



University of
Massachusetts
Amherst

Where in the Word is the Udi Clitic?

Item Type	article;article
Authors	Harris, Alice Carmichael
DOI	https://doi.org/10.2307/417136
Download date	2024-08-12 03:47:16
Link to Item	https://hdl.handle.net/20.500.14394/32369

WHERE IN THE WORD IS THE UDI CLITIC?

ALICE C. HARRIS

Vanderbilt University

This article shows that endoclitics do exist in Udi, a language of the North East Caucasian family, and this fact poses a challenge to the lexicalist hypothesis. Clitics may be positioned between the morphemes of complex verb stems and immediately before the final segments of monomorphemic verb stems. The author argues, on the basis of accepted tests for wordhood, that complex verb stems are single words, not phrases. On the basis of criteria developed by Zwicky and Pullum (1983), it is argued that the clitics of Udi are true clitics. An analysis of the placement of clitics in various positions inside verb stems is proposed in optimality theory. The author shows that phonological phenomena do not provide an alternative basis for positioning these clitics and concludes that clitics in Udi are a counterexample to the lexical integrity hypothesis.*

1. INTRODUCTION. Some linguists have claimed that endoclitics, that is, clitics inside words, do not exist (e.g. Klavans 1979 and 1985). I will show that in Udi, a language of the Lezgian subgroup of the North East Caucasian family, certain clitics do occur inside words, as well as at their edges. I show that Udi has both intermorphemic clitics, occurring between morphemes, and intramorphemic clitics, ones that occur inside a monomorphemic verb stem.

A variety of frameworks adhere to the LEXICAL INTEGRITY HYPOTHESIS, and it has been widely discussed. Di Sciullo and Williams state it as in 1.

- (1) Words are ‘atomic’ at the level of phrasal syntax and phrasal semantics. The words have ‘features’, or properties, but these features have no structure, and the relation of these features to the internal composition of the word cannot be relevant in syntax—this is the thesis of the atomicity of words, or the lexical integrity hypothesis, or the strong lexicalist hypothesis. (1987:49)
- ‘[S]yntactic atomicity’, that is, the inability of syntactic rules to ‘analyze’ the contents of X^0 categories. (1987:47)

[W]ords are opaque to all sentence-level operations or descriptions. (1987:52)

The lexical integrity hypothesis has been interpreted in a number of ways, but its essence is that the structure of the word is not ‘visible’ to the syntax. This article shows that in Udi the internal structure of the word is, in fact, relevant to rules of syntax, the rules that position clitics. The internal structure of the word, even the phonological structure of the morpheme, must be visible to the rules that position clitics.¹

* This material is based upon work supported by the National Science Foundation under Grant No. SRB-9710085 and by the International Research and Exchanges Board under the ACLS-Academy of Sciences Exchange with the Soviet Union (1989). The research reported here could not have been completed without the patient help of my consultants, Luiza Nešumašvili, Dodo Misk’ališvili, Nana Agasišvili, Caco Čik’vaiže, Valya Broyani, Alvina Mat’ilian-Barxudarašvili, Simon K’isbabašvili, and Zina Silik’ašvili. Many other people facilitated my fieldwork in various ways and I am grateful also to them. I am grateful to Wolfgang Schulze for providing invaluable commentary on a related manuscript. Parts of this article have been presented at the Conference on Non-Slavic Languages of the Commonwealth of Independent States and the Balkan Republics held at the University of Chicago in 1997, at the Eastern States Conference on Linguistics held at Yale University in 1997, at the 1999 Linguistics Colloquium at the University of North Carolina at Chapel Hill, and at the University of Canterbury, where I was a Visiting Erskine Fellow. I am grateful to all of these audiences for helping me to understand the Udi data, especially to Stephen Anderson, Geraldine LeGenre, and Gerdt Webelhuth. Naturally all errors are my responsibility.

¹ Dobrin (1998) argues that in Arapeshan gender endings are based on phonological form and that ‘this situation can be expressed by recognizing direct reference to phonological form, in violation of lexicalism’

In order to establish that clitic pronouns (person markers, PMs) do occur inside words in Udi, I must demonstrate three things: (i) where these clitics occur under various conditions, (ii) that they are indeed clitics, and (iii) that the units inside of which they occur are indeed words. I will first introduce some general characteristics of Udi grammar that are necessary for interpreting the examples.

The description below is based on the Vartašen dialect, formerly spoken in the village of Vartašen in Azerbaijan and still spoken in the village of Okt'omber (Zinobiani) in the Republic of Georgia.² The Vartašen dialect is also represented in texts published by Schiefner, Dirr, and Jeiranišvili. Examples from texts have the advantage of demonstrating that the placement of the PM, together with other aspects of the sentence, is natural, although we normally get minimal pairs only through elicitation. Accordingly, I have cited examples from both types of sources, including a text I collected, 'Taral' (abbreviated T).

2. THE STRUCTURE OF UDI.

2.1. CASE MARKING AND WORD ORDER. Basic case marking in Udi is ergative; that is, subjects of transitive verbs are in the ergative case, while subjects of intransitives and direct objects are in the absolutive case, as illustrated in 2.

- (2) a. iš-en tängä peškäš-ne-b-e äyel-ax³
 man-ERG money.ABSL gift-3SG-DO-AORII child-DAT
 'The man gave money to the child.'
 b. äyel baf-ne-d-e kur-a-boš
 child.ABSL fall-3SG-LV-AORII hole-DAT-in
 'The child fell into a hole.'

Definite direct objects occur instead in the dative case, which moreover has two forms, one ending in -Vx, the other without the x (Pančviže 1940). The example in 3 is identical to that in 2a, except that the direct object is in the dative in 3, and the noun is understood as definite.

- (3) iš-en täng-in-a(x) peškäš-ne-b-e äyel-ax
 man-ERG money-OBL-DAT gift-3SG-DO-AORII child-DAT
 'The man gave the money to the child.'

Only in the direct object is definite formally distinguished from indefinite. A variety of other word orders are also possible, such as that in 4.

- (4) iš-en äyel-ax täng-in-a(x) peškäš-ne-b-e
 man-ERG child-DAT money-OBL-DAT gift-3SG-DO-AORII
 'The man gave the money to the child.'

Subjects of certain intransitives are in the ergative case.

(1998:61), concluding that 'morphosyntax cannot be categorically denied direct access to a word's phonological form' (1998:78).

² As far as I am aware, all aspects of the description also apply to the other dialect, the Niĵ dialect, except that there is a possibility that the occurrence of the PM is not obligatory in the Niĵ dialect.

³ The following abbreviations are used in glosses: ABL ablative, ABSL absolutive case, AOR aorist, CAUS causative, DAT dative, ERG ergative case, FUT future, GEN genitive, IMPER imperative, INF infinitive, INST instrumental, LV light verb, NEG negative, OBL oblique stem, PL plural, PRES present, PTCL particle, PTCLPL participle, Q 3rd person singular PM used in content questions, SG singular, SUBJV subjunctive. Udi does not indicate gender in pronouns or PMs; in the English translation, the feminine is used unless the context requires the masculine pronoun. In other contexts, the abbreviations FOC focus, FocC constituent in argument-focus, IncE incorporated element, PM person marker (clitic), and TAM tense-aspect-mood, and TOP topic, are used.

- (5) äyel-en ɸne-ne-xa⁴
 child-ERG cry-3SG-SAY.PRES
 'The child is crying.'

2.2. PERSON MARKERS. A basic property of Udi clitic pronouns (PMs) is that they encode the person and number of the subject. Three complete sets of pronouns are used; the major allomorphs of the most frequently used set are given in Table 1.⁵

	CLITIC PMS	INDEPENDENT PRONOUNS
1 SG	-zu, -z	zu
2 SG	-nu, -n, -ru, -lu	un
3 SG	-ne, -le, -re, -n	meno, kano, šeno
1 PL	-yan	yan
2 PL	-nan, -ran, -lan	van, efan
3 PL	-q'un	met'oɣon, kat'oɣon, šet'oɣon

TABLE 1. General person markers in Udi, compared with selected independent pronouns.

The allomorphy in the form of the PMs, shown in Table 1, is conditioned by the phonology and is entirely regular. For example, the third person singular *-ne* assimilates to an immediately preceding *l* or *r*, accounting for the occurrence of *-le* and *-re*. The independent pronouns of first and second persons are the ergative/absolute forms. The third person independent pronouns are (in order) proximate, distal, and remote deictics, which are also used as personal pronouns. Listed here in the third person singular are the nominative case forms, in the third person plural the ergative case forms; these provided the respective sources for diachronic formation of the third person PMs.

In content questions, the morpheme *-a* substitutes for a third person singular PM, as shown in 6; 6b can be compared with the corresponding statement in 12.

- (6) a. ma-q'un aš-b-esa?
 where-3PL work-DO-PRES
 'Where do they work?'
 b. ma-a aš-b-esa?
 where-Q work-DO-PRES
 'Where does she work?'
 c. *ma-ne aš-b-esa?
 where-3SG work-DO-PRES
 'Where does she work?'

Examples 6b and 6c show that in content questions the usual third person singular PM, *-ne*, does not occur, but *-a* occurs instead. Example 6a illustrates that this substitution does not take place in other person/number combinations. Apart from the fact that it occurs only in content questions and substitutes only for the third person singular, the distribution of *-a* is identical to that of the other clitics described here (Harris 1992).

2.3. THE STRUCTURE OF THE VERB IN UDI. The finite verb consists minimally of a verb stem and a tense-aspect-mood (TAM) suffix. For example, the verb form *uɣ-sa* consists just of the verb stem *uɣ-* 'drink' and the present tense marker, *-sa*; the verb form *box-e*

⁴ Sublinear dots indicate pharyngealization of vowels. Apostrophes indicate glottalized consonants. A colon marks a so-called intensive consonant.

⁵ All of the relevant facts regarding clitics in Udi are discussed in greater detail and in a diachronic context in Harris 1998; the other sets and their use is described in chap. 2 of that work.

consists of the verb stem *box-* 'boil' and the aorist II marker, *-e*.⁶ These are simplex verbs, with a verb stem consisting of a single morpheme. Most of the verbs in the language are complex.

Most complex verbs are based on one of a handful of light verbs, but other simplex verbs can also be used in complex verbs. Some light verbs occur also as independent verbs, and in examples I gloss them with their meanings as independent verbs, but in small capitals when they function as a light verb. For example, *b-* 'do, make' can occur as an independent simplex verb or as a light verb within a complex verb, as in *aš-ne-b-e* [work-3SG-DO-AORII] 'she worked'. Two light verbs, *Ø/(e)c-/γ-/k-* (multiply suppletive according to TAM) 'come, go' and *-č-/š-* 'carry', are used primarily with locative preverbs; even though they are not used independently, their meanings are clear, and they are glossed. Still others cannot occur as independent verbs and have no clear meaning, and I gloss them simply LV. The light verbs are listed in 7.

(7) Light verbs

b- 'do, make', *bak-* 'be, become', *Ø/(e)c-/γ-/k-* 'go, come', *-č-/š-* 'carry', *-d-/t'-* LV, *-k'-* LV, *(e)x-/p-/luk'-lup-* 'say'

Light verbs incorporate (a) nouns, (b) adjectives, (c) locative preverbs, (d) ancient simplex verb stems, (e) borrowed verbs, (f) infinitives (to form causatives), or (g) elements that are not used independently and cannot be identified. An example of each type of INCORPORATED ELEMENT (IncE), in order, is given in 8, and the type of element incorporated is identified to the right of the example. I gloss the unidentifiable element in 8g, *bas-*, with the meaning of the unit consisting of the incorporated element and the light verb, *bas-k'-* 'lie (down)'.⁷

- (8) a. . . . *pašcaγ-un γar-muγ-on lašk'o-q'un-b-esa* (D 62:34)⁷ noun
king-GEN boy-PL-ERG wedding-3PL-DO-PRES
'The king's sons marry [the girls they had rescued].'
- b. *äyel kala-ne-bak-e* adjective
child.ABSL big-3SG-BECOME-AORII
'The child grew (up).'
- c. *zu-al buzaq'sa ta-γ-a-z* locative preverb
I-AND want thither-GO-SUBJV-1SG
'I want to go too.'
- d. *nana-n tur-ex oc'-ne-k'-e* ancient simplex verb
mother-ERG foot-DAT wash-3SG-LV-AORII
'Mother washed her foot.'
- e. *sa adamar-en fikir-re-b-sa te . . .* (D 68:6) borrowed verb
one person-ERG think-3SG-DO-PRES that
'A person thinks that . . .'
- f. *nana-n äyel-ax ak'-es-ne-d-e k'uč'an* infinitive
mother-ERG child-DAT see-INF-3SG-CAUS-AORII puppy.ABSL
'The mother showed a puppy to the child.'
- g. *pašcaγ-un γar bas-ne-k'-esa* (D 61:23) unidentified
king-GEN boy.ABSL lie-3SG-LV-PRES
'The king's son lies down.'

I argue below (§5) that the complex verb is a single word.

⁶ The names used here for the various TAM categories are based on those in Pančviže 1974.

⁷ Examples identified as D are from Dirr 1928, DG are from Dirr 1904, J are from Jeiranišvili 1971; the first number indicates the page, and the second the line. Sf examples are from Schiefner 1863; here the first number is that of the text, the second that of the line. Examples not so identified are from my own fieldwork.

3. FOCUSED CONSTITUENTS. Like clitic pronouns in some other North East Caucasian languages, those in Udi also function as markers of focus (Harris 1996).⁸ In Udi, the PM is enclitic to constituents in focus (FocC) in argument-focus structures. (I follow the terminology and analysis of focus set forth in Lambrecht 1994.) Examples are given in 9, where the focused constituent is in capitals in the translation line. Questioned constituents, the answers they elicit, and negatives are obligatorily in focus.

- (9) a. me išq'ar-mux mano äiz-i-q'un karx-esa?
 this man-PL.ABSL which village-DAT-3PL live-PRES
 'WHICH VILLAGE do these men live in?'
 b. okt'omber-a-q'un karx-esa
 Oktomber-DAT-3PL live-PRES
 'They live in OKT'OMBER.'
 c. okt'omber-a te-q'un karx-esa
 Oktomber-DAT NEG-3PL live-PRES
 'They do NOT live in Okt'omber.'

Note that the PM *-q'un* 3PL is enclitic to the word in focus in 9b,c, and to the phrase in focus in 9a.

An example with argument-focus is associated with a presupposition containing a variable that corresponds to the constituent in focus. In some textual examples the presupposition is especially well supported by what precedes it. In 10, the presupposition made by the servant is *someone caused the misfortune*; in 10b *un* 'you' satisfies the variable in the presupposition. The focus structure is given in 10c; see Lambrecht 1994, ch. 5 on this way of presenting focus structure.

- (10) a. She cried and ran to the servant to tell of her misfortune and misery. The
 servant was so mean and angry, she said
 b. xatin-ax un-nu b-e, un-al duz-b-a (Sf xv:21)
 misfortune-DAT you-2SG do-AORII you-and right-DO-IMPER
 ''It was you who caused the misfortune, you make it right.''
 c. [xatin-ax [un-nu]FOC b-e]TOP

In general, the focused constituent immediately precedes the verb, as in the examples above. Since in other languages there are instances where a clitic is syntactically associated with one constituent, for example to its right, but is phonologically associated with another, on its left, one may wonder whether this is true in Udi. With the question-words that mean 'why', the focused constituent may be separated from the verb, as in 11.

- (11) a. ek'alu-n okt'omber-a are?
 why-2SG Okt'omber-DAT come.AORII
 'WHY did you come to Okt'omber?'
 b. ek'alu-a vi viči are?
 why-Q your brother.ABSL come.AORII
 'WHY did your brother come?'

The fact that the PM can be separated from the verb in examples such as 11 establishes that the PM is not syntactically associated with the verb, but with the focused constitu-

⁸ Somali and other East Cushitic languages have a similar system, in which an agreeing form (or personal pronoun) follows the focused constituent (Hetzron 1974, Svolacchia et al. 1995). In Somali and a number of the North East Caucasian languages other than Udi, the agreement marker is affixed (or cliticized) to a focus particle; in Udi there is no separate particle. Anderson and Zide (1999) describe a somewhat similar placement of a marker of subject agreement in Kherwarian and North Munda languages.

ent. Additional evidence related to this issue is discussed in §7.5 and in Harris 1998: ch. 3.

4. WHERE IN THE VERB IS THE UDI CLITIC? I follow Lambrecht 1994 in assuming that every sentence must have focus. In Udi, when there is no argument-focus, the PM is associated with the verb. The exact position of clitics, including the PMs, is determined by a complex set of conditions. For example, in certain TAM categories (the future II, subjunctive I, subjunctive II, and imperative), the PM is enclitic to the verb, as illustrated in 8c and 42a, b; this condition takes precedence over all others, including that requiring the marking of argument-focus. Only the conditions effecting endoclititicization are discussed further here; all conditions are discussed in detail in Harris 1998.

First, in most TAM categories (present, imperfect, aorist I, aorist II, perfect, particle conditional, future I, conditional I) PMs are endoclititic in a complex verb stem, occurring between the IncE and the light verb. The examples in 12–14 illustrate this.

- (12) zavod-a aš-ne-b-sa
 factory-DAT work-3SG-DO-PRES
 ‘She works in a factory.’
- (13) nana-n buɣa-ne-b-e p’ə ačik’alšey
 mother-ERG find-3SG-DO-AORII two toy
 ‘Mother found two toys.’
- (14) baba-n äyel-ax ači-p-es-ne-d-e nard
 father-ERG child-DAT game-SAY-INF-3SG-CAUS-AORII backgammon.ABSL
 ‘The father got the child to play backgammon.’

In 12–13 the PM, *-ne-*, occurs between the IncE, *aš-* or *buɣa-*, and the light verb, *b-* ‘do, make’. The infinitive in the causative construction in 14 is a special case of a nominal element incorporated into the verb. That is, in this example the IncE itself is complex; it consists of a complex verb (*ači-p-* ‘play’, where *-p-* is the light verb) plus *-es*, the formant of the infinitive. Thus, under circumstances that require endoclititicization, the PM occurs between the IncE and the (outermost) light verb in a complex verb.

In contrast, in a simplex (monomorphemic) verb stem the PM is endoclititicized immediately before the final segment of the verb stem. Examples of monomorphemic verb stems are given in 15.

- (15) *ayz-* ‘stand up’, *ač-* ‘get lost’, *ak-* ‘see’, *ap-* ‘ripen’, *aq-* ‘take, receive’, *bak* ‘be, become; be possible’, *bap-* ‘arrive’, *bašq-* ‘steal’, *bəɣ-* ‘look (at), watch’, *biq-* ‘catch; build’, *bos-* ‘throw away, throw’, *boš-* ‘sate’, *box-* ‘cook, boil’, *bəq-* ‘gather, harvest’, *buq-* ‘want, love’, *burq-* ‘begin’, *čalx-* ‘recognize, know’, *čuk-* ‘pick, etc.’, *duɣ-* ‘beat’, *ef-* ‘keep’, *lax-* ‘put’, *muč-* ‘kiss’, *sak-* ‘push’, *t’it-* ‘run’, *uk-* ‘eat’, *uk-* ‘say’ (future stem of suppletive *p-* ‘say’), *uɣ-* ‘drink’.

Examples 16–19 illustrate this position of the PM. In these and other sentences below, a discontinuous morpheme is glossed twice, with glosses that are identical except for subscript numbers.

- (16) paščay-un ɣar-en gölö bə-ne-ɣ-sa met’a-laxo (D 62:27)
 king-GEN boy-ERG much look₁-3SG-look₂-PRES this.GEN-on
 ‘The prince looks at this for a long time.’
- (17) kayuz-ax a-z-q’-e
 letter-DAT receive₁-1SG-receive₂-AORII
 ‘I received the letter.’

- (18) q'ačay-γ-on bez tänginax baš-q'un-q'-e
 thief-PL-ERG my money.DAT steal₁-3PL-steal₂-AORII
 'Thieves stole my money.'
- (19) ba-ne-k-sa sa pašč'aγ-k'ena adamar (J 169:3)
 be₁-3SG-be₂-PRES one king-like person.ABSL
 '[Once upon a time, there] is a person like a king.'

Thus, in most TAM categories, when there is no argument-focus, a PM occurs inside the single morpheme that comprises the verb stem of a simplex verb.

5. ARGUMENTS FOR THE WORD STATUS OF COMPLEX VERBS IN UDI. I argue that each complex verb is a single word: the stem of a complex verb consists of (a minimum of) two morphemes—a light verb and an incorporated element, either (a) a noun, (b) an adjective, (c) a locative preverb, (d) an ancient simplex verb stem (e) a borrowed verb, (f) an infinitive, or (g) an element that is not used independently and cannot be identified, as exemplified in 8 above. I am arguing against an analysis on which the light verb and the incorporated element are two independent words, comprising a phrase rather than a word. The latter analysis could provide an attractive explanation of why the PM occurs in a position that appears to be the middle of a word. This is not an argument against a syntactic treatment of incorporation (such as head-to-head movement), but only against an analysis that fails to recognize as a word the verb stem consisting of an incorporated element and the light verb that incorporates it.

5.1. LEXICAL LISTEDNESS, NONCOMPOSITIONALITY, AND NONOCCURRENCE AS INDEPENDENT WORDS. The meaning of many complex verbs is not the sum of the meanings of the parts. For example, *xam-p-* 'shave' seems to be composed of *xam* 'pasture, never plowed land' and *-p* 'say'. Related to this is the fact that in some instances the incorporated element does not occur independently, as in *furu-p-* 'stroll; hunt for'. In some instances, as mentioned above, the light verb does not occur independently, as in *čev-k'* 'send out, let go', *gam-d-* 'heat up, warm up'. Thus, the meaning of complex verbs in Udi is not straightforwardly compositional; rather, the meaning of many complex verbs is idiosyncratic. Listedness (or noncompositionality) is neither a necessary nor a sufficient condition to establish that an individual unit is a word, rather than a phrase (see Di Sciullo & Williams 1987), but the listedness of at least SOME such units is probably viewed by many as a necessary condition to establish the members of the class of such units as words. That is, many linguists would expect at least some complex verbs to be noncompositional in meaning.

5.2. INPUT TO DERIVATIONAL PROCESSES. Words or stems, not phrases, are the input to derivational processes; thus, from the phrase *do homework*, we do not get the derived form **homework-doing* (*child*). In Udi, complex verbs are the input to the full range of processes that create new words from verbs. Udi has an infinitive in *-es* and a masdar (deverbal noun) in *-(e)s-un*; complex verbs occur in both forms, as the examples in 20 demonstrate.⁹

⁹ All of the processes described in this section are derivational, according to the criterion of changing word class. For example, the infinitive and masdar of 20 are nominals derived from verb bases, and they decline as nominals, as shown in the paradigm below.

- (i) ABSL tämiz-b-es-un 'to clean, cleaning'
 ERG tämiz-b-es-un-en
 GEN tämiz-b-es-un-un
 DAT tämiz-b-es-un-a(x)
 ABL tämiz-b-es-un-axo

According to Jeiranišvili (1971:104–7), while the more complex masdar has a complete declension, that of

(20) INFINITIVE	MASDAR	
čixar-k'-es	čixar-k'-es-un	'to end, come to a close'
täyn-b-es	täyn-b-es-un	'to establish, install'
ta-š-es ¹⁰	ta-š-s-un	'to take, carry'

In the first example in 20, *k'* is one of the light verbs that does not occur independently; in the second example, *täyn-* is an IncE that does not occur independently; and in the third example, both morphemes of the verb stem fail to occur independently. These facts ensure that the outputs are not phrases; they must be words.

Similarly, complex verbs occur in both participial forms—the future participle in *-al(a)* and the past participle in *-i*.

(21) FUTURE PARTICIPLE		PAST PARTICIPLE
ta-š-al(a)	'to take, for taking'	ta-š-er-i 'taken'
šam-k'-al(a)	'to be killed, for killing'	šam-p-i 'killed'
boš-t'-al(a)	'to be planted/buried, for planting/burying'	boš-t'-i 'planted, buried'

Like the examples in 20, those in 21 contain at least one morpheme that cannot occur independently, thus establishing that these examples cannot be phrases.

Complex verbs occur in the full range of converbs (also called gerundials or absolutes), including those listed in 22.¹¹ These are more difficult to translate satisfactorily into English than are other categories.

(22) FINALIS CONVERB ¹²	
ta-š-s-an	'to take'
furu-p-s-an	'to hunt; to stroll'
ser-b-es-t'-es-an	'to have [something] fixed, to cause [something] to be fixed'
CONVERB OF SIMULTANEOUS ACTION	
bač'uk'-t'-a-xun	'while lighting, setting fire to'
č'e-bak-a-xun	'while passing, while going (across)'
furu-k'-a-xun	'while hunting; while strolling'
CONVERB OF POSTERIOR ACTION	
but'-k'-ama	'before covering, closing'
käy-bak-ama	'until it becomes dawn, until dawn breaks'
č'oč'a-bak-ama	'until it becomes red (i.e. until it browns, of food)'

In the last category of 22, I have only a single example, *but'-k'-ama* 'before covering, closing' that contains a morpheme, *k'*, that cannot occur independently; thus the last two examples in the last category are arguably phrases, because both the IncE and the light verb can occur independently. Nevertheless, all of the remaining examples in

the infinitive lacks the genitive case. (See also Pančviže 1974:194–95.) The participles in 21 are deverbal adjectives; like other adjectives in Udi, they do not decline, but they do modify nouns: *ašbal adamar* 'working man' (cf. *aš-b-* 'work'), *kar-x-i adamar* 'a man who has lived' (cf. *kar-x-* 'live') (examples from Pančviže 1974:201, 203). The converbs in 22 are derived adverbs, lacking the TAM suffixes and PMs that characterize verbs. Like the infinitive and the masdar, the deverbal nominals in 23 decline; see Pančviže (1974:120) for examples. The criterion of changing word class is widely accepted; it is cited in textbooks such as O'Grady et al. (1997:142) and Jannedy et al. (1994:135), and is the only criterion discussed by Spencer (1991:9).

¹⁰ The verb *ta-š-* 'take, carry' uses the light verb of transitive motion (i.e. meaning 'bring', 'take', 'take up', etc. depending on the incorporated preverb), which has the allomorphs *-š*, *-č*.

¹¹ A converb is a deverbal adverb most often used in a subordinate clause. In many instances the converb indicates simultaneous action, as in 'As I was **going through** the cemetery, I saw that . . .' (D 60:18, my translation), or 'While the bottom of the pilaf is **browning**, let's put butter in another pot—small' (fieldwork), where the words in bold translate converbs.

¹² I follow Schulze's (1982:187–89) analysis of converbs and approximately his terminology.

20–22 illustrate clearly the fact that complex verbs, ones that cannot be phrases, are the input to derivational processes.

Thus, all of the productive deverbal word formation rules apply to complex verbs; what is more, an unproductive process does so. Jeiranišvili (1971:29) and Pančviže (1974:120) describe a fossilized formant, *-un*,¹³ which forms a noun from a verb; these are idiosyncratic in meaning, but generally the noun refers to the result of the action expressed by the verb. Pančviže remarks that there are so few of these forms that one can count them on one's fingers. Among these fossilized forms are ones based on complex verbs, including those in 23, each of which contains at least one morpheme that occurs only as part of complex verbs.

- (23) č'e-p-un 'thrown out' (cf. č'epsun 'throw out')
 k'ač'-k'-un 'chewing gum' (cf. k'ač'pesun 'chew')
 kiš-p-un 'wood shaving, chip' (cf. kišpesun 'plane')

The fact that complex verbs such as those in 23 are among those preserved in a fossilized form confirms that complex verbs are single words.

5.3. NEGATION. In Udi the negative particle must immediately precede the verb, as shown in 24.¹⁴

- (24) a. merab-en ait-ax e-ne-f-sa
 Merab-ERG word-DAT keep₁-3SG-keep₂-PRES
 'Merab keeps his word.'
 b. merab-en ait-ax te-ne ef-sa
 Merab-ERG word-DAT NEG-3SG keep-PRES
 'Merab does not keep his word.'
 c. merab-en te-ne ef-sa ait-ax
 Merab-ERG NEG-3SG keep-PRES word-DAT
 'Merab does not keep his word.'
 d. *merab-en te-ne ait-ax ef-sa
 Merab-ERG NEG-3SG word-DAT keep-PRES
 'Merab does not keep his word.'
 e. *ait-ax te-ne merab-en ef-sa
 word-DAT NEG-3SG Merab-ERG keep-PRES
 'Merab does not keep his word.'

In 24b,c, the negative immediately precedes the verb; in 24d, e it does not. Examples 25b, c show that the negative immediately precedes the complex verb as well.

- (25) a. nana-n buɣa-ne-b-e p'a ačik'alšey
 mother-ERG find-3SG-DO-AORII two toy.ABSL
 'Mother found two toys.'
 b. nana-n te-ne buɣa-b-e p'a ačik'alšey
 mother-ERG NEG-3SG find-DO-AORII two toy.ABSL
 'Mother did not find two toys.'
 c. *nana-n buɣa te-ne b-e p'a ačik'alšey
 mother-ERG find NEG-3SG DO-AORII two toy.ABSL
 'Mother did not find two toys.'

¹³ Other morphemes with the same form (*-un*) include an allomorph of the genitive case marker, (related to this) a formant of denominal adjectives (Pančviže 1974:127), and part of the productive formant (*-(e)s-un*) of the complex *masdar*.

¹⁴ Under certain specified conditions it is possible for the negative to immediately follow the verb form (Harris 1998), but this does not provide a test and is thus not discussed further here.

Sentence 25c is ungrammatical because the IncE *bɥga-* is separated from its light verb *-b-* by the negative marker. This establishes that complex verbs such as *bɥga-b-* ‘find’ cannot be separated and supports the claim that they are single words.

5.4. ANAPHORIC ISLANDS. Complex verbs are anaphoric islands. In the forms in 26 and 27, the incorporated elements cannot be replaced with pronouns, as shown in the b examples. (The light verb *-ɣ-* in 26 is one stem of a suppletive verb ‘come, go’ whose directionality is defined by a locative preverb. The pronoun used in 27b is the distal deictic, which functions also as a personal pronoun. The form used is the absolutive, the case that is usual in complex verbs.)

- (26) a. *ta-z-ɣ-o*
thither-1SG-GO-FUT I
‘I will go’
b. **mya-z-ɣ-o*
here-1SG-GO-FUT I
‘I will come here’
(27) a. *baf-ne-d-e*
fall-3SG-LV-AOR II
‘s/he fell’
b. **šono-ne-d-e*
it-3SG-LV-AOR II

It is a general property of words that they are anaphoric islands, and therefore parts of words cannot be replaced by proforms. The ungrammaticality of 26b and 27b shows that the complex verbs *ta-ɣ-* ‘go (thither)’ and *baf-d-* ‘fall’ are single words, not phrases.

5.5. QUESTIONING. Sentence constituents can be questioned, but parts of words cannot.

- (28) a. Elise outwitted Russell.
b. **What did Elise out- Russell?*
c. **What did Elise -witted Russell?*
d. *?*Elise outwhatted Russell???*

In Udi it is impossible to question parts of complex verbs.

- (29) a. *manana-n kaɣəz k'al-le-p-e*
manana-ERG letter.ABSL read-3SG-SAY-AOR II
‘Manana read a letter.’
b. **manana-n kaɣəz ek'a-a/ne(-) p-e*
manana-ERG letter.ABSL what-Q/3SG say-AOR II
(30) a. *lay-ne-č-esa*
up-3SG-CARRY-PRES
‘She takes it up.’
b. **ma-a-č-esa / *ma-ne-č-esa*
where-Q-CARRY-PRES /where-3SG-CARRY-PRES
‘Where is she taking it?’

Attempts to question parts of the complex verbs in 29a and 30a are ungrammatical. This supports the view that these are words, not phrases.

5.6. CONJOINING. Words, phrases, and clauses can be conjoined, but parts of words cannot.¹⁵ Thus, in English we can conjoin the words in 31a, but not the parts of words in 31b.

¹⁵ Booij (1985), Bresnan and Mchombo (1995:188–89), and others show that in some languages conjoining parts of a word may be possible if the parts constitute phonological words separate from the stem.

- (31) a. Elise outran and outwitted Russell.
 b. *Elise out-ran and -witted Russell.

We cannot conjoin the word parts *-ran* and *-witted*.

In Udi, words, phrases, and clauses are most often conjoined without an overt conjunction, though it is possible to use the conjunctions *-al* or *-q'an* 'and'. In compounds, too, juxtaposition often indicates conjoining, e.g. *nana-baba* [mother-father] 'parents', *xunč'i-vič'i* [sister-brother] 'siblings'. Parts of complex verbs, however, cannot be conjoined.

The sentences in 32 represent conjoined clauses; 32b, c show that *lay-* 'up' and *bay-* 'in' cannot be conjoined, with or without an overt conjunction.

- (32) a. *lay-ne-č-esa, bay-ne-č-esa*
 up-3SG-CARRY-PRES in-3SG-CARRY-PRES
 'She takes it up, [and] she takes it in.'
 b. **lay-bay-ne-č-esa* (= **lay-bay-ne č-esa*)
 'She takes it up and in.'
 c. **lay-q'an-bay-ne-č-esa* (= **lay-q'an-bay-ne č-esa*)
 up-and-in-3SG-CARRY-PRES

But it is possible to conjoin (compound) finite verb forms based on complex verb stems; see Harris 1998. The impossibility of conjoining the locative preverbs *lay-* 'up' and *bay-* 'in' supports the view that these are parts of words, not themselves independent words.

5.7. GAPPING. In many languages, a verb can be gapped at its second occurrence in a sentence, but a part of a word can never be. In Udi, too, a verb can be gapped.

- (33) a. *manana-n kayəz k'al-le-p-e, dodo-n-an roman*
 manana-ERG letter.ABSL read-3SG-SAY-AORII Dodo-OBL-ERG novel.ABSL
k'al-le-p-e
 read-3SG-SAY-AORII
 'Manana read a letter, [and] Dodo read a novel.'
 b. *manana-n kayəz k'al-le-p-e, dodo-n-an roman*
 'Manana read a letter, [and] Dodo a novel.'

While a whole complex verb can be gapped, as in 33b, part of one cannot, as shown by 34b.

- (34) a. *manana-n kayəz cam-ne-p-e, dodo-n-an roman*
 manana-ERG letter.ABSL write-3SG-SAY-AORII Dodo-OBL-ERG novel.ABSL
k'al-le-p-e
 read-3SG-SAY-AORII
 'Manana wrote a letter, [and] Dodo read a novel.'
 b. **mana-n-an kayəz cam-ne-p-e, dodo-n-an roman k'al*

Even though the material gapped in 34b is morphologically identical (the *-ne-/le-* alternation is phonological) to part of what remains, the result is ungrammatical. This provides further evidence that complex verbs such as *cam-p-* 'write' and *k'al-p-* 'read' are each single words, not phrases.

5.8. CONCLUSION. Units composed of IncE and a light verb comprise a single word according to widely accepted criteria based on phenomena such as compositionality and absence of independent occurrence, acting as input to derivational rules, placement of negation, anaphoric islandhood, questioning, conjoining, and gapping. The significance of this is that clitic PMs occur between the morphemes of these words; they cannot be explained away as occurring between the words of a phrase.

6. ARE SIMPLEX VERBS MONOMORPHEMIC? Simplex verbs are here DEFINED as ones that are monomorphemic, but, given the unusual claim that a clitic is placed inside a morpheme in examples such as 16–19, the reader has every reason to wonder whether the verbs analyzed here as monomorphemic might really be complex.

As it turns out, there is indeed reason for suspicion, but only from a diachronic point of view. Jeiranišvili (1956) showed that a number of verbs in Udi begin with a *b-* that represents the fossilized marker of the neuter gender class. Most of the languages of the North East Caucasian family, including most of those of the Lezgian subgroup, to which Udi belongs, have gender-class agreement, and it is well accepted that this system of gender-class agreement must be reconstructed to Proto-North-East-Caucasian. In this system, intransitive verbs agree with their subjects, transitive verbs with their direct objects; each of these nominals is in the absolutive case. Gender-class agreement is indicated by a prefix on the verb indicating the class and number of the absolutive nominal. In contemporary Udi, verb stems that preserve this archaic marker are mostly of the form *bVC*; these may go back to two or even three morphemes historically.

Jeiranišvili lists the following verbs as preserving the class marker *b-*: *bak-* 'be, become; be possible', *bap-* 'arrive', *baq-* 'have, get; be (located)', *baq'* 'hold, contain', *basak-* 'put in, push in', *bask'-* 'lie down', *b-* 'make, do', *bes-* 'ask, request', *bəγ-* 'look at, watch', *bi-* 'die', *biq'-* 'catch; build', *bit-* 'fall', *bit'-* 'sow', *bok-* 'burn, grill', *bot'-* 'wound, cut', *bos-* 'throw away', *box-* 'boil', *bəq'-* 'gather, harvest', *bu-* 'be, have', *buq'-* 'want, love'. While some of these may have a history different from that posited by Jeiranišvili, there is little doubt that he is correct in the overall point he was making; that is, the class marker *b-* is preserved in word-initial position in a number of Udi verbs. Because this fossil is so widespread in the language, we cannot assume that any verb stem in Udi that begins with *b-* is historically monomorphemic.

Although some of the verbs in *b-* clearly originated as two or three morphemes, they are synchronically monomorphemic. There is no sense in which *box-* 'boil', for example, is composed of meaningful parts. In the meaning cited, *x* cannot occur without both *b* and *o*, nor do any of these parts alternate synchronically. From a synchronic point of view, *box-* and the other verbs in *b* are monomorphemic just as English *newt*, reanalyzed from *(a)n ewt ((a) eute)*, is synchronically monomorphemic.

Few diachronically simplex verbs remain in Udi; I find only about twenty verbs that are both diachronically and synchronically simplex. The verbs listed in 15 are synchronically monomorphemic. The importance of this is that PMs occur inside these morphemes. (A more complete treatment of the diachronic issues is contained in Harris 1998 especially chs. 8 and 9.)

7. PMs ARE CLITICS. My claim that in Udi clitics invade complex and simplex verb stems is not convincing unless I establish that the PMs that do so are indeed clitics. Zwicky and Pullum (1983:503–4) provide six criteria that have become classic for distinguishing between clitics and affixes.

A. Clitics can exhibit a low degree of selection with respect to their hosts, while affixes exhibit a high degree of selection with respect to their stems.

B. Arbitrary gaps in the set of combinations are more characteristic of affixed words than of clitic groups.

C. Morphophonological idiosyncrasies are more characteristic of affixed words than of clitic groups.

D. Semantic idiosyncrasies are more characteristic of affixed words than of clitic groups.

E. Syntactic rules can affect affixed words, but cannot affect clitic groups.

F. Clitics can attach to material already containing clitics, but affixes cannot.

Applying these criteria to the facts of Udi, I argue that PMs in Udi are indeed clitics.

7.1. CRITERION A. Zwicky and Pullum consider a morpheme that occurs with a word from any major form class to be less selective than one occurring only with words from a single form class. A morpheme that occurs with any word from a major form class is in turn less selective than a morpheme occurring only with some subclass of words from this form class. In Udi, PMs may attach to any word from any major form class—verbs (35), nouns (36), adverbs (37), adjectives (38), pronouns (39), or postpositions (40).

- (35) a. äzbär cam-zu-p-e
poem.ABSL write-1SG-SAY-AORII
'I wrote a poem.'
- b. čubux čomox t'oyol čurp-i imux-ne-lax-o [T]
woman.ABSL door.DAT near stand-PTCPL ear-3SG-put-FUT I
'The woman, standing near the door, heard.'
- (36) a. äyel-en p'a eš-ne aq'-e
child-ERG two apple.ABSL-3SG take-AORII
'The child took TWO APPLES.'
- b. sa jähil γar-re biq-sa xibq'o manat-al (D 61:1)
one young boy-3SG catch-PRES sixty ruble-on
'He hired A YOUNG BOY for sixty rubles.'
- (37) a. irähät-en mya-ne bist'a cil-l-ux
peasant-ERG here-3SG sow.PRES seed-OBL-DAT
'The peasant sows the seeds HERE.'
- b. išu qoš-ne qay-bak-e k'wa
man.ABSL back-3SG back-BECOME-AORII house.DAT
'The man returned BACK home.'
- (38) a. un šavat'-t'u yoxsan äybetjär?¹⁶
you.ABSL pretty-2SG or ugly
'Are you pretty or ugly?'
- b. beš mässäb š:el-le (D 69:18)
our religion.ABSL good-3SG
'Our religion is good.'
- (39) a. vax-ne ta-š-ša ič k'wa
you.DAT-3SG thither-CARRY-PRES self's house.DAT
'She takes YOU to her house.'
- b. zu š:u-zu (Sf xii:1)
I.ABSL who-1SG
'WHO I am' (title of a text)
- (40) a. xe-n-en-k'-ne tay-sa (Sf xv:4)
water-OBL-ERG-for-3SG thither-PRES¹⁷
'She went FOR WATER.'

¹⁶ Udi uses a zero copula commonly, and in this construction, the unmarked position of the PM is enclitic to the predicate nominal, as in 38; 39b illustrates a zero-copula structure with argument focus on the question word.

¹⁷ I have adopted Pančviže's analysis of this postposition here, namely that the postposition *-k'* governs the ergative case (here *-en*). Alternatively, *-enk'* may be analyzed as a postposition from a synchronic point of view. In either case, it is clear that a postposition is involved here, and that the PM *-ne* attaches to it.

- b. *hovuz-un-t'öyöl-le ar-e* (DG 85:7)
 pool-GEN-near-3SG come-AORII
 'She went NEAR THE POOL.'

The fact that a PM may attach to any word of any major word class establishes, according to Zwicky and Pullum's first criterion, that it is a clitic, rather than an affix.

7.2. CRITERION B. There is one particular host word, the prohibitive negative particle *nu/nut'/nut*, to which PMs cannot attach, and this makes it APPEAR that there is an arbitrary gap in the set of host-clitic combinations. The system behind this gap is briefly explained below.

Four TAM categories—the future II, subjunctive I, subjunctive II, and imperative—take *nu/nut'/nut* as a marker of negation. These same four TAM categories require that the PM be enclitic to the verb, and this rule takes precedence over all others that position the PM (see Harris 1998); that is, in these categories argument-focus cannot be marked by PM placement, and the PM cannot be endoclititic.

- (41) a. *irähät-en cil bi-ne-t'-e*
 farmer-ERG seed.ABSL SOW₁-3SG-SOW₂-AORII
 'The farmer sowed seed.'
- b. *irähät-en cil te-ne bit'-e epsen*
 farmer-ERG seed.ABSL NEG-3SG SOW-AORII this.year
 'The farmer did not sow seed this year.'
- c. **irähät-en cil nut'-t'-e bit'-e epsen*
 farmer-ERG seed.ABSL NEG-3SG SOW-AORII this.year
 'The farmer did not sow seed this year.'
- (42) a. **irähät-en cil bi-ne-t'-al*
 farmer-ERG seed.ABSL SOW₁-3SG-SOW₂-FUTII
 'The farmer will sow seed.'
- b. *irähät-en cil bit'-al-le*
 farmer-ERG seed.ABSL SOW-FUTII-3SG
 'The farmer will sow seed.'
- c. **irähät-en cil te-ne bit'-al epsen*
 farmer-ERG seed.ABSL NEG-3SG SOW-FUTII this.year
 'The farmer will not sow seed this year.'
- d. **irähät-en cil te bit'-al-le epsen*
 farmer-ERG seed.ABSL NEG SOW-FUTII-3SG this.year
 'The farmer will not sow seed this year.'
- e. *irähät-en cil nut' bit'-al-le epsen*
 farmer-ERG seed.ABSL NEG SOW-FUTII-3SG this.year
 'The farmer will not sow seed this year.'

The sentences in 41 provide an example of a TAM category that does not require encliticization of the PM; 41b and 41c show that this tense cooccurs with the general negator *te*, not the prohibitive *nu/nut'/nut*. In contrast, the sentences of 42 are in the future II, one of the TAM categories that always require encliticization of the PM: 42a shows that in this category the PM indeed cannot be endoclititic; in 42a the PM occupies the same position as in 41a, but in one is grammatical and the other ungrammatical. Example 42b illustrates the only grammatical position for the PM in the named TAM categories, namely enclitic to the verb. Examples 42c, d show that the general negator *te* cannot precede the future II, whether the PM is positioned as required in the future II or as in other tenses (cf. 41b). Example 42e shows why *nu/nut'/nut* never hosts the

PM: in the only TAM categories that govern this negative marker, the PM is required to be enclitic to the verb.

The question of concern here is whether there is an arbitrary rule or condition that prohibits the combination **nut'-t'e* [NEG-3SG], or whether the impossibility of this combination is due to more general conditions and rules (see Zwicky & Pullum 1983: 505 on the interpretation of this criterion). The nonoccurrence of **nut'-t'e* is due to the facts that

- (i) *nut'* and its allomorphs occur only with the four TAM categories named,
- (ii) those same TAM categories require the PM to be enclitic to the verb,
- (iii) the rule cited in (ii) takes precedence over all other rules that effect PM placement.

All three conditions are general, and the effect of the set of rules is general, not arbitrary. There is no rule that proscribes certain combinations, such as **nut'-t'e*; rather all combinations with *nut'* are impossible, including, for example, **nut'-zu*, **nut'-q'un*, **nut'-a*. The general set of rules (i–iii) do not refer to specific combinations or even to specific PMs. Thus, there are no arbitrary gaps in the set of expected combinations, and according to criterion B, Udi PMs are clitics.

7.3. CRITERION C. According to criterion C, in clitic groups, allomorphs are ‘distributed by general rules referring to phonological and morphological properties of the host’ (Zwicky & Pullum 1983:505). Affixes, in contrast, are characterized by unexpected phonological forms, irregularities, and suppletion. Zwicky and Pullum consider it a morphophonological idiosyncrasy when either (i) hosts are affected by the bound morphemes attached to them, or (ii) the phonological form of the bound morpheme cannot be predicted by general rules (1983:505). In the case of the Udi PMs, the host morpheme is never changed by the cliticization of the PM. Allomorphs of the PMs are the result of consonant assimilation, simplification of consonant clusters, and vowel elision. While some of these processes are optional, this fact cannot be interpreted as evidence that these are in any sense idiosyncrasies. There are no unexpected forms or irregularities.

Zwicky and Pullum, referring to the English clitics *'s* and *'ve*, write that ‘Hosts are unaffected by these clitics, and the clitics themselves have allomorphs distributed by general rules referring to phonological and morphological properties of the hosts’ (1983: 505). This statement is equally applicable to Udi PMs.

7.4. CRITERION D. This criterion also indicates that PMs in Udi are clitics. There are no instances where the combination of host + PM has a meaning that is not predictable from the sum of its parts.

7.5. CRITERION E. There are few phenomena that satisfactorily test criterion E in Udi. There are no movement rules of the type that occur in English—no subject-verb inversion, no passivization, though gapping does shed some light on the status of PMs. Example 43 appears to show that a PM is gapped with its host FocC, here *evaxt'* ‘when’.

- (43) a. *evaxt'-t'u čax-exa čur-ax, evaxt'-t'u tast'a*
 when-2SG milk-SAY.PRES COW-DAT when-2SG give.PRES¹⁸
k'ok'oc'-a k'ač'?
 chicken-DAT feed.ABSL
 ‘When do you milk the cow(s), and when do you give feed to the chicken(s)?’

¹⁸ The form *tast'a* is from *ta-d(-e)sa* ‘give’, where *ta-* is a locative preverb, *-d-* is a light verb, and *-esa* is the present tense suffix. In this environment, *d* regularly becomes *t'* and metathesizes with *s*.

- b. evaxt'-t'u čax-exa čur-ax da k'ok'oc'-a k'ač' tast'a?
 when-2SG milk-SAY.PRES COW-DAT and chicken-DAT food.ABSL give.PRES
 'When do you milk the cow(s) and give feed to the chicken(s)?'

In 43a no gapping has applied; in 43b 'when' has been gapped, and it appears that the PM has been gapped with it. Thus it appears that a syntactic rule has affected a clitic group. Other examples suggest otherwise.

- (44) a. eš-zu gir-b-e, merab-en ar-ne gir-b-e
 apple-1SG gather-DO-AORII merab-ERG pear-3SG gather-3SG-DO-AORII
 'I gathered APPLES, [and] Merab gathered PEARS.'
 b. *eš-zu gir-b-e, merab-en ar-ne
 apple-1SG gather-DO-AORII merab-ERG pear-3SG
 'I gathered APPLES, [and] Merab PEARS.'
 c. eš-zu gir-b-e, merab-en ar
 apple-1SG gather-DO-AORII merab-ERG pear
 'I gathered apples, [and] Merab pears.'

Example 44 shows that the PM is not simply gapped with its host FocC. In 44a there is no gapping, and in 44b the verb (*girbe*) alone is gapped, leaving the PM *-ne*. In 44c, both PM and verb are absent. In 45 neither the verb nor the FocC is gapped, but the subject and the PM appear to be.

- (45) me mal-ur-ux bütün döv-ürγ-on yaq'č'ebakal-γ-oxo-q'un
 this goods-ABSL all dev-PL-ERG traveller-PL-ABL-3PL
 fuq'-p-e, ič-urγ-ox-al bes-b-e [D 62:15]
 steal-SAY-AORII those-PL-DAT-AND kill-DO-AORII
 'The devs [mythical beings] had stolen all these goods FROM TRAVEL-
 LERS, and [they] had killed them too.'

Examples 44, 45 and numerous similar examples do not support an analysis under which the clitic group is gapped. Together with 43, they suggest instead that a constraint on the process that introduces the PM or on gapping itself prevents the occurrence of a PM in a clause in which the verb, the subject, or the FocC (if indeed there is one) has been gapped under identity with a constituent in the preceding clause. I conclude that contrary to the first impression given by (43), gapping does not affect the host + PM; thus, this criterion, too, shows PMs to be clitics, not affixes.

7.6. CRITERION F. According to criterion F, clitics can be added to material already containing clitics, but affixes cannot. This entails that if a PM follows an enclitic, the PM must also be a clitic. In Udi there are productive enclitics which PMs may follow, showing that the PMs are themselves clitics. In order to demonstrate this, I must first establish that these are also clitics. (Although there are at least four such clitics, I describe only one here.)

The enclitic *-y/-i* PAST is different from TAM suffixes in at least two respects. First, *-y/-i* can occur with zero copulas. Udi copula clauses ordinarily lack an overt copula, and this entails that TAM suffixes not occur. Examples of *-y/-i* PAST with a zero copula can be seen in 47b,c. Second, the irregular verb *bu-* does not take any TAM suffixes (e.g. **bu-(e)sa* 'is, has', **bu-e* 'was, had') but it does occur with *-y/-i*, as illustrated in 46.

- (46) sa kašib-un xib xinär-re bu-y (DG 88:1)
 one poor-GEN three girl-3SG be-PAST
 ‘A poor man had three daughters.’

This enclitic occurs at least with verbs, nouns, and adjectives, as illustrated in 47.

- (47) a. te p’ə xinär-ax ačes-b-a-ne-y (DG 88:6)
 that two girl-DAT lose-DO-SUBJV-3SG-PAST
 ‘That he might lose the two girls.’
 b. čubuxal gölö šel-uk’la adamar-re-y (J 169:24)
 woman very good person-3SG-PAST
 ‘The woman was a very good person.’
 c. haq’ullu-ne-y (D 61:7)
 clever-3SG-PAST
 ‘He was clever.’

The fact that the past marker *-y/i* occurs with words from a variety of form classes, as in 47, indicates that it is a clitic (criterion A). There are no unexplained gaps in the set of combinations (criterion B): With verbs, the *-y/i* past tense clitic can occur with any TAM suffix except those of the imperative and the hortative. I assume that the exception is due to the absence of a semantic reading for such combinations. There are no morphophonological idiosyncrasies (criterion C); for speakers this morpheme is realized as a glide when it follows a vowel, though some linguists write it as *-i* in all environments. The meanings of the TAM categories have not been closely studied (but see Pančviže 1974:155ff and Schulze 1982), but it appears to me that there are indeed semantic idiosyncrasies in the combinations (criterion D). For example, the combination of the subjunctive I (in *-a*) and *-y/i* PAST, which is termed the subjunctive II, is not a past subjunctive as might be expected. In spite of this apparent idiosyncrasy in the combination of *-y/i* with other TAM markers, this criterion alone does not seem to me to be an adequate reason to think that *-y/i* ‘PAST’ is a suffix, rather than an enclitic. I know of no evidence relevant to criterion E for *-y/i*, and evidence relevant to F is presented in 47 and 48–50. Thus, all diagnostics except D indicate that *-y/i* PAST is a clitic, and on this basis I conclude therefore that *-y/i* does indeed have this status.

The clitic *-y/i* may follow the PM, as in 47, but the sentences in 48 show that the PM can instead follow *-y/i* PAST in the subjunctive II.

- (48) ek’a-te γač’-k’-a-y-z, zap’-nu-k’-o! [D 62:4]
 what.ABSL-PTCL bundle-LV-SUBJV-PAST-1SG pull-2SG-SAY-FUT I
 ‘Whatever I may tie [on], you will pull it up!’
 (49) ägär un bezi bak-a-y-n [J 174:8]
 if you.SG mine be-SUBJV-PAST-2SG
 ‘If you will be mine.’
 (50) šu usin e-γ-a-y-n, ta-š-al-le
 who.ABSL fast hither-go-SUBJV-PAST-3SG thither-CARRY-FUTII-3SG
 ‘Who[ever] comes early will win.’

According to Zwicky and Pullum’s (1983) criterion F, the fact that the PMs in 48–50 follow *-y/i*, indicates that the PM, too, is a clitic.

7.7. SUMMARY. The results of the tests provided by Zwicky and Pullum’s six criteria are summarized in Table 2.

CRITERION	RESULTS TO STATUS OF PMs
A (degree of selection)	clitics
B (gaps in the set of combinations)	clitics
C (morphophonological idiosyncrasies)	clitics
D (semantic idiosyncrasies)	clitics
E (effect of syntactic rules)	clitics
F (order of clitics and affixes)	clitics

TABLE 2. Results of the application of criteria A–F to Udi PMs.

It is clear on the basis of criteria A–F that PMs in Udi are clitics.^{19, 20}

8. ANALYSIS. Udi is unusual both in that clitics occur inside verb stems and in that PMs occur in different positions under various circumstances. Not every theory will find these facts easy to account for, and the goal of this section is to show how some of the unusual facts of Udi might be accounted for in OPTIMALITY THEORY (OT). The two specific problems addressed here are (i) how the framework of OT might be adapted to place the Udi PM between the IncE and the light verb when that is appropriate, yet inside the verb stem, when that is appropriate (and to avoid incorrect placement), and (ii) how to place the PM in the correct position in the simplex verb stem.

McCarthy and Prince (1993) propose a constraint family of the general form Align (Cat1, Edge1, Cat2, Edge2), where Cat1 and Cat2 are phonological or morphological categories, and Edge1 and 2 may be left or right. McCarthy and Prince discuss, *inter alia*, the application of this constraint to problems of infixation, but they make no suggestion about using it to account for endoclitisis. Nevertheless, it is this constraint family that is the key to the OT analysis proposed here.²¹

As we have seen above (e.g. 12–14), the PM occurs between the IncE and the light verb; this is accounted for here by means of the constraint Align-PM-IncE.

(51) Align-PM-IncE

Align (PM, L, IncE, R)

The constraint in 51 places the left edge of the PM at the right edge of the IncE, so that the PM follows the IncE like beads on a string. IncE is a morphological category, which is explicitly allowed by McCarthy and Prince for Cat2 (or Cat1). The names I propose for constraints are not of the form used by most other authors. The need for the third element, here IncE, will become clear below in 52; the second element, PM,

¹⁹ One might also claim to find endoclititics in European Portuguese, but the morphemes at issue have a number of Zwicky and Pullum's (1983) affix characteristics, including a high degree of selection with respect to their stems (criterion A), morphophonological idiosyncrasies (C), semantic idiosyncrasies (D), affect by syntactic rules (E) (Crysmann 1997).

²⁰ As can be seen from examples in this paper, PMs in Udi are grammatical agreement markers in the sense of Bresnan & Mchombo 1987; the PMs do not represent arguments in the sense of Jelinek 1984. Similar processes in other languages are often referred to as 'clitic doubling', but Suñer 1988 has argued (of Spanish dialects) that this should be considered agreement. Everett (1996), however, has suggested that clitics, agreement, and pronouns are all epiphenomena, and he has defined clitics and agreement such that they differ only in that the former are adjuncts to their hosts (not *m*-subcategorized by them), while agreement is included in its host (*m*-subcategorized by it) (1996:45–46). I find this aspect of Everett's argument convincing and, since Udi PMs are not *m*-subcategorized by their hosts, as can be seen from the examples in §3, I conclude that on this basis as well they must be considered clitics.

²¹ Even though Udi PMs occur in a position that resembles second position in the word, the OT approaches developed in Anderson 1997 and Roberts 1997 to account for second position clitics in Pashto do not work well in Udi.

is needed because Udi must have additional constraints to account for other clitics not discussed in this paper (see Harris 1998).

Candidates	Align-PM-IncE
a. +ne-ta -d-e	!
b. [☞] ta +ne-d-e	
c. ta -d+ne-e	!

TABLEAU 1. Endocclisis in a simplex verb stem, {ta-d-, -e, ne}.

Tableau 1 illustrates the application of this constraint alone with *ta-ne-d-e* ‘she works’, where *ta-d-* ‘give’ is a complex verb stem consisting of two bound morphemes, *-e* is the formant of the aorist, and *ne* is the third person singular PM. A vertical line indicates the right edge of the IncE, and a plus sign the left edge of the PM. In Tableau 1 and subsequent tableaux, the input is stated in the title line as the unordered set {Verb stem, Af₁, Af₂, . . . , Cl₁, Cl₂, . . .}. Candidates a and c both fatally violate Align-PM-IncE, leaving candidate b.

We have also seen (16–19) that the PM occurs inside the simplex verb stem, and I propose to account for this by means of the constraint Align-PM-Verbstem.

(52) Align-PM-Verbstem

Align (PM, R, Verbstem, R)

Note that this constraint, unlike Align-PM-IncE, does not align the right edge of one morpheme with the left edge of another, like beads on a string. Constraint 52 is partially parallel to constraint 53, proposed to account for infixation in Tagalog (McCarthy & Prince 1993:80).

(53) Align ([um]_{Af}, L, Stem, L)

Thus 52 explicitly draws a parallel between infixation and endocclisis. Note however the important difference between 53 and 52, aside from the fact that 53 refers to an affix, while 52 refers to a clitic: the Udi PM, since it occurs in some instances outside the verb stem, cannot be part of it, while the Tagalog affix in 53 is part of the input verb stem. Constraint 52 assumes that two segments cannot occupy the same moment in time (or space) and that this formulation thus forces the PM to the left or right of the segment whose right edge defines the right edge of the verb stem (see also McCarthy & Prince 1993:89–90).

Candidates	Align-PM-Verbstem
a. ne+bak -e	bak !
b. b-ne+ak -e	ak !
c. [☞] ba-ne+k -e	k
d. bak -ne+e	ne !

TABLEAU 2. Endocclisis in a simplex verb stem, {bak-, -e, ne}.

Tableau 2 illustrates the application of 52 alone in the simplex verb stem *bak-* ‘become, be’; + marks the right edge of the PM, and | the right edge of the verb stem.

All candidates violate Align-PM-Verbstem, and their violations can be measured in terms of the number of segments preventing the right edge of the PM from coinciding with the right edge of the verb stem. For example, in candidate a, three segments, *b*, *a*, *k*, come between the PM *ne* and the right edge of the verb stem. McCarthy and Prince (1993) define a notion of minimal violation. In candidate c, for example, the violation of Align-PM-Verbstem is minimal, a single segment; it is therefore the optimal candidate. Note that in candidate b, it is two segments of the verb stem that prevent the right edge of the PM from coinciding with the right edge of the verb stem, while in candidate d it is the two segments of the PM itself that do so.

An alternative approach might be proposed, namely a constraint of the form in 54 to replace 52.

(54) Align (PM, R, C]_{Vst}, L)

Constraint 54 aligns the right edge of the PM with the left edge of the consonant that is final in the verb stem. The constraint refers to a consonant, rather than to a segment in general, because the two verb stems in the language that end in vowels are exceptions with respect to PM placement. I have avoided the formulation in 54 because C]_{Vst} is neither a grammatical nor a prosodic category, the two types of categories permitted in alignment constraints in McCarthy & Prince 1993.

Align-PM-IncE may appear to be superfluous, since Align-PM-Verbstem alone could obtain the appropriate results in Tableaux 1 and 2. Tableau 3 shows why both constraints are necessary and why the former is higher ranked than the latter.

Candidates	Align-PM-IncE	Align-PM-Verbstem
a. +ne-qay -bak-e	neqay !	qaybak
b. qay +ne-bak-e		bak
c. qay -b+ne-ak-e	b !	ak
d. qay -ba+ne-k-e	ba !	k
e. qay -bak+ne-e	bak !	ne

TABLEAU 3. Endoclesis in complex stem with longer incorporating verb, {qay-bak-, -e, ne}.

(55) Align-PM-IncE >> Align-PM-Verbstem

Tableau 3 contains a longer incorporating verb, *qay-bak-* 'return'. In this tableau, + again marks the left edge of the PM, and | the right edge of the IncE. In Tableau 3, candidates a, c, d, and e all violate Align-PM-IncE, and all candidates violate Align-PM-Verbstem. Candidate b is optimal because it does not violate the higher ranked constraint, Align-PM-IncE. This tableau shows that Align-PM-Verbstem is outranked by Align-PM-IncE; if the reverse were true, the ungrammatical candidate d would be selected because of its minimal violation of Align-PM-Verbstem. The contrast between *ba-ne-k-e* 'she became' of Tableau 2, with an intramorphemic PM, and *qay-ne-bak-e* 'she returned' of Tableau 3, with an intermorphemic PM is accounted for here by the fact that Align-PM-IncE cannot apply in the former, where there is no IncE.

9. A PHONOLOGICAL APPROACH? The analysis proposed above accounts for the Udi data, but it does so in a way that raises important questions. One of the strengths of OT has been the fact that it accounts for infixation through relative ranking of phonological

constraints (see for example, McCarthy & Prince 1993), and recent analyses of clitics also propose some sort of phonological analysis (e.g. Roberts 1997). Some authors (for example, Booij 1985, Bresnan & Mchombo 1995) have suggested that the lexical integrity principle holds only with respect to morphological structure, not phonological structure. If the process that places PMs in Udi is a phonological one, it would not be a counterexample to lexical integrity on this approach. We must, therefore, explore the possibility of an analysis based on phonological phenomena.

The position into which the PM is placed cannot be characterized by counting syllables. From the left, the PM may be placed after the first syllable, as in 56a, or after the second, as in 56b.

- (56) a. aš.**ne**.be b. xa.bar.**re**.a.q'e (also xa.ba.**re**.a.q'e)
 aš-ne-b-e xabar-re-aq'-e
 work-3SG-DO-AORII news-3SG-take-AORII
 'she worked' 'she asked'

Example 56 also shows that the PM is placed before the final syllable in some instances (56a), and before the penultimate in others (56b). In still other instances, the PM forms parts of more than one syllable, as in 57a, or is not a complete syllable, as in 57b,c,d.

- (57) a. xa.bar.**q'u.na**.q'e
 xabar-q'un-aq'-e
 news-3SG-take-AORII
 'they asked'
 b. aš.**neb**.sa c. a.čiz.pe d. be.**ney**.sa
 aš-ne-b-sa ači-z-p-e be-ne-γ-sa
 work-3SG-DO-PRES play-1SG-SAY-AORII look₁-3SG-look₂-PRES
 'she works' 'I played' 'she looks'

Consider the hypothesis that PMs are placed, at least in part, to maximize CV structure. This appears to be true of 56a,b and 57a,b. But consider (57c,d): **a.či.pe* would preserve the integrity of the morphemes and maximize open syllables in 57c, as would the form **a.či.zu.pe*, in which syncope has not reduced the PM. In 18, this hypothesis would lead us to expect **baš.(q')q'u.ne*, which would preserve the integrity of the morphemes, instead of *baš.q'un.q'e*, which does not. Thus, the position of the PM is not straightforwardly predicted from these phonological phenomena.

In many examples the PM in Udi immediately follows a stressed syllable, as illustrated in 58–59, which contain a question word and another focused constituent, respectively.

- (58) un eká-n maslahát-b-esá? (D 67:8)
 you.ERG what-2SG advise-DO-PRES
 'WHAT do you advise?'
 (59) γar-í ci Arzumán-ne bak-é (D 70:2)
 boy-GEN name.ABSL Arzuman-3SG be-AORII
 'The boy's name was ARZUMAN.'

We may therefore hypothesize that the PM always immediately follows word stress, but the examples in 60–61 disprove this hypothesis. In these examples, a syllable of the stem (60, 61a) or a syllable constituting a suffix (61b) intervenes between the stress and the PM.

- (60) ekáluγ-nu myá ar-é (D 62:7)
 why-2SG here come-AORII
 'WHY have you come here?'

- (61) a. šét'in-al hamétar-re b-esá (D 60:30)
 he.ERG-AND this.way-3SG do-PRES
 'He, too, behaves THIS WAY.'
 b. ič-ux-ne γač'-p-i (D 62:16)
 self-DAT-3SG tie-SAY-AORI
 'He tied HIMSELF on.'

A further problem faced by any approach that would place the clitic on the basis of stress is that assignment of stress is not consistent in the Udi verb. If there is no argument focus, stress may fall either on the last syllable of the verb stem, or on the TAM suffix. For example, in the present tense of *beγ-* 'look, see' with a third person singular subject, stress may fall on the simplex verb stem (D 62:27) or on the present tense suffix, *-(e)sa* (D 60:7); in the aorist I of *ta-d-* 'give', stress is attested on the (last syllable of the) IncE (D 62:35) or on the aorist I suffix, *-i* (D 60:19); in the subjunctive I, first person singular subject of *ta-š-* 'take', stress may be on the first syllable (D 61:26) or on the last (DG 90:1); in the subjunctive II of *bak-* 'happen', stress may fall on the first syllable (D 61:21) or on the last (DG 93:9). Thus stress itself may occur in either of two locations in a single verb form, though more commonly stress falls on the last vowel of the stem. Although the PM most often immediately follows the stressed syllable, it does not invariably do so.²²

I conclude both from these facts and from those in 60–61 that the position of the Udi PM is not predictable on the basis of stress alone. Further, I conclude that we cannot predict the position of the PM on any phonological basis. For this reason it cannot be argued that Udi PMs are not counterexamples to the lexical integrity hypothesis; they cannot be correctly positioned on the basis of phonological phenomena, but can be properly placed on the basis of morphological conditions described in previous sections.

10. CONCLUSION. I have demonstrated here that claims that endoclitics do not exist (Klavans 1979, 1985, and other works) are untrue, and conclude that endoclitics must after all be recognized as a possibility in natural languages.

The lexical integrity hypothesis states, in various formulations, that rules of syntax do not have access to the morphological structure of a word. We have seen that PMs are enclitic to focused phrases (see §3, especially ex. 9a) and that they must be positioned on a variety of syntactic constituents; I conclude from this that the rules that position PMs are syntactic rules. Since PMs occur between morphemes (in complex verb stems) and inside morphemes (in simplex verb stems), the rules that position them there must have access to the morphological structure of those verbs. Hence, syntactic rules—those that position the PM—must have access to the internal structure of verbs. Phonological phenomena cannot account for the position of PMs, and they are therefore real counterexamples to the lexical integrity hypothesis. The notion of lexical integrity cannot be maintained in the forms quoted in 1 or similar formulations; even morphological integrity, though maintained in most languages, is not a principle that applies in all languages.

REFERENCES

- ANDERSON, GREGORY D. S., and NORMAN H. ZIDE. 1999. Recent advances in the reconstruction of the Proto-Munda (Austroasiatic) verb. Paper presented at the International Conference on Historical Linguistics 14, Vancouver.
 ANDERSON, STEPHEN R. 1997. Towards an optimal account of second position phenomena.

²² Harris 1998 proposes a diachronic analysis which takes changes in stress into account.

- Yale a-morphous linguistics essays, ed. by Lizanne Kaiser, 1–27. New Haven: Department of Linguistics, Yale University.
- BOOIJ, GEERT. 1985. Coordination reduction in complex words: A case for prosodic phonology. *Advances in nonlinear phonology*, ed. by Harry van der Hulst and Norval Smith, 143–60. Dordrecht: Foris.
- BRESNAN, JOAN, and SAM A. MCHOMBO. 1987. Topic, pronoun, and agreement in Chicheŵa. *Language* 63.741–82.
- , ———. 1995. The lexical integrity principle: Evidence from Bantu. *Natural Language and Linguistic Theory* 13.181–254.
- CRYSMANN, BERTHOLD. 1997. Cliticization in European Portuguese using parallel morphosyntactic constraints. *Proceedings of the LFG '97 Conference*, ed. by Miriam Butt and Tracy Holloway King. CSLI Publications. [<http://csli-publications.stanford.edu/hand/miscpubsonline.html>]
- DIRR, ADOLPH. 1904. *Grammatika udinskogo jazyka*. (Sbornik materialov dlja opisanija mestnostej i plemen Kavkaza 33.) Tbilisi.
- . 1928. *Udische Texte*. *Caucasica* 5.60–72.
- DISCIULLO, ANNA-MARIA, and EDWIN WILLIAMS. 1987. *On the definition of word*. Cambridge, MA: MIT Press.
- DOBRIN, LISE M. 1998. The morphosyntactic reality of phonological form. *Yearbook of Morphology* 1997.
- EVERETT, DANIEL L. 1996. *Why there are no clitics: An alternative perspective on pronominal allomorphy*. Arlington, TX: Summer Institute of Linguistics.
- HARRIS, ALICE C. 1992. The particle *-a* in Udi. *The non-Slavic languages of the USSR: Linguistic studies*, ed. by Howard I. Aronson, 135–56. Chicago: Chicago Linguistic Society.
- . 1996. Focus in Udi. *NSL.8: Linguistic studies in the non-Slavic languages of the Commonwealth of Independent States and the Baltic Republics*, ed. by Howard I. Aronson, 201–20. Chicago: Chicago Linguistic Society.
- . 1998. *Endoclitics and the origins of Udi morphosyntax*. Nashville: Vanderbilt University, ms.
- HETZRON, ROBERT. 1974. An archaism in the Cushitic verbal conjugation. *Quarto Congresso internazionale di studi etiopici*, vol. 2, Sezione linguistica, 275–81. Roma: Accademia nazionale dei Lincei.
- JANNEDY, STEFANIE; ROBERT POLETO; and TRACEY L. WELDON. 1994. *Language files: Materials for an introduction to language and linguistics*. 6th edn. Columbus: Ohio State University Press.
- JEIRANIŠVILI, EVG. 1956. *Gramat'ik'uli k'lasis gakvavebuli nišnebi udur zmnebsa da nazmnar saxelebši* [Fossilized markers of grammatical class in Udi verbs and deverbal nouns]. *Iberul-k'avk'aziuri enatmecniereba* 8.341–62.
- . 1971. *Udiuri ena* [The Udi language]. Tbilisi: Universit'et'i.
- JELINEK, ELOISE. 1984. Empty categories, case and configurationality. *Natural Language and Linguistic Theory* 1.39–76.
- KLAVANS, JUDITH L. 1979. On clitics as words. *Chicago Linguistic Society* 15(2).68–80.
- . 1985. The independence of syntax and phonology in cliticization. *Language* 61.95–120.
- LAMBRECHT, KNUD. 1994. *Information structure and sentence form: Topic, focus and the mental representation of discourse referents*. Cambridge: Cambridge University Press.
- MCCARTHY, JOHN J., and ALAN S. PRINCE. 1993. Generalized alignment. *Yearbook of morphology* 1993.79–153.
- O'GRADY, WILLIAM; MICHAEL DOBROVOLSKY; and MARK ARONOFF. 1997. *Contemporary linguistics: An introduction*. 3rd edn. New York: St. Martin's.
- PANČVIŽE, VL. 1940. *Ak'uzat'ivis genezisisatvis udur enaši* [On the origin of the accusative in the Udi language]. *Enimk'is Moambe* 5–6.137–152.
- . 1974. *Uduri enis gramat'ik'uli analizi* [A grammatical analysis of the Udi language]. Tbilisi: Mecniereba.
- ROBERTS, TAYLOR. 1997. The optimal second position in Pashto. New Brunswick, NJ: Rutgers Optimality Archive, paper #ROA-174-0297 [<http://rucss.rutgers.edu/roa.html>].
- SCHIEFNER, A. 1863. *Versuch über die Sprache der Uden*. *Mémoires de l'Académie impériale*

- des sciences de St.-Pétersbourg, VII série, tome vi, no. 8. St. Petersburg: Kaiserliche Akademie der Wissenschaften.
- SCHULZE, WOLFGANG. 1982. *Die Sprache der Uden in Nord-Azerbajdžan*. Wiesbaden: Harrassowitz.
- SPENCER, ANDREW. 1991. *Morphological theory: An introduction to word structure in generative grammar*. Oxford: Blackwell.
- SUÑER, MARGARITA. 1988. The role of agreement in clitic-doubled constructions. *Natural Language and Linguistic Theory* 6.391–434.
- SVOLACCHIA, MARCO; LUNELLA MEREU; and ANNARITA PUGLIELLI. 1995. Aspects of discourse configurationality in Somali. *Discourse configurational languages*, ed. by Katalin É. Kiss, 65–98. Oxford: Oxford University Press.
- ZWICKY, ARNOLD M., and GEOFFREY K. PULLUM. 1983. Cliticization vs. inflection: English *n't*. *Language* 59.502–13.

Linguistics Program
Vanderbilt University
Nashville, TN 37235
[alice.c.harris@vanderbilt.edu]

[Received 3 May 1999;
accepted 22 October 1999]