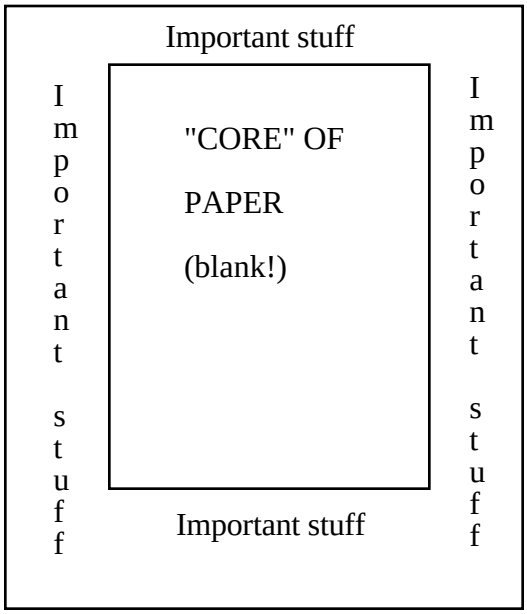
On the Linguistics of Marginality: The Centrality of the Periphery

Brian D. Joseph*
The Ohio State University

1. Introduction

Had I been more clever, and had the editors allowed it, this paper would appear entirely as marginalia, or at least with the more important points off in the margins, as in (1), where the ideal format for this paper is given:

(1)



Such a format would be highly iconic, for I claim here that despite the fact that it has long been recognized that some aspects of language are quite central to the system while others are more marginal or peripheral, and despite the fact that the core has always attracted more intense attention among linguists than the periphery has, nonetheless there is much of value in considering the margins and the peripheral aspects of language.

The reasons for the interest in the core are obvious and perfectly understandable. For instance, the core typically lends itself to a greater degree of generalization while the periphery often is the epitome of "anti-generalization", in that many peripheral facts in a language are best treated as nothing more than a list of exceptional behaviors. Also, the core is where one typically finds linguistic universals and universal grammar — as defined by those elements common to all languages, the points on which all languages intersect[1] — whereas what is typically

relegated to the margins, to the periphery of our science, are those aspects that are highly language particular.

Moreover, marginal phenomena are often hard to integrate into broader, more far-reaching analyses.

For example, with regard to phonology, certain lexical classes that are marginal in terms of their frequency, their function, and their semantics, such as onomatopoeia and iconic vocabulary in general, often show phonological properties that are not found in more frequent and functionally and semantically more indispensable lexical items; thus in Sanskrit, geminate aspirated stops, e.g. **-khkh-** or **-jhjh-**, occur only in the peripheral class of onomatopes, as in (2a), while in non-iconic words, an aspirated stop can only be geminated with its unaspirated counterpart, e.g. **-cch-** or **-tth-**, as in (2b).

(2) a. doubly aspirated, in onomatopes: a**khkh**aī-kr- 'make a noise of surprise or joy', ja**jhjh**aī- 'making a splashing sound (of rushing waters)'
b. nonaspirate-aspirate, in nononomatopes (Whitney 1889, §154): a**cch**ā 'to(ward)' (not *a**chch**ā), ar**tth**a- (optional phonetic realization of ar**th**a- 'goal' (Whitney 1889, §228), but not **ar**thth**a-)

Similarly, expressive and iconic words and foreign vocabulary (i.e., borrowings), i.e., elements that are at the margins of the lexicon in some sense, often contain segments which are not found elsewhere in the language, and which are thus marginal themselves; such is the case, for example, in Maori (see (3a)), in Sanskrit (see (3b)), in literary Macedonian (see (3c)), and in various Iroquoian languages (see (3d)):

(3) a. Maori: "some interjections expressing disgust appear to have vowels not normally part of the Maori system [e.g. [**æ**]] and can end with a voiceless glottal (sometimes as far forward as velar) fricative" (Bauer 1993:579), and the "combinations *wo, wu, who, whu* occur only in English loanwords" (Bauer 1993:542)
b. Sanskrit **jh** primarily in affective vocabulary, e.g. **jhat** 'quick as a wink', **jh**am**jh**ā 'roaring of the wind', **jh**ar**jh**ara- 'low rumbling sound; a drum'
c. Literary Macedonian vocalic /r/ (cf. Friedman 1994:252) "occurs in final position only in a few foreign and onomatopoetic words", and one of the limited contexts for schwa is "for dialectal effect in words of Slavonic or Turkish origin" (where "dialectal effect" can be interpreted as "peripheral" vis-à-vis the literary standard as the "center")
d. Iroquoian (e.g. Mohawk, Cayuga, Seneca) labial stops series are absent "in regular vocabulary, but labials occur freely in expressive terms" (Mithun 1982:53)

More generally, as Mithun (1982:49n.4) points out, there are many languages and language families that show "distinctive phonological, morphological, and syntactic behavior" for expressive vocabulary,

including Salishan, Finnish, Semai, Bahnar, Garo, Chadic, Nguni, Shona, and Yao.

Integrating facts like these into an account of the systematic generalizations evident in core vocabulary can lead to an analysis with lexical stratification, a listing of exceptions to otherwise valid generalizations, and internal inconsistencies, i.e. one that is aesthetically somewhat unsatisfying.

The temptation to ignore such phenomena, therefore, often is great, yet it is argued here that by doing so, we as linguists do our science a grave disservice and, worse, miss out on a source of information that is highly revealing about the nature of language in general.

Still, it is possible to compile a fairly long, though admittedly heterogeneous, list of linguistic marginalia and phenomena that relate to linguistic marginality in some way or another that have occupied many linguists. The collection in (4) is such a list, intended as representative not exhaustive. In some instances, the domain itself is marginal and remains so despite considerable interest in it, as with (4a) — sound symbolism, onomatopoeia, and ideophones; in other instances, the domain is marginal but has been brought into the "mainstream", as with (4b) — language games — since theorists, in phonology at least, have made use of such external evidence to judge the correctness of analyses (see Churma 1985 and Yip 1982, for instance); for some domains, the marginality is a reflection of social marginality, as with (4c) — the language of gays, thieves' slang, etc. — and (4d) — endangered languages (see Hale et al. 1992) — while for others, the social context of contact among members of different groups fosters the development of the domain, as with the simplified registers, less-than-fully-formed grammars, and loan words of (4e), (4f), and (4g); finally, some subdisciplines, as a matter of methodology, regularly rely on some level of marginality, as with psycholinguistics (see (4h)) or comparative reconstruction (see (4i)):

(4) a. language games and secret languages (Pig Latin, etc.)
b. sound symbolism, onomatopoeia, ideophones
c. language of marginalized social groups (gays, Gypsies, to some extent children, thieves, etc.)
d. endangered languages (inasmuch as they are marginal in the overall context of the sociology of the world's languages and often marginalized within their larger social context)
e. simplified registers (baby talk, foreigner talk, pidgins,[2] etc.)
f. emergent language (e.g. language of (young) children, or of second-language learners)
g. loan words and their integration into the borrowing language
h. to some extent, a lot of research in psycholinguistics, in the sense that psycholinguistic experiments create data that would not otherwise exist; the results are peripheral to naturalistic data, but it is the job of the psycholinguist (one generally quite well done) to show how this inherently peripheral activity (a speaker reacting to a stimulus in a controlled setting) can be the basis for a lesson about core linguistic processes, by developing a coherent theory of linguistic competence

i. linguistic reconstruction (relevant here is Meillet's dictum about the importance of reconstructing from irregularities, i.e. from synchronically marginal phenomena; also, note the use of "core" vocabulary as a convenient starting point in language comparison, thereby relegating other lexical comparisons to the periphery).

A thread running through the examination and characterization of many of these domains is the consideration of markedness[3] and language typology, especially insofar as they allow for the identification of some phenomena or language type as unusual in some respect and thus peripheral.

In what follows, evidence is presented to show that the examination of many of these areas has led to just as many useful insights as the study of the core alone has yielded, and perhaps even more. The observations of two Chicago linguists are relevant here and provide some (additional) authority for the discussion of the periphery to follow: McCawley (1988:154) has noted that "the core/periphery distinction will remain on shaky grounds until conclusions of some substance are drawn about the periphery", and Janda (1985:171), in discussing echo-questions, observed that "marginal data ... [can] acquire a curious aura of centrality".[4]

2. Some Claims Relevant to the Place of Marginality

The following claims provide the basis for the discussion in this paper:

(5) Under a different view of "universal grammar" from the one articulated above, even the most marginal phenomenon, one that is found in just one language, must be encompassed by universal grammar

(6) There are marginal phenomena that recur in various languages in interestingly parallel ways that, as a result of this parallelism, tell us something significant about language in general

(7) There are marginal phenomena that give a basis for very important lessons about the nature of language in general, about particular languages, and about the way we need to approach the examination and analysis of languages.

Although the third claim, that in (7), is the primary focus here, a few words about the first two are in order.

2.1. Claim #1—How Broad is Universal Grammar?

The view expressed in (5) essentially says that universal grammar can be seen as the *union* of traits found in all languages, not the intersection (as given above in section 1); in this view, there is no real "core", except as defined by markedness or frequency. Nonetheless, features found in any language become significant, even if only in a single language.[5] A dramatic case involving a sparsely attested sound

that has come to light recently is discussed by Ladefoged & Everett 1996.

They note a marginal, but possible and attested, sound, a "voiceless apico-dental plosive [followed by a] voiceless labio-labial trill" ([t̪B]) that is rare in the two Chapakuran languages in which it is attested (Wari' and Oro Win) and otherwise unattested cross-linguistically, and they ask pointedly: "How frequent does a sound have to be before it can be considered as part of the phonological system of a language? How widespread in the world's languages does a sound have to be before we take note of it when devising a universal feature set?"

Their answer to the first question is most certainly right but still a bit disappointing; they note that this unique sound, despite occurring in just a handful of words, "is not limited in the types of words in which it can occur [and so] is definitely a part of the languages investigated" (p. 798), yet this implies that restricted lexical distribution could be a reason for excluding a sound from consideration as "a part of [a] language". That is, with this statement, they seem to be suggesting that lexical peripherality can be interpreted as linguistically irrelevant; that interpretation is challenged below after a look at some other phonological marginalia.[6]

Regarding their answer to the second question, again it is most certainly correct but this time not at all disappointing. They look to the International Phonetic Alphabet as a justification for distinguishing between core and periphery with regard to sounds cross-linguistically; they note that the IPA division into sounds with their own symbols and sounds characterizable through the use of diacritics gives us the answer with regard to a rare sound like the one under discussion: "we must not overlook it, but we must also acknowledge that it is not part of the regular sounds commonly found in languages" (p. 799).

In a sense, then, distinguishing core from periphery, at least as far as the range of human linguistic sounds is concerned, but perhaps more generally too, becomes a matter of frequency, and thus markedness becomes crucial; however, markedness here is not a metric by which elements are ruled in or out of the realm of "possible human language" but rather it becomes a probabilistic measure that ties in with how common an element is cross-linguistically.

2.2. Claim #2—Cross-linguistic Parallels at the Margins

Ladefoged & Everett's implication about lexical distribution, it turns out, is directly linked to the second claim I wish to explore here.

As noted in (3), and as the Chapakuran sound they discuss illustrates also, it is often the case that a particular sound or class of sounds will be lexically restricted in a language. One such case I have discussed at some length elsewhere is the Modern Greek apico-dental affricate [ts];[7] as noted in (8), [ts] shows a skewed lexical distribution, being generally restricted to occurrence in marginal lexical categories, that is words and forms that are not essential to the function of language as a source of communication among fully competent users,[8] such as sound symbolic combinations ((8a)), interjections ((8b)), calls to animals ((8c)), onomatopoetic words and derivatives therefrom

((8d)), ideophonic adverbials ((8e)), child-language forms as conventionalized by adults for use to and around children ((8f)), and a whole host of highly expressive, playful, and generally slangy words which add color to language ((8g)):[9]

(8) a. occurs in several sound symbols, e.g. (not exhaustive listing): **tsi-** 'small, narrow, thin', (as in **tsi**tóno 'stretch', **tsí**xla 'thin woman', **tsí**ros 'thin person'); **tsV-** 'sting, tease, bite, burn' (as in **tsú**zo 'sting', **tsi**m(b)úri 'tick' ("small stinging insect"), **tsí**kna 'smell of meat or hair burning'

b. interjections, e.g. prí**ts** 'so what?!; who cares?!', **ts** 'NEGATION' (actually an apico-dental click, but conventionally represented in this way; cf. also **ts**uk as a conventionalization of this noise), **ts**á 'noise used in peek-a-boo game' (with variant *dzá*).

c. calls to animals, e.g. gú**ts** 'call to pigs', **ts**ú(nk)s 'call to donkeys', ó**ts** 'whoa!'

d. onomatopes and derivatives, e.g. **ts**ák 'crack!' (cf. **ts**akízo 'I break'), krí**ts**-krí**ts** 'crunch!' (cf. kri**ts**anízo 'I crunch'), má**ts**-mú**ts** 'kissing noise', **ts**iú-**ts**iú 'bird's chirp', plí**ts**-plá**ts** 'splish-splash!', ɣrá**ts** 'scratching sound' (with variants xrá**ts**, krá**ts**, and krá**ts** krú**ts**, and derivative ɣra**ts**unízo 'I scratch')

e. ideophonic adverbials, e.g. **ts**áka-**ts**áka 'immediate quick action; straightaway; directly', **ts**úku-**ts**úku 'steadily and surely, with a hint of secretive activity', **ts**áf-**ts**úf 'in an instant'

f. adult conventionalized child-language forms, e.g. **ts**á**ts**a 'aunty', **ts**i**ts**í 'meat' (also adult slang for 'breast'), **ts**ís(i)a 'peepee', pí**ts**i-pí**ts**i '(act of) washing'

g. expressive, playful, slangy words, e.g.: **ts**ambunízo 'whimper; prate; bullshit', **ts**alavutó 'do a slovenly job', **ts**ókaro 'vulgar woman' (primary meaning: 'wooden shoe'), **ts**irízo 'screech', **ts**ili(m)burʤó' gallivant; fart about', **ts**i**ts**iʤí '(stark) naked'.

While [ts] can be found in some perfectly ordinary and non-marginal sorts of words, such as *étsi* 'so, thus' or *tsiménto* 'cement', the overwhelming preponderance of its lexical occurrences is in words such as those in (8).

Moreover, there are other indicators of marginal status for [ts] and its voiced counterpart [dz]. As noted in Joseph 1994a, [ts] (and [dz]) are exceedingly rare in all sorts of basic (i.e. "core") vocabulary: only 3 out of 100 body-part words in Greek have a [ts] (or [dz]) in them; not one of 55 kinship terms has a [ts] and only one has [dz]; of 19 basic (and some not-so-basic) color terms, not one has [ts] or [dz]; and, a "Swadesh" list of 207 basic vocabulary items yields only one word with [ts] and none with [dz].

In addition, various frequency count statistics point in the same direction. A phoneme frequency count for [ts] (together with [dz]) done by Householder, Kazazis, and Koutsoudas 1964, building on one by Mirambel 1959, working on a "normal" corpus of lengthy passages of connected prose, found a frequency of occurrence of 0.07% for [ts] and [dz] combined, lower than any other sounds in

Greek. By contrast, a count for [ts]/[dz] based on the set of interjections and onomatopes given in Householder, Kazazis, and Koutsoudas 1964, that is on a corpus of words that are expressive in nature, finds a much greater frequency of occurrence of 4.1% for these sounds in this lexical domain. Similarly, a count of the initial segments of entries in a listing of nicknames from Kefallonia (Lorendzatos 1923) — another corpus of inherently expressive items — yielded a frequency of 6.4% for [ts]/[dz] combined.

Thus by a variety of indicators, [ts] and [dz] come out as marginal in the overall phonological system of Modern Greek. An especially intriguing dimension to this marginality is the iconicity that [ts] participates in, in the sense that this marginal sound occurs in a number of words referring to various marginal social groups, such as in the characterization of physical deformities of one sort or another and thus referring to people who are physically marginalized, as in (9a); in words for gypsies, a marginalized group that in Greece typically lives at the fringes of mainstream society, as in (9b); and in descriptions of negative character traits, the sort which would place an individual at the margins of society, as in (9c):

(9) a. (-)**ts-** in words for various deformities: **ts**evʒós 'lisping', **ts**átra-pátra 'stumblingly (especially of speech)', ku**ts**ós 'lame', ka**ts**íʒa 'balding, scurvied head', ka**ts**o- 'wrinkledy', **ts**imblíaris 'bleary-eyed'
b. (-)**ts-** in words for 'gypsy': **ts**igános, a**ts**íganos, ka**ts**ívelos
c. #**ts** in words for character flaws or negative traits: **ts**apatsúlis 'slovenly', **ts**ulís 'untidy person', **ts**igúnis 'miserly', **ts**úla 'loose-living or low-class woman; slut', **ts**ifútis 'skinflint'.[10]

The situation in (9) wherein a marginal segment is iconically involved in the designation of marginal or peripheral groups is found in other languages, and is not just a peculiarity of Modern Greek. As indicated in (10), the reconstructed word in Proto-Indo-European (PIE) for an outsider, a non-IE-speaker, *barbaro-, employed two sounds, *b and *a, which seem to have been marginal in PIE in the sense that there are not a lot of cognate forms pointing to a need to reconstruct *b[11] and *a has a restricted distribution:[12]

(10) PIE *barbaro- 'non-IE-speaking, an outsider to PIE society', reconstructible based on Greek **bá**r**ba**ros 'non-Greek; speaking an unintelligible (i.e. non-Greek) language', Latin **ba**r**ba**rus 'foreign, strange', Sanskrit **ba**r**ba**ra- 'stammering; non-Aryan', etc.

Similarly, as indicated in (11), Nichols 1986 has described the use of the lexically restricted — and thus irregular and somewhat marginal — phonological feature of pharyngealization in Chechen-Ingush in words that refer to boundaries of various sorts, i.e. elements on the margins in some sense:

(11) a. the numerals ‘1’ and ‘100’ (conceptual boundaries, the bounds of native numerals in these languages)
b. various words for sharp ends, tips, and points
c. some words for physically and/or socially marginal people, in particular the words for ‘klutz’, ‘hulk’, ‘giant’, and ‘illegitimate child’.

The lessons to be learned from this extended discussion of the lexical distribution of [ts] in Modern Greek and of other marginal elements in other languages are follows (see also Joseph 1992): we can make sense of the situations in (8) - (11) by recognizing the role of iconicity in language; a recurring trait across different languages and cultures would seem to demand an explanation in terms that refer to what it means to be a human speaker of a language embedded in a common human environment. Languages have marginal features; societies have marginal members, social borderers. Hence, finding that speakers might utilize a feature of their linguistic system as an icon of a feature of their social system — associating marginal social groups with marginal linguistic elements — is entirely in keeping with the notion that in essence language users do not impute arbitrariness to language and to linguistic signs; rather they seek some motivation for the signs, even if the resulting reflection is extra-systemic in nature. Linguistic marginalia here provide a crucial insight into the nature of language.

Thus, this recurring phenomenon of marginal elements referring to marginal social groups surely cannot be accidental, but must respond to something very basic about the way speakers use language to signify. This recurring pattern, isolated and marginal in each of languages in which it occurs but not in terms of its cross-linguistic manifestations, really brings the periphery right into center stage, and thus paradoxically must be taken to be a central part of what an overall view of language should be like — in developing a notion of what the basic design of language entails, which after all is one possible interpretation of what Universal Grammar provides, it seems we must allow for speakers to be able to utilize linguistic marginalia to satisfy their expressive purposes.

Moreover, if speakers do exploit marginality, then it seems that it is hard to accept the view that sounds that occur just in onomatopes and the like are not “a part” of a language; they surely are part of the linguistic material that speakers have available to them and are aware of, and as such must be taken into account in a serious way by linguists.

3. Claim #3—The Value of Marginalia

I move on now to my third claim, concerning the value of marginal data and the linguistic peripheries for our understanding of various phenomena in particular languages and for shedding light on the nature of language in general. As before, I give several case-studies, each with a point.

3.1. Reflexives in English

The analysis of pronominal forms with *self* in English has generally focused on their behavior in "core" reflexive pronoun uses, i.e. the most common ones, such as the object of a verb where they are coindexed with another argument of the same predicate, and in exempt anaphor constructions (Pollard & Sag 1994) where they act as pronouns licensed nonsyntactically, as in (12a) and (12b), and debate in recent years has centered on reconciling their dual status.

(12) a. Ordinary reflexives: He hurt **himself**
b. "Exempt anaphors" (Pollard & Sag 1994):
The portrait of **himself**$_{i/*j}$ hanging in Clinton's$_i$/Lincoln's$_j$ study is very flattering

Recently, though, Golde 1997 has argued that "self-pronouns should not be considered ambiguous between reflexives and exempt anaphors (as is assumed in e.g. Pollard and Sag 1994) but rather as unambiguous lexical items which are licensed by different syntactic and non-syntactic constraints (the view taken by Kuno 1987 and Reinhart and Reuland 1993, for example)", and has drawn attention to more marginal, but nonetheless real, uses such as those in (13a-c), which she collected from (early) modern American English, in (13d) from current usage, and (13e), cited from Joseph 1979, that show that both uses of self-pronouns "should be analyzed compositionally, as a possessive pronoun followed by the nominal *self*" rather than as "instances of a pronoun with a particular morphological marking".

(13) a. I shall end my letter, my dear Mr. Franklin, with a personal application to **your** proper **self** (B. Franklin, *Autobiography of Benjamin Franklin* (1771))
b. When Atmananda sensed that she was not **her** usual, happy **self**, he did not openly communicate his displeasure (M. Laxer, *Coming of Age in a Destructive Cult* (1993))
c. ...one thought **one's self** dealing with honest men or with rogues (H. Adams, *The Education of Henry Adams* (1905))
d. Somebody must have hurt themself/*?themselves
e. (Nurse to patient:) We seem a bit displeased with ourself / *ourselves today

The availability of this compositional analysis for both "core" uses of *self*-pronouns is significant, but from a methodological standpoint, it is significant that the crucial evidence comes from marginal constructions. This may in fact be the norm in syntactic investigations, for the tests of analyses often come from pushing the syntax to its limits, seeing how constructions fare at the margins of acceptability, for instance, and Golde's work on reflexives is a case in point. The margins thus prove to be decisive in her analysis.

3.2. Morpholexically Particular Phonology

As a second example, I turn to the *a*/*an* alternation in the English indefinite article. It is a marginal one in English morphophonology,

inasmuch as it is the only case, synchronically speaking,[13] of an alternation of Ø with *-n* present in the language. However, despite the fact that it is isolated, this alternation is quite regular, and with the exception of some inter- and intra-speaker variability observable before initial *#h* (e.g. *a(n) historic occasion*), the distribution can be stated in purely phonological terms, with *a* occurring before consonants (e.g. *a book*/**an book*) and *an* before vowels (e.g. *an elephant*/*??a elephant*[14]). Thus this alternation tells us that there can be phonological effects that are tied to a single lexical item. However, if phonology is a domain of maximal generality, with only phonological features mattering for phonological alternations, it is hard to see this alternation as purely a matter of phonology, even though it does pertain to the pronunciation of a morpheme in different phonologically-characterizable environments

There is thus a paradox of sorts, for the *a/an* alternation is regular (or very nearly so), and so is the sort of phenomenon that is typically captured by a rule, yet it is lexically particular to the extreme, being limited to a single item. We therefore have to reconcile its lexical specificity with its apparent rule-governed-ness.

A way to accomplish this end is to think of the rule governing a/an not as a (general) phonological rule, e.g. Ø —> n /æ__ #V, but instead as a morphological rule, e.g. // a // —> // an // /__ #V. This means, though, that morpheme-specific phonological effects need to be recognized, and these effects can be built in as a process as part of the realization rule for the indefinite article morpheme. Such a treatment takes the alternation out of the realm of phonology and places it in the morphology, where it seems to properly belong.

Thus, the solution is to see this alternation as an inherently morphological phenomenon, i.e. the result of a process by which the indefinite article morpheme is realized, or spelled out. That might seem like a relatively benign step to take, but it is really a dramatic one, for it opens the door up for viewing virtually all apparent phonological alternations as part of the morphological processual rules by which morphemes are realized. That is, once we admit that we really have to recognize lexically particular phonology, then if we have to recognize it for one form, why not generalize that view to other alternations even if they are more widespread. In such a view, velar softening, for instance, which "mediates" between allomorphs such as *opaque* [owpeyk-] and *opac-* [owpæs-], the latter conditioned by the suffix *-ity*, can be recast not as a phonological rule but instead as part of the suffixation operation, a process that is invoked by a particular type of suffixation. In this view, a lot of what is typically referred to as phonology is instead treated as morphology, and the role of phonology proper is more restricted.

This one little set of forms, therefore, isolated though they are within English, gives us reason to rethink the way we account for morphophonemic alternations, and the way we conceive of the relationship between morphology and phonology and the place of morphology within a grammar.

3.3. Recently Borrowed Verbs in Bulgarian

As another case, let us consider the treatment of recently borrowed verbs into Bulgarian, as discussed by Jill Neikirk Schuler 1996. Her examination of how these new verbs are handled with respect to aspect, generally considered to be a critical grammatical category in Bulgarian, is very revealing as to the actual status of aspect. Bulgarian is a language in which verbs typically occur in aspectual pairs, in an imperfective and a perfective form, though it is important to note that there are several native Slavic "biaspectual" verbs where the same form serves for both aspects, e.g. *moga* 'be able' and *menja* 'change'. Nonetheless, the native biaspectuals constitute a closed class, so that when borrowed verbs, which generally arrive with just a single aspectual form, enter the language, something has to give, so to speak; either new aspectual forms must be created, through a type of nativization, or new entries into the class of biaspectuals must be tolerated, thus indicating a change in the nature of aspect as a grammatical category.

Based on a questionnaire, Neikirk Schuler found great variation, of a very revealing nature, in the way Bulgarian speakers treat newly borrowed verbs. 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, and interestingly, responses generally were not uniform for all verbs or for all speakers. For some verbs, most notably *korespondiram* 'correspond', all 49 speakers she worked with did not allow a prefixed perfective form. Her findings for other verbs showed considerable variability as to whether a prefixed perfective form was tolerated and if it was, what prefixes were allowable, as indicated in (14), where the numbers in parentheses show the number of responses for a given prefix or for no prefix, for two verbs in her sample:

(14) a. *komentiram* 'comment': prefixed perfective *iz* (8), *ot* (7), *do* (2), *pro* (1), unprefixed (31)
b. *kodirvam* 'code': prefixed perfective *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 formally aspectualize borrowed verbs, no 100% true generalization can be made over all speakers nor over all borrowed verbs. She concludes that loan words are marginal in the language yet allow for a very important insight concerning aspect: they are marginal in that they enter at the peripheries of the grammar, yet their treatment by speakers suggests that "all ... Bulgarian ... verbs [i.e. those fully part of the language] have aspect, but not all verbs in ... Bulgarian [i.e. in the language to any extent] ... have aspect". In a sense, then, marginal elements, namely loan words, reveal that the native biaspectuals, also somewhat marginal in that they

are a restricted closed class, occupy a more important place in the grammar than one might at first suppose.

Thus aspect cannot be considered an obligatory category in Bulgarian, for there are verbs in the language, especially new (and thus marginal) ones as well as old (but still somewhat marginal) ones for which it is not obligatorily marked. Significantly for my purposes here, this lesson is learned from the treatment of loan words, items on the margins of the grammar, and the consideration of a marginal class of native words.

3.4. Hyperforeignization—Marginal but Rule-Governed

I turn now to a different angle on foreign words as my next case-study. Janda, Joseph, & Jacobs 1994 call attention to a phenomenon, which they refer to as "hyperforeignization", whereby speakers (re)shape words perceived to be foreign in ways that do not nativize them, but rather "anti-nativize" them, i.e. make them sound more foreign, even if the foreign sound is not technically appropriate to the target language. For example, for many American English speakers, *smorgasbord* is pronounced with initial [ʃm-] even though the proper Swedish pronunciation has [sm-]. This hyperforeignization process typically affects words that are recent loans or are perceived to be loans (e.g. foreign proper names); such words are at the peripheries of the language, and so in a real sense, perhaps, are not fully part of the borrowing language. Yet this phenomenon is important, for it gives evidence of rule-governed linguistic behavior, in sense of being a "strong pattern, an observed regularity, something that can be formulated",[15]. One such rule can be informally stated as "characterize as foreign through the use of palatal sibilants", as with *smorgasbord*, or with the common pronunciations of *Taj Mahal* with [taʒ] instead of the correct [taʤ], or of Beijing as [beyʒɪŋ] instead of the more accurate [beyʨɪŋ]; even more significantly, these "rules" are quite limited in scope, suggesting that linguistic generalizations are best treated as very "localized" in nature, i.e. as "local generalizations", rather than as wide-ranging ones, a point brought out by the material in the remaining sections.[16]

3.5. Phonesthematic Attraction—Robustness at the Margins

As a final example involving sounds, consider the phenomenon of "phonesthematic attraction", which shows that even marginalia can have some robustness. This term refers to situations in which sound symbolic clusters of words, marginal though they may be, nonetheless can draw other words into their "orbit", so that these other words change their form in the direction of the sound symbol. A case in point (discussed in Hock & Joseph 1996:293) comes from the history of the cluster of words in English that end in [-æg] and have the meaning of "pertaining to slow, tiring, tedious action", i.e. *drag*, *fag*, *flag*, *lag*, and *sag*. As indicated in (15a-d), the first four of these words have varied origins, some being inherited and some being borrowings, but all etymologically have a [g] in them; still, by the 15th century, they were present in English and presumably formed a cluster of related words, with [-æg] thus being analyzable as a sound symbol, or

phonestheme. As noted above, sound symbolism is generally considered to be a marginal phenomenon, at least to the extent that it is somewhat unsystematic — for example, *bag* is not a member of the set despite fitting the formal criterion of ending in [-æg], nor is *droop* despite its appropriate semantics; nonetheless, this cluster of words proved to have some strength, for they pulled the semantically similar *sacke* in (e) into their form, resulting in *sag*:

(15) a. *drag* 'lag behind' < ME *draggen* < either OE *dragan* or ON *draga* 'drag, pull'
b. *fag* 'exhaust, weary, grow weary', presumably < ME *fagge* droop'
c. *flag* 'hang limply; droop', probably of Scandinavian origin, from a word akin to Old Norse *flögra* 'flap about'
d. *lag* 'fail to keep up; straggle' < earlier English *lag* 'last person', ME *lag-* 'last', possibly from Scandinavian
e. *sag* 'sink; droop' < 16th century Engl. *sacke*, ultimately probably of Scandinavian origin, compare Swed. *sacka* '(to) sink'

Thus despite being marginal and restricted to just a few lexical items, the [-æg] 'tiring' sound symbol showed enough life within its limited domain to influence the form of *sacke* not long after this word entered the language as a borrowing. Being on the margins, therefore, does not relegate a linguistic phenomenon to being a 98-pound weakling!

3.6. Marginal Robustness in Morphosyntax

Evidence of the sort of limited robustness shown by [-æg] comes in morphosyntax also, and examples in this domain too can lead to some important lessons about the nature of linguistic generalizations. A particularly interesting case is the development of weak subject pronouns in Modern Greek.

As the result of several developments in the Medieval Greek period, Greek now 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 ...?'; these are exemplified in (16):

(16) 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

The ultimate source of this construction in Greek is language contact,[17] for it appears that *ná* is a borrowing from South Slavic and that the original syntax in Greek with *ná* was as in (17a), with a deictic element followed by a weak accusative (ACC) pronoun, a widespread South Slavic pattern, as exemplified by Bulgarian in (17b) but with parallels in Macedonian and Serbo-Croatian:

(17) a. ná ton 'Here he is'
here-is him/ACC.MASC.SG.WEAK

b. eto go ‘Here he is’
here-is him/ACC.MASC.SG.WEAK

The steps that then led to forms like *tos* can be summarized as follows: by an extension within Greek alone, strong accusative forms became possible after *ná*, so that sequences like *ná aftó* ‘Here it (lit. “this-thing”) is!”. Then the *ná + ACC* pattern was reinterpreted as *ná + NOM(INATIVE)*, presumably through the medium of strong neuter forms whose NOM and ACC were identical, so that these several patterns were coexistent. A proportional analogy involving strong and weak masculine (M) accusative forms as well as the innovative use of a strong nominative form after *ná* led to constructions such as *ná tos*, with a new category of weak nominative pronoun. This final step is schematized below in (18):

(18) ná aftón : ná ton :: ná aftós : X, X --> ná tos
M.ACC.STR M.ACC.WK M.NOM.STR M.NOM.WK

Finally, in the development that is of particular significance here, via an analogical extension the highly restricted pattern of *ná* plus weak nominative then spread to use with *pún* ,[18] thus showing some very limited robustness for a marginal construction. The result, moreover, was a very localized generalization regarding the use of weak subject pronouns, namely with these two semantically linked predicates, ‘where-is/are?’ and ‘here-is/are’.

Hittite offers a highly relevant parallel to Greek *tos*, in that weak nominative pronouns in Hittite (e.g. masculine singular *-as*) seem to have arisen analogically (so Garrett 1990) like *tos* did, but have a wider distribution, occurring only with intransitive verbs. The Greek case provides a suggestive basis for understanding how the quite general and systematic use of weak nominative pronouns in Hittite with intransitive verbs could have arisen in a piece-meal fashion, spreading from verb to verb, just as *tos* has spread in a limited way in Greek, until a generalization over the class of verbs participating in this construction was possible.

3.7. The Emergence of Morphological Paradigms

Finally, there is evidence that morphological paradigms, just like the syntactic “paradigm” *ná tos* / *pún dos* in Greek, can be built up piece-by-piece, and need not emerge fully formed from the heads of speakers. For one thing, there are defective paradigms that seem never to have been filled out and which have arisen piecemeal, e.g. Ancient Greek ἠμί ‘I say’, which was synchronically defective, with only 1st and 3rd person singular present and past forms, as in (19):

(19) 1SG.PRES ē-mi 1SG.PST ē-n
3SG.PRES ē-si 3SG.PST ē

Interestingly, moreover, this verb seems to have been built up with a single form as the starting point. Only the 3SG.PST is an inherited

form; the others would be expected to have turned out differently from their presumed Pre-Greek forms, as shown in (20):

(20) Pre-Greek	1SG.PRES	*ē̆g-mi	should give Greek	ē̆g-mi*
	3SG.PRES	*ē̆g-ti		ē̆k-si*
	1SG.PST	*ē̆g-m		ē̆g-a*
	3SG.PST	*ē̆g-t		ē̆

and so must have been re-formed with regular Greek verbal inflexional morphology for each form added to the 3SG.PST base. Thus, when the conditions for new paradigms are manifested, e.g. a split from an existing paradigm, the new paradigm can emerge just through a single form.[19]

Such may well be the case also with Modern American English *hafing to* , where all there seems to be at present is the starting point for a new, fully elaborated paradigm. In particular, some American English speakers have separated the obligational *have to* ([hæftu]) from the main/auxiliary verb *have*, a separation which is evident in the gerund form *hafing to* (Joseph 1994b). These speakers appear to have based the innovative gerund form on the [hæf-] of the surface form [hæftu], thus severing the connection with the original base verb *have* ([hæv]); apparently, though, they do not have a more elaborated "paradigm", only the gerund,[20] yet, the gerund alone gives evidence of a new but highly restricted, defective, and thus marginal, verb.

4. Some general conclusions

What these case studies show is that speakers have a considerable amount of knowledge that is highly particularized, and more specifically that is keyed to the behavior of particular lexical items; this observation suggests that the real basis for generalizations — which are usually thought of as representing the "core" of grammar and of language — is a set of extensions over an ever-wider set of particular facts. Linguistic generalizations, in this view, are really post-hoc generalizations that summarize a situation, more or less like redundancy rules ; they are generally static, but it seems that speakers can exploit them dynamically, under certain conditions.[21]

In a sense, then, all linguistic knowledge starts out as marginal — only after enough bits of information have accumulated and the pieces begin to be fit into a system do broad "generalizations" emerge. The smaller, more local, generalizations are what speakers exploit dynamically, as the case of phonesthematic attraction shows, and the broader generalizations, perhaps, are more static phenomena.

Linguistic marginalia, therefore, lead to this crucial insight into the nature of language and linguistic competence, and as such, are deserving of a more prominent place in linguistic debate and analysis.

---

*The experiences that have shaped my thinking on this subject extend over two decades and involve more people than I could possibly recognize here, yet I would like to acknowledge the recent very important and very useful comments from the

CLS audience, and in particular Alexis Dimitriadis, Matthew Dryer, Victor Friedman, Eric Hamp, Richard Janda, Knud Lambrecht, Salikoko Mufwene, and Haj Ross. I alone bear responsibility for what I may have done with their advice.

[1]See Mufwene 1990 for some discussion of different views of Universal Grammar, including this one.

[2]Note that I do not include creole languages here, on the assumption that once a language is a full-fledged one in terms of its range of communicative functions, its degree of expressivity, its place in a speech community, and its having native speakers who transmit the language in a normal fashion, then it is not marginal in any sense. Pidgins, however, as a type of simplified register, are in a different category.

[3]See for instance Battistella 1990, 1996 and Andrews 1990 for recent views on markedness theories in general; Mufwene 1990 has some interesting observations on markedness and universal grammar.

[4]My thanks to Rich Janda for bringing these articles to my attention; I note with some pride that though now at Chicago, Janda was a faculty member at the Ohio State University when he wrote the paper cited here.

[5]Matthew Dryer, in his talk at this year's CLS conference, gave several examples of linguistic features found in just one language; for instance, a bilabial trill is attested only in Kele, a New Guinea language, and only the Australian language Nunggubuyu shows a four-fold phonemic contrast in which dental, alveolar, palato-alveolar, and retroflex stops are differentiated. See Dryer (this volume) for more examples and further discussion.

[6]I hope I am not being unfair to Ladefoged and Everett with this interpretation of their statement. If they did not intend what I read out of (or into) their statement, still it raises the important question of what it means to be "part of a language".

[7]I label [ts] an affricate here, realizing that there is some controversy in Greek, as perhaps with similar sounds in virtually every language that has them, as to whether they are true affricates or instead are consonant clusters. See Joseph & Philippaki-Warburton (1987: 231-2, 238) for some discussion of the analytical ambiguities.

[8]This is not to say that language is only for adults talking to adults for purely informational purposes, but rather to suggest that the practice among linguists is generally to make such an assumption. Quite to the contrary, in my view, these lexical categories to a large extent make language fun, give it life and color, and allow for individuality in expressiveness.

[9]See Joseph 1994a for further discussion of this case and additional references.

[10]This word bears some relationship to Turkish *çıfıt* 'Jew, stingy', and interestingly, Jews represented a marginal group in Ottoman society; although the standard view is that the Greek form is a borrowing from the Turkish (so Andriotis (1983: s.v.)), Victor Friedman (p.c., 4/19/97) has informed me that the Turkish form may well be a borrowing from Greek, since Turkish also has *Cühudi* (from Persian but ultimately from the Semitic designation, such as Hebrew *yehuda*) as a learned form for 'Jew'. Whatever the direction of the historical relationship between Greek and Turkish with regard to this word, the reference to marginally acceptable personality traits is present synchronically in Modern Greek and that is what is most relevant for the claims about [ts] put forward here.

[11]This fact is of course one of the reasons that some scholars have doubted the existence of *b in PIE or have claimed that a reevaluation of the phonetics of the reconstructed sound system is needed; I happen to believe that the evidence for *b is sufficient to warrant reconstructing such a sound — for me, a sound can be part of the language even if it is marginally attested. See Joseph 1985 for some discussion.

[12]That is, *a (importantly, not the *a*-vowel that results from laryngeal coloring) occurs mainly before or after *s and any of the PIE gutturals; see Meillet 1964 for discussion.
[13]Though historically parallel to *a/an*, the *my/mine* alternation is not parallel at all in modern usage, since *my* and *mine* are different words, each with its own distributional range. Occasional simple possessive uses of *mine* for *my*, as in *Mine eyes have seen the glory...* ,can be treated as belonging to a different register from the one under consideration here.
[14]One can in fact hear phrases like *a elephant*, especially if there is a pause between the article and the noun (e.g. when a speaker is searching for a word) or in rapid or casual speech as a sort of speech error; the possibility of using *a* before a vowel-initial word is probably a function of //a// (now) being the basic form, and //an// a variant of it.
[15]I am indebted to Eric Hamp (p.c. 9/95) for this characterization.
[16]See also Joseph & Janda 1988, Joseph 1996, and the references there.
[17]See Joseph 1981, 1994c for more detailed discussion.
[18]*pun* itself is an innovative predicate formed from a contraction of *pú* 'where?' and *íne* 'is/are'
[19]That the 3rd person should be that single-form basis is the insight known as "Watkins' Law" (from Watkins 1962, see also Collinge 1985 for discussion and references).
[20]As an alternative to saying that [hæf-] is a defective verb occurring only in the gerund, one could say that *hafing* speakers have a verb [hæf-] and a verb [hæv-] that overlap in their paradigms in all forms except the gerund; see also Pullum 1997:88 for some discussion.
[21]For more discussion about the nature of linguistic generalizations, see Joseph 1996 and the references there.

References

Andrews, Edna. 1990. Markedness theory — the union of asymmetry and semeosis in language. Durham: Duke University Press.

Andriotis, Nikolaos. 1983. Etimolojikó leksikó tis kinís neoelinikís. Thessaloniki: Idrima Manoli Triandafilidi.

Battistella, Edwin L. 1990. Markedness: the evaluative superstructure of language. Albany: State University of New York Press.

Battistella, Edwin L. 1996. The logic of markedness. New York: Oxford University Press.

Bauer, Winifred. 1993. Maori. London & New York: Routledge.

Churma, Donald G. 1985. Arguments from external evidence in phonology. New York: Garland Publishing Co. [Published version of 1979 Ohio State University Ph.D. dissertation].

Collinge, N. E. 1985. The laws of Indo-European. Amsterdam: John Benjamins.

Friedman, Victor. 1994. Macedonian. The Slavonic languages, ed. by B. Comrie & G. Corbett, 249-305. London & New York: Routledge.

Garrett, Andrew. 1990. The origin of NP split ergativity. Language 66.2.261-296.

Golde, Karin. 1997. The value of marginal data: insights into the compositional nature of *self*-pronouns. Unpublished OSU manuscript.

Hale, Kenneth, Michael Krauss, Lucille J. Watahomigie, Akira Y. Yamomoto, Colette Craig, LaVerne Masayesva Jeanne, & Nora C. England. 1992. Endangered languages. Language 68.1.1-42.

Hock, Hans Henrich & Brian D. Joseph. 1996. Language history, language change, and language relationship. An introduction to historical and comparative linguistics. Berlin: Mouton de Gruyter.
Householder, Fred W., Kostas Kazazis, & Andreas Koutsoudas. 1964. Reference grammar of literary dhimotiki. Bloomington: Indiana University.
Janda, Richard D. 1985. Echo-Questions are evidence for what? CLS 21.171-188.
Janda, Richard D., Brian D. Joseph, & Neil Jacobs. Systematic hyperforeignisms as maximally external evidence for linguistic rules. The reality of linguistic rules, ed. by S. Lima, R. Corrigan, & G. Iverson, 67-92. Amsterdam: John Benjamins.
Joseph, Brian D. 1979. On the agreement of reflexive forms in English. Linguistics 17.519-523.
Joseph, Brian D. 1981. On the synchrony and diachrony of Modern Greek NA, Byzantine and Modern Greek studies 7.139-154.
Joseph, Brian D. 1985. Proto-Indo-European consonantism: methodological and further typological concerns. Papers from the 6th International Conference on Historical Linguistics, ed. by J. Fisiak, 313-321.. Amsterdam: John Benjamins.
Joseph, Brian D. 1992. Greek perspectives on the question of the arbitrariness of linguistic signs. Modern Greek Studies Yearbook 7.335-351.
Joseph, Brian D. 1994a. Modern Greek *ts*: beyond sound symbolism. Sound symbolism, ed. by L. Hinton, J. Nichols, & J. Ohala, 222-236. Cambridge: Cambridge University Press.
Joseph, Brian D. 1994b. Diachronic explanation: putting speakers back into the picture. Explanation in historical linguistics, ed. by G. Davis & G. Iverson, 123-144. Amsterdam: John Benjamins.
Joseph, Brian D. 1994c. On weak subjects and pro-drop in Greek. Themes in Greek Linguistics, ed. by I. Philippaki-Warburton, K. Nicolaidis, & M. Sifianou, 21-32. Amsterdam: John Benjamins.
Joseph, Brian D. 1996. How general are our generalizations? What speakers actually know and what they actually do. Paper read at 13th Annual Meeting of the Eastern States Conference on Linguistics; to appear in ESCOL '96 (Proceedings of the 13th Eastern States Conference on Linguistics).
Joseph, Brian D. and Richard D. Janda. 1988. The How and Why of Diachronic Morphologization and Demorphologization. Theoretical morphology: approaches in modern linguistics, ed. by M. Hammond & M. Noonan, 193-210. San Diego: Academic Press.
Joseph, Brian D. and Irene Philippaki-Warburton. 1987. Modern Greek. London: Croom Helm Publishers.
Kuno, Susumu. 1987. Functional syntax. Chicago: University of Chicago Press.
Ladefoged, Peter & Daniel Everett. 1996. The status of phonetic rarities. Language 72.4.794-800.
Lorendzatos, P. 1923. Simbolí is ta neoeliniká paronímia. Leksikografikón arxíon tis mésis ke néas elinikís 6.40-71.
McCawley, James. 1988. Comments on Noam A. Chomsky, "Language and problems of knowledge". University of Chicago Working Papers in Linguistics 4.148-156.
Meillet, Antoine. 1964. Introduction à l'étude comparative des langues indo-européennes. University: University of Alabama Press.
Mirambel, André. 1959. La langue grècque moderne. Description et analyze. Paris: Librairie Klincksieck.

Mithun, Marianne. 1982. The synchronic and diachronic behavior of plops, squeaks, croaks, sighs, and moans. International Journal of American Linguistics 48.1.49-58.
Mufwene, Salikoko. 1990. Creoles and universal grammar. Linguistics 28.783-807.
Neikirk Schuler, Jill. 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.
Nichols, Johanna. 1986. The semantics of pharyngealization in Chechen and Ingush. Paper read at the Conference on Sound Symbolism, University of California, Berkeley (January 1986).
Pollard, Carl J. & Ivan Sag. 1994. Head driven phrase structure grammar. Chicago: University of Chicago Press.
Pullum, Geoffrey K. 1997. The morpholexical nature of English *to*-contraction. Language 73.1.79-102.
Reinhart, Tanya & Eric Reuland. 1993. Reflexivity. Linguistic Inquiry 24.657-720.
Watkins, Calvert W. 1962. Indo-European origins of the Celtic verb. Vol. I: The sigmatic aorist. Dublin: Institute of Advanced Studies.
Whitney, William D. 1889. Sanskrit Grammar. Cambridge: Harvard University Press.
Yip, Moira. 1982. Reduplication and C-V skeleta in Chinese secret languages. Linguistic Inquiry 13.4.637-661.