Peter W. Smith
University of Connecticut

LAGB 2013
August 2013

Evidence for second position cliticization within a word

1. Introduction

In this talk, I study two complex systems of clitic placement, and show that they provide evidence that clitics can target second position **within a word.**

I show that once we recognize that word internal second position is a possible clitic position, we can account for the complex systems of Udi (Nakh-Dagestanian, Harris 2002) and Sorani Kurdish (Samvelian 2007, a.o.).

2. System 1: Udi

Udi, as described in detail by Harris (2002) has an extremely complicated system of clitic placement.

Clitics which mark person and number features of the subject of the clause can appear in a wide variety of positions.

(1) a. baba-n ęš nut ečal-**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$_1$-3PL-steal$_2$-AORII
'The thief stole my money.'

Within the chaos however, there is a predictable system of clitic placement. Harris gives a list of seven descriptive rules, which are hierarchically ranked and cover all cases of clitic placement.

Rule 1: Clitics 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: 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 forms of category A and category B, clitics are enclitic to the entire verb form.
Rule 7: Clitics are endocliticized immediately before the final consonant in monomorphemic verbstems.

Application of rule 1 prevents rule 2 from applying, and so on.

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 tense-aspect-mood (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 talk will be the rules which Harris claims produce endoclitics. These are rules 4, 7 and 6. I will claim that they all fall under the same rule of placement, which is second position within the complex containing the verbal head.

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

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

(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) 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.'

(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.

(8) a. q'ačaɣ-ɣ-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. kaɣuz-ax a-**z**-q'-e
letter-DAT receive$_1$-1SG-receive$_2$-AORII
'I received the letter.'

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:[2]

[2] 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.

> 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:

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

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

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

In sum, Udi clitics can appear in various positions within the sentence. When they appear on the verb (and not on a predicate nominal or focused element), they occupy various positions within the verb:

- After the incorporated element.
- At the end of the verb.
- Inside the verbal root.

3. Approaches to account for Udi

Harris (2002) gives a system of OT alignment constraints in order to account for the Udi data.

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

(12) Align-PM-*al/a*[3]
Align (PM,L,-*al/-a*,R)
Read as: "align the left edge of the person marker to the right edge of -*al/-a*"

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

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

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

[3] 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.

However, as can be seen from the constraints, they merely restate the descriptive rules of placement in terms of alignment constraints. There is little predictive power to the constraints aside from capturing the Udi data.

Furthermore, Harris' constraint (15) which governs root-internal clitic placement is problematic. It directly permits clitics to be placed inside another morpheme. This is a powerful addition to UG.

This seems too permissive, given that Udi is the only documented case of a clitic appearing internal to another morpheme, and it happens in only a subset of environment in which these clitics occur.

In Smith (to appear), I argue that this move is unwarranted from a cross-linguistic perspective. If UG did allow for clitics to be placed within other morphemes, then we would expect more cases cropping up in the languages of the world. To the best of my knowledge, this is elsewhere unattested.[4]

If we can account for the Udi data without allowing the grammar to directly place clitics within other morphemes, then we are able to keep any theory of clitics more constrained.

Other approaches to the Udi data have taken this approach. For instance, Luís & Spencer (2005) have claimed that what appears to be clitics appearing internal to simplex morphemes in Udi is actually clitics appearing between two elements of a complex morpheme.

4. Second position placement in Udi

I claim that the key to understanding the Udi data, in all of the cases covered by Harris' rules 4-7 is recognizing that their default position of placement is second position within the word.

Once we recognize this default placement rule, then we are able to principally predict where the clitics appear in Udi verb forms.

What makes this analysis tenable is the fact that deviations from second position within the word are entirely predictable and easily captured if we assume, following Arregi & Nevins (2012), that clitics can be moved from their original placement sites in order to comply with the morphotactics of a language.

[4] Pashto could potentially be brought up as a counterexample to this claim, based on claim of Tegey (1977), where the clitic *me* in *á-me-xistələ* 'I was buying them' ostensibly appears internal to the verb root *axistələ*. Kaisse (1981) however, provides compelling arguments that *axistələ* is bimorphemic, with *a* constituting a bound prefix. This means then that the clitics which mark the subject agreement in Pashto, such as *me,* then simply allow for the rare, but attested (see Smith to appear for an overview), possibility of word internal intermorphemic placement.

4.1. Complex verbs

Second position placement immediately accounts for the complex verb forms of Udi, where the clitic appears between the incorporated element and the verbal root.

(16) 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.'

(17)

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

Second position placement is transparent here, since the clitic follows the first morpheme of the word.

4.2. Simplex verbs

Placing the clitic in second position with a simplex verb however gives the wrong result. In the following, if second position placement were observed here, then we would expect to see (18), which is ungrammatical, instead of (19), which we do observe:

(18) *bak-**ne**-sa sa paščʼaγ-kʼena adamar.
be-3SG-PRES one king-like person.ABSL

(19) ba-**ne**-k-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.'

In these instances however, we can appeal to the morphotactic rules of Udi and see that the clitic is actually placed by the grammar in second position but is then forced to move away since it violates the morphotactics of Udi.

The violation in question is that as follows:

(20) **Nothing can intervene between the verb root and the TAM morpheme in Udi**.

This condition is rigidly obeyed in Udi. Nothing ever intervenes between the root and the TAM morpheme.

If clitics are placed in second position within a word in Udi, and the morphotactic rule in (20) is operative in the language, then there is a clear conflict; placement in second position will interrupt the adjacency of verb root and TAM morpheme.

In order to save the derivation from crashing, I propose that the clitic is metathesized inside the verbal root, allowing the right edge of the root and the left edge of the TAM morpheme to be adjacent.

Assuming that this requirement is evaluated at the point of vocabulary insertion (VI) of the clitic, following standard Distributed Morphology (Halle & Marantz 1993) assumptions on VI, such that it proceeds from the root outwards (Bobaljik 2000, Embick 2010), the phonological information of the root will already be present during the derivation when the clitic comes to be metathesized.

Consider the following derivation:

(21) ba-**ne**-k-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.'

(22)	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/

This then adheres to the requirement in (20) above, and there is nothing to crash the derivation.

4.3. Word final placement

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:[5]

[5] 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.

> 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:

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

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

(24) 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.) 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.

How does this fit in with the current analysis?

⇨ 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.[6]

Firstly consider the derivation for *k-e-ne* ('he eats') (from (23b) above):

(25)	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.

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

⇨ 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 (24) 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$_2$)

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:

(26)	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/

5. System 2: Sorani Kurdish

Claiming that second position within a word is a theoretical possibility allowed for by UG with respect to clitics requires that other cases are found which display the same requirement.

Such a case is found in Sorani Kurdish (Indo-Iranian), where we again seem to find word internal second position placement.

Samvelian (2007), Bonami & Samvelian (2008) and Walther (2012) show that, like Udi, Sorani Kurdish has a system of clitic placement where there *prima facie* doesn't seem to be any uniformity regarding where certain clitics are placed by the grammar.

The clitics in question mark the phi-features of the subject of the clause.[7]

When there is material available (for instance, an overt subject or object), the clitics appear as enclitic to the first constituent within the VP (Samvelian 2007):

[7] For reasons of time I focus here only on the past tense forms of the verb. Sorani Kurdish is complicated in that in the present tense, the features of the subject are marked by a verbal affix, and the clitic marks the object features, whilst this pattern is reversed in the past tense. Walther (2012) proposes that this is a system of morphological reversal in the sense of Baerman (2007), but it could potentially simply reflect a pattern of ergative split in the language (J. Bobaljik, p.c.).

(27) ba dûrbin-**mân** dît-in
with binoculars-**1PL** see-3PL
'We saw them with binoculars.'

When there is no material preceding the verb, the clitics appear internal to the verb form, in varying positions:

(28) na-**m**-xwârd (second position)
NEG-**1SG**-eat.PAST
'I did not eat'

(29) nard-in-**î** (third position)
send.PAST-3PL-**3SG**
'He sent them'

(30) xward-bû-in-**î** (fourth position)
eat.PAST-PART-3PL-**3SG**
'He had eaten them'

(31) xward-bû-**man**-in (third position)
eat.PAST-PART-**1PL**-3PL
'We had eaten them'

Samvelian (2007) briefly considers an analysis of second position within the verb, but rejects it on the grounds that it cannot account for the forms in (29-31).

However, we again can see that there are predictable patterns, and that the default position seems to be second position within the word. This is shown in (32-35) (from Thackston 2006):

> **"When the logical agent affix does not precede the verb (i.e. if only the verb, or verb + logical subject, and no other element is present), the logical subject agent affix is suffixed to the verb (=verbal root - PWS), and the logical object (affix - PWS) follows the agent affix (Thackston 2006: 47)."**

(32) dît-**im**-î
saw-1SG-2SG
'I saw you'

(33) dît-**yân**-im
saw-3PL-1SG
'They saw me'

(34) ná-**m**-dît-î(t)
NEG-1SG-saw-2SG
'I didn't see you'

(35) ná-**yân**-dît-im
NEG-3PL-saw-1SG
'They didn't see me'

As can be seen, when there is no morpheme preceding the verb root, the clitic goes after it. However, when the negative prefix is added, then the clitic moves between it and the verb, and now precedes the verb.

Departures from second position are seen in two contexts (perhaps more, see Walther 2012):

- 3sg clitics always follow object markers.
- Clitics never intervene between the verb and its participle marker.

Like in Udi, we can think of these as part of the morphotactic requirements of Sorani Kurdish. We can then say that the clitic *is* placed by the grammar in second position in the word, but can be moved later on in the derivation in order to satisfy the surface requirements of the language.

Hence, we find the following forms:

(36) xward-bû-**man**-in (third position)
eat.PAST-PART-**1PL**-3PL
'We had eaten them'

(37) xward-bû-in-**î** (fourth position)
eat.PAST-PART-3PL-**3SG**
'He had eaten them'

6. Discussion

In this talk I have showed that there exist clitic systems which place clitics, as a default rule, as second position within a particular word.

The notion 'word' is problematic, but it may be possible to analyze this as being second position with a complex head.

We can think of the domain of cliticization in both Udi and Sorani Kurdish as being the V^0 (unless the clitic is placed somewhere else such as a focused element in Udi, or the material preceding the verb within VP in Sorani Kurdish).

If this is the case, then it raises interesting implications for any theory of cliticization.

For instance, Anderson's (1992) claim that clitics are phrasal affixes appears to be too strong in this regard, since there are clearly clitics which target the internal structure of a complex head.

It also raises interesting implications for the study of clitics in general; for instance, does 'second position' in the sense discussed fall under the same process as second position more generally, such as in Serbo-Croatian?

Peter W. Smith
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.

Bonami, O. & Samvelian, P. (2008). *Sorani Kurdish person markers and the typology of agreement.* Paper presented at the 13th International Morphology Meeting, Vienna.

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.

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.

Smith, P.W. (to appear). 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.