Peter W. Smith
University of Connecticut

BLS 39
February 2013

On the cross-linguistic rarity of endoclisis*

1. Introduction.

The broad goal of this talk is to consider the status of true endoclisis with respect to theories of Universal Grammar.

Endoclitics are clitics that are neither proclitic nor enclitic. In other words, they appear inside words. If we look at the typology of endoclitics we find two types:

- *Intermorphemic placement*, as exemplified by European Portuguese (from Anderson 2005):

(1) 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

- *Intramorphemic placement*, as (purportedly) exemplified by Pashto (2), and the main case study of this talk, Udi, (3).

(2) a. á-me-xistələ
buy$_1$-1SG-buy$_2$

b. á-me-ɣustə (Pashto)
wear$_1$-1SG-wear$_2$

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

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

Clitics that are placed intermorphemically will not concern us here. Whilst they do appear to violate Lexical Integrity (Di Sciullo & Williams 1987), there are a variety of ways across different frameworks to account for their placement, see for instance Crysmann (1997) on European Portuguese (in HPSG).

* I would like to thank Alice Harris in particular for invaluable help with respect to her knowledge of Udi. I would also like to thank Jonathan Bobaljik, Beata Moskal and Susi Wurmbrand for discussion and pointers on the ideas here. All errors are of course my own.

Intramorphemic placement of clitics is more challenging for linguistic theory since it would involve a complex interaction of morphology, syntax and phonology, which is impossible to model directly in some (but crucially not all) frameworks.

My aim in this talk is not to argue *for* a specific framework[1], but instead to argue *against* a realizational approach to the problem, along the lines of Optimality Theory (OT).

⇨ Specifically, I will argue that an approach to cliticization that allows the syntax to place clitics directly with respect to phonological information is too permissive, and as such is inappropriate for a theory of endoclisis (see Kaisse 1981, and below for a similar claim based on Pashto).

2. The nature of the Udi problem.

Udi (Northeast Caucasian) is a Caucasian language spoken mainly in parts of Azerbaijan and surrounding areas.

It has a system of subject agreement clitics that have an extremely complex system of placement. As outlined in detail by Harris (2000, 2002), person marker clitics are placed according to syntactic information.

The person marker clitics (PMs) under discussion are as follows:[2]

(4)

	General	Inversion
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

[1] Though, in order to present an analysis of the Udi data, I must of course adopt a framework. For concreteness, I set my analysis within *Distributed Morphology* (DM, Halle & Marantz 1993), but this is not directly relevant to the matter at hand. For further discussion of why I adopt DM, I refer to discussion in Smith (forthcoming).

[2] This is an incomplete paradigm of the clitics. There is also a class of possessive clitics and a question clitic, but their placement remains the same so they are not important here. The inversion class of PM clitics mark the subject of experiencer predicates, see Harris (1984) for details.

2.1. They really are clitics.

Important for this discussion is the conclusion that the PMs are clitics, and not affixes. Harris (2000, 2002) provides extensive discussion that leads her to the conclusion that this is the case, showing that they conform to the properties of clitics outlined by Zwicky & Pullum (1983), and do not act like affixes:

Criterion A: Clitics can exhibit a low degree of selection with respect to their hosts, while affixes exhibit a high degree of selection with respect to their stems.
Criterion B: Arbitrary gaps in the set of combinations are more characteristic of affixed words than of clitic groups.
Criterion C: Morphophonological idiosyncrasies are more characteristic of affixed words than of clitic groups.
Criterion D: Semantic idiosyncrasies are more characteristic of affixed words than of clitic groups.
Criterion E: Syntactic rules can affect affixed words, but cannot affect clitic groups.
Criterion F: Clitics can attach to material already containing clitics but affixes cannot.

⇨ Harris shows that in all these tests, the PMs behave uniformly as clitics and not affixes.

2.2. They are placed with respect to syntactic information.

Harris proposes a system of rules that describes placement of the PMs:[3]

Rule 1: PMs are final in the V*x*[4] 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.

Rules 1 to 3 all require direct reference to syntactic information.

[3] I have modified the wording slightly for clarity here.

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

⇨ The problem that this causes is that it suggests that the clitics are placed in their positions during the syntactic derivation. Rules 4, 5 and 7 are then problematic because it allows the syntax to generate endoclitics. Importantly, for many theories the internal structure of words is inaccessible to the syntactic derivation, with words built coming out of the lexicon.

The contexts that concern us presently are those of Rule 4 and 7. Illustrative examples are given below.

The structure of complex verbs in Udi is:

(5) IncE-(PM)-light verb-TAM suffix

In (6) we see the PM (boldface) between the incorporated element and the light verb:

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

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

c. 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.'

Where there is a monomorphemic verb, i.e. not an incorporating light verb, the PM marker appears before the final consonant of the root. In this instance, the structure is therefore, where $Root_2$ consists of the final consonant and any subsequent vowels of the root:

(7) $Root_1$=PM=$Root_2$-TAM suffix

(8) a. pasčaγ-un γar-en gölö **be**-ne-**γ**-sa met'a-laxo verb = *beγ*
king-GEN boy-ERG much $look_1$-3SG-$look_2$-PRES this.GEN-on
'The prince looks at this for a long time.'

b. kaγuz-ax **a**-z-**q'**-e verb = *aq'*
letter-DAT $receive_1$-1SG-$receive_2$-AORII
'I received the letter.'

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:

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

(10) Align-PM-*al/a*[5]
Align (PM,L,-*al/-a*,R)
Read as: "align the left edge of the person marker to the right edge of -*al/-a*"

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

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

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

The constraints that govern endoclisis are (12) and (13).

Align-PM-IncE, (12), aligns the left edge of the PM to the right edge of the incorporated element in complex verbs.

Align-PM-Verbstem, (13), aligns the right edge of the PM to the right edge of the verbstem.

2.3. Directly allowing endoclisis?

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

Below I will offer a more economical account of Udi, where Harris' rules 4, 5, 6 and 7 can be collapsed into one 'elsewhere' placement rule.

The true motivation for moving away from Harris' account is not purely on grounds of economy considerations. 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.

[5] 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.

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 this move is not warranted. The Udi PM endoclitic placement can be analyzed in terms of encliticism and surface readjustment to a position inside the morpheme.

3. Udi Analysis.

3.1. Outline of the system, assumptions and rules.

Having shown that there is cross linguistic reason to be suspicious of endoclisis arising directly in Udi, it remains to show that there is another way of capturing the Udi data. There have been a few recent attempts to capture the Udi data in a variety of non-OT frameworks, for instance Crysmann (2000) (HPSG), Luís & Spencer (2006) (PFM) and Wescoat (2009) (LFG).

For my analysis I adopt Distributed Morphology (Halle & Marantz 1993). DM is a non-lexicalist theory of morphology where word formation is done within the syntax. As such, this allows us to see syntactic structure within words.

⇨ Importantly, using DM, we can assume that the clitics can be placed syntactically, but still have access to the internal structure of words.

Working within DM, we can also have a more predictive set of rules compared to Harris'. Seeing the internal structure of words allows us to formulate the default rule of (16) below, which replaces Harris' rules 4,5 and 7 (rule 6 constitutes exceptions to rule 7, which I discuss below):

(14) PMs are enclitic to the first element within the complex head containing the verb.

I therefore assume the following ranked set of placement rules:

(15) a. PMs are enclitic to the TAM categories Future II, subjunctive I, subjunctive II and imperative.
b. PMs are enclitic to focus.
c. PMs are enclitic to predicate nominals.
d. Elsewhere, PMs are enclitic to the first element within the complex head containing the verb.

⇨ This system is immediately attractive since PMs are uniformly enclitic. Even the cases of endoclisis that Harris observes are assumed to be underlyingly enclisis.

3.2. The elsewhere placement rule.

3.2.1. Complex verbs.

The elsewhere placement rule allows us to easily conflate rules 4 and 5 of Harris'. It is not at all uncommon to find second position clitics (so called Wackernagel clitics). They are widely attested at the sentential level in diverse languages, from Serbo-Croatian (see for instance Boskovic 2001), and as we saw above, in Pashto (Roberts 1997, Anderson 2005).

This rule will uniformly place the clitic after the incorporated element within complex verbs:

(16) 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.'

(17)

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

3.2.2. Transitivity alternations.

Recall Harris; rule 5:

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

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 intrasitive:

(18) a. a-t'u-k'-sa
see$_1$-3SG-see$_2$-PRES
'sees'

b. ak'-ne-sa
see-3SG-PRES
'shows, is visible'

(19) a. bo-ne-x-sa
boils$_1$-3SG-boils$_2$-PRES
'boils, cooks'

b. box-ne-sa
boils-3SG-PRES
'boils (intr.)

(20) a. bǫ-ne-q'-sa
gather$_1$-3SG-gather$_2$-PRES
'gathers'

b. bǫq'-ne-sa
gather-3SG-PRES
'gathers, is gathered'

It appears as though the shift in transitivity is somehow marked by altering the properties of the clitic. However, Harris shows that the intransitive form really contains a suppletive light verb 'go' that is null in this instance. The pattern therefore directly fits in with the complex verbs outlined above:

(21)

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

3.2.3. Simplex verbs.

With respect to simplex verbs we seem to face a problem. Placement of the PM in second position in the V^0 complex would predict that the PM appear between the simplex root and the TAM suffix. However, this is not the case and it is in this instance that we find true endoclisis, where the clitic is positioned inside the verbal root:

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

(23) kaɣuz-ax **a**-z-**q'**-e verb = *aq'*
letter-DAT receive$_1$-1SG-receive$_2$-AORII
'I received the letter.'

In these cases, I propose that the intramorphemic placement is not a case of syntactic endoclisis; rather the PM is inserted in second position, but moves leftward to repair a morphotactic violation.

I assume that there is a language specific requirement of Udi that the verbal root and TAM suffix are surface adjacent to each other, (24).

(24) *root-X-TAM suffix

⇨ 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 here triggers a repair. The clitic is metathesized leftwards inside the final consonant, leaving the right edge of the root to be adjacent to the TAM suffix (metathesis as a morphotactic repair has been recently employed by Arregi & Nevins 2012 in Basque):

(25)

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

Thus, there is no sense in which the PM clitics are true endoclitics, they only are on the surface because they have shifted inwards. The syntax does not place the PM within a morpheme but rather the mapping from morphology to phonology is responsible.

We can see that this is on the right track by looking at the exceptions to endoclisis that Harris groups under Rule 6. There are two groups of verbs that do not conform to the expected (surface) endoclisis like the verbs above. The category B verbs Harris mentioned seem to be true, non-predictable lexical exceptions to the rule, and I don't discuss them further here.

The Category A verbs are far more interesting. The verbs forming this category are either single consonant verbs, (26), such as *b-* 'do, make', *p-*'say' (suppletive) and *k-* 'eat', or they are CV roots, (27), such as *bu-* 'be', *bi-* 'die' (suppletive).

⇨ In both these instances the PM is placed in verb final position:

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

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

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

Harris accounts for these cases (p.c.) by saying that there is a pre-condition on endoclisis that the stem be an open syllable. Where this fails, the clitic goes outside, which we can consider the final alignment constraint of Harris' system (though she doesn't formulate it as such). But this is clearly a descriptive constraint, without any reason that the clitic goes finally instead of simply appearing outside the root.

However, in the approach outlined above, we can predict why the PM appears word finally. It cannot stay in its inserted position, since it would intervene between the root and TAM suffix. We now see that metathesis must apply to remove the material (PM) that interrupts the adjacency between root and TAM.

But, suppose that there is a restriction on metathesis that it cannot move an element inside a syllable that is CV. There is no other option other than to apply metathesis rightwards outside of the TAM suffix.[6,7]

(28)

Input to cliticization		Adjacency violation and repair		Output of cliticization
_____bi-esa	→	*bi-zu-esa	→	bi-esa-zu
die-PRES		die-1SG-PRES		die-PRES-1SG

4. 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 sufffixation - but we would certainly expect other cases elsewhere.

It has been claimed that there exists another instance of endoclisis, in Pashto (see Tegey 1977). The relevant data is in (29,30), where in the imperfective form stress is either penultimate/final or initial:

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

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

(30) 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)"

[6] I assume (as must Harris) that PMs are not allowed to be proclitics. There are no proclitics in Udi (Alice Harris, p.c.) suggesting that there is a language wide ban on proclisis.

[7] Some may question why the PM does not metathesize inside the TAM suffix, particularly where the phonological form *-esa* is used for present tense. This would create the ill-formed **bi-e-zu-sa.* Within DM, this is in fact predicted if vocabulary insertion applies cyclically. At the point where the grammar can see that the root and TAM suffix are non-adjacent (i.e. at spell-out of the PM), the phonological information of the root is visible, but only the syntactic features of the TAM suffix, and no phonological information, so if rightward metathesis applies, it will move the PM outside the TAM suffix as a whole. For more discussion of this matter in Udi, I refer to Smith (forthcoming), and on the interleaving of morphology and phonology see Bobaljik (2000).

At first glance, the data might seem problematic for my claim that clitic placement cannot directly break up a morpheme.

➪ However, the placement of clitics in Pashto is clearly governed by stress (Tegey 1977, Roberts 1997, Anderson 2005), which does not cause so much of a problem here. As shown above, I allow surface readjustments to break up morpheme integrity.

But, this analysis 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:

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

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

However, Kaisse (1981) and Roberts (1997) (see also Anderson 2005) all show that the *a*-class verbs in (16,17) 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:

(32) a. á-me-xistələ
PREFIX-1SG-buy

b. á-me-ɣustə
PREFIX-1SG-wear

The Pashto data then no longer cause a problem. 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, and so adds credence to the analysis presented here for Udi.

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, (16).
ii. The adjacency requirement of root and TAM suffix, (26).
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 as a pattern seen in languages, but we must further recognize that it is predicted to be rare since it requires a number of factors to come together.

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 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.
Bošković, Ž. (2001). On the nature of the syntax-phonology interface. Cliticization and related phenomena. Amsterdam: Elsevier.
Crysmann, B. (1997). Cliticization in European Portuguese using parallel morpho-syntactic constraints. *Proceedings of the LFG97 conference.*
Crysmann, B. (2000). On the placement and morphology of Udi Subject Agreement. http://www.coli.uni-sb.de/ crysmann/papers/UdiAgr.html
Di Sciullo, A. and Williams, E. (1987). On the definition of Word. Cambridge, MA: 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. (1984). Case marking, verb agreement and inversion in Udi. In Perlmutter, D. and Rosen, C. (eds.) *Studies in Relational Grammar 2,* pp. 243-258.
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.
Kaisse, E. (1981). Separating phonology from syntax: a reanalysis of Pashto cliticization. Journal of Linguistics 17, 197-208.
Luís, A. & Spencer, A. (2005). Udi clitics: A generalized paradigm function morphology approach. Available at http://privatewww.essex.ac.uk/~spena/papers/UDI.PDF
Prince, A. & Smolensky, P. *Optimality theory: constraint interaction in generative grammar.*
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. (2013). The indirect nature of endoclisis. *Talk to be given at the 37th Penn Linguistics Colloquium, March 22-24, Philadelphia, PA.*
Tegey, H. (1977). *The grammar of clitics: evidence from Pashto and other languages.* Doctoral dissertation, University of Illinois at Urbana-Champaign.
Wescoat, M. (2009). *Udi person markers and lexical integrity. Proceedings of the LFG09 conference.*
Zwicky, A. & Pullum, G. (1983). Cliticization vs inflection: English n't. *Language 59, 502-513.*