



A Number of Roots Found in the First Periods of Language Establishment and their Manifestation in the Languages of Other Systems

Kubra Guliyeva

Philology, Azerbaijan State University of Economics

DOI: 10.26417/a9bnbk52

Abstract

The Turkic language family is distinguished globally by its profound lexical richness and complex morphological history. Historically, the assumption within general linguistics has been that a significant portion of Turkic vocabulary—particularly terms relating to specialized domains of life, agriculture, and household administration—originated as borrowed lexical items meant to enrich the linguistic fund and expand synonymy. However, rigorous comparative-historical analysis reveals that many lexemes categorized as foreign borrowings, or attributed exclusively to a single language family (such as Indo-European), are in fact deeply embedded within the indigenous Turkic lexicon. This manuscript challenges prevailing etymological assumptions by examining ancient word roots that manifest across multiple, genetically distinct language systems, suggesting an origin in the earliest periods of human language establishment. By exploring the phonetic evolution and semantic dispersion of specific roots—namely *pa-/ba-* (foot/leg), *ter-/der-* (tree/forest), and *id-/iz-* (owner/master)—this paper positions these units as potential Nostratic or macro-family cognates rather than mere areal borrowings. Employing materials primarily from modern Azerbaijani, its dialects, and ancient Turkic monuments, this study juxtaposes these roots against their Indo-European, Uralic, and Dravidian counterparts. The findings advocate for a re-evaluation of lexical attribution in comparative linguistics, emphasizing that shared root bases across languages of different systems are frequently the result of primordial linguistic divergence rather than secondary historical contact.

Keywords: first establishment period of languages, languages of other systems, macro-comparative linguistics, Nostratic theory, similar word-base, Turkic languages, Indo-European languages, comparative-historical method.

Introduction

The identification of identical or structurally analogous word roots across geographically and typologically distinct language families remains one of the most compelling frontiers in modern historical linguistics. A substantial inventory of homophonous and synonymous roots exists across languages of different systems. Traditionally, these parallels have been explained away through the lens of contact linguistics as historical borrowings (loanwords) resulting from trade, conquest, or cultural exchange. However, a deeper chronostratigraphic analysis suggests that these lexical parallels are the surviving remnants of parallel usage dating back to prehistoric epochs. The historical development of these words, their semantic convergence or divergence, and their varied derivational trajectories compel scholars to ask a fundamental question: to which foundational proto-language do these pervasive roots originally belong?

The presence of shared root morphemes across disparate language families can be interpreted as a natural consequence of primordial language divergence. Within the framework of macro-comparative linguistics, such forms are often designated as "Nostratic" or "Eurasianic" roots—terms referring to hypothetical macro-families encompassing Indo-European, Uralic, Altaic (including Turkic), Dravidian, and Kartvelian languages (Illich-Svitych, 1971; Greenberg, 2000). Historical and archaeological research indicates that over a developmental span of 35,000 to 40,000 years, early human language grew increasingly complex. By the subsequent millennia, driven by demographic expansion and massive migratory waves, these early proto-languages began to fracture (Pletneva, 1976:32). These gradual disruptions birthed distinct proto-languages that differed fundamentally in their initial phonetic form and morphological structure, ultimately crystallizing into the language families recognized today. The preservation of cognate roots across languages of vastly different systems can logically be explained by this shared, albeit remote, genetic ancestry.

Furthermore, as prominent foundational linguists such as Wilhelm von Humboldt and August Schleicher have theorized, the migratory patterns of ancient tribes facilitated complex processes of language convergence and divergence. As tribes dispersed across Eurasia, the resulting geographic isolation and subsequent secondary contacts triggered profound typological shifts. Languages evolved distinct morphological

profiles, separating into isolating (amorphous), incorporating, fusional (inflectional), agglutinative, analytic, and synthetic systems (Yarkova, 1961:109-116).

Acknowledging these diachronic processes allows contemporary linguists to accurately trace the etymological origins of universally retained root bases. This paper provides a comparative explanation of several foundational roots used in parallel forms and synonymous meanings across Turkic languages and language families of other systems. The linguistic data is primarily drawn from the modern Azerbaijani language, its rich dialectal varieties, and broader ancient Turkic materials, analyzed alongside Indo-European and other Eurasian language data.

Theoretical Framework: Distinguishing Cognates from Borrowings

Before embarking on specific etymological analyses, it is necessary to establish the methodological criteria used to differentiate between genetic cognates (words inherited from a common ancestral language) and loanwords (words acquired through language contact).

Turkic languages are inherently characterized by the expansive richness of their lexical vocabulary. Throughout their history, Turkic languages have absorbed terminology related to statecraft, religion, and the arts from Persian, Arabic, and later Russian. These borrowed words were assimilated to enrich the lexicon and expand semantic nuance. Because of this well-documented history of lexical receptivity, early Turcologists and Indo-Europeanists often defaulted to classifying any shared vocabulary between Persian (an Indo-European language) and Turkic as a Persian borrowing into Turkic.

However, this unidirectional assumption systematically overlooks the indigenous derivational productivity of the Turkic root. If a root is highly productive in a language—yielding dozens of derivative nouns, verbs, and adjectives—it is highly probable that the root is native to that language, or at least inherited from a prehistoric macro-family, rather than a recent borrowing. In many cases, words currently accepted as the product of a foreign language are actually universal lexical units belonging simultaneously to many languages of the world. To demonstrate this, we will analyze three highly productive roots: *pa-/ba-*, *ter-/der-*, and *id-/iz-*.

The Root *pa-* / *ba-* (Foot / Leg / Lower Extremity)

The word *paça* is found in modern Azerbaijani, as well as in the vast majority of contemporary Turkic languages. It is utilized to denote "the part of the leg from the knee to the hip" or, more broadly, the lower limbs of an animal. Traditional lexicography almost universally categorizes this word as a borrowing of Persian origin. For instance, the monumental works of early comparative linguistics, including

V.V. Radlov's *Dictionary Practice of the Turkic Adverbs*, L.Z. Budagov's *Comparative Dictionary of Turkic-Tatar Adverbs*, and E.V. Sevortyan's *Etymological Dictionary of Turkic Languages*, all record *paça* as a Persian loanword. The standard morphological breakdown provided by these scholars traces the word to the Persian root *pa* (meaning "foot" or "leg") combined with the Persian diminutive suffix *-ça*.

While it is undeniable that the Persian root *pa* means "foot" and that the word *piyada* (pedestrian)—used in modern literary Azerbaijani—is derived directly from this Persian root, this explanation is incomplete when applied to the broader Turkic lexicon. A critical sociolinguistic principle dictates that core vocabulary, particularly words designating fundamental body parts and basic household items, are highly resistant to borrowing. Such lexemes are typically created indigenously.

Operating from this methodological standpoint, there is substantial evidence that the root *ba-* (meaning "foot" or "step") was an active, indigenous morpheme in ancient Turkic languages. At present, tens of words derived from this root are actively preserved across the Turkic language family, distinctly visible in the modern Azerbaijani literary language, as well as in its regional dialects and accents.

In Azerbaijani dialects, we find a highly productive cluster of words derived from this same *ba-* root:

- *badax* (to pull the foot / to trip)
- *badağlamaq* / *badamaq* (to tie the feet of an animal)
- *badalaq* (to tangle legs in wrestling, a tripping move)
- *badıc* (high-socks covering the lower leg)
- *badış* (socks worn on the calf without a pedestal)
- *papış* (knitted house shoes / slippers) (ADDL, 1964:49-51)

Furthermore, foundational body-part and footwear terminology such as *bazu* (arm/limb), *barmaq* (finger/toe), *başmaq* (shoe), and *baldır* (calf) demonstrate a morphological reliance on a primordial *ba-* root. In ancient Turkic languages, lexemes such as *başa* (to set legs), *başak* (shoes), and *başğak* (the part of the leg from hip to knee) emerged directly from this source (Clouston, 1972). Modern Turkish words like *pabuç* (shoe) and *bacak* (leg) also trace their origins back to this archaic root.

The historical phonological development of the word *paça* within Turkic languages likely occurred in two parallel evolutionary tracks, stabilizing according to the following phonetic shifts:

1. *başğaq* → *baçak* → *bacak* → *paçak* → *paça*
2. *başğak* → *baçak* → *bacak* → *badak* → *adak* / *azak* / *atak* → *ayak* / *ayaq*

Given this immense derivational depth, it is illogical to conclude that the Turkic peoples borrowed a Persian root for their basic anatomical terminology. Instead, we must conclude that the root *pa-* / *ba-* is of common Nostratic origin. It existed not only in the Indo-Iranian branch of Indo-European but as a universal base. This is overwhelmingly supported by its phonetic manifestations across multiple distinct language families: the reconstructed Proto-Indo-European root **ped-* / **pod-* yields the French *pié* / *pe* (Azerbaijani-French Dictionary, 2007:63), the German *Fuß* / *be* (Russian-German Dictionary, 1981:230), the Persian *pa* (Russian-Urdu Dictionary, 1958:388), the English *foot* / *fe* (English-Russian Dictionary, 1971:218), and the Romanian *picior* / *pi* (Russian-Romanian Dictionary, 1954:473). The widespread distribution of the *p-* / *b-* / *f-* labial initial for "foot" indicates a pre-historical macro-family connection rather than a localized medieval borrowing.

The Root *ter-* / *der-* (Tree / Forest / Support)

Another lexical unit that bridges the Turkic languages with languages of other systems is the ancient word *tereğ* (and its variant *terek*). Preserved primarily in the archaic lexemes and dialects of the Azerbaijani language, this word denotes the phenomenon of a "tree" or "wood."

The semantic and phonetic parallels of *tereğ* in antiquity are staggering. The root appears in Sumerian as *taru* (meaning "forest"); in ancient Sanskrit/Vedic Indian as *daru* or *dru* (meaning "wood" or "tree"); in the Zoroastrian *Avesta* as *dauru* / *dru*; in Gothic as *triu*; in modern English as *tree*; in Russian as *derevo*; and in Anglo-Saxon with the specialized meaning of "gum tree" as *teru* (Musayev, 1968:19). Modern Indo-European linguistics universally reconstructs this back to the Proto-Indo-European root **deru-* or **doru-* (meaning "tree" or "wood," yielding derivatives associated with steadfastness and truth).

Ancient written Turkic monuments reveal that at a specific stage of linguistic development, the semantic scope of *terek* was restricted in certain regions, shifting from the generalized meaning of "tree" to the highly specific "poplar" (*qovaq*) (DTS, 1969:533). However, today, in a number of Turkic languages (e.g., Kumyk, Karachay-Balkar), the word *terek* / *dereg* has retained or regained its original semantics and is used as the standard literary synonym for "tree."

The root *tir-* / *ter-* has remarkably preserved its phonetic integrity and core semantics across the Eurasian landmass. Its presence in the oldest written monuments of the Turkic peoples (such as the Orkhon inscriptions) attests to its absolute antiquity

within the Turkic continuum. Analyzing this wide distribution, K. Musayev concludes that while the word *terek* is formally cohesive across Turkic languages (from a lexical standpoint), semantic shifts (e.g., "tree" vs. "poplar") effectively categorize the Turkic languages into distinct dialectal and historical groups (Musayev, 1984:142-143).

The directionality of transmission has been a subject of intense academic debate. K. Musayev posits that the Persian word *derext* (tree) is a borrowing from the ancient Turkic languages. Conversely, R. Akhmetyanov argues the opposite, suggesting an Indo-Iranian origin that penetrated Turkic (Akhmetyanov, 1980:87-95). However, when we apply the criterion of derivational productivity, the argument for an indigenous or deeply inherited Turkic root strengthens. The Turkic languages exhibit immense derivational capacity stemming from the *tir-/ter-* root, implying native mastery.

Currently, a vast array of lexemes has been generated from this single root across the Turkic family:

- *tirə-* (verb: to set, to prop up)
- *tirəqu* (noun: column, support)
- *tereg* (noun: a thing used for support)
- *tereqlik* (noun: birch forest)

In the modern Azerbaijani language specifically, the root manifests in vital structural and positional vocabulary: *dirək* (pole/pillar), *dirsək* (elbow—the body's structural pivot), *tir* (beam/joist), *darağacı* (gallows/wooden frame), *dorağacı* (mast), *tirlənmək* (to stiffen), *durmaq* (to stand / to be established), *dirəmək* (to prop up), *dırmaşmaq* (to climb), and *dirənmək* (to resist / stand firm). The semantic transition from "tree/wood" to "pillar/support" and ultimately to verbs denoting "standing" and "resistance" reflects a profound, indigenous cognitive linguistics paradigm that could not have been achieved through the superficial borrowing of a Persian noun. Therefore, the *ter-/der-* root must be recognized as an archetypal, Nostratic unit.

The Root *id-* / *iz-* / *ir-* (Owner / Master / Lord)

A third compelling example of macro-comparative resonance is the root *idi* (or *idi* / *iye* in various Turkic reflexions). This lexeme is widespread across the Turkic languages and carries the foundational meaning of "sahib" (owner, master, lord, or possessor). While it has assumed subtle semantic variations throughout Turkic history—sometimes denoting a spiritual master or a tutelary deity of a specific locale (e.g., the *yer-su* spirits)—it stabilized overwhelmingly in the sense of "owner."

The reach of this root extends far beyond the Altaic sphere, appearing in ancient Uralic, and significantly, in Dravidian languages. Linguists have heavily debated the phonological shifts of this root, particularly regarding the $z \sim y$ and $d \sim y$ sound coincidences observed in comparative Indo-European and Altaic studies.

V. Bang was among the first to systematically analyze this term, noting that the word inherently accepts possessive suffixes, suggesting that the absolute root of the word is simply *id* (Bang, 1921:2). Building upon this, K. Menges synthesized monument inscriptions, historical dictionaries, and comparative data to suggest an even broader East Asian connection. Menges proposed that the root *id* shares a Mongol origin and potentially established early links with the Chinese linguistic sphere (Menges, 1976:104).

In his seminal paper entitled *"Turkic İdl 'Sir', its some reflexes in the Turkic languages and parallels in other language families,"* Menges traces the existence of this term into the Korean and Japanese languages, hypothesizing a genetic or areal link to the Korean word *ezen* (master/lord). Furthermore, when examining the $t \sim r$ consonant coincidences highlighted by scholars of Common Dravidian and Proto-Dravidian (such as Zvelebil), a compelling case emerges that, driven by regular phonetic laws, this root shares a genetic semantic core with titles of lordship across the Eurasian continent.

Synthesizing the theoretical models proposed by Menges, Baskakov, Ramstedt, Cincius, and Zvelebil (Starostin et al., 2003), it becomes evident that the phonological variations *id* $\sim$ *idz* $\sim$ *iz* $\sim$ *ir* $\sim$ *it* all stem from a single, primordial word. These are merely dialectal reflexes that solidified as different language families branched off from the macro-family tree. As the process of language separation accelerated during the Neolithic period, each linguistic branch adopted and standardized a specific phonetic variant. The subsequent phonetic forms of the word encountered in later written records are the natural result of standard diachronic phonological shifts (such as rhotacism and lambdacism, which are heavily documented in Altaic linguistics).

Conclusion

The evidence presented regarding the lexical roots *pa-/ba-*, *ter-/der-*, and *id-/iz-* systematically challenges the traditional approach of classifying shared Eurasian vocabulary strictly as localized loanwords. By analyzing the immense derivational productivity of these roots within the Turkic languages—manifesting in hundreds of nouns, verbs, and adjectives related to basic anatomy, fundamental flora, and core societal concepts—it is clear that these are not superficial borrowings.

Instead, the striking phonetic and semantic parallels found in Indo-European, Uralic, Dravidian, and Altaic languages point toward an incredibly ancient, shared linguistic heritage. These are remnants of the first periods of language establishment, echoing from a time prior to the massive geographical dispersals that fractured the Nostratic macro-family.

Today, the advanced development of historical and comparative linguistics, augmented by computational models and cross-disciplinary archaeological data, both allows and obliges us to conduct research that transcends rigid language-family boundaries. In an era defined by modern globalization and integration processes, returning to the primordial roots that once united disparate human populations is of profound scientific and philosophical importance. This trajectory of historical-comparative research will continually contribute to uncovering the universal archetypes utilized across languages of different systems, ultimately revealing the deeply interconnected nature of human linguistic evolution.

References

- [1] Pletneva, S. A. (1976). *Khazars*. Moscow.
- [2] Yarkova, V. N. (1961). On the principles of determining the morphological type of the language. *Morphological typology and the problem of classification of languages*. Moscow.
- [3] *Dialectological dictionary of Azerbaijani language*. (1964). Baku.
- [4] *Azerbaijani-French dictionary*. (2007). Baku.
- [5] *Russian-German dictionary*. (1981). Moscow.
- [6] *Russian-Urdu dictionary*. (1958). Moscow.
- [7] *English-Russian dictionary*. (1971). Moscow.
- [8] *Russian-Romanian dictionary*. (1954). Moscow.
- [9] Musayev, K. M. (1968). Some issues of the lexical-semantic differentiations and integration of the Turkic languages. *Turkological researches*, Moscow.
- [10] Akhmetyanov, R. G. (1980). "Forest", "tree" and "bars" of the Turks. *Sovetskaya Tyurkologiya (CT)*, No. 5.
- [11] Bang, W. (1921). Vom Köktürkischen zum Osmanischen, 4. *Abhandlungen der Preußischen Akademie der Wissenschaften (APAW)*, No. 2.
- [12] Menges, K. G. (1976). Turkic İdl "Sir", its some reflexes in the Turkic languages and parallels in other language families. *Turcologica*, Leningrad.

- [13] Illich-Svitych, V. M. (1971). *Opyt sravnenija nostraticheskix jazykov* [An attempt at a comparison of the Nostratic languages]. Nauka.
- [14] Clauson, G. (1972). *An Etymological Dictionary of Pre-Thirteenth-Century Turkish*. Oxford University Press.
- [15] Greenberg, J. H. (2000). *Indo-European and Its Closest Relatives: The Eurasiatic Language Family* (Vol. 1). Stanford University Press.
- [16] Starostin, S., Dybo, A., & Mudrak, O. (2003). *Etymological Dictionary of the Altaic Languages*. Brill.