How General are our Generalizations? What Speakers Actually Know and What They Actually Do

Brian D. Joseph
The Ohio State University

1. Introduction

As linguists, we know what our job is: we describe and analyze and ultimately account for the linguistic behavior that speakers exhibit. That certainly sounds easy, especially since we are all pretty much committed to the idea that language is a rule-governed system and that we can make general statements, what we call "generalizations", and especially since we can point to evidence that speakers follow "rules", and that they evince rule-governed behavior.

For instance, it is commonly noted that when as English speakers we encounter a noun for the first time, whether a borrowing from a foreign language or from another dialect of English, whether a neologism such as an acronym or clipping, or even an obscure word we have not run into before, we know how to form the plural, and moreover, the distribution of the allomorphs we use, namely [s]/[z]/[əz], is the same as we use in words that have been a part of English for generations (not all necessarily "native" but all part of English for a long time). Some examples are given in (1a), with a relatively new loan word, two recent acronyms, a recent clipping, a reverse-slang form (based on spelling) from British English (thus new to at least this speaker of American English), and a unit of foreign currency (from Thailand), along with an obscure but native word for a Middle English letter; note that the plurals are parallel to the forms in (1b) which all have a longer history in English:

(1) a. taco–[z], SIMM-[z], VHS-[əz], condo-[z], yob-[z], baht-[s], yogh-[s]
b. toe-[z], whim-[z], horse-[əz], bat-[s], rock-[s]

The "rule-governed-ness" of the plural emerges in the fact that plural forms such as those in (1a) are unexceptional — there appears to be a straightforward way of forming them, and there cannot be a precise model for a word like *baht*, for instance, since it is lexically isolated within English, so instead, the argument goes, it must be the case that a pattern is summoned up to generate the necessary plurals, and when we start talking about "patterns", we are really talking about rules! Plural-formation, in English, therefore, gives evidence of rule-governed behavior, and thus a generalization is possible over the distribution of [s]/[z]/[əz] as a marker of plurality.

Such examples of regularity allowing for generalizations are not restricted to individual words nor to borrowings, though we will see that borrowings are especially interesting in this regard. In particular, regularities can be found in the way English speakers put words together into sentences, so that speakers — and linguists — can readily

discern that the strings in (2) are well-formed while those in (3) are not:

(2) a. The fat cat fell onto the mat
b. The hungry bear roamed through the forest.
c. The eloquent attorney stood before the judge

(3) a. *The cat fat fell onto the mat / The fat cat onto the mat fell
b. *The bear hungry roamed through the forest / The hungry bear through the forest roamed
c. *The attorney eloquent stood before the judge / The eloquent attorney before the judge stood

From examples like these, we extract the regularities that adjectives precede their nouns in English and that prepositional phrases cannot precede verbs in the predicate; these regularities constitute generalizations over the set of possible phrases (noun phrases and verb phrases) and thus over the set of possible sentences as well, and can be stated as in (4):

(4) a. $ADJ - N]_{NP}$ / $*N - ADJ]_{NP}$
b. $V - PP]_{VP}$ / $*PP - V]_{VP}$

Similarly, a generalization can be made over the combinations of locative-like expressions in English involving a state of being vis-à-vis institutions, as in (5):

(5) in school, in college, in jail, in court

such that the noun is always unarticulated, i.e. lacking a definite article (with a definite article, of course, the phrases are acceptable, but have a different interpretation, referring to a specific location rather than a general state of being).

But, it can easily be demonstrated that not all of these generalizations are not completely accurate, for all speakers and for all possible combinations of elements. In particular, there are some speakers who treat combinations like *attorney general* as a Noun - Adjective sequence, showing plural marking on *attorney*, as in (6a), and there are speakers who have a definite article in the institutional locative expression with the noun *hospital*, as in (6b):

(6) a. attorney general / attorneys general
b. in the hospital

Thus on the face of it, the facts in (6) tell us that the generalizations stated in (4a) and evident in (5) are not 100% true; they may reflect the typical case but do not unequivocally delimit the range of possibilities. Now, as is well-known, several mitigating factors might be cited in defense of the generalizations and to counter the “attack” on them that (6) poses. In particular, it can be noted that *attorney general* is the result of the borrowing of a French phrase and represents a French patterning for the placement of adjectives; moreover there are many speakers of English who have “regularized” this noun phrase and

mark *general* with plural inflection, indicating that they now treat *general* as a noun and *attorney* as an adjective, or else treat the combination as a compound noun. Similarly, *in hospital*, without the article, is standard usage in a number of dialects of English, including British and Canadian English.

(7) a. attorney generals
b. in hospital [British/Canadian usage]

A question to ask at this point is what is the significance of the fact that there are some speakers who follow the generalization and some who do not? Is American English any less a natural human language because its speakers say *in the hospital* but *in school*? Are Canadian English speakers "better" speakers of a natural language because they generalize over *school* and *hospital* (etc.) in institutional locatives? And what of the fact that *attorney general* is a French-ism in English, especially for speakers with *attorneys general* as the plural? Is it to be discounted? Is significance only to be attributed to the speakers with the plural *attorney generals*? What about *attorneys general* speakers — are they not speakers of English?

2. The Main Point — Lexically Particularistic Knowledge

My main point here is that while linguistic generalizations are important and to some extent reflect speakers' linguistic competence, we should not take our generalizations too seriously; they are important guides as to what is *generally* true for a language, but like Sapir's grammars, they "leak", and we should not be surprised to find behavior that is contradictory to apparent generalizations, not only across speakers but within speakers as well. Moreover, the type of information that speakers have access to again and again seems to be facts about particular lexical items — specific morphemes, specific words, specific phrases, and I would argue that such information, this *lexically particularized* knowledge, gives us crucial insights into the nature of what speakers know about their language.

3. Lexical Particularism Exemplified

Part of what is at issue with *attorneys general* versus *attorney generals* is what may be termed the extent of nativization, i.e. the degree of assimilation of a borrowed element into the native language patterns. But, as the existence of speakers with the plural *attorneys general* shows, sometimes borrowing can lead not to assimilation but to change in the native patterns. Moreover, it shows that speakers can vary in their treatment of such forms. In a recent study on aspect formation with borrowings in Slavic languages, Neikirk (1996) has found a similar variability in the way Russian, Macedonian, and Bulgarian speakers deal with the fact that verbs in their languages typically have aspectual pairs, an imperfective and a perfective form, yet borrowed verbs enter the language with just one form. She found that some Bulgarian speakers create (or at least tolerate) a prefixed perfectivized aspect form to go along with a borrowed imperfective form, whereas other speakers

reject such forms and use the ostensible imperfective, the unprefixed borrowing, as a biaspectual form; interestingly also, of the 13 verbs she examined, although some were treated uniformly by all 49 speakers she tested in her questionnaires, e.g. *korespondiram* 'to correspond' (in (8a), with no prefixed form for any speaker), there was considerable variation on the majority of the verbs, as indicated in (8b), and importantly, of the verbs that showed variation, not a single one was treated in the same way by all speakers, and not all speakers had the same patterns of responses; she reports similar results for questionnaires she administered to Russian and to Macedonian speakers (the numbers in parentheses in (8b) indicate the number of times a given response occurred in her data set from among the various conceivable options):

(8) a. No prefixed perfectives: korespondiram 'correspond' (49)
b. *komentiram* 'comment': *iz* (8), *ot* (7), *do* (2), *pro* (1)
unprefixed (31)
kodirvam 'to code': *za* (16), *ot* (13), *raz* (4), *pre* (3), *de* (2)
unprefixed (11)

Thus while there is a tendency towards nativization and regularization of these borrowed verbs, by the creation of perfective forms that thereby assimilate the verb to the predominant native pattern of an imperfective and a matching, prefixed, perfective formation, so that we can generalize that speakers tend to aspectualize formally borrowed verbs, no 100% true generalization can be made over all speakers nor over all borrowed verbs. There may be subgeneralizations regarding the use of particular prefixes and the languages seem to be moving in the direction characterized by Neikirk as follows: "all Russian/Bulgarian/Macedonian verbs have aspect, but not all verbs in Russian/Bulgarian/Macedonian have aspect", but her basic investigation reveals the sort of situation that I would argue is typical in a language — generalizations that are not 100% true and speakers who show lexically particularized knowledge of their language.

So loan words can be revealing of native linguistic competence, but not always in the way that the "rule-governed" sorts among us would expect — or prefer. It is important to note also that there are cases in which evidence from the treatment of loan words seems at first to lend itself to use in support of a given statement of a generalization, but on further inspection turns out to be ambiguous as to the status of that putative generalization.

For instance, while it seems to be true, as the examples in (1) above show, that all borrowed nouns that get an *s*-plural do follow the distribution of [s]/[z]/[əz] evident in "native" (or "nativized") English nouns — though more on that below — not all borrowed nouns in English get an *s*-plural, at least not at first, as shown by the forms in (9), all cited as possible plural forms in *AHD* (and note the plural form *baht-s* cited in (1a)):

(9) schemata, antennae, criteria, baht

There is of course some variability as to nativization, in terms of both when individual items become nativized and what forms individual speakers use for particular words (*schemas* vs. *schemata*, *criterions*

vs. *criteria*, etc.) and there is some polarization of forms (e.g. *TV antennas* but *insect antennae*), but that is precisely the point!

Even more dramatic perhaps regarding the ambiguous status of generalizations in the face of loan words is the situation with the accent of neuter nouns in Modern Greek. Greek neuter nouns in *-i* present a regularity with regard to accent placement, as indicated in (10): although nouns in general can be accented on any of the last three syllables (note e.g. the feminine nouns in *-a* listed in (10a): *peripétia* 'adventure' vs. *omorfiá* 'beauty' vs. *∂imokratía* 'democracy'), and although neuter nouns in general can have antepenultimate accent (e.g. the forms in (10b): *ónoma* 'name', *jípe∂o* 'field', *rápsimo* 'sewing'), neuter nouns in *-i* do not in general show antepenultimate accent placement (and thus have only penultimate or ultimate accent), as the forms in (10c) and (10d) indicate — actually, disyllabic neuter *-i*-nouns show penultimate or ultimate accent, and trisyllabic or longer show penultimate accent.[1] Since other accent placements are possible for nouns, if neuter *-i*-nouns systematically delimit accentual possibilities, then that is a significant regularity in the language.

(10) a. peripétia 'adventure', ∂imokratía 'democracy', omorfiá 'beauty'
b. ónoma 'name', jípe∂o 'field', rápsimo 'sewing'
c. máti 'eye', spíti 'house' vs. malí 'wool', spirí 'grain'
d. vivliaráki 'booklet', tetrá∂i 'notebook', molívi 'lead', kremí∂i 'onion', farági 'gorge', sfugári 'sponge', skulíki 'worm'

There is, of course, an historical explanation for the absence of antepenultimately accented neuter *-i*-stems, given in (11): most derive from earlier diminutives in *-ion*, a suffix which was either stressed on *-i-* itself or induced an accent on the syllable before it (the antepenultimate), thus *...VC'VCion* or *...VCVC'ion*; in Post-Classical Greek, such nouns underwent a reduction in the final syllable, first through the loss of the *-o-* in *-ion*, so that it became *-in*, and then the loss of the final *-n*, giving *-i*, so that one-time antepenultimate-accent nouns became penultimately accented. Also, borrowings ending in consonants that had accent on the final syllable often were nativized as neuter *i*-stem nouns (e.g. Turkish *lukúm* —> Greek. *lukúmi* 'Turkish delight'), giving them penultimate accent in Greek.

Still, regarding the strength of this generalization synchronically for current Modern Greek, and synchronically for a relatively recent point in the past, one can ask what to make of the Turkish loan-word given in (12):[2]

(12) *fíldisi* 'ivory' (from Turkish compound *fil-di̇-i*, literally *fil* 'elephant' + *di̇* '*tooth*, with the 3SG.POSS marker *-i* joining together the parts of the compound, thus: "elephant-tooth")

As far as accent is concerned, *fíldisi* actually retains the Turkish accent, which is strongest on the first syllable, with a secondary stress on the final syllable (thus: *fíl-di̇-i*), so that we can ask whether *fíldisi* can simply be ignored as a foreign word (recall Neikirk's observation

about all verbs in Slavic versus all Slavic verbs!), as is often done (in a more or less dismissive way, as might be done with *attorneys general*) so that we can then say the generalization is valid as far as "real" Greek nouns are concerned; or, does the fact that *fíldisi* was not nativized to *fildísi** or turned into a feminine noun[3] mean that the generalization is not valid, the reasoning being that if the generalization had some reality to speakers, they would not have tolerated this word in violation of it. In a sense, then, the evidence of *fíldisi* can cut either way as far as the validity of the generalization is concerned. It really comes down to the extent to which we are willing to accept counterevidence to observed regularities and whether we are more interested in the exceptions or in the more overarching generalizations.

However, to return to native or nativized English for the moment and to return to the question of plurals, one aspect of the distribution of the allomorphs of [s] turns out to be more interesting than hinted at earlier. Thorp (1990) reports on a very clever study, in which she used sibilant-final words, both real words, e.g. *Moses*, and nonsense words, e.g. *Dwoozes*, and presented them to some 20 or more English speakers, asking them to form plurals (for a proper name, the plural would either be a family plural, like "The Joneses", or a reference to multiple characters with the same name), and she found great variability from speaker to speaker, and within individual speakers, e.g.:

(13) Subject 1: PL Moses-[Ø] / Dwoozes-[iz]
Subject 4: PL Moses-[Iz] / Dwoozes-[Iz]

i.e. Subject 1 formed the plural differently for *Moses* and for *Dwoozes*, and differed from the way Subject 4 formed the plurals. This is a task that fairly tests native speaker competence, for the need to form such plurals can arise — for instance, what would the family plural for attorney Susan Thomases, whose name already looks like a plural? Would it be a zero-plural [taməsəz] or an affixal plural [taməsəzəz], or what? Or if one were talking about singer Kenny Rogers and pitcher Steve Rogers, how many [radʒərz]/[radʒərzəz] (or whatever) would be under consideration?

Moreover, the picture gets even more interesting when one considers, as Thorp did, other morphemes with the allomorphs [s]/[z]/[əz] in English, e.g. the markers for nominal possessives and for verbal 3SG present forms (in order to form verbs from these, imagine a derivation with *out-*, e.g. *to out-do someone*, *to out-Moses Moses*,[4] or the like), and found even more striking discrepancies between and within speakers, e.g.:

(14) Subject 1:
POSS: Moses-[Ø] / Dwoozes-[Iz]
VB: Moses-[Ø] / Dwoozes-[Iz]
Subject 4:
POSS: Moses-[Ø] / Dwoozes-[Iz]
VB: Moses-[Iz] / Dwoozes-[Iz]

Thus even though the usual case is for these three morphemes to share the same distribution of allomorphs, with the canonical form after

sibilants being [-Iz] (or some central vowel), these speakers used different strategies for different lexical items and for different morphemes, and moreover no two speakers in her sample had the same responses to all items in the list; there was a distinct pattern for each speaker and each speaker showed considerable internal variation.

To reiterate my main point, what such facts show us is that generalizations often are not all that general, and that speakers may vary internally in their own usage and externally against the usage of others, all in a lexically particular way.

This point can be brought out forcefully by an observation made by Pagliuca and Mowrey (1987). We might well think that one place in the grammar where real generalizations are possible is in the area of phonemics, in that declaring there to be a phoneme X in a language seemingly commits the linguist to the claim that speakers can generalize over a class of phones; as far as English is concerned, for instance, it means that we treat [p] and [p^h], as well as different tokens of [p^h] as being the "same" in some invariant sense, as indicated in (15).

(15) Phonemic Generalizations
a. [p] in *spin* / = [p^h] in *pin*/*pun*, etc.
b. [p^h] in *pin* = [p^h] in *pan* = [p^h] in *pun*, etc.

However, if lexical diffusion (e.g. Wang 1969) is the correct view of the implementation of sound change, whereby a change diffuses through the class of lexical items containing phoneme X rather than affecting all tokens of X at the same time, then speakers seem not to generalize even over all instances of phoneme X in a language, in that they change only some at a time, and not all. It is of course possible to take a contrary view regarding lexical diffusion,[5] but if lexical diffusion has any validity, it seriously challenges the notion of a phoneme as a generalization over a class of phones.

I have dwelt here on examples from phonology and morphology, but recall that one of my first examples involved the internal syntax of noun phrases in English. Syntax provides further cases of non-general generalizations, in which speakers give evidence of lexically particularized information, thereby showing this phenomenon to be a pervasive one at all levels of grammar.

For example, Maurice Gross has discussed his efforts, along with others, to produce a generative grammar of French, and reports (Gross 1979, and see references given there) that the grammar they devised, in particular the part dealing with complementation possibilities for verbs, "contains about 600 rules and conditions of application ... We attempted to verify systematically the applicability of these rules to more than 12,000 lexical items ... if we compare the syntactic properties of any two lexical items — it is observed that no two lexical items have identical syntactic properties. If we compare ... the domains of the rules, the result is the same". Gross takes this outcome as a stunning "failure of generative grammar", as his title indicates, and uses the lexically particularistic nature of these facts to challenge a basic tenet of generative grammar, namely that a theory of grammar should be based on a theory of learnability; yet, these facts are learned, without any generalizations being evident, suggesting that

speakers do not need generalizations as much as linguists think they do, or as much as linguists do!

A similar result has been reached by Kim (1996), an examination of the historical development of English impersonal verbs. She found that even when the overall class of impersonal verbs of Old English (OE) is broken down into several different subclasses, based on their semantics and general syntactic patterns, the individual members of each subclass still show differences in matters of detail in their syntax, and more important, do not behave uniformly in their later diachronic development into Middle English (ME) and on into Modern English. For example, as summarized by the table in (16), of the semantically similar RUE-class verbs, OE *hreowan* / ME *reuen* 'to rue', OE *sceamian* / ME *shamen* 'to shame', and OE *eglian* / ME *eilen* 'to ail' show uniformity in that they all can occur with an oblique experiencer and a nominative cause argument, and this uniformity continues into Middle English (see (16a)); at the same time, though, in both Old English and Middle English, these verbs differed somewhat from one another (see (16b)), e.g. RUE and SHAME but not AIL occur with oblique experiencer and genitive cause, and RUE but not SHAME and AIL occurs with oblique experiencer and an accusative cause; thus synchronically, these verbs at either stage were far from uniform and showed lexically governed differences. Diachronically also these verbs show individual behavior, as indicated in (16c), for while all three could occur with an oblique experiencer and a finite clause as causal argument in Old English, AIL loses that ability in Middle English even though the others retain it; also, while none of these verbs could occur with a nominative experiencer and an infinitival causal argument in Old English, RUE and SHAME, but not AIL, gain this syntactic possibility in Middle English.

(16) a. OE/ME:

	hreowan/*reuen*	*sceamian*/*shamen*	*eglian*/*eilen*
OBL Exp. + NOM Cause	X / X	X / X	X / X

b. OE/ME:

	hreowan/*reuen*	*sceamian*/*shamen*	*eglian*/*eilen*
OBL Exp. + GEN Cause	X / X	X / X	Ø / Ø
OBL Exp. + ACC Cause	X / X	Ø / Ø	Ø / Ø

c.

OE	*hreowan*	*sceamian*	*eglian*
OBL Exp. + FIN.CLAUSAL Cause	X	X	X
—> ME	X	X	Ø
NOM Exp. + INF.CLAUSAL Cause	Ø	Ø	Ø
—> ME	X	X	Ø

And, Kim found similar non-uniform, individualized behavior for the representatives of other classes of impersonal verbs in Old English and in Middle English, and in the transition between these stages.

Again, the lesson to be drawn is that speakers have lexically particularistic linguistic knowledge, and do not form broadly applicable generalizations — the verbs in the same semantic class are not

necessarily parallel to one another in any stage of English and they show individual development on into Middle English. While some generalizations are possible, for example, there is an overall tendency towards the elimination of impersonal constructions between Old and Middle English, the individual behavior of particular verbs makes it impossible to state 100% true generalizations of a broad scope.

4. The Evidence from Neologisms Revisited

At this point, it is useful to reconsider some of the evidence we examined at the outset that appeared to support the notion of rule-governedness. We said that plural formation for neologisms like those in (1), or more accurately, in the light of the alternate plurals in (9), the distribution of the allomorphs of the // s // plural marker, provides evidence for rule-governed behavior.

Yet, it can be argued that even apparent "rule-governed" cases, like *tacos*, *SIMMs*, etc. need not be interpreted that way. In particular, there is *a priori* no reason not to treat the plurals to neologisms simply as being formed by analogy to existing words, like *toe*, *whim*, etc., by what may be schematized as a proportional analogy, as in (17); and it need not be a rhyming word, though that is an easy model to look to, but rather the basis for the analogy could be parallelism in stem-final segments, or even in stem-final phonological features, as in (17b) or (17c):

(17) Neologistic plurals by analogy:

a. toe : toe-[z] : : taco : X, X —> taco-[z]
whim : whim-[z] : : SIMM : X, X —> SIMM-[z]

b. ... -m# : ...-m-[z] : : [sIm] : X, X —> [sIm]-[z]

c. ...$W_{[+voiced]}$: ...$W_{[+voiced]}$-[z] : : ...$m_{[+voi]}$: ...m-[z]

Viewed in this way, one could say that analogy, rather than a rule *per se*, is at work in the formation of neologistic plurals. One might well ask, at this point, what the difference is between analogy and a rule, and that, again, is precisely the point: one way of interpreting the notion of "rule" is that it is a summation, a kind of shorthand so to speak, for a series of analogies; in a sense, analogies define the patterns, the regularities, that we tend to interpret as "rules".

Generalizations, in such a view, become really a matter of retrospectively summing up a set of analogical extensions from a model to a novel "target"; after enough such extensions, there is the appearance of rule-governed behavior, but even the novel formations could just be on-going examples of the analogical extension process.

What all this suggests is that speakers' knowledge really is not so general, but rather is focused in a lexically particularistic way; generalizations are possible, but they are more a retrospective formulation. In this regard, the results seen above from Neikirk, Kim, and Gross are significant, for in them, essentially each lexical item had its own synchrony, its own synchronic state, thereby giving each its own history too.

This view by which speakers in essence learn individual facts about lexical items and/or phrases, and extend that knowledge can be

seen to be operative in the spread of innovative constructions, both those that have a basis in language contact and those that result from language-internal innovations.

For example, several languages in the Balkans, e.g. Albanian, Bulgarian, and Greek, have a mildly productive syntactic construction consisting of a verb followed immediately by its negation, i.e. *Verb-NOT-Verb*, meaning 'whether one Verb-s or not', as exemplified in (18a); the likely source for this pattern is the expression *WANT-NOT-WANT* 'whether one wants to or not', which is the best lexical match found for this pattern in the various languages (and is typologically paralleled by English *willy-nilly*), shown in (18b).

(18) a. Greek *fíji ∂e fíji* 'whether one leaves or not'
Bulgarian *pie ne pie* 'whether one drinks or not'
b. *WANT-NOT-WANT*: Greek *θéli ∂e θéli*, Bulgarian *ste ne ste* / Albanian *donin s'donin* 'whether one/they want(s) to or not'

Thus it appears to be the case that one instance of the construction, the *WANT-NOT-WANT* one, was borrowed, and from that basis in at least some of the languages, a pattern grew, through extensions to other verbs. Thus here too we have a lexically particularized starting point for a "generalization", i.e. a syntactic pattern, through essentially an item-by-item spread, with a phraseological borrowing as the initial locus for the pattern.

Similarly, Modern Greek has a third person nominative weak pronoun that occurs in two and only two expressions, a deictic with the meaning 'Here is/are ...!' and a locative interrogative with the meaning 'Where is/are ...?', as exemplified in (19):[6]

(19) a. ná tos 'Here he is!'
here-is he/NOM.MASC.SG.WEAK
b. pún dos 'Where is he?'
where-is he/NOM.MASC.SG.WEAK

As discussed elsewhere (Joseph 1981, 1994), the ultimate source of this construction in Greek is language contact, for it appears that *ná* is a borrowing from South Slavic and that the original syntax in Greek with *ná* was deictic followed by a weak accusative (ACC) pronoun, a widespread South Slavic pattern. In what may have been an extension within Greek alone, strong accusative forms became possible after *ná*. What happened then in Greek is that *ná + ACC* was reinterpreted as *ná + NOM(INATIVE)*, through the medium of strong neuter forms whose NOM and ACC were identical, so that these several patterns were coexistent. An analogy involving strong and weak masculine (M) accusative forms together with the strong nominative form now possible with *ná* led to constructions such as *ná tos*, with a new category of weak nominative pronoun. These steps are summarized in (20), with the innovative analogical creation *tos* based on the inherited strong forms of the 3rd person pronoun sketched in (20c):

(20) a. Original syntax: *ná* + WEAK.ACC, e.g. *ná ton* 'Here he is'

b. 3SG.NTR.ACC *aftó* ‘it’, in *ná aftó* ‘Here it is’, reanalyzed as NOM (since NTR.NOM = NTR.ACC)

c. ná aftón : ná ton :: ná aftós : X, X —> ná tos
M.ACC.STR M.ACC.WK M.NOM.STR M.NOM.ACC

From that starting point, this construction-type has experienced a spread, minimally to be sure but a spread nonetheless, to the semantically similar *pún* ‘where is/are?’ (made up of *pú* ‘where’ and a reduced form of *íne* ‘is/are’), giving the *pún dos* of (19b). The result is a modest generalization of a syntactic pattern over *ná*/*pún*, representing a pair of semantically related forms, but again it is a generalization that has been created retrospectively, as it were, through the spread of a type from one lexical starting point to another.

5. Are Any Generalizations Possible?

So, in the face of all of this evidence, do we have any reason to believe that it is possible to make generalizations at all? I would argue that the answer is yes, as long as we are clear as to the scope of our generalizations. In a sense, there is a tension between speakers’ learning of individual facts about their language and their desire to extend that knowledge, to use existing knowledge in learning new facts (a process which is at the heart of analogy — interpreting the novel or unfamiliar in terms of already-existing knowledge, what is already known and familiar), and thereby to generalize.

Most generalizations, therefore, should be recognized as being truly local in nature, that is, they have a restricted scope, and where linguists’ generalizations go astray is in not being sufficiently localized.

In the discussion earlier about American English *in school* / *in the hospital* as opposed to other dialects with *in school* / *in hospital* can now be elaborated on somewhat. In particular, a similar point can be made with directional phrases, as in (21), for where American English has *to school* but *to the hospital*, other dialects have unarticulated nouns in these phrases, e.g. Australian English *to hospital* (cf. *to school*, *in school*/*in hospital*)[7] What this means, then, is that there is a generalization here in American English, but it is a somewhat limited one: the generalization in question concerns the behavior of the noun *hospital* in American English in such phrases, for *hospital* is consistently irregular, and thus defines a subregularity of its own. One might well say that a generalization over a single item is not very general, and that is true, but there is, nonetheless, a regularity in the behavior of *hospital*, a regularity of a lexically particularistic sort.

This type of situation, in which data reflects highly localized generalizations rather than broadly general ones, would seem to defy any formalization, but since it is, I claim, such a common one, a place must be made in linguistic theory for such arrays of facts. Indeed, a construct I have proposed together with Richard Janda in a series of papers (Janda & Joseph 1986, 1989, 1992, 1995, as well as Joseph & Janda 1988), provides the necessary tool.[8]

Such situations can be characterized and captured synchronically via the notion of CONSTELLATIONS, generalizations that

are not wide-ranging ones, but rather are localized or fragmented; we have defined this construct as in (21):

(21) The **constellation**: a group of elements which share at least one characteristic property of form but are distinguished by individual idiosyncrasies, both of form and of function, that prevent their being collapsed with one another.

Various constellations have been discussed over the past decade, including the following umlaut in modern High German (Janda 1982, Janda & Joseph 1986), aspiration processes in Sanskrit (Janda & Joseph 1988), the multiple realizations of negation in Greek (Janda & Joseph 1995), the multiple realizations of irrealis marking in Ukrainian (Sydorenko 1996), and reduplication in Sanskrit (Janda & Joseph 1986). I exemplify this construct with the first case that was presented in the literature, namely reduplication in Sanskrit.

As shown in (22), Sanskrit reduplication, which is associated with various functions (e.g. perfect tense, desiderative formation, etc.), although claimed by many linguists to be a CV- prefixing process, with universal association principles essentially taking care of the rest, there are in fact a number of differences in the various manifestations of reduplication. As (22a) shows, there are actually at least 8 different template shapes, of which CV- is only one (admittedly the most common one but the others are well-represented); also, as indicated in (22b), not all reduplication is prefixal — a small number of roots form infixal reduplicated stems, as with *arh-* in (22b), where *-ji-* is the infixed reduplicating syllable. Finally, as indicated in (22c), there are some formations that are completely parallel to reduplicative ones, occurring in functions in which reduplication is expected, but the prefixed element happens to show no phonetic similarity with the root, meaning that Sanskrit has non-copying reduplication as well as the more usual copying reduplication. At the same time, however, all Sanskrit reduplication categories follow the same pattern with regard to reduplication of clusters with an initial sibilant — sibilant + stop clusters are reduplicated with the stop, while sibilant + resonant clusters are reduplicated with the sibilant, as indicated in (22d).

(22) a. Differences in reduplication template shape: V-, VV-, VC-, CV-, CVV-, CVC-, CVCV-, and CVCVV-
b. Differences in reduplicand placement: most are prefixal, but some are infixed, e.g. root *arh-* 'deserve' —> desiderative *ar-**ji**-h-(iṣa)-*, though)
c. Some apparent reduplicands show no copying, even though they are parallel to copying reduplicative formations, e.g. *arh-* 'deserve' —> perfect *ṣn-rh-* (parallel to *añj-* 'anoint' —> perfect *ṣn-añj-*)
d. At the same time, all types of Sanskrit reduplication follow constraint of reduplicating stop (T) in sibilant (S) + stop clusters but reduplicating the sibilant in sibilant + resonant (R) clusters, i.e. T ... ST- vs. S ... SR-

So, what generalization is there about Sanskrit reduplication? My claim is that there are generalizations, but they are best stated in

purely local terms, looking at particular grammatical categories as well as categories and sub-categories of root types; then it can be stated that CV- is the typical reduplication for perfect tense, VC- is typical for vowel-initial roots, and so on. However, there is no single generalization to be made.

If that is the case, is there any reason to speak in terms of "reduplication" in Sanskrit as if it were a unified phenomenon, or a unifiable phenomenon? Here the answer is yes, due to commonality in Sibilant-Stop/Resonant cluster reduplication. Thus reduplication in Sanskrit shows unity (based on (22d)) in diversity (as defined by the differences in (22a-c)), which is the defining characteristic of constellations.

6. Conclusion

It should be clear now where I stand on the question of how general our generalizations are: they are as general as speakers allow them to be, and that can be very ungeneral or quite broadly general. There are rules, and there are regularities in language, but when one examines where these rules come from, it is often from the cumulative effect of particularized extensions from one lexical item to another. Since this rule-formation process is an on-going one, a synchronic glimpse of a language is always going to capture the language with at least some incomplete generalizations. Thus if we as linguists are attempting to mirror speakers' knowledge of their language through our grammars, we should be prepared to have less-than-fully-general generalizations, and also subregularities that are defined on a very localized basis; in short, we should expect to find, and thus to have in our grammars, both *local generalizations* and *constellations* — if speakers are able to stand having them, then so should we as linguists!

Endnotes

1. I am not sure of the exact source of this generalization, but I heard it first from Angeliki Malikouti-Drachman, who mentioned it, and the exceptional *fíldisi*, in discussion at the First International Conference on Greek in 1993; a formal statement may well appear in print in one of her many works on Greek and I regret not having a precise citation. Nespor & Ralli 1994 mention this generalization but without reference to *fíldisi*.

2. Some apparent counterexamples can be discounted as irrelevant. For instance *xáiði* 'caress' (actually a variant form for *xáði*) and *vóiði* 'ox' (also a variant form, for *vóði*), are only seemingly trisyllabic; the internal *-i-* in these words forms a diphthong with the preceding vowel, so that they are more properly cited as *xájði/vójði*. Similarly, compounds with *méli* 'honey', e.g. *roðómeli* 'rose-flavored honey' (for which there is a more usual variant *roðómelo* that is not in conflict with the generalization) are true counterexamples from a synchronic standpoint, for the *-i* in *méli* is parallel to the *-i* in (10c-d) (cf. genitive singular *meljú*, like *tetraðjú*) ; diachronically, however, the *-i* in *-meli* is not from an earlier *-ion* (note Ancient Greek genitive *mélit-os*, with

stem *melit*-) and thus was not part of the formative processes that led to the accentual generalization.
3. Greek has feminine nouns in *-i*, which have a different declension from the neuter -i-nouns and do not follow the accentual generalization, e.g. *ðínami* 'power' (Genitive singular: *ðínamis* or *ðinámeos*), *rí›ani* 'oregano' (Genitive singular: *rí›anis*).
4. That is, to be more like Moses than Moses himself.
5. As I myself would, though mine may well be a minority opinion.
6. The initial [d] in *dos* is the result of a voicing assimilation triggered by the preceding nasal. Masculine forms are used in the examples since they are the only ones that overtly show the relevant distinction between nominative and accusative forms.
7. My understanding is that British English here uses *into hospital*, so no generalization is possible in that dialect since a different preposition is used. My thanks to John Payne for this information.
8. Given the venue for this paper, I am pleased to be able to note that three of the papers cited on constellations were presented at ESCOL conferences (ESCOL '85, '88, '91), and appear in the respective conference proceedings.

References

AHD = The American Heritage Dictionary of the English Language. Third Edition (1992). Houghton Mifflin Co.

Gross, Maurice. 1979. "On the failure of generative grammar". *Language* 55.5.859-885.

Janda, Richard D. 1982. "Of Formal Identity and Rule-(Un)Collapsibility: On Lost and Found Generalizations in Morphology". *WCCFL* 1.179-197.

Janda, Richard D. & Brian D. Joseph. 1986. "One Rule or Many? Sanskrit Reduplication as Fragmented Affixation". *OSU Working Papers in Linguistics* 34.84-107 (earlier version in *ESCOL '85*, pp. 103-119).

Janda, Richard D. & Brian D. Joseph. 1989. "In Further Defense of a Non-Phonological Account for Sanskrit Root-Initial Aspiration Alternations". *ESCOL '88*, pp. 246-260.

Janda, Richard D. & Brian D. Joseph. 1992. "Meta-templates and the Underlying (Dis-)Unity of Sanskrit Reduplication". *ESCOL '91*, pp. 160-173.

Janda, Richard D. & Brian D. Joseph. 1995. "On Constellations in Morphology: Crosslinguistic Evidence and a Case Study from Modern Greek". University of Chicago/Ohio State University unpublished manuscript.

Joseph, Brian D. 1981. "On the Synchrony and Diachrony of Modern Greek *na*". *Byzantine and Modern Greek Studies*, 7.139-154

Joseph, Brian D. 1994. "On Weak Subjects and Pro-Drop in Greek". In I. Philippaki-Warburton, K. Nicolaidis, & M. Sifianou (eds.) *Themes in Greek Linguistics (Papers from the First International Conference on Greek Linguistics, Reading, September 1993)*. Amsterdam: John Benjamins Publishers (1994), pp. 21-32.

Joseph, Brian D. & Richard D. Janda. 1988. "The How and Why of Diachronic Morphologization and Demorphologization". In M.

Hammond & M. Noonan (eds.) *Theoretical Morphology*. New York: Academic Press, pp. 193-210.

Kim, Hyeree. 1996. *The Synchrony and Diachrony of English Impersonal Verbs: A Study in Syntactic and Lexical Change*. Ph.D. Dissertation, The Ohio State University.

Neikirk, Jill M. 1996. *From Adaptation to Nativization: A Synchronic Analysis of the Category of Aspect in Borrowed Verbs in Russian, Bulgarian and Macedonian*. Ph.D. Dissertation, The Ohio State University.

Nespor, Marina & Angela Ralli. 1994. "Stress Domains in Greek Compounds: A Case of Morphology-Phonology Interaction". In I. Philippaki-Warburton, K. Nicolaidis, & M. Sifianou (eds.) *Themes in Greek Linguistics (Papers from the First International Conference on Greek Linguistics, Reading, September 1993)*. Amsterdam: John Benjamins Publishers (1994), pp. 201-208.

Pagliuca, William & Richard Mowrey. 1987. "On natural classes". Paper presented at University of Wisconsin-Milwaukee Linguistic Symposium on Categorization, April 1987.

Sydorenko, Halyna. 1996. *The Atypical Morpheme in Ukrainian*. Ph.D. Dissertation (in progress), The Ohio State University.

Thorp, Patricia. 1990. "The collapsibility of rule sets of the morpheme constellation *-s*". University of Pennsylvania unpublished manuscript.

Wang, William S.-Y. 1969. "Competing changes as cause of residue". *Language* 45.9-25.

ABSTRACT: In this paper I bring several perspectives from language change and diachronic linguistics more generally to bear on the question of how broad linguists' generalizations about speakers' linguistic behavior should be, and the extent to which they correspond to what speakers actually do themselves. Examples are drawn from the development of Greek, of English, and of Slavic, and from processes of language contact, in the Balkans and elsewhere, to address this issue. The methodological import of this investigation is to raise the question of whether linguists' analyses ought to match more closely with what speakers do as they use and change their language, and to bring diachronic and language-contact evidence into the picture in resolving this question.