
The Problem and the Approach

At its core, this is a book about what a word is. Many linguists have assumed the correctness of the Lexical Integrity Hypothesis, the hypothesis that words are composed according to morphological principles that differ in kind from the syntactic principles responsible for the composition of sentences. According to this hypothesis, the morphological composition of a word is not accessible to the rules of syntax, and items manipulated by syntactic rules (i.e. words) cannot be manipulated by the morphology (see, for example Di Sciullo and Williams 1987, Bresnan and Mchombo 1995).

The facts of Udi challenge these basic assumptions. In this language, person markers (PMs) may occur between the bound morphemes of a verb, as for example the third person singular PM *-ne-* in *aš-ne-b-sa* [work-3SG-DO-PRES] ‘she works’, which may not be so surprising, but also inside a monomorphemic verb, as in *e-ne-f-sa* ‘she keeps (it)’, where *-ne-* splits the unanalyzable morpheme *ef-* ‘keep’. The problem is that the rules that position *-ne-* and other PMs must be in part syntactic rules, given that PMs may occur on words outside the verb, as in (1). (In (1a) *-ne* is attached to the noun *yaq* ‘road’ used adverbially, while in (1b) it is attached to the noun *ait*, the direct object of the sentence.)

- (1) a. *yaq*’-a-*ne* *ba-st*’a [T 9]
 road-DAT-3SG in-LV-PRES
 ‘On the road he opens it.’
- b. *merab-en* *ayt-ne* *ef-sa*
 Merab-ERG word.ABSL-3SG keep-PRES
 ‘Merab keeps his word.’

But if the rules are syntactic, the Lexical Integrity Hypothesis claims that they do not have access to the internal structure of a word and therefore cannot position the PM inside the verb, as in the examples *aš-ne-b-sa* ‘she works’ and *e-ne-f-sa* ‘she keeps (it)’, cited above. Put differently, in Udi we find endoclitics both intermorphemically (between the morphemes that compose a complex verb word) and intramorphemically (directly inside a single morpheme). Udi PMs have many properties traditionally associated with affixes, such as marking agreement¹ between subject and verb. Yet unlike most familiar affixes, their position in

¹ The notion that these morphemes mark ‘agreement’ is adopted from more traditional grammars uncritically until §5.5, where the notion is discussed. Use of the term is not intended to imply (a) that

the verb is anything but fixed. If linguists are to understand the limits on what constitutes a word, we must come to grips with the data from Udi.

1.1. Problems to Be Addressed and Goals

The fundamental problem that motivated this research is the placement of PMs in Udi. Although it has long been recognized that the PM can occur in various positions, it has not been possible until now to predict where the PM will occur in a given clause. The PM may be word-final in the verb, as in *bošt'-al-je* [plant-FUT-3SG] 'she plants it', or it may occur between morphemes in the verb, as in *box-ne-sa* [boil-3SG-PRES] 'it boils'. In some instances the PM may occur inside the verb root, as in *a-ne-q'-sa* 'she takes, receives', where the morpheme *aq'* is split by the PM *-ne-* [3SG]. In other examples the PM attaches to a word other than the verb, as in (1) and (2).

- (2) *nana-n* *äyel-ax-ne* *oc'k'alde*
 mother-ERG child-DAT-3SG bathed
 'The mother bathed the child.'

Note that the fact that linguists have never specified the rules for positioning PMs means that it is impossible to teach someone how to form even one grammatical sentence of the language. The fact that there are at least eight published full-scale grammars or grammatical sketches of the language (Schiefner 1863, Dirr 1904, Bouda 1939, Jeiranišvili 1971, Gukasjan 1974, Pančviže 1974, Schulze 1982, and Schulze-Fürhoff 1994b), as well as numerous scholarly articles about it, none of which has stated the position of the PM precisely, indicates how elusive the goal of specifying the position of the PM has proven to be. Thus my first goal is to describe the position in which PMs occur under specific grammatical conditions; this is partly accomplished in Chapter 3, where it is shown that PMs are hosted by focused constituents. The description of the positions in which PMs occur is completed in Chapter 6.

The theoretical context in which the position of Udi clitics becomes important is the claim by Klavans (1979, 1985) that endoclitics, clitics that occur inside words, cannot exist. In fact, almost all linguists believe that it is impossible for clitics to occur inside roots. Although it has been suggested that there are endoclitics in European Portuguese, Crysmann (1997) has shown that the morphemes at issue are not clitics according to the criteria established by Zwicky and Pullum (1983). It has also been claimed that endoclitics exist in Pashto, but Kaisse (1981, 1985: 132–43)

these morphemes 'belong to' the verb or to imply (b) any particular analysis of the status of these morphemes as clitics vs. affixes. A position on the first issue, what word hosts the agreement markers, follows as a consequence of the discussion in Chapter 3; no assumption is made independently of that analysis. The second issue is discussed in Chapter 5. The assumption that these markers are examples of agreement, rather than constituting arguments, is consistent with the fact that they cooccur with coreferential independent full noun-phrase or pronominal arguments.

showed that the clitics at issue always follow a morpheme; they do not occur inside a morpheme. (See further discussion in Chapter 6.)

A related goal of this book is to describe the diachronic origins of each of the positions of PMs, explaining how the use of that position came to be part of the grammar; this is completed in Chapter 12, and most of the chapters leading up to it contribute to the same goal. In §8.3 it is shown that Proto-Lezgian, the subgroup proto-language from which Udi descended, had gender-class agreement; and later in that chapter we see that the PMs of Udi are a relatively recent innovation. The newness of these markers makes it especially puzzling that they would occur in the interior positions described in the preceding paragraph. In Chapter 9 I consider the hypothesis that PMs were ‘trapped’ in an interior position when two independent words were reanalyzed as a complex verb, as well as the hypothesis that the position of the innovative PMs represents an ancient position inside verbs that can be reconstructed to PL or even to Proto-North East Caucasian (PNEC). It is especially problematic to find morphemes that are clearly innovations inside a word, since there is good reason to believe that morphemes of this type, if they move at all diachronically, move to the periphery of the word (Haspelmath 1993a).

Part of the explanation of the origins of PM positions lies in the reconstruction of the focus cleft. Some of the other languages of the North East Caucasian (NEC) family have productive focus clefts or monoclausal focus constructions derived from them. Dargi retains both, as illustrated in (3).

- (3) 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 FOC[AUX-2SG] brothers.ABSL bring-PTCPL.SG
 ‘You brought the brothers.’
 c. x'o-ni sabi uzbi arku-si [Kazenin 1994, 1995]
 2SG-ERG FOC[AUX] brothers bring-PTCPL.SG
 ‘You brought the brothers.’

(3a) is a simple sentence with no focus marked. In (3b, c), focus is placed on ‘you’. In Chapter 10 I argue that (3b) is a cleft, with the first two words, *x'o saj-ri* ‘you are’ forming the matrix clause, and that (3c) is monoclausal.² The sentence is thus comparable to English *It was you who brought the brothers*. In Chapter 10 focus clefts are tentatively reconstructed for Pre-Udi.

² If I understand Kazenin (1995) correctly, he argues that sentences comparable to (3b) are not clefts because the copula determines the case of the focused constituent (absolutive). He seems to claim that this is not characteristic of clefts. However, that would be incorrect; in unreanalyzed clefts, the focused constituent is subject of the copula, and this grammatical relation determines its case. Being uncertain of whether Kazenin views this Dargi construction as a cleft, I have provided my own analysis.

Additional goals of this book include presentation of material relevant to other current theoretical debates in synchronic morphosyntax. First, in Chapter 5 it is argued that PMs in Udi are clitics, not affixes, according to criteria set out in Zwicky and Pullum (1983) and in other works. Klavans (1979 and 1985) claims that endoclitics (clitics that divide a lexical item) cannot exist, but Jeiranišvili (1971: 83) and Pančviže (1974: 148–50) claim that Udi PMs do divide simplex verbs in examples such as those in (4), where the stems *buq'*, *čuk*, and *aq'* are considered unanalyzable single morphemes.

- (4) a. *bu-za-q'-sa*
 love₁-1SG-love₂-PRES
 'I love him/her'
- b. *a-ne-q'-sa*
 take₁-3SG-take₂-PRES
 'She takes it'
- c. *ču-ne-k-sa*
 break₁-3SG-break₂-PRES
 'She breaks it'

Synchronically the elements glossed as a verb with the subscripts $\langle_1 \rangle$ and $\langle_2 \rangle$ are parts of a single morpheme, made discontinuous by the imposition of the PM inside it. However, it is possible that (4a) could be explained historically as consisting of a fossilized gender-class marker (*b(u)-*) with a root *q'*. In a number of verbs in Udi, *b-* occurs as a fossil of the agreement system inherited from Proto-NEC and Proto-Lezgian, in which a prefix marked the grammatical gender-class of the absolutive nominal (subject of an intransitive or direct object) (Jeiranišvili 1956). Generally a vowel occurs between the gender-class marker (CM) and the verb root under circumstances that are not yet fully understood. Thus *b-* and *u-* in (4a) were in all likelihood historically a morpheme (or two morphemes) distinct from (the ancestor of) the root *-q'* 'love, want' (see also Nikolayev and Starostin 1994: 645, Schulze 1988: 202). A related explanation may apply to (4b); it could consist of the (different) historical root *q'* 'take', together with a (different) vowel belonging historically to the ancient gender agreement system. However, there appears to be no comparable explanation for (4c) and a handful of similar forms. It appears that neither *ču-* nor *-k* can be related historically to other morphemes. Thus, the following issues are addressed. Are PMs in Udi clitics or affixes? Do *čuk* 'break' and similar stems consist of a single morpheme (an unanalyzable base), or does a closer inspection reveal that they are etymologically two (or more) morphemes? If this is a single morpheme, why do PMs split it, making it discontinuous? Aspects of the structure of the verb relevant to this issue are presented in Chapter 4, and the position of the PM inside the verbstem³ is described in Chapter 6.

³ I use 'verbstem' here as a technical term in distinction to '(verb) stem'. While 'stem' is used in different ways by different authors, in relation to a verb it most often means a base (usually more

A second synchronic issue dealt with is the theoretical account of the positions of clitics in Udi. In Chapter 7 an account in the framework of Optimality Theory is proposed, and it is shown that this approach can account elegantly for the complex set of requirements and options for placement of the Udi PM.

Although not taken up as a theoretical issue, the structure of complex verbs also figures in an important way in this work (described primarily in Chapter 4, and here and there in Chapter 6). Udi provides evidence relative to causatives, potentials, and incorporation of nouns and other elements. It distinguishes true complex verbs from lexicalized phrases, which are so similar that they require the use of specific diagnostics.

The final major goal is presentation of data and discussion of issues relevant to a theory of diachronic morphosyntax. Harris and Campbell (1995, chapter 7) make specific predictions relative to the simplification of biclausal structures. In Chapter 10 of the present work it is suggested that Udi clefts were simplified as monoclausal focus constructions. Comparative data from other NEC languages were first presented by Kazenin (1994), when Harris and Campbell (1995) was already in press. Thus the NEC data could not be taken into account in that work, and they provide data relevant to the hypotheses formulated there.

Second, changes in case-marking patterns result in part from changes in the structure of complex verbs. Udi inherited the ergative-absolutive case marking characteristic of NEC languages, with subjects of transitive verbs in the ergative case, and direct objects and subjects of intransitives in the absolutive case. Udi retains that system to an extent, but it has introduced two changes not shared by its sister languages. One of these differences is that in Udi, while direct objects may be marked with the absolutive, according to the inherited system, they may alternatively be marked with the dative, according to an innovative system. This is illustrated in (5) and discussed at greater length in Chapter 11, where it is shown that the dative case marking of the direct object has developed the function of marking definiteness. (In the remainder of this chapter, Udi verbs are segmented and glossed only as relevant to the point at hand.)

- (5) a. *xinār-en šum banest'a*
 girl-ERG bread.ABSL bake
 'The girl is baking bread.'
 b. *xinār-en šum-ax banest'a*
 girl-ERG bread.DAT bake
 'The girl is baking the bread.'

Dative direct objects may precede or follow the verb, and they appear to have the same privileges of occurrence that absolutive direct objects have (but see

complex than a root) to which additional affixes are added. I also refer to a 'present stem', 'aorist stem' etc. in this sense (see especially §2.4). The verbstem contrasts with this and is described more precisely in Chapter 4.

further below). How did they arise diachronically in Udi? In Chapter 11 I show that they developed as a ‘second direct object’, illustrated in (6).

- (6) *me čuby-ox axri q'onay-q'un-b-esa* [J 170]
 this woman-DAT finally guest-3PL-make-PRES
 ‘Finally they make a guest of this woman.’ ‘. . . receive this woman as a guest.’

As shown there, *q'onay* ‘guest’ was historically the direct object of *b-* ‘make, do’, and *me čuby-ox* ‘this woman-DAT’ was an oblique complement expressed in the dative (roughly ‘make a guest of this woman’). When *q'onay-b-* was reanalyzed as a unitary verb, it was natural to interpret as its direct object the nominal represented in this example by *me čuby-ox* ‘this woman-DAT’. The use of dative marking for direct objects was then extended, beside absolutely marked direct objects. Therefore, variable use of dative or absolutive for direct objects in other constructions is in part explained diachronically through the ‘second direct object’ construction, though we will see that contact is likely to have played a role as well.

A second change Udi has introduced in the inherited case marking system involves the use of the historical ergative case with verbs that are, loosely speaking, active intransitives (unergatives). This change may have been influenced by neighboring Georgian, which also uses the historical ergative case with verbs of a similar semantic type in certain tenses; however, it is possible to identify an internal mechanism of change, whether or not there was Georgian influence. It is natural to hypothesize that this phenomenon is related to incorporation, since many active intransitive verbs in Udi are of the incorporated type illustrated in (7), using the same simplex verbs used to form transitive verbs, while most inactive intransitives use the intransitive formants *-bak* ‘be, become’ or *-c* (polysemous), as in (8).

- (7) *äyel-en qone-ne-xa*⁴
 child-ERG cry-3PL-SAY.PRES
 ‘The child is crying.’
- (8) *lek'er q'ari-ne-bak-e*
 dishes.ABSL dry-3SG-become-AORII
 ‘The dishes dried.’

The structure of (7) was historically something like ‘the child says a crying’ and it was natural for the clause to be treated as transitive, and thus for the subject to be marked with the ergative case. From this sort of construction developed ergative case marking with a restricted set of intransitive verbs.

The third issue in the theory of historical syntax to be taken up here is the use of Internal Reconstruction and the Comparative Method in syntax. In Chapters

⁴ The verb ‘say’ is easily identified synchronically, and its meaning in complex verbstems should not be taken literally. This is discussed in detail in §9.3.1.1.

8–12, the use of these methods in the reconstruction of the syntax of the North East Caucasian family is illustrated, according to principles discussed in Harris and Campbell (1995). In §8.2 cognate sentences, a phenomenon disputed by Lightfoot (1979: 8), are used in comparative reconstruction. In §§8.4–5, the pronominal origins of the PMs are identified, as well as the origin of one special PM, *-a*. In Chapter 9 I reconstruct the structure of the PL verb, and in Chapter 10 the Pre-Udi focus cleft.

Lastly, my goals also include relating these diachronic phenomena in NEC languages to the universals of language change; they thus include further an explanation of how endoclitics (word-internal clitics) originate and an explanation of how an ergative-absolutive case system could develop intransitive verbs with ergative case subjects.

Udi was selected as the basis for a study in diachronic morphology and syntax because (a) the structures of the Daghestanian languages are different in many ways from those of more familiar languages and therefore can contribute significantly to our understanding of language universals, (b) Udi has a complex morphology, which supplies ready evidence of syntactic relations, and (c) Udi has undergone significant morphosyntactic change.

1.2. What is Udi?

Udi is a language of the Lezgian subgroup of the North East Caucasian language family. It is characterized by a number of interesting features, including some that play little or no role in this book. For example, its phonetic inventory contains glottalized consonants, [p', t', k', q', c', č'], and pharyngealized vowels, [i, ɛ, a, ɔ, ʊ] (see Fähnrich 1975).

Typologically, the language is agglutinative. In the verb, tense-aspect-mood (TAM) is marked by a set of suffixes, while person and number are marked by the clitics that are the main topic of this work. (Person and number are not marked by separate morphemes, as they are in the most highly agglutinative languages.) In *aq'-al-zu* 'I will take, receive', the future is marked by the suffix *-al* and the first person singular subject by the enclitic *-zu*. In the noun, case and number are marked by distinct suffixes: *äyel-uy-on* 'child-PL-ERG'.

Like other languages in its family, Udi has so-called dual-base noun declension. This term indicates a declension in which (some) nouns have one base for the absolutive and ergative cases and a different base, derived from the ergative, for other cases. Table 1.1 provides an illustration of the singular. Dual-base declension is somewhat different for certain other nouns.

Some Udi verbs, including verbs of motion, are characterized by directional preverbs. Only a few verbs have a monomorphemic base, such as *aq'* 'take, receive'. The majority of verbs in the language are composed of an incorporated element and a light verb, e.g. *aš-b-* 'work' (literally, 'work-DO'). The structure of the verb is treated in detail in Chapter 4.

TABLE 1.1. *Dual-base declension in Udi*
(*Jeiranišvili* 1971: 59 and *Sixarulize* 1987: 43)

Nominative	mex 'sickle'
Ergative	mex-en
Genitive	mex-n-ay
Dative	mex-n-u(x)

The inherited case marking is ergative, with subjects of transitive verbs in the ergative case, and with subjects of intransitive verbs and direct objects in the absolutive case. However, definite direct objects are marked with the dative case, which is otherwise used for indirect objects and certain locative functions. These constructions are illustrated and described in detail in §2.2. Subjects of certain verbs that appear to be intransitive may be in the ergative case; the historical explanation of this is discussed in Chapter 11.

Like other languages of the North East Caucasian family, Udi makes extensive use of non-finite verb forms, including participles, infinitives, masdars, and converbs. A frequent construction with a participle is illustrated in (10).

- (10) pačay-un čubux tac-i ayz-un boš bay-ne-sa [T 2]
king-GEN woman.ABSL go-PTCPL village-GEN in in-3SG-PRES
'The king's wife, having gone, enters the village.'

In this construction, the subject of the first verb, the participle, is also subject of the second; the participle is also used in other constructions.

Word order in Udi is generally SOV; this is illustrated and described in §2.7.

1.3. A Brief Overview

1.3.1. *Guide for the Reader*

The reader interested only in synchronic theoretical issues, not their historical explanation, needs to read Chapters 2–7. Chapter 2 provides background on basic points of Udi grammar relevant to the issues discussed here, including the use of the several sets of PMs. In Chapter 3 it is shown that the position of the PM is involved in marking focus. It is the combined effect of Chapters 4–6 that establishes that Udi possesses endoclitics. In Chapter 4 it is argued that complex verbs are single words; were they not, in some instances the PM would occur between words, rather than between morphemes of a single word. In Chapter 5 it is shown that PMs are clitics; were they not, we would have examples of infixes, rather than of endoclitics. Chapter 6 establishes informal rules for the placement of PMs and shows that they may occur between the morphemes of complex verbs and be internal to the roots of simplex verbs. In Chapter 7 I argue that a hierarchical set of violable rules provides a simple account of the placement of PMs,

while other theoretical approaches cannot provide an elegant account of these facts in Udi.

The reader interested only in diachronic theoretical issues must nevertheless acquaint himself with their description, which can be found in Chapters 2–6. Issues in diachronic morphosyntax are discussed in Chapters 8–12. §8.2 provides a new illustration of the use of the Comparative Method in syntax, while §§8.3–4 exemplify use of both the Comparative Method and Internal Reconstruction in morphosyntax. In Chapter 10 I discuss the origin of the monoclausal cleft, described in Harris and Campbell (1995, chapter 7). Chapter 9 discusses the problem of how the innovative PM could move inside existing verbstems, and Chapter 12 provides a summary discussion of how the various PM orders came about historically. Issues in the diachrony of case marking are discussed in Chapter 11.

Naturally I hope that most readers will be interested in both synchronic and diachronic aspects of this book.

1.3.2. Sources of Data

The problems with the use of texts are well known. If an example does not occur, we do not know whether its absence is due to its being ungrammatical or whether it represents an accidental gap. For this reason, and to obtain minimal pairs, I have relied primarily on elicitation for my analysis. I have elicited data (a) by asking for translations from Georgian (using bilingual consultants), (b) by presenting Udi sentences for grammaticality judgement, (c) by asking consultants to make up a sentence containing a particular form, and through other miscellaneous means. In this language, translation cannot be used effectively to determine word order and certain other phenomena, while Udi sentences I have formulated may present other problems.

On the other hand, texts provide advantages over elicited data in some respects. First, in texts we find constructions that we would not know existed if we relied on elicitation alone. The best example of this in Udi is compound verbs (§5.1.5), which could not have been elicited. Second, under most circumstances we find more natural examples in texts. For both reasons, I started my research with close study of texts and have elicited additional texts from consultants.

Selection of examples to cite is a problem distinct from analysis. In this work I have cited elicited examples when no textual example was available and when I needed pairs that differed minimally. On the other hand I have cited text examples to show the naturalness of a construction and so that my data can be verified. The latter is important in a work on Udi, while it would not be in one on English, German, or Italian. In order to verify the data in the field, the linguist would have to travel to the former Soviet Union and have a good knowledge of Udi, Russian, Georgian, or Azerbaijani. Text examples, on the other hand, can be verified with only a trip to the library and knowledge of German (for Bouda 1939, Dirr 1928, and Schiefner 1863) or Russian (for Bežanov 1888, 1902, Dirr 1904). I

have prepared some brief texts for publication on the web, with glosses and translation in English to make verifiability that much easier.

1.4. History of Research⁵

1.4.1. Position of Person Markers

Although a considerable amount of linguistic research has been done on Udi (e.g. Dirr 1904, Gamq'relize 1983, 1984, Gukasjan 1974, Jeiranišvili 1971, Pančviže 1974, Schiefner 1863, Schulze 1982), PMs have never been fully described. Concerning the location of the PM, Schiefner (1863: 25), the first to describe the language, writes only

The person marking occurs either after the verb or added between the two elements of the complex verb; third, however, the personal pronoun that belongs to the verb can occur enclitic to a preceding word. [My translation]

Dirr (1904: 49), in his description of the location of PMs ('pronominal infixes'), seems to be the first to identify the secondary function of the position of the PM as indicating what he refers to as 'logical stress':

These pronominal infixes are usually located between the root and other elements of the verb, . . . but very often they separate from the verb and adjoin to other words of the sentence, mainly to interrogative, negative, or prohibitive particles, and in general to words on which the logical stress falls. [My translation]

Elsewhere he notes that they may also be final in the verb form (Dirr 1904: 39).

Jeiranišvili (1971: 83–4) provides a more extensive discussion, distinguishing the position of PMs in the Vartašen dialect from that in the Niĵ dialect. Regarding the former he writes in part:

The position in which the case-changing pronominal person markers are used is as follows: 1) in static verbs it is always final, for example: *bu-ne* (is) [for glosses, see below] . . . 2) in the root of dynamic (transitive and intransitive) verbs it is used in various places; for example . . . *ba-ne-ksa* 'is, happens', *č'e-ne-baksa* 'go across, convert', . . . *baksa-ne* 'is, happens'. . . . As we see, the person marker is never used in initial position in any verb: When used as a prefix it must nevertheless follow some other independent (morphemic) element. . . . It is interesting that grammatical meaning is changed by a change in the position of the person marker in particular groups of verbs . . . ; compare for example *box-ne-sa* 'boils (INTR)'—*bo-ne-xsa* 'boils (TRANS)'. . .

As shown in the adduced examples, in composite verbs [complex verbs] the person marker is used either finally, as a suffix (more precisely, as a postpositional element): *aš-besa-ne* 'works' . . . ; or it turns up between the initial (nominal, adverbial) component and the root-form of the auxiliary verb that follows it; for example: *aš-ne-bsa* 'works', and thus in all other verbs of complex structure. [My translation]

⁵ For a general history of research on Udi, see Schulze (1982: 8–11).

(Jeiranišvili does not provide morpheme-by-morpheme glossing; a more complete glossing of the structure of the verbs cited is as follows:

<i>ba-ne-k-sa</i>	<i>bak</i> ‘become’, <i>-ne-</i> 3SG, <i>-sa</i> PRES	‘it is, becomes, happens’
<i>č’e-ne-bak-sa</i>	<i>č’e-</i> ‘down’	‘it goes across, converts’
<i>box-ne-sa</i>	<i>box-</i> ‘boil’	‘it boils’
<i>aš-b-esa-ne</i>	<i>aš-</i> ‘work’, <i>-b-</i> ‘do’, <i>-(e)sa</i> PRES	‘she works’.

Pančviže (1974: 148–50) observes that the PM is ordinarily what he calls an infix; that is, it is both preceded and followed by other parts of the verb complex. However, with certain verb forms it may be the [final] suffix; his paradigm of verb forms (Pančviže 1974: 155) helps to make this statement more specific, for the forms listed there show the PM as the final or penultimate suffix only in the subjunctive I, subjunctive II, imperative, and future II (eg. *kar-x-al-zu* ‘I will live’, where *-zu* marks the first person singular subject). He points out that the PM often precedes the verbal base, as in *te-z-cam-p-i* ‘I did not write’, literally ‘not-1SG-writing-SAY’, where the verbal base is *cam-p-* ‘write, writing-SAY’. He adds that sometimes the PM is entirely outside the verb, as in (11), where it is attached to the subject *sa q’oĵa kaft’ar-* ‘an old woman’.

- (11) *sa q’oĵa kaft’ar-re pašč’ayun-t’oĵoĵ e-sa*⁶
 one old woman-3SG king-to come-PRES
 ‘An old woman comes to the king.’
 (Bežanov 1888: 8, 7; Niĵ dialect; cited by Pančviže 1974: 149)

Pančviže (1974: 149) suggests that we would have expected instead *sa q’oĵa kaft’ar pašč’ayun-t’oĵoĵ e-ne-sa*, with the PM between the two elements of the verb base, as it ordinarily is.

Schulze (1982: 168–9) formalizes the discussion as a list of environments in which the agreement markers (which he calls PZ, my PM) may occur in relation to the verb (V), tense marker (T), nominal incorporated into the verb (Vn), and auxiliary (HV). Table 1.2 provides Schulze’s formulae and his example of each.

TABLE 1.2. *Patterns of morpheme order in Udi (from Schulze 1982: 168–9)*

1. V-T-PZ	p-i-ne	‘he said’
2. V-T ¹ -PZ-T ²	p-e-ne-i	‘he said’ ⁷
3. V ¹ -PZ-V ² -T	ba-ne-k-i	‘he was’
4. Vn-PZ-(H)V-T	aš-ne-biq’-e	‘he took a job’
5. Vn-(H)V-T-PZ	laš:q’o-bak-a-nan	‘you should marry’
6. . . . X-PZ-V-T	ġar-re-bak-i	‘he was a boy’

⁶ The word for ‘king’ varies by speaker; Schiefner (1863) records *pađšax*, Dirr (1904) *pačšax* or *pačšay*, Dirr (1928) *p’asčax*, Jeiranišvili (1971) *pašč’ay*, and the consultant who wrote ‘Taral’ *pačay*.

⁷ Schulze (1982) does not translate this form at all; in Schulze-Fürhoff (1994b: 476) he glosses it as a perfect, ‘he has said’. In my view, the TAM forms have not been studied thoroughly enough for us to distinguish adequately among the meanings of the several pasts.

While these sources describe a wide variety of locations available for PMs, they provide little in the way of statements to predict where PMs will occur under various circumstances. That is, these sources do not state *when* the PM will occur in *which* position. Schulze-Fürhoff (1994b: 476), on the other hand, states

There is no evidence of functional criteria for the use of the single slots. Sentence intonation may play a role (when the verb is final, the structure VERB-TM-PAM [i.e. VERB-STEM-TAM-PM] is preferred).

In Chapters 3 and 6 I establish the criteria that determine the position of the PM, showing that for many speakers these are not preferences, but absolute rules, while for some speakers variation is permitted in two out of seven rules. Verb-final order plays no role in this (see examples above), and in Chapter 6 I show also that intonation does not determine the position of the PM.

1.4.2. *The Structure of Complex Verbs*

The most important aspects of the structure of complex verbs were described in the very first work on Udi grammar, Schiefner (1863: 23–5). Schiefner notes that there are few simplex (monomorphemic) verbs in Udi; among them he includes *aq'* ‘take, receive’, *uk-* ‘eat’, *tit'* ‘run’, and others. The majority of verbs are what Schiefner calls compounds, which I refer to as complex verbs. He points out that these contain the verb *besun* ‘make, do’, *pesun* ‘say, do’, *desun*, *t'esun*, or less frequently *k'esun* (or its variant *xesun*), where *-(e)sun* forms a masdar (deverbal noun, often used as the citation form); I refer to these verbs henceforth as light verbs. These light verbs combine with nouns (e.g. *q'i-besun* ‘be afraid’, with *q'i* ‘fear’), including ones in the ergative-instrumental case (e.g. *q'uful-en-b-esun* [castle-INST-DO-MAS] ‘lock up’, with *q'uful* ‘castle, fortress’). They may also combine with adjectives (e.g. *ayu-b-esun* ‘make bitter’, with *ayu* ‘bitter’), or with adverbial elements (e.g. *ala-b-esun* ‘raise’, with *ala* ‘up, above’). The light verbs combine with verbs in the infinitive form (in *-es*); e.g. *ac-es* ‘disappear, be lost’ combines with *b-esun* ‘make, do’ to form *aces-besun* ‘destroy’. The light verb *desun*, *t'esun* forms causatives in the same way; e.g. *ot'bes-t'esun* ‘make someone ashamed’, based on *ot'-besun* ‘shame oneself’, based in turn on *ot'* ‘shame NOUN’.

Schiefner observes further that a number of verbs other than light verbs also incorporate nouns (e.g. *čubux-aq'-sun* ‘marry’ from *čubux* ‘woman’ and *aq'-sun* ‘take, receive’) or adverbial particles, many of which are however no longer independent words (e.g. *ba-sak-sun* ‘push in, stick in’ from *ba-* ‘in’ and *sak-sun* ‘push’).

Dirr (1904: 39–41) divides Udi verbs into three types: (a) Simple verbs, e.g. *aq'sun* ‘take, receive’, (b) verbs composed of a root, a preverbal element, and the formant of the masdar *-(e)sun* (*esun*, *b(e)sun*, *p(e)sun*, *t'esun*, *desun*, *k'esun*, *xesun*, *kesun*); e.g. *ba-p-sun* ‘put in’, with *ba-* ‘in’, and (c) verbs composed with a root verb and another constituent. In the third type the verb can combine

- (i) With a noun: *aš-besun* ‘work’ (with the noun *aš* ‘work’)
- (ii) With an adjective: *xuru-bsun* ‘reduce to fragments’ (*xuru* ‘small’)
- (iii) With an adverb: *oq'a-sak-sun* ‘push down’ (*oq'a* ‘down’)

- (iv) With a non-Udi verb: *buyurmış-besun* ‘fasten’; or with an Udi verb in *-es*, the formant of the infinitive: *ap’es-besun* ‘cook TRANS’ (*ap’sun* ‘cook INTR’)
- (v) With an interjection: *vay-besun* ‘moan, suffer’ (*vay* ‘woe!’).

Dirr (1904: 39) seems to have been the first to point out that the roots of simple verbs (his type (i)) are divided in two parts by the pronominal ‘infixes’; e.g. *a-ne-q’-sa* ‘she takes, receives’ from *aq’-* ‘take, receive’, *-ne-* 3SG, and *-sa* PRES.

Jeiranišvili (1971: 72–9) lists thirty simplex verbs, yet concedes (74) that many of these are no longer used this way. For example, he lists *oc’-* as a simplex root and provides examples of finite forms, such as *oc’-ne-sa* ‘it is washed’, *oc’-ne-c-e* ‘it was washed’, *oc’-ey-al-le* ‘it will be washed’. The root *oc’* actually occurs primarily in *oc’-d-*, a complex verb meaning ‘wash’ (see further below).

One subgroup of Jeiranišvili’s simplex category (numbers 25–30: *č’esun* ‘go out’, *baysun* ‘go in’, *esun* ‘come’, *taysun* ‘go (away)’, *laysun* ‘go up’, *cisun* ‘go down’) represent directional movement. He suggests that the adverbial portions of these were once independent words and that they are now similar to preverbs of other languages.

Among complex verbs, Jeiranišvili singles out a group that have an initial consonant which is a fossilized element, possibly with a vowel, followed by a root: *ba-k-* ‘be, become; be able’, *ba-q-* ‘have, get’, *bi-x-* ‘bear, be born’, *bi-q’-* ‘catch, hold; build’, *bo-k’-* ‘burn’, *bo-š-* ‘sate, saturate’, *bo-t’-* ‘cut, break’, *bo-s-* ‘throw’, *bo-x-* ‘boil’, *bo-q’-* ‘gather’, etc. Elsewhere Jeiranišvili (1956) has identified this *b-* as the inherited marker of the neuter gender-class agreement (see §4.2 for discussion).

Jeiranišvili (1974: 77–9) notes that either the incorporated element of a complex verb or the verbal component may be limited to use in these complexes. The complex verb *aš-b-* ‘work’ provides an example of a complex in which both components are still used independently, *aš* ‘work’ as a noun, and *b-* (masdar *besun*) as a verb ‘make, do’. An example in which the incorporated element is still used independently, while the verb is not, is *žol-d-* ‘cork up, stop up’ (*žol* ‘cork, stopper’); *-d-* is not used today as an independent verb. Jeiranišvili (1971: 78) identifies *šam-p-* ‘kill’ as consisting of productive *p-* ‘say’ (see note 4) and an element *šam* which is no longer used outside this complex.

Jeiranišvili further notes that the nominal constituent can itself be morphologically complex. Among his examples are *azaru-bak-* ‘become ill’ (based on the underlying form *azar-lu* ‘ill’ related to *azar* ‘illness, need’) and *išex-tay-* ‘marry (said of a woman)’ (where *išex* is the dative case form of *išu* ‘man’, and *tay-* is ‘go’).

Schulze (1982: 147–51) recognizes two types of Udi verb structure: simple and complex. He is careful to define the simple verb structures as those that cannot be analyzed synchronically, and points out that a number of these are complex from a diachronic point of view. Among the latter he points out two subtypes. The old locative preverbs, he says, combine with the suppletive verb base *e(γ)* ‘go’ to form verbs of directional movement, such as *lay-* ‘go up’, *bay-* ‘go in’. Schulze also draws attention to the fossilized gender-class marker (CM) *b-*, which shows up in numerous simple verbs with the structure *BVC*, such as *bak-* ‘be’, *bix-* ‘bear, give birth to’, *bok-* ‘burn’.

Complex verbs generally consist of two parts: a nominal base and a verbal element. Schulze notes that the nominal base is generally in unmarked form when it combines with a recognized auxiliary verb (HV, light verb); on the other hand, when it combines with a simple verb base the nominal base may be inflected. Formally these are of three types—those formed with one of the intransitive auxiliaries (see below), those formed with one of the transitive auxiliaries, and those formed with an independent simple verb (1982: 149):

Vn + HV ⁱ	<i>gogin-baksun</i> 'be blue'
Vn + HV ^t	<i>gogin-besun</i> 'make blue'
Vn + V	<i>xabar-aq'sun</i> 'ask' (literally 'take question')

A fourth type consists of combinations with a 'quasi-nominal', which was originally verbal, but which no longer occurs independently: *γač'-esun* 'be bound', *γač'-pesun* 'bind'. In this case the auxiliaries function to make the quasi-nominal bases into transitive or intransitive verbs.

While the intransitive HVs are only *e(γ)-* 'go' and *bak-* 'be, become', Schulze argues that the transitive HVs include practically all of the stops of the phoneme inventory of Udi: *b-*, *d-*, *p-*, *k-*, *q-*, *p'-*, *t'-*, *k'-*, *q'-* (1982: 149). The identical character of these stems becomes clear, he says, from the fact that transitives formed with any of those named above can be opposed by an intransitive formed with *-esun*.⁸

Schulze (1982: 150) lists the following verbs as ones which also allow incorporation of a nominal element: *aq'-* 'take', *duγ-* 'beat', *tast'-* 'give', *sak-* 'throw', *zap-* 'pull', *biq'-* 'catch'.

Lastly, there is a group of composed verbs whose quasi-nominal base is formed with the Turkish morpheme *-miş* (*-miş*, *-muş*, *-müs*), formant of the past participle. In Udi this element is borrowed with specific verbs; it does not exist as an independent morpheme. Schulze-Fürhoff (1994b) presents the same view in an abbreviated form.

Other aspects of the history of research on Udi are discussed below as they become relevant.

1.5. The Theoretical Basis of This Work

1.5.1. Diachronic Syntax and Morphology

The theoretical framework for the historical analysis undertaken here is provided by Harris and Campbell (1995). Of particular importance is that in historical

⁸ Schulze (1982: 150) provides evidence to support this for verbs in *k'-*, *p'-*, *d-*, and *b-*. However, in the Vartašen dialect, the word Schulze cites for *p'-* is not *γač'-p'-esun* 'bind' but *γač'()-p-esun*; see dictionary entries in Jeiranišvili (1971: 233) and Fähnrich (1999: 35) and textual examples such as D 62: 17, DG 96: 10, as well as examples such as D 62: 12, D 62: 4 that show the suppletion *-exa/p-/(u)k'-*, which characterizes the light verb 'say'. Thus, it may be that in the Nij dialect there are more light verbs than in the dialect described here.

syntax and morphology we are dealing with just three mechanisms of change: reanalysis, extension, and borrowing.

Reanalysis is defined in Harris and Campbell (1995: 50) as 'a mechanism which changes the underlying structure of a syntactic pattern and which does not involve any modification of its surface manifestation'. Underlying structure is specified as including '(i) constituency, (ii) hierarchical structure, (iii) category labels, and (iv) grammatical relations'. Surface manifestation, on the other hand, includes '(i) morphological marking, such as morphological case, agreement, and gender-class, and (ii) word order'. Extension is defined as 'a mechanism which results in changes in the surface manifestation of a pattern and which does not involve immediate or intrinsic modification of underlying structure' (Harris and Campbell 1995: 51). Borrowing is 'a mechanism of change in which a replication of the syntactic pattern is incorporated into the borrowing language through the influence of a host pattern found in a contact language'.

The three mechanisms identified apply in both syntax and morphology. It is recognized as especially important that a complete account not limit itself to the morphological source of the phonological content of a morpheme (for example, the independent pronoun *zu* 'I' as the source of the first person singular subject PM *zu*), but consider also the syntactic structure of the phrases or clauses involved in change.

In linguistics of Caucasian languages it has often been a problem that much of the syntax and some morphology is attributed to borrowing, with little evidence to support specific individual claims. I consider it as essential to present specific evidence of borrowing as to present specific evidence of inheritance or of the application of another mechanism of change. Often there is *prima facie* evidence of borrowing (that is, a similar construction exists in a contact language), but this fact alone does not actually prove that borrowing was the source of the construction. In thoroughly studied instances of contact, one can often identify an internal mechanism of change, which likely applied because of the influence of the contact languages. Because it is often difficult to find specific evidence of borrowing, it is necessary to recognize that contact may have been one of several causes of change. It is important to note that even the proven fact that a syntactic construction is due to the influence of contact languages does not absolve the linguist of the responsibility to account for the mechanisms through which it became part of the language and through which it developed further. In addition, if a construction was borrowed early and used in a proto-language, then it must be reconstructed to that proto-language; that is, the term proto-language does not include only what is inherited, but everything found in that stage of the language, whether acquired at some earlier time through borrowing or made up of inherited elements. It may be difficult or impossible to determine whether a borrowed construction was part of a proto-language or was later borrowed individually by daughter languages, but this does not mean that it should not be our goal to determine this.

Reconstruction in syntax is reconstruction of patterns, not reconstruction of specific sentences that have existed in the past (Harris and Campbell 1995, chapter 12). Syntactic patterns, like morphological patterns, can be reconstructed through application of the Comparative Method, with rigorous safeguards for comparability of data (*loc. cit.*, Harris 1985: 21–33).

Universal properties of the reduction of biclausal structures to monoclausal ones and of changes in case-marking patterns are two issues given particular attention in Harris and Campbell (1995, in chapters 7 and 9, respectively), and these two are particularly relevant to the Udi changes described in this book. Here these two questions arise especially in Chapters 10 and 11, and more background on theoretical issues is provided there.

1.5.2. *Theoretical Underpinnings*

To the extent possible, I have kept the descriptive chapters (Chapters 1–6) free of the assumptions of any specific syntactic or morphological theory, though I have assumed syntactic constituency and hierarchy.

In my analysis of focus, I have relied heavily on Lambrecht (1994), who takes the approach that every sentence has focus; my analysis has also been influenced by recent work on focus in Hungarian and a number of other diverse languages (references in Chapter 3).

In analyzing complex verbs, I have been influenced by the framework set out in Ackerman and Webelhuth (1998), which provides a theory of predicates. Most pertinent to the present problems is that the approach in their work provides the basis for treating an auxiliary and a dependent verb form as a predicate. Seen through the diachronic lens of Harris and Campbell (1995), this means that an ‘auxiliary’ and a ‘dependent’ verb may originate as independently viable verbs, each governing the syntax of a clause in a biclausal structure. As a consequence of reanalysis, the once independent verbs may take on their ‘auxiliary’ and ‘dependent’ verb status, now forming a single predicate. Through further reanalysis and through extension, these may, over time, become a single verb. Though more visible, the many small changes that create a single verb from a two-part predicate are more superficial than the single reanalysis that makes a biclausal structure into a monoclausal one, and turns two verbs into a single predicate. It is proposed in Chapter 10 that such a series of changes took place in Udi.

To determine whether PMs and other ‘particles’ are clitics, I have turned first to Zwicky and Pullum (1983), but also have considered important work by other specialists in this area. The analysis of the position of clitics in the word and in the phrase in Udi is based on Optimality Theory. The theory developed in Harris and Campbell (1995) is assumed in dealing with historical issues, though in many instances I have not discussed here the relevance of Udi data to that framework.

The data presented here represent a challenge to the Lexicalist Hypothesis. Some theoretical aspects of this are discussed briefly in Chapter 7 within the

framework of Optimality Theory. That analysis has drawn most on McCarthy and Prince (1993), together with other works cited there.

This work is intended to be primarily descriptive and to provide an explanation of why Udi is the way it is, and thus of why it provides a counterexample to the Lexicalist Hypothesis.