



# Explaining the Origins of the Morphosyntax of Aluan and Udi

Alice C. Harris  
UMass Amherst

# Introduction

- Aluan (Caucasian Albanian) is an ancient language of the Lezgian branch of the Nakh-Daghestanian family.
- Aluan was the direct ancestor of modern Udi or its very close relative.
- We know Aluan from palimpsest texts dating from between 7<sup>th</sup> and 10<sup>th</sup> c. CE.
- The texts were found in 1975 and deciphered, prepared, and published in Gippert et al. 2008.



# Aluan

- Aluan was an important language, spoken by thousands over a large area.

# Aluan

- Aluan was an important language, spoken by thousands over a large area.
- It had its own writing system, its own orthodox church, and its own patriarch.

# Aluan

- Aluan was an important language, spoken by thousands over a large area.
- It had its own writing system, its own orthodox church, and its own patriarch.
- Udi is a small language spoken in one village in Georgia and two in Azerbaijan.

# Why would we want to know about Aluan?

- Udi has sets of clitics, PMs, indicating person and number. These are placed by a set of rules so complex it needs some historical explanation.

# Why would we want to know about Aluan?

- Udi has sets of clitics, PMs, indicating person and number. These are placed by a set of rules so complex it needs some historical explanation.
- Aluan has what is clearly an earlier version of the same clitics, plus an additional set of clitics indicating person, number, and gender.



# Why would we want to know about Aluan?

- Udi has sets of clitics, PMs, indicating person and number. These are placed by a set of rules so complex it needs some historical explanation.
- Aluan has what is clearly an earlier version of the same clitics, plus an additional set of clitics indicating person, number, and gender.
- In this paper I will focus on the origin of the Aluan system. How would a language acquire a system so complex?

# Hypothesis (in general terms)

- Aluan first made person markers (PMs) enclitic in the verb.

# Hypothesis (in general terms)

- Aluan first made person markers (PMs) enclitic to the verb.
- Later PMs came to attach to the stressed syllable, and stress was based in part on information structure, with focused elements stressed.

# Hypothesis (in general terms)

- Aluan first made person markers (PMs) enclitic in the verb.
- Later PMs came to attach to the stressed syllable, and stress was based in part on information structure, with focused elements stressed.
- Perhaps even before attested Aluan, the system was grammaticalized, leading to the complex system that remains today.

# Organization of the presentation

- Description of the background grammar and clitic systems in Aluan (part 2)
- Description of the *Sprachbund* of which Aluan was a part (part 3)
- Development of the placement of clitics (part 4)
- *Sprachbund* again (part 5)
- Summary of what led to this strange placement (part 6)

Part 2

Background on Aluan grammar



# Aluan grammar

- Flexible SOV word order



# Aluan grammar

- Flexible SOV word order
- Ergative case marking





# Aluan grammar

- Flexible SOV word order
- Ergative case marking
- Differential argument marking (DAM) of direct object



# Aluan grammar

- Flexible SOV word order
- Ergative case marking
- Differential argument marking (DAM) of direct object
- Simplex and complex verbs, the latter consisting of an incorporated element (IncE) and a light verb (LV)

# Placement of Aluan person markers (PMs)

1. If the verb stands in the future II, the present, the imperfect, or the imperative, the PM must be enclitic to the verb.

# Placement of Aluan person markers (PMs)

1. If the verb stands in the future II, the present, the imperfect, or the imperative, the PM must be enclitic to the verb.
2. a. Agreement clitics are enclitic to a clausal negator.  
b. Agreement clitics are enclitic to a question word.  
c. Agreement clitics may be enclitic to a focused constituent.



# Placement of Aluan person markers (PMs)

1. If the verb stands in the future II, the present, the imperfect, or the imperative, the PM must be enclitic to the verb.
2.
  - a. Agreement clitics are enclitic to a clausal negator.
  - b. Agreement clitics are enclitic to a question word.
  - c. Agreement clitics may be enclitic to a focused constituent.
4. PMs occur between the IncE (noun, adjective, adverb, simplex verbstem, borrowed verb, unidentified element, or locative preverb) and the light verb. (Based on Harris 2021)



# Aluan Rule 1

If the verb stands in the future II, the present, the imperfect, or the imperative, the PM must be enclitic to the verb.

(1) aḵa-za (John 9: 11)

aḵa=za

see.PRS=I.DAT

‘(and now) I see’



## Aluan Rule 2a

Agreement clitics are enclitic to a clausal negator.

- (2)    te-**za**-aḱē                                    sa (Galatians 1: 19)  
         te=**za**        aḱē                                    sa  
         NEG=**I.DAT** see.PST.PST                    one  
         ‘I saw no (other) one’

There are a few exceptions to this rule in Aluan.

# Interaction between Rules 1 and 2a

(3) te-n-aa-**v**<sup>ʼ</sup>a (Matthew 22: 29)

NEG=NEG=know.PRS=**you**.PL.DAT

‘you do not know’



## Aluan Rule 2b

Agreement clitics are enclitic to a question word.

- (4) *ha-šow-ne hăya-hē* *ihesownax* *beši*  
hašu=**ne** hăy-a-h-e-y *ihesun-ax* *beši*  
who=**3** believe-TV-BE-TV.PST-PST hearing-DATIII us.GEN  
'who has believed hearsay of us?' (John 12: 38)

There are many exceptions to this rule.

## Aluan Rule 2c

Agreement clitics may be enclitic to a focused constituent.

(5) *ćowdowaxoṣṭayzow aci-bāhē-ḵa* (J 6: 42)

*ćowdowaxoṣṭay=zu aci-bāhe-y=ḵa*

heaven.from=I down-go.PST-PST=QUOT

“I have come down FROM HEAVEN”

It is difficult to determine how consistently this rule is followed.

## Aluan Rule 4

PMs occur between the IncE ...and the light verb.

(6) Aḵowḵ-ḡa-v<sup>ʿ</sup>a-hē (Matthew 24: 30)

Aḵuḵ=ḡa=v<sup>ʿ</sup>a=he-y

appearance=SUBJ=2PL.DAT=be-PST

‘Be it known to you’

aḵowḵ-ihesown ‘be revealed, become manifest’

=ḡa forms an optative or future

There are many exceptions to this rule.

The default position is enclitic to the verb

- (7) hükel-biyay-**ne**                      ta-båhesown  
     hükel-biya-y=**ne**                    ta-båhesown  
     memory-DO.PST-PST=**3**        exodus  
     '[he] mentioned (remembered) the exodus'  
     (Hebrews 11: 22)

# What needs to be explained

The rules for placement of PMs depend on three different kinds of characteristics of clauses:

- the TAM category
- the information structure, specifically focus, and the stress
- the internal structure of the verb.

It seems strange that a language would develop such heterogeneous rules for a common purpose.

# Part 3

*Sprachbund*



# *Sprachbund*

## Core languages

- Aluan
- North-Eastern Neo-Aramaic
- certain “aberrant Armenian dialects”
- some of the Tatic (NW Iranian) languages, including Northern Tati and some varieties of Talyshi

(Noorlander & Stilo 2015, Stilo & Noorlander 2015)

# Some shared characteristics of the *Sprachbund*

- development of new TAM forms on the basis of non-finite forms
- use of past converter
- complex vs. simplex verbs
- development of new agreement markers based on independent pronouns
- placement of clitics (Noorlander & Stilo 2015, Stilo & Noorlander 2015)





# Development of new TAM forms

- (a) NON-FINITE FORM ON IMPERFECTIVE STEM + AUXILIARY VERB *\*-a* 'be' IMPERFECTIVE
- (b) NON-FINITE FORM ON PERFECTIVE STEM + AUXILIARY VERB *-y* (< *\*-* *(i)hē*) 'be' PERFECTIVE
- based on reconstruction and studies by Alekseev (1985) and Maisak (2008)
- *ihesown* 'be' in Aluan has the present (imperfective) *aha* and the past (aorist) *(i)hē*.

# Use of past converter

Present stem		Aorist (past) stem	
Present $-a$	$he\dot{q}-a=zu$ 'I take'	Aorist $-e/a$	$he\dot{q}-a-y=zu$ 'I took'
Imperfect $-a=h\bar{e}$	$he\dot{q}-a=h\bar{e}=zu$ 'I was taking'	Perfect $-e/a=h\bar{e}$	$he\dot{q}-a-y=h\bar{e}=zu$ 'I had taken'

# Complex vs. simplex verbs

- Complex verb: incorporated element + light verb (+TAM, clitics)

e.g. hükel-biya-y=ne                      hükel-biyesown 'to remember'

memory-DO.PST-PST=3

's/he remembered' (Hebrews 11: 22)

- Simplex verb: root (+TAM, clitics)

e.g. zow iğa=zow                      iğesown 'to go'

I go.PRS=1

'I go' (John 8: 21)

# Development of new agreement markers

- This kind of agreement is not characteristic of N-D languages.
- Some may take issue with my referring to clitics as agreement.
- At least first and second person PMs developed from pronouns.

	Singular		Plural	
	Independent pronoun	PM	Independent pronoun	PM
1 <sup>st</sup>	zow	=zow	žan	=žan
2 <sup>nd</sup>	vown	=vown/nown	v <sup>ɕ</sup> an	=v <sup>ɕ</sup> an/nan

# Placement of clitics

- The topic of this presentation.

# Part 4

## The placement of clitics

# General hypothesis (repeated)

- (Pre-)Aluan first made new person markers (PMs) enclitic in the verb.
- Later, PMs came to attach to the stressed syllable, and stress was based in part on information structure, with focused elements stressed.
- Perhaps even before attested Aluan, the system was grammaticalized, leading to the complex system that remains today.

## Evidence from Aluan and Udi

- Stress is not marked for Aluan in Gippert et al. 2008.



# Evidence from Aluan and Udi

- Stress is not marked for Aluan in Gippert et al. 2008.
- In Udi, clitics immediately follow stress in most but not all situations (Harris 2002: 138-143).



# Hypothesis (elaboration 1)

Negatives (*te* and *ma*) bore sentence stress, as did question words and other focused constituents in Aluan.

# Support for this hypothesis

This is still true in Udi.

(8) té=**ne**      ta-d-e                      (Dirr 1928: 60: 10)

NEG=3SG   give-LV-AORl

‘he did not give’

# Support for this hypothesis

This is still true in Udi.

(8) té=**ne** ta-d-e (Dirr 1928: 60: 10)

NEG=3SG give-LV-AORl

‘he did not give’

(9) ek’á=**qun** b-e (Dirr 1928: 60: 15)

what=3PL do-AORl

‘what they had done’



## Support for hypothesis (elaboration 1)

- (9) íčux=**ne**    ɣač'-p-i                      (Dirr 1928: 62: 17)  
self=**3sg**   tie-LV-AORII  
'he tied himself on'

Thus, Udi supports the hypothesis that (Pre-)Aluan put stress on negatives, question words, and other focused constituents.

## Hypothesis (elaboration 2)

- The auxiliary used with the present (imperfective) stem in (Pre-)Aluan was stressed; that used with the past (perfective) stem was not.

## Hypothesis (elaboration 2)

- The auxiliary used with the present (imperfective) stem in (Pre-)Aluan was stressed; that used with the past (perfective) stem was not.
- This hypothesis explains why the PM is always enclitic to the verb in the present, imperfect, imperative, and future II, but not in other TAM categories.

## Support for this hypothesis (elaboration 2)

(11) kéf-b-a-nan (Dirr 1928: 60: 3)

enjoy-DO-IMPER-**2PL**

‘enjoy yourselves!’

(12) bes-b-ál-le (Dirr 1928: 61: 25, 61:37)

kill-DO-FUTII-3SG

‘he will kill’

Stress in future II is a relic of earlier word-final stress in these four TAMs. Stress in other imperfective TAMs assimilated to that in perfective TAMs – on the IncE.



- Stress on the final syllable in those TAMs based on the present (imperfect) stem but not on that of TAMs based in the past (perfect) stem explains the difference between Rule 1 and all other rules.

- Stress on the final syllable in those TAMs based on the present (imperfect) stem but not on that of TAMs based in the past (perfect) stem explains the difference between Rule 1 and all other rules.
- Stress on negatives, question words, and other focused elements explains Rules 2a, 2b, and 2c, respectively.

- Stress on the final syllable in those TAMs based on the present (imperfect) stem but not on that of TAMs based in the past (perfect) stem explains the difference between Rule 1 and all other rules.
- Stress on negatives, question words, and other focused elements explains Rules 2a, 2b, and 2c, respectively.
- What explains the advent of Rule 4, which places the PM between the incorporated element (IncE) and the light verb (LV) in complex verbs?

# The advent of Rule 4

- Up to this point, Pre-Aluan, like most other languages, had respected the integrity of the word.

# The advent of Rule 4

- Up to this point, Pre-Aluan, like most other languages, had respected the integrity of the word.
- Something changed to lead the language to experiment with violating the integrity of the word by placing PMs enclitic to the stressed morpheme, the incorporated element.

# The advent of Rule 4

- Up to this point, Pre-Aluan, like most other languages, had respected the integrity of the word.
- Something changed to lead the language to experiment with violating the integrity of the word by placing PMs enclitic to the stressed morpheme, the incorporated element.
- In the Aluan of the palimpsest, this placement is limited to clauses containing the optative clitic =*q̣a*.

(13) aḵowḵ-ḡa-n-hē  
aḵuḵ-ḡa-n-h-e-y  
appear-**OPT-3**-BE-TV.PST-PST  
'will appear'

(Matthew 24: 30)

We can assume that the position was later generalized, since it is obligatory in Udi if Rules 1-3 do not apply.



We can assume that the position was later generalized, since it is obligatory in Udi if Rules 1-3 do not apply.

What could have led to the violation of the integrity of the word?

We can assume that the position was later generalized, since it is obligatory in Udi if Rules 1-3 do not apply.

What could have led to the violation of the integrity of the word?

- Internal: The optative clitic =*q̇a* plus the PM could have been trapped between the IncE and the light verb in the process of univerbation.

We can assume that the position was later generalized, since it is obligatory in Udi if Rules 1-3 do not apply.

What could have led to the violation of the integrity of the word?

- Internal: The optative clitic =*qa* plus the PM could have been trapped between the IncE and the light verb in the process of univerbation.
- External: Aluan could have been influenced by one of the other languages in its *Sprachbund*.



Evidence to support the internal explanation: =*qa*  
+ the PM might have been **trapped** during  
univerbation

There is general evidence that univerbation is not  
complete in Aluan.

- (14) {zow n}ow-**baa**-z-oowxow                      **b<sup>ɸ</sup>a<sub>ɿ</sub>x** (J 12: 47)  
zu        nu-**baa**-z-o-uxu                              **b<sup>ɸ</sup>a<sub>ɿ</sub>x**  
I.ERG    NEG-DO.PRS-1SG-3SG.M-DIR              **judgment**  
‘I do not judge him’

But I know of no specific evidence of trapping clitics.



Evidence to support the internal explanation:  
=*q̣a* + the PM were trapped in univerbation

There is general evidence that **tmesis** of many kinds was common in Aluan (Harris 2022).

(15) il'ow-al-hanayṭ'a-gāen-ḵe-zow-pē (II Cor 4: 13)

il'u-al-hanay-ṭ'a-gāen-ḵe-zu-pē

word-ADD-**which**-3N.SG.GEN-bec.-SUB-1SG-SAY.PST

'[I believe,] because of which I have spoken, too'

In an environment with so much tmesis, tmetict clitics seem bound to occur, attracted to the stressed IncE.



Evidence to support the external explanation:  
Other languages in the area had tmesis.

- Ancient Greek (Homeric) – long before attested Albanian
- Avestan
- Georgian



Some other languages in the *Sprachbund* place clitics between constituents of complex verbs

- In Gorani Hawramani Luhon, a dialect of Northeastern Neo-Aramaic (NENA), the PM can be enclitic to the IncE in a 2-word complex verb (Noorlander 2021: 77, quoting others).



Some other languages in the *Sprachbund* place clitics between constituents of complex verbs

- In Gorani Hawramani Luhon, a dialect of Northeastern Neo-Aramaic (NENA), the PM can be enclitic to the IncE in a 2-word complex verb (Noorlander 2021: 77, quoting others).
- In N. Talyshi, N. Tati, etc. Set 2 PMs, if not fronted, occur between a verb root and the associated enclitic auxiliary (Stilo 2008: 377).



# Section conclusion

- Stress on the final syllable in those TAMs based on the present (imperfect) stem but not on that of TAMs based in the past (perfect) stem explains the difference between Rule 1 and all other rules.
- Stress on negatives, question words, and other focused elements explains Rules 2a, 2b, and 2c, respectively.
- An incipient ability of the PMs to encliticize to the stressed morpheme of a complex verb led to the development of intermorphemic endocclisis.
- Endocclisis was doubtless influenced by tmesis and specifically endoclititic PMs in contact languages.

# Part 5

Broader implications of the *Sprachbund*



Some other languages of the *Sprachbund* place clitics before the verb under certain circumstances

- In some “aberrant” Armenian dialects, clitics can also be prefixes or suffixes, or, in one variety enclitic to the focused constituent (Bezrukov & Dolatian 2020) – Hamshen, Gyumri, and Akhalkalaki.
- N. Talyshi, N. Tati, and some other Iranian languages in this area have fronting of Set 1b and 2 PM enclitics to hosts outside the verb (Stilo 2008).
- In some NENA dialects, some clitics are “semi-mobile” (*op cit*, 94).

Some other languages have variable stress placement in the verb

- Pashto imperfective aspect stresses the final morpheme of the verb, while perfective stresses the first. Clitics follow the stress in the verbs (Tegey 1977: 85-86).
- In Persian, stress falls on different syllables in different TAM categories (Persian Online 2007).
- In Vafsi, a Tatic language of central Iran, stress falls on different syllables in different TAM categories (Soltani).

None of these Iranian languages is believed to have been in the *Sprachbund* or in contact with Aluan.

# Part 6

## Conclusions

# Conclusions

- (Pre-)Aluan made new person markers (PMs) enclitic in the verb.
- Later, PMs came to attach to the stressed syllable, and stress was based in part on information structure, with focused elements stressed.
- Negatives (*te* and *ma*) bore sentence stress, as did question words and other focused constituents in Aluan.
- The auxiliary used with the present (imperfective) stem in (Pre-) Aluan was stressed; that used with the past (perfective) stem was not.
- Perhaps even before attested Aluan, the system was grammaticalized, leading to the complex system that remains today.

# How this explains the facts

- Stress on the final syllable in those TAMs based on the present (imperfect) stem but not on that of TAMs based in the past (perfect) stem explains the difference between Rule 1 and all other rules.

# How this explains the facts

- Stress on the final syllable in those TAMs based on the present (imperfect) stem but not on that of TAMs based in the past (perfect) stem explains the difference between Rule 1 and all other rules.
- Stress on negatives, question words, and other focused elements explains Rules 2a, 2b, and 2c, respectively.



# How this explains the facts

- Stress on the final syllable in those TAMs based on the present (imperfect) stem but not on that of TAMs based in the past (perfect) stem explains the difference between Rule 1 and all other rules.
- Stress on negatives, question words, and other focused elements explains Rules 2a, 2b, and 2c, respectively.
- Stress on the IncE in complex verbs in perfective TAMs, together with a general acceptance of tmesis, led to clitics occurring between the IncE and the LV in complex verbs.

# *Sprachbund*

As argued by Noorlander & Stilo 2015, Stilo & Noorlander 2015, and others, the *Sprachbund* of which it was a member played an important role in the development of these characteristics in Aluan, in particular with respect to

- stress assignment in various TAMs
- attraction of clitics to stress
- assignment of stress to negatives and focus
- tmesis.



# Thank you!

[acharris@linguist.umass.edu](mailto:acharris@linguist.umass.edu)



# Selected References

Bezrukov, Nikita, and Hossep Dolatian. 2020. "Mobile Affixes Across Western Armenian: Conflicts Across Modules," *University of Pennsylvania Working Papers in Linguistics*: Vol. 26 : Iss. 1 , Article 6. Available at: <https://repository.upenn.edu/pwpl/vol26/iss1/6>

Gippert, Jost, Wolfgang Schulze, Zaza Aleksidze, Jean-Paul Mahé, eds. 2008. *The Caucasian Albanian Palimpsests of Mount Sinai*. (Monumenta Palaeographica Medii Aevi : Series Ibero et Caucasica SIBE 2.) Turnhout: Brépols. (in 2 volumes).

Harris, Alice C. 2002. *Endoclititics and the origins of Udi morphosyntax*. Oxford: Oxford University Press.

Majsak, Timur. 2008. Glagol'naja paradigma udinskogo jazyka (nidžskij dialekt). *Udinskij sbornik: grammatika, leksika, istorija jazyka*, ed. by M.E. Alekseev, T.A. Majsak, D.S. Ganenkov, Ju.A. Lander, 96-161. Moskva: Akademia.

Mohammadirad, Masoud. 2020. *Pronominal clitics in Western Iranian languages: Description, mapping, and typological implications*. Diss. Université Sorbonne Nouvelle Paris 3.

Noorlander, Paul M. and Donald Stilo. 2015. On the Convergence of Verbal Systems of Aramaic and its Neighbours. Part I: Present-Based Paradigms. *Neo-Aramaic and its Linguistic Context*, ed. by Geoffrey Khan and Lidia Napiorkowska, 426-52. Piscataway, NJ: Gorgias Press.

*Persian Online: Grammar & Resources*. 2007.

[https://sites.la.utexas.edu/persian\\_online\\_resources/](https://sites.la.utexas.edu/persian_online_resources/), University of Texas at Austin. Accessed 24 June 2022. Liberal Arts Instructional Technology Services.

Stilo, Donald. 2008. Two sets of mobile verbal person agreement markers in the Northern Talyshi language. *Aspects of Iranian linguistics*, ed. by Donald L. Stilo, Vida Samiiian, Simin Karimi, 363-390. Newcastle upon Tyne: Cambridge Scholars Publishing.

Stilo, Donald and Paul M. Noorlander. 2015. On the Convergence of Verbal Systems of Aramaic and its Neighbours. Part II: Past Paradigms Derived from Present Equivalents. *Neo-Aramaic and its Linguistic Context*, ed. by Geoffrey Khan and Lidia Napiorkowska, 453-84. Piscataway, NJ: Gorgias Press.