

On the Cross-Linguistic Rarity of Endoclis^{*}

PETER W. SMITH

University of Connecticut

1 Introduction

This paper discusses the phenomenon of endoclis^{*} and why it is so rarely found across languages. Endoclis^{*} refers to the situation where a clitic appears neither as a proclitic at the beginning of a word nor an enclitic at the end, but in fact appears internal to the word itself. As a phenomenon, it is found in remarkably few languages around the world. In fact, as Harris (2002) points out, in various frameworks it is considered to be impossible. However, as further shown by Harris, there do exist cases where it seems undeniable that clitics appear internal to a word. Harris makes this claim based on data from Udi (Northeast Caucasian) and she goes through in detail that the relevant elements under consideration are in fact clitics, and moreover that they clearly appear word internally. Consider for instance the following two examples. In (1a), the clitic *ne*, expressing 3rd singular agreement with the subject, appears internal to the monomorphemic verb *bey* ‘look’, causing the verb to be discontinuous (as indicated by subscripting on the gloss). In (1b), the clitic *q’un*, this time expressing 3rd plural agreement with the subject, appears in a complex verb construction. Here, it lies in between the light verb *b* ‘do’ and the noun *laš’k’o* ‘wedding,’ which is incorporated by the light verb (in the examples, and henceforth in the paper, the clitic will be indicated by underlining):

^{**} For various discussion and comments in putting this work together, I thank Jonathan Bobaljik, Alice Harris, Beata Moskal, Susi Wurmbrand and the audience of BLS 39. All errors are of course mine alone.

- (1) a. pasčay-un γar-en gölö be-ne-γ-sa met'a-laxo
king-GEN boy-ERG much look₁-3SG-look₂-PRES this.GEN-on
'The prince looks at this for a long time.'
b. pasčay-on γar-muy-on lašk'o-q'un-b-esa
king-GEN boy-PL-ERG wedding-3PL-DO-PRES
'The king's son's married.'

Once we bite the bullet and recognize that Udi shows a clear instance of a clitic appearing inside another word, and further still another morpheme (though this in itself is not without objection - see Luís & Spencer 2006), then a host of questions remain to be explained. Firstly, how do we best capture this phenomenon? The obvious parallel to link endoclisism to is infixation (see Yu 2007 for a comprehensive overview of infixation), the well known instance of an affix appearing internal to a root instead of being either a prefix or a suffix. If it can be shown that endoclisism and infixation show the same patterns, then there is strong evidence that there is one mechanism that underlies each of them, and the handful of cases of endoclisism that we see in the literature and all future ones ought to be conflated under whatever mechanism underlies infixation. In fact, there already exists a number of analyses of infixation encompassing different theoretical frameworks, for instance Prince & Smolensky (1993), McCarthy & Prince (1995) in Optimality Theory (OT), Halle (2001) in Distributed Morphology (DM), Yu (2007) in Sign Based Morphology, etc. Supposing it were to be the case that endoclisism showed the same distribution as infixation, then pending counterexamples, we face no problem. However, if there are differences between the two that can't *a priori* be handled by independent properties of clitics vs affixes, for instance (non)-categorical selection, then we need to see how far these differences go.

The second question that needs to be addressed is: why is endoclisism so rare? Yu (2007) notes that infixation, even though it is attested in over 100 languages, is rare in comparison to prefixation and suffixation. Now, this may reflect functional pressures in keeping the integrity of morphemes intact (Anderson 2005), but it still remains the case that infixation is possible and allowed by Universal Grammar. Supposing that infixation and endoclisism come from the same mechanism, then we reasonably expect a comparable rate of endoclisism to arise. However, as will be discussed, there are strikingly few cases of endoclisism that we know of, and only a handful of clear ones. This may of course represent oversight, misanalysis or fuzzy cases being analyzed as something else, but the point remains that in comparison to infixation, the paucity of endoclitics worldwide should give us pause before admitting it as an operation of UG.

This paper attempts to answer both of these questions. I will propose that endoclisism is not a *direct* operation of UG, but can arise *indirectly* due to morphological readjustments. Much of the discussion centers around the most robust case of endoclisism that is present in the literature, that of Udi. I will show that the endo-

clitics that we see in this language are not endoclitics in any deep sense, but that they are forced into their surface position by the morphotactic requirements of Udi. This analysis opens the door for an answer to the second question, that of why endoclitics are so rarely found. The answer that will be offered is that it takes a confluence of factors for endocclisis to arise. The analysis of Udi is augmented by discussion of other cases of endocclisis that have been claimed in various places to exist.

2 Subject clitics in Udi

The clitics which will be of interest to us are subject marker clitics in Udi. These clitics mark the agreement features of the subject, which can be seen in (2) below:

- (2) a. q'ačay-γ-on bez tänginax bašq'al-q'un
thief-PL-ERG my money.DAT steal-FUTII-3PL
'Thieves will steal my money.'
- b. nana k'wa-ne
mother.ABS house.DAT-3SG
'Mother is at the house.'

Harris (2000, 2002) shows that the distribution of these clitics is extremely complex. Their presence in the sentence is obligatory, yet they appear in a wide variety of positions. Harris shows that these positions can be described by a system of seven hierarchically ranked rules, given below:

- (3) Rule 1: Subject clitics (SCs) are final in the Vx ¹ if the verb is in the future II, the subjunctive I, the subjunctive II, or the imperative.
Rule 2: SCs occur enclitic to a focused constituent.
Rule 3: In clauses with zero copulas, SCs are enclitic to predicate nominals.
Rule 4: SCs are endoclititic 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, SCs are endoclititic occurring between the verbstem and the present tense marker.
Rule 6: With verbs forms of category A and category B, SCs are enclitic to the entire verb form.
Rule 7: SCs are endoclititicized immediately before the final consonant in monomorphemic verbstems.

¹ Harris uses the notation Vx to mean the complex consisting of the verb and negative.

For reasons of space I do not wish to give a comprehensive review of the arguments that these elements are clitics as opposed to affixes. Harris goes through this in great detail in Harris (2000, 2002) and I refer the reader to these works for the full arguments. The most obvious of these that can be seen from the data used in this paper is that they have freedom of attachment to various syntactic categories, but Harris goes through a wealth of tests that show their clitic status.

If we look at the distribution of the clitics in the rules, we can see that there is an interesting alternation; in some rules the clitic appears as an enclitic and at other times it appears as an endoclititic. The cases that will be the major focus of this work will be rules 4 through 7. Here we can see that the clitic sometimes appears inside a word, and at other times it appears at the end. The concept of ‘word’ is traditionally quite hard to define (for some frameworks more than others), and I do not wish to commit to anything here, or offer any thoughts of my own. The important parts about the clitics in Udi is that even though cross-linguistically we see clitics at peripheral positions, in Udi, they are clearly non-peripheral. What I will take to be the thing of interest about Udi is that the clitics can appear at a non-peripheral position in a complex head. In what follows, I will loosely use the term ‘word’, to mean multiple elements that are spelled out together as part of a complex head.

We can see this in the case of Rule 4, where the clitic is positioned between an incorporated element, and the incorporating light verb. This gives the structure of complex verbs that we see in (4), and some real examples seen in (5):

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

- (5) a. äyel kala-ne-bak-e
 child.ABS big-3SG-BECOME-AORII
 ‘The child grew up.’
 b. nana-n tur-ex oc’-ne-k’-e
 mother-ERG foot-DAT wash-3SG-LV-AORII
 ‘Mother washed her foot.’

The other case of endocclisis that we see in Udi comes from cases where the clitic appears internal to a monomorphemic verb. These are the cases that Harris captures under her Rule 7. In these instances, the subject clitics move before the final consonant of the verb. This produces a discontinuous verbal root. In the glosses I follow Harris by indicating this with subscripting:

- (6) a. kayuz-ax a-z-q’-e
 letter-DAT receive₁-1SG-receive₂-AORII
 ‘I received the letter.’

- b. q'ačay-γ-on bez tänginax baš-q'un-q'-e
thief-PL-ERG my money.DAT steal₁-3PL-steal₂-AORII
'Thieves stole my money.'

Rule 6 refers to simplex verbs as well, and this rule can be seen as an exception to the general pattern of clitic placement with respect to simplex verbs, that of Rule 7. In these cases the clitic appears at the end of the verb, outside of the affix which expresses tense-aspect-mood (TAM):

- (7) a. b-esa-ne b. bi-esa-zu
make-PRES-3SG die-PRES-1SG
'She makes.' 'I am dying.'

If we compare the cases of simplex verbs, we can see that the difference between the two is the phonological shape of the verb root. If the root consists of a single consonant, or a CV syllable, then the clitic will not appear internal to the root, but outside the TAM suffix. If however, the clitic minimally consists of or ends in a closed syllable, then the clitic can appear inside the final consonant. Below, I will provide an account of this phonological sensitivity.

The final rules Harris gives consist of cases where the clitic will surface as an enclitic if certain elements are present in the sentence. Rule 1 states that wherever the verb is in the future II, subjunctive I, subjunctive II or imperative form, the clitic will appear at the end:

- (8) q'ačay-γ-on bez tänginax bašq'al-q'un
thief-PL-ERG my money.DAT steal-FUTII-3PL
'Thieves will steal my money.'

Where none of these forms are relevant, the clitic will be enclitic to a constituent that is in focus:

- (9) tāzā k'oǰ-q'un biq'-e išq'ar-muy-on
new house-3PL build-AORII man-PL-ERG
'The men build a new house.'

Finally, where there is a zero copula in the sentence, the clitic will appear enclitic to the predicate nominal:

- (10) nana k'wa-ne
mother.ABS house.DAT-3SG
'Mother is at the house.'

Postponing discussion of Rule 5 to footnote 3, once we look at the above distribution of the subject clitic, we can notice two things. Firstly, it is an extremely complex system of clitic placement, which in itself warrants attention, even without the fact that it apparently contains endoclititics. Secondly, we note that endoclititics only arise in the absence of any other rules, suggesting that if a clitic would have any kind of inherent specification, it would be to appear non-peripherally. The categories that can be seen as ‘attracting’ subject clitics - certain TAM suffixes, focus, predicate nominals - all then force the clitic to appear enclitic to the entire form.

Finally, we may need to recognize two separate types of endocclisis. Firstly, there are cases where the clitic appears word internally, but its position is inter-morphemic, in the sense that it appears in between individual morphemes. Secondly, there is also cases where we are dealing with intramorphemic placement of clitics, where the clitic is splitting up a single morpheme. As we will see in section 4, Udi is perhaps unique in the world’s language in having a case of intramorphemic clitics.

3 Two possible analyses

In this section I discuss two possible analyses of the Udi facts. Firstly I discuss representational approaches to the problem, such as Harris’ own, and Yu’s (2007) claim that his analysis of infixation is consistent with the Udi data. Secondly, I present my own analysis, which is derivational in nature and couched within Distributed Morphology (Halle & Marantz 1993). I show that this analysis is able to catch the same data, but crucially is able to capture all the cases of endocclisis together. This analysis leads us to further question the Udi data and how much of a reflection of UG it really is. This question will be further discussed in section 4.

3.1 A Representational Approach

Harris explains the position of the clitic through Optimality Theoretic (OT) alignment constraints. She proposes that in order to account for the full paradigm of clitic placement, we must appeal to different constraints that apply depending on what is in the sentence. Harris proposes the following constraint system, with the rankings below:

- (11) Align-PM-*al/a*²
 Align (PM,L,-*al/-a*,R)
 Read as: “align left edge of the person marker to right edge of -*al/-a*”

² Whilst Harris writes the constraint in terms of phonetic content, she does so for parsimony. *al/a* are the usual exponents of future II, subjunctive I, subjunctive II and imperative.

- (12) Align-PM-FocC
Align (PM,L,FocC,R)
- (13) Align-PM-IncE
Align (PM,L,IncE,R)
- (14) Align-PM-Verbstem
Align(PM,R,Verbstem,R)
- (15) Align-PM-*al/a* » Align-PM-FocC » Align-PM-IncE » Align-PM-Verbstem
(TAM) (focus) (complex verbs) (simplex verbs)

The above constraints work in parallel and ensure that the place of the clitic adheres to the descriptive rules given above. Since the constraint that refers to aligning the clitic to the Future II and other specific TAM suffixes is ranked highest among them all, this ensures that the output that best satisfies this will be picked, i.e. the left edge of the clitic will align with the right edge of the TAM suffix. Therefore, the clitic will be enclitic to the entire verb forms when they are in these conjugations. Similarly with when the clitic is attached to focus, this is because the higher ranked constraint, which pertains to the TAM suffixes, is irrelevant without them being in the sentence. So, the sentence where the clitic is placed on focus ‘wins’.

The constraint which is truly interesting for our current purposes is that given in (14). This is the one that places the clitic inside the verbal root (Harris labels it *verbstem*, but the terminology doesn't matter much). This constraint forces the right edge of the clitic to be aligned with the right edge of the verbal root. Now, there isn't any way to fully satisfy this constraint, so the minimal violation will be placing the clitic as close to the right edge as possible.

The constraints do indeed place the clitic correctly; however, they are unsightful in that they merely restate the descriptive rules and offer little in the way of explanation for why the clitic shows such a complicated distribution. Moreover, in taking the tack that she does, Harris essentially encodes into UG the possibility that a clitic is placed by the syntax inside another morpheme. Harris argues that it must be the case that clitic placement is syntactic, since the placement of the clitic is sensitive to the elements of the entire sentence, instead of simply being about which elements are in a particular word. It hardly needs pointing out that this is an extremely powerful device to allow into the toolkit of UG, and it is something that is impossible to model within many theories of natural language.

Yu (2007) offers a similar approach to Udi, saying that the facts are consistent with his theory of infixation, which proposes that affixes can subcategorize for phonological information, and that they can therefore be aligned such that the

right edge of the clitic is adjacent to the left edge of the final consonant. Yu's approach faces the same problems as Harris' does, in that it allows into UG the opportunity to place clitics directly inside morphemes. As we'll see in section 4 this is problematic since Udi appears to be the only exponent of this, questioning its status as a UG operation. Furthermore, Yu's approach can offer no more of an explanation the distribution of the clitic, since it relies on other constraints, which presumably must also restate the descriptive rules.

3.2 A Derivational Analysis

In contrast to the approach of Harris, I present here a derivational account. A derivational account of this data allows us to assume that the clitic is placed in a position that is different from where it surfaces and there is some intervening process that moves the clitic into the place where it surfaces. Before getting into the analysis, firstly consider what benefits this could have. Harris' approach entails that there is some operation of syntax that is able to place a clitic directly inside another morpheme. Such an assumption is an extremely powerful mechanism to allow into the toolkit of universal grammar. It may end up being correct to do this, but there is another option before we make this move. It could also be the case that the clitic is placed somewhere else by the syntax and then is moved into the root internal position by some surface readjustment. To the extent that this approach is tenable, we do not need to say that the syntax can ever place a clitic inside another morpheme. We'll see in the next section that this is a desirable result when we look at a wider typology of endoclititics, but for now consider the analysis itself.

My analysis proposes that the subject marker clitics in Udi are a special case of second position clitics, but instead of being second position within a phrasal domain, they are in fact second position within a complex head. The nature of second position is something that I wish to leave slightly open (for a concrete analysis I refer the reader to Smith in prep.) since this is an issue that goes well beyond what I have space to discuss here. My major claim here is that the distribution of Udi can be captured once we understand that Harris' rules 4 through 7 all derive from a single default placement position of the Udi clitic (16d below), and then the differences in where the clitic is realized come about through the morphotactic considerations of Udi forcing the clitic to surface somewhere else.

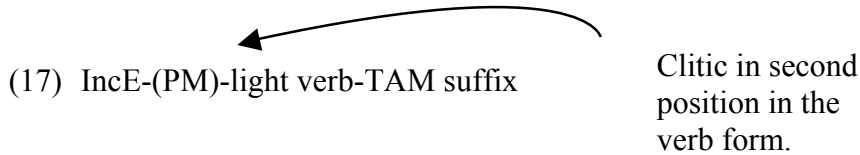
Specifically, I propose that we can simplify Harris' seven rules into the following four, with (16d) being the innovation here:

- (16) 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.

In addition to remaining non-committal to how ‘second position’ is to be captured, I also leave discussion of what it is that underlies (16a-c) to further research. What I want to focus on in this article is the cases where the clitic surfaces as an endoclititic. Since (16a-c) all deal exclusively with enclitics, they are irrelevant to the discussion at hand of how endocclisis as a general phenomenon is to be analyzed.

(16d) is then presented here as the default rule of subject clitic placement in Udi; where no other category ‘attracts’ the clitic, the clitic will be positioned in second position inside the complex head that spells out the verb. This analysis yields an immediate benefit since it allows us to account for the cases of complex verbs where the clitic is transparently in second position within the verb. Recall that complex verbs in Udi consist of the schema given in (17), repeated from (4) above:



When the clitic is hosted by complex verbs it appears in second position, as we might expect given the default rule of placement proposed here. We also account for the position of the clitic in the intransitive forms of the transitivity alternations that Harris accounted for under Rule 5. As discussed above, they are exactly the same as the complex verb cases.

- (18) äyel kala-ne-bak-e
 child.ABSL big-3SG-BECOME-AORII
 ‘The child grew up.’

However, in the cases of simplex verbs, second position placement seems to make entirely the wrong predictions. Recall from the discussion in section 2 that when the clitic is to be hosted by a simplex verb, there are two possible locations. If the verb consists of a single consonant, or is an open syllable, then the clitic appears at the end of the entire verb form, outside the TAM suffix. If the verb is a closed syllable, then the clitic will appear internal to the root itself, creating a discontinuous root, with the clitic located before the final consonant of the verb. It is clear that neither of these positions is second position within the verb form, as rule (17d) requires. If (17d) applies as is, then we would expect that the clitics would appear between the monomorphemic verb root and the TAM suffix, contrary to fact:

Predicted:

- (19) *q'ačay-γ-on bez tänginax bašq'-q'un-e
thief-PL-ERG my money.DAT steal-3PL-AORII

Found:

- (20) q'ačay-γ-on bez tänginax baš-q'un-q'-e
thief-PL-ERG my money.DAT steal₁-3PL-steal₂-AORII
'The thieves stole my money.'

(16d) however, since it is couched within a derivational approach, only refers to where the clitic is placed by the syntax, and says nothing about any other movements that may arise throughout the rest of the derivation. I propose that there is an extra movement of the clitic away from its original position, which causes the clitic to be moved into its surface position. In the case of simplex verbs which can host an endoclititic, then (so, ending in a closed syllable), this means that the clitic is placed between the verb and the TAM suffix and is then moved internal to the verb root. In the cases where the clitic cannot be hosted by the verb (C or CV roots), the clitic again gets positioned between the root and the TAM suffix, but this time gets moved rightward outside the TAM suffix.

Two things must be answered at this point. Firstly, why would the clitic be placed in one position and then moved to another? Secondly, what is the process that moves the clitic? The answer to the second must either be a process of the morphology or of phonology, since where the clitic moves to is sensitive to the phonological information of the root. It is a crucial assumption of DM that syntax makes reference only to abstract feature bundles, with the phonological exponents of lexical items only being inserted in a post-syntactic morphology module. It follows from this that the syntax cannot make reference to phonological information, as there is no phonological information in the derivation until it is inserted by the morphological component.

Returning to the question of why a clitic would be placed in one position and then moved to another, here I propose that there is a conflict between the placement rule of (16d) and the morphotactic rules of Udi. (16d) will place the clitic in between the root and the TAM suffix, however, if we take a wider look at Udi, we find that nothing ever intervenes in between the root and the TAM suffix. I take this to mean not only that nothing is allowed to intervene between the root and TAM in Udi, but that anything - crucially involving clitics - that is placed there will be moved to ensure that in the surface representation, the root and the TAM suffix will be adjacent. The following morphotactic rule is in effect in Udi.³

³ Something that might seem an exception to this rule are the cases described by Harris' rule 5. In these, it appears as though in the intransitive form of the relevant verbs the clitic lies between the root and the TAM suffix as in the following:

(21) *root-X-TAM suffix

I assume, following Arregi & Nevins (2012) who propose something in the same vein for Basque, that clitics can be moved by morphological metathesis from one position to another in the morphology. Note that the violation could be equally as well repaired by deletion of the clitic, since this would allow the verb root and the TAM suffixes to be adjacent. However, subject clitics are obligatory in Udi, showing that this particular repair strategy is never taken. To see how all of this works, consider how we arrive at the surface form in (22).

- (22) q'ačay-γ-on bez tänginax baš-q'un-q'-e
 thief-PL-ERG my money.DAT steal₁-3PL-steal₂-AORII
 'Thieves stole my money.'

What we are interested in is the form *baš-q'un-q'-e*, which spells out the verb and the clitic. Following DM assumptions, in the output of syntax into morphology, the verb consists of a root morpheme and the feature bundle expressing the TAM information, aoristII. Then, since none of the more specific rules apply, the clitic is placed in the second position of the complex head of the verb when all the elements undergo linearization, which happens to be between the root and TAM suffix. Left in this position, it will cause a violation of the morphotactic rules of Udi, however I do not assume that it is repaired at this point. Rather, I take evaluation of the violation to occur at the point the clitic undergoes vocabulary insertion (VI). Following Bobaljik (2000) and Embick (2010) a.o. I assume that VI proceeds from the root outwards, so spell-out of the verb can be seen as iterative application of VI.

Firstly, the phonological exponent of the root is inserted. It is only after this has taken place that the clitic comes to be spelled out. At this point, the grammar recognizes that leaving things as they are would yield an output that does not satisfy the morphotactics of verb formation and so enacts the repair. The chosen repair for Udi, as mentioned above, is morphological metathesis, and this moves the clitic the minimal position leftward to ensure that the right edge of the root will be adjacent to the left edge of the TAM suffix. Placing the clitic inside the final consonant allows this, so the form is able to converge satisfying the morphotactics of

(i) box-ne-sa
 boil-3sg-pres
 'it boils (intransitive)'

Harris however shows that the intransitive forms of these verbs are formed with a light verb 'GO', which is suppletively null in the present tense. The structure of the verb is then abstractly boil-3SG-GO-PRES. Thus, the clitic does not appear between the root and TAM. This shows that (21) is not a surface phonological constraint, but rather refers to the position of the morphemes, without reference to their phonological exponents.

Udi. This creates the discontinuous root *baš-q'un-q'*. Finally, the TAM suffix undergoes VI, and the final representation of *baš-q'un-q'-e* is derived.

- (23) i. linearization of elements: $\sqrt{\text{STEAL-}[\mathbf{3PL}]\text{-}[\text{PRES}]}$
 ii. VI of root: $/\text{bašq}'\text{-}[\mathbf{3PL}]\text{-}[\text{+PRES}]$
 iii. VI of clitic: $/\text{bašq}'\text{-}/\mathbf{q'un}\text{-}[\text{+PRES}]$
 iv. metathesis repair: $/\text{baš-q'un-q}'\text{-}[\text{+PRES}]$
 v. VI of TAM: $/\text{baš-q'un-q}'\text{-sa/}$

This explanation nicely explains why the clitic moves inside the verb root in Udi, and ends up breaking up the integrity of a simplex verb. However, as discussed there are also cases where the clitic does not appear internal to a simplex verb, as what I have just outlined would lead us to expect. In these cases, we find that the clitic moves to the end of the verb form. Harris explains these cases by proposing that only closed syllables can host clitics in Udi. A verb that is of shape CV or C is not of the correct shape to host a clitic, and therefore they appear at the end of the verb form. I follow Harris in assuming this to be the case, however I take the restriction to be a constraint on the repair that moves the clitic. It can only apply in cases where the destination of the clitic will be hosted inside a syllable. Put another way, (morphological) metathesis cannot move anything inside a (C)V syllable in Udi.

Why then does the clitic move outside of the TAM suffix? This finds a natural explanation under the approach given here. Instead of having to stipulate the word final position of the clitic, we can see it as an instance of rightwards metathesis. Leftward metathesis fails since the host is not large enough⁴, but if the clitics remains between the root and the TAM suffix the resulting form would violate the morphotactics of Udi. We expect then that the clitic would be metathesized rightward, which then allows the root and TAM to be adjacent. However, it still begs the question of why the clitic goes all the way to the end and why not inside the TAM suffix, like it does with the root. This is actually predicted within DM, since the repair is enacted at the point of VI of the clitic. As shown in the above derivation, this means that the phonological information of the root is in the derivation (recall the VI proceeds from the root outwards) when the clitic moves. However, since the clitic undergoes VI before the TAM suffix does, then the phonological information of the TAM suffix is simply not there. TAM at that stage remains a feature bundle. Therefore, when the clitic comes to be metathesized, and cannot go leftwards, the only remaining position for it to go is outside of the TAM suffix. This is shown in the following derivation:

⁴ Nor can the clitic go at the beginning of the word, since proclitics are not allowed in Udi.

- (24) bi-esa-zu
die-PRES-1SG
'I am dying.'
- (25) i. linearization of elements: $\sqrt{\text{DIE}}\text{-}[\text{1SG}]\text{-}[\text{+PRES}]$
ii. VI of root: /bi/-[1SG]-[+PRES]
iii. VI of clitic: /bi/-/zu/-[+PRES]
iv. metathesis repair: /bi/-[+PRES]-/zu/
v. VI of TAM: /bi-esa-zu/

4 Endocclisis in a Wider Context

As mentioned before, once we look at languages around the world, we find that endocclisis is a very rare phenomenon. One of the best known cases comes from so called 'mesocclisis' in European Portuguese (discussed in Anderson 2005), where the clitic appears between the verb root and agreement morphology. The relevant examples in (26a') and (26b') below:

- | | | | |
|---------|--------------------|-----|-------------------------|
| (26) a. | dariamos | a'. | dár-te-íamos |
| | give.1PL.COND | | give-1PL-1PL.COND |
| b. | perceberás | b'. | percerbér-me-ás |
| | understand.2SG.FUT | | understand-2SG-2.SG.FUT |

Whilst there is not space for me to provide an analysis of this language here, the data from European Portuguese show a clear case of intermorphemic placement. In (26a') for instance, the root *dár* is separated from the rest of its agreement morphology, which is spelled out by *-íamos*, by the clitic *te*.

Another case of endocclisis is found in Sorani Kurdish (Samvelian 2007, Bonami & Samvelian 2008, Walther 2012), a language which involves a similarly complex system of placing the relevant clitics as Udi. Again, for reasons of space I cannot delve too deeply into the language, but I will pick out three relevant remarks. Firstly, Sorani Kurdish has clitics which go into the second position within the word. This, recall, is what I am proposing is the case for Udi, only Udi is not as transparent as Sorani Kurdish, since the clitics are sometimes subject to further movement. Illustrative examples are given below, taken from Samvelian (2007), with the clitic representing the agreement features of the subject:

- (27) a. na-m-xwârd
NEG-1SG-eat.PAST
'I did not eat.'

- b. na-m-da-xwârd
NEG-1SG-PROG-eat.PAST
'I was not eating.'
- c. nard-man-in
send.PAST-1PL-3PL
'We sent them.'

The second thing of note about Sorani Kurdish is that independent morphotactic rules can force movement of the clitic. (28) shows an idiosyncratic rule of Sorani Kurdish which requires a 3sg clitic to occur after the object agreement, forcing the clitic to move to the end of the verb form, not in the usual second position:

- (28) nard-in-î
send.PAST-3PL-3SG
'He sent them.'

Finally, just like in Udi there is a morphotactic requirement that the root and TAM be together, in Sorani Kurdish the stem and past participle cannot be separated, forcing the clitic to surface ostensibly in 3rd position in the word:

- (29) a. nârd-uw-tân-in
send.PAST-PART-2PL-3PL
'you have sent them.'
- b. *nard-tan-uw-in
send.PAST-2PL-PART-3PL

Again, when looking at Sorani Kurdish, we see clear cases of endocclisis where the clitic is going inside a word. From these data, it seems clear that we need to recognize the existence of clitics which go inside a word, and allow for inter-morphemic placement of clitics within a complex head. Under DM assumptions, this is not too surprising, since the syntax is the input to word formation; words do not enter the syntax fully formed, but rather are built. This means that the syntax can in principle place clitics between morphemes that comprise a single word. More interesting is intramorphemic placement, which as Halle (2001) points out is impossible to model within DM, since this would mean moving some element inside the terminal node that comprises a feature bundle. It is difficult to even conceive how this would happen in the grammar.

The final example of endocclisis which I would like to bring up is that of Pashto. Pashto is interesting since it has been argued (Tegey 1977) to be a case of intramorphemic placement of a clitic, just like Udi. The discussion centers around the position of a subject clitic in the imperfective form of certain verbs. In Pashto, the position of the clitics is sensitive to stress, with clitics following the first

stressed constituent in the sentence. Yet, sentences can also be single words. In the usual case, clitics appear after the verb.

However, for certain verbs there is an optional stress shift, which moves stress from final or penultimate position, to the first vowel in the verb. Where the stress is final or penultimate, the clitic appears enclitic to the entire verb. However, after the optional stress shift the clitic can appear internal to the verb form:

- (30) a. axistólə me
buy 1SG
'I was buying them.'
- b. á-me-xistələ
buy₁-1SG-buy₂
'I was buying them.'
- (31) a. ayustə me
wear 1SG
'I was wearing it.'
- b. á-me-yustə
wear₁-1SG-wear₂
'I was wearing it.'

Looking at these data, it may seem as though we have another language like Udi, where the integrity of a single morpheme can be broken up by a clitic. However, the case is not quite as clear as the glosses make out. Tegey claims that *axistələ* and *ayustə* 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).

However, various authors have taken issue with this claim and argued that the *a-* is actually a bound prefix that carries no independent meaning. For instance, Kaisse (1981) argues precisely this, and claims that this allows us to explain both the fact that the clitic appears in this position, but also the fact that there are only 9 verbs in the language that begin with a vowel. This approach then brings Pashto in line with having a word-internal second position like Sorani Kurdish above. It also makes the above verbs consistent with other forms in Pashto, where clitics can follow a stressed prefix. Furthermore, it is strongly supported by verbs which begin with a consonant that show the same stress shift. Here, we find that the clitic remains at the end of the verb form, and doesn't break up the integrity of the morpheme:

- (32) a. pərebdə me
beat 1SG
'I was beating him.'
- b. pərebdə me
beat 1SG
'I was beating him.'

With Kaisse's analysis in mind (see also Roberts 1997 as well as discussion in Anderson 2005 and Yu 2007), we can then analyze the Pashto cases as follows, not with the discontinuous roots as seen before, but rather with a compound verb:

- (33) a. á-me-xistələ b. á-me-yustə
 PREFIX-1SG-buy PREFIX-1SG-wear

The point which I wish to highlight in this section is that not only is endocclisis a strikingly rare phenomenon in itself - which is true since in addition to the languages given here, there is really only a handful of other cases to my knowledge - but that the specific type of endocclisis exhibited by Udi, intramorphemic placement, is unattested in any other language that we know of. This may of course represent simply a gap in the typology, and the fact that we don't know of any more cases simply reflect the rarity of the process to begin with. However, on the basis of the data, and the fact that there is an equally viable analysis of Udi where clitic placement is intermorphemic, then we can make the stronger claim that intramorphemic placement of clitics is not a possibility of UG.

This claim has implications beyond simply arguing for a derivational versus a representational approach of the Udi data. It also tells us something quite deep about clitics that must be built into any theory. It is well known that the distinction between clitics and affixes is often fuzzy at best, and despite repeated attempts in the literature to find definitional criteria for one or the other (see for instance Zwicky & Pullum 1983), cliticization versus affixation is decided on gradient diagnostics, not absolute judgements. However, to the extent that the argument I put forward here is correct, then we can observe a real, clear difference between clitics and affixes; affixes apparently can break up the integrity of a morpheme but clitics cannot. There are various consequences to this difference, and many questions to be asked, but I leave this open for now to future research.

5 Conclusions

In this paper I have discussed the status of endocclisis within UG. I have argued that whilst we ought to recognize that clitics can appear in a position that is non-word peripheral, whenever they appear inside a word they overwhelmingly are positioned between morphemes. Udi presents a complication to this relatively simple picture, in that we seem to find clitics placed internal to another morpheme. This led Harris (2002) and Yu (2007) to create theories of cliticization which in essence allows for the recognition of the cliticization equivalent of infixation. However, I have proposed that this is too hasty. The move to allow UG to position clitics directly inside other morphemes is not warranted in two respects. Firstly, it is not necessary from the Udi data alone, which as I have shown is equally explainable on an account where the root internal position is an epiphe-

nomenon, arising from intermorphemic placement of the clitic and then movement inside the root to satisfy morphotactic constraints. Nor is the move warranted on cross-linguistic considerations. As pointed out, there are only a handful of cases that we know of where a clitic appears internal to a word, and Udi seems to be the only case that involves a clitic being inside another morpheme. Weakening our theories of UG to allow for clitics to be placed inside a morpheme without something else forcing it there then massively overgenerates the range of clitic patterns that we find around the world.

In contrast, on the approach here we expect intramorphemic placement of clitics to be rare, since it requires a confluence of factors to come together. Firstly, the clitic must be placed in a position inside a word, which in itself is rare. Secondly, it must violate some morphotactic requirement specifically in the position that it's placed. Finally, the repair for the language needs to be a displacement operation, instead of simply deleting the clitic altogether.

References

- Anderson, Stephen. 2005. *Aspects of The theory of Clitics*. Oxford studies in theoretical linguistics, Oxford University Press.
- Arregi, Karols. and Nevins, Andrew. 2012. *Morphotactics: Basque Auxiliaries and The Structure of Spellout*. Dordrecht, Springer.
- Bobaljik, Jonathan. 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, Olivier. & Samvelian, Pollet. 2008. *Sorani Kurdish Person Markers and The Typology of Agreement*. Paper presented at the 13th International Morphology Meeting, Vienna.
- Embick, David. 2010. *Localism Versus Globalism in Morphology and Phonology*. Cambridge, MA: MIT Press.
- Halle, Morris. 2001. Infixation Versus Onset Metathesis in Tagalog, Chamorro, and Toba Batak. In *Ken Hale: A life in Language*, Cambridge, MA: MIT Press.
- Halle, Morris. and Marantz, Alec. 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, Alice. 2000. Where in the Word is the Udi Clitic? *Language* 76, 593-616.
- Harris, Alice. 2002. *Endoclitics and The Origins of Udi Morphosyntax*. Oxford, Oxford University Press.
- Kaisse, Ellen. 1981. Separating Phonology from Syntax: A Reanalysis of Pashto Cliticization. *Journal of Linguistics* 17, 197-208.
- Luís, Ana. & Spencer, Andrew. 2006. Udi Clitics: A Generalized Paradigm Function Morphology Approach. *Essex Research Reports in Linguistics* 47, 47-59.

Peter W. Smith

- Prince, Alan. & Smolensky, Paul. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers Optimality Archive #537, <http://roa.rutgers.edu/files/537-0802/537-0802-PRINCE-0-0.PDF>.
- Roberts, Taylor. 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, Pollet. 2007. What Sorani Kurdish Absolute Prepositions Tell Us about Cliticization. *Proceedings of the Texas Linguistic Society IX*.
- Tegey, Habibullah. 1977. *The Grammar of Clitics: Evidence from Pashto and Other Languages*. Doctoral dissertation, University of Illinois at Urbana-Champaign.
- Walther, Geraldine. 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](http://www.linguist.univ-paris-diderot.fr/~gwalther/homepage/Publications_(fr)_files/MMM8_GWalther.pdf)
- Yu, Alan. 2007. *A Natural History of Infixation*. Oxford, Oxford Studies in Theoretical Linguistics.
- Zwicky, Arnold. & Pullum, Geoffrey. 1983. Cliticization vs Inflection: English n't. *Language* 59, 502-513.

Peter W. Smith
University of Connecticut
Department of Linguistics
365 Fairfield Way, Unit 1145
Storrs, CT 06269-1145

peter.w.smith@uconn.edu