Peter W. Smith
University of Connecticut

PLC 37
March 2013

The Indirect Nature of Endoclisis*

1. Introduction

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) a. 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.'

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.'

(2) a. äyel kala-**ne**-bak-e (endoclitic)
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 $steal_i$-3PL-$steal_2$-AORII
'The thief stole my money.'

2. Udi endoclitics

Udi has a system of clitics showing subject agreement. They mark person, number and case features of the subject of the clause.

(3)

	Absolutive/Ergative	Dative
1SG	-zu, -z	-za
2SG	-nu, -n, -ru, -lu	-va
3SG	-ne, -le, -re	-t'u
1PL	-yan	-ya
2PL	-nan, -ran, -lan	-va, -vạn
3PL	-q'un	-q'o

* I'd like to thank Alice Harris for invaluable discussion on the Udi data and proposal presented here. Thanks also to Jonathan Bobaljik, Beata Moskal and Susi Wurmbrand for comments and pointers. All errors mine.

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.

The system of subject clitic placement in Udi is extremely complex. Harris shows that there is a ranked hierarchy of placement positions (PM = person marker = subject clitic):

Rule 1: PMs are final in the V*x*[1] 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.

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):

(4) a. baba-n ẹš 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):

(5) 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.'

And so on...

[1] Harris uses the notation V*x* to mean the complex consisting of the verb and negative.

The part that I will focus on in this talk 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 the 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)

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

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

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

(6-8) 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)

(9) a. q'ačaɣ-ɣ-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.'

b. kaɣuz-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 constriant here:

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

(10)

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'

(11)

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

(12)

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

3. The problem of endoclitics

Endoclitics pose various theoretical problems. Harris points out that because the clitics require access to syntactic information (i.e. focus), they must be placed by the syntax. But in the model she assumes, this means that the syntax will have access to the internal structure of words, in violation of lexical integrity (DiSciullo & Williams 1987).

This problem in large part disappears if we adopt the assumptions of Distributed Morphology (DM, Halle & Marantz 1993) where the lexicon is not some component before the syntax, but rather word formation is done within the syntax.

If Harris is correct and clitic placement is syntactic, then some of these problems begin to go away. For instance with complex verbs, the subject marker no longer causes any issues since it is placed between morphemes, something fairly easily accounted for in DM.

(13) IncE-(PM)-light verb-TAM
(TAM henceforth is shorthand for tense-aspect-mood suffix)

However, there still remains a problem; with the contexts of simplex verbs stems such as below (14a = 9a, 14b = 12a):

(14) 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 thief stole my money.'

b. bo-**ne**-q'-sa
gather$_1$-3SG-gather$_2$-PRES
'gathers'

The morphemes which the clitic places itself in are unambiguously monomorphemic according to Harris (though see Luís & Spencer (2006) for a suggestion that they're bimorphemic).

Placing the clitic intramorphemically within the syntax does cause a major problem for most frameworks, especially and including distributed morphology, as noted by Morris Halle when discussing infixation:

> "Since the phonetic exponents of morphemes are thus not present in the syntax, it is literally impossible within the syntax to infix /um/ or /in/ before the first vowel of the Tagalog stem. (Halle 2001:153)"

4. 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 Chamorro/Tagalog/Toba Batak infixation.[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 (16d) below:

(16) 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.

[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. For discussion, see Smith (forthcoming).

For the remainder of this talk, I concentrate on how (16d) interacts with other aspects of Udi morphotactics in order to produce the illusion of endoclisis.

(16d) immediately accounts for the placement of the clitic in complex verbs, since it will be positioned after the incorporated element:

(17) 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.'

(18) 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 (EXS), 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:

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

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

(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.'

⇨ 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:

(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.

Placing the clitic in second position violates this adjacency, as we see above:

(22) *bak-ne-sa
root-clitic-TAM

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 (Embick 2010, Bobaljik 2012 *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 (20) above is derived:

(23)	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):

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

Below is the derivation for the verb form *a-z-q'-e* (receive$_1$-1SG-receive$_2$-AORII) in the following:

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

(26)
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/

4.3. Transitive/intransitive alternations

The elsewhere placement rule of the clitic allows us to make sense of the intransitive/transitive alternations in (10,11,12), repeated below:

(10)

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'

(11)

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

(12)

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 (27 below the derivation of (10b)):

(27)

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.

(28)

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:

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

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

(30) 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. A. Harris (p.c.) 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.

[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.

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 (29b) above):

(31)	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 (32) below, following Harris & Halle (2005) (again, grey shading indicates deletion):

(32) 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 (30) above, where *-esa,* the exponent of present tense is the correct syllable structure (CVC) to host a clitic, yet we don't find *bi-e-zu-sa* (= die-pres$_1$-1sg-pres$_2$)

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

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

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:

(33)	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 (34) below:

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

5. Implications for UG and endoclisis in general

The Udi data are particularly interesting since they present the only clear case in the literature of true endoclisis, i.e. positioning of a clitic within a morpheme.

There are scattered reports of other cases of endoclitics, but they involve placement of clitics intermorphemically. See for instance European Portuguese (Anderson 2005), Pashto (Tegey 1977, Roberts 1997) and Sorani Kurdish (Bonami & Samvelian 2008, Walther 2012).

Harris' treatment in terms of OT alignment constrains builds into UG the ability of the grammar to directly generate endoclitics.

⇨ This is a powerful addition to the theory of UG to capture the facts of just one language. We might expect more cases to arise, contrary to observation (see Smith forthcoming for more discussion on this point).

In contrast, on the approach outlined here we seem to make the correct cut.

⇨ Endoclisis (at least *true* endoclisis) can arise but only due to a confluence of language specific factors. We then expect it to be rare cross-linguistically); it is not part of UG.

⇨ Endoclisis in Udi arises due to the interaction of second place clitic positioning, the violation of a morphotactic rule, and the availability of metathesis as a repair.

6. Conclusions

➩ In this presentation I have presented a derivational DM analysis of Udi endoclitics.

➩ This analysis captures the same facts as Harris', but offers greater predictive power.

➩ Endoclisis arises indirectly, due to a repair operation that moves the clitic in order to satisfy morphotactic requirements.

➩ The Udi facts do not require us to recognize endoclisis as an aspect of UG.

References

Anderson, S. (2005). *Aspects of the theory of clitics.* Oxford studies in theoretical linguistics, 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.
Bobaljik, J. D. (2012). *Universals in comparative morphology.* MIT Press.
Bonami, O. & Samvelian, P. (2008). *Sorani Kurdish person markers and the typology of agreement.* Paper presented at the 13th International Morphology Meeting, Vienna.
Di Sciullo, A. and Williams, E. (1987). On the definition of Word. Cambridge, MA: MIT Press.
Embick, D. (2010). Localism versus globalism in morphology and phonology. MIT Press.
Halle, M. (2001). Infixation versus onset metathesis in Tagalog, Chamorro, and Toba Batak. In *Ken Hale: A life in language.*
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.
Harris, J. & Halle, M. (2005). Unexpected plural inflections in Spanish: Reduplication and metathesis. *Linguistic Inquiry 36,* pp.195-222.
Luís, A. & Spencer, A. (2005). Udi clitics: A generalized paradigm function morphology approach. Available at http://privatewww.essex.ac.uk/~spena/papers/UDI.PDF
Noyer, R. (1992). Features, positions and affixes in autonomous morphological structure. Doctoral disseration, 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
Smith, P.W. (forthcoming). On the cross-linguistic rarity of endoclisis. *Proceedings of the 39th meeting of the Berkeley Linguistic 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
Wojdak, R. (2005). The linearization of affixes: evidence from Nuu-chah-nulth. Doctoral dissertation, University of British Columbia.
Yu, A. (2007). *A natural history of infixation.* Oxford Studies in Theoretical Linguistics.