A Corpus-based Study of Endoclitic $=\hat{i}\hat{s}$ in Kurdish

Sina Ahmadi

Antonios Anastasopoulos Géraldine Walther

George Mason University Fairfax, VA, USA {sahmad46,antonis,gwalthe}@gmu.edu

Endoclitics and mesoclitics, clitics that appear within their hosts, are typologically rare phenomena found only in a few languages such as Udi (W. Smith, 2014), Pashto (Kopris, 2009), Degema (Kari, 2002) and Andi (Maisak, 2021). This paper follows Anderson (2005) and (Walther, 2012) in using the term endoclitic to mean any word-internal cliticization. Kurdish, an Indo-European language of >25M speakers, is known for displaying a number of morphological complexities related to endoclitics. While Sorani Kurdish endoclitic pronominal person markers have been extensively discussed in the literature (Walther, 2012; Bonami and Samvelian, 2008; Allahweisiazar et al., 2022), to the best of our knowledge, the endoclitic $=\hat{i}s$ 'also/even/too' and its variants have not been documented in any variants of Kurdish. Yet, Sorani Kurdish data shows that the distribution of the endoclitic $=\hat{i}\hat{s}$ is influenced by the complex placement properties of the language's enclitic pronominal person markers. This paper investigates the interdependent complexities of Sorani Kurdish endoclitic markers. Using a corpus-based approach, we compare the usage and distribution of complex Sorani Kurdish $=\hat{i}s$ to variants found in other Kurdish varieties which do not display the same complex person maker placement patterns (see Table 1): Kurmanji and Zazaki $j\hat{i}$ and $z\hat{i}$ (also $z\hat{i}$) appear as free morphemes; Gorani $=\hat{i}\varphi$ and Southern Kurdish $=\hat{i}\varphi$ appear as enclitics and follow oblique morphemes or precede pronominal suffixes, but do not interact with enclitic person markers (MacKenzie, 1966).

Language	Variant	Endoclitic
Kurdish	Kurmanji	$j\hat{i}$ (ڑی)
	Sorani	(ش) ۽= / (يش) =
	Southern	(ش) s- ((یش) - îş
Zaza-Gorani	Gorani (Hawrami)	چ) / -ç (چ) -îç
	Zazaki	$\overline{z\hat{i}}$ (zi)

Table 1: The $=\hat{i}\hat{s}$ endoclitic in variants of Kurdish and in Zaza-Gorani

Sorani Kurdish enclitic pronominal marker placement has been shown to be dependent on complex subject vs. object person/number properties and tense interactions (Bonami and Samvelian, 2008; Walther, 2012). Sorani Kurdish $=\hat{i}\hat{s}$ appears to be linked to those variable placements, as illustrated in Figure 1. Specific morphosyntactic features determine the insertion of the endoclitic, but realizations also differ between Northern Sorani (e.g., Mukri, examples 1 and 2) vs. Southern Sorani (e.g., Ardalani, examples 3 and 4). It is worth mentioning that we follow Matras (2019)'s classification of Sorani.

(1)	hat îş im. hat=îş-im. come=also-pst.1sg. '(I) also came'			(3)	$hatim \hat{ig}.$ $hat-im = \hat{ig}.$ come-pst.1sg=also. '(I) also came'
(2)	 <i>helîşhatim.</i> <i>hel=îş-hat-im.</i> up=also-flee-pst.1sg. '(I) also fled away' 			(4)	helhatim îş . hel-hat-im=îş. up-flee-pst.1sg=also. '(I) also fled away'
	bird im bird im in bird im in bird im in bird is im	e e in	ewe	ewe	past stem of BIRDIN (to take, to get) I got I got them I got them to/with I got them to/with again I got them also to/with again

 $\begin{array}{c}
 0 \\
 1 \\
 2 \\
 3 \\
 4 \\
 5 \\
 6
 \end{array}$

8 heł

Figure 1: The placement of the endoclitic =is (in green boxes) and agent marker (in grey boxes) with respect to the base and each other in a verb form in Northern Sorani

ewe

e ewe

I did not get them also to/with again

I was not getting them also to/with again

I was not lifting them also to/with again

bird in

bird

in e ewe

im [

ne de bird in

im de

ne

im

ne

We collected data for Gorani, Sorani and Southern Kurdish and analyzed the presence of $=\hat{\imath}$ both as an enclitic and as an endoclitic within special verb forms. In Northern Sorani, our findings show that $=\hat{\imath}$ tends to attach to the leftmost morpheme within the verb itself regardless of the transitivity of the verb. This is akin to the behavior of pronominal endoclitics in past transitive verbs (W. Smith, 2014). On the other hand, in Southern Sorani, particularly Ardalani, it tends to appear at the end of the verb. This is similarly observed in Southern Kurdish and Gorani. In all the varieties, the placement of $=\hat{\imath}$ within a phrase varies based on the intended meaning or emphasis on a specific word. But based on extensive corpus analysis, we also show how the frequency of $=\hat{\imath}$ is affected by the complex patterns of pronominal endoclitics and by more or less straightforward interactions between the personal markers and $=\hat{\imath}$.

This work documents a new and rare phenomenon and analyzes it using a data-driven corpus approach. The collected data will also be useful for further analyses in the field.

References

- Allahweisiazar, G., Gholami, V., Mohammadi Bolbanabad, S., and Varzandeh, O. (2022). Mukri Pronominal Enclitics distribution in VP: An Optimality Theoretic. Zabanpazhuhi (Journal of Language Research), 14(45).
- Anderson, S. R. (2005). Aspects of the Theory of Clitics. Number 11. Oxford University Press on Demand.
- Bonami, O. and Samvelian, P. (2008). Sorani Kurdish person markers and the typology of agreement. In 13th International Morphology Meeting.
- Kari, E. E. (2002). On endoclitics: Some facts from Degema. Journal of Asian and African Studies, 63:37–53.
- Kopris, C. (2009). Endoclitics in pashto: Can they really do that. In Third Workshop on Computational Approaches to Arabic Script-based Languages. Machine Translation Summit XII, Ottawa, Ontario, Canada.
- MacKenzie, D. N. (1966). The Dialect of Awroman (Hawraman-i Luhon): Grammatical Sketch, Texts, and Vocabulary. E. Munksgaard.
- Maisak, T. (2021). Endoclitics in Andi. Folia Linguistica, 55(1):1–34.
- Matras, Y. (2019). Revisiting Kurdish dialect geography: findings from the Manchester Database. *Current issues in Kurdish linguistics*, 1:225.
- W. Smith, P. (2014). Non-peripheral cliticization and second position in Udi and Sorani Kurdish. In *Natural Language and Linguistic Theory*. (Date accessed: 12.05.2020).
- Walther, G. (2012). Fitting into morphological structure: accounting for Sorani Kurdish endoclitics. In *Mediterranean Morphology Meetings*, volume 8, pages 299–321.