



Endoclitics in Pashto

Noor Ahmad Aail¹, Mohammad Daud Mangal²

^{1,2} Assistant Professor, at Paktia University, Education faculty, Pashto language and Literature Department, Gardiz, Paktia, Afghanistan

Abstract

A cross-linguistically extremely rare kind of clitic, the endoclititic, happens in Pashto. Like infixes, endoclitics can be embedded within a word, in any case, by parting words separated into isolated nonadjacent pieces which themselves may not have any importance. Not at all like infixes, be that as it may, endoclitics are not affectations; their importance is irrelevant to that of their host word. This paper investigates about the portion of the issues endoclitics cause for preparing Pashto, both composed furthermore, spoken.

Keywords: cross-linguistically, infixes, Endoclitics, investigate

1. Introduction

This paper will investigate at the phenomenon of Pashto endoclitics, which are liable to both prosodic and syntactic constraints. These clitics challenge the perspective on prosody as being subsidiary from the grammar (for example Selkirk 1984) and the Principle of Lexical Integrity (Bresnan and Mchombo 1995) ^[7] in that Pashto permits clitics to be embedded into the morphological word. In any case, these difficulties can be settled by expecting an engineering that perspectives sentence structure and prosody as autonomous yet collaborating elements of language. Of seeing the prosodic projection as being founded on the syntactic tree, prosody and sentence structure are seen as equivalent accomplices attempting to line up with one another however much as could reasonably be expected. This paper presents information appearing that it is the prosodic segment that must be represented the position of the clitics WITHIN words in action word beginning Pashto sentences, which prompts the end that in instances of 'contradiction', the prosodic segment may overshadow the syntactic segment, despite the fact that this incorporates an infringement of the Guideline of Lexical Integrity. A formal record is given inside Lexical-Functional Grammar (LFG). Clitics have been characterized from multiple points of view, both phonologically and grammatically, frequently as semi-independent structures which join to phrases rather than words. The specialized subtleties of various definitions are not pertinent for this paper; here clitics can be portrayed just as a grammatical form somewhere close to joins and particles, appended to hosts like fastens, yet simultaneously autonomous words, similar to particles. An English model would be the possessive 's. Rather than appending to a thing alluding to the holder, it really appends at the part of the bargain holder thing phrase. For instance, the Queen of England's cap places 's at the part of the bargain, toward the part of the bargain thing Queen.

The two most normal sorts of clitic over dialects are enclitics, which connect toward the end their host (parallel to additions or postpositions), and proclitics, which join toward the start (parallel to prefixes or relational words). Pashto has a few proclitics, including! (PERF), "# (NEG),

\$(1), %& (2), and %! (3), however it is the following sort that is of enthusiasm here. The third kind is the endoclititic, which connects inside a word (like infixes). These don't just get embedded inside a word at a linguistic limit, in which case they would just be fastening, but instead they can part morphemes into separate pieces (called partials here). Some portion of a morpheme may wind up in one fractional while the rest of the morpheme may wind up in another, conceivably isolated by numerous different words. In etymological hypothesis they are commonly viewed as inconceivability, abusing lexical respectability (Kopris what's more, Davis, 2005) ^[9]. This hypothetical inconceivability may clarify why the main dialects asserted to have endoclitics are Pashto, Udi (Harris, 2002) and Degema (Kari, 2003) ^[1 & 2]. Rather than a hypothetical dialog of how endoclitics can exist by any means, the spotlight here will be on reasonable issues of experiencing them in Pashto, particularly in the composed language.

2. Materials and methods

Data collection

Data collection includes some Pashto online news from different sources such as Taand.com, Baheer.com, (<http://www.baheer.com/baheer/>) Ziar.net, Tolafghan (www.tolafghan.com/), and Benawa (www.benawa.com/) these are Pashtun sites which has different publications and constituents produced in-house for corpus building and linguistic analysis, totaling around 1.3 million words. These online sources are mixture from the dialects, while in-house materials are predominantly from the different parts of Afghanistan such from the Western (Kandahari, Southern) dialect, with the considerable volumes of Eastern (Jalalabad, Northern) and to the smaller extent of southern (Paktia, Central) dialects as well. Different dialects difference can affect the association and behavior of endoclitics in the ways beyond the possibility of this research. For example, for some speakers from Kandahar at least negative proclitic has some endoclititic properties.

3. Pashto Endoclitics

Pashto is an Iranian language spoken in Afghanistan and

parts of Pakistan. Clitics are very normal in this language and have been depicted altogether by Tegey (1977) [4]. This paper centers around one gathering of clitics that have extraordinary properties which challenge the basic comprehension of the cooperation of morphology, language structure and phonology. This gathering incorporates individual pronouns, modals and adverbials ((1)):

Pers-pronoun	Number	Modal	Translation	Adverbials	Translation
Me	1. Sg				
De	2. Sg				
Am/amo	3. Sg	Ba	Will, should,	Xo	Indeed,
Am/mo	1. Pl	De	may, should,	No	Really,
Ye	2. Pl		let		then.
	3. Pl				

These clitics carry on similarly as to their position. On the off chance that more than two clitics cooccur, they are put in a fixed layout. The requesting inside this format is of no importance to this paper and will in this manner not be examined. From the start, these clitics appear to be typical second position clitics: they are set after the principal expression of a sentence (clitics are underlined).

When multiple endoclitics occur, they follow a strict internal ordering (Tegey, 1977) [4]:

(2) angur ye rawr.

Grapes he brought

‘He brought grapes.’ (Tegey 1977, 138) [4].

There are although two different endoclitics with the form of *ye*, the one only might be happen at the time. The kind of the endoclitic has no behavior on its collation. The two adverbials are the reverse ends of the list, and the two models are interspersed between the pronominals.

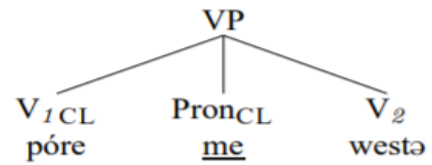
4. The Prosodic Constraints

In spite of the fact that there are two distinctive endoclitics with the shape #\$, just one may show up at once. The sort of the endoclitic doesn't matter to its requesting. The two adverbials are at far edges of the rundown, and the two models are blended among the pronominals:

The prosodically unstressed material before the verb word powers the clitic to show up at the correct edge of the expression, despite the fact that Pashto is a genuinely unbending action word last language. The significance of prosody moves toward becoming significantly progressively clear when seeing action word starting sentences. Like South-Asian dialects all in all, Pashto is a conflict dropping language (Butt 2007) [8]. Sentences can along these lines comprise of just an action word and a clitic. The endoclitics show up in these short sentences with regards to an angle caused pressure moving; the perfective part of the action word is shaped with the assistance of an action word inside pressure move. Concerning the stress move, Pashto action words fall generally into three classes, contingent upon their promise inside structure. Class 1 action words are monomorphic. In the imperfective, they bear weight on the last foot; the clitic is set after the action word ((6a)). In the perfective, class 1 action words take on a perfective prefix *w@-* that bears the principle stress. For this situation, the clitic happens after the prefix ((6b)):

(3) a. imperfective (4) perfective

Taxnwala me wa me taxnawala (wataxnawla me) Tickle I
 PREF I tickle I was tickling her (tagey, 1977:86) [4]. ‘I ticked (her) (tagey, 1977:92) [4].



Albeit by and large class 2 action words are seen as ONE unit, one could contend that the prefix may be a clitic what's more, has its very own terminal hub — much more so with class 3 action words, where the components of the action word development (=complex predicate) can be obviously distinguished as independent words. In the event that we see the clitic as a postlexical utilitarian component with its own syntactic hub, the information in (7) would therefore prompt a portrayal comparable to (8):

Aside from the three classes presented above, there is another gathering of action words that can have elective worry in the imperfective, yet structure the perfective with the perfective prefix of class 1 (*w@-*), in this manner embracing properties of each of the three classes. Inside this gathering, there are action words that start with consonants, which don't demonstrate any exceptional conduct in the imperfective: regardless of whether the pressure is on the front vowel, the clitic is set after the action word. In any case, there are nine action words in this gathering with an underlying vowel *a-*, which demonstrate a very unmistakable conduct with respect to the discretionary pressure move in the imperfective:

(5) a. imperfective—stress on second foot b. imperfective — stress on first foot
 Axistele me me xistele Buy I? I buy
 I was buying them I was buying them (Tegey 1977, 89) [4].

It has been contended that the *a-* was initially a prefix/clitic (Kaisse 1981, Anderson 2005) [5 & 13]. Yet this can't be affirmed for each of the an action words — besides, synchronically, the underlying *a-* never again has an unmistakable prefixfunction. Aside from class 2 and 3 action words, which can be examined as having two terminal hubs and in this manner permit the clitic to be embedded postlexically with a claim terminal hub (as has been illustrated in (8)), this turns out to be somewhat troublesome when managing the an action words or class 1 action words by and large, in light of the fact that the Guideline of Lexical Integrity would be damaged. Another alternative is consider the clitic as being produced in the vocabulary, as a piece of the morphological word itself, in this manner forestalling the infringement. Be that as it may, there is proof supporting the reality that the clitic is embedded into the action word POSTLEXICALLY. In the perfective, the an action words take the perfective prefix *w@-* like all different class 1 action words. As opposed to the consonant-introductory action words, nonetheless, perfective an action words show vowel mixture, a procedure that is expected to occur in the vocabulary. In (10a), the contiguousness of the perfective prefix *w@-* and the *a-* causes a combination: *w@-a-*. At the point when a clitic is embedded after the perfective prefix, the vowel mixture is as yet present ((10b)), giving proof that the clitic has really been embedded into the word postlexically

(5) a. Vowel coalescence — without clitic b. Vowel coalescence — with clitic

Ta ye waxla wa ye xla You it buy PREF it buy You buy it

(Tegey 1997, 149) [4]. You buy it (Tegey 1997, 163) [4].

6. Second Position

Pashto endoclitics like to be in second position in a sentence, with the admonition that "second position" might be characterized in different ways. There are four unique classes of action word that act contrastingly within the sight of endoclitics, particularly in recognizing imperfective and perfective structures: straightforward, subordinate, An underlying, and doubly unpredictable. Straightforward action words, for example, +, - 'beat' recognize perfective from imperfective structures by the expansion of the perfective proclitic, -. On the off chance that a non-endoclitic pronoun like ', '3sg' comes initially, the endoclitic (here '!') will pursue right away:

Table 1: I beat him

هغه مي ووايه			
هغه	مي	=و	وايه-ه
3sg	-beat	=PERF	1sg 3sg

Note that the action verb *ووايه* is touching, and parsing is clear. In the event that the basic pronoun *هغه* is evacuated, the endoclitic should in any case be in second position. To achieve this, it is embedded between the (stress-bearing) perfective proclitic *و* with the rest of the verb in the that clause.

Table 2: I beat him

و مي ووايه	
و مي ووايه-ه	
3sg -beat 1sg PERF	

Note that now the perfective marker *و* is no longer connected to the action word, in spite of the fact that the remainder of the action word is as yet adjacent, and effectively parsable. Assuming even the perfective marker *و* is expelled, coming about imperfective angle, the endoclitic will in any case be in second position. This time, the essential sentence structure rule that action words are last will be disregarded, and the endoclitic will be last.

Table 3: I was beating him.

وايه مي	
وايه ه مي	
1sg 3sg -beat	

In spite of the fact that the strange word request should be tended to, the action word is as yet adjacent and promptly parsable. Subsidiary action words (Tegey and Robson, 1996) [16] fuse a thing or modifier into an assistant in the imperfective, however split them separated in the perfective, making a sort of parting action word.

Table 4: I was making it worse

ما وړاناوه	
ما وړان-او-ه	
3sg -do -worse 1sg	

In the above table 4, the imperfective of *وړانول* 'make worse' incorporates the adjective *وړان* 'worse' into a shortened form of the auxiliary *كول* 'do' resulting in *وړاناوه*.

Table 5: I made it worse

ما وړان كړ	
ما وړان كړ	
do.PERF.3sg -worse 1sg	

In the perfective although, as in the above 5, the adjective *وړان* is separated and there is fully auxiliary *كړ*. Unlike the basic verbs, there is also no perfective *و*.

Table 6: I was making it worse, endoclitic

وړاناوه مي	
وړان او-ه مي	
1sg 3sg do worse	

Table 7: I made it worse, endoclitic

وړان مي كړ	
وړان مي كړ	
Do.PERF.3sg 1sg worse	

Table 8: I was buying them

ما خيستل	
ما خيستل	
Buy.3sg 1sg	

In the imperfective (table 6), the endoclitic takes second position after the verb (which integrates the adjective), violating basic word order, while in the perfective (table 7) the endoclitic seems after the non-incorporated adjective. In terms of parsing, derivative verbs pose no particular problems, as long as incorporation in the imperfective can be handled.

A-initial verbs (Tegey, 1977) [4] are also a type of splitting verb, but not in a semantically or morphologically natural manner. In the presence of an endoclitic, the initial / of these verbs can split off from the rest of the root. As with simple verbs, Ainitial verbs also take *و* in the perfective.

Table 9: I bought them.

ما واخيستل	
ما واخيستل	
Buy.3sg PERF 1sg	

Note in tables 9 and 10 that the imperfective and perfective forms are parallel to those of simple verbs. However, when an endoclitic is added, unexpected changes occur.

Table 10: I was buying them, endoclitic final

اخيستل مي	
اخيستل مي	
1sg buy.3sg	

Table 11: I was buying them, endoclitic medial

ا مي خيستل	
ا مي خيستل	
Buy?.3sg 1sg?	

Table 12: I bought them, endoclitic

وامي خيستل	
وامي خيستل	
Buy?.3sg 1sg ? PERF	

The underlining in tables 10 through 12 shows the focused

on syllable. In the imperfective, if the last syllable of the action word is focused on, the endoclititic accept second position after the action word (table 10). Notwithstanding, if the main syllable is focused on, the endoclititic again shows up after it, yet by constraining that syllable to isolate from the remainder of the action word (table 11). The underlying / which is anything but an important prefix, remains without anyone else. This causes issues for parsing, in that two trivial strings from various positions in the sentence must be distinguished as parts of one entirety. In the perfective (table 12), the marker ۛ pulls the ځو both structure another single introductory string, ځ. This draw even happens when an endoclititic can show up second without causing a split.

7. Segmentation and Tokenization

From one viewpoint, tokenization (discovering sentence and word limits) isn't influenced by the nearness of endoclititics. They typically are set off by blank area recorded as a hard copy, as are effectively recognized as individual strings, with the admonition that due to the idea of Pashto content in utilizing both associating and non-interfacing letters, endoclititics finishing in a non-associating ! might be composed without a space character, depending on the peruser to see the non-association as space. From another point of view, in any case, they are progressively troublesome in that they make issues for division (morphology, discovering roots) by the formation of non-word strings (partials). Utilizing lexical turn upward to decide whether a string is a word will fizzle on the grounds that the word partials made by endoclititic addition won't typically be in the dictionary, and those found in the dictionary will be homographs. The %& of table 16 is a homograph of a female name, and the) 1 of table 17 is a homograph of a word signifying 'a few'.

Basically applying morphology isn't compelling since a word is part into isolated words, rather than fastens being included. Fragmenting # \$ from table 16 may discover a substrings comparing to the second individual particular addition, #, yet the remaining \$ can't be utilized for finding the action word '%7& in the vocabulary (notwithstanding the morphology activities as of now required to perceive sporadic - \$%& as '%7&). Regarding the partials as a basic compound, as English blackbird, is additionally not compelling, since the partials have no significance to be intensified, in expansion to indistinguishable morphology issues from previously.

Another issue now and again shows up due to the nature of the Pashto composing framework. Since it is a variation of Arabic content, numerous vowels are unwritten, particularly word inside. At the parts of the bargains words, where postfixes for individual, number, sexual orientation, case, tense and viewpoint are discovered, endeavors are made to show generally unwritten vowels. Whenever an endoclititic parts a word, it is conceivable that a vowel which is unwritten in the entire word moves toward becoming composed toward the part of the bargain fractional. Think about these two variations of the doubly unpredictable (stress-moving) action word!"#\$% 'thump down', one with an endoclititic and one without:

Table 18: *I knocked them down*

without endoclititic	ما ځملول
with endoclititic	ځه مي ملول

Note that in the principle model there is no vowel shown between the consonants ځه and م. Be that as it may, in the second, where the endoclititic ۛ has split the action word in the two partial (after the pushed syllable), the main halfway presently finishes in the vowel letter ۛ. Regardless of whether treated as basic morphology or as exacerbating, the additional letter in the incomplete must be considered. Luckily, that the letter is normally (maybe consistently). Obviously, since this change just applies to composing, discourse acknowledgment would not have to address the "new" vowel. Portioning partials as unfound strings can be effective, insofar as there are strategies in the accompanying parsing stage to recoup the words that have been part.

8. How to Parse Them?

Expecting an agreeable phase of tokenization and division, one conceivable way to deal with parsing the action word partials coming about because of Pashto endoclititics is to treat them as irregular strings. Rejoining the partials while fixing potential spelling changes is direct, as long as the partials can be distinguished in that capacity. Area 5 recommended that morphology alone will fall flat, and that is on the grounds that it can't bargain with various word tokens one after another. Be that as it may, if partials can be recognized in that capacity, as opposed to state as obscure appropriate names, at that point there is the chance to assemble them back.

The issue at that point is the manner by which to recognize what to put together? Step by step instructions to realize that unfound strings are partials as opposed to different questions? The key is the event of at least one endoclititics. In the event that no endoclititics are found, than unfound strings can't be partials, and must be treated in the typical way (for example as appropriate names). On the off chance that endoclititics are found, at that point unfound strings can possibly be partials, particularly in the event that one of the unfounds is toward the end of the sentence and the other unfound is before the endoclititics. The probability is expanded if the unfound going before the endoclititic(s) is short, especially just a solitary syllable. Short perceived strings going before the endoclititics may likewise be in reality partials, with just homographs perceived, if the string in the action word position is unfound.

In the event that unfounds satisfy these necessities, at that point they can be tried as partials. In the event that there are two unfounds, they should be converged all together and after that tried with standard morphological handling, including testing both a string with all characters of the partials and a string with the last, of the primary halfway dropped. On the off chance that solitary the last word is unfound, at that point it should be tried a similar way, yet with a generally perceived string situated before the endoclititics as the potential first fractional.

Coming back to table 16, the string - . will be unfound, while the string "/will be perceived as a legitimate name, 'Bow'. Between them, the parser will perceive two endoclititics,)and (. The presence of the endoclititics and a last unfound string can at that point trigger the converging of that unfound - . with the short string "/going before the endoclititics (even despite the fact that officially perceived as a name). Applying morphology to the subsequent - . "/permits the unfound to be divided as an arched action word. On the off chance that transitivity data is incorporated into the lexical passage, the subsequent sentence will be linguistically stable

as the expulsion of the correct name will lessen the number of contentions to two, coordinating the transitivity of the action word.

Luckily for content examination, and sadly for discourse acknowledgment, making partials through the utilization of endoclititics is progressively normal in spoken than composed Pashto. Formal composed content has a low recurrence of partials, while discourse has a higher recurrence. Then again, discourse acknowledgment does not have to address spelling changes of specific partials, aside from to the extent that interpreting them legitimately.

Table 19: *I was buying them*

with endoclititic	امي خيستل
without endoclititic	ما اخیستل

The way that partials are less visit recorded as a hard copy implies that speakers can discover ways around utilizing them. This raises another plausibility, that of changing over sentences with partials into counterparts without. One path is to abstain from utilizing endoclititics, and the other is rethinking with the end goal that the action word isn't split into partials.

Table 20: *I was buying them*

with partials	امي خيستل
without partials	هغه مي اخیستل

In the above table 19, the 1sg endoclititic *مي* is replaced by the basic and simple 1sg pronoun *ما*. Because there is no endoclititic, there are no partials, and the verb be-comes contiguous. Where the first example has contiguous. Where the first example has two partials, / and *خیستل*, the second just has the complete verb *اخیستل*.

Table 20 alternately demonstrates a revamp due to the expansion of another pronoun, *هغه*, at first. This permits the endoclititic to show up in second position without expecting to part the action word. Despite the fact that these techniques are the ones probably utilized by speakers in abstaining from producing partials, endeavoring to utilize them in parsing existing sentences keeps running into a similar fundamental issue as in the past: how to recognize partials and consolidation them back together. Improvement or interchange decision of pronoun in a current sentence does not contact the partials in composed content, just the psyches of the speakers.

8. Conclusion

Endoclititics are cross-etymologically an exceedingly uncommon marvel, however they exist in Pashto and at the point when experienced should at present be parsed. Albeit no single explicit arrangement has been given in this paper, different useful methodologies have been introduced including perceiving unfound strings (particularly single syllables) in the nearness of endoclititics as potential partials, enabling them to be remerged for lexical query. As endoclititics exist on the limit of morphology and linguistic structure, the parsing of endoclititics must additionally include both morphology and sentence structure.

9. References

1. Alice C Harris. Endoclititics and the Origins of Udi Morphosyntax. Oxford University Press, Oxford, 2002.

2. Ethelbert Emmanuel Kari. Clitics in Degema: A Meeting Point of Phonology, Morphology and Syntax. Research Institute for Languages and Cultures of Asia and Africa, Tokyo, 2003.

3. Craig A Kopriss, Anthony R Davis. Endoclititics in Pashto: Implications for Lexical Integrity. Presented at the Fifth Mediterranean Morphology Meeting, Sept. 15-18, 2005, Fréjus, France, 2005.

4. Habibullah Tegey. The Grammar of Clitics: Evidence from Pashto and Other Languages. International Center for Pashto Studies, Kabul, 1977.

5. Anderson S. Aspects of the Theory of Clitics. Oxford University Press, 2005.

6. B`ogel T, Butt M, Kaplan R, King TH, Maxwell J. III. Prosodic Phonology in LFG A New Proposal. Proceedings of LFG09, 2009.

7. Bresnan J, Mchombo S. The Lexical Integrity Principle: Evidence from Bantu. Natural Language and Linguistic Theory. 1995; 13(2):181-254.

8. Butt M. The Role of Pronominal Suffixes in Punjabi. In Zaenen *et al.* (eds.), Architectures, Rules and Preferences. CSLI Publications, 2007, 341-368.

9. Dost A. A Domain-Based Approach to 2P Clitics in Pashto. Proceedings of the Texas Linguistics Society IX Conference, 2005.

10. Franks S, King TH. A Handbook of Slavic Clitics. Oxford University Press, 2000.

11. Halpern A. On the Placement and Morphology of Clitics. CSLI Publications, 1995.

12. Hock H. Who's on First? Toward a Prosodic Account of P2 Clitics. In A. Halpern and A. Zwicky (eds.), Approaching Second Second Position Clitics and Related Phenomena. CSLI Publications, 1996, 199-270.

13. Kaisse E. Separating Phonology from Syntax: A Reanalysis of Pashto Cliticization. Journal of Linguistics. 1981; 17:197-208.

14. Lahiri A, Plank F. What Linguistics Universals can be true of. In S. Scalise, E. Magni and A. Bisetto (eds.), Universals of Language Today. Springer, 2009, 31-58.

15. Roberts T. The Optimal Second Position in Pashto. In G. Booij and J. van de Weijer (eds.), Phonology in Progress Progress in Phonology. HIL Phonology Papers III. Holland Academic Graphics, 1997, 367-401.

16. Tegey H. The Grammar of Clitics: Evidence from Pashto and Other Languages. Ph.D. thesis, University of Illinois, 1977.

17. Van der Leeuw F. Clitics Prosodic Studies. Holland Academic Graphics, 1997.

18. Zec D, Inkelas S. Prosodically Constrained Syntax. In S. Inkelas and D. Zec (eds.), The phonology-syntax connection. CSLI publications, 1990, 365-378.