

Morphological and phonological origins of Albanian nasals and its parallels with other laws

Lindon Dedvukaj & Patrick Gehringer*

Abstract. The Albanian language is traditionally divided between the Gheg dialect to the geographic north and the now Standard Tosk dialect to the geographic south. Recent literature of the historically isolated dialect of Malsia Madhe (Dedvukaj 2022) has revealed a subdialect which has not undergone the specific phonological sound changes seen in both the Standard Tosk and Modern Gheg dialects. The Tosk dialect is distinct from the dialects of Gheg and Malsia Madhe (Malsia) in that it contains homorganic nasal-stop clusters in positions where they did not occur in various Proto-Indo-European (PIE) reconstructed forms. Three historical processes and the distinct ways in which nasal-stop clusters appear in Tosk are discussed: homorganic nasal assimilation that occurred in the 16th-18th century, the insertion of epenthetic stops due to sonority constraints and analogical extension, and constraints on the analogy that can be attributed to the Obligatory Contour Principle (OCP), which restricts Tosk Albanian to one nasal-stop cluster within a single morpheme.

Keywords. Albanian phonology; morphology; Tosk Albanian; Gheg Albanian; Malsia Madhe Albanian; Old Gheg and Old Tosk; Proto-Albanian; Proto-Indo-European; nasal-stop clusters; prefixal *en-*; homorganic nasal assimilation; apheresis; epenthetic stops; Obligatory Contour Principle; Grassmann's Law; Rosenthal's Law

1. Introduction. Albanian is an Indo-European language, traditionally divided between the northern Gheg dialects and southern Tosk dialects. Standard Tosk is the national variety, and its dialects are the most widespread within Albania and in other countries (western Macedonia, Northern Greece, Southern Italy). Gheg is spoken in the geographic north of Albania and in the countries of Kosovo, Serbia, Macedonia, and Montenegro. Related to Gheg is the historically isolated Malsia Madhe Dialect (Malsia), which preserves more features of Proto-Albanian.¹

Across the Albanian dialects, the principal nasals are /m/ and /n/. The alveolar nasal also has an allophone [ɲ] that appears before velar consonants. Additionally, the nasals /m/ and /n/ appear in homorganic nasal-stop clusters [mb] and [nd] in Tosk Albanian. Both the phonemic nasals and homorganic clusters occur word-initially, word-medially, and word-finally (see Table 1). The nasals /m/ and /n/ continue earlier nasals alone, whereas many of the homorganic clusters are not so straightforwardly accounted for in Indo-European (IE). The main focus of this paper is the origin of the homorganic nasal-stop (NT) clusters in Tosk Albanian. The analytical claim is that these clusters can be broken down into two groups.

* Special thanks to Brian D. Joseph (The Ohio State University), Björn Köhnlén (The Ohio State University), and Samuel Rosenthal (Oakland University), for their help and feedback. Authors: Lindon Dedvukaj, The Ohio State University (dedvukaj.3@osu.edu), Patrick Gehringer, University of Kentucky (pge228@g.uky.edu).

¹ The main corpus of data and the analysis are taken from Dedvukaj (2022), adapted in Dedvukaj and Gehringer (2022).

- (1) The majority of word-initial NT clusters are produced by a morpho-phonological process whereby the Proto-Albanian preposition **en* developed into a prefix, attaching to stems with initial stops (e.g. /ɛn/ + /paj/). The unstressed initial vowel deleted through a regular sound change (apheresis), and the nasal consonant assimilated to the place of articulation of the following stop, which subsequently voiced: /ɛn/ + /paj/ = [mbaj] ‘hold, carry’ (§3).
- (2) The second group of NT clusters formed by stop epenthesis can be divided into three subgroups: those phonetically motivated in (a) and (b), and those that arose through analogy in (c). The relevant processes are discussed in more detail in §3-4.
- (a) Medial epenthetic stops which obey the Syllable Contact Law ([pɛnd-ə] ‘feather’, [-ə] indefinite feminine article), and epenthetic stops which appear to be in word-final position due to a -∅ indefinite article in masculine nouns ([katund-∅] ‘town’) (§4.2).
- (b) Epenthesis due to mistiming ([əmbəl] ‘sweet’) (§4.3).
- (c) Word-initial epenthetic stops that have developed into NT clusters, possibly through analogy with the apheretic clusters ([mbet] ‘remain, stuck’) (§4.4).

	[m]	[n]	[mb]	[nd]	[ŋg]
Initial	[mal] ‘mountain’	[nata] ‘the night’	[mba] ‘carry’	[nder] ‘honor’	[ŋga] ‘where from’
Medial	[amuł] ‘stagnant water’	[ana] ‘the side’	[əmbəl] ‘sweet’	[pɛndə] ‘feather’	[kuŋguł] ‘pumpkin’
Final	[fɛstim] ‘celebration’	[cɛn] ‘dog’	[krimb] ‘worm’	[vɛnd] ‘place’	[ləŋg] ‘juice’

Table 1. Positional distribution of standard Albanian nasals

2. Attested nasals and gaps. The first step is to confirm which nasals are attested in the Indo-European family. This includes checking Tosk Albanian against Proto-Indo-European (PIE) reconstructions and other Albanian dialects.

2.1 ATTESTED NASALS IN INDO-EUROPEAN. The following examples demonstrate the origins of nasals in Tosk Albanian directly derived from PIE. Examples (3-5) demonstrate bilabial nasals in word-initial, medial, and final positions, and (6-8) demonstrate alveolar nasals in the same environments. (“Albanian” will be used if the term is consistent cross-dialectally, otherwise the dialectal variant will be specified).

- (3) #m PIE **méh₂tēr-* ‘mother’ > Albanian *motër* ‘sister’
- (4) VmV PIE **h₁ŋh₃-men-* ‘name’ > **h₁ŋmen-* > Tosk *emër* ~ Gheg *êmen* ‘name’
- (5) m# PIE **g^{wh}er-mo-* ‘fire’ > Proto-Albanian **džerm* > Tosk *zjarm* ~ Malsia *gjarm* [jarm] ‘fire’
- (6) #n PIE *(*s*)*nusós* ‘daughter in law’ > Albanian *nuse* ‘bride, young woman, younger daughter in law’
- (7) VnV Proto-Albanian **taruna* ‘brain’ > Gheg *trûni* ‘the brain’²
- (8) n# PIE **H₁én-* ‘in’ > Old Tosk/Old Gheg *en-* ~ Malsia *en* ‘in’

2.2 UNATTESTED NASALS IN INDO-EUROPEAN. Examples (9-12) demonstrate where NT clusters can appear in Tosk Albanian and where they are not attested in PIE reconstructions or in the relevant IE languages. The bilabial distributions demonstrate that the NT clusters can occur in the same position as the IE nasals, and without attestation in those positions elsewhere in the IE Family. (The alveolar nasals have the same distributions as the bilabials).

- (9) PIE **mer-* ‘to die’ > Tosk *mbaron* ~ Gheg *maron* ‘finished, destroyed, dying, withering’
- (10) PIE **b^her-* ‘to carry’ > Tosk *mbaj* ‘hold, carry’ ~ Malsia *baj* ‘carry’
- (11) Sanskrit *amel* ‘sour, acid’; Tosk *ëmbël* ~ Gheg *âmel* ‘sweet’
- (12) PIE **k^wrmis* ‘worm, insect’ > Tosk *krimb* ~ Gheg *krym* ‘worm, maggot’

The analysis will attempt to explain the possible origins of the Tosk Albanian NT clusters in word-initial, medial, and final positions. While previous work has identified that Gheg reduces the Tosk word-initial clusters to nasals (Gjinari 1989, S. Demiraj 2013, Shkurtaj 2016), this does not account for the appearance of homorganic clusters in other environments. This also does not explain the different distribution of the word-initial stops in Malsia in contrast to the Tosk NT clusters, which alters the diachronic analysis.

3. Nasal-stop clusters by apheresis

3.1. PREFIXAL EN- REANALYSIS. There are several word-initial NT clusters in Tosk Albanian that are neither loanwords nor possess a word-initial vowel. These occur within the categories of labials, alveolars, and velars (13-15).³

- (13) PIE **b^her-* ‘carry’ > Tosk [mbaj] ‘carry, hold’
- (14) PIE **ten-* ‘to stretch’ > Tosk [ndej] ‘stretch, spread’
- (15) PIE **g^hod^h-* ‘to join, unite’ > Tosk [ŋge] ‘time, chance, opportunity’

² The denasalization (rhotacism) of intervocalic Proto-Albanian **-n-* to Tosk *-r-* prevents there being any clear examples, but there are cases like *punë* ‘work, labor’ < *(*s*)*puđ-no-* (cf. Grk *σπευδω* ‘rush, hasten’) and *dhënë* ‘given’ < **dH₃-no-*, where rhotacism has been blocked.

³ Xhaferaj (2022: 45) notes that the occurrence of the Albanian groups /mb/, /nd/, /ng/, and /mp/ in the initial position is a feature that distinguishes Albanian from the other Balkan languages.

Another complicating factor is that when examining the word initial homorganic NT clusters, a significant portion of them appear as nasals in Gheg, and as stops in Malsia (Table 2).

	Malsia	Gheg	Tosk (Standard)
Bilabials	p, b	m	mb
Alveolars	t, d	n	nd
Velars	k, g	k, ŋ, ŋg	ŋg

Table 2. Albanian dialectal distribution of homorganic stops and nasals

This is most distinguishable throughout the Albanian dialects in the word-initial bilabial, alveolar, and velar categories (16-21).⁴

- | | | | |
|------------------|--------------------|---------------------------|--------------------|
| (16) ‘to wrap’ | a. Malsia [pʃtjɛʃ] | b. Gheg [mʃtjɛʃ] | c. Tosk [mbəʃtjɛʃ] |
| (17) ‘to fill’ | a. Malsia [buːʃ] | b. Gheg [muːʃ] | c. Tosk [mbuʃ] |
| (18) ‘to touch’ | a. Malsia [teːʃ] | b. Gheg [neːʃ] | c. Tosk [ndeʃ] |
| (19) ‘to divide’ | a. Malsia [dɔː] | b. Gheg [dɔː], [ndɔː] | c. Tosk [ndarə] |
| (20) ‘to freeze’ | a. Malsia [kr̥iː] | b. Gheg [ŋr̥iː], [ŋgr̥iː] | c. Tosk [ŋgr̥iː] |
| (21) ‘to drive’ | a. Malsia [grah] | b. Gheg [grah] | c. Tosk [ŋgas] |

When examining the same words from their PIE origins and across the IE Family (22-27), Malsia appears to have preserved the IE voiced and voiceless stops. Tosk has homorganic NT clusters, leaving Gheg in a transition state.

- (22) PIE **peh*₂- ‘protect’ > Malsia [pɔːj] ‘to hold, keep’ ~ Tosk [mbaj] ‘hold, carry’ ~ Gheg [mɔːj]
- (23) PIE **b^her*- ‘bear, carry’ > Malsia [bɔːj] ‘carry’ ~ Tosk [mbaj] ‘hold, carry’ ~ Gheg [mɔːj]
- (24) PIE **ten* > ‘stretch, tighten’ > Malsia [tãːn] ‘push, press’ ~ Tosk [ndej] ‘stretch’⁵
- (25) PIE **deh*₂- ‘share, divide’ > Malsia [dɔː] ‘split, cut, divide’ ~ Tosk [ndaj] ~ Gheg [dɔː]
- (26) Hittite /*k^wls-*/ ‘to carve’; Malsia [kuːʎ] ‘to thrust, put on point’ ~ Tosk [ŋgul]⁶
- (27) PIE **g^hod^h-* (o-grade of **g^hed^h-*) > Malsia [gæː] ‘time, chance, opportunity’ ~ Tosk [ŋge]⁷

⁴ Malsia data taken from Dedvukaj (2022); Shkurtaj (1967, 1974, 1975, 2016).

⁵ Hyllested and Joseph (2022): PIE **ten*- ‘to stretch’ > Albanian *n-de(n)**j*

⁶ See Kloekhorst (2008: 492-93) for the etymology of Hittite *gulš-^{zi}*.

⁷ Cf. Latvian *gads* ‘year’

The shift from a word-initial proto stop to a nasal-stop cluster in Tosk is evident. Previous reconstructions accounted for this by positing a prefix *en-*, derived from an earlier Proto-Albanian preposition *en* ‘in’ (28-30 from Orel 1998).

(28) Proto-Albanian **en-busa* > Tosk [mbuʃa] ‘to fill’ (Malsia [buʃa])

(29) Proto-Albanian **en-tenja* > Tosk [ndej] ‘to stretch’ (Malsia [tã:n])

(30) Proto-Albanian **en-kula* > Tosk [ŋgula] ‘to thrust, put on point’ (Malsia [ku:ʌ])

Having introduced a vowel and a nasal in the form *en-*, an unstressed word-initial vowel preceding a NT cluster appears. The Malsia forms with initial stops appear to preserve the phonemic environment that was present in Proto-Albanian (4th-7th CE) and PIE (22-27). Take for example the formation of the Tosk Albanian adverb *nga* [ŋga] ‘where, from where’. *Nga* is a shortened form of **ën-ka* ‘where, from where’. The preposition **ën* ‘in’ is added to the old relative adverb *ka* ‘where, from where’ (see Orel 1998: 292, Fortson 2010: 455). Only in Malsia has the Proto-Albanian preposition *en* ‘in’ been preserved. In the other Albanian dialects this has been reanalyzed as a verbal prefix, thus losing its prepositional quality. Malsia *en* ‘in’ and *ka* ‘where, from where’ are separate lexical terms preserved from Proto-Albanian, whereas in Tosk Albanian these two terms attached and ultimately led to the formation of *nga* < **ën-ka*. Note that there is progressive voicing assimilation of the stop and regressive place assimilation of the nasal in Modern Tosk.

If Malsia preserved the preposition *en* ‘in’ and the word-initial stops from Proto-Albanian, then the question that remains is during what period of Albanian did Tosk reanalyze the preposition *en* as a prefix? It is possible to position this change during the period of Old Albanian (16th-18th CE), which encompasses both dialects in the literary period of Old Gheg and Old Tosk. The chronological progress can be exemplified with the PIE terms **b^her-* ‘carry’ and **peh₂-* ‘protect’, as shown in Figure 1.

In Proto-Albanian (4th-7th CE), these terms yield **baj* ‘carry’ and **paj* ‘hold’. The Proto-Albanian preposition *en* ‘in’ from PIE **h₁en* ‘in’ developed into a verbal prefix by the time of Old Gheg and Tosk (16th-18th CE). The verbal prefix *en-* attached to the verbs forming **enbaj* and **enpaj*. The Old Gheg priest Gjon Buzuku wrote the form of <enbaj>, which means both ‘to carry’ and ‘to hold’ (B. Demiraj 1997: 86). The progressive voicing assimilation of the stop and regressive place assimilation of the nasal occurred in Modern Tosk, as well as in Old-Gheg three centuries earlier, but the orthography of <enbaj> demonstrates the writer Buzuku’s knowledge that *en-* is a prefix and distinct from the verbal root. Old Gheg and Old Tosk erased the distinction between **en-paj* ‘hold’ and **en-baj* ‘carry’, which merge into **em-baj* ‘carry, hold’. As the distinction between the verbal prefix and the root continued to weaken, the unstressed word-initial vowel deleted through apheresis. /en/ now /n/ is reanalyzed into the root and incorporated into the underlying form, yielding Modern Tosk *mbaj* ‘carry, hold’. Modern Gheg *maj* is a reduction from the cluster *mb-*.

	16 th CE	17 th CE	18 th CE	21 st CE	
‘to lean on’	p ështetëtë (Buzuku)	mp shtet (Budi)	-----	p shtet (Malsia)	mb ështet (Standard Tosk)
‘to fill’	ënb ushën (Buzuku)	mb bush (Budi)	mb bush (Bogdani)	b ush (Malsia)	mb bush (Standard Tosk)
‘feel, hear, understand’	ënd ien (Buzuku)	nd ien (Budi)	nd ien (Bogdani)	t ie (Malsia)	nd ien (Standard Tosk)
‘to, cut, split, divide’	da (Buzuku)	dān (Budi)	nda- (Variboba)	da (Malsia)	nda (Standard Tosk)
‘satiare’	1) ënginj ën, g injën (Buzuku)	nginj , g īne (Budi)	-----	[gĩ:j] (Malsia)	[ŋgij] (Standard Tosk)
	2) gl īrë (Matrënga)				

Table 3. Diachronic development of *en-* from Old Albanian to Modern Albanian

Buzuku’s usage of the prefix *en-* varied. Interestingly, he even wrote the same term with and without the prefix (*ënginjën ~ ginjën* ‘satiare’). The same variation is evident with the Gheg author Budi in the 17th century (*nginj ~ gīne*). The Old Tosk writer Matrënga wrote the form *glīrë* without an initial nasal, whereas in Modern Tosk the prefix has been fully reanalyzed into the root as a homorganic cluster [ŋg] in [ŋgij].⁸ The forms in the Malsia dialect show evidence of an earlier isolation and preservation of the Proto-Albanian stops (also evidenced in Figure 1).

3.2. HISTORICAL APHERESIS AND THE FORMATION OF INTIAL NASAL-STOP CLUSTERS. The foremost attested method for explaining the emergence of word-initial NT clusters in Tosk Albanian, is through the regular loss of initial unstressed vowels (also known as apheresis). The Proto-Albanian reconstructions (31-33) demonstrate the loss of the initial unstressed vowels in Modern Albanian forms. Examples 34-36 demonstrate how word-initial NT clusters were formed through the same process in Latin loans.

(31) Proto-Albanian **apa* > Albanian *pa* ‘without, before, then’⁹

(32) Proto-Albanian **awantha* > Gheg *vâth* ~ Tosk *vëth* ‘earring’¹⁰

(33) Proto-Albanian **ōlenā* ‘elbow’ > Gheg *llânë* ~ Tosk *llërë* ‘forearm’¹¹

⁸ Old Albanian data taken from Schumacher and Matzinger (2013)

⁹ Orel (1998: 307). Cf. Sanskrit *āpa* ‘away, off’, Greek *ἀπό* ‘from, Gothic *af* ‘from’.

¹⁰ Witczak (2016) “The earliest Albanian loanwords in Greek”.

¹¹ See Fortson (2010: 451).

- (34) Latin *imperātor* ‘emperor’ → Tosk *mbret* ‘king’
 (35) Latin *indīcāre* ‘to impose, inflict’ → Tosk *ndikoj* ‘to influence’
 (36) Latin *ingrātus* ‘wretched, unhappy’ → Tosk *ngratë* ‘id.’

To summarize, the process of apheresis was productive in Albanian, evident in Proto-Albanian reconstructions and loan terms. The Old Albanian *en-* reanalysis and historical apheresis accounts for the vast majority of Tosk word-initial NT clusters not present in IE.

4. Nasal-stop clusters by sonority driven/motivated epenthesis. Epenthetic stops (also known as excrescence) surface in Tosk Albanian word-initially (37), medially through a mistiming of the articulatory gestures (38), and medially on a morpheme boundary (39). Example 40 is also a medial cluster but appears word-finally in the indefinite form due to a -Ø suffix in masculine nouns.

- (37) PIE **mer-* ‘to die’ > Tosk *mbaron* ‘finished, destroyed, dying, withering’
 (38) Proto-Albanian **amla* ‘sweet’ > Tosk *ëmbël* [əmbəl] ‘sweet’¹²
 (39) Latin *pinna* ‘feather’ → Tosk *pendë* [pəndə] ‘feather’
 (40) Middle Greek κατούνα ‘tent, camp’; Tosk *katund* ‘village’

There are few options for how to explain the appearance of word-initial, medial, and apparent word-final NT clusters. It could be generated by a phonotactic constraint in Albanian such as metrical or syllabification requirements, a sonority constraint, or it could be an analogical extension of apheresis which generates word-initial NT clusters.

4.2 MEDIAL CLUSTERS. Analysis reveals that these clusters can be motivated to optimize sonority. Albanian is not phonotactically stringent in terms of syllabification or sonority profiles, allowing up to (C)(C)CC structured onsets (*çndryshk* ‘unrust’) with a high degree of freedom in what type of consonants are allowed within that structure (Hysa, 2019: 192); nonetheless certain conditions can optimize these. First focusing on word-medial constructions, we can compare Gheg and Tosk equivalents (41), showing that the NT cluster in Tosk corresponds to the nasal in Gheg, requiring stop epenthesis to achieve the Tosk form.

- (41) ‘feather’ Gheg [pẽ:nə]/[pẽ:nnə] ~ Tosk [pəndə]

The medial cluster in (41) occurs at the syllable boundary between the nominal root and the case suffix. Knowing that we are inserting a stop word medially between the nominal root (*pen-*) and the case suffix -*ë* (singular feminine indefinite) it is possible to motivate epenthesis (42).¹³

- (42) (a.) ‘feather’ /pen.-/ → /pen.T+ə/ [pən.də] NOM.INDEF.SG.
 (b.) ‘the feather’ /pen.-/ → /pen.T+a/ [pən.da] NOM.DEF.SG.

The Syllable Contact Law in Murray and Vennemann (1993) states that at points of syllable contact, a language will attempt to maximize the sonority drop from the preceding syllable to the following syllable. This is what appears to be occurring between the nominal root

¹² Tomaschek (ZG 1875, 529) reconstructs Proto-Albanian **amla* (in Orel 1998: 91).

¹³ See Newmark (1982: 159-61) for the Albanian feminine and masculine singular noun declensions.

and case suffix. The sonority drop from the nucleus of syllable-1 is maximized by the insertion of a stop before rising again with the onset of syllable-2. In Figure 2, the Latin loan *pinna* ‘feather’ would presumably surface in Tosk Albanian as **penë* but in order to maximize the sonority drop from syllable-1 *pen-* to syllable-2 *-ë*, Tosk inserts a stop that receives laryngeal and place features from the nasal /n/.

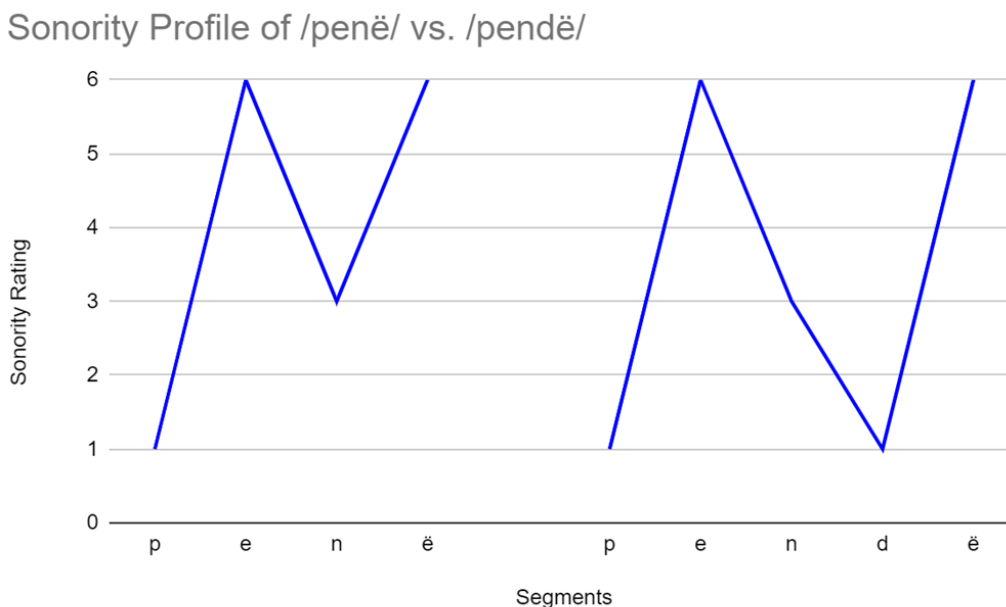


Figure 2. Comparing the intersyllabic sonority drop of *penë* vs. *pendë*

The same process occurs with the masculine nouns, which receive a null case $-\emptyset$ in the indefinite forms, but $-i$ in definite forms, allowing for the same syllable contact driven epenthesis (43).

- (43) (a.) ‘worm’ /krim.-/ → /krim.T+ \emptyset / [krim.b] NOM.INDEF.SG
 (b.) ‘the worm’ /krim.-/ → /krim.T+i/ [krim.bi] NOM.DEF.SG

Also note the masculine definite forms in (44-46), where the NT clusters in Tosk appear in the medial position.

- (44) ‘the cinder’ [ʃkru:**m**-i] (Malsia/Gheg) [ʃkr**umb**-i] (Tosk)
 (45) ‘the village’ [katu:**n**-i] (Malsia/Gheg) [kat**und**-i] (Tosk)
 (46) ‘the nation’ [ko:**m**-i] (Malsia/Gheg) [ko**mb**-i] (Tosk)¹⁴

4.3 WRONG TIMING EPENTHESIS.¹⁵ An instance of a Standard Tosk medial cluster /-mb-/ that cannot be attributed to syllable contact epenthesis is the adjective *ëmbël* [əmbəl] ‘sweet’. This surfaces in Malsia and Gheg as *âmel* [ʃ:məl] with a medial nasal [m]. Tomaschek (ZG 1875, 529) reconstructs a Proto-Albanian **amla* cognate with Sanskrit *amlá* ‘sour’ (cited in Orel 1998: 91). B. Demiraj (1997: 169-70) reconstructs **h₂en-m(e)lit-*, which would still position /b/ as an

¹⁴ Cf. Ancient Greek *κώμη* /kɔ:mɛ:/ ‘village’.

¹⁵ Thank you to Brian D. Joseph for his recommendation of “wrong timing epenthesis” (p.c. December 2022).

epenthetic stop; **h₂en-m(e)lit-* > Gheg *âm(e)li* suggesting that the NT cluster in Tosk is the result of an epenthetic stop.¹⁶

The medial cluster does not occur at a syllable boundary marked by a final article as in §4.2. Rather than positing analogy, it is possible to attribute the change to a miscontrol of the articulatory gestures. Hock (2021: 115-16) calls this “wrong timing epenthesis”, citing the development of an oral stop between a nasal stop and a [-nasal] consonant. Examples of this type of change include pre-Greek **anros* > *andros* ‘of man’ and pre-Greek **amrotos* > *ambrotos* ‘immortal’. The switch from *-n-* to *-r-* in *andros* is not completed at once. Hock (115) states, “Instead, the gesture of nasality is discontinued before the stop articulation comes to an end. The result is a stretch of an oral dental stop.”

In the Tosk Albanian feminine singular *e ëmbël*, masculine singular *i ëmbël*, and masculine plural *të ëmbël*, the cluster *-mb-* is not adjacent to */-l/*. However, in the feminine plural *të ëmbra*, the epenthetic stop surfaces between a nasal and liquid consonant. The Gheg form *të âmla* provides the environment where the nasal *-m-* and liquid *-l-* are adjacent and a miscontrol of the articulatory gestures could have occurred (47).

(47) **të amla* > *të ëmbra*

Interestingly, Murray and Vennemann (1983: 520) state that according to the Syllable Contact Law a hypothetical pattern *am.la* is a preferred consonantal pattern to a form *at.ia*, because the consonantal strength values of *l* and *m* differ. Analogy for the origin of the word-initial clusters will be posited in the following section, but in this instance, it appears as if the epenthetically generated stop in the feminine plural *të ëmbra* may have spread to all case forms of the adjective.¹⁷

In the Italo-Albanian dialects (a Tosk variety), there are forms with epenthetic stops distinguishable from the Standard Tosk forms with a lone nasal. These include *zëmbra* < *zemra* ‘the heart’, and *dimbri* < *dimri* ‘the winter’ (see Çabej 2017: 89). The epenthetic stops in these forms occur in definite forms where the nasal [m] and rhotic [r] are adjacent. These are two other cases just as **të amla* > *të ëmbra* that can be attributed to a mis-control of the articulatory gestures.

4.4. INITIAL NASAL-STOP CLUSTERS BY ANALOGY TO APHERESIS. The last group that cannot be accounted for with syllable contact epenthesis are the word-initial NT clusters. Examples (48-50) show that these initial clusters are historically from word-initial nasals. In fact, the insertion of a stop as seen in the Tosk forms actually interrupts the optimal sonority profile (see Figure 3 for Gheg *ner* ~ Tosk *nder*).

(48) PIE **mer-* ‘die’ > Gheg *maron* ~ Tosk *mbaron* ‘destroyed, dying, withering’

(49) PIE **men-* ‘remain’ > Gheg *met* ~ Tosk *mbet* ‘remain, stuck’

¹⁶ Compare to the origin of the nasal vowel in Gheg *âsht*. Hamp (1980) cites Tosk *është* and Gheg *âsht* < **ensti*, where this derives from a phrase with the adverb **H_een*, the “locative of a defective noun ... in late Indo-European.”

¹⁷ Xhaferaj (2022: 44) also cites the difficulty in articulating the initial groups /mr-/ and /ml-/. In Standard Albanian, insertion of an epenthetic vowel /ë/ between a nasal and liquid has become more frequent.

(50) PIE **h₂ner* ‘man, power, force’ > Albanian *njer* ‘man’ > Gheg *ner* ~ Tosk *nder* ‘honor’¹⁸

Sonority Profile of /ner/ vs. /nder/

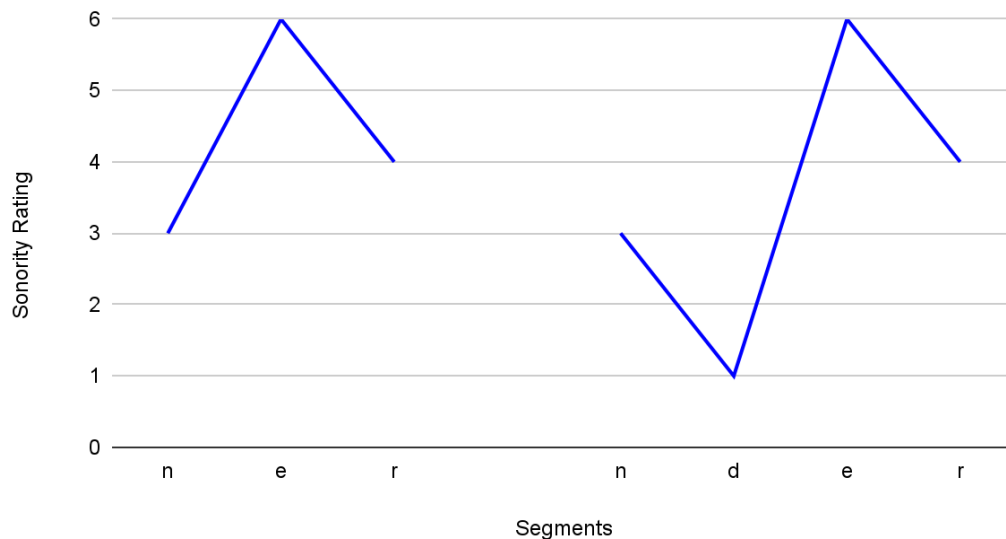


Figure 3. Syllable Sonority profile of Gheg *ner* and Tosk *nder*

There is however a model in the productive apheresis group in Albanian (§3). This large group of words has undergone a process of cluster formation. Analogy with the apheretic group would assume that a stop is inserted in order to form the target word-initial cluster. It is also possible that sociolinguistic pressures to differentiate from clusterless Gheg speakers has reinforced this analogy in some forms and resulted in hypercorrection in Tosk. Gheg NT clusters that have recently arisen may also be due to this form of hypercorrection.¹⁹

5. Constraining Albanian nasal-stop clusters. Importantly, NT cluster formation through apheresis, sonority repair, wrong timing epenthesis, analogical epenthesis/hypercorrection, and any phonemic clusters should not generate beyond what exists in Tosk Albanian. Examples (51-55) models a synchronic phonology in Tosk Albanian at the time of these changes. Note that only one nasal-stop cluster is permitted in the examples.

- (51) ‘spider’ /merimangə/ [merimangə] *[mberimangə]
 (52) ‘king’ /mbɾet/ [mbɾet] *[mbɾend]

¹⁸ Camarda (181) positions this as a derivation from Albanian *njeri*, while Miklosich (*Rom. Elemente* 31); Meyer (WZB. 298-299) consider it a borrowing from Latin *honōrem* (cited in Orel 1998: 283-84). A Latin loan is phonologically plausible. A derivation from Albanian *njeri* ‘man, human’ is also phonologically and semantically plausible (man = honor). Common Albanian phrases elude to this connection such as *Nderi Kombit* (Gheg: *Neri Komiti*) ‘honor of the nation, hero’, and *i pa njer* ‘man without honor’ (literally: without a person). For this analysis, the specific etymology is not important, rather that both possibilities position an original single nasal /n/ (Alb. *njer* or Latin *honōrem*), with an epenthetic /d/ in Tosk [nder].

¹⁹ Variation has also arisen in the Gheg dialects with the addition of clusters, albeit minor. One clear example that may be attributable to hypercorrection is found in *Gheg Albanian Reader* by Linda Mëniku. The standard genitive case *gjinore* is written as <gjindore> with a medial cluster -nd- (see Mëniku 2008: xix).

(59) Greek /tʰrɪk-s/ θρίξ ‘hair’; /tʰrɪkʰ-es/ τρίχες ‘hairs’

(60) Sanskrit **phal** ‘burst’; **phalati** (present), **paphala** (perfect)

Other typological similarities include Lyman’s Law in Japanese, which prevents sequential voicing in Japanese (*rendaku*). When two words are compounded, the initial voiceless consonant of the second element voices if it does not already contain a voiced obstruent (examples 61-62 taken from Vance et al. 2021). In this regard, it is similar to Albanian’s OCP constraint only taking effect within the domain of the morpheme.

(61) /asa/ ‘morning’ + /ϕuro/ ‘bath’ → /asa-buro/ *asa+ ϕuro ‘morning bath’

(62) /umi/ ‘sea’ + /kame/ ‘turtle’ → /umi + game/ *umi-kame ‘sea turtle’

A similar phenomenon is also seen in Austronesian and Australian languages. Blust (2012) describes dissimilation patterns that affect geminates or nasal clusters, where there is a dispreference for more than one marked element. When there are two marked elements, one is eliminated, which Blust compares to Grassmann’s Law. In summary, it is possible to constrain the generation of Tosk Albanian NT clusters with the Obligatory Contour Principle. With this evidence the following law can be adopted.

- **Rosenthal’s Law:** In Albanian, only one instance of a nasal-stop cluster is allowed within a morpheme, preserving any clusters belonging to the phonemic word at the expense of clusters generated by subsequent morpho-phonological processes.²³

6. Conclusion. The nasal-stop clusters [mb], [nd], and [ŋg], in Standard Tosk Albanian appear in word-initial, medial, and final positions. By comparing the dialectal variants, we are able to show that the Malsia Madhe dialect preserved word-initial stops from Proto-Albanian and Proto-Indo-European. The word-initial nasal-stop clusters in Standard Tosk Albanian are thus explainable by a rule that eliminated unstressed word initial vowels (16th-18th CE). This caused verbs with a prefix *en* ‘in’ to reanalyze the nasal into the beginning of the verb (undergoing place assimilation with the following stop). Thus, **en-paj* ‘to hold’ > *mbaj* ‘to hold, carry’. The same rule eliminated word-initial vowels from Proto-Albanian and in several loan words. Most cases of stop epenthesis in the word-medial environment are accounted for with the Syllable Contact Law, maximizing the sonority drop at the syllable boundary (*penë* > *pendë*). The medial cluster that could not be attributed to the Syllable Contact Law appears to be a case of wrong timing epenthesis (*ëmbël*). Instances of word-initial nasal-stop clusters that cannot be explained by either of the previous processes, may be attributable to analogy with the apheresis rule, which was highly productive in Albanian. These processes have their nasal-stop clusters limited by the Obligatory Contour Principle. This new constraint named Rosenthal’s law, accounts for the appearance of only one nasal-stop cluster within the domain of a morpheme in Albanian.

²³ Samuel Rosenthal (Oakland University) first proposed the Obligatory Contour Principle (OCP) and Grassmann’s Law as a way to constrain the appearance of nasal-stop clusters in Albanian (p.c. March 2022). For this reason, it is appropriate to name this constraint after him.

References

- Blust, Robert. 2012. One mark per word? Some patterns of dissimilation in Austronesian and Australian languages. *Phonology* 29(3). 355–381. <https://doi.org/j364>.
- Çabej, Eqrem. 2017. *Italoalbanische Studien. Wiener Dissertation aus dem Jahr 1933*. Wiesbaden: Harrassowitz Verlag.
- Dedvukaj, Lindon. 2022. *The Malsia Madhe dialect of Albanian: A living fossil of ancient Europe*. Rochester, MI: Oakland University MA thesis.
- Dedvukaj, Lindon & Patrick Gehringer. 2022. Re-evaluating Albanian's place in Indo-European studies. *Proceedings of the Linguistic Society of America (PLSA)* 7(1). 5178. <https://doi.org/10.3765/plsa.v7i1.5178>.
- Demiraj, Bardhyl. 1997. *Albanische Etymologien*. (Untersuchungen zum albanischen Erbwortschatz). Amsterdam/Atlanta: Rodopi B.V.
- Demiraj, Shaban. 2013. *Gjuha Shqipe dhe Historia e Saj*. Tirana, AL: Onufri.
- Elsie, Robert. 2003. *Early Albania. A reader of historical texts, 11th-17th Centuries*. Wiesbaden: Otto Harrassowitz Verlag.
- Fortson, Benjamin W. IV. 2010. *Indo European language and culture. An introduction* (2nd edn.). Malden, MA: Wiley-Blackwell.
- Gjinari, Jorgji. 1989. *Dialektet e Gjuhës Shqipe*. Tirana, AL: Akademia e Shkencave e RPS të Shqipërisë. Instituti i Gjuhësisë dhe i Letërsisë.
- Hamp, Eric P. 1980. Albanian *është*. In Kathryn Klar et al. (eds.), *American Indian and Indoeuropean studies: Papers in honor of Madison S. Beeler*. 337–46. The Hague: Mouton.
- Hock, H. Hans. 2021. *Principles of Historical Linguistics* (3rd edn). Berlin: De Gruyter.
- Hyllested, Adam & Brian D. Joseph. 2022. Albanian. In Thomas Olander (ed.), *The Indo-European family*, pp. n/a. Cambridge: Cambridge University Press.
- Hysa, Merita. 2019. The beginning of the syllable in Albanian. *Quaderni di Linguistica e Studi Orientali/Working Papers in Linguistics and Oriental Studies* 5. 175–193. <https://doi.org/10.13128/QULSO-2421-7220-25965>.
- Kloekhorst, Alwin. 2008. *Etymological dictionary of the Hittite inherited lexicon*. Leiden: Brill.
- Mëniku, Linda. 2008. *Gheg Albanian reader*. Hyattsville, MD: Dunwoody Press.
- Murray, W. Robert & Theo Venneman. Sound change and syllable structure in Germanic Phonology. *Language* 59(3). 514-528. <https://doi.org/10.2307/413901>.
- Newmark, Leonard, Philip Hubbard & Peter Prifti. 1982. *Standard Albanian. A reference grammar for students*. Stanford, CA: Stanford University Press.
- Orel, Vladimir. 1998. *Albanian Etymological Dictionary*. Leiden: Brill.
- Shkurtaj, Gjovalin. 1967. *E folmja e Kastratit*. Tirana, AL: Studimet Filologjike.
- Shkurtaj, Gjovalin. 1974. *E folmja e Hotit*. Tirana, AL: Dialektologjia Shqiptare II.
- Shkurtaj, Gjovalin. 1975. *E folmja e Kelmendit*. Tirana, AL: Studimet Filologjike.
- Shkurtaj, Gjovalin. 2016. *Dialektet e Shqipes*. Tirana, AL: Albanian University Press.
- Schumacher, Stefan & Joachim Matzinger. 2013. *Die Verben des Altalbanischen. Belegworterbuch, Vorgeschichte und Etymologie*. Wiesbaden: Harrassowitz.
- Vance, J. Timothy, Shigeto Kawahara & Mizuki Miyashita. 2021. The diachronic origins of Lyman's Law: Evidence from phonetics, dialectology, and philology. *Phonology*. 38(3). 479–511. <https://doi.org/10.1017/S0952675721000270>.
- Witczak, Krzysztof T. 2016. The earliest Albanian loanwords in Greek. In (ed. n/a), *1st International Conference on Language Contact in the Balkans and Asia Minor*, 40–42. Thessaloniki: Institute of Modern Greek Studies.
- 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>.