

## Re-evaluating Albanian's place in Indo-European studies

Lindon Dedvukaj & Patrick Gehringer\*

**Abstract.** The Albanian language has traditionally been marginalized in Indo-European Studies because it has been perceived as a minor player in the reconstruction of Indo-European vocabulary with the least number of Indo-European cognates (Mallory and Adams 2006: 26-7). This is due to the extensive borrowing in the Tosk standard Albanian dialect used by previous literature. However, it will be demonstrated that the Malsia Madhe dialect of Albanian better preserves the phonotactics and lexicon of Proto-Albanian, and Proto Indo-European. The Malsia Madhe dialect of Albanian gives an adequate understanding of the ancient history of Albanian and its place in Indo-European studies. Through this dialect many utilities are gained that can tie it to Ancient Indo-European languages such as Illyrian and Messapian.

**Keywords.** Albanian language; Proto Albanian; Malsia Madhe Albanian; Tosk Albanian; Proto-Indo-European; Illyrian; Messapic; phonotactics; Albanian phonology; Albanian lexicon; sound change; borrowings

### 1. Introduction

1.1. BACKGROUND. Albanian is an Indo-European (IE) language located in southeastern Europe and is believed to be the last surviving member of the Illyrian Family. The language has two main dialects: Tosk is tied to the geographic south, and Gheg to the north (see Figure 1). The majority of previous research on Albanian has used the standard Tosk dialect. Due to Tosk's (and sometimes Gheg's) extensive borrowings from Greek, Latin and Turkish, Mallory and Adams's (2006: 26) historical analysis of Albanian in relation to other IE languages has been hindered. This is due to Tosk's preference for borrowed words rather than its own Proto Albanian (PALb.) lexicon. Further difficulties in analyzing Albanian arise from multiple phonotactic innovations in the Tosk dialect. While these are fairly well known by researchers, previous work has at times not accounted for them, which has affected the way the transition between Proto-Indo-European (PIE) and Albanian has been documented.

The Malsia Madhe dialect of Albanian (MMA) would traditionally be described as a sub-dialect in the Northwestern Gheg region; however, the dialect preserves many unique features which point to MMA being more conservative than other dialects. The phonological conservatism of MMA in comparison to the modern Gheg and Tosk dialects will tie MMA closer to PIE. MMA is also lexically more conservative in its preservation of older forms that are potentially descended directly from PIE in opposition to Gheg and Tosk. Finally, the proposed conservatism of MMA can be utilized for potential further avenues of research in Indo-European studies. The goals are summarized in the following three proposals:

- (1) MMA is phonologically more conservative in its preservation of the Proto-Albanian (PALb.) and PIE phonotactics than the Tosk dialect.
- (2) MMA is lexically more conservative and maintains unborrowed PALb. and Illyrian words more consistently than Tosk.
- (3) MMA should be an important dialect used when evaluating Albanian in the IE language family based on the first two proposals.

---

\* Authors: Lindon Dedvukaj, Oakland University ([ldedvukaj@oakland.edu](mailto:ldedvukaj@oakland.edu), [dedvukaj.3@buckeyemail.osu.edu](mailto:dedvukaj.3@buckeyemail.osu.edu)) & Patrick Gehringer, Oakland University ([patrickgehringer.scholar@gmail.com](mailto:patrickgehringer.scholar@gmail.com)).



Figure 1. Map of traditional Albanian dialect boundaries

1.2. REGIONAL HISTORY. The historical approach to the geographic area of Malsia Madhe is naturally different from other Albanian speaking areas, with the region undergoing both times of high and low prestige. This begins with seclusion from the Roman Empire. The geographic area of Malsia Madhe corresponds with the ancient Illyrian tribe of the Labiatae. This prestigious territory was very important to the ancient Illyrians, due to the fact that in the early wars with Rome, the Labeates accepted the role of a privileged vassal state rather than being submitted to war (Buxhovi 2013: 216). During the Roman occupation of Illyria, the Illyrian born Roman Emperor Diocletian created the state of *Praevalitana* that encompassed the area of the Labeati tribe and the cities of Scodra and Lissus (modern Shkodër and Lezhë, Albania) (Wilkes 1992: 210). This same area in modern Albania and Montenegro survived until the Avar and Slavic invasions of the 6<sup>th</sup> and 7<sup>th</sup> centuries C.E.

The Illyrians retreated into the mountains and retained their identities as Albanians, only to reappear in written sources of the 11<sup>th</sup> and 12<sup>th</sup> centuries (Fine Jr. 1991: 2, 34). From the 11<sup>th</sup> century onwards, Albania was known as *Arbenia*. The name *Ἀρβανων* appears in *The Alexiad*, written by the Byzantine princess Anna Comnena (Elsie 2003: 8). With the conquest of the Ottoman Turks in the late 15<sup>th</sup> century, the natives of Malsia Madhe once again took refuge in the mountains, existing independently from the Ottomans.

The 20<sup>th</sup> century for the Albanians of Malsia Madhe was just as difficult as the previous five hundred years under Ottoman rule. During the 1913 Conference of Ambassadors in London, the Western European powers determined that the Albanians of Malsia Madhe were to be divided between Montenegro and Albania. The northern tribes of Gruda, Trieshi, Koja Kuqit, and Guci

were forcefully annexed to Montenegro. Hoti was divided between the two nations, while Kelmendi, Kastrati, and Shkreli remained under Albania (Jacques 1995: 336-37).

**2. Proposal 1 - Phonological conservatism & similarities to PIE.** There are three major phonological differences between MMA, Gheg, and Tosk. These are homorganic nasal assimilation/alternation, intervocalic nasal rhotacism, and the manner assimilation of plosives before fricatives. Each of these are well documented changes, but they have not been consistently accounted for in previous research. Furthermore, previous research did not have access to some of the context provided by the MMA data, and therefore needs to be amended in order to account for the insights it provides. The intention is to demonstrate where MMA expands or facilitates the data for historical analysis.

## 2.1. HOMORGANIC NASAL ALTERNATION

2.1.1. HOMORGANIC NASAL ALTERNATION IN ALBANIAN. This is a highly visible feature of Tosk that differentiates it from MMA. Tosk exhibits a nasal in front of plosives, providing the following alternation in (4):

- (4) a. /p, b/ (MMA) ~ /mb/ (Tosk) ~ /m/ (Gheg)  
b. /t, d/ (MMA) ~ /nd/ (Tosk) ~ /n/ (Gheg)  
c. /k, g/ (MMA) > /ŋg/ (Tosk) ~ /ŋ/ (Gheg)

There have been various approaches on how to interpret this alternation. Orel (1998 xv.) states “Tosk dialects preserve groups *mb*, *ngj* and *nd* which assimilated to *m*, *nj*, and *n* in Geg [Gheg],” which is demonstrated in (5a). Demiraj (2013: 280) states that the groups *mb*, *nd*, *ng* evolved to *m*, *n*, and *ŋ* in the northern speaking areas as seen in (5b). Beci cited in (Klein et al. 2018: 1803) states that Gheg and Tosk may be distinguishable in the correspondence of Gheg *p* ~ Tosk *mb*. Shkurtaj (1975) lists many of the MMA forms from the Kelmendi tribe but followed previous linguistic laws in their assessment. While this past approach proved true in the relation between Gheg and Tosk, Malsia Madhe does not follow this previous pattern of (5a) or (5b), but still retains the distribution seen in (5c) Gheg.

### (5) Previous explanations for nasal alternation

- a. Orel: /mb, ngj/, nd/ (Tosk) > /m, nj, n/ (Gheg)<sup>1</sup>  
b. Demiraj: /mb, n, ng/ (Tosk) > /m, n, ŋ/ (Gheg)<sup>2</sup>  
c. Beci: /mb/ (Tosk) ~ /p/ (Gheg)<sup>3</sup>

MMA completes the context. MMA preserves words where the bilabials /p/ and /b/ regularly surface as /m/ and /mb/ in modern Gheg and Tosk. Dedvukaj (2022) states that nasalisation is a feature that began in Old Tosk and Old Gheg, but that MMA preserves the original plosives as they might have appeared in Proto-Albanian. Note that examples 1-28, 30-65, 69-70 are data from Dedvukaj (2022) that have been adapted and processed for the conduct of this paper.

2.1.2 ALTERNATION IN BILABIAL PLOSIVES. In the following examples (6-18), a clear alternation between the free bilabial in MMA, the free nasal in Gheg, and the homorganic nasal in Tosk, will be demonstrated. The first column represents the words in MMA, Gheg, and Tosk, the second

---

<sup>1</sup> (Orel 1998: xv)

<sup>2</sup> (Demiraj 2013: 280)

<sup>3</sup> (Klein et. al. 2018: 1803)

column contains the corresponding glosses, column three the phonological alternation between the dialects, and finally the reconstructed Proto-Albanian sound.

(6)	ba:j~ma:j~mbaj	‘carry’	b~m~mb	*b
(7)	pa:n~ma:n~mban	‘holding’	p~m~mb	*p
(8)	ba:n~ma:n~mban	‘carrying’	b~m~mb	*b
(9)	pe:t/me:t~met~mbet	‘stuck, remain’	p/m~m~mb	*p
(10)	pro:n~mroj~mbroj	‘defend, protect’	p~m~mb	*p
(11)	pi:~mĩ:~mbi	‘over, above’	p~m~mb	*p
(12)	pʃtet~mʃtet~mbəʃtet	‘support, rely on’	p~m~mb	*p
(13)	pʃtjet~mʃtjet~mbəʃtjet	‘wrap/cover’	p~m~mb	*p
(14)	prɔ:m~mrɔm~mbrəmɔ	‘evening/last night’	p~m~mb	*p
(15)	py:t~myt~mbyt	‘drown’	p~m~mb	*p
(16)	bu:f~muʃ~mbuʃ	‘fill’	b~m~mb	*b
(17)	pe:ri~mri~mbəri	‘arrive’	p~m~mb	*p
(18)	psỹ:~mäsỹj~mbäsỹj	‘attack’	p~m~mb	*p

In order to establish the Proto-Form, PIE reconstructions are compared to its reflexes found in Albanian in (19). Here, PIE \*b<sup>h</sup> corresponds more strongly to MMA /b/ than an /m/ or /mb/. (19a) demonstrates the less efficient chronology of \*b<sup>h</sup> > mb > m~b, while in (19b) there is the more economic derivation of \*b<sup>h</sup> > b > mb > m. With this evidence, the Proto-Albanian plosive can be reconstructed to match the forms attested in MMA.

- (19) a. PIE <sup>\*\*b<sup>h</sup>er-</sup> ‘carry’ > [mbaj] ‘hold/carry’ (Tosk) > [maj] (Gheg), [ba:j] (MMA)  
 b. PIE <sup>\*b<sup>h</sup>er-</sup> ‘carry’ > [ba:j] ‘carry’ (MMA) > [mbaj] ‘hold/carry’ (Tosk) > [maj] (Gheg)

2.1.3. ALTERNATION IN ALVEOLAR PLOSIVES. The same type of alternation in the bilabial plosives occurs amongst the alveolar plosives. Where Gheg and Tosk level these sounds to /n/ and /nd/ respectively, MMA preserves the alternation between the voiced and voiceless plosive. The first set of data represents the alternation in the word initial position, and the second set represents alternation word medially.

(20)	ktej~knej~kendej	‘over here, this way’	t~n~nd	*t
(21)	tryʃɛ~nryʃɛ~ndryʃɛ	‘different’	t~n~nd	*t
(22)	tõ:t~tõ:t/tõnd~tõndɔ	‘yours’	t~t/nd~nd	*t
(23)	teʃ~neʃ~ndeʃ	‘to touch’	t~n~nd	*t
(24)	tɛrno~nroj~ndəroj	‘switch, change, turn’	t~n~nd	*t
(25)	tu:k~(?) <sup>4</sup> ~nduk	‘pluck’	t~(?)~nd	*t
(26)	dɔ:~dɔ:/ndɔ~ndare	‘split, divide, separate’	d~d/nd~nd	*d
(27)	atej~anej~andej	‘thence, over there’	t~n~nd	*t
(28)	kadɔʎ~ŋgadaʎ <sup>5</sup>	‘slow, slowly’	d~d	*d

<sup>4</sup> The Gheg entry is *shkul*, which does not match the other forms. Incidentally, in MMA this term refers to defecation (Dedvukaj 2022).

<sup>5</sup> The apparent governing of the homorganic insertion rule is discussed later, but /d/ not receiving a nasal is regular (see Table 1).

When the forms found in each dialect are compared with PIE, MMA is closer to PIE than the other dialects, which corroborates the bilabial data.

- (29) a. \*PIE \**teng*- ‘think/feel’ > [ndjɛjə] ‘feel’ (Tosk)  
 b. PIE \**teng*- ‘think/feel’ > [tiɛ] ‘hear, feel’ (MMA) [ndjɛjə] ‘feel’ (Tosk)
- (30) PIE \**deh<sub>a</sub>*(i)- ‘cut, divide’<sup>6</sup>/\**dā*(i)- ‘split, cut’ > Illyrian *dae-tor* ‘slicer, divider’<sup>7</sup>> PAIb \**daja* ‘divide’<sup>8</sup> > MMA *da(j)* [dɔː] ‘split, divide, cut’ ~ Tosk *ndaj* [ndaj]

2.1.4. ALTERNATION IN VELAR PLOSIVES. The same alternation and explanations are seen with the velar plosives. MMA and Tosk alternate /k/ with /ŋg/ both word initially and word medially just as with bilabial and alveolar plosives.

- |      |                  |                              |       |    |
|------|------------------|------------------------------|-------|----|
| (31) | ka~ŋga           | ‘where, from where’          | k~ŋg  | *k |
| (32) | kadɔɫ~ŋgadaɫə    | ‘slow, slowly’               | k~ŋg  | *k |
| (33) | grah~ŋgas        | ‘drive’                      | g~ŋg  | *g |
| (34) | kuɫ~ŋguɫ         | ‘thrust, stab, put on point’ | k~ŋg  | *k |
| (35) | kɪː~ŋgrɪə        | ‘cold, freeze’               | k~ŋg  | *k |
| (36) | katro~ŋgatəro    | ‘to fight, confuse’          | k~ŋg  | *k |
| (37) | kū:kut~kuŋguɫ    | ‘pumpkin’                    | k~ŋg  | *k |
| (38) | ktej~knej~kendej | ‘over here/this way’         | k~k~k | *k |
| (39) | lā:k~ləŋg        | ‘juice, liquid, broth’       | k~ŋg  | *k |

When the dialects are compared to PIE in examples (40, 41), MMA still deviates less from PIE when compared to Tosk.

- (40) PIE \**katu* ‘fight’<sup>9</sup>/\**katu* ‘fight’<sup>10</sup>> MMA *katrrua/katrrro* ‘fight, confuse, entangle’ ~ Tosk and Gheg *ngatërro* ‘fight, confuse, entangle’
- (41) PIE \**uleik*<sup>w</sup>/\**wlik*<sup>w</sup> ‘to make moist’ > MMA *lâk* [ɬɬ:k] ‘juice, liquid, broth’ ~ Tosk *lëŋg* [ɬəŋg] ~ Gheg *lang* [ɬəŋg]<sup>11</sup> See Dedvukaj (2022) for etymological analysis.

2.1.5. SUMMARY FOR HOMORGANIC NASAL ALTERNATION. In most instances, one can see a very clean and regular alternation between the dialects in relation to how they realize their plosives as outlined below in (42-44). Overall, PIE exhibits voiced or voiceless plosives in the same locations where MMA does, while these have leveled in Gheg and Tosk on account of the innovation of homorganic nasal assimilation.

- (42) Albanian bilabial plosives  
 a. MMA: /p/, /b/                      b. Gheg: /m/                      c. Tosk: /mb/
- (43) Albanian alveolar plosives  
 a. MMA: /t/, /d/                      b. Gheg: /n/                      c. Tosk /nd/

<sup>6</sup> (Mallory & Adams 2006: 469)

<sup>7</sup> (Krahe 1955: 71)

<sup>8</sup> (Orel 1998: 54)

<sup>9</sup> (Mallory & Adams 2006: 535)

<sup>10</sup> (Pronk 2019)

<sup>11</sup> Cf. MMA *lâk* to Latin *liquor* ‘a liquid’, Tocharian A *lyĩktsi* ‘to wash’, Tocharian B *laikanaträ* ‘washes’, and Old Irish *fliuch* ‘moist’ (De Vaan 2016: 345).



- (55) kadɔλ~ηgadaλə ‘slow, slowly’ k~ηg d~d  
 (56) katro~ηgatəro ‘to fight, confuse’ k~ηg t~t  
 (57) pʃtjɛt~mʃtjɛt~mbəʃtjɛt ‘wrap/cover’ p~m~mb t~t

/ka.daλ/	DEP-σ <sub>2</sub>	*Plosive =(CV)_V	DEP-σ <sub>1</sub>	*N <sub>[α place]C<sub>[α place]</sub></sub>	*N <sub>[α voice]C<sub>[α voice]</sub></sub>	IDENT <sub>Plosive Voice</sub>
ka.daλ-		**!				
nka.daλ-		*	*	*!		
ηka.daλ-		*	*		*!	
nga.daλ-		*	*	*!		*
☞ ηga.daλ-		*	*			*
ka.ndaλ-	*!	*				
nka.ndaλ-	*!					
ηka.ndaλ-	*!		*			
nga.ndaλ-	*!		*			
ηga.ndaλ-	*!		*			

Table 1. Tableau for governing homorganic nasal insertion

The second group displays where nasalisation appears to occur inconsistently in Tosk, applying to one plosive but not the other, often incongruent with the mechanism discussed above.

- (58) tɔ̃:t~tɔ̃:t/tɔ̃nd~təndə ‘yours’ t~t/nd~nd \*t  
 (59) kɪ̃:~ηgrɪ̃ə ‘cold, freeze’ k~ηg \*k  
 (60) kũ:kut~kuŋgʉt ‘pumpkin’ k~k \*k  
 (61) lã:k~ləŋg ‘juice, liquid, broth’ k~ηg \*k  
 (62) pɾɔ:m~mɾɔm~mbrəmə ‘evening/last night’ p~m~mb \*p

In MMA and sometimes in Gheg one can see both long vowels and nasalised vowels where in Tosk there is a nasal. Because of this context, it can be inferred that Proto-Albanian had nasal vowels in place, but these were lost in Tosk in the likely process of nasal vowel unpacking (see Dedvukaj 2022).

- (63) ner~nder~nder ‘honor, respect’ n~nd~nd \*n  
 (64) mɛno~meno/mendo~mendo ‘to think’ n~n/nd~nd \*n  
 (65) nim/dim~nim/njime~ndihmə ‘help, aid’<sup>14</sup> n/d~n/nj~nd \*n/\*d/\*nj

<sup>14</sup> MMA nim may be a borrowing from Greek *ονίνιμι* [oninimi] ‘to be of use, help, avail,’ which is derived from PIE \*h<sub>3</sub>neh<sub>2</sub> (Beekes 2016: 1083).

More difficulties seem to appear in the data above where Tosk and some Gheg nasal correspondences do not line up to the rest of the paradigm (ie. not stop~nasal~homorganic nasal). These forms appear to align with a Proto-Albanian \**n* (cf. PIE \**h<sub>a</sub>ner*<sup>15</sup>, \**men*<sup>16</sup>). Realizations of what appears to be a homorganic nasal can generally be delegated to hyper corrective analogy.

In conclusion, a model can be established of the consistent correspondences between MMA, Gheg, and Tosk plosives, nasals, and homorganic nasals, with MMA corresponding closest to PIE reconstructions.

2.2. INTERVOCALIC NASAL RHOTACISM. The second change is much simpler than homorganic nasal insertion, well documented, and unique to Tosk. In an intervocalic environment, words extant in Albanian ca. 800-1000 C.E. (Demiraj 2006: 101-2) rhotacised from /*n*/ to /*r*/, whereas Gheg and MMA maintain the nasal expected from PIE. This change is well known, but occasionally generates some opacity when analyzing potential cognates.

- (66) a. PIE \**gónu* ‘knee’ > [gǰũ:ni] ‘knee’ (MMA)  
 b. PIE \**gónu* ‘knee’ > [ǰuri] ‘knee’ (Tosk)
- (67) a. PIE \**h<sub>1</sub>nomn* ‘name’ > [ǣmɛn] ‘name’ (MMA)  
 b. PIE \**h<sub>1</sub>nomn* ‘name’ > [ɛmɔr] ‘name’ (Tosk)
- (68) a. PIE \**ǵr̥h<sub>a</sub>nóm* ‘grain’ > [gruni] ‘wheat’ (MMA) [gruri] ‘wheat’ (Tosk)  
 b. PIE \**ǵr̥h<sub>a</sub>nóm* ‘grain’ > [gruri] ‘wheat’ (Tosk)

2.3 ASSIMILATION OF PLOSIVES BEFORE FRICATIVES. Another difference between Tosk and MMA is seen in (69-70), being the manner assimilation of plosives before fricatives. Some of these examples are from words in daughter languages such as Homeric Greek and Latin. These languages provide a written record of sounds thousands of years before any written records of Albanian appear. This makes cognates and borrowed words a good gauge of how Albanian compares to earlier descendants of PIE. But again, when analysis of the phonotactic correspondences between Albanian and other IE languages are considered, MMA provides an accurate representation of historical Albanian where the standard Tosk does not.

- (69) a. πτώζ ‘cower, hide’ (Homeric Greek)<sup>17</sup>      b. [pʃɛ] ‘hide/conceal’ (MMA)  
 c. [fʃɛ] (Tosk)
- (70) a. *pascalis* ‘pasturing/grazing’ (Latin)      b. *pasq* (Old Church Slavonic)  
 c. *pasti* (Romanian)      d. [pʃikjɛ] (MMA)  
 e. [fʃikjɛ] (Tosk)

In both examples, it is debatable whether the Albanian forms were borrowed from the form in (a). In (69), Albanian modifies the Homeric Greek voiceless alveolar plosive to better agree with Albanian sonority preference, with fricativisation of the second element. This increases the sonority dispersion in the initial demisyllable compared to a flat sonority contour from the Greek #*stop+stop* sequence (Xhaferaj 2022). MMA maintains the initial plosive and the /*p*/ undergoes manner assimilation in Tosk to match the fricative.

<sup>15</sup> \**h<sub>a</sub>nér* ‘man, hero’ > *ner* ‘honor’ (MMA); cf. *nér* ‘hero’ (NWels); *ner* ‘chief’ (Umb); *anér* ‘man’ (Luv); *nar* ‘man’ (Phryg); *ayr* ‘man, person’ (Arm); *annar-* ‘man’ (Mallory 2006: 203)

<sup>16</sup> \**men-* ‘think, consider’

<sup>17</sup> (De Vaan 2008: 449)

In (70), Albanian applies vowel metathesis to better accommodate the syllable structure. Albanian has very loose constraints governing the stacking of consonants in the onset (Hysa 2019). This is a process that can drive some words to condense their consonants to maintain open syllables wherever possible (Xhaferaj 2022). Interestingly, these changes seem to be historical, as modern Tosk is described as both allowing *#stop+stop* and *#fricative+fricative* structures, but their attestation in the lexicon remains infrequent (Hysa 2019: 179-180). The same Tosk fricativization occurs to the initial plosive as in (69). In both examples, MMA maintains the unassimilated plosive, preserving the older form in contrast to Tosk innovation.

Wherever possible, regular sound changes and attested phonologies have been accounted for. In some data, there are exceptions or rules unaccounted for despite there being very striking potential cognates. Some are part of sound changes that are unaccounted for in Albanian studies due to lack of data. For others, it should be kept in mind that “Albanian,” like many languages, is a modern agglomeration of what were once the local languages and dialects of Illyria, languages for which there are little or no records. Many sound changes are likely idiosyncratic to various ancient regions made opaque by being incorporated into a larger language.

In conclusion, MMA exhibits phonological conservatism by preserving PIE plosives in comparison to Tosk/Gheg via homorganic nasal assimilation as well as the preservation of PIE *\*n* in contrast to the Tosk rhotic. Finally, MMA demonstrates greater faithfulness of initial plosives while Tosk conducts manner assimilation in the context of a following fricative. Altogether, where Tosk and occasionally Gheg fail to maintain older phonological forms, MMA often succeeds, providing context for the distribution and chronology of these processes.

**3. Proposal 2 - Lexical conservatism & similarities to PIE.** Provided with the knowledge about the phonological conservatism of Malsia Madhe Albanian, the lexical conservatism can also be demonstrated. One of the primary difficulties of studying Albanian in the past was the paucity of words not borrowed into the language. By accepting the lexical conservatism of MMA, a larger body of data is available to work with, which can be used to reinforce previous claims and motivate new analyses.

The first set of data presents etymologies established for Gheg and Tosk dependent on loans from neighboring languages, which is presented in (71-73a). The corresponding MMA words are presented in (71-73a), which are also good candidates for being Albanian reflexes of PIE vocabulary.

(71) ‘sand’

a. Latin *arena* → *rëra/rana* (Tosk & Gheg)<sup>18</sup>

b. PIE *\*m<sub>1</sub>d<sup>h</sup>o/h<sub>a</sub>-*<sup>19</sup> ‘clay’/ *melh<sub>1</sub>* ‘soft’ > *mulin* ‘sand’ (MMA)

(72) ‘suckle/nurse’

a. PIE *\*d<sup>h</sup>eh<sub>1</sub>* > *tít<sup>h</sup>e* ‘nurse,’ /*titt<sup>h</sup>ós* ‘teat, nipple’ (Classical Greek) → *thith* ‘to suck’ (Tosk)

b. PIE *\*d<sup>h</sup>eh<sub>1</sub>* ‘suck’ > *dergo* ‘suck’ (MMA)<sup>20</sup>

<sup>18</sup> (Demiraj 2006: 102)

<sup>19</sup> *\*m* > *m*; *\*l̥* > *ul*, *lu*, *il*, *li* (may account for *-uli-*), the final *-n* may be a suffix or may have some relationship with Hittite *wilan/ulin* ‘clay’. This is one of many words in MMA that do not necessarily seem to come from an Anatolian language, but there are patterns that seem more typical of Anatolian than other IE families.

<sup>20</sup> *\*d<sup>h</sup>* > *d*; *\*eh<sub>1</sub>* > *o*, *ua*; the /*r*/ is mysterious, for /*g*/ compare to Old Swedish *dæggja* ‘suck’.

- (73) ‘well’ (n.)  
 a. Latin *puteus* → *pus* (Tosk)<sup>21</sup>  
 b. PIE *\*upo* ‘under’ > *ubel*<sup>22</sup> ‘well’ (MMA)

Altogether, MMA continues to preserve older sounds and older words than attested in modern Tosk or Gheg, demonstrating a lexical conservatism. This is critical. While phonological conservatism can provide some additional context to things such as the distribution and chronologies of linguistic innovations, it principally facilitates historical analysis by avoiding many innovations that might cause errors in analysis. In general, this can be avoided anyway if the investigator is thorough and cautious with their data. Lexical conservatism is important because it provides access to words that are not found in the other dialects (e.g. *mulin* ‘sand’), making this a very important feature unique to Malsia Madhe Albanian.

**4. Proposal 3 - Using Malsia Madhe for historical analysis.** Phonological and lexical conservatism in MMA is capable of providing additional data and a potential insight into earlier forms of Albanian and Illyrian. This section provides some uses, as well as items that warrant further investigation. These suggestions will be minimal, as they deserve their own papers. The primary intention for MMA is to provide additional data and historical context for studying Albanian. The lexical and phonological conservatism allows for more connections to PIE and its daughters. When used alongside the Gheg, and Tosk dialects, a more thorough internal reconstruction of Proto-Albanian can be made.

MMA also provides additional data to the discussion on Albanians relationship with the Messapian language dating to ca. 7th-2nd centuries BCE in Apulia, Italy. Initial forays into studying the language by Seliščev (1931), Barić (1955), Georgiev (1960), and Popović (1960) led to the opinion that Albanian bore no relation to the Illyrian and Messapian languages. However, more recent works by Matzinger have given indication of a connection between the two languages. Matzinger cites similarities between Gheg *brini* ‘horn’ and the Messapian city of Brindisium (Maztinger, 2005: 35). In his more recent work cognates include Messapian *biles* ‘son’ and *bilia* ‘daughter’ with Albanian *bir* ‘son’ and *bija* ‘daughter’ (Maztinger 2019: 144).<sup>23</sup> This aligns more closely to both Messapian and Albanian being related Illyrian languages. Given this, it becomes possible to further compare the languages, particularly with MMA and its noted linguistic conservatism more likely to preserve cognate forms. For potential new cognates between Messapian, Illyrian, and MMA (see Dedvukaj 2022).

By establishing and strengthening these connections between Albanian and Messapian using MMA, a comparison between the two Illyrian languages and any Proto-Albanian forms generated with Albanian internal reconstruction, can be used to better fit these languages into the scheme of Indo-European. The addition of MMA and comparison with the other dialects of Albanian could help open up some of the discrepancies in the data likely generated by the interaction of various Illyrian tribes and their dialects, yielding clues as to what these unattested languages and dialects were like and how they might alter our wider views.

---

<sup>21</sup> (Orel 1998: 351)

<sup>22</sup> Orel provides the etymologies for both (a) and (b). The *-l* suffix is a diminutive. There is no attested historical Albanian intervocalic voicing rule currently known, although there may have been one in the Illyrian dialects or languages that this word may be from.

<sup>23</sup> Note that these words do not undergo homorganic nasal insertion in Tosk or Modern Gheg. There is no real phonological explanation other than these are very common and stable words.

**5. Conclusions.** Previous research into historical Albanian has been limited by the language's phonological innovations and its extensive borrowing from its neighbors and thus has been widely dismissed as a viable tool for studying the Indo-European family. By comparing lexical entries from across the Albanian dialects of Malsia Madhe, Gheg, and Tosk, in comparison to Proto-Indo-European and its daughters, Malsia Madhe has a phonology and lexicon more akin to these languages and less affected by innovations and borrowing. This indicates a linguistic conservatism in Malsia Madhe that can be used to provide additional data and context to historical analysis of Albanian.

This data in conjunction with the other dialects can be used to explore historical Albanian and obtain a better understanding of its place within the Indo-European family of languages. In the future, this can manifest in terms of exploring internal reconstruction of Proto-Albanian, or in better understanding the phonological history of the language and its relationship with its neighbors ranging from its lexical donors to how it might relate to Messapian or the Illyrian language family.

## References

- Beekes, Robert. 2016. *Etymological dictionary of Greek* (Vol. 1 and 2). Leiden: Brill.
- Buxhovi, Jusuf. 2013. *Kosova. Dardania in ancient and medieval times* (Vol. 1). Houston: Jalifat Publishing.
- Byron, Janet. 1976. *Selection among alternates in language standardization: The case of Albanian*. Berlin: De Gruyter. <https://doi.org/10.1515/9783110815931.fm>.
- Çabej, Eqrem. 2017. *Italoalbanische Studien: Wiener Dissertation aus dem Jahr 1933* (Albanische Forschungen 34). Wiesbaden: Harrassowitz Verlag.
- Clackson, James. 2007. *Indo-European linguistics. An introduction*. Cambridge, UK: University of Cambridge Press.
- Copeland, Mel. 2020. *Copeland English-Hittite dictionary, with concordance*. Manuscript. [http://www.maravot.com/English-Hittite\\_Dictionary.pdf](http://www.maravot.com/English-Hittite_Dictionary.pdf).
- Dedvukaj, Lindon. 2022. *The Malsia Madhe dialect of Albanian: A living fossil of ancient Europe*. Rochester, MI: Oakland University MA thesis. <http://hdl.handle.net/10323/11861>.
- Demiraj, Shaban. 2006. *The origin of the Albanians: Linguistically investigated*. Tirana, AL: Academy of Sciences of Albania.
- Demiraj, Shaban. 2013. *Gjuha shqipe dhe historia e saj*. Tirana, AL: Onufri.
- De Vaan, Michiel. 2016. *Etymological dictionary of Latin and the other Italic languages*. Leiden/Boston: Brill.
- Drizari, Nelo. 1957. *Albanian-English and English-Albanian dictionary*. New York: Ungar.
- Elsie, Robert. 2003. *Early Albania: A reader of historical texts 11th-17th Centuries*. Wiesbaden: Harrassowitz.
- Elsie, Robert. 2016. *Early Albanian bible translations in Todhri script*. London: Centre for Albanian Studies.
- Fine, V.A. John, Jr. 1991. *The early medieval Balkans. A critical survey from the sixth to the late twelfth century*. Ann Arbor: The University of Michigan Press.
- Hammond, Nicholas G. L. 1994. *Illyrians and north-west Greeks*. In D. M. Lewis, John Boardman, Simon Hornblower & M. Ostwald (eds.), *The Cambridge ancient history* (Volume 6). 422–443. Cambridge: Cambridge University Press.
- Hoffner A. Harry Jr. 2008. *A grammar of the Hittite language, Part I: Reference grammar*. Winona Lake, IN: Eisenbrauns.

- Hysa, Merita. 2019. The beginning of the syllable in Albanian. *Quaderni di Linguistica e Studi Orientali* 5. 175–193. <https://doi.org/10.13128/QULSO-2421-7220-25965>.
- Jacques, E. Edward. 2009. *The Albanians. An ethnic history from prehistoric times to the present* (Vol. 2). Jefferson, NC: MacFarland & Company.
- Klein Jared, Joseph Brian, Fritz Matthias. 2018. *Handbook of comparative and historical Indo-European linguistics* (Vol. 3). Berlin/Boston: Walter de Gruyter.
- Krahe, Hans. 1955. *Die Sprache Der Illyrier. Erster Teil: Die Quellen*. Weisbaden: Verlag Otto Harrassowitz.
- Mallory J. P. & Douglas Q Adams. 2006. *The Oxford Introduction to Proto-Indo-European and the Proto-Indo-European world*. New York: Oxford University Press.
- Matzinger, Joachim. 2005. Messapisch und Alabnisch. *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 2. 29–54.
- Matzinger, Joachim. 2019. Messapisch. Wiesbaden: Reichert Verlag.
- Mëniku, Linda. 2008. *Gheg Albanian reader*. Hyattsville, MD: Dunwoody Press.
- Morwood, James. 2005. *Oxford Latin desk dictionary*. New York: Oxford University Press.
- Morwood, James & John Taylor (eds.). 2002. *The pocket Oxford Classical Greek dictionary*. New York: Oxford University Press.
- Orel, Vladimir. 1998. *Albanian etymological dictionary*. Leiden: Brill.
- Pronk Tijmen. 2019. Proto-Indo-European \*a. *Indo-European Linguistics* 7(1). 122–163. <https://doi.org/10.1163/22125892-00701002>.
- Schumacher Stefan, Matzinger Joachim. 2013. *Die Verben des Altalbanischen. Belegwörterbuch, Vorgeschichte und Etymologie*. Wiesbaden. Harrassowitz.
- Shkurtaj, Gjovalin. 1975. *E folmja e Kelmendit*. Tirana, AL: Studimet Filologjike.
- Trumper, John. 2018. Some Celto-Albanian isoglosses and their implications. In Grimaldi Mirko, Lali Rosangela, Franco Ludovico & Benedetta Baldi (eds.), *Structuring variation in Romance linguistics and beyond: In honour of Leonardo M. Savoia* (Linguistics Today 252), 379–389. Amsterdam: John Benjamins. <https://doi.org/10.1075/la.252.26tru>.
- Valpy, Francis Edward J. 1828. *An etymological dictionary of the Latin language*. London: Baldwin & Co.
- Van Den Hout, Theo. 2011. *The elements of Hittite*. Cambridge: Cambridge University Press.
- Wilkes, John. 1992. *The Illyrians*. Cambridge MA: Blackwell.
- Xhaferaj, Artan. 2022. The sonority dispersion principle in Albanian. *European Journal of Social Science Education and Research* 9(1). 40–47. <https://doi.org/10.26417/994fto61>.