On the Cross-Linguistic Rarity of Endoclisis[*]

PETER W. SMITH
University of Connecticut

1   Introduction

This paper discusses the phenomenon of endoclisis, and why it is so rarely found across languages. Endoclisis 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 *beɣ* '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 *lašk'o* 'wedding,' which is incorporated by the light verb:

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

   b. 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.'

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 endoclisis to is infixation (see Yu 2007 for a comprehensive overview of infixation), the well known instance of an affix ap-

---

pearing internal to a root instead of being either a prefix or a suffix. If it can be shown that endo-clisis 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 endoclisis that we see in the literature, and all future ones ought to be conflated under whatever mechanism underlies infixa-tion. In fact, there already exists a number of analyses of infixation encompassing different theo-retical frameworks, for instance Prince & Smolensky (993), McCarthy & Prince (1995) in Opti-mality Theory (OT), Halle (2001) in Distributed Morphology (DM), Yu (2007) in Sign Based Morphology etc. Supposing it were to be the case then the endoclisis showed the same distribu-tion as infixation, then pending counterexamples, we face no problem. However, if there are dif-ferences between the two, that can't *a priori* be handled by independent properties of clitics vs affixes (non)-categorial selection for instance, then we need to see how far these differences go.

The second question that needs to be addressed, is why is endoclisis so rare? Yu (2007) notes that infixation, even though it is attested in over 100 languages, is rare in comparison to prefixa-tion and suffixation. Now, this may reflect functional pressures in keeping the integrity of mor-phemes intact (Anderson 2005), but it still remains the case that infixation is possible and al-lowed by Universal Grammar. Supposing that infixation and endoclisis come from the same mechanism, then we reasonably expect a comparable rate of endoclisis to arise. However, as will be discussed, there are strikingly few cases of endoclisis 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 endoclisis 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 endoclisis that is present in the literature, that of Udi. I will show that the endoclitics 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 offered is that it takes a confluence of factors for endo-clisis arise. The analysis of Udi is augmented by discussion of other cases of endoclisis that have been claimed in various places to exist.

2   Subject Clitics in Udi

Harris (1984, 2000, 2002) outlines the distribution of subject marker clitics in Udi. These clitics mark the agreement features of the subject, and their presence is obligatory in the sentence. Har-ris shows that even though there is an extremely complex distribution of the clitics, this distribu-tion is describable by a system of seven descriptive rules. I list the rules first, in (2), before giv-ing examples and explanation after.

(2)   Rule 1:Subject clitics (SCs) are final in the Vx[1] if the verb is in the future II, the subjunc-tive I, the subjunctive II, or the imperative.

---

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

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 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, SCs are endoclitic 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 endocliticized immediately before the final consonant in monomorphemic verbstems.

These rules are ranked hierarchically, such that application of rule 1 bleeds application of rule 2 and so on. What this means is that whenever the verb is in the Future II, Subjunctive I, the Subjunctive II or Imperative conjugation, the subject marker clitic will appear enclitic to the entire verb form. This is shown below in (3):

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

     b. eγ-a-q'un?
         come-subjvI-3pl
         'Will they come?'

If none of these tense-aspect-mood (TAM) categories are present in the derivation then the clitic will attach to a focused constituent. In addition to general focus, Rule 2 also refers to negation and question constituents:

(4) a. nana-n   te-ne bụγa-b-e   p'ạ   ačik'alšey
         mother-erg neg-3sg   find-do-aorII   toy.absl
         'Mother did not find two toys.'

     b. täzä   k'oǰ-q'un   biq'-e išq'ar-muγ-on
         new   house-**3PL**   build-AORII   man-PL-ERG
         'The men build a new house.'

Next, Rule 3 applies to instances where there is a zero copula. Then, the predicate nominal is the clitic's host:

(5) nana   k'wa-ne
     mother.absl   house.dat-3sg
     'Mother is at the house.'

The truly interesting cases of Udi clitic placement come when the clitic is placed internal to a word itself. This is seen firstly in complex verbs. Udi complex verbs have the form in (6), where

a light verb incorporates a some element and the meaning of the verb is comes from the entire form. Rule 4 says that when none of the above rules are applicable, then the subject clitic will appear between the light verb and the incorporated element, as is shown in (7):

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

(7)   a.   äyel   kala-**ne**-bak-e
             child.ABSL 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 cases that are described by Rule 5 are where there is a transitive/intransitive alternation that appears at first glance to be marked by a shift in the placement of the clitic. In the intransitive form of the verb, the clitic appears internal to the verb root, (8a), whereas in the intransitive form the clitic appears between what appears to be the root and the TAM suffix, (8b):

(8)   a.   bo-ne-x-sa
             $boil_1$-3sg-$boil_2$-pres
             'he boils, cooks (transitive)'

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

    As noted by Harris however, this alternation is misleading, and is in fact representative of a wider pattern in the language. The intransitive cases should be conflated under Rule 4, since they are in fact complex verbs. Harris shows that changing the TAM conjugation of the verb form reveals the presence of a light verb GO, that is suppletively null in the present tense. Thus, when the light verb becomes visible, we see the clitics appearing regularly between the incorporated verb and the light verb, or as in the case of (9b) following the regular rules (not in (9b) that the TAM of the verb is FutureII, so the clitic appears at the end of the form):

(9)   a.   box-ne-c-e
             boil-3SG-GO-AORII
             'it boiled'

       b.   box-eγ-al-le
             boil-go-futII-3sg
             'it will boil.'

    The transitive form, where the clitic appears internal to the verb root are conflated under Rule 7, which we'll get to shortly. Harris' penultimate rule consists of exceptions to Rule 7. There are

some verbs which are simply exceptions, which cannot be predicted, and there are also cases of verb roots which Harris claims are phonologically too small to host a clitic. In these cases, the clitic appears enclitic to the entire verb form, outside the TAM suffix. There are two types of verbs that fall under this rule in addition to the true exceptions, and Harris argues that these are exceptions to Rule 7 simply because they *cannot* follow it. Since Rule 7 applies to monomorphemic verbs that end in a consonant (note that the clitic must go inside the final consonant in the definition of the rule), and places the clitic inside the root, if the root is either too small or does not contain a final consonant, then Rule 7 cannot apply. This is what we find, with verbs of phonological shape C and CV disallowing endoclisis and instead placing the clitic after the TAM suffix:

(10) a. b-esa-ne
make-pres-3sg
'She makes.'

b. bi-esa-zu
die-pres-1sg
'I am dying.'

Finally, when none of the other rules apply, and the verb is monomorphemic ending in a consonant, the clitic goes *internal* to the verb root, before the final consonant. As mentioned earlier, this causes the verbal root to be discontinuous:

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

b. q'ačaɣ-ɣ-on   bez   tänginax   baš-q'un-q'-e
thief-PL-ERG   my   money.DAT   steal$_1$-3PL-steal$_2$-AORII
'Thieves stole my money.'

If 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 endoclitics. Secondly, we note that endoclitics 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 word internally. 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 endoclisis. Firstly, there are cases where the clitic appears word internally, but its position is intermorphemic, 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 with 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 endoclisis 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:

(12) Align-PM-*al/a*[2]
    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)

(16)   Align-PM-*al/a* » Align-PM-FocC » Align-PM-IncE » Align-PM-Verbstem
= constraint for   (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 (15). This is the one that places the clitic inside the verbal root (Harris labels it verbstem, but the terminology

---

[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.

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 rules constraints do indeed place the clitic correctly, however they are uninsightful in that the 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 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 to 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, then 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 endoclitics, 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 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 (17d) being the innovation here:

(17) 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 is it is that underlies (17a-c) to further research. What I want to focus on in this article is the cases where the clitic surfaces as an endoclitic. Since (17a-c) all deal exclusively with enclitics, they are irrelevant to the discussion at hand of how endoclisis as a general phenomenon is to be analyzed.

(17d) 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 (18), repeated from (6) above:

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

When the clitic is hosted to 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.

(19) äyel   kala-ne-bak-e   Clitic in second
   child.ABSL   big-3SG-BECOME-AORII   position in the
   'The child grew up.'   verb form.

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. However, (17d), 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 endoclitic 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), then 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 (17d) and the morphotactic rules of Udi. (17d) 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 then is in effect in Udi:

(20) *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 (21). 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 when the clitic undergoes vocabulary insertion (VI). Following Bobaljik (2000, 2012) 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 an 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

(23) i. linearization of elements: √DIE-[1SG]-[+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   Endoclisis in a Wider Context

As mentioned before, once we look at languages around the world, we find that endoclisis is a very rare phenomenon. One of the best known cases comes from so called 'mesoclisis' in European Portuguese, (Roberts 1977, Anderson 2005), where the clitic appears between the verb root agreement morphology. The relevant examples in (24a') and (24b') below:

(24) a. daríamos
       give.1PL.COND

      a'. dár-te-íamos
         give-1PL-1PL.COND

    b. perceberás
      understand.2SG.FUT

      b'. percerbér-me-ás
        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 (22a') 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 endoclisis 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 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:

(25) 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. (26) 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:

(26) nard-in-î
      send.past-3pl-3sg
      'He send 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:

(27) a. nârd-uw-tân-in
         send.past-part-2pl-3pl
         'you have sent them.'

      b. *nard-tan-uw-in
         send.past-2pl-pp-3pl

Again, when looking at Sorani Kurdish, we see clear cases of endoclisis 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 intermorphemic 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 endoclisis 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:

(28) a. axistə́lə   me   b. á-me-xistələ
         buy   1SG   buy$_1$-I-buy$_2$
         'I was buying them.'   'I was buying them.'

(29) a. aɣustə́   me   b. á-me-ɣustə
         wear   1SG   wear₁-1SG-wear₂
         'I was wearing it.'   '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 *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)

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 <mark>verb.</mark> 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:

(28) a. pərebdə́ me   b. pə́rebdə   me
         beat   1SG   beat   1SG
         'I was beating him.'   '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:

(29) a. á-me-xistələ   b. á-me-ɣustə
         PREFIX-1SG-buy   PREFIX-1SG-wear

The point which I wish to highlight in this section is that not only is endoclisis 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 endoclisis 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 endoclisis 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 be 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 epiphenomenon, 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

Do references

Peter W. Smith
University of Connecticut
Department of Linguistics
Oak Hall (East SSHB)
365 Fairfield Way, Unit 1145
Storrs, CT 06269-1145

peter.w.smith@uconn.edu