

The Malsia Madhe Dialect of Albanian

Lindon Dedvukaj (PhD Candidate* The Ohio State University)
dedvukaj.3@buckeyemail.osu.edu

I. Introduction

1. Malsia Madhe is a historically isolated region of Northern Gheg Albanian.
2. The geographic region of northern Gheg Albania is cited as a true refuge area, comparable to the Caucasus and western Himalayan reaches (see Coon 1950: 10).
3. Culturally Isolated: With the furniture, costumes, and technology that the Gheg Albanians of the mountains used, until the mid-20th century they were still living in the Hallstatt Iron Age, where no technological changes have occurred (Coon 1950: 40).
4. Malsia Madhe Albanian not accounted for in previous Albanian historical linguistics analyses.
5. This historic isolation has preserved ancient elements of the Albanian language.



Figure 1: Albanian dialectal distribution: Gheg Dialects RED, Tosk Dialects GREEN



Figure 2: Traditional Albanian Dialect Map with Malsia Madhe

II. Objectives

1. Malsia Madhe Albanian (MMA) is an unknown variant within Albanian language studies.
(Introduces new information that alters previous analysis)
2. MMA provides evidence for a better preservation of the Proto Albanian and Proto Indo-European plosives than the other Albanian dialects.
3. MMA can be used to more accurately determine Albanian's place in Indo-European which has remained an enigma.

III. The creation of Homorganic nasal-plosive clusters in Tosk Albanian

1. Previous Research:

- In *Sistemi I Numerimit Të Gjuhës Shqipe* (1997), B. Demiraj cites specific dialectal terms, which exhibit word initial variation such as: *prapa* ~ *mbrapa* and *përoj* ~ *mbroj*.

- B. Demiraj (1997: 105-6) also references the linguist Bopp who states that: word-initial nasal sequences with /m/ are a later homorganic addition.
- Newmark et. al. (1982: 116) describes verbal prefixes in Standard Tosk: the prefix *m-* surfaces before a labial consonant and *n-* elsewhere; the nasal prefixes cause the initial voiceless plosives of the underlying stem to voice to *-d-*, *-g-*, and *-gj-*.

2. Homorganic Nasal Assimilation

- A nasal sound such as the labial nasal [m] assimilates to another plosive at the same place of articulation [p, b], forming the nasal clusters *mp* or *mb*.
- This also occurs with the historical formation of the clusters *np*, and *nk*, assimilating to [mb] and [ŋg]. These undergo the chronological changes of *np* > *mp* > *mb* and *nk* > *ŋk* > *ŋg*.
- Three types of assimilation patterns have arisen in Standard Tosk:
 - Bilabials /p, b, m/ > [mb]
 - Alveolars /t, d, n/ > [nd]
 - Velars /k, g/ > [ŋg] <ng>

(a) Bilabials: PIE **nepōtṛiā* > *mbesa* ‘granddaughter, niece’

MMA/Tosk *nip* ‘grandson, nephew’

Tosk/Gheg *mbesa* ‘granddaughter, niece’.

MMA *nipja* ‘granddaughter, niece’ (feminine suffix *-ja*, see PIE **nepōt-ṛiā*).

- Feeding chronology: PIE **nepōtṛiā* > ***npōtia* (syncope of *-e-*)
***nepōtṛiā* > *mbesa* (np > *mp > mb)
- Other changes: **nepōtṛiā* > *mbesa* (vowel change **ō* > *-e-*)
**nepōtṛiā* > *mbesa* (spirantization of *t* > *s*)¹

(b) Alveolars: MMA *tesh* ‘touch’ ~ Tosk *ndesh*

Tosk *ndesh* ‘touch’ developed through the chronological change: **tesh* > **en-tesh* > *ndesh*.

- Cf. Ancient Greek *θίξις* [tʰiksɪs] ‘touch’ (see Beekes 2016: 549).
 Latin *tetigī* ‘to touch’ (see De Vaan 2016: 606).²

¹ This form of spirantization is also evident in the Old Gheg past participle *pasē* ‘had’ < **pot-to-* (see Fortson 2010: 449).

² Greek *ξ* [ks] surfaces in Albanian as *sh* [ʃ]. This is also evident with Homeric Greek *πτῶξ* [ptō:ks] ‘cower, hide’ > MMA *pshe* ‘hide, conceal’. (*πτῶξ* > ***pōsh* > ***pesh* > MMA *pshe* ~ Tosk *fshe*). Also note Latin *exemplum* ‘example’ > Albanian *shembull* (Latin *x* [ks] > Alb. *sh* [ʃ]).

(c) Velars: MMA *kha* ‘where, where from’ Tosk *nga*

- In Fortson (2010: 455), “... *nga* is shortened from **ën-ka* ‘where, from where’, with **ën* ‘in’ added to the old relative adverb *ka* ‘where, from where’ ...”
- **ën-ka* > Tosk *nga* [ŋga] (MMA *en* ‘in’, *kha* ‘where, where from’)
- PIE **h₁en* > MMA *en* ‘in’ (**ën* in Fortson 2010)
(Cf. Latin *en*, Greek *έν*)
Modern Tosk/Gheg *në*.³
- *en* + *ka* has yielded Tosk *nga* through loss of the initial unstressed vowel *ë*.⁴

3. Nasal Vowel Unpacking (Result of the loss of vowel length in Tosk Albanian)

Before Plosives

(a) Reflexive pronoun ‘yours’ MMA *tât* [t̃ɔ:t] ~ Tosk *tëndë* [təndə].

- In this case the earlier nasal vowel [̃ɔ:] has unpacked or separated into a vowel and a nasal consonant of the same place of articulation as the consonant it precedes: [̃ɔ:] > [ən]/_t = [ənd] (**tât* [t̃ɔ:t] > *tëndë* [təndə]).
- See Paradis and Prunet (2000) for nasal vowel unpacking.

Before Fricatives

(b) MMA *vâth* [ṽɔ:θ] ‘earing’, Tosk *vëth* [vəθ].

- Doric Greek *άávθα* ‘a kind of earing’ from Proto Albanian **awantha* (Witczak 2016: 40)

³In Tosk literature an isolated variant <inj> surfaces with an unclear meaning. B. Demiraj (1997: 207) cites Çabej and Meyer who equate <inj> to the preposition ‘in’ similar to Greek *έν* from **en*. There are three potential origins of the Modern Tosk/Gheg preposition *në* ‘in, into, to, towards’. It may be a metathesis shift from an earlier Albanian *en* ‘in, into, on, to, towards’ (preserved in MMA *en*). Second, it may be an effect of Slavic influence in the area. The preposition *në* is similar to the many variants of Slavic *na* ‘in, on, at’. See Mallory (1989: 76) for Old Church Slavonic *na*; Russian *na*; Serbo/Croatian *na*; Bulgarian *na*. The third and most compelling possibility is that it originated from the Italian articulated prepositions. The Albanian writers of the 16th-17th centuries were Catholic priests who in most cases were bilingual. The Italian masculine singular *nel* ‘in the’, is a contraction of the preposition *in* and the masculine singular article *el* (e.g., feminine singular *nella* < *in la*). Italian *in* < Latin *en* < Proto Italic **en* ‘in’ < PIE **h₁én* ‘in’). Albanian *në* may have developed from this paradigm.

⁴ Unstressed initial vowels delete in Albanian. Note Latin *arena* > Gheg *rana* ~ Tosk *rëra* (S. Demiraj 2006: 102). The same process developed through Old Albanian *én* + *ká* > **enká* > ***nká* > Tosk *ngá*.

(c) MMA *pês* [pǣ:s] ‘five’ Tosk *pës* [pəs]

- Proto Indo-European **penk^{we}* ‘five’

(d) MMA *mâz* [mǣ:z] ‘foal’, Tosk *mëz* [məz]

- Messapian (Illyrian) *Manzana* ‘horse deity’⁵

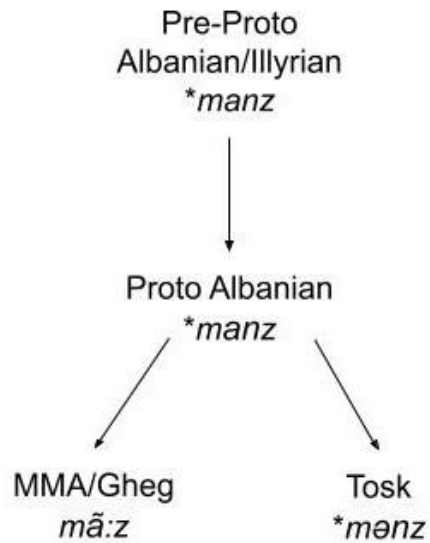


Figure 3: Expected outcome of Proto Albanian **manz* > Tosk Albanian **menz*

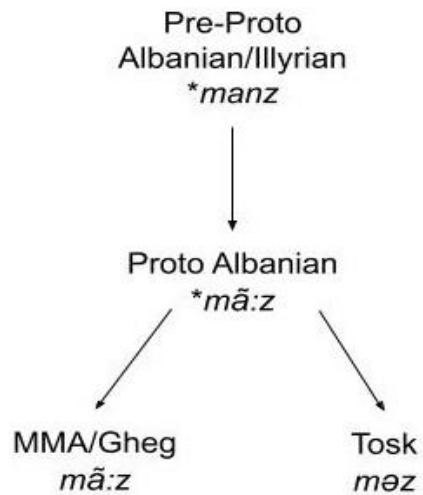


Figure 4: Proposed outcome of Proto Albanian **mā:z* > Tosk Albanian *mēz*

⁵ The 16th century writer Buzuku writes *manz*, which appears to be influenced by Italian *manzo* ‘ox, steer’, originally from Latin/Pre Illyrian influence.

(e) Latin *conventus* ‘assembly’ > MMA *kûvêt* [kũ:vǣ:t] ‘to talk, discuss’
Tosk *kuvend* [kuvend]

- Modern day Albanian parliament *Kuvendi I Shqipërise*.
- Latin <on> yields MMA [ũ:] through compensatory lengthening.
Tosk loses vowel length and nasality from Proto Albanian (see Klein et. al. 2018)
- Latin *conventus* > Proto Albanian **kûvêt* [kũ:vǣ:t] > Tosk *kuvend*.
Tosk *-vend* < Proto Albanian **-vǣ:t*.
- Expected outcome is Tosk **kunvend* not *kuvend*.

4. Epenthesis (Excrescence - the addition of a consonant)

(a) PIE **men* ‘remain, stay’ > MMA *met* [me:t] ~ Tosk *mbet*

- Cf. Latin *maneo*, Greek *meno*, and Sanskrit *man* ‘delay’⁶

(b) MMA *maro/maron* ~ Tosk *mbaron* ‘finished, destroyed, dying, withering’,

- Kloehurst (2008: 8) cites Hittite *mer-^{zi}/mar-* ‘to disappear’ as cognate with the PIE root **mer-* ‘to die’ (IE: Sanskrit *mar* ‘to die’; Avestani *mar-*; Greek *ἄμβροτος* ‘immortal’, (Hesiod *ἔμωπτεν* ‘has died’); Latin *morior* ‘to die’; Old Church Slavonic *mrěti*; Lithuanian *miřti*; Gothic *maurþr* ‘murder’; Armenian *meřanim*).
- In Beekes (2016: 904) Greek *μαραίνω* ‘to quench, destroy, die away, wither’ has no concrete etymology. Delg thinks that the root is connected to Latin *morior*. Albanian *maron* and Greek *μαραίνω* pose the closest semantic meanings to Hittite *mer-^{zi}/mar-* ‘to disappear’.

Preservation of single nasals in Malsia Madhe Albanian and epenthetic plosives in Tosk

(c) *Loanwords*

- Greek kokkúmēlon > MMA kumulla ~ Tosk kumbulla ‘plum’
- Greek palame > MMA pllâma ~ Tosk pëllëmbë ‘palm’
- Latin scamnum > MMA shkam [ʃkam] ~ Tosk shkëmb [ʃkəmb] ‘cliff, stone’

⁶ The Albanian adjectival suffix (-t/-të) in MMA *met* from the PIE verbal adjective suffix *-tós*, displaces the nasal *-n* in PIE **men* rather than yielding **ment/*mbent*. E.g., Albanian *preh* ‘cut’ < MMA *preh-t* ‘sharp’.

(d) *Cognates*

- Greek *μαραίνω* ‘to quench, destroy, die away, wither’
MMA *maro/maron* ‘finished, destroyed, dying, withering’
Tosk *mbaron*
- Middle Greek *κατόννα* ‘tent, camp’ (Tsakonian in origin)
MMA *katun* ‘village, town’
Tosk *katund*
- Romanian - *scrum* ‘ash’
MMA *shkrum* ‘cinder’
Tosk *shkrumb*

(e) Italic dialects of Albanian (Tosk variant) (see Çabej 2017: 89)

- MMA/Gheg *dimni* ~ Tosk *dimri* > Alb-Italic *dimbri* ‘the winter’
- Albanian *zemer* ~ Alb-Italic *zëmbra* ‘heart’

III. Proto Albanian

1. ‘to fill’

Proto Albanian **en busa* > *mbusha* Tosk (nb > mb)

MMA *busha*

Cognate to Greek *βυνέω* < **βυσ-ν-έω* (see Orel 1998: 253)

2. ‘in the evening’

Proto Albanian **en-prama* > *mbrëma/mbrama* Tosk/Gheg (Orel 1998: 252)

MMA *en prama/prama*

- Modern Albanian *mirë mbrëma* ‘good afternoon’
MMA *mir prama*
<myreprama> Arnold von Harff (1497) (see Elsie 1984: 118)

3. ‘to stretch out, to spread’

Proto Albanian **en tenja* > *ndej/ndêj* (Tosk/Gheg) (Orel 1998: 283)

MMA *tên* ‘to tighten, push, press’, *têj* ‘to sit’

Greek *τείνω* [tinɔ:] ‘stretch

PIE **ten-* ‘to stretch’ > Albanian *n-de(n)j* (see Hyllested and Joseph 2022)⁷

⁷ For growing evidence of a subgroup consisting of Albanian and Greek please see: Hyllested, Adam and Joseph, D. Brian. 2022. ‘Albanian’. To appear in *The Indo-European family*. Thomas Olander, ed., Cambridge University Press.

4. ‘to thrust in, to stick in’

Proto Albanian **en-kula* > *ngula* Tosk (see Orel 1998: 297)

MMA *kula*

Cf. to Hittite *gulš-* /k^wls/ ‘to carve, engrave, inscribe, write’ (see Kloekhorst 2008: 492), Greek *κολάπτω* [kolapto] ‘to peck, strike, carve, engrave’ (see Beekes 2016: 734-35)⁸

- Hittite is famously known for its cuneiform script (from Latin *cuneus* ‘wedge’), where a wedge was pressed into a soft clay tablet leaving a triangular impression (Van den Hout 2011: 9).
- Hittite *gulš-* reflects the process of ‘carving, recording’ where the scribe would have to keep pushing or sticking the wedge into the soft clay. MMA *kula* ‘to thrust, to stick in’ reflects the semantic meaning of Hittite *gulš-*.
- This close correspondence raises the question of the origin of Greek *κολάπτω* [kolapto] ‘to peck, strike, carve, engrave’ (deriv. *κολαπτός* ‘carved out’, *κολαπτήρ* ‘chisel’, *δια-κολαπτηρίζω* ‘to engrave with a chisel’).⁹

IV. Developing a Chronology

1. 16th century to Modern Albanian

(a) *peshtetete* ‘he/she is leaning on’ (Buzuku 16th century),¹⁰ MMA *pshtetet*

mpshtet (Budi 17th century)

mbështet (Modern Tosk)

Chronology: *shtat* ‘body’ > *p(e)shtet* > *mpshtet* > Modern Tosk *mbështet*

- Albanian *shtat* ‘body’ cognate with Greek *στατός*, Latin *status*, Sanskrit *sthitá*.
Prefixal *p-* + *shtat* ‘body’ > *pshtet* ‘to lean, support’
(Also note *p-* + *arrit* > *perrî* with an internal vowel change of *a* > *e*).¹¹

⁸ Albanian traditionally has been viewed as a Satem language mainly due to the outcome of the PIE palatal dorsals. In Satem languages, the outcome of **k̑* is a fricative, whereas other such as Demiraj have stated that Albanian is neither Centum nor Satem, as there is a tripart division of the palatal and labio-velars. e.g., *thom* ‘I say’ < PIE **keHsmi*, Albanian *si* ‘how’ < PIE **k̑iH*, and Albanian *kush* ‘who’ < PIE **k̑i* + **so*. There is also evidence of the preservation of the PIE labiovelar **k̑*, not just in Albanian *kush* ‘who’ from **k̑i* + **so* (Cf. Hittite **k̑is*), but also with *kul* < **k̑l* (Cf. Hittite *gulš-* /k^wls/)

⁹ See Beekes (2016: 734-735) who cites Frisk and Delg in assuming a derivation from an IE root **kelh₂-*, and that it replaces a primary verb by adding *-άπτω* after verbs with similar meaning. However, Beekes also cites as Pre Greek and assumes a derivation from *κόλαφος* ‘punch, buffet’.

¹⁰ The Old Albanian forms from the authors Buzuku, Bogdani, Budi, and Matranga are cited from Schumacher and Matzinger (2013) unless stated otherwise.

¹¹ Cf. *arri/arrit* ‘to reach, achieve, arrive’ > MMA *p-errî* ‘to arrive’ ~ Tosk *mb-ërri*.

(b) *pson* ‘to learn’ teach’ (MMA)

/ëmpson/ (Buzuku 16th century)

/mson/ /mpson/ (Budi 17th century)

/mpson/ (Bogdani 17th century)

mbëson/ *mëson*, Modern Tosk

mëson Modern Gheg

(16th century) /ëmpson/ > (17th century) /mpson/ > Modern Albanian *mbëson*/ *mëson*

(c) *pshtjell* ‘wrap, cover’ (MMA)¹²

pshtjell (Buzuku 16th century)

pështilëtë (Arvanitika Albanian: Tosk variety spoken in Greece)

mbështjell Tosk

(d) *gî* [gĩ:] /*gîn* ‘sate/satiated’ (MMA), *ngij* [ŋgij] (Tosk)

<gline> intended for <giine> (Buzuku 16th)

gliræ /*glirë*/ (Matranga late 16th century)

gine/*ngign* /*nginj*/ (Budi 17th)

/*nginj*/ (Bogdani late 17th)

Chronology: (16th century *g-*); (17th century *g-/ng-*); (Modern Tosk *ng-*)

¹² MMA *pshtjell* ‘wrap’ consists of the prefix *p-* and the root *shtjell* ‘to wind up, collect’ cognate with Greek *στέλλω* ‘to put in order, make ready’, (see Beekes 2016: 1397-98, B. Demiraj 1997: 376, Mallory and Adams 2006: 295).

(d) PIE **b^her* ‘carry, bring’ > MMA *ba(j)* ‘carry’

PIE **peh₂* ‘to shepherd, protect’ > MMA *paj* ‘hold’ (*pa* ‘to see’, *pae* [pæ] (2nd sing).

(Buzuku) *enbaj* ‘carry, hold’, *pae*, *pe* [pe:] ‘seen’ (B. Demiraj 1997: 86, 312)

Tosk *mbaj* > Gheg *maj*

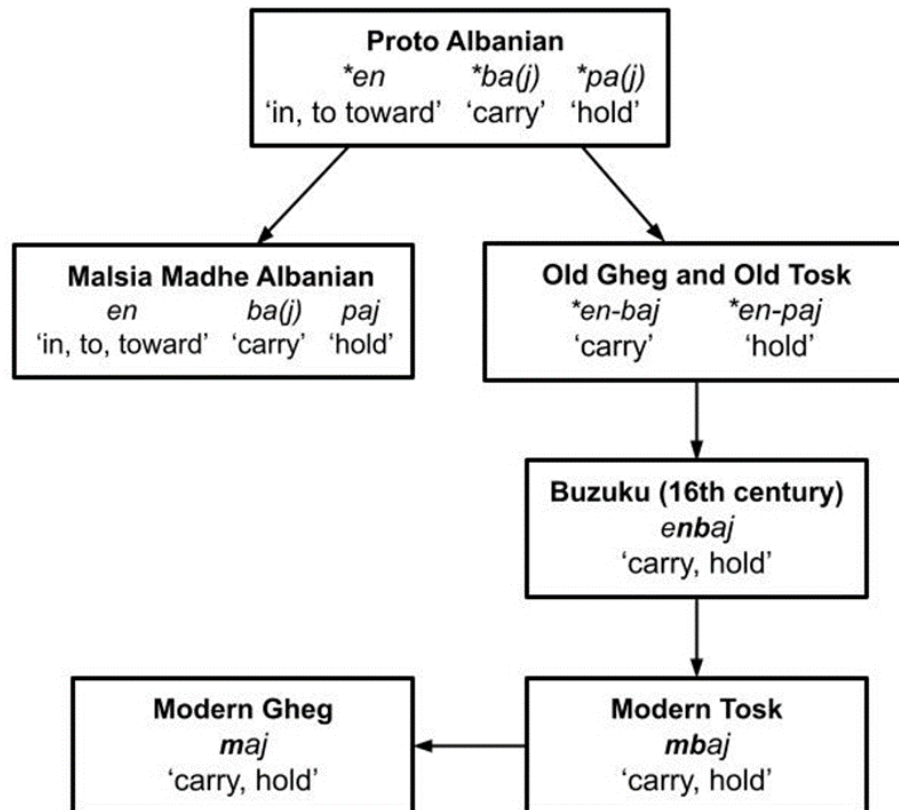


Figure 5: Formation of Tosk *mbaj* and Gheg *maj* ‘carry, hold’

2. Indo-European cognates

(a) MMA *pshî* ‘to clean, sweep’

- Tosk and Gheg variants of *fshij*, *mbëshi(n)j*, and *m(ë)shi(n)j*.
- Budi 17th century uses *mpshi* (see Schumacher and Matzinger 2013: 93).
- **pshi* > *fshij* via a process of manner assimilation of plosives before fricatives: *p* > *f* / *_sh*.
- In Mallory and Adams (2006: 376), Tosk *fshij* is connected with Ancient Greek *psáō* ‘to rub’. MMA *pshî* provides further evidence for a cognate with Ancient Greek *psáō*.

V. Conclusion

1. *Homorganic Nasal Assimilation*

- Homorganic nasal assimilation through prefixes may have begun in the period of New Albanian (16th-20th century). However, due to lack of written literature past the 16th century, it may also have begun in Middle Albanian.
- The writer Buzuku ministered in Venice and Bogdani wrote his work in exile in Padua (Schumacher & Matzinger 2013: 47). This is very well the period when the preposition *en* ‘in’ could have become a verbal prefix due to outside influence (possibly Italian *in-* from Latin *en*) as there is great variation between different authors.

2. *Epenthetic Plosives*

- Epenthetic plosives could be due to analogical extension or hypercorrection
- Nasal-plosive clusters have become quite frequent in Modern Albanian.
- At time epenthetic plosives appear to be very recent changes, occurring decades apart:

Newmark et. al. (1982: 134) *gjinore* (genitive case)

Mëniku (2008: xix.) *gjindore*

- Growing evidence of a “Markedness/Grassman’s Law” pattern in Albanian.
- *en-* prefix always guarantees a nasal plosive cluster on the leftmost plosive.
e.g., *en + paj > mbaj*
en + ten > nden
en + kul > ngul
- In a term with two nasal-plosive clusters the first cluster is reduced.
e.g., Literary Modern Albanian *mbrenda* ‘inside, within’ is colloquially said as *brenda*.

3. *Loss of Nasal Vowels in Tosk*

- Loss of vowel length in Tosk would result in the loss of nasal vowels. In Klein et. al. (2018: 1733-36) Pre-Proto Albanian **e* > Proto Albanian **ẽ* > Gheg *ẽ*.
- Proto Albanian nasal and long vowels survived in Gheg but were lost in Tosk.

Resulting sound changes due to the loss of vowel length in Tosk Albanian.

(a) Intervocalic Nasal Rhotacism (Flapping)

MMA	Tosk	
[trũ:ni]	[truri]	‘the brain’
[bð:ni]	[bəri]	‘do, make’
[prũ:na]	[prura]	‘brought’
[ũ:ni]	[uri]	‘hunger’
[gjø:ni]	[juri]	‘the knee’

(b) Before a Plosive (Unpacking)

MMA	Tosk	
[λð:t]	[ləndə]	‘tree/acorn’
[nð:t]	[nəntə]	‘nine’ ¹³
[λλ:k]	[λəng]	‘juice, liquid’
[tð:t]	[təndə]	‘yours’
[pæ:tua]	[pənduar]	‘to regret’

(c) Before a Fricative

MMA	Tosk	
[mð:z]	[məz]	‘foal’
[vð:θ]	[vəθ]	‘earring’
[græ:θ]	[grəθ]	‘wasp’
[pæ:s]	[pəs]	‘five’

(d) Open Syllables

MMA	Tosk	
[kũ:]	[ku]	‘where’
[vĩ:]	[vijə]	‘line’ ¹⁴

¹³ There are many proposed chronologies for how PIE **neun̥(-ti)* > Tosk *nëntë* ‘nine’ (see B. Demiraj *Sistemi I Numerimit Të Gjuhës Shqipe* 1997: 73-80). PIE **n̥* > Alb. *a*, which has created a problem for how the final *-n* surfaced in *nëntë*. With MMA *nât* [nð:t] there is evidence for an alternate possibility. The nasal vowel in MMA arose from the preceding nasal consonant. (e.g., *thmî* ‘child’, *nâr* ‘under’ *mâ* ‘above’.) PIE **neun̥* > Pre-Proto Albanian **naŋa(-ti)* > Proto Albanian **naate*. With this analysis the medial **u* would have dropped prior to the change of **u* > *v* (see La Piana’s hypothesis in B. Demiraj 1997: 75-6). A medial *v* would not have easily dropped from Albanian (Cf. Latin *novem*, Old Indic *náva*). Nasality in Proto Albanian would have arisen from the preceding *n-* yielding **nâte* ‘nine’ > MMA *nât*. In the period when Tosk lost vowel length, the vowel would have unpacked or segmentalized into two parts: an oral vowel *ë*, and a nasal consonant *n*, yielding *nëntë*. Gheg *nândë* can be explained as a hybrid form, preserving the old Gheg nasal vowel and borrowing the Tosk cluster *-nt-* with subsequent voicing to *-nd-*.

¹⁴ In Orel (1998: 508) *vijë* is cited as a borrowing from Latin *vīa* ‘road, way’. The MMA form *vî* [vĩ:] with a final nasal vowel would require a Proto form of **vina*, which cannot confidently be seen as a borrowing from Latin. Further investigation is required.

Malsia Madhe/Gheg Albanian – *Faleminers!*

Tosk Albanian – *Faleminderit!*

Thank you!