Explanation in Historical Linguistics

Edited by Garry W. Davis and Gregory K. Iverson

Offprint
Diachronic explanation:  
Putting speakers back into the picture

BRIAN D. JOSEPH
The Ohio State University

1. Introduction

I begin with a parable. The course numbering system for linguistics courses at The Ohio State University (OSU) includes the following course listings:

(1) Linguistics 201 — Introduction to Language  
    Linguistics 601 — Introduction to Linguistics  
    Linguistics 261 — Language and Social Identity  
    Linguistics 661 — Introduction to Sociolinguistics  
    Linguistics 861 — Seminar in Sociolinguistics  
    Linguistics 271 — Language and Mind  
    Linguistics 671 — Introduction to Psycholinguistics  
    Linguistics 871 — Seminar in Psycholinguistics  
    Linguistics 609 — Morphology  
    Linguistics 809 — Seminar in Morphology.

These courses admit of two generalizations, a “topic generalization” by which the general topic is indicated by the tens and digits (e.g., -09 for morphology, -71 for psycholinguistics, etc.) and a “level generalization” by which the level is indicated by the hundreds place, i.e., 200 = lower level undergraduate, 600 = upper-level undergraduate and beginning graduate, and 800 = upper-level graduate (seminar). At the same time, though, there are some other courses, given in (2):

(2) Linguistics 801 — Historical Linguistics I  
    Linguistics 802 — Historical Linguistics II
Linguistics 602 — Introduction to Syntax
Linguistics 820 — Seminar in Syntax

that deviate from this pattern; for example, the pairs 601/801, 602/802, and 602/820 do not follow the topic generalization evident in (1).

As it turns out, the pattern noted is one that to a certain extent was deliberately imposed on the courses, in two ways. First, the pattern with regard to level is dictated by the University and so cannot be violated; it is simply a general fact about courses at OSU. Second, the pattern with regard to topic is the product of careful planning and conscious decision-making by the framers of the courses in 1988, since Linguistics 809, 861 and 871 are the most recent additions to the course list; in a like manner, the anomaly of the 601/801 pair stems from the fact that Linguistics 801 was created many years ago before there was any real pattern to the course numbering, and so on.

To bring this parable around to the matter of explanation in historical linguistics, the following question can be asked: Which account gives the better explanation of why the courses are numbered as they are? Does the diachronic account explain the distribution better, taking into consideration as it does when each course was created and what went into the decision to number it in a particular way, or does the synchronic account, which aims for a generalization that is currently valid concerning the relation between and among the different course numbers? One might even say that the “level-generalization” is in a sense irrelevant, since it is part of an obligatory pattern — the “universal grammar” of course numbering at OSU — to which all OSU courses must conform to be acceptable OSU courses, so that the question really boils down to which account of the topical pattern is more explanatory.

In a sense, both accounts have validity and both offer insights into the situation. The synchronic account certainly provides a true generalization — for at least one layer of course offerings — and the diachronic account does indeed give some insight into why such a generalization is possible and why there are exceptions to this generalization. Thus it would seem that neither account alone explains everything one might want to know about this (fascinating) topic, and one might further conclude that the two accounts actually complement one another nicely in adding to the overall understanding of this (crucial) area of inquiry.

It should be clear what the point of this parable is, for when one turns to real language phenomena, the same situation is found: while it might be said that there is always an historical explanation for some particular state of affairs in a language, it is equally true that relying solely on historical explanation means ultimately that an historical explanation is no explanation (so Janda (1990)); if everything is to be explained in that way, then there is no differentiation possible among various synchronic states even though some conceivable synchronic states never actually occur. Moreover, universal grammar, to the extent that it can be given meaningful content, is in a sense achronic, for it is valid at particular points in time for any given synchronic stage but also valid through time in the passage from one synchronic state to another. Thus in addition to the diachronic perspective, a synchronic perspective is needed as well, for that enables one to say that a given configuration of facts exists because it gives a reason for the existence of a pattern — a generalization over a range of data — in a given system at a given point in time.

It is well-known of course, as Labov (1989) has recently reiterated, that there is a tradition of separating synchrony out from diachrony in linguistic scholarship, and of treating these two aspects of linguistic investigation as important each in its own right but collectively opposed to one another in a “ne’er-the-twine-shall-meet” relationship. As Labov (p. 85) notes, for instance:

one of the strongest arguments for the separation of synchronic and diachronic linguistics is that children do not know the history of the language they are learning. As the grammar of the language must be the rule system that is learned and internalized by the language learner, and the child is ignorant of its history, it follows that historical linguistics is irrelevant for students of synchronic linguistics.

Labov was summing up a position that he himself does not necessarily hold to, but rather one that is prevalent in the linguistic community. However, despite this prevailing view, there are strong indications for the complementarity of the two enterprises, and for their symbiosis.

While it is undeniable that speakers impose a synchronic analysis of some sort on the forms of their language, i.e. that they organize their linguistic data synchronically, it is fair to say that synchronic states and diachronic changes have a special relationship. Trivially, in order to understand diachrony, it is necessary to understand what the starting point and what the end point of a given state of affairs are like, and that understanding requires a synchronic perspective over each point. Also, though, as is well known from the work of Labov and others following him, synchronic
variation is often the starting point of language change — though stable variation is certainly possible and may even be the norm, as noted by Labov (1989:87) — and can be the endpoint of change as well, as is the case with the predictable variation that arises in instances of conditioned phonological splits, or the less predictable variation that results from the analogical spread of a new inflectional marking that might give, for example, lexically determined variation within a subclass.

This proposed complementarity of synchronic and diachronic accounts should not be taken to mean the two are the same — they are different sides of the same coin perhaps, but they are not identical by any means. It is reasonable to suppose, for instance, that they are subject to different types of evaluation metrics: a synchronic account should aim for economy and typically, for example, avoids positing the same rule or constraint at two different points in a grammar or derivation (though the validity of the usual interpretations of “economical” in this context is far from a foregone conclusion, as suggested by Janda & Joseph (1987, 1991)), while a diachronic account, being interested in determining what actually happened over some time interval, should aim for the truth, even if it is messy and even if it might entail positing the operation of, for instance, the same sound change at two different adjacent time periods, perhaps 50 or 100 years apart.

Moreover, and this point relates to the notion of universals as achronic, what is diachrony except a succession of synchronies, a movement through successive synchronic states, a situation that can be represented schematically and somewhat idealistically — in that it omits synchronic variation — as in (3)?

(3)  

\[
\begin{array}{c|c}
D & L_1 \text{ Synchronic Stage 1} \\
I & L_2 \text{ Synchronic Stage 2} \\
A & L_3 \text{ Synchronic Stage 3} \\
C & L_4 \text{ Synchronic Stage 4} \\
H & \vdots & \text{Synchronic Stage } n \\
R & \ddots & \text{Synchronic Stage } m+1 \\
O & \ddots & \text{Synchronic Stage } n+1 \\
N & \vdots & \text{Synchronic Stage } n \\
Y & \vdots & \text{Synchronic Stage } n+1 \\
\end{array}
\]

Such a view provides a basis, for example, for treating the evidence from language change as being relevant to the development of linguistic theory, i.e. to an understanding of the ways in which languages are alike and the ways in which they differ, for the constraints imposed by universal grammar at any synchronic stage will necessarily then be the same constraints that govern the passage from one state to another, i.e. diachrony. Diachronic principles or generalizations do not exist, then, outside of the synchronic processes of grammar formation at synchronic stage after synchronic stage.

More important, though, this view allows for the correction of what seems to be a serious flaw in many accounts of language change, namely the fact that all too often, even though language change necessarily involves change in the speech habits of speakers, linguists have treated speakers as if they were somehow irrelevant to the process of language change. The view espoused here of the relation between synchrony and diachrony is entirely consistent with the notion stated in Joseph & Janda (1988:194), and most recently reiterated in Janda (1990), that language change always takes place in the present, i.e. it always occurs in some speaker’s (or group of speakers’) present, and thus will always be subject to the constraints that shape synchronic grammars. More significantly, too, language change will always be shaped by what speakers do in their grammar formation process. Perhaps framing the question in this way makes it seem very obvious, almost trivial and tautologous, that speakers must have an important role, but it is not at all obvious, it seems, to linguists who operate with a notion of “diachronic principle” or “diachronic tendency”, independent of substantive language universals, that somehow hovers over language use over extended periods of time and guides the way language develops.

When taken seriously, this view that language change takes place in the present for some (group of) speakers turns out to have important consequences. A good starting point for the discussion is provided by the observation that for all the fact that language change involves changes in the ways real speakers use their language, various methodologies in historical linguistics and schemes for diachronic explanations have treated speakers as if they were somehow beside the point. Several case-studies can then be considered which suggest that there is a real need to bring speakers back into the diachronic picture, and the consequences of such a decision for the proper view of synchronic linguistics can be explored.
2. Methodological flaws tied to the failure to recognize the role of speakers

As mentioned above, various specific methodologies and general approaches in historical linguistics have fallen into the trap of neglecting the role of the speaker. Thus, despite the ground-breaking study of Labov (1963), which enriched to such a great extent the understanding of the spread of change and more generally of the relevance of social factors — and therefore of individuals in social groups — in language change, the "language change as grammar/rule change" approach of generative historical linguistics, as epitomized by King (1969) but as practiced by many others in the 1960s and early 1970s, treated change as driven more by abstract properties of grammars, such as formal simplicity of rule statements or ordering relations between rules, than by speakers' actual linguistic behavior, e.g. as represented by the surface forms they produce. Although this is not the place to review the relative merits of the different approaches, the summary comment found in Labov (1989:96) bears repeating: "there is little evidence to support the notion of a language-learning faculty isolated from social and historical developments", i.e. attempting to study language change and development without reference to social factors, and therefore to the role of the individual speaker in a speech community, is pointless.

2.1 A similar problem with another specific method in historical linguistics has been pointed out by Janda (1990). He notes that glottochronology must be considered suspect since it rests on the assumption of a panteleological perspective that no speaker of a language could possibly take, i.e. how could a speaker of English in 1991, for instance, know whether he or she must replace so many lexical items in order for English over a thousand-year period (a period starting and ending when?) to meet its quota for lexical replacement as mandated by glottochronology? Language use in a speech community does not necessarily lend itself to probabilistic projections of the sort an actuary might develop for insurance risks, for example.

2.2 Another methodological arena in which recognition of the role of speakers has been late in coming is explanation via recourse to language contact. While this may not hold for all such accounts in all cases of language contact, in many well-known ones, for example those involving the Balkan Sprachbund, it is not uncommon to find facile invocations of situations being caused by contact between language X and language Y. An example is Weigand (1925:viii) who explained various "non-Slavic" features in Bulgarian, such as the postposed article, the restricted use of the infinitive, and the loss of case-inflexion in the nominal system, as being the result simply of Romanian influence:

Alles das, was dem heutigen Bulgaren seine Sonderstellung unter den slavischen Sprachen gibt, wie nachgestellter Artikel, Beschränkung im Gebrauch des Infinitivs, Verlust der Kasusflexion und manches andere kann nur durch rumänischen Einfluss in das Bulgareische gelanget sein, denn das Altbulgarische kannte dies alles noch nicht.

Such an attribution of the influence of one language over another seems plausible in the abstract — after all, early stages of Bulgarian did not have the typical Balkan features that later stages attest, while Romanian may have developed those features relatively early on in its development from Vulgar Latin — but founders in the concrete. In general, such accounts fail to recognize that the contact is not really between the languages but is rather actually between speakers of the languages in question, and thus factors such as the prestige of one language in the other speech community, the communicative necessities that would arise in a contact situation, and the effects of language shift, are often not considered. In this case, it is not at all clear that Romanian ever had a position of influence and prestige among Bulgarian speakers that would have allowed for the introduction of Romanian features into the native usage of Bulgarians, and moreover, at least one of the features shared by these two languages, e.g. the postposed definite article, seems not to have a connection with communicative necessities that would arise in a contact situation; finally, while a language-shift explanation, better known as a substratum account, could explain the postposed article, there is no direct evidence for such a shift in the case of Bulgarian and Romanian.

There are of course exceptions to this trend towards nonexplanation in the Balkans without consideration of speakers. Civ’jan (1965:9), for example, has stated that "the stimuli for the formation of the [Balkan Sprachbund] are to be sought not in history, but in the daily necessity of the use of language as a means of communication". Also, the account of the restriction of the use of the infinitive in the Balkans presented in Rozenericg (1969, 1976) — an account which, as discussed by Joseph (1983:201-209), is the most plausible starting point for understanding this phenomenon — is predicated on the communicative needs that emerge in
bi- and multi-lingual contact situations; as Rozenvejg (1976:42) puts it:

the period of transition [from] ... Common Slavic to the South Slavic languages ... witnessed not only a great increase in the numbers of speakers, but a great change in their ethnic composition as well. In short, a basically monolingual society was becoming a bi- or multilingual one.

Such an account thus provides the point of departure for the incorporation of more sociolinguistically realistic models of the Balkan Sprachbund. Still, one general consequence of accounts such as Weigand's and others like it has been an overemphasis on comparisons of standard languages rather than regional dialects, even though the contact between individuals, in certain parts of the Balkans at least, more typically involved nonstandard dialects.

2.3 A similar situation is found with substratum explanations as a subclass of language contact accounts. Most typically, in putative substratum cases, it often turns out that no evidence of the substratum effect is apparent until long after the putative contact took place; for instance in some accounts of the loss of the infinitive in Greek, the substratum effect would have to have persisted for several thousand years before surfacing. Now, while Labov 1989 has argued that certain nonfunctional constraints on language use, e.g. on <t/d> deletion in English final clusters, can be transmitted diachronically, the transmission is from speaker to speaker, i.e. within a social context, not imposed onto speakers by some disembodied abstract entity “LANGUAGE”.

A similar example is the case of the weakening of gamma (the voiced velar stop) in Pamphylian Greek in words such as the adjective mégas ‘great’ and derivatives (e.g. proper name Méds for Megás), which Szemerényi (1968) has suggested is the result of a Luwian substratum in southwest Asia Minor. Here the substratum can be identified and there is even some reason to believe, based on onomastic evidence, that there were Luwian speakers in that part of Asia Minor, possibly even at the relevant time, and that Luwian did weaken voiced velars in the context /e- _a/ (e.g. if the derivation of Cuneiform Luwian tiyammi ‘earth’ from *dheg’hom- is correct). Thus, a substratum account in this case at first glance meets some of the typical objections. However, as Wallace (1983) has shown, that is not enough. The substratum account could only make sense if the conditions on g-deletion were identical in the two languages, so that speakers of Luwian (or a Luwian-derived language such as Lycian) in shifting to Pamphylian Greek carried over their native speech habits into their pronunciation of Greek, but in fact, as Wallace demonstrates, the contexts are not identical; Luwian g-deletion occurs word-initially (e.g. Cuneiform Luwian ıssarı- ‘hand’ from *g’hesr-) and post-consonantally (e.g. Cuneiform Luwian parri- ‘high’ from *bherg’h-i-), while Pamphylian retains initial g-, as in geras ‘gift of honour’. Thus here too, the recognition of the fact that contact takes place between speakers, thereby putting a speaker-oriented interpretation on what a substratum entails, leads to a better understanding of what actually happened, in this case that the Luwian change and the Pamphylian change are unrelated to one another. The undesirable alternative that might come from treating language contact as something that takes place between languages as opposed to speakers is that one might be led to say, erroneously, that the Pamphylian change could have been the borrowing, in some sense, of a sub-part of the Luwian g-deletion process.

3. Still further evidence — what speakers do versus what linguists do

These examples from language contact phenomena show the potentially detrimental effect that can arise from overlooking the role of the speaker. Similarly, even when looking at the internal history of a language, even when examining a language on its own terms, it cannot be denied that linguists can learn from looking to speakers, in particular to the evidence that speakers’ analyses present, for these analyses often show real linguistic insight as speakers’ creativity in dealing with the data of their language comes to the fore. Speakers can and do come up with analyses that make sense and to a certain extent are almost forced on them by a certain configuration of facts, and such reanalyses often tell the linguist just what is, or will be, going on in the language. For example, as noted by Jeffers & Lehiste (1979:64-5) and discussed more extensively by Anderson (1988:329-331), after the effects of a sound change reducing final syllables there was a reanalysis that yielded the plural marker -ir in Old High German out of a neuter noun-forming suffix *-iz- in Proto-Germanic, deriving from Proto-Indo-European *-es-, and an earlier marking of -Ø versus *-fern for singular versus plural; the new plural marker -ir could then spread, as it did in the case of the originally zero-marked plural wort ‘words’, which gave later (e.g. Modern High German) Wörter, with the -ir- suffix. The steps for this reanalysis can be summarized as follows, with the noun lamb- ‘lamb’ as
exemplar:

\[
\begin{array}{lcl}
\text{Stage I} & \text{Stage II} & \text{Stage III} \\
\text{SG *lamb-iz-Ø} \rightarrow & \text{lamb-Ø-Ø} \rightarrow & \text{lamb-Ø} \\
\text{PL *lamb-iz-a} \rightarrow & \text{lemb-ir-Ø} \rightarrow & \text{lemb-ir} \\
\end{array}
\]

Given that the only formal difference between singular and plural at Stage II was zero versus -ir, and that the Ø’s in the display have no real content serving only to indicate potential slots for morphemes, it is easy to see how a speaker might have been led to analyze -ir as the marker of plural number; in fact, one might well wonder about a speaker who did not make such an obvious analysis.

Similarly, it is worth considering in this context the phenomenon of folk etymology. This type of change is one in which speakers do not “perform” as linguists might expect, in that properly monomorphic words can be analyzed as polymorphemic or opaque polymorphemic words can be analyzed in a more transparent fashion; examples of the former type include sparrow-grass from asparagus, and woodchuck from Cree oček, while examples of the latter type include hangnail from earlier (Middle English) agnail (from an Old English form meaning ‘painful prick in the flesh’), and the like. Folk etymology often represents a reasonable attempt on a speaker’s part to make sense of, i.e. to render transparent, a sequence that is opaque for one reason or another, e.g. because it is a borrowing and thus has no synchronic parsing in the receiving language. As such, it shows speakers actively working to give an analysis to data that confronts them, even if such a confrontation leads to a change in the input data. Moreover, folk etymology demonstrates that speakers take what the surface forms are — an observation which becomes important later on as well — and work with that, so that while they are creative, they are not really looking beyond the immediate phonic shape — and, in some instances also, the meaning — that is presented to them.

Once speakers are put back into the picture, though, and their analyses of data become valuable for insights into how linguistic data is to be dealt with, then, based on examples such as those to be presented below, it becomes clear that even though the analyses often show insight and creativity, speakers are not what Lightner (1975:634-5) once referred to as “perfect” speaker/hearers, i.e. speakers who know everything, literally, about their language, including all possible derivational relationships and thus all the etymologies of all words, and the like. As Lightner (1975:634) put it:

Chomsky has written (1965:3) that “linguistic theory is concerned primarily with an ideal speaker-listener ... who knows its language perfectly” ... But ... the problem [is one] of what it means to know a language “perfectly.” I think that, to understand this, one must take the task of writing a grammar to be the task of writing the most all-inclusive grammar possible; the result will presumably be something like the grammar of the ideal speaker-hearer who knows his language “perfectly.”

Contra Lightner, however, it can be argued that once one starts talking about what real speakers do, i.e. about “real people, real problems”, to borrow from the characterization of Watkins (1989:798), one has to reckon with speakers who, unintentionally of course, might analyze an element of their language in such a way that it differs from its historically justified analysis, even when the historically “correct” analysis is reasonably well justified synchronically as well. More important, though, is the fact that such reanalyses sometimes run counter to synchronic analyses that would appear to be fairly obvious in terms of their internal composition or derivation. In what follows, several cases are presented in which such transparent and seemingly well-motivated synchronic analyses are overlooked by speakers, who treat them as if they involved totally opaque segmentation, for instance, and end up reanalyzing the data in a way that, while certainly plausible, is not an analysis that a linguist necessarily would come up with.

It is important to note that reanalysis is quite common — as Janda (1984) pointed out, everything present in a language at a given time either derives from an inheritance, a borrowing, or a reanalysis of existing material, once these three sources are allowed to yield productive means of creating new linguistic expressions — and in addition reanalysis always involves a break with an analysis that is historically prior, and thus was the proper one at some pre-stage. Still, not just any reanalysis is of interest here. The obvious ones that are forced upon speakers, e.g. the German plural marker reanalysis and the various folk etymologies mentioned above, are perhaps what would be expected to occur, but the ones in which transparent connections are ignored in favor of more opaque connections reveal something very interesting about the limitations on speakers’ abilities to reach “perfection” in their linguistic competence.

3.1 The first case is a trivial, but nonetheless telling, instance of a lexical reanalysis. The expressions pitch-black and pitch-dark, for many speakers, are to be analyzed as having some relation with the phrases black as pitch / dark as pitch. Indeed, Lightner (1976:181-2) has argued that in an adequate
grammar of English, the compound must be so derived, in order to explain why "pitch [should] appear only before black and dark", and he asks "Why not before pink, white, and plaid? Why not before sincere, awkward, blue, plaid, and after go, think, escape?", answering his question with the suggestion that "this problem can be avoided if pitch-black/-dark are derived from something like black/dark as pitch, where pitch is a noun meaning 'a black or dark, viscous, sticky substance obtained as a residue in the distillation of organic materials, especially tars'" and continuing "under this analysis, the grammaticality of sequences like pitch-sincere ... will depend on the grammaticality of sincere as pitch"; similarly, he notes that given that sticky as pitch is well-formed, "therefore, pitch-sticky must be generated as grammatical". Despite Lightner's account, it must be pointed out that for some speakers of English, the expression pitch-black is not treated as derived from black as pitch, even though pitch is quite black and even though there are structurally parallel compounds consisting of a noun with a salient color property as first member and the relevant color as second member, such as snow-white or blood-red. Instead, these speakers have analyzed the pitch in pitch-black as nothing more than an intensifier, and thus are able to produce forms such as pitch-red for 'extremely red'. While one can certainly make sense of this analysis, and can see it as being motivated by the meaning — in that pitch-black indeed means 'very black', since pitch is so black — and perhaps by an unfamiliarity with pitch, still it runs counter to the obvious and perhaps the easiest analysis of the expression and thus is of interest to anyone exploring what the imperfect speaker, i.e. a real speaker, does with his/her language.

3.2 In a sense, then, any case in which synchronic traces are present in a language but are ignored by speakers will be of interest here. One fairly common occurrence of this type comes when the synchronic traces of earlier connections involve derivational relationships as opposed to paradigmatic relationships. In particular, it is not at all unusual for speakers to loosen derivational bonds while maintaining strong paradigmatic bonds between and among forms, so that analogies that operate intra-paradigmatically often do not extend to derivationally related forms. Two such cases from the history of Greek can be cited as examples.

3.2.1 For instance, in Pre-Greek, the paradigm for 'one' in the neuter gender included a nominative *hém and a genitive *hem-ós, with a relatively clear, though admittedly unproductive, relationship with the feminine form *hm-ía. However, when the neuter paradigm was changed by a regular sound change of m → n/-.# that affected the nominative and by an analogy (paradigmatic levelling) that affected the genitive, so that the forms hén / hen-ós resulted, the feminine was unaffected, remaining as *hmía (later giving the attested form mia). Similar developments are seen with kthôn / kthonós 'earth' (NOM/GEN), which derived from an m-stem noun (compare Sanskrit ksam-) but whose derived adjective kthamalós 'near/on the ground' was not affected by the levelling in the noun paradigm.

3.2.2 In like fashion, the Ancient Greek adjectival paradigm for 'fourth' included a masculine form (nominative) téartos and a feminine (nominative) téartē, in which an accent shift was induced by long ultima in the feminine. This accent shift was levelled out in Modern Greek, a language with no distinctive vowel quantity, in the adjectival paradigm, yielding téartos / téarti (MASC/FEM, with i in the feminine by regular sound change from earlier ē) but the form of 'fourth' in the derived expression téartē (hēméra) 'Wednesday' (literally 'fourth (day)') remained unaffected by the levelling, and came into Modern Greek as téarti 'Wednesday', with accent on the penultimate and not on the antepenultimate as in the related paradigmatic feminine adjective.

One might of course argue that in these cases the fact that these represent derivational relationships which are not productive is the key to understanding these facts. However, it can be countered that productivity is a relative matter, and that derivational relationships can often be quite salient even when unproductive. For example, even though VERB-NOUN compounds are not productive in English today, as the ungrammaticality of a putative noun *picknose 'someone who picks his nose' shows, many speakers nonetheless perceive a relationship between the compound agent noun pickpocket and the verb phrase to pick someone's pocket (though, when asked, my sons, aged 4 and 9 at the time, did not, partly, they said, because they had never heard the verb phrase before, offering instead a verb to pickpocket). What is apparent here is that speakers are focussing more on the immediate (the paradigmatic) than the distant (the derivational), just as in folk-etymology it seems that they focus on the surface forms of words, i.e. on the phonetic shape that is immediately before them.

3.3 Yet another example comes from the English hortative let's, which
seemingly derives in a straightforward way from a reduction of *let us*. However, when one examines a few additional facts concerning *let's*, they show together that this obvious analysis of *let's* is not the correct one. Presumably, then, at some point, speakers must have lost sight of the obvious connection and allowed *let's* to go its own way, and to drift apart from *let us*. Thus, there are several ways, some of them somewhat subtle in nature, in which the ostensibly contracted form *let's* differs from its putative source *let us*.

First, *let's* does not even behave like a contracted form. In particular, even though contracted forms generally do not occur before a “deletion site” (i.e. a zero anaphor), as in (5a), such is not the case with *let's*, as indicated in (5b):

(5)  a. Have you seen him? Yes, I have Ø / *I've Ø
b. Shall we go? *Let's Ø!

Second, for many speakers, *let's* negates differently from *let us*. While it is possible simply to use *not* after *let's*, as after *let us*, as in (6a), many speakers use *don't* to negate *let's* (with variable ordering), a possibility that is not found with *let us*, on the relevant reading, as indicated in (6b) and (6c):

(6)  a. *Let's not go / Let us not go
     b. *Let's don't go / Don't let's go
     c. *Let us don't go / *Don't let us go.

Third, there are other cooccurrence facts that differentiate *let's* from *let us*. In particular, ostensible subject nominals can be found after *let's* but not after *let us*, as shown in (7):

(7)  a. *Let's us go / *Let us us go
     b. *Let's you and I go / *Let us you and I go
(though the type of (7b) is possible as an afterthought or in a poetic register, as in T. S. Eliot's *Let us go then, you and I*, from *The Love Song of J. Alfred Prufrock*). Especially interesting in this regard are sentences such as (8), which are acceptable for some speakers and, strikingly, have no first person plural referent at all:

(8) *Let's you and him go.*

Given these differences, it seems that *let's* must be separated from *let us*, despite the obvious character of a connection between them, a connection that is salient and recognizable to native speakers, even if not supported by the linguistic evidence. The significance of this separation is that obvious synchronic connections do not always keep historically related forms together; individual surface forms can each follow their own path of development, as speakers allow them to diverge.

3.4 Similarly also, there are speakers of American English who, instead of the gerund/participial *having* to from the necessitative *to have* to have innovated the form *hafing* to ([haɪfɪŋ tu]). The source of this innovation seems to lie with the surface form of *have to*, which is generally pronounced [haɪftə], as reflected by the common spelling *hafta*. That is, in *hafing to*, *haf-* has apparently been extracted out of the surface form *hafta*, and used as the basis for the creation of the innovative gerund/participial form. In that way, *hafing to* shows the effects of a reanalysis in which speakers totally overlooked the seemingly “easy” and reasonably well-motivated connection between *hafta* and *have to*.

That is, speakers with the innovative *hafing* to seem to have focussed more on the surface form *hafta* than the putative source *have + to*, even though there are good grounds for connecting *hafta* with *have to*. In particular, *hafta* and *have to* share inflectional forms, e.g. 3SG *hasnaad*, and *hafta* can be separated, in which case the pronunciation must be *[hæv ... ]* and not *[hæf ... ], as shown in (9):

(9)  a. You only have ([hæf] / *[hæv]) to ask and it'll be yours
    b. You have ([hæv] / *[hæf]) only to ask and it'll be yours.

Moreover, the voicing assimilation process needed to give [*f*-] from /-vt-/ is relatively trivial, and is independently motivated — assuming that the [*f*-] / [*t*-]/ combination involves some clitic readjustment — by the absence of [*vt-] clusters word-internally. It should be noted, though, that the voicing assimilation must be lexically specific, since it is not found with other verb + *to* combinations, as shown by (10):

(10) *I love to ([ləv tu] / *[ləf tu]) shop.*

Still, despite a relatively transparent connection with *have*, some speakers have reanalyzed *hafta* as (underlyingly) *haf* - *ta* thus yielding the innovative form, [haɪfɪŋ ta], based on the surface manifestation of the phrase. Once again, then, speakers have moved away from transparency, moving instead in the direction of increased opacity in the matter of the formal relationship between various semantically related elements.
3.5 This failure to see connections that seem quite transparent is not restricted to “imperfect” speakers of Modern English; an example can be cited from Sanskrit in which an originally reduplicative element was reanalyzed as having no phonological connection with the root it attached to, even though the phonological connection was not terribly obscure.\(^5\)

The most typical stem for the perfect tense in Sanskrit has reduplication, of varying shapes but still with some obvious phonological relation — though not always identity — between the reduplicative syllable and (part of) root, as in (11):\(^6\)

\[
(11) \begin{align*}
\sqrt{\text{pat}} & \quad \text{'fly'} \rightarrow \text{PERF pa-pat}- \\
\sqrt{\text{tsar}} & \quad \text{‘approach stealthily’} \rightarrow \text{PERF ta-tsår}- \\
\sqrt{\text{bhid}} & \quad \text{'split'} \rightarrow \text{PERF bi-bhed}- \\
\sqrt{\text{sthå}} & \quad \text{'stand'} \rightarrow \text{PERF ta-sthå}-. 
\end{align*}
\]

Early on in the history of Sanskrit, the root \(\sqrt{\text{an}}\)- ‘attain’ had a reduplicative perfect stem \(\text{an-an} \text{-s} \)-, and other roots with the phonological shape of short a followed by a nasal did too, as shown in (12), where the latter two are somewhat later forms found in the native grammatical literature:

\[
(12) \begin{align*}
\sqrt{\text{aĩj}} & \quad \text{'anoint'} \rightarrow \text{PERF an-aĩj}- \\
\sqrt{\text{aŋ}} & \quad \text{'move'} \rightarrow \text{PERF an-āŋ}- \\
\sqrt{\text{aŋc}} & \quad \text{'bend'} \rightarrow \text{PERF an-ānc}-.
\end{align*}
\]

What is interesting about this subclass of perfect stems is that despite the rather transparent phonological connection they show between the reduplicative element \(\sqrt{\text{an}}\)- and the nasal-containing root with these verbs, and despite the fact that copying reduplication in the perfect is the norm, the formative \(\sqrt{\text{an}}\)- spread as a nonreduplicative perfect-stem marker onto roots with \#ar- or \#r-, without a nasal, as shown in (13):

\[
(13) \begin{align*}
\sqrt{\text{rdh}} & \quad \text{'thrive'} \rightarrow \text{PERF an-rdh}- \\
\sqrt{\text{arh}} & \quad \text{'deserve'} \rightarrow \text{PERF an-th}-. 
\end{align*}
\]

In fact, this perfect-stem formation process was given by the native Sanskrit grammarians as the norm for roots with this shape and, again, according to the grammarians, was possible even for one root without \(-r/-l\)- or a nasal, namely \(\sqrt{\text{aks}}\)- ‘attain’ (with a perfect \(\sqrt{\text{an-aks}}\)-).\(^7\)

It is possible that some dissimilatory effects may have been operative in spread of \(\sqrt{\text{an}}\)- onto \(-r/-l\)-roots, since the sequence [...] r [...] is disfavored — though not impossible — in Sanskrit phonology. If so, however, it is not clear why \(\sqrt{\text{an}}\)- would have spread to \(\sqrt{\text{aks}}\)- (though see note 7). Similarly, if dissimilation is held responsible, then it is not obvious why the dissimilation was to \(-n\)- and not to \(-l\)- such as occurred in the another reduplicative formation, the so-called “intensive” conjugation, of \(\sqrt{\text{t(ch)}}\)- 'go', i.e. \(\sqrt{\text{al-ar}}\)-, as opposed to the (admittedly later) perfect \(\text{an-arch} \)-.

It can be concluded, then, that \(\sqrt{\text{an}}\)-amś- was reanalyzed as nonreduplicative, i.e. as just having a prefix \(\sqrt{\text{an}}\)-, and not a (reduplicative) copy-prefix, as the perfect-stem formative, and that this reanalysis took place despite the apparent transparency of \(\sqrt{\text{an}}\)-amś- as involving reduplication. Once again, then, speakers lost sight of a seemingly obvious analysis and restructured in the direction of greater opacity.

3.6 Finally, as a suggestive example from a non-Indo-European language,\(^8\) an apparent opaque reanalysis of a transparent form can be cited from Korean. Stem-final clusters of consonant plus \(s\) may generally remain or optionally undergo reduction when case markers are added on; for example, \(\sqrt{\text{kap} / [\text{ts]}-\text{p}}\) ‘price’ has the nominative form \(\sqrt{\text{kap} / [\text{ts]}-\text{i}}\) or optionally \(\sqrt{\text{kap} / [\text{ts]}-\text{b}}\). The noun stem \(\sqrt{\text{tol}}\)- ‘anniversary’ is spelled as if it were \(\sqrt{\text{tos}}\)-, suggesting it may have had a final \(-s\)- in earlier stages of Korean; if so, it quite likely had variant realizations, parallel to those found with \(\sqrt{\text{kap}}\)-, i.e. \(\sqrt{\text{tol} / [\text{k}]}\) / \(\sqrt{\text{tol}}\). In Modern Korean, the \(-s\)- never is pronounced in this word, suggesting a reanalysis, i.e. a restructuring or relexification, to underlying \(\sqrt{\text{tol}}\)-, even though, presumably, the \(-Cs\)- / \(-C\)- alternation was regular and productive and therefore eminently recoverable and transparent. Unless there was a separate and very specific sound change affecting \(-Is\)- sequences, which is entirely plausible but seems not to have occurred, it would appear that in this case, too, a reanalysis of what is seemingly a transparent relation between two forms occurred, with the result being greater opacity in the relationship.

4. Conclusion

The upshot of all these examples, besides showing the benefits of reintroducing speakers into diachronic formulas for understanding language change, is that the last several especially allow for the argument to be made that speakers pay much more attention to their immediately apparent data, in particular to surface forms of words and expressions, than they do to more “distant” data of the sort that a linguist — or a perfect speaker —
might be aware of. This is not a new observation; indeed, as recently reiterated in Derwing & Skousen (1989:58), Anttila (1972:349) noted that “memory or brain storage is on a much more extravagant scale than we would like to think; even the most ‘obvious’ cases [of linguistic relations] can be stored separately”. These opaque reanalyses all depend ultimately on separate storage for the elements involved, on their being divorced from one another in speakers’ grammars.

It can thus be argued, based on these and other similar examples, that speakers in the process of using — and thus of changing — their language often act as if they were in a fog, by which is meant not that they are befuddled but that they see clearly only immediately around them, so to speak, and only in a clouded manner farther afield. They thus generalize only “locally”, in the sense of Joseph & Janda (1988), and not globally over vast expanses of data, and they exercise their linguistic insights only through a small “window of opportunity” over a necessarily small range of data. How else can the persistent phenomenon of opaque reanalyses of apparently transparent configurations of facts be explained? Presumably, what the linguist thinks might be salient is not always so salient to native speakers!

Nonopaque reanalyses do occur, so speakers must be capable of viewing and keeping in mind chunks of data beyond what is immediately before them, but such is not always the case, and it is the opaque reanalyses that are the most revealing of what actually goes on with speakers. That is, when reanalyses occur, they are not always in the direction of simpler grammars overall but rather are often complicating, in a global sense, even if they are simplificatory in a local sense.

This view has been shown to have the value of leading to a deeper understanding of at least certain types of language change, i.e. at least the case-studies presented here. At the same time, though, putting the speaker and his/her abilities back into diachronic accounts also has important implications for the proper construction of synchronic accounts. In particular, if one adheres to the view that grammars ought to mirror speakers’ actual capabilities and not a somewhat idealized construction of them, then clearly cases such as those presented here show that the typical types of evaluation metrics that linguists use to argue for the proper formulation of a fragment of a grammar cannot (always) be maintained; the grammars linguists construct, therefore, ought to be allowed to reflect uneconomical “solutions”, at least in diachrony, but also, given the relation between synchrony and diachrony argued for here, in synchronic accounts as well.

Derwing & Skousen (1989:59), in putting forth views similar to the one advocated here, though on different grounds, give the following assessment, which provides a good closing to the present discussion:

Whose grammars are learned or acquired by speakers, after all — linguists’ grammars, or the grammars (if we may call them that) that are actually present in individual brains? And whose grammars, after all, actually play a role in the very real processes of language production and comprehension that real speakers and hearers actually engage in? And whose grammars go wrong in any of the various aphasic disorders? Surely, if linguists are content to write arbitrary grammars, based on arbitrary theories and judged by arbitrary decision criteria, they remove themselves by default from those very areas of explanation that most of them, we gather, have hoped might lift their field out of the backwater of its taxonomic past.

Paying attention to what speakers do diachronically and figuring them into the overall picture should therefore be the basis for diachronic explanation as well as for synchronic explanation.

Notes

* I would like to thank several people who contributed, directly or indirectly, to this paper. I owe a significant intellectual debt to Rich Janda, who, as always, has proved to be an important stimulus for the ideas contained herein. My thinking on this subject owes much also to the work of Raimo Anttila and to Joki Schindler. The presentation of this paper at the Milwaukee conference provoked numerous interesting comments, for which I thank Raimo Anttila, John Haiman, Larry Hutchinson, David Lightfoot, and Karin Michelson; many suggestions of additional instances of opaque reanalyses were made, which I plan to follow up on in Joseph (forthcoming). Finally, I would like to thank the organizers of the conference, and especially Garry Davis, for giving me the opportunity to hold forth on a topic so dear to my heart.

1. As Labov (1989:85) has put it: “every construction, every word, every sound and vocal gesture of [speakers’] local dialect is the product of an historical evolution. All language is an historical residue, except perhaps for that shimmering target of formal linguistics, the principles of innate and universal grammar”.

2. Maher, in commentary included in Anttila (1976:322), notes the occurrence of this reanalysis of pitch in pitch-black citing his son’s use of pitch-white for ‘very white’; see also Anttila (1985:20) for further discussion. I myself have heard the form pitch-red used by an adult speaker, a red-head, to refer to the bright red color of his hair when he was young.

3. My thanks to Joy Hoyte for bringing sentences like (8) to my attention.

4. I have heard [hafig] from a few adult speakers in Columbus, and others have reported hearing it too. My younger son Adam, when he was three and a half years old, produced
a slightly different but clearly related form, [hafta], which incorporated the [t] of *hafta* as well and thus also showed no overt connection with the putative base verb *have*.

5. I am indebted to Joki Schindler for this example, which he presented over a dozen years ago in a class that had a considerable impact in helping to shape my understanding of the difference between what speakers do with their language and what linguists do.

6. For some recent discussion of reduplication in Sanskrit, with extensive bibliography, see Janda & Joseph (1986, 1989); the standard account of the facts is to be found in Whitney (1964). The symbols used in the transliteration follow standard conventions for Sanskrit.

7. It must be noted, though, that *vakṣ- ‘attain’ is historically related to the nasal-containing root *vanē* cited above, so that the occurrence of the perfect marker *ān- with *vakṣ- may be its occurrence with *ānē*. Still, the chronology of the appearance of *ān- with these roots argues against that, since it is likely that a synchronic connection between *vanē* and *ānē* was not very salient.

8. I say “suggestive” only because I cannot vouch for the facts in the way that I feel I can with English, Greek, or Sanskrit; my sources are Cheun (1975) and Oh (1990). I am indebted to Mira Oh for bringing this example to my attention.

9. As noted earlier, Janda & Joseph (1987, 1991) call into question the validity of the usual interpretations of ‘ecomonical’ in the context of the evaluation metrics for grammars and for individual analyses of facts from a language.

References


