The Challenge of Typologically Unusual Structures[1]

Alice C. Harris
SUNY Stony Brook
acharris@notes.cc.sunysb.edu

1. Introduction

Among the types of explanation that have been offered for typologically unusual structures are claims that the structure is rare because

- our innate endowment discourages this structure (perhaps as part of a more general feature)
- this structure does not function well
- this structure cannot be acquired easily by children
- this structure is not easily processed.

All of these proposed explanations share several problems. (i) In some cases there is no direct evidence to indicate what information our innate endowment provides about the structure at issue. In some cases evidence that the specific structure functions poorly, or is difficult to acquire, or is difficult to process is also lacking. (ii) In many instances, the reasoning that supports the proposed generalization is circular: This structure is rare because it does not function well (or is difficult to acquire, or is difficult to process, or is not part of our innate endowment), and we know this because the structure is rare. (iii) In many instances, including those discussed below, the unusual structure has existed for a very long time. If it is not easily acquired (or not easily processed, or not part of our innate endowment, or dysfunctional) how do we explain its longevity? (iv) None of the explanations summarized above explains why a few languages do have the structure or feature at issue. If it is not easily processed (or is not innate, or does not function well, or is difficult to acquire) how and why do some languages manage to have this feature or structure? If one or all of the explanations above are correct, we must still explain under what circumstances a dispreferred structure or feature may exist and under what circumstances it may not.

In this paper I argue that in many instances there is a different kind of explanation for typologically unusual features or structures. In many cases such a structure is the result of a complex series of very ordinary diachronic changes. I am suggesting that there is nothing unusual in any of the changes; the only thing unusual is the fact that all occur together here, and in a manner and order that produce this system.

---

[1] A different version of this paper was presented at a workshop, "Explaining Linguistic Universals: Historical Convergence and Universal Grammar", held at the University of California, Berkeley in March 2003, and a more complete version of it will be published with the papers from that conference.
The research reported here was supported in part by the National Science Foundation under grant BCS 0215523; gathering and analysis of data were supported by earlier grants, including a National Science Foundation National Needs Postdoctoral Fellowship (1978–79), the American Council of Learned Societies' exchange with the Academy of Science of the USSR (administered by the International Research and Exchanges Board, 1981, 1989), and National Science Foundation grants BNS-7923452, BNS-8217355, and SRB-9710085.

G. Booij, E. Guevara, A. Ralli, S. Sgroi & S. Scalise (eds.), *Morphology and Linguistic Typology, On-line Proceedings of the Fourth Mediterranean Morphology Meeting (MMM4) Catania 21-23 September 2003*, University of Bologna, 2005. URL http://morbo.lingue.unibo.it/mmm/

In section 2 below, I describe one typologically rare structure and propose that its rarity is not explained as the result of our innate endowment, its inability to function, the difficulty of its acquisition, the difficulty of processing it, or by any universal rule specifically outlawing it. Rather, it is suggested, it results from a complex sequence of quite ordinary diachronic events. The structures at issue are the endoclitics of Udi, a language of the North East Caucasian family. In section 3, I briefly review some additional structures from other languages and suggest similar explanations for their relative uncommonness.

2. Endoclitics

A set of person-number clitics (PM's), a subjunctive clitic *-q'a*, and a now moribund conditional clitic *-gi* occur in a number of positions in Udi, a member of the Lezgian subgroup of the North East Caucasian language family. As illustrated below, these may occur enclitic to the verb form (1), enclitic to a negative (2), question word (3), or other focused constituent (4), between morphemes in the verb form (5), inside the root of monomorphemic verbs (6), and in other positions.

(1) [h]at'ia xe bak-al-**[l]e**[2]
right.there water be-FUTII-**3SG**
'There will be water there'

(2) juɣab-a te-**ne** ta-d-e
answer-DAT NEG-**3SG** give-LV-AORII
'He did not give the answer'

(3) ek'a-**va** buq'-sa?
what-**INV2SG** want-PRES
'What do you want?'

(4) gölö kala häzirluɣ-**ne** bak-sa
much big preparation-**3SG** be-PRES
'There is much preparation'

(5) bar-k'-en ta-q'a-**n**-c-i
permit-LV-HORT go-SUBJV-**3SG**-LV-AORI
'Let us permit [her] to go'

(6) a-**ne**-q'-o sa kisak' q'əzəl
take$_1$-**3SG**-take$_2$-FUTI one purse gold
'She takes a purse of gold'

[2] Abbreviations used in glossing include ABSL absolutive, AOR aorist, COP copular, DAT dative, ERG ergative, FM focus marker, FUT future, HORT hortative, INV inversion, LV light verb, NEG negative, PRES present, PTCPL participle, SG singular, SUBJV subjunctive. The following additional abbreviations are used in structures: Agmt agreement, FOCC$_i$ focused constituent, INCE incorporated element, PM person marker, SUBJ subject, SUF suffix.

The PM's are in bold, and each is third person singular *-ne*, except in (3), where a special form of the second person singular is used; in (1) *-ne* assimilates and reduces, and in (5) it reduces to *-n*, a change that is regular after the subjunctive clitic *-q'a*. The conditions on their occurrence in each of these positions are stated in Harris (2002, Chapter 6) and more briefly in Harris (2000). All examples in this paper are from a text, "Taral", collected in 1989 and not yet published.

Enclitics are not typologically unusual, and it is only the endoclitics in (5) and (6) that need in some sense to be explained. These are endoclitics, not infixes, according to a variety of criteria widely accepted in the field (Harris 2000, 2002: 94–114).[3] Similarly, the sequences within which they occur are words, not phrases, according to well established tests (Harris 2000, 2002: 76–87). I suggest that most languages lack endoclitics because their origin requires a number of steps, which do not often occur together in the necessary order.

A number of languages of the North East Caucasian family have focus cleft constructions similar to the one illustrated in (7b) from Dargi (examples from Kazenin 1994, 1995; see also Kazenin 2002).

(7) a. x'o-ni uzbi arkul-ri
2SG-ERG brothers.ABSL bring.PAST-2SG
'You brought the brothers.'

b. x'o saj-ri uzbi arku-si
2SG.ABSL FM[COP-2SG] brothers.ABSL bring-PTCPL.SG
'YOU brought the brothers.' 'It was YOU that brought the brothers.'

(8) Dargi
[$_S$ FocC$_i$ Copula-Agmt$_i$ [$_S$ .... Verb]]
SUBJ PARTICIPLE

Because this construction is so easily borrowed, it cannot be reconstructed to the proto-language; but it is likely that Udi had this construction, widespread among other languages of its family. Udi lost the inherited gender-number agreement, and the agreement shown in (7) is a language-specific development of Dargi. In Udi, it is likely that a pronoun coreferential to the focused constituent (FocC) introduced the embedded clause.

(9) [FocC$_i$ Copula [$_S$ that$_i$ .... Verb]]
SUBJ PTCPL

Udi lost its copula in ordinary equational sentences. Although this may have occurred before the structure shown in (9), it is shown here as following, in (10).

(10) [FocC$_i$ ∅ [$_S$ that$_i$ .... Verb]]
SUBJ COPULA PTCPL

[3] Even an analysis that claims that there is no such thing as a clitic, only affixes, must explain why there are not more languages with "infixes" that can also occur outside the verb, as this one can. That is, the need for typological explanation remains even to the linguist who denies the existence of clitics.

Diachronically, biclausal focus structures are often reanalyzed as monoclausal (Harris and Campbell 1995: Chapter 7), and this very common change occurred also in Udi, yielding the structure shown in (11).

(11) [FOCC -PM ... Verb ...]
FINITE

The PM in (11) is derived from the pronoun indicated as 'that$_i$' in (10). In the first person singular, for example, the independent pronoun is *zu* and the PM is *zu*. The second person forms have undergone some changes, and the third person forms are not yet well understood. Third person forms occurring in sentences with the structure of (10) may have been *t'e* 'that' or **no*.[4] The independent pronouns of (10) cliticized to the focused constituent in a way that is known to occur; for example, in Somali, subject pronouns cliticized to a focus marker in the formation of the focus construction (Harris and Campbell 1995 and sources cited there).

While the structure in (11) is attested in 19th century Udi, it has been replaced almost entirely with the structure in (12), where the focused constituent occurs immediately before the verb.

(12) [... FOCC -PM Verb ...]
FINITE

This is a common position for a focused constituent, occurring, for example, in Hungarian, Korean, and Armenian (Kiss 1995, Lambrecht and Polinsky 1997: 197).

For some combinations, the structure in (12) was reanalyzed as a lexicalized phrase, and this in turn was reanalyzed as a complex verb. The structure in (12) was not itself lost and continued to exist beside the reanalyzed structure. Lexicalized phrases and complex verbs consisting of noun-verb or adjective-verb are very common in the Lezgian subgroup, to which Udi belongs, and indeed in the family as a whole. For example, one or both of these constructions are found in the following other Lezgian languages: Lezgi, Tabasaran, Rutul, Tsaxur, Budukh, Khinalug, and Archi. In Udi, the lexicalized phrases were formed from (12), with the focused constituent becoming the incorporated element (IncE), and the verb becoming a light verb (LV) in many instances, as in (13).

(13) [INCE-PM-LV]$_V$

During the process of univerbation, or consolidation of a verbal element and incorporated element, the PM, enclitic to the incorporated element, was trapped between these two lexical elements. A similar process in Indo-European has been discussed by Jeffers and Zwicky (1980), Klavans (1979), and Watkins (1963, 1964), among others, and this is discussed as a general process in language in Yu (2003). (13) represents the structure of the verb in (5) above, one of the unusual structures we are trying to explain.

[4] The form *t'e* in the modern language occurs only before a noun, e.g. *t'e išu* 'that man'; ***no*** occurs in the modern language only as parts of deictic pronouns – *meno* 'this one', *kano* 'that one (close by)', *šeno* 'that one (distant)'.

The last structure to explain is that illustrated in (6), in which a monomorphemic verb root is divided by a PM. This developed, at least in part, through analogy to the structure in (13). All of the light verbs in Udi, except *-bak-* 'be, become', consist of a single consonant. Most of the time, then, the PM in (13) occurs between the incorporated element and a consonant, followed by the tense-aspect-mood suffix.

(14) [INCE-PM-C-SUF]$_V$

The structure of monomorphemic verbs can be analyzed on this pattern:

(15)	[INCE- PM- C-SUF]$_V$	[CV- PM- C -SUF]$_V$
	ci- ne- p- e	bẹ- ne- ɣ- e
	down-3SG-LV-AORII	see$_1$- 3SG-see$_2$-AORII
	'she poured down'	'she saw'

Speakers can analyze the structure on the left in (15), exemplified by the example on the left, as the structure on the right (that is, in terms of sound segments instead of morphs) and apply this analysis to the example on the right. The difference is that in the example on the left, the incorporated element and the light verb (usually of the form -C-), are different morphs in the stem, whereas in the structure on the right the CV- and -C- are in a single morpheme.

One way of taking stock of why these structures in Udi are typologically uncommon is to examine why the same thing did not happen in its sister languages. Proto-North-East-Caucasian had gender agreement, but not person agreement. Udi and Tabasaran are the only two languages in the family that have (independently of one another) created complete new person agreement systems from pronouns, though some of the other languages have some more limited innovative person agreement. The agreement markers in the other languages are affixes, not clitics as in Udi. It appears in structures of some of the other languages that agreement affixes there have also been trapped, but because they are affixes, this same process in the other languages of the family has created infixes, not endoclitics. So it is the combination of the fact that Udi created new agreement marking from pronouns, the fact that these markers are clitics, and the fact that the language has undergone extensive univerbation that has led to its being unique in its family in having the structure in (13), illustrated by (5). Note that it is the retention of the structure in (12), illustrated by (4), together with certain other structures, that keep these PM clitics from becoming affixes.

Since the Romance languages have well known person-number clitics that some analyze as marking agreement, another way of taking stock of why these structures in Udi are typologically uncommon is to examine why the same thing did not happen in the Romance languages. The simple answer is that although the formation of complex verbs is quite common, it has not occurred in the recent history of the Romance languages, and thus there has been nothing to trap the clitics.

Although analogy is known to be a very common diachronic process, the application of it described above may seem unusual, but that is only because few languages have the structure on the left in (15). Without this key analogue, it is clear that this particular use of analogy cannot be applied.

We can summarize this discussion by listing the changes that led to the intermorphemic clitic in (13) and (5).

(16) Changes involved in the development of the intermorphemic clitic:

a. development of focus cleft

b. loss of copula

c. use of pronoun to introduce the embedded clause

d. loss of the inherited agreement system

e. development of person-number clitics out of independent pronouns

f. univerbation

g. maintenance of structures such as (12), which prevent the clitics from being reanalyzed as affixes.

Thus, it appears that a complex sequence of common changes is responsible for the development of this structure in Udi. While each change is common, the combination appears to be uncommon and does not occur elsewhere in the family.

But the fact that Udi underwent such a complex development does not prove that this is the only route to developing endoclitics. Part of explaining why endoclitics are typologically unusual involves examining whether there are other possible historical routes to this same structure. Probably there are. However, to maintain agreement markers as clitics entails, by most definitions of clitics, that the markers occur in some instances in some other position, as in (1–4) here, to prevent them from being reanalyzed as affixes. For example, when we compare endoclitics with infixes, we see that the occurrence of the former in other positions is the only characteristic that distinguishes them. The complex origin described above (and in more detail in Harris 2002) accounts for the occurrence of Udi clitics in these various positions, while a simpler history would not. In other languages, it is most likely that only innovative agreement markers would be clitics, for eventually clitics are usually reanalyzed as affixes. The only known source of endoclitics is entrapment in the course of univerbation or some similar process. It may be possible for another change to have the same outcome, but there is no reason to believe that it would be a simpler process than entrapment in Udi.[5] Thus it seems that at least (16e–g), or substitutes for them, are likely to be present most of the time, and other changes parallel to (16a–d) would most likely be required to set the stage for these, including getting the elements into the order required. Thus, while the changes summarized here are probably not the only possible route to the formation of endoclitics, it is unlikely that any other route would be significantly simpler.

If our innate endowment discourages endoclitics, the only evidence of this we have is that they are uncommon among languages of the world. As an explanation of their infrequency, this is circular. There is no specific evidence that this structure does not function well or is difficult to acquire or difficult to process, since these issues have not been researched. On the other hand, there is good reason to believe it is primarily

[5] Yu (2003) proposes four mechanisms for the creation of infixes, and one might assume that any one of these might in principle create endoclitics as well. His four are entrapment, metathesis, reduplication mutation, and prosodic stem association.

the complexity of the history of Udi clitics that has insured that they would occur in a variety of positions and in this way has prevented their being reanalyzed as infixes. The complex history thus explains the typological rarity of this structure; it also explains why endoclitics do occur, in spite of their rarity. Other accounts cannot accomplish this.

3. Other Unusual Structures

While infixes and circumfixes are not as unusual as endoclitics, they are less common than either prefixes or suffixes. On the approach taken here, the reason is clear. Existence of a prefix or suffix requires only the creation of that affix – one historical step. In contrast, an infix would seem to require two steps – creation of a prefix or suffix, together with some mechanism for getting that affix inside the word (see note 5); some of the processes described by Yu (2003), however, are considerably more complex than this. A circumfix in most instances requires three steps – creation of a prefix, creation of a suffix, and the linking of these two morphemes into one. This is probably not the only way in which a circumfix can be created, but it is likely that any route to formation of a circumfix will be more complex than formation of a simple prefix or suffix.

In the paper cited in note 1, I have shown that the infrequency of a very different kind of structure is likewise best explained in terms of the many changes required to create it. This is the case system in Georgian, where three different tense-aspect-mood characteristics of verbs are associated with three different case patterns for their arguments. Again, while there may be alternative routes to such systems, it is unlikely that the creation of a system with three distinct case-marking patterns will ever be simple.

I am by no means suggesting that the relative frequencies of all structures are determined by the complexity of the processes that create them. For example, we assume that the creation of prefixes and suffixes are parallel processes, yet suffixes are believed to be more common. Historical complexity cannot account for this and other facts. Yet it seems that in a number of instances, infrequent structures are infrequent simply because their creation requires more steps than that of more common structures.

References

Harris, A.C. (2000), *Where in the word is the Udi clitic?*, "Language" 76, 593–616.

Harris, A.C. (2002), *Endoclitics and the origins of Udi morphosyntax*, Oxford, Oxford University Press.

Harris, A.C. & L. Campbell (1995), *Historical syntax in cross-linguistic perspective*, Cambridge, Cambridge University Press.

Jeffers, R.J. & A.M. Zwicky (1980), *The evolution of clitics*, in E. Closs Traugott, R. LaBrum & S. Shephard (eds.), *Papers from the 4th International Conference on Historical Linguistics*, Amsterdam, John Benjamins, 221–232.

Kazenin, K.I. (1994), *Focus constructions in Daghestanian languages and the typology of focus constructions*, unpublished ms., Moscow State University.

Kazenin, K.I. (1995), *Focus constructions in North Caucasian languages*, unpublished ms., Moscow State University.

Kazenin, K.I. (2002), *Focus in Daghestanian and word order typology*, "Linguistic Typology" 6, 289–316.
Kiss, K.É. (ed.)(1995a), *Discourse configurational languages*, Oxford, Oxford University Press.
Klavans, J.L. (1979), *On clitics as words*, Chicago Linguistic Society Parasession volume, 68–80.
Lambrecht, K. & M. Polinsky (1997), *Typological variation in sentence-focus constructions*, in K. Singer, R. Eggert & G. Anderson (eds.), *CLS 33: The panels*, Chicago, Chicago Linguistic Society, 189–206.
Watkins, C. (1963), *Preliminaries to a historical and comparative analysis of the syntax of the Old Irish verb*, "Celtica" 6, 1–49.
Watkins, C. (1964), *Preliminaries to the reconstruction of Indo-European sentence structure*, in H.G. Lunt (ed.), *Proceedings of the ninth international congress of linguists*, The Hague, Mouton, 1035–45.
Yu, A.C. (2003), *The morphology and phonology of infixation*, Ph.D. Dissertation, University of California, Berkeley.