Bringing Order(ing) into Chaos: the Placement of Udi Endoclitics*

Peter W. Smith, University of Connecticut
peter.w.smith@uconn.edu

1. Introduction

- Do we need timings of operations in the grammar or can we deal with an all-at-once effort?
- Even if something can in principle be captured both with and without recourse to ordering of operations, does ordering (or not) buy us anything?

➩ **In this talk I show that having some order of operations buys us a great deal in analyzing an extremely complex system of clitic placement in Udi (Nakh-Daghestanian).**

I will show that allowing certain operations to be interleaved with others in a derivational model of morphology allows us to unify numerous parts of the Udi placement paradigm in a principled way. These aspects become coincidental if we do not accept the presence of ordering in the grammar.

Roadmap

§2: Overview of the Udi data
§3: Second position placement
§4: The phonological asymmetries
§5: Conclusions

2. The Udi subject marker clitic paradigm

Udi is rare among the languages in the world in that it contains endoclitics. These are different from the more familiar enclitics and proclitics in that they appear *internal* to a word, and not lying at the periphery.

Only a handful of other languages exhibit these types of clitics, for instance Sorani Kurdish (Samvelian 2007, Walther 2012), Pashto (Tegey 1977, Roberts 1997), European Portuguese (Anderson 2005), Old Irish (Kern 2011).

* Thanks to Jonathan Bobaljik, Alice Harris, Beata Moskal and Susi Wurmbrand for various discussions on the material here.

Endoclitics are exhibited in Udi by the clitics which mark the agreement features of the subject, as is shown below (data from Harris 2002, as is all Udi data unless otherwise noted):

(1) a. äyel kala-**ne**-bak-e
child.ABSL big-3SG-BECOME-AORII
'The child grew up.'

b. q'ačaγ-γ-on bez tänginax baš-**q'un**-q'-e
thief-PL-ERG my money.DAT steal$_1$-3PL-steal$_2$-AORII
'The thief stole my money.'

In (1a) the clitic *ne* lies internal to the complex verb, between the incorporating verb *bak* and the verb *kala* which is incorporated.

In (1b) the clitic lies internal to the verb root itself, which is realized discontinuously (as indicated by the subscripts in the gloss).

These are not the only positions that the clitic lies in. Harris (2000, 2002) shows that there are seven distinct positions that can be identified for the clitic:

Rule 1: Clitics are final in the V*x* if the verb is in the future II, the subjunctive I, the subjunctive II, or the imperative.
Rule 2: Clitics occur enclitic to a focused constituent.
Rule 3: In clauses with zero copulas, clitics are enclitic to predicate nominals.
Rule 4: Clitics are endoclitic in a complex verbstem, occurring between the Incorporated element (IncE) and the light verb or verb root.
Rule 5: For verbstems of class M, in the intransitive, clitics are endoclitic occurring between the verbstem and the present tense marker.
Rule 6: With verbs that are either phonologically C or CV, clitics are enclitic to the entire verb form.
Rule 7: Clitics are endocliticized immediately before the final consonant in monomorphemic verbstems.

These rules are hierarchically ordered. So, whenever Rule 1 can apply, it does so. Rule 2 can only apply if Rule 1 can't and so on.

(2) a. baba-n ęš nut eč-al-**le** k'wa
father-ERG apple.ABSL NEG bring-FUTII-3SG house.DAT
'Father will not bring apples to the house.'

b. nu aq'-a-**n** box-ala k'ok'oc'-ax
NEG take-SUBJI-2SG boil-PTCPL chicken-DAT
'You should not take the chicken that it to be cooked.'

(3) nana-n ten-**ne** bụγa-b-e p'ạ ačik'alšey
mother-ERG NEG-3SG find-DO-AORII two toy.ABSL
'Mother did not find two toys.'

➩ The cases that I will focus on in this talk are those covered by Rules 4 through 7. Here, we find that the clitics begin to switch between being enclitics and endoclitics.

With a complex verb, the clitic goes between the incorporated element and incorporating verb (Rule 4):

(4) äyel kala-**ne**-bak-e
child.ABSL big-3SG-BECOME-AORII
'The child grew up.'

(5) nana-n tur-ex oc'-**ne**-k'-e
mother-ERG foot-DAT wash-3SG-LV-AORII
'Mother washed her foot.'

(6) pasčaγ-on γar-muγ-on lašk'o-**q'un**-b-esa
king-GEN boy-PL-ERG wedding-3PL-DO-PRES
'The king's son's married.'

In some verbs there seems to be a difference based on transitivity of the verb. In the transitive form the clitic goes inside the verb root, whilst in the intransitive form the clitic appears to go between the verb and its tense-aspect-mood (TAM) suffix (Rule 5):

(7)

Transitive	Intransitive
a. a-**t'u**-k'-sa see$_1$-3SG-see$_2$-PRES 'he sees'	b. ak'-**ne**-sa see-3SG-PRES 'it shows, is visible'

(8)

a. bo-**ne**-x-sa boils$_1$-3SG-boils$_2$-PRES 'he boils, cooks'	b. box-**ne**-sa boils-3SG-PRES 'it boils (intr.)'

(9)

a. bọ-**ne**-q'-sa gather$_1$-3SG-gather$_2$-PRES 'he gathers'	b. bọq'-**ne**-sa gather-3SG-PRES 'it gathers, is gathered'

When the verb consists solely of a consonant or CV sequence, then the clitic goes to the end of the verb form (Rule 6).

(10)

a. b-esa-**ne** make-PRES-3SG 'she makes'	b. k-e-**ne** eat-AORII-3SG 'she ate'

(11) bi-esa-**zu**
die-PRES-1SG
'I am dying'

Finally, when none of the other rules apply, then the clitic goes internal to the verb root itself. This causes the root to be realized discontinuously (Rule 7):

(12) q'ačaγ-γ-on bez tänginax baš-**q'un**-q'-e
thief-PL-ERG my money.DAT steal$_1$-3PL-steal$_2$-AORII
'The thieves stole my money.'

(13) kaγuz-ax a-**z**-q'-e
letter-DAT receive$_1$-1SG-receive$_2$-AORII
'I received the letter.'

The parallels to infixation are immediately apparent.

However, it is not clear that we should treat these cases in the same way. In Smith (to appear) I argue that there is reason to be suspicious of importing approached to infixation to endoclisis.

Udi appears to be the only known case of a clitic appearing internal to another morpheme, so allowing for the same treatment as infixation has UG overgenerating in a large way.

If this is correct then we need to look for an alternate analysis of these clitics.

3. Second position placement

I propose that the key to understanding the placement of Udi endoclitics is to recognize that there is a default rule of placement that positions the clitics in second position within the verb.

That is, when the clitic comes to be placed, it will be so according to the following rule:

(14) Linearization of subject clitics: elsewhere case
clitic + [$_V{}^0$ X ≫ (Y)...] → [$_V{}^0$ X ≫ clitic ≫ (Y)...]

Second position within the verb is also seen in Sorani Kurdish, where the boldface clitic is always the second morpheme within the verb:

(15) a. dît-**yân**-im
saw-3PL-1SG
'They saw me'

b. ná-**yân**-dît-im
NEG-3PL-saw-1SG
'They didn't see me'

c. da-**m**-xwârd
PROG-1SG-eat.PAST
'I was eating.'

d. na-**m**-da-xwârd
NEG-1SG-PROG-eat.PAST
'I was not eating.'

3.1. Complex verbs

Second position placement is transparently true in the case of complex verbs (see (4-6) above).

Complex verbs in Udi are formed of an incorporated element and a light verb, and the TAM suffix. Thus, the spellout of the verb has the following basic structure:

(16) [$_V{}^0$ IncE - light verb - TAM]

When the clitic comes to be placed then, it will position itself between the element that is incorporated, and the light verb.

(17) clitic + [$_V{}^0$ IncE - light verb - TAM] → [$_V{}^0$ IncE - clitic - light verb - TAM]

Note though that this is a default rule, and will only apply in the absence of a more specific one. For instance, when the TAM morpheme is in one of the categories of Rule 1, the clitic will go there, and not in second position:

(18) bez vič-en aš-b-al-**le** zavod-a
my brother-ERG work-DO-FUTII-3SG factory-DAT
'My brother will work in a factory.'

3.2. Simplex verbs

Whilst second position within the verb is transparently true in the case of complex verbs, it seems to make the wrong prediction for simplex verbs.

Unlike the complex verbs above, these roots are not decomposable into separate morphemes.

Second position placement in the case of the simplex root *bak* ('be') would give the (ungrammatical) following:

(19) *bak-**ne**-sa sa pašč'aγ-k'ena adamar.
be-3SG-PRES one king-like person.ABSL

Instead, we find that the clitic goes internal to the simplex verb root:

(20) ba-**ne**-k-sa sa pašč'aγ-k'ena adamar.
be$_1$-3SG-be$_2$-PRES one king-like person.ABSL
'[Once upon a time, there] is a person like a king.'

⇨ What I propose here is that second position placement is the correct analysis, but we must recognize that there is an extra repair operation in play. The clitic is placed between *bak* and *sa* (as in (19)), but needs to move from that position.

The clitic moves because in that position it interrupts the adjacency between the verb root and the TAM suffix. This is a violation of the morphotactics of Udi.

There is a surface constraint on Udi, such that (some version of) the following holds:

(21) *root-X-TAM

⇨ This requirement is surface true in Udi, since there exists no case where the verb and TAM suffix are non-adjacent.

⇨ I propose that we view this as a real, deep property of Udi.

But.....
We are left with a conflict. The placement rule of (14) wants to place the clitic in second position, but (20) won't allow it to appear there.

⇨ In these instances, the clitic is moved leftwards, internal to the verb root. This allows the right edge of the root to be adjacent to the right edge of the TAM suffix, and (21) is satisfied.

Consider the (simplified) derivation in (23) for how the verb in (22) is derived:

(22) pasčaɣ-un ɣar-en gölö be-**ne**-ɣ-sa met'a-laxo
king-GEN boy-ERG much look$_1$-3SG-look$_2$-PRES this.GEN-on
'The prince looks at this for a long time.'

(23)

Input to cliticization		Adjacency violation and repair		Output
beɣ-sa	→	*beɣ-ne-sa	→	be-ne-ɣ-sa
look-PRES		*look-3SG-PRES		look$_1$-3SG-look$_2$-PRES

⇨ **Viewed like this, the surprising situation where the clitic is placed internal to another morpheme is in reality the clitic placed intermorphemically followed by metathesis of the clitic to a position inside the root to save a potential morphotactic violation.**

3.3. Transitive/Intransitive alternations

Whilst it appears as though a clitic shift marks transitivity in (7-9) above, these are actually regular cases. The present tense of the intransitive forms contains a null form of a suppletive light verb, which incorporates the verb seen in the intransitive.

This can be seen by changing tense of the intransitive form (light verb is underlined):

(24) a. box-e̲ɣ̲-al-**le**
boil-GO-FUTII-3SG
'it will boil'

b. box-**ne**-c̲-e
boil-3SG-GO-AORII
'it boiled'

The intransitive/transitive alternation in these cases is simply the addition of a light verb in the intransitive, which alters what it means to be second position.

(25)

Input to cliticization		Output
ak'-∅-sa	→	ak'-**ne**-∅-sa
see-GO-PRES		see-3SG-GO-PRES

The transitive variants then involve second position placement, and leftward metathesis:

(26)

Input to cliticization		Adjacency violation and repair		Output
ak'-sa	→	*ak'-**t'u**-sa	→	a-**t'u**-k'-sa
see-PRES		*see-3SG-PRES		see_1-3SG-see_2-PRES

4. The two asymmetries

Everything seems straightforward so far. A strange situation where a clitic is apparently placed internal to another morpheme is nothing more than the resolution of a (somewhat rare) placement rule conflicting with the surface requirements. However, the story gets more complicated.

There is a curious phonological sensitivity at play.

4.1. Clitic placement doesn't care about phonology - apart from when it does

A glance at the above rules show that phonological considerations don't seem to play a role in the placement of the clitic in Udi.

Rule 1 makes reference to syntactic/morphological notions such as subjunctive, futureII and imperative, irrespective of the phonological shape of the exponents.

Similarly, rules 2 and 3 make reference to notions of focus and predicate nominals, both syntactic notions and not phonological.[2]

But, there is an instance where the phonological information of a host becomes important.

Rules 6 and 7 actually both cover simplex verbs,[3] and it is possible to write the descriptively true statement, producing asymmetry #1:

[2] Harris (2002) notes that although there is a tendency for some of these elements to receive stress, an account based on stress determining the placement of the clitic is not possible. For further discussion I defer to Harris (2002: 139-143).

[3] Ignoring some genuine exceptions that Harris puts under rule 6, which I, like her, ignore for the purposes here.

Asymmetry #1: with monomorphemic verbs, if the verb consists minimally of shape (C)VC (i.e. there is a postvocalic consonant), then the clitic is positioned immediately before the final consonant. If there is no postvocalic consonant (the verb is C or CV), the clitic appears enclitic to the TAM morpheme.

Thus, we now have the phonological shape of the verb root influencing the position of the clitic.

We seem to be able to predict at this point why the clitic appears enclitic to the TAM suffix in the following:

(27) k-e-**ne**
eat-AORII-3SG
'she ate'

Note that there appears to be a general prohibition against proclisis in Udi, since there exist none in the language.

(28)	i. input to cliticization:	k-e
	ii. second position placement:	k-**ne**-e
	iii. metathesis repair:	k-e-**ne**

Here the clitic positions itself at the end of the verb form. We can see this as the same process that would otherwise place the clitic inside a verb. Leftward metathesis fails because in order to allow the right edge of the root and the left edge of the TAM suffix to be adjacent, it would cause *ne* to be a proclitic moving leftwards.

Rightward metathesis on the other hand satisfies both the adjacency requirement of root +TAM ((21) above) and the enclitic nature of the clitic.

So, we can see why the clitic appears at the end of the verb form - it's simply the repair going the other way to avoid proclisis.

Problem solved?

⇨ Almost.

4.2. The second asymmetry

A problem for this approach is that there are instances where there is a potential consonant in the TAM morpheme. We might expect the clitic to attach before this consonant:

(29) b-esa-**ne**
make-PRES-3SG
'she makes'

(30) bi-esa-**zu**
die-PRES-1SG
'I am dying'

The present tense exponent *esa* contains *s,* which we expect to be a potential host for the clitic, if the clitic is making the minimal movement away to allow the adjacency of root +TAM. But the clitic rigidly goes outside the entire TAM suffix.

(31) *bi-e-**zu**-sa
die-PRES$_1$-1SG-PRES$_2$

This is not simply a further rule that places the clitic at the end of the verb form when none of the others apply however, since in these instances, it can appear between the TAM suffix and the past clitic *-y*:

(32) b-esa-**ne**-y
do-PRES-3SG-PAST
'she was doing'

Asymmetry #2: in repairing a morphotactic violation, the clitic can target the phonological information of the root, but not affixes.

4.3. Resolving the asymmetries

I propose that asymmetry #2 can be explained by adopting a strictly derivational model of the morphology, such as DM, where some operations happen before others.

I assume Distributed Morphology (Halle & Marantz 1993), where spell-out of words involves insertion of phonological exponents replacing phonological features.

⇨ In DM, is assumed that VI proceeds cyclically from the root outwards (Bobaljik 2000, Embick 2010 a.o.).

Bobaljik (2000) shows that this model allows phonological asymmetries to be explained in a principled way.

I propose that it is this order of operations - specifically the fact that VI of the root occurs before VI of TAM suffixes - that allows us to explain asymmetry #2 in Udi.

Concretely, I propose that the repair operations are interleaved with vocabulary insertion of items, and that the morphotactic violations of Udi are evaluated <u>at the point of VI of the violator</u>.

What this means is that when the clitic causes a morphotactic violation after being placed in second position within a complex verb form, it won't be moved until it undergoes VI.

➩ This means that it will undergo metathesis at some intermediate point of spell out. Crucially, the phonological information of the root is already present in the derivation when the clitic is moved.

Consider first the 'regular' cases, where the clitic is moved inside a simplex verb:

(33) ba-**ne**-k-sa sa pašč'aγ-k'ena adamar.
be$_1$-3SG-be$_2$-PRES one king-like person.ABSL
'[Once upon a time, there] is a person like a king.'

(34)

i. input to cliticization:	√BE-[+PRES]
ii. second position placement:	√BE-[3SG]-[PRES]
iii. VI of root:	/bak/-[3SG]-[+PRES]
iv. VI of clitic:	/bak/-/ne/-[+PRES]
v. metathesis repair:	/ba-ne-k/-[+PRES]
vi. VI of TAM:	/ba-ne-k-sa/

(35) kaγuz-ax a-**z**-q'-e
letter-DAT receive$_1$-1SG-receive$_2$-AORII
'I received the letter.'

(36)

i. input to cliticization:	√RECEIVE-[+AORII]
ii. second position placement:	√RECEIVE-[1SG]-[+AORII]
iii. VI of root:	/aq'/-[1SG]-[+AORII]
iv. VI of clitic:	/aq'/-/z/-[+AORII]
v. metathesis repair:	/a-z-q'/-[+AORII]
vi. VI of TAM:	/a-z-q'-e/

Note that the morphotactic violation is only evaluated at stage iv. This allows us to capture asymmetry #2, since the clitic moves when the phonological information of the root is in the derivation, but crucially *not* the phonological information of the TAM suffix.

Consider the relevant derivations below:

(37)

i. input to cliticization:	√EAT-[+PRES]
ii. second position placement:	√EAT-[3SG]-[+PRES]
iii. VI of root:	/k/-[3SG]-[+PRES]
iv. VI of clitic:	/k/-/ne/-[+PRES]
v. metathesis repair:	/k/-[+PRES]-/ne/
vi. VI of TAM:	/k-e-ne/

(38)	i. input to cliticization:	√DIE-[+PRES]
	ii. second position placement:	√DIE-[1SG]-[+PRES]
	iii. VI of root:	/bi/-[1SG]-[+PRES]
	iv. VI of clitic:	/bi/-/zu/-[+PRES]
	v. metathesis repair:	/bi/-[+PRES]-/zu/
	vi. VI of TAM:	/bi-esa-zu/

⇨ Therefore even though the TAM suffix has the potential to host the clitic being of correct phonological shape, since the exponent of it hasn't entered into the derivation by the time the clitic moves, there is no way that the clitic can position itself inside. All that is seen is a feature bundle.

The analysis given above relies crucially on the steps of spell out happening *in that order*.

⇨ If the morphotactic violation were evaluated earlier, before any element has undergone VI, then it is not possible to capture the 'standard' case where the clitic moves internal to the verb root.

⇨ If the clitic moved after all elements have undergone VI, then it is not possible to explain the fact that phonological information of the root can be targeted, but not the TAM suffix.

⇨ Without a cyclic model of spellout where some operations are ordered before others, all of this comes out as a coincidence.

5. Conclusions

In this talk I have presented an analysis on the placement of Udi endoclitics.

It has been shown that their complicated placement is not a list of rules of placement (*contra* Harris 2002), but rather a single, default rule of placement, plus a repair operation of metathesis which works differently according to the phonological shape of the root.

This phonological sensitivity has been shown to come from the structure of spellout.

Under DM assumptions about late insertion of exponents, and the cyclic nature of spellout such that it proceeds outwards from the root, we capture the fact that metathesis of the clitic can only target the phonological structure of the root, but not the TAM suffix.

Peter W. Smith
University of Connecticut
peter.w.smith@uconn.edu
http://homepages.uconn.edu/~pws10003

References

Anderson, S. (2005). Aspects of the theory of clitics. Oxford, Oxford University Press.

Arregi, K. and Nevins, A. (2012). Morphotactics: Basque auxiliaries and the structure of spellout. Springer.

Bobaljik, J.D. (2000). The ins and outs of contextual allomorphy. In K. Grohmann and C. Struijke, eds., *University of Maryland Working Papers in Linguistics,* volume 10, 35-71.

Embick, D. (2010). Localism versus globalism in morphology and phonology. MIT Press.

Halle, M. and Marantz, A. (1993). *Distributed morphology and the pieces of inflection.* In Hale, K. and Keyser, S. (eds.), The view from building 20. Cambridge, MA: MIT Press.

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

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

Kern, G. (2011). Incliticization? Apparent word-internal clitic placement in Pashto and Old Irish. Ms. MIT.

Roberts, T. (1997). *The optimal second position in Pashto.* Rutgers Optimality Archive #174. http://roa.rutgers.edu/files/174-0297/174-0297-ROBERTS-0-0.PDF

Samvelian, P. (2007). What Sorani Kurdish absolute prepositions tell us about cliticization. *Proceedings of the Texas Linguistic Society IX.*

Smith, P.W. (to appear). On the cross-linguistic rarity of endoclisis. To appear in *The Proceedings of the 39th Meeting of the Berkeley Linguistics Society*

Tegey, H. (1977). *The grammar of clitics: evidence from Pashto and other languages.* Doctoral dissertation, University of Illinois at Urbana-Champaign.

Walther, G. (2012). Fitting into morphological structure: accounting for Sorani Kurdish endoclitics. http://www.linguist.univ-paris-diderot.fr/~gwalther/homepage/Publications_(fr)_files/MMM8_GWalther.pdf