Peter W. Smith
University of Connecticut

GLOW 36
Lund, April 2013

Endoclisis (only) by Way of Repair*

1. Udi subject marker clitics

Harris (2000, 2002) describes the system of subject agreement clitics in Udi (Northeast Caucasian). Interestingly, the clitics in this system vary between being enclitics and endoclitics:

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

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

They mark person, number and case features of the subject of the clause.

The clitics are obligatory, and always are either enclitic or endoclitic. There are no proclitics in the language, showing a general ban against proclisis in Udi.

Harris shows that there is a complex ranked hierarchy of placement positions (PM = person marker = subject clitic):

Rule 1: PMs 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: PMs occur enclitic to a focused constituent.
Rule 3: In clauses with zero copulas, PMs are enclitic to predicate nominals.
Rule 4: PMs 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, PMs are endoclitic occurring between the verbstem and the present tense marker.
Rule 6: With verbs forms of category A and category B, PMs are enclitic to the entire verb form.
Rule 7: PMs are endocliticized immediately before the final consonant in monomorphemic verbstems.

Application of rule 1 prevents rule 2 from applying.

* Thanks to Jonathan Bobaljik, Beata Moskal, Susi Wurmbrand and audiences at BLS 39 and PLC 37 for invaluable comments and feedback. Special thanks are due to Alice Harris for sharing her knowledge of Udi with me.

So, if the verb is in the future II tense, subjunctive I,II or imperative form, then the subject marker appears enclitic to the verbal complex (hence application of rule 1):

(3) 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.'

If none of these TAM categories are present, then the clitic attaches to the constituent in focus (rule 2):

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

And so on...

The part that I will focus on in this paper will be the rules which Harris claims produce endoclitics. These are rules 4, 5 and 7. I will claim that they all fall under the same rule of placement, which is second position within the complex containing the verbal head.

Firstly, rule 4, with Harris' OT alignment constraint:

> Rule 4: PMs are endoclitic in a complex verbstem, occurring between the Incorporated element (IncE) and the light verb or verb root.
>
> Align-PM-IncE
> Align (PM,L,IncE,R)

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

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

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

(5-7) show that in the absence of focus and relevant TAM suffixes, the clitic places itself in between the incorporated element and light verb.

When none of the other alignment rules apply, rule 7 kicks in and the clitic is placed inside the verbal root by the Align-PM-verbstem constraint:

> Rule 7: PMs are endocliticized immediately before the final consonant in monomorphemic verbstems.
>
> Align-PM-Verbstem
> Align(PM,R,Verbstem,R)

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

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

The paradigm governing rule 5 is the alternation between transitive and intransitive forms. Some examples are listed below. The *a* examples are transitive, *b* examples intransitive (see section 4.3 for why there is no alignment constraint here):

> Rule 5: For verbstems of class M, in the intransitive, PMs are endoclitic occurring between the verbstem and the present tense marker.

(9)

	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'

(10)

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

(11)

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

Harris accounts for the distribution of PMs in Udi with a system of OT alignment constraints. They are ranked to ensure that wherever a relevant TAM suffix is in the sentence, it bleeds

placement of the clitic to focus, and so on. The two that govern endoclisis are Align-PM-IncE and Align-PM-Verbstem:

(12) Align-PM-*al/a* ≫ Align-PM-FocC ≫ Align-PM-IncE ≫ Align-PM-Verbstem
= constraint for (TAM) (focus) (complex verbs) (simplex verbs)

(13) Align-PM-*al/a*[1]
Align (PM,L,-*al/-a*,R)
Read as: "align the left edge of the person marker to the right edge of -*al/-a*"

(14) Align-PM-FocC
Align (PM,L,FocC,R)

(15) Align-PM-IncE
Align (PM,L,IncE,R)

(16) Align-PM-Verbstem
Align(PM,R,Verbstem,R)

2. Should we allow the grammar to directly generate endoclitics?

Harris' system works, but is entirely descriptive in that there is one constraint per rule.

By proposing Align-PM-Verbstem, Harris, and the representational nature of standard OT (Prince & Smolensky 1993, McCarthy & Prince 1995) allows the syntax to directly position clitics inside morphemes according to their PF alignment preferences.

Yu (2007), working in a different framework, also proposes that endoclitics should be directly generable.

This however is a very powerful device to allow into the toolkit of universal grammar. There is little other evidence that I am aware of of any other syntactic object moving inside another host.

The question is whether this move is warranted. If it really is warranted, then we need to find a way of modeling intramorphemic placement, but also find appropriate ways to constrain it. If it is not warranted, we need to find another way of capturing the Udi data.

⇨ I argue that recognizing endoclisis within UG is not warranted. The Udi data can be analyzed in terms of enclisis and surface readjustment to a position inside the morpheme.

[1] Whilst Harris writes the constraint in terms of phonetic content, she does so for parsimony. *al/a* are the exponents of future II, subjunctive I, subjunctive II and imperative.

3. Endoclisis elsewhere?

If UG does permit intramorphemic placement of clitics, then we would expect to find other cases elsewhere.

This doesn't need to be at the same frequency of enclitics and proclitics - infixation is rare in comparison to prefixation and suffixation - but we would certainly expect other cases in other languages.

3.1. Intermorphemic placement

There are scattered reports of endoclitics in languages other than Udi. Generally these take the form of clitics that are placed within morphemes.

For instance, European Portuguese (Anderson 2005) has subject agreement clitics which appear between the verb root and the agreement (as in (a', b') below):

(17) a. daríamos
give.1PL.COND

a'. dár-**te**-íamos (European Portuguese)
give-1PL-1PL.COND

b. perceberás
understand.2SG.FUT

b'. percerbér-**me**-ás
understand-2SG-2.SG.FUT

Sorani Kurdish (Bonami & Samvelian 2008 and Walther 2012) also appears to show endoclitics, again appearing intermorphemically:

(18) a. ward-**man**-in
eat-**1PL**-3PL
'We ate them.'

b. na-**m**-xward-in (Sorani Kurdish)
NEG-**1PL**-eat-3PL
'I didn't eat them.'

Though it is interesting in itself that clitics can be placed in these positions - it goes against the extremely robust tendency that clitics appear outside affixes (Zwicky & Pullum 1983) - their placement is not difficult once we make the DM assumption that word formation is done syntactically.

3.2. Intramorphemic placement?

It has been claimed that there exists another instance of "true" endoclisis, in Pashto (see Tegey 1977) where a clitic goes inside a morpheme.

The relevant data is in (19,20), where in the imperfective form stress is either penultimate/final or initial:

(19) a. axistə́lə me
buy 1SG
'I was buying them.'

b. á-me-xistələ
buy$_1$-I-buy$_2$
'I was buying them.'

(20) a. aɣustə́ me
wear 1SG
'I was wearing it.'

b. á-me-ɣustə
wear$_1$-1SG-wear$_2$
'I was wearing it.'

The claim by Tegey is that *axistələ* and *aɣustə* are monomorphemic.

> "It is important to bear in mind that in such instances the clitics are placed after a phonological segment which constitutes part of the root (i.e. '*a'* - PWS), and which is not a separate morpheme. (Tegey 1977:89)"

The placement of clitics in Pashto is clearly governed by stress (Tegey 1977, Roberts 1997, Anderson 2005).

But, an analysis where the clitic is placed solely in the phonology next to the stressed syllable is not sufficient. When the verb in question is not one of the *a-* initial verbs, the clitic follows the stressed *morpheme*, not the syllable:

(21) pərebdə́ me
beat 1SG
'I was beating him.'

b. pə́rebdə me
beat 1SG
'I was beating him.'

Indeed, Kaisse (1981) and Roberts (1997) (see also Anderson 2005) all show that the *a*-class verbs in (19, 20) above are actually complex, and that the *a-* part that Tegey claimed was part of the root is actually a prefix. This gives the verbs the structure in (19) instead, in the relevant cases:

(22) a. á-me-xistələ
PREFIX-1SG-buy

b. á-me-ɣustə
PREFIX-1SG-wear

The Pashto data then no longer constitute a case of intramorphemic placement as we see that the clitic aligns according to the stressed morpheme.

Pashto then also does not force us to recognize endoclisis as an operation of UG.

Moreover, there doesn't appear to be any known case apart from Udi which casts serious doubt on whether endoclisis truly is part of UG at all.

With Udi being the only exponent of endoclisis, it is worth investigating whether the facts can be captured using a system of enclisis and readjustment of the position of clitic positions, something which is independently attested across languages (see Arregi & Nevins 2012).

4. Udi Analysis

I claim that what appears to be a endoclisis in Udi is actually enclisis + surface readjustment. In this way, I am following Halle's treatment of infixiation in Chamorro, Tagalog and Toba Batak.[2]

4.1. Movement of clitics and elsewhere placement in second position

In what follows, I will show that the cases covered by Harris' rules 4-7 are all covered by a default rule of placement, shown in (23d) below:

(23) a. PMs are enclitic to the TAM categories Future II, subjunctive I, subjunctive II and imperative. (= Harris' Rule 1)
b. PMs are enclitic to focus. (= Harris' Rule 2)
c. PMs are enclitic to predicate nominals. (= Harris' Rule 3)
d. Elsewhere, PMs are enclitic to the first element within the complex head containing the verb.

For the remainder of this talk, I concentrate on how (23d) interacts with other aspects of Udi morphotactics in order to produce the illusion of endoclisis.

(23d) immediately accounts for the placement of the clitic in complex verbs, since it will be positioned after the incorporated element:

(24) pasčaɣ-on ɣar-muɣ-on lašk'o-**q'un**-b-esa
king-GEN boy-PL-ERG wedding-3PL-DO-PRES
'The king's son's married.'

(25)

Input to cliticization		Output of cliticization
lašk'o-b-esa	→	lašk'o-**q'un**-b-esa
wedding-DO-PRES		wedding-3PL-DO-PRES

4.2. Metathesis as a repair

Simply placing the clitic in second position works for the cases of complex verbs, however it is not a complete analysis for the cases where the clitic is positioned intramorphemically, because second position would still be after the root in these examples.

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

[2] Though see Yu (2007) for a critique of Halle's approach to infixation. It is not however clear that Yu's criticisms carry over to the present treatment of endoclisis, since this assumes that they should be accounted for under the same theory.

(26) *<u>bak</u>-**ne**-sa sa paščʼaγ-kʼena adamar.
be-3SG-PRES one king-like person.ABSL

Instead, the clitic must obligatorily go inside the root, so we find (27) instead:

(27) <u>ba</u>-**ne**-<u>k</u>-sa sa paščʼaγ-kʼena adamar.
be$_1$-3SG-be$_2$-PRES one king-like person.ABSL
'[Once upon a time, there] is a person like a king.'

⇨ Here I propose that the clitic gets placed intramorphemically due to a confluence of three factors:

i. second position placement of the clitic.
ii. second position placement in these cases causes a morphotactic violation.
iii. the violation triggers a repair.

The morphotactic violation caused by clitic placement is that it interrupts a requirement of Udi that verb root and TAM suffix are adjacent:

(28) *root-X-TAM

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

Placing the clitic in second position would give ***bak-ne-sa***, violating the adjacency requirement.

When this adjacency is violated, the morphology enacts a repair to allow convergence. The chosen repair for Udi is metathesis, which moves the clitic to allow the right edge of the root to be adjacent to the TAM suffix.

⇨ Morphological metathesis as a repair has been recently argued for in Arregi & Nevins (2012) for Basque clitics.

⇨ I assume that metathesis moves the clitic the minimal amount required to allow the (right edge of the) root and TAM suffix to be adjacent, thus adhering to the morphotactic requirements of Udi.

⇨ Further I assume that this violation is evaluated at the point of spell-out of the clitic (I return to this below).

⇨ Assuming that vocabulary insertion (VI) proceeds from the root outwards (Bobaljik 2000, Embick 2010 *inter alia*), this means that at the point where the morphotactic violation is seen, the phonological features of the root are in the derivation.

For a sample derivation consider how *ba-ne-k-sa* (be$_1$-3sg-be$_2$-PRES) in (27) above is derived:

(29)

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/

For concreteness, I follow Harris & Halle's (2005) approach to metathesis as reduplication (see also Arregi & Nevins 2012), so step (v) above is more accurately represented as the following (grey shading indicates deletion in the following):

(30) ba[k⟩⟨ne]-[+PRES] → ba-kne-kne-[+PRES] → ba-ne-k-[+PRES]

4.3. Transitive/intransitive alternations

The elsewhere placement rule of the clitic allows us to make sense of the intransitive/transitive alternations in (9,10,11), repeated below:

	Transitive	Intransitive
(9)	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'
(10)	a. bo-**ne**-x-sa boils$_1$-3SG-boils$_2$-PRES 'he boils, cooks'	b. box-**ne**-sa boils-3SG-PRES 'it boils (intr.)'
(11)	a. bo-**ne**-q'-sa gather$_1$-3SG-gather$_2$-PRES 'he gathers'	b. boq'-**ne**-sa gather-3SG-PRES 'it gathers, is gathered'

Harris shows that the intransitive variants are formed with a suppletive light verb *go*, which is phonologically null in the forms in (9-11). Clitic placement is then in the regular second position and follows the complex verb analysis of section 4.1 ((31) below the derivation of (9b)):

(31)

Input to cliticization	⟶	Output of cliticization
ak'-∅-sa		ak'-ne-∅-sa
see-GO-PRES		see-3SG-GO-PRES

The transitive variants then involve second position placement, and leftward metathesis as in section 4.2.

(32)

Input to cliticization	→ Adjacency violation and repair →	Output of cliticization
ak'-sa	*ak'-t'u-sa	a-t'u-k'-sa
see-PRES	*see-3SG-PRES	see$_1$-3SG-see$_2$-PRES

4.4. Greater predictive power

In the previous two sections I showed that a derivational approach to Udi clitic placement is possible and there isn't a need to resort to a system of descriptive OT alignment constraints to capture the data, as Harris does. In this section I show that my system, coupled with the architecture of DM, has more predictive power than Harris'.

Placing the clitic in second position allows us to have one rule of placement for both complex verbs and simplex verbs. In Harris' system they were previously separate alignment constraints. Since we need an additional rule of metathesis however, there is little that we can point to in favoring one theory over the other.

However, there are cases where endoclisis fails where we would expect it in simplex roots. These require an extra constraint in Harris' system, whereas it falls out naturally from mine.

These are the cases described by Harris' rule 6, category A:[3]

> Rule 6: With verbs forms of category A and category B, PMs are enclitic to the entire verb form.

Some examples of these forms are given below; note the final position of the clitic (boldfaced) within the verbal complex:

(33) a. b-esa-**ne**
make-PRES-3SG
'she makes'

b. k-e-**ne**
eat-AORII-3SG
'she ate'

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

Harris doesn't give a constraint to capture these cases, but there must be an alignment constraint which places the clitic enclitic to the verb form where no other rule can apply. Alice Harris (p.c.)

[3] Category B verbs are described by Harris as true lexical exceptions requiring special treatment. They are a small class and I do not consider them further.

says that this would apply when the verbstem is either too small to accommodate a clitic (single consonant roots cannot have clitics inside them) or is an open syllable.

The approach given here however predicts that the clitic would appear verb finally in these cases without saying anything extra if in cases where metathesis is prevented from applying leftward it applies minimally rightward instead.

⇨ Recall that there is a general prohibition on proclisis in Udi. This may be due to them having a suffixal nature (for instance Noyer 1992, Wojdak 2005).

⇨ Also, assume that metathesis cannot apply to open syllables.[4]

Firstly consider the derivation for *k-e-ne* ('he eats') (from (33b) above):

(35) 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/

At the point where the metathesis repair would be triggered, moving the clitic leftward in the regular manner would cause it to be a proclitic. Therefore, metathesis instead applies rightwards moving the clitic outside the TAM suffix.

The crucial step of metathesis, in step (v), is more accurately represented as (36) below, following Harris & Halle (2005) (again, grey shading indicates deletion):

(36) k-[ne⟩⟨[+PRES]] → k-[ne[+PRES]]-[ne[+PRES]] → k-[+PRES]-ne

⇨ Crucially this leaves the right edge of the root to be adjacent to the TAM suffix, satisfying the morphotactic requirements of Udi, with the output being /k-e-ne/ (eat-PRES-3sg).

An interesting consequence of the above is that we can see why the clitic goes to the final position in the verbal complex, even when the exponent of TAM is in principle large enough to host the clitic.

⇨ For instance in (34) above, where *-esa,* the exponent of present tense, has the correct syllable structure to host a clitic, yet we don't find *bi-e-zu-sa* (= die-PRES$_1$-1sg-PRES$_2$)

[4] This amounts to saying that open syllables cannot host clitics, which is an equally required constraint in Harris' approach.

This is because at the point that the metathesis repair is enacted, only the morphosyntactic features of the TAM suffix are present within the derivation. As the phonological exponent is missing, the operation cannot be sensitive to the phonological structure of the exponent (in the spirit of Bobaljik 2000). This is shown in the derivation of *bi-esa-zu* (die-PRES-1sg) below:

(37) 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/

The crucial step of metathesis, in step (v), is more accurately represented as (38) below:

(38) bi-[zu⟩⟨[+PRES]] → bi-[zu[+PRES]]-[zu[+PRES]] → bi-[+PRES]-zu

5. Conclusions

The analysis presented here shows that Udi does not force us to recognize endoclisis as an operation of UG.

This theory makes the correct cut. It allows the endoclisis seen in Udi, but only indirectly by surface readjustments.

Endoclisis arises in Udi due to an interaction of three quirks of Udi:

i. The elsewhere placement rule of PM clitics, (23).
ii. The adjacency requirement of root and TAM suffix, (28).
iii. The availability of metathesis as a repair to adjacency violations (as opposed to, say, deletion of the PM clitic).

We can recognize that endoclisis can occur *indirectly* as a pattern seen in languages, but we must further recognize that it is predicted to be rare, requiring a number of factors to come together.

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

Selected references (please email me for the complete list)
Arregi, K. and Nevins, A. (2012). *Morphotactics: Basque auxiliaries and the structure of spellout*. **Bobaljik, J.D. (2000).** The ins and outs of contextual allomorphy. **Embick, D. (2010).** Localism versus globalism in morphology and phonology. **Halle, M. (2001)**. Infixation versus onset metathesis in Tagalog, Chamorro, and Toba Batak. **Harris, A. (2002).** *Endoclitics and the origins of Udi morphosyntax.* **Harris, J. & Halle, M. (2005).** Unexpected plural inflections in Spanish: Reduplication and metathesis. **Roberts, T. (1997)**. *The optimal second position in Pashto.* **Tegey, H. (1977)**. *The grammar of clitics: evidence from Pashto and other languages.* **Walther, G. (2012).** Fitting

into morphological structure: accounting for Sorani Kurdish endoclitics. **Yu, A. (2007).** *A natural history of infixation.*