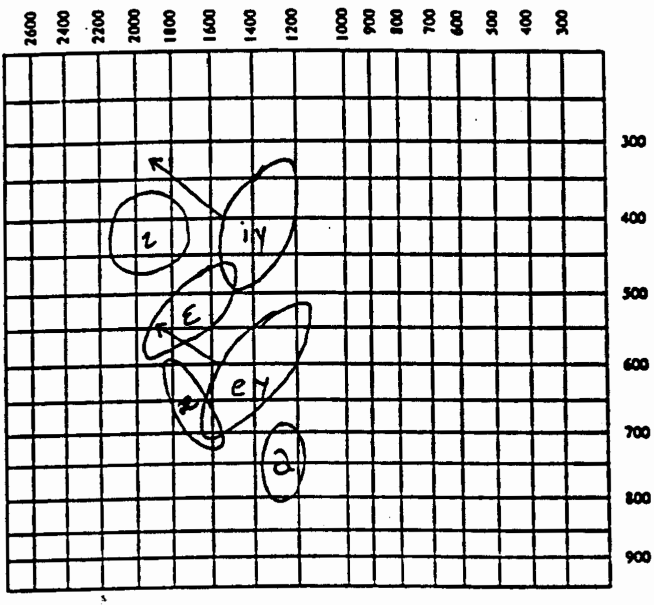
The tense-lax distinction in these statements is the one that distinguishes opposing vowel classes like English [i, ei, u] vs. [ɪ, ɛ, ʊ]. In English these classes are defined by a combination of phonetic and phonological features. Vowels in the tense class are typically longer and more peripheral, while the lax vowels are shorter in duration and more central: thus [i] is fronter and higher than [ɪ], while [u] is backer and higher than [ʊ]. In addition, English has a systematic phonological distinction: only tense vowels appear in final open syllables: thus *he* and *hay* are legitimate English words, while [hɪ, hɛ] are not possible.

An illustration of the LY&S principles at work can be found in the vowel realizations of current Australian English (AusE), which are strikingly shifted away from their expected positions. After the completion of the major vocalic changes of Middle and Early Modern English (such as the Great Vowel Shift), the major English vowel phonemes are generally understood to have occupied positions in F1-F2 space like those described above: /iy/ and /ih/ are high-front, /ey/, /eh/ mid front, /uw/, /uh/ high back and in each pair the tense member was more peripheral than the lax member.

In Australian English, however, all of these phonemes are displaced in a chain shift. According to the desciption of Mitchell and Delbridge 1965, vernacular speakers show the diphthongal or 'tense' vowel class with centralized and lowered nuclei: /iy, uw/ -> [əy, ´w] and /ey, ow/ -> [ʌy, ʌw]. This is consistent with LY&S principle II: although originally tense, these vowels have developed 'lax' (i.e. centralized) nuclei, which are consequently expected to fall. The front and back high vowels, in particular, are merely recapitulating comparable developments in the Great Vowel Shift.

However, what of the original lax vowels of English: /ih, eh, uh/ etc.? My impressionistic phonetic analysis of these, when I first encountered them, was that they were RAISED. I even experienced various phonemic confusions in conversation, such as interpreting an Australian utterance of the word *possession* as equivalent to my American English *position,* thus misconstruing a raised /eh/ as /ih/. On first glance, this seemed to be a violation of LY&S principle II: if these vowels were lax, they should not have raised.
To investigate this apparent violation, Barbara Horvath and I conducted an acoustic study of Australian English speakers. The F1-F2 measurements of the front vowels for a typical male speaker in our study are shown in Figure 2. These reveal the solution to the problem. Focussing on the peripheral/central dimension, we find that the AusE /ih/ is more peripheral (fronter and higher) than the nucleus of /iy/ and /eh/ is more peripheral than /ey/. Thus if we select peripherality as the defining property of tenseness, these vowels are, in effect, 'tense' in AusE! Consequently, they are rising in accordance with LY&S principle I.

**Figure 2: Peripheral and non-peripheral front vowels in Australian English**



This approach is fully consistent with LY&S's analysis. They emphasize peripherality as the defining characteristic of tense vowels and restate principles I and II as follows:

> I'. In chain shifts, peripheral vowels rise.
> II'. In chain shifts, non-peripheral vowels usually fall. (1972: 106)

So, what has happened in AusE is that these members of the tense and lax vowel classes have changed places on the peripherality dimension. The formerly 'tense' upgliding diphthongs /iy, ey, uw, ow/ have acquired centralized, 'lax' nuclei and are falling on a non-peripheral track. This vacated the periphery of the vowel space, leaving /ih, eh, uh/ to become phonetically tense and then to rise along a peripheral track. Consequently, we may conclude that LY&S were right and their principles are, in fact, NOT contradicted by the AusE situation.

But more importantly for the post-Saussurean case, we can also conclude that the LY&S principles characterize a dynamic vector which is somehow built into the vowel system of human language. Given a particular vowel system with a particular synchronic structure, we can predict likely directions of change. Of course, we cannot yet predict WHEN it will change, or if it will remain stable; neither can we say exactly by what mechanism of production or perception these principles operate, or why, at the present state of our knowledge, we need the qualifier "usually" in principle II. But we can rule out many changes that could be imagined, but are evidently impossible, such as [i -> e -> æ]. Hence, the facts again contradict Saussure's position. If LY&S's principles universally constrain linguistic operations, then they are in some sense available to the 'operators' and the argument that understanding a synchronic state requires ignorance of diachrony becomes absurd.

**Structural constraints on morphosyntactic change.** This idea of structural 'vectors', dynamic principles that say not only where the language is, but also where it is going, has also been advanced in morphology and syntax. I will illustrate with the problem of how functional categories are maintained in the face of linguistic change. I will cite two theories of how maintenance works, one from Kiparsky (1982) and one of my own. Both appeal to universal linguistic processes present in synchrony as well as diachrony.

Kiparsky's proposal is that there is a general principle, the Distinctness Condition (DC), that defends morphological categories from erosion by language change: "There is a tendency for semantically relevant information to be retained in surface structure. ... It characteristically originates as a blocking of rules in environments in which their free application would wipe out morphological distinctions on the surface." (1982: 87-89)

The DC thus makes predictions about the future course of linguistic change. Whenever a strongly functional morphological category is threatened by some other change, that change will be blocked. In other words, the DC spells out certain types of change that are prohibited, presumably on universal grounds. If this is true, it would clearly have to constrain synchrony as well as diachrony

In several papers I have given a different interpretation of functional constraints. My work suggests that the DC is NOT operationalized as a constraint on real-time PRODUCTION of language. Rather, functional distinctions are maintained through the normal operations of perception and acquisition. The argument runs as follows (cf. Guy, to appear).

In production, data from several sources suggest that apparently functional constraints on variable processes are actually formal constraints, conditioned by morphological structure. Whenever there is a mismatch between form and function, the formal structure is what governs the patterns of variation. Thus in English -t,d deletion, there is an apparently functional constraint by which past tense verb forms like *talked,* *raced, planned,* undergo less deletion than monomorphemic words like *east, old, act.* This preserves the functional tense marker represented by the *-ed* suffix and avoids homonymy with the present tense. But this could just as well be expressed as a formal constraint, roughly, don't delete after a morpheme boundary (*#_#). The question of whether the productive phonology responds to the form or the function can only be decided by looking at other cases where form and function make different predictions.

Such a case is supplied in English by the past participles in *have walked, have planned.* Fortuitously, these are formally identical for most verbs to the regular past tense form, but the functional load of the participial *-ed* affix is extremely low (deletion of the affix in a sentence like *I've miss' my bus* produces no dysfunctional ambiguity whatsoever). So on functional grounds these affixes should be freely deletable, but if the rule is sensitive to the boundary, they should behave like past tense verbs. What do the data show? Some of my results appear in Table 3, which has been replicated several times in various English dialects: past participles are deleted at a low rate, like the formally equivalent past tense verbs. Therefore, the rule is actually conditioned by form, not function.

| | N | % Deleted | Factor weight |
|---|---|---|---|
| Monomorphemes (e.g. *mist, pact*) | 739 | 38.6 | .64 |
| Irregular Past (e.g. *lost, kept*) | 74 | 35.1 | .60 |
| Regular Past (e.g. *missed, packed*) | 157 | 19.1 | .41 |
| Past Participles (e.g. *have missed*) | 74 | 17.6 | .35 |

**Table 3. English -t,d deletion in 4 morphological classes.**
(from Guy, to appear)

However, a functional constraint must still affect perception. If speakers delete past tense markers, or any other markers of morphological categories, they run the risk of misconstrual by hearers. And when the hearer misconstrues the deleted form, that form DOES NOT COUNT in their perception of whether the deletion rule applies to that category. If you hear someone say *'I always miss my bus'*, you normally perceive that as present tense and NOT as a token of past tense *missed* with deletion applying to remove the marker. So in the ordinary course of events, hearers will not PERCEIVE violations of the distinctness condition and it will appear to them as if Kiparsky was right. Children learning the language, in attempting to make their output match the input they perceive, will attempt to construct grammars that have formal constraints emulating the distinctness condition, like 'don't delete in the context #_#'. This process is not perfectly efficient; occasionally their formal constraints will pick up and protect categories with low functional load, like the English past participle, which is just getting a free ride on the past tense form. But this process should operate in all languages at all times and by this means, pseudo-functional constraints will constantly appear and determine the course of language change.

Numerous other examples of general principles governing morphosyntactic change could be adduced. Kroch's (1989) work on the development of periphrastic *do* constructions in English and Tarallo's (1995) treatment of a series of changes affecting word order and object pronoun use in Brazilian Portuguese, both show how syntactic changes can be drivẹn in a specific direction as a consequence of their linguistic embedding. Naro and Lemle (1976) and Naro (1981) argue that saliency of morphological distinctions has a directional effect on change from below, such that they begin in unsalient environments where distinctions are minimal and reanalyses particularly plausible. And

Guy (1981) argues further that saliency affects targeted changes in the reverse direction, developing first in salient environments.

### 2.3 The post-Saussurean synthesis

So what may we conclude from all this? There is ample evidence from a variety of sources to show that it is incorrect to assume that speakers know nothing about change, that their experience of language is of a frozen and static slice of time. Rather, they can tell from their own experience and from observing the social distribution of forms in their speech communities, which way things are heading. Furthermore, the existence of structural directionalities in linguistic change, of general principles governing linguistic evolution, suggests that Saussurean absolutism is incorrect. Rather, to give a coherent account of synchrony, we must often make reference to diachrony and the forces that give rise to change are inherently present in the structure of language, even when it is viewed as a series of frozen moments. Adequate linguistic theories, therefore, must take note of this and cannot be designed in ways that ignore the dynamic nature of language. Hence we can and must abandon the first Saussurean dichotomy.

## 3 Integrating langue and parole

Now let us consider the second Saussurean dichotomy: *langue* vs. *parole*. This defines a fundamental opposition between, first, what is considered the essential SYSTEM of language — the abstract mental construct of processes and elements that defines what is possible in a language and comprises what we would now call the generative capacity of a speaker — and second, the operations and products of that system, the actual usage of language by speakers. This opposition has been reformulated in the course of the century: Saussure's distinction between langue and parole has now largely been subsumed by Chomsky's contrast between competence and performance. But the elements of the distinction remain the same. On the one hand there is the abstract, not directly observable construct: the grammar, langue, competence; and on the other hand there is the concrete, observable sum of language production: parole, performance, utterances.

This dichotomy has colored much of linguistic thought in this century and it is still a vital organizing principle for the conduct of the discipline. For many linguists, it affects how they conceive of the object of study, what they see as data, what methods they use and how the discipline is organized. For both Saussure and Chomsky, the distinction offers a specific rationale for the fragmentation of the field: the main business of linguistic theory is to account for langue/competence and hence the study of phenomena which may be defined as arising in parole/performance is really something else, not strictly speaking a part of linguistics. Saussure does allow that one might imagine a separate linguistics of parole, but the only kind of linguistics he does himself is the linguistics of langue. And Chomsky is more extreme: he appears to rule performance right out of the discipline. Thus he says, in *Aspects* (1965: 4): "Observed use of language... surely cannot constitute the actual subject matter of linguistics, if this is to be a serious discipline." He does allow that performance data "may provide evidence as to the nature of [competence]" and he even envisions the development of a theory of performance in *Aspects* chapter 1.2; but what he calls 'linguistic theory' is really concerned with "discovering a mental reality underlying actual behavior". In this view, our only interest in what people actually do with language is that it "may" provide a means towards this end of investigating competence, which is the real business of the field.

Many consequences that follow from this position, but the one of greatest importance for a sociolinguistic audience is what it means for variation. For both Saussure and Chomsky, the SYSTEMATIC properties of language are seen to lie in langue/ competence. And systematic, for them, means invariant and categorical. One of the great themes in the history of linguistics has been the search for INVARIANCE. The paradigmatic example of this in phonology is the development of the concept of the phoneme: out of a plethora of different phonetic realizations (allophones), we assemble a single invariant unit, the phoneme, which may appear different ways in different contexts, but which is the mentally real unit of grammar in which lexical items are represented, on which phonological rules operate, etc. And syntax has analogous theoretical constructs: early transformational syntax linked different surface structures to a common deep structure by means of transformational rules. In each case, doing linguistics consisted of finding unity out of diversity and the system of language lay in the invariant, unified, 'deep' constructs, not in the variable 'surface' realizations.

In this view, then, variability is not part of langue or competence. The items and processes of the linguistic system were either invariant or, in a limited range of cases, they were 'optional'. Optional processes were ones that could occur or fail to occur, but their occurrence was either random, or else governed by non-linguistic factors outside of the scope of the discipline. Therefore variability lay in performance and was not a subject for linguistic theory. Sociolinguistic variability in particular is not part of this view of linguistics. Thus, we get Chomsky's famous quotation: "Linguistic theory is concerned with the ideal speaker-listener, in a completely homogeneous speech community, who know its language perfectly..." (1965: 3). This implies, of course, that virtually *everything* investigated by sociolinguists and variationists is NOT a concern of linguistic theory.

Now to sociolinguists, such a position seems just silly. But Chomsky and Saussure are intelligent men, who presumably do not take such positions frivolously. Before attacking them one ought to understand why they take these positions. What was the point of setting up the langue/parole dichotomy and what uses has it been put to?

I think there were several reasons for this distinction. First, it is a simplifying assumption. This is a fair approach to any problem: faced with something too big to deal with all at once, you break it down into smaller problems, ignoring some aspects while you work on others. All sciences do this and it often results in significant insights. Thus Newton's theory of motion works quite well if one ignores friction and relativistic effects. So excluding parole was a way of setting aside a lot of complex stuff to allow progress on other issues. And in this it has worked; the theoretical work in 20th century linguistics that has been carried out under this framework has certainly achieved major advances. However, we make a serious and obvious mistake if we elevate this simplifying assumption to the status of a theoretical principle.

Second, this distinction did worthwhile work as a conceptual tool in advancing theoretical ends for both Saussure and Chomsky. When Saussure enunciated the langue/parole distinction, linguistics was engaged in a struggle to differentiate itself as a discipline from traditional studies like etymology and philology that were principally historicist in perspective and focussed on the concrete products of the linguistic system. Asserting the primacy of langue was an important theoretical step, emphasizing the existence of an abstract system with its own rich synchronic structure, that language was more than just utterances. This helped to clarify the distinct intellectual content of what has become modern linguistics and to establish the fundamental notion that the mental processes which produce speech are real and complex and worthy of scholarly attention. Similarly, for Chomsky, spelling out the concept of competence helped to underscore the generative capacity of the linguistic system and to push linguistic theory from what had become a relatively static view of linguistic structure to a more dynamic and challenging pursuit of a 'generative grammar', that could seek to model the infinite creative capacity of language. Both of these were radical positions in their time and the langue/parole and competence/performance oppositions were useful tools in the polemics that their authors were engaged in.

Third, the dichotomy constitutes a compelling analogical extension of the basic analytical operations of generalization and abstraction. If we generalize about a set of observations, say, 'regular English verbs take *-ed* in the past tense' and reify that generalization in the form of abstract categories and rules, then it becomes reasonable as a next step to say that the abstractions ARE the system and the observed productions are just the output of that system. But we run certain risks in taking that step. In some ways it makes as much sense to say that the system lies within the productions themselves. Words, for example, seem to have an essentially concrete existence and it doesn't buy us much understanding of them to try to see them as products of a separate system in which they have a separate abstract identity. And related disciplines that deal with cognitive capacities do not commonly make such a distinction: thus the psychology of vision does not seek to distinguish a visual competence from a visual performance. Rather, performance itself is the object of study.

All of these motives for the establishment of this dichotomy were worthwhile, but none of them provide sufficient reason for maintaining it if there is evidence against it or if it is not serving a useful end. None of these things, neither simplifying assumptions, nor theoretical talking points, nor analogical extensions, have any privileged position as analytical primitives. If our understanding has progressed to the point where we can handle more complex cases, we can abandon simplifying assumptions; if the theoretical debates of the past are no longer issues, the tools used in those debates may need to be replaced by something else. And if a different model of reality is more useful to explain additional facts, why not use it? In particular, we should consider what can be gained from a model of language that begins with the goal of treating language as an integrated phenomenon and seeks to account for the broadest range of facts about what is is and how it is used; in other words, one that supersedes the competence/performance, langue/parole dichotomy.

I will argue that this dichotomy is no longer being used to good ends and that in fact it has been used to pose barriers to progress towards an integrated understanding of language. Let us consider some examples. One unfortunate consequence of this dichotomy was raised at the outset: that is, the fragmentation and circumscription of the horizons of the field. Linguistic theory declares itself responsible only for langue/competence and leaves the vast territory of parole/performance uncharted. The relationship between the theoretical models of the moment and the real performances of speakers is left undefined and uninvestigated and indeed, unimportant for most theoreticians. Another adverse usage of the dichotomy is that it devalues the richest source of empirical evidence about language, namely the vast continuous production of utterances and discourses by human speakers. Since data from language use must come, by definition, via performance, it is treated as not necessarily relevant to the development of linguistic theory. Many kinds of data that might be brought to bear on a theoretical argument are therefore treated with suspicion; and whenever the facts contradict a theory, there is always this potential move available to the theoretician to dismiss the facts as some kind of 'performance phenomenon'.

This has made much work in linguistics dangerously unempirical and unscientific. In place of empirical observation, this tradition has substituted a curious 'empiricism' of intuition, which is supposed to give the native speaker direct access to competence. But this is a very dubious claim. The rationale offered for this step is that performance is subject to, in Chomsky's words, "grammatically irrelevant conditions [such] as memory limitations, distractions, ... and errors", which we can filter out of the data by relying on native-speaker intuitions. But intuitions such as grammaticality judgments and the like are arguably just as much a performance or product of the system as an utterance is and they are probably subject to their own kinds of errors and other irrelevant conditions. They certainly are malleable, as anybody who's ever had a syntactic discussion along the lines of 'can you get this?' can attest.

It is important to note that, in this framework, the models that linguists construct of competence — the mental grammars that we hypothesize — do not actually 'generate' language usage at all. A grammar is functionally limited to defining 'all and only the grammatical utterances of a language' and its relationship to actual productions of natural language is oblique and ill-defined. Chomsky spells this out in *Aspects* (1964: 9): "A generative grammar is NOT a model for a speaker or hearer... When we say that a sentence has a certain derivation with respect to a particular generative grammar, we say nothing about how a speaker of hearer might proceed ... to construct such a derivation." This is seen as a virtue by many theoretical linguists, in that such theories represent the idealized abstract 'knowledge of language' in its purest form, which is theoretically equally useful for modelling any kind of linguistic phenomenon, including perhaps production, perception and acquisition. But since such a theory explicitly avoids saying anything about *any* kind of language use, it might also turn out to be universally useless.

So at bottom, what does this dichotomy consist of? It is a way of thinking about language that allows us to: (1) divide the evidence; (2) divide the field; and (3) advance certain theoretical aims. The fundamental point of the dichotomy is therefore segregative, just like the synchony/diachrony distinction. It comes down to a claim that these two aspects of language are different in their essence and need different theoretical treatments. This difference is conceived of as somewhat less symmetrical than synchrony/diachrony. Whereas that opposition is supposed to involve mutual irrelevance, most theoreticians imagine, as we have noted, that models of competence will eventually serve as the basic components of models of performance. But the bottom line is still essential difference.

To undermine that position, let us do as we did for the first dichotomy and look for ways in which the two aspects are integrated and interdependent. The fundamental issue here, as for the first dichotomy, turns out to be the treatment of variation. The basic way in which variation integrates competence and performance is, in the terminology of Weinreich, Labov and Herzog, 'orderly heterogeneity'.

## 3.1 'Orderly heterogeneity': regular properties of parole

Orderly heterogeneity amounts to the observation that parole has regular properties. As we have seen, the search for invariance that characterized so much of the mainstream of linguistic research lead ultimately to the attribution of invariance to the linguistic system itself, with the concomitant assumption that parole/performance was disorderly, a heterogeneous grab-bag encompassing speech errors and memory lapses along with other phenomena such as dialect diversity, social stratification and gender and ethnic differences, all of which linguists would not be responsible for if they narowed their focus to the fictional homogeneous world of Chomsky's ideal speaker-listener. But decades of sociolinguistic research reveal that the heterogeneous reality of language use is also orderly: it has structure and patterning, follows rules, reveals a system. Social stratification, for example, shows that the class, gender and ethnic differences between speakers in a community are actually organized into a vast structured pattern governed by shared evaluations of the variables.

What is more, this research shows that within-speaker variability also exhibits orderly heterogeneity: there are orderly and systematic, but non-categorical (i.e. quantitative) patterns in linguistic usage at all levels of linguistic structure, which characterize all speakers of a language. Thus every English speaker ever investigated shows variable deletion of clustered final coronal stops; furthermore, they all show more deletion before a following consonant than a vowel; e.g., *wes' side* but *west end.* And such orderly heterogeneity is not limited to phonology. For example, all but the most highly educated Brazilians show variable plural marking of verbs with plural subjects and all show more plural marking when the subject precedes the verb (*Eles chegaram*) than when it is postposed (*Chegaram eles*). Finally, this orderliness is, in statistical terms, non-random: that is, although individual tokens of a variable may not be predictable, the frequency of a type in a corpus can be predicted probabilistically with great precision.

The challenge that these finding present to a theory postulating invariant competence is, **where do such non-categorical regularities arise?** If they are systematic and regular, why aren't they governed by the linguistic system, i.e., the grammar? But if, on the other hand, by virtue of their variable, non-categorical nature they are excluded from an invariant grammar, then where is the system that generates them?

In light of this evidence of non-categorical regularity, only two conclusions are possible. One, which is the conclusion drawn by most variationists, is that these regular features of language are accounted for by the same thing that linguists postulate to govern other linguistic regularities, namely competence; hence competence must include a variable, even a quantitative, component. The principle model in use today that adopts this approach is the so-called 'variable rule' model of Labov, Cedergren and Sankoff and their associates. This model postulates that choice points in the grammar — points where optional elements, processes, rules, are selected — can be associated with probabilities, weighted according to context. Such a grammar will state not only that a form is grammatical or ungrammatical, but also whether it is likely or unlikely. The model further assumes that knowledge of these weightings is part of a speaker's linguistic competence.

## 3.2 The Counter-Reformation of competence

In opposition to the variationist view is an alternative conclusion drawn by linguists who wish to maintain the postulate of invariant competence. This conclusion is that the quantitative regularities of orderly heterogeneity arise outside of competence and the invariant grammar. In that case, some other 'explanation' must be offered for their regularity. One that was noted above is that 'optional' processes in the grammar are applied randomly in production. This is the statistical meaning of the 'free variation' concept of structuralist and generative theories. But this account fails immediately in the face of the non-random character of variation. In its stead, however, a few other proposals have been advanced. I will consider three such approaches, which keep systematic variation outside the grammar by attributing it to: (1) extralinguistic universals of production; (2) a separate 'grammar' of performance; or (3) competition among invariant grammars. In what follows I sketch these 'solutions' of orderly hetereogeneity and show their limitations.

**Production universals.** One early approach was the suggestion that patterned variation derived automatically from universal but trivial, essentially non-linguistic aspects of production and performance, such as the operations of the articulators, constraints of memory, functional considerations, etc. (cf. for example, Kiparsky 1982). If this were true, then the constraint patterns on variation that are observed in language use would have to be uniform in operation across dialects and indeed across languages. This prediction has, by now, been thoroughly falsified. It turns out, instead, that variable constraints have the same kind of patterning that categorical constraints exhibit: some are universal and some language-specific. I will consider two examples here.

First consider the hypothesis that phonological constraints on variability arises from some universal articulatory or production-related conditions. This explanation has been thoroughly considered in connection with the English deletion of clustered final coronal stops, a.k.a. -t,d deletion. This process, as I mentioned, is constrained by the following segment, so that there is maximum deletion before obstruents, intermediate deletion rates before glides and liquids and low rates before vowels. This pattern has been variously explained as a consequence of sonority factors, resyllabification and articulatory effort, although the details of the various explanation need not concern us here. Whatever the cause, the putative universality of the pattern permits an analysis in which the effect is non-phonological and indeed, non-linguistic, in the sense of lying outside the grammar. But just how universal are these patterns?

Some aspects of this constraint are evidently universal: for example, all English speakers appear to delete more before consonants than before vowels and the same constraint ranking is found for deletion of final consonants in other languages, such as final -s deletion in Spanish and Portuguese. But two details of this constraint are clearly language- or dialect-specific and hence should belong to the learned phonology of the grammar. One is the by now well-known fluctuation in the effect of a following pause on -t,d deletion. Following pause (i.e. phrase-final position) has in every dialect a consistent position in the constraint hierarchy, but this position differs from dialect to dialect. Thus for the Philadelphia speakers studied in Guy (1980), pause was a conservative environment, associated with low rates of deletion. But the New Yorkers examined in the same study all showed pause to be a favorable environment for deletion. The group figures for this pattern are shown in Table 4 and Table 5 shows the summary of the rankings of these constraints in the data of the individuals involved. Subsequent replications have confirmed these results and shown similar dialect-internal consistent targets for the pause environment for other speech communities. (For example, AAVE speakers consistently have high rates of deletion before pause, like New Yorkers.) These results lead us to conclude that the pause effect has a dialect-specific target for each speech community; therefore it cannot be a universal, but must be learned by language acquirers from exposure to data, just like other language-specific features of competence. In other words, this variable feature is part of the grammar.

| | --------- Following Context ---------- | | |
|---|---|---|---|
| | Obstruent | Vowel | Pause |
| 19 Philadelphians | 1.0 | .40 | .19 |
| 3 New Yorkers | 1.0 | .56 | .83 |
| | (Varbrul1 factor weights) | | |

**Table 4. Dialect differences in following context effect on -t,d deletion** (from Guy 1980)

| | Number of speakers for whom the effect of following: | | |
|---|---|---|---|
| | Vowel > Pause | Pause > Vowel | |
| Philadelphians | 18 | 1 | |
| New Yorkers | 0 | 4 | (signif.: p<.01) |

**Table 5. Ordering of following segment effects on -t,d deletion iη New York and Philadelphia** (Guy 1980).

A second detail of the following segment constraint on -t,d deletion is also language-specific. In recent work, I have given an autosegmental account of -t,d deletion that explains certain aspects of the process in terms of resyllabification of the final stops as onsets of the following syllables (as has transparently occurred in phrases like *band-aid*). This analysis made the prediction that following /l/ and /r/ would have different effects on the deletion rate, because resyllabified /tr-, dr-/ onsets are possible in English, but language specifically, /tl-, dl-/ onsets are not. This prediction has been strongly confirmed in a series of recent studies; Table 6 shows the figures from Guy (1991). Hence, once again, we would conclude that the process is constrained by a specific feature of English phonology, not some articulatory universal.

| | -------------- Following Segment ---------------- | | | | |
|---|---|---|---|---|---|
| | Obstruent | /l/ | Glide | /r/ | Vowel |
| Rate of deletion | .66 | .80 | .57 | .42 | .19 |
| | (Varbrul2 factor weights) | | | | |

**Table 6. Following segment effect on -t,d deletion contrasting /l/ and /r/** (from Guy 1991).

**Performance grammar.** Thus, the attempts to remove systematic variability from the grammar by attributing it to nebulous production universals has failed. But other approaches have risen in its stead. A second possible argument that has sometimes been advanced in order to keep the grammar invariant is to attribute the regularities of variability to a hypothesized separate (but so-far, uninvestigated) system governing production, a kind of 'grammar of performance'. This could have its own language-specific characteristics, thus circumventing the above failings of universalist explanations. However, this approach is itself suspect because the units and explanatory principles required in such a performance grammar keep looking eerily like the units and principles required to account for categorical phenomena. We have already seen an example that illustrates this point. If the avoidance of *tl- and *dl- sequences that is evidenced for a variable process in Table 6 is due to a separate *performance* constraint on English and is NOT related to the *categorical* constraint against such onsets that appears in the competence grammar of English, this would be a very surprising coincidence indeed.

Another example appears in the analysis of preceding context effects on -t,d deletion. Deletion rates have long been known to depend in part on place and manner features of the preceding context: for example, English speakers always delete more after /s/ (e.g. *mist, west*) than after /l/ (e.g. *melt, old*). Recent work by Charles Boberg and myself shows that this effect can be reduced to a simple similarity function between the preceding segment and the target coronal stop, involving the features [cor], [cont], [son], (and also [voice]). Figures from Guy and Boberg (1994) are given in Table 7. When the target shares like values of these features with the trigger, deletion is favored, but when target and trigger disagree for some feature value, deletion is disfavored. Thus higher rates of deletion in a word like *west* arise because the preceding /s/ shares the feature values [+cor, -son] with the target /t/, while lower rates of deletion in *melt* arise because the /l/ shares only the feature [+cor].

This, any phonologist will note, looks like an OCP effect. The Obligatory Contour Principle in current theory is postulated to prohibit same-tier sequences of adjacent identical autosegments, segments, or even features. It is the constraint that explains facts like the English prohibition of geminates, which is why final -t,d are never preceded by another -t or -d (when this would arise thru affixation, for example, epenthesis occurs to break up the sequence; viz, *painted, raided*, etc. **paint#t, *raid#d.*) This prohibition is categorical, applying to lexical entries as well as surface forms. Guy and Boberg (1994) suggest that it is this same principle, operating probabilistically, that accounts for the data in Table 7.

| **Preceding Segment** | **N** | **% Deleted** | **Factor weight** |
|---|---|---|---|
| Three shared features: | | | |
| /t, d/ [+cor, -son, -cont] | (categorical absence by geminate prohibition, i.e. 1.0) | | |
| Two shared features: | | | |
| /s, z, ʃ, z/ [+cor, -son] | 276 | 49 | .69 |
| /p, b, k, g/ [-son, -cont] | 136 | 37 | .69 |
| /n/ [+cor, -cont] | 337 | 46 | .73 |
| One shared feature: | | | |
| /f, v/ [-son] | 45 | 29 | .55 |
| /l/ [+cor] | 182 | 32 | .45 |
| /m, ŋ/ [-cont] | 9 | 11 | .33 |
| No shared features | | | |
| /r/ ? | 86 | 7 | .13 |
| vowels --(nearly categorical retention, i.e. 0.0) | | | |

**Table 7. -t,d deletion by preceding segment classes**
(from Guy and Boberg 1994)

One of the remarkable points about this finding is that the effects of all the individual features involved are approximately equal. The differences between the various two-feature and one-feature classes in Table 7 are not significant and a feature-by-feature analysis given in Table 8 reveals that all favoring features get a Varbrul weight of about .64. (The slightly lower value for [son] is due to unavoidable interaction in the Varbrul algorithm between that feature and the [voice] feature.) Hence it is not the NATURE of the features that affect deletion, rather it is simply their identity with the target. The effect is simple and cumulative: the more features shared by target and trigger, the more like deletion will occur.

| **Features of Prec. Seg.** | | **Factor Weight** |
|---|---|---|
| Sonority | [-son] | .58 |
| | [+son] | .42 |
| Coronal Place | [+cor] | .65 |
| | [-cor] | .35 |
| Continuancy | [-cont] | .65 |
| | [+cont] | .35 |
| Voice (prec. sonorants only) | [αvoice] | .64 |
| | [-αvoice] | .36 |

Features of deletion target: /t,d/ = [-son, -cont, +cor, αvoice]

**Table 8. Feature analysis of preceding segment effect**
(from Guy and Boberg 1994)

Now, if the patterns in Tables 7-8 arise not because of the OCP constraint on competence, but stem from a separate performance OCP constraint, it should come as a theoretical surprise, a random coincidence that the two are so similar in nature and direction of effect. Pursuing the competence/performance distinction throughout our accounts of linguistic facts would probably lead to such twinning of constraints on an elaborate scale. Each competence constraint, summarizing invariant facts, would have a separate but equal performance twin, which accounted for variable facts. Such a result is manifestly absurd and we should be deeply suspicious of a conceptual framework that leads to it, if only by Occam's Razor; such theoretical *apartheid* has no more merit than its social counterpart.

**Grammar competition.** Finally, let us consider a third approach to the justification of invariant competence. This one is actually fairly old and has been amply debated in sociolinguistics and related fields (see, for example, the 'mixed grammars' explanation of creole continua), but is enjoying new attention among theoretical linguists, largely because of the rise of Optimality Theory. This approach hypothesizes that several different categorical grammars are involved in the production of variability, with each grammar generating discretely and invariably a different one of the variable outcomes. As Kiparsky puts it in a recent formulation (1993, 1994): "Assumption: Variation comes from competition of grammatical systems (in the individual or in the community), not from a probabilistic component in the rules of the language."

In some versions of this position whole competing grammars are envisioned; other formulations, for cases where only small segments of the grammar differ, postulate alternating or underdetermined specifications within the otherwise categorical grammar. But in either case, the multiple grammars, or the multiple states of the underdetermined grammar, generate different, discrete outputs under similar conditions; these appear superficially to involve variation as we know it — quantitatively conditioned alternations between related linguistic elements. However, in this model variation is really located in the selection from among these competing grammars, or the instantiation of the various possibilities in the underdetermined grammar.

This kind of approach has been particularly evident in recent treatments of variable phenomena within the framework of Optimality Theory (OT). As a sizable number of scholars have noted by now, the constraint hierarchy postulated in OT lends itself well to modelling variation, because the theory does not impose any principled limitations on what sequential orders the various constraints adopt; indeed the theory explicitly attributes differences between languages to different orders of the universal constraint inventory. If such is possible for different languages, why not for different regional dialects, social dialects, stylistic differences, or within-speaker variability generally? A series of papers in the last two years have explored this option. Iverson and Lee (1994) attribute differences between two dialects of Korean and between two stylistic varieties within one dialect, to differing orders of a pair of constraints. Rose (1995) similarly accounts for dialect differences in Gurage and Anttila (1994, 1995) takes this approach for variable productions in Finnish.

Several papers in this vein have made specific quantitative predictions. In Kiparsky's recent work referred to above, the general quantitative prediction is formulated as follows: "the more grammars that derive a form (i.e the more systems of ranked constraints in which it is the optimal output) the more frequent it is." This provides a kind of probabilization of outputs by means of variable constraint orders within the individual. On any given occasion the speaker will presumably select one particular order and give a deterministic output, but since the individual commands a number of orders, the overall distribution of outputs will depend on how many of these orders select each form.

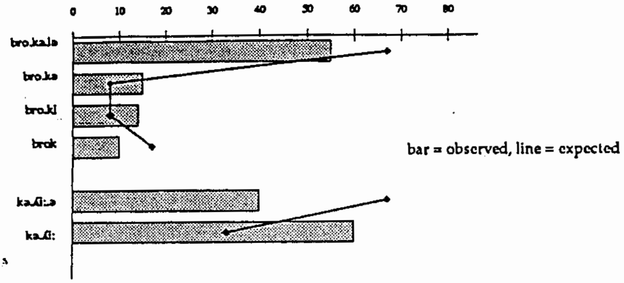
I will exemplify this approach with data from Nagy and Reynolds (1994), whose analysis is very similar to Kiparsky's. In their model, certain optimality constraints can float within a certain range on the hierarchy. At various points in its range it licenses different surface outputs, depending on what other constraints it eclipses or is eclipsed by. Counting up licensed forms across all possible orders appears to correctly predict the observed rates of apocope in Nagy's Faetar data. Their results are illustrated in the Tableaus in Figure 3. The constraint ALIGN-PROSODIC.WORD can rank anywhere from above NO.CODA to below *SCHWA and HNUC, which also may vary in order. This formulation yields 12 possible constraint orders, each of which would deterministically select an optimal candidate form. I give example tableaus for a couple of these orders. The complete set shows that, of the 12 possible orders given by their floating constraint, 8 of them select /brokələ/, two select /brok/ and one each pick /brokl/ and /brokə/. This predicts the outputs should occur in the fractions: 2/3, 1/6, 1/12, 1/12. The graph at the bottom of the figure shows the actual frequency counts of the forms and they correlate fairly well with the predictions. Unfortunately, for some other words they look at, the correlations between model and data are not so good. I've included their figures for one other word, /kelyia/, where their procedure predicts proportions that are diametrically opposed to the observed data. Nevertheless, the results to date merit further investigation.

**Figure 3: Optimality analysis of apocope in Faetar**
(from Nagy and Reynolds 1994)

$$\text{*CXONS} \gg \left\{ \begin{array}{c} \cdots\cdots\cdots\cdots \text{ALIGN-PRWD} \cdots\cdots\cdots\cdots \\ \text{NOCODA} \gg \text{PARSE} \gg \text{ONS} \gg \left\{ \begin{array}{c} \text{*SCHWA} \\ \text{HNUC} \end{array} \right\} \end{array} \right\}$$

| Input: /bro.kə.lə/ | LEX WORD | FILL | *CXONS | *CODA | ALIGN PRWD | PARSE | ONS | HNUC | *SCHWA |
|---|---|---|---|---|---|---|---|---|---|
| brok | | | | *! | | *** | | | |
| -> bró.kə | | | | | * | ** | | | * |
| bró.kl | | | | | * | ** | | *! | |
| bró.kə.lə | | | | | **! | | | | ** |

| Input: /bro.kə.lə/ | LEX WORD | FILL | *CXONS | *CODA | PARSE | ALIGN PRWD | ONS | *SCHWA | HNUC |
|---|---|---|---|---|---|---|---|---|---|
| brok | | | | *! | *** | | | | |
| bró.kə | | | | | **! | * | | *! | |
| bró.kl | | | | | **! | * | | | * |
| -> bró.kə.lə | | | | | | ** | | ** | |



These approaches mark a significant departure in the theoretical treatment of variation, in that they seriously engage the quantitative findings and seek to offer explanations of non-categorical regularities in language use. This is a very important step towards reconciling competence and performance and turning away from the misuses of this distinction that I have outlined above. However, I think the basic strategy, which relies on mixing the output of contrasting discrete grammars to generate the observed ratios, is flawed. There are several reasons for this. First, I find it implausible to suggest that human speakers would construct multiple grammars to encompass the facts of one language; I strongly suspect that all our mental apparatus relating to the use of language is highly integrated. Now this particular objection is partially obviated by the reformulation in Kiparsky and Nagy and Reynolds that does not postulate entire grammars but merely alternating or underdetermined versions of some subset of the grammar (in the OT analyses, this is the varying orders of a subset of constraints). This strategy clearly locates variation WITHIN the grammar, but tries to eliminate quantification. More accurately, it reduces variation to random selection and derives specific quantities in the outputs as an epiphenomenon, arising haphazardly from the way the constraint hierarchy determines optimal forms. This result is a sophisticated version of the 'free variation' concept: outputs alternate randomly. As noted above, that explanation fails because variation is not random. I strongly suspect that this OT version will fail for the same reason. It remains to be seen whether quantitative results along these lines will work out more like Nagy and Reynolds' results for /ƀrokələ/, which fit well, or like their results for /kelyia/, which are very poor. (I should add, however, that Antilla's Finnish results, using a similar model, do give an extremely impressive conformity between predicted and observed frequencies.)

A further reason for concern about the OT approaches is based on the kind of evidence that we have seen in Table 7. The difference between the various categories there does NOT arise from differing orders of two or more constraints, rather it arises from different numbers of times a SINGLE constraint is violated for a given form. The full range of facts in Table 7 can be handled as part of a single cumulative sequence: preceding vowels share NO features with the target -t,d and hence never incur OCP violations and trigger no deletion; then with increasing numbers of shared features there is increasing deletion, until all features are shared with a preceding -t,d at which point there is categorical prohibition. But this is not because OCP, or the competence grammar to which it belongs, is intrinsically categorical; rather it is merely the upper end of a quantitative continuum.

If we try to account for this result in an OT analysis, we do NOT get a correct quantitative prediction by mixing the outputs of discrete grammars. The most likely candidates for OT constraints governing this alternation would be some version of the OCP, prohibiting same-tier sequences of adjacent identical features, say *[αF] [αF] and the PARSE constraint, which accords higher optimality to those outputs that instantiate more of the underlying form. If these two constraints varied in order, we'd have either OCP >> PARSE, which would select categorical deletion, or PARSE >> OCP, which selects categorical retention. Therefore, if speakers selected randomly between these orders, we would predict a 50% deletion rate predicted for ALL word-final -t,d's preceded by any consonant , regardless of what that consonant was. Instead, as we have seen, the data clearly show two discrete levels of deletion, differentiated by the number of features that violate the OCP. Since this pattern arises from more or fewer violations of ONE CONSTRAINT, there is no way to twiddle the order of that constraint to generate these results.

Now, there is one way around this difficulty via a device that has been proposed in OT, which is to 'explode' a unitary constraint into a 'family' of constraints, which may then each occupy a dişcrete point in the hierarchy. Thus OCP could be exploded into a set of constraints, perhaps one for each feature, so that there would be an OCP-cor, OCP-son, OCP-cont. This would quantitatively differentiate the relevant sets of preceding segments, but not in a way that was consistent with the data in Tables 7-8. If the three OCP-quarks and PARSE were randomly ordered, there would be 24 possible orderings; these would select deletion for words with one shared feature between target and trigger 12 times, while for two-shared-feature cases, deletion is selected 16 times. In Nagy and Reynolds' approach, this would predict deletion percentages of 50% and 66.7% respectively, which are at odds with the observed values of 30.5% for the one-feature cases and 45.4% for the two-feature cases. Furthermore, the method does not predict categorical absence of the geminate cases: deletion in a -tt cluster would only be selected in 75% of the tableaux.

The more serious objections to this procedure, however, are logical ones. Exploding the OCP loses the very generalization that it, or any constraint, is supposed to capture and if OT allows unlimited decomposition of its putatively universal constraints, it will ultimately become so flexible as to be effectively content-free. Furthermore, the explosion procedure still fails to explain the quantitative uniformity of the OCP effect: Tables 7 & 8 show that that the identity of the violated feature makes no difference in the deletion rate. If the several OCP quarks are disconnected from each other and permitted to occupy different points on the constraint hierarchy, the theory offers no principled reason why they should not be subject to different limitations on their hierarchical position, hence yielding different quantitative effects. Only the concept of a single unified OCP that has a constant effect regardless of what particular feature sequence violates it correctly predicts the quantitative results. So I conclude that an adequate grammar must, at some point, require a quantitative component, which recognizes that two violations of a constraint are worse than one violation. And if we must have some quantitative component in the grammar, we have come around to the other side of the variationist position articulated by Weinreich, Labov and Herzog, namely that variability is INHERENT in language, that language users are exquisitely sensitive to frequency and that the manipulation of frequency is part of competence.

### 3.3 A post-Saussurean synthesis of langue and parole

To summarize the foregoing discussion, variation studies show:

- the existence of orderly heterogeneity,
- that this order is governed by the same principles that are used in theories of competence to account for invariant facts,
- and that theoretical devices for excluding orderly heterogeneity from the grammar either fail, or undermine the very invariance and universality that they are meant to defend.

The consequences of these findings are that a dichotomy that postulates that order is a property of langue and disorder a property of parole, cannot be sustained; orderly hetero–geneity, locates regular, systematic structure within the domain of parole. Furthermore, the evidence for inherent variability locates variability and heterogeneity within the domain of langue/competence. So ultimately the dichotomy fails, or at least fails to be a very useful tool at our present state of understanding. It doesn't buy us much and it hinders progress on understanding languge as an integrated phenomenon.

## 4 Design principles for an integrated theory

In sections §2 and §3 I have argued that the conceptual distinctions of Saussurean linguistics, the oppositions of synchony/diachrony and langue/parole, are both contradicted by the facts of linguistic variation. Neither dichotomy permits an integrated account of the facts and both obscure rather than clarify important aspects of the nature of language. Therefore, instead of wasting further efforts debating dichotomous positions like those quoted above from Saussure, Chomsky and Kiparsky, I think linguistics needs to move towards a conceptual framework that achieves a Hegelian synthesis of these two pairs of theses and antitheses. In other words, we must adopt a Post-Saussurean view of language and linguistics. I would like to suggest some design principles for such an integrated view.

First, such an integrated theory will declare itself responsible for language use as well as for hypothesized abstract knowledge and will seek to account for the broadest possible range of facts about language. It will thus address Hymes' 'communicative competence' as well as Chomsky's grammatical competence and also linguistic performance of all kinds.

Second, an integrated theory must be designed to account for both the invariant aspects and the patterned variability of language. It will therefore embrace orderly heterogeneity and inherent variability as central features of human language. Indeed, these steps should not be seen as radical departures for the field, but rather merely as a belated recognition of reality. This is a minimum requirement for linguistics to achieve the most elementary level of 'observational adequacy'.

Third, in recognizing variability, it seems inevitable that an integrated theory will require non-categorical but non-random statements and properties — in other words, quantification. On present evidence, the appropriate quantification will be probabilistic.

Fourth, a coherent theory of language must account for both change and stability. This is intimately related to the treatment of variation, in that change and variation are in an important sense the same thing, merely viewed from different temporal points of view.

Fifth, in recognizing change, we will necessarily have to develop linguistic descriptions that incorporate a dynamic, directional element, that contain, in effect, linguistic vectors, combining a description of a state with an account of the forces that incline those states toward change and indicate directions of movement within the system.

Sixth and finally, an integrated theory must tackle head-on the question of similarity and difference at the level of GRAMMAR. What range of diversity can be accommodated by a single grammar? When do accumulated differences become so substantial, either in the course of change, or in the comparison of dialects, that it becomes necessary to treat the varieties involved as having different grammars? And what are the intermediate stages between 'same' and 'different'? Can we speak of grammars of an entire speech community, or of 'pan-lectal' grammars as C.-J. Bailey does, or even of 'pan-temporal' grammars? The opposition between the continuous and the discrete is still another dichotomy requiring synthesis.

Many papers at this conference (NWAVE-24) are heading in this direction; indeed, NWAVE has always been a forum in which Post-Saussurean elements could be found. But my point is that we cannot continue to make progress on many important issues while we continue to phrase our questions and analyses in Saussurean terms. It is unproductive, for example, to see ourselves as working on *parole*, while other scholars study *langue* in a separate and possibly unrelated way. Such a conceptualization may redress the historical imbalance, but it does not lead to an integrated understanding.

## 5 Conclusion

To conclude, one could say that linguistics has placed itself in a situation rather like that of the group of blind men who encountered an elephant. One of them gets ahold of the tusk and concludes that elephants are hard and pointy; another grasps a leg and concludes that elephants are tall and cylindrical, like tree-trunks. A third touches the trunk and says that elephants are sinuous and twisty, like snakes, while a fourth feels the tail and concludes that elephants are thin and hairy, like ropes. Linguistics has this same kind of fragmented view of language and a principle cause of this fragmentation is that Saussurean linguistics has made a virtue of segregating the field into different disciplines each grasping different pieces of the beast. It is time to take off our blindfolds and begin to see the whole elephant.

## References

Anttila, Arto (1994). "Deriving variation from grammar: A study of Finnish genitives." Paper presented at NWAV 23, Stanford University.

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

Guy, Gregory R. (1980). "Variation in the group and the individual: the case of final stop deletion", in William Labov, ed., 1980, *Locating Language in Time and Space* (New York: Academic Press), 1-36.

Guy, Gregory R. (1991). "Explanation in variable phonology: an exponential model of morphological constraints." *Language Variation and Change* 3:1-22.

Guy, G.R., B. Horvath, J. Vonwiller, E. Daisley and I. Rogers (1986). "An intonational change in progress in Australian English." *Language in Society* 15, 1:23-52.

Guy, Gregory R. and Charles Boberg (1994). "The obligatory contour principle and sociolinguistic variation", *Toronto Working Papers in Linguistics: Proceedings of the CLA 1994 Annual Meeting* (Toronto: University of Toronto Press).

Iverson, Gregory K. and Shinsook Lee (1994). "Variation as optimality in Korean cluster reduction." Paper presented at ESCOL-94 (Eastern States Conference on Linguistics), University of South Carolina.

Kiparsky, Paul (1982). *Explanation in Phonology* (Dordrecht: Foris).

Kiparsky, Paul (1993). "Variable rules." Paper presented at ROW-1 (Rutgers Optimality Workshop), Rutgers University.

Kiparsky, Paul (1994). "An OT perspective on phonological variation." Paper presented at NWAV-23, Stanford University.

Kroch, Anthony (1978). "Towards a theory of social dialect variation." *Language in Society* 7:17-36.

Labov, William, Malcah Yaeger and Richard Steiner (1972). *A quantitative study of sound change in progress* (Philadelphia: U.S. Regional Survey).

Nagy, Naomi and William Reynolds (1994). "Optimality theory and variable word-final deletion in Faetar." Paper presented at NWAV-23, Stanford University.

Naro, Anthony J. and Miriam Lemle (1976). "Syntactic diffusion", in S.B. Steever et al., eds., *Papers from the Parasession on Diachronic Syntax* (Chicago: Chicago Linguistic Society), 221-40.

Naro, Anthony J. (1981). "The social and structural dimensions of a sound change." *Language* 57:63-98.

Rose, Sharon (1995). "Dialect variation and transparency: The role of constraints." Paper presented at the Canadian Linguistics Association, Université du Québec à Montréal.

Saussure, Ferdinand de. (1983). *Course in General Linguistics* (London: Duckworth).

Tarallo, Fernando (1995). "Turning different at the turn of the century: 19th Century Brazilian Portuguese", in G.R. Guy, C. Feagin, D. Schiffrin and J. Baugh, eds.,*Towards a social science of language: Papers in honor of William Labov* (Amsterdam and Philadelphia: John Benjamins), 199-220.

Weinreich, Uriel, William Labov and Marvin Herzog (1968). "Empirical foundations for a theory of language change," in W. Lehmann and Y. Malkiel, eds., *Directions for Historical Linguistics* (Austin: University of Texas Press).