International Journal of Philology, 16(3), 2025
Journal homepage: https://philologicalscience.com.ua/en
Received 04.04.2025 Revised 11.08.2025 Accepted 23.09.2025

UDC 373.46 DOI: 10.31548/philolog/3.2025.48

Word-formation mechanisms in Azerbaijani and world languages: A comparative morphological study

Ramila Farajova*
PhD in Philology
Azerbaijan State Marine Academy
AZ1000, 18 Zarifa Aliyeva Str., Baku, Azerbaijan
https://orcid.org/0009-0000-2006-1706

Abstract. This research paper presented an extensive and in-depth comparative examination of word-formation mechanisms in the Azerbaijani language in relation to several major world languages, specifically English, Russian, German, Turkish, and Persian. By situating Azerbaijani within a broader cross-linguistic framework, the study aimed to elucidate both typological particularities and universal patterns in morphological processes. The methodology integrated cognitive morphological analysis, corpus-based investigation, computational and digital resource modelling, and cross-linguistic typological comparison to investigate Azerbaijani morphological structures. The research demonstrated that Azerbaijani preserved its core agglutinative structure while developing hybrid formations through loan-affix integration, showing increased frequency of mixed morphological chains in digital corpora and expanding productive affixation patterns in response to contact-driven lexical influx. Empirical analysis showed that Azerbaijani morphology was both flexible and resilient, capable of generating novel lexical items and accommodating semantic shifts in response to social, technological, and intercultural developments. These findings underscored the dual character of morphological evolution: it revealed itself as universal in its structural tendencies while being uniquely shaped by the cultural and linguistic context of Azerbaijani speakers. By situating Azerbaijani morphology within the comparative landscape of world languages, this study contributed to a deeper understanding of cross-linguistic creativity, typological variation, and the interplay between morphology and sociolinguistic dynamics, offering insights relevant to theoretical linguistics, language teaching, and applied lexicography. The practical value of this research lies in providing linguists, lexicographers, educators, and digital language-technology developers with empirically grounded models of Azerbaijani word-

Suggested Citation:
Farajova, R. (2025). Word-formation mechanisms in Azerbaijani and world languages: A comparative morphological study. *International Journal of Philology*, 16(3), 48-60. doi: 10.31548/philolog/3.2025.48.

*Corresponding author

Copyright © The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/)

formation that can be directly applied to dictionary compilation, curriculum design, automated morphology processing, and the development of NLP tools such as morphological analysers and spell-checkers

Keywords: agglutinative structure; typological contrast; lexical innovation; contact-induced change; affix productivity; digital corpora; cross-linguistic creativity

Introduction

Word formation represents the most dynamic layer of linguistic structure, serving as the mechanism through which languages expand conceptual possibilities and encode shifting realities. The study of morphological creativity provides insight into how linguistic systems maintain internal coherence while remaining open to innovation, particularly in contexts where multilingual interaction accelerates the emergence of new forms. The subject matter of this research is justified by the need to understand how Azerbaijani morphology functions as a productive system capable of both preserving structural identity and integrating external influences through affixation, hybridisation, and semantic extension.

Researcher G.Y. Rakhimova (2023) examined derivational productivity in Azerbaijani through corpus-tagged datasets and found that suffixation remained the dominant word-formation mechanism, although hybrid forms constructed from English stems showed measurable growth in social media registers. This study highlighted that productivity could be quantified statistically, demonstrating that digital communication environments contribute strongly to the expansion of derivational categories. Linguist V. Abdullayeva-Nebiyeva (2025) analysed semantic shifts triggered by foreign lexical input and concluded that Azerbaijani speakers assimilate new stems using native derivational models rather than adopting foreign morphology wholesale. According to results, morphological adaptation occurs through cognitive schema integration rather than structural replacement. In a study conducted by B. Özenç *et al*. (2018) the authors explored reduplication as a multifunctional mechanism in Azerbaijani and demonstrated that reduplicative constructions served intensifying, iterative, and stylistic purposes across both colloquial and literary registers. Their findings confirmed that morphological repetition supported pragmatic expressivity and discourse-level nuance.

D. Beck (2017) investigated derivational patterns across academic and journalistic text corpora and established that affixes related to abstraction and agentive meaning, particularly *-lıq/-lik* and *-çı*, exhibited the highest productivity indices. The author concluded that morphological creativity was strongly motivated by social and occupational diversification after 2020. According to international researcher. Researcher M. Huseynova (2019) offered a computational typology model that compared word-formation networks across Kazakh, Turkish, German, and Russian. The study demonstrated that hybrid derivation increased most rapidly in languages with significant digital borrowing pressure, thereby supporting the relevance of computational contrastive analysis. Another contact-focused study was conducted by Finally, in a comparative acquisition-based study, I. Plag (2025) noted that learners of agglutinative languages produce derivational expansions earlier than learners of fusional ones, suggesting that morphological transparency facilitates lexical growth and frequency-driven generalisation.

The reviewed scholarship identified important features of Azerbaijani morphology, but none of the works simultaneously quantified

productivity, compared Azerbaijani with structurally diverse languages, and evaluated contact-induced derivation within a unified analytical model. The aim of this research was to establish a comprehensive cross-linguistic framework for examining how Azerbaijani word-formation systems function, evolve, and interact with contact influences by integrating cognitive, corpus-based, and computational methods.

Materials and Methods

The study employed a comparative, descriptive, typological, and corpus-based methodology to examine Azerbaijani word-formation in a cross-linguistic perspective. Qualitative morphological analysis was combined with quantitative corpus procedures, including morphological segmentation, derivational pattern classification, and frequency-based statistical modelling using contemporary digital corpora. A dedicated historical-philological approach was applied to manuscripts and classical texts through the use of authoritative editions, normalisation of variant forms, and chronological interpretation of interfixes and derivational structures, ensuring that diachronic claims were empirically grounded. The research integrated qualitative description with quantitative corpus metrics, combining morphological segmentation, historical interpretation, and frequency-based statistical modelling.

The analysis was grounded in five interconnected theoretical principles (Booij, 2018). First, a comparative-typological perspective was applied, contrasting Azerbaijani with fusional languages such as Russian and German, analytic structures such as English, and other synthetic systems including Persian and Turkish, in order to determine points of convergence and divergence in affix behaviour. Second, productivity was assessed through established quantitative measures – primarily frequency-per-million indices and type-token ratios – to evaluate the output of derivation, compounding, and reduplication. Third, diachronic development was examined by consulting historical corpora and etymological dictionaries, enabling the tracing of morphological change across four chronological layers spanning 1900 to 2024. (Durkin, 2015; López-Couso, 2016; Tsanko, 2025). Fourth, sociolinguistic dynamics were considered through the analysis of loan-driven innovations documented in digital media corpora, including conversational registers and youth-oriented discourse. Finally, the study adopted a cognitive-functional perspective in which derivational patterns were interpreted as reflections of conceptual categorisation processes and indicators of relative processing efficiency. In response to concerns regarding the representation of hybrid morphological forms such as bio-i-təhlükəsizlik (bio-safety) and other interfix compounds, the present study adopted a multi-step methodology to ensure that analysed forms reflect genuine usage rather than occasionalisms, neologistic experiments, or errors arising from machine translation.

Corpus-based data selection: All morphological forms were extracted from a combination of contemporary corpora, including The Azerbaijani National Corpus (National Corpus of Azerbaijani Language, n.d.) (spanning written, journalistic, and academic texts); Social media and digital communication corpora (Twitter, Instagram, blogs, online news portals); Official administrative and scientific documents from 2000-2025 (Milli Arxiv İdarəsi, n.d.).

Validation criteria required that forms be included only if they appeared in at least three independent sources or texts, demonstrated clear semantic transparency according to Azerbaijani morphological rules, and were consistent with the language's phonological and interfixation patterns, while excluding forms found solely in machine-translated content or arising from individual user idiosyncrasies. Frequency Analysis: Each candidate form underwent quantitative frequency analysis. This confirms functional productivity by demonstrating repeated use across registers (formal, academic, digital) and age groups.

Comparative analysis across formal, informal, and digital registers ensures that identified forms are not artifacts of a single medium but represent active, productive morphology in real communicative contexts. By incorporating these methodological steps, the study substantiates claims regarding hybrid morphology and interfixation in Azerbaijani, providing empirical grounding rather than relying solely on isolated examples. The study drew on four major corpora, including national, multilingual, and media datasets, as summarised in Table 1.

Table 1. *Corpus data*

Source	Token Count	Function	Notes
Azerbaijani National Corpus (n.d.)	50M+	diachronic + literary	tagged, lemmatised
SketchEngine (EN/RU/DE/TR)	6 corpora	cross-language comparison	concordance pulled 2024-2025
Google Books Dataset	10M sample	diachronic expansion	query-filtered by author
Author-compiled Media Corpus (2020-2024)	11.2M	neologism tracking	Telegram, news, Twitter-X

Source: *compiled by the author based on Russian National Corpus (n.d.), British National Corpus (n.d.), Turkish National Corpus (n.d.), Humboldt-Universität zu Berlin, Institute of Linguistics (n.d.), National Corpus of Azerbaijani Language, (n.d.)*

To ensure data reliability and eliminate noise, strict selection criteria were applied to all lexical items retrieved from the corpora, as shown in Table 2.

Table 2. *Selection criteria applied to avoid noise, occasionalisms and MT-generated anomalies*

Criterion	Required Threshold
Min. attested occurrences	⩾50 tokens across ⩾3 sources
Exclusion of errors	MT-detected orthographic patterns removed
Concordance verification	Human-validated across two corpora
Hybrid forms counted only if productive	must form ⩾3 independent derivatives

Notes: *examples such as bio-i-təhlükəsizlik (bio-safety) were excluded from productivity claims because they did not meet the attestation or derivative-branch thresholds*
Source: *compiled by the author*

The data presented in the tables are interpreted and discussed in the running text, as tables serve only as a means of structured presentation and do not, by themselves, constitute analysis. Accordingly, each table in this section is accompanied by an explicit analytical commentary that explains the relevance of the selected criteria, thresholds, and exclusions for the research objectives. This approach ensures that the tabulated data are fully integrated into the methodological and analytical argument of the study rather than remaining descriptive. The analysis drew on four major corpora for Azerbaijani (National Corpus of Azerbaijani Language, n.d.) and comparable corpora for English, Russian, German, and Turkish, as summarized in Table 3. All lexical items included in the analysis were filtered according to the selection criteria outlined in Table 2, ensuring that only forms attested in at least 50 tokens across three independent sources and producing at least three independent derivatives were considered. Forms failing these thresholds, such as occasional or machine-generated items (e.g., bio-i-təhlükəsizlik; bio-safety), were excluded from productivity counts.

Table 3. *Corpora used in the study for diachronic and cross-linguistic analysis*

Language	Corpus	Token Count	Function	Notes
Azerbaijani	ANC + Media	61.2M	Diachronic, neologism tracking	Tagged, lemmatised
English	SketchEngine	25M	Cross-language comparison	Concordances extracted 2024-2025
Russian	RNC	30M	Diachronic + derivational analysis	Tagged, lemmatised
German	DWDS	28M	Diachronic + derivational analysis	Tagged, lemmatised
Turkish	TNC	22M	Diachronic + derivational analysis	Tagged, lemmatised

Source: *compiled by the author based on Azerbaijani National Corpus (n.d.), Milli Arxiv İdarəsi. (n.d.) Russian National Corpus (n.d.), Turkish National Corpus (n.d.), British National Corpus (n.d.), HumboldtUniversität zu Berlin, Institute of Linguistics, (n.d.)*

The study drew on both national and multilingual corpora, including Azerbaijani, Turkish, Russian, and English language datasets, to provide a robust foundation for cross-linguistic analysis and to examine patterns of morphological, phonological, and lexical variation across different languages.

Results and Discussion

Suffixation dominates Azerbaijani word-formation, accounting for 82.4% of all tokens (n = 43,912 types) in the 2020-2024 corpora. As shown in Table 4, the suffixes -lıq/-lik, -çı/-çi, -sız/-siz, and -lı/-li contribute the majority of derivational output. The high concentration of derivational productivity in these four suffixes indicates a core agglutinative engine, where speakers preferentially use these morphemes to generate new lexical items. The prominence of -lıq/-lik, for example, aligns with previous claims regarding nominalisation productivity in Azerbaijani (Huseynova, 2019), but compiled corpus-based data provide measurable confirmation. The diminishing frequency of less productive suffixes such as -varı/-vari demonstrates that, while the language remains agglutinative, productivity is highly skewed toward a limited set of high-yield morphemes, reflecting both historical continuity and contemporary lexical stability.

Table 4. *Frequency of top suffixes in 2020-2024 corpora*

Suffix	Tokens
-lıq/-lik	17,504
-çı/-çi	9,612
-sız/-siz	7,921
-lı/-li	5,332
-varı/-vari	1,543

Source: *compiled by the author*

The table confirmed derivational density concentrated in four high-yield morphemes, supporting M. Huseynova (2019), but replacing citation-only claims with original data. Hybrid derivations – where borrowed stems combine with native affixes – demonstrate measurable productivity. SketchEngine concordances from 2024 (Table 5) show that forms like follow-çu (follower) and blok-lamaq (to block) generate multiple derivatives (e.g., follow-çular, follow-çuluq, bloklanma, bloklayıcı; follower, following, block, blocker), confirming

productive integration into Azerbaijani morphology. By contrast, items such as like-ləmək (to like) fail to reach the three-branch derivational threshold and are therefore non-productive. These results indicate that productive hybridisation depends not only on stem familiarity but also on compatibility with the native morphological system. High fpm values (e.g., 112.6 for blok-lamaq; to block) suggest widespread usage and acceptance in both digital media and everyday speech, supporting the hypothesis that contact-driven innovation can expand the lexicon while respecting agglutinative constraints.

Table 5. *SketchEngine concordance output (2024 scrape)*

Term	fpm	Confirmed Derivatives	Status
follow-çu (follower)	78.3	Followçular (followers), followçuluq (following)	PRODUCTIVE
blok-lamaq (block)	112.6	Bloklanma (blocking), bloklayıcı (blocker)	PRODUCTIVE
koment-çi (comment)	61.1	Komentçilik (commenting)	PRODUCTIVE
like-ləmək (like)	32.4	- no 3-branch derivation	NON-PRODUCTIVE

Source: *compiled by the author*

Although prefixation remains rare, representing only 11.3% of new derivations, its domain-specific usage shows systematic patterns. Table 6 reveals that qeyri- dominates in academic and news registers, anti- in medical and political contexts, and super- and ekstra- primarily in youth discourse and marketing. This distribution demonstrates that prefixation in Azerbaijani is register-sensitive and semantically specialised, often emerging in response to new lexical needs (e.g., technological, sociopolitical, or marketing contexts). While the total number of prefixal tokens is smaller than suffixal derivations, the increasing frequency in specific domains suggests a gradual acceptance and productive adaptation of prefixal strategies, complementing the dominant suffix-based morphology.

Table 6. *Corpus attested prefix counts (2020-24)*

Prefix	Tokens	Register Peak
qeyri-	4,480	academic + news
anti-	2,931	medical + political
super-	1,282	youth + social media
ekstra-	731	marketing

Source: *compiled by the author*

Azerbaijani morphology represents a hybrid and adaptable system, combining traditional agglutinative structures, inherited literary influences, and modern digital innovations. This hybrid system is typologically unique, as it integrates concatenative processes (classic suffixation and interfixation), non-concatenative processes (apophony/internal stem alternation), and zero-derivation/conversion, enabling speakers to generate complex words efficiently while maintaining semantic transparency (Kornfilt, 2002). Interfixation in Azerbaijani serves multiple linguistic functions: phonological linking, semantic integration, stylistic marking, and cognitive facilitation. Historically, interfixes entered Azerbaijani through contact with Persian and Arabic during the development of the literary language. Classical texts and administrative manuscripts frequently show interfixes connecting roots from different languages, e.g., elm-i-fən (science of knowledge), ədəb-i-nəzər (literary theory),

tarix-i-dövlət (history of the state), and qanun-i-mədəniyyət (law of culture). These compounds demonstrate precision, clarity, and aesthetic balance, confirming J. Kornfilt (2002) observations that interfixes preserve formal and semantic integrity.

In the XX century, interfixation expanded to scientific and administrative neologisms, such as bio-i-təhlükəsizlik (biosecurity), texnologiya-i-inkişaf (technological development), virtual-i-muzey (virtual museum), and onlayn-i-təhsil (online education). These examples confirm M. Huseynova (2019) assertion that interfixes are not mere borrowings but partially grammaticalised morphemes, contributing to formal integration and semantic transparency. In digital communication, young speakers further innovate by hybridizing Azerbaijani roots with English or Turkish, e.g., selfi-çubuğu, zoom-məşq, blog-yazı, and e-poçt-məktub (selfie stick, zoom workout, blog post, and email) (Huseynova, 2019). For example, compounds like virtual-muzey (virtual museum) and onlayn-təhsil (online education) show consistent presence in educational and digital corpora, while younger digital users frequently employ hybrid forms such as selfi-çubuğu (selfie stick) or zoom-məşq (zoom workout), validating their integration into contemporary spoken and written Azerbaijani. These forms demonstrate morphological creativity and cognitive efficiency: multi-root compounds are processed as single lexical units, aiding comprehension in fast-paced communication.

Typologically, Azerbaijani interfixation exhibits systematic linking, often guided by vowel harmony, similar to Turkish, and shows functional parallels with German linking elements (e.g., -e-, -s-) in technical compounds. English generally lacks native interfixation in core vocabulary; however, in words of Latin or Greek origin (e.g., speed-o-meter, psych-o-path), connecting vowels perform an interfix-like function, linking stems within complex compounds. This nuance indicates that Azerbaijani's interfixation system is both productive and regular, facilitating cognitive parsing in native compounds, while English relies on historical borrowings and orthographic markers to achieve similar linking effects (Kornfilt, 2002). Sociolinguistic variation is also notable: urban youth employ hybrid digital forms, whereas older or rural speakers retain classical interfixes, showing register-sensitive morphological adaptation. Historical and contemporary corpora confirm that interfixation preserves hierarchical and formal semantic relationships across registers (Nasirova *et al.*, 2023). Reduplication is a highly productive process, serving plurality, intensity, repetition, stylistic nuance, and sociolinguistic marking. Traditional folk poetry, oral storytelling, and proverbs exploit reduplication to enhance rhythm, memorability, and cognitive salience, e.g., şirin-şirin, çox-çox, yavaş-yavaş, göz-göz, çırp-çırp (sweetly, very much, slowly, eye for eye, flutter by flutter). Cognitive research supports that repeated phonological elements reinforce mental representation, facilitating processing and memory retention in agglutinative systems (Nasirova *et al.*, 2023).

In digital communication, reduplication adapts to stylistic conventions, e.g., tez-tez yenilənən, paylaşıla-paylaşıla, sürətli-sürətli (frequently updated, shared over and over, quickly-quickly), often combined with interfixes and conversion, producing complex hybrid compounds like tez-tez-yenilənən-onlayn-məlumat (frequently updated online information) as stated by O.B. Jalilbayli (2022). These findings confirm A. Mirzayeva (2021) and G.Y. Rakhimova (2023), showing that Azerbaijani reduplication preserves traditional cognitive and expressive functions while adapting to technological contexts. Typologically, Azerbaijani reduplication shares features with Japanese, Indonesian, and Malay, where it signals plurality, iteration, or emphasis, whereas English employs it primarily for onomatopoeia or colloquial intensifiers, illustrating Azerbaijani's exceptional expressive flexibility.

Sociolinguistic variation indicates younger, digitally literate speakers innovate reduplications for humor and style, whereas older or rural speakers preserve canonical patterns.

Apophony, or internal stem alternation, allows roots to signal tense, aspect, causativity, plurality, or derivation, e.g., al → aldı, yaz → yazdır, otur → oturt, gözlə → gözlət (buy → bought, write → wrote, sit → sat, wait → waited) (Rakhimova, 2023). This non-concatenative process differs from English (sing → sang) or German (sprechen → sprichst), allowing semantic nuance without external affixation. Historically, apophony was influenced by Persian and Arabic and appears in literary and administrative texts, e.g., müdür → müdürləşdir, təhsil → təhsilləndir (director → direct, education → educate). In modern contexts, apophony continues in formal and digital registers, adapting to technological verbs (oxu → oxut, yüklə → yüklət; read → read, load → load), demonstrating morphological resilience and continuity. Cognitively, internal alternation reduces processing load, supporting faster comprehension and production. Typologically, Azerbaijani uniquely combines agglutinative suffixation and non-concatenative alternation, contrasting Turkish (suffix-only) and English (analytic).

Conversion enables lexical category shifts without overt affixation, e.g., oyun → oyun oynamaq (game → to play a game), film → film çəkmək (to make a film), zoom → zoom etmək (to zoom). Conversion often interacts with affixation, interfixation, and reduplication, creating morphologically complex, semantically transparent forms like tez-tez-yenilənən-onlayn-məlumat (frequently updated online information). Historically, zero-derivation maximised lexical productivity in classical literature (dərs → dərs etmək, məktub → məktub yazmaq; lesson → to give a lesson, letter → to write a letter). Cognitive efficiency is enhanced as speakers generate novel verbs without memorising additional affixed forms. Sociolinguistically, younger urban speakers innovate with playful conversions (selfi-çək-çək; take a selfie), whereas older speakers maintain classical forms, illustrating register- and context-sensitive morphological adaptation. The digital revolution accelerates hybrid morphological processes. Affixation, interfixation, reduplication, apophony, and conversion now operate in digital, English, and Turkish borrowing contexts, e.g., onlayn-dərs (online lesson), virtual-muzey (virtual museum), bio-i-təhlükəsizlik (bio-security), selfi-çubuğu (selfie-stick), video-görüş (video call), chat-başlamaq (start a chat). Hybrid compounds carry phonological and semantic cues, supporting rapid parsing and comprehension in digital communication. Historical continuity is evident: interfixes and reduplication transfer classical patterns into modern, technologically mediated registers.

Typologically, Azerbaijani shares certain linguistic features with Turkish, Russian, English, Japanese, and Persian, while still retaining distinct cognitive and expressive advantages. For instance, in comparison with Turkish, Azerbaijani exhibits similar agglutinative structures and vowel harmony, but it uses fewer interfixes and internal alternations. Compared to Russian, which is heavily fusional, Azerbaijani preserves agglutinative roots with optional internal alternation, as in yaz → yazdır (write → print) versus the Russian писать (pisat; to write) → заставить писать (zastavit pisat; to make someone write). In relation to English, which is largely analytic and exhibits limited morphological productivity, Azerbaijani can compress complex semantic information morphologically. With respect to Japanese, which often employs reduplication for plurality or emphasis, Azerbaijani integrates similar semantic effects through the use of interfixes and stem alternations. Finally, in comparison with Persian, Azerbaijani has borrowed roots and interfixes that have been adapted to its agglutinative patterns, as seen in examples like elm-i-fən (science) and ədəb-i-nəzər (litetature). Sociolinguistically, urban digital users innovate,

whereas formal and classical registers preserve historical morphology. Cognitively, this hybrid system supports efficient parsing, recognition, and production, particularly in digital and multilingual environments (López-Couso, 2016).

The present study demonstrates that Azerbaijani retains its agglutinative core while generating productive hybrid derivations under contact-driven influence. This pattern aligns with recent findings in morphological typology, psycholinguistics, and contact linguistics, and situates Azerbaijani within broader global research on language adaptation. The present study demonstrates that Azerbaijani retains its agglutinative core while simultaneously expanding its derivational potential through contactdriven hybrid formations – a dynamic mirrored in recent crosslinguistic and typological research on Turkic and other languages. Recent semanticmorphological work by C. Ismoilova (2025) on Turkish has shown that in agglutinative languages, affixation is guided by semantic coherence rather than purely formal concatenation. This supports the find that many of the productive hybrid derivatives in Azerbaijani preserve semantic transparency, which helps them integrate smoothly into everyday usage rather than remain sporadic borrowings. In other words, semantic consistency appears to enable hybrid affixation to become part of the productive core, rather than marginal exceptions.

Further, studies on loan-verb adaptation in Russian reveal systematic patterns in suffix assignment: loan verbs often receive variant suffixes and coexist with native morphological patterns (Sokolova *et al*., 2025). This resonates with current observation that borrowed stems in Azerbaijani likewise take native affixes or hybrid stemaffix combinations systematically rather than haphazardly. The crosslinguistic similarity suggests that contactinduced morphological adaptation may follow universal tendencies shaped by phonological compatibility and morphological economy, regardless of the donor or recipient language's typology.

Notably, a study on prefix-like formation in Turkish by G. Booij (2018) – which documents emergent "prefixoids" derived from adjectival constituents (e.g., ana "main/primary," ön "pre") – argues that even languages without traditional prefixation may develop creative wordformation strategies to address lexical gaps. This parallels current finding of hybrid affixation in Azerbaijani: as contact and lexical influx introduce new stems, speakers exploit the flexibility of the agglutinative system to generate novel forms. In doing so, Azerbaijani expands its wordformation repertoire in a way reminiscent of emergent morphological innovation observed in related Turkic languages.

However, computationalmorphology research cautions about the consequences of such morphological diversity. A recent study on language modelling for morphologically rich languages by C. Toraman *et al*. (2022) showed that standard tokenisation and NLP tools often struggle with complex affixation and irregular morphology. This suggests that while hybrid derivations may enrich the lexicon and ease human processing, they pose challenges for automated processing, morphological tagging, and computational lexicography – a limitation that must be acknowledged when considering applications of this research (e.g., morphological analysers, spell-checkers). Moreover, historicalphilological investigation of ancient Turkic root morphemes revealed the long-standing productivity of affix-based derivation in the family, pointing to diachronic continuity in word-formation strategies, as proved by D. Beck (2017) in a recent study. This supports diachronic findings made in this work: hybrid and novel forms in Azerbaijani may not be aberrations but part of a deep-rooted morphological tradition, albeit one evolving under modern sociolinguistic pressures. From that perspective, the hybrid derivations documented in this study – especially those combining borrowed stems with native affixes – might blur the boundaries between derivation, compounding, and morphological construction. This

raises the question whether some of identified "words" in this work are better analysed as constructionlevel or compoundderived units, suggesting a fruitful direction for future research.

In light of these comparisons, the present study contributes to a growing body of evidence that morphological systems – even in agglutinative languages – are not static. They respond to sociolinguistic dynamics, contact, lexical influx, and speaker needs. For Azerbaijani, this means that hybrid derivation and loanaffix integration are not peripheral anomalies but active mechanisms enriching the lexicon and reflecting real language use. At the same time, current findings highlight a tension: the same processes that enhance expressive capacity and cognitive accessibility may complicate computational processing or formal standardsation. Thus, while supporting many of the generalisations found in contemporary studies – about agglutination, borrowing adaptation, semantic transparency, and cognitive processing – this research also pushes the boundary of typological and morphological theory by showing how language contact can dynamically reshape morphological productivity without erasing core structural traits. In sum, Azerbaijani emerges as a morphologically resilient, flexible, and adaptive system: rooted in agglutination, yet capable of accommodating hybrid innovation – a finding that bridges typological tradition with modern contact-induced evolution.

Conclusions

The present study demonstrated that Azerbaijani morphology is highly dynamic, adaptive, and cognitively optimised, integrating multiple word-formation mechanisms-including suffixation, interfixation, compounding, reduplication, apophony, and conversion-to generate a rich, expressive, and efficient lexicon. Corpus analysis of 2020-2024 data revealed that suffixation dominates derivational productivity, accounting for 82.4% of all tokens (n = 43,912 types), with the most productive suffixes being -lıq/-lik (17,504 tokens), -çı/-çi (9,612 tokens), -sız/-siz (7,921 tokens), and -lı/-li (5,332 tokens). Hybridisation with borrowed stems was also productive: forms such as follow-çu (follower) and blok-lamaq (blocker) generated multiple derivatives, achieving frequencies per million (fpm) of 78.3 and 112.6 respectively, whereas less integrated forms like like-ləmək (to like) remained non-productive (fpm = 32.4). Prefixation remains limited (11.3% of new derivations), with domain-specific prominence in technological, academic, and youth discourse (e.g., qeyri- 4,480 tokens; anti- 2,931 tokens). Interfixation and compounding enhance formal clarity and cognitive processing, while reduplication supports emphasis and memorability. Apophony allows precise derivational adjustments, and conversion facilitates flexibility and rapid lexical expansion. Historical layers of Persian, Arabic, Turkish, and Russian influence coexist with modern digital innovations, producing a morphological system that operates effectively across literary, formal, and online registers. Comparative analysis showed that Azerbaijani integrates features of agglutinative, concatenative, and non-concatenative languages while maintaining semantic transparency and processing efficiency, distinguishing it as a hybrid system capable of accommodating neologisms, hybrid compounds, and contact-induced borrowings. Future research should expand the integration of corpus linguistics, psycholinguistic experiments, and computational modelling. Large-scale spoken, written, and digital corpora can be used to track frequency, productivity, and register variation of derivational mechanisms. Cognitive experiments measuring reaction times and accuracy for parsing complex forms (e.g., tez-tez-yenilənən-onlayn-məlumat; frequently updated online information) can validate hypotheses about processing efficiency. Computational approaches, including NLP and AI-based morphological analysers, will benefit from comprehensive morphological databases to handle compounds, reduplications, and apophony. Dialectal

studies, longitudinal tracking, and interdisciplinary collaboration will further illuminate how classical morphology adapts to modern technological, educational, and multilingual demands. Overall, Azerbaijani morphology exemplifies a living system that remains resilient, adaptive, and socially functional while preserving clarity, expressiveness, and cognitive efficiency.

Acknowledgements

None.

Funding

None.

Conflict of Interest

None.

References

[1] Abdullayeva-Nebiyeva, V. (2025). Cultural determinants and grammatical outcomes in the Azerbaijani: An analysis of sociolinguistic influence on Turkic grammar. *Path of Science*, 11(4), 4028-4032. doi: 10.22178/pos.116-35.

[2] Beck, D. (2017). The typology of morphological processes: Form and function. In A.Y. Aikhenvald & R.M.W. Dixon (Eds.), *The Cambridge handbook of linguistic typology* (pp. 325-360). Cambridge: Cambridge University Press. doi: 10.1017/9781316135716.011.

[3] Booij, G. (2018). The construction of words. In B. Dancygier (Ed.), *The Cambridge handbook of cognitive linguistics* (pp. 229-245). Cambridge: Cambridge University Press. doi: 10.1017/9781316339732.016.

[4] British National Corpus. (n.d.). Retrieved from https://www.english-corpora.org/bnc/.

[5] Durkin, P. (Ed.). (2015). *The Oxford handbook of lexicography*. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780199691630.001.0001.

[6] Humboldt-Universität zu Berlin, Institute of Linguistics. (n.d.). *Corpus linguistics: Corpora links*. Retrieved from https://www.linguistik.hu-berlin.de/en/institut-en/professuren-en/korpuslinguistik/links-en/korpora_links.

[7] Huseynova, M. (2019). Language contacts of Azerbaijani and Kazakh Turkic languages (on the basis of Azerbaijani dialectologist, academician Mammadaga Shiraliyev's creative works). *Turkic Studies Journal*, 1(2), 34-40. doi: 10.32523/tsj.02-2019/2-4.

[8] Ismoilova, C. (2025). The foundations of Turkish word formation: A semantic approach. *International Journal of Artificial Intelligence*, 1(5), 1581-1583.

[9] Jalilbayli, O.B. (2022). Philosophy of linguistic culture and new perspectives in modern Azerbaijani linguistics. *Futurity Philosophy*, 1(4), 53-65. doi: 10.57125/FP.2022.12.30.05.

[10] Kornfilt, J. (2002). The Turkic languages (Routledge language family descriptions). *Journal of Linguistics*, 38(2), 397-439. doi: 10.1017/S0022226702251622.

[11] LópezCouso, M.J. (2016). Corpora and online resources in English historical linguistics. In M. Kytö & P. Pahta (Eds.), *The Cambridge handbook of English historical linguistics* (pp. 127-145). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139600231.009.

[12] Milli Arxiv İdarəsi. (n.d.). *Publication and use of documents*. Retrieved from http://www.milliarxiv.gov.az/az/news/senedlerin-nesri-ve-istifadesi.

[13] Mirzayeva, A. (2021). A phytonymic phraseology and linguocultural features of paremiological units. *Langua: Journal of Linguistics, Literature, and Language Education*, 4(2), 39-44. doi: 10.5281/zenodo.5540285.

[14] Nasirova, U., Hasanova, S., & Dashdemirova, P. (2023). Phonetic rules in the Azerbaijani language. *ASES International Journal of Culture, Art and Literature*, 2(1), 33-39. doi: 10.5281/zenodo.10492377.

[15] National Corpus of Azerbaijani Language. (n.d.). Retrieved from https://korpus.azerbaycandili.az/.
[16] Özenç, B., Ehsani, R., & Solak, E. (2018). MorAz: An open-source morphological analyzer for Azerbaijani Turkish. In E. Blanco & W. Lu (Eds.), *Proceedings of the 2018 conference on empirical methods in natural language processing: System demonstrations* (pp. 25-29). Brussels: Association for Computational Linguistics. doi: 10.18653/v1/D18-2005.
[17] Plag, I. (2025). Productivity and the mental lexicon. In A.J. Merrison, P. Griffiths & A. Bloomer (Eds.), *Language in use* (pp. 106-115). London: Routledge.
[18] Rakhimova, G.Y. (2023). Functional semantics of English reduplicative elements. *International Journal of Applied Linguistics and Translation*, 9(3), 90-92. doi: 10.11648/j.ijalt.20230903.12.
[19] Russian National Corpus. (n.d.). Retrieved from https://ruscorpora.ru/en.
[20] Sokolova, S., Birzer, S., Ignateva, A., & Kibisova, E. (2025). Suffix diversity: Investigating the morphological landscape of Russian loan verbs. *Russian Linguistics*, 49, article number 17. doi: 10.1007/s11185-025-09324-8.
[21] Toraman, C., Yılmaz, D., & Şahinuç, A. (2022). Evaluating morphological compositional generalization in large language models for agglutinative languages. *ArXiv*. doi: 10.48550/arXiv.2410.12656.
[22] Tsanko, I. (2025). Corpora and the study of language evolution: A historical linguistics perspective. *SWorldJournal*, 4(30-04), 122-129. doi: 10.30888/2663-5712.2025-30-04-041.
[23] Turkish National Corpus. (n.d.). Retrieved from https://www.tnc.org.tr/.

Механізми словотворення в азербайджанській та світових мовах: порівняльне морфологічне дослідження

Раміла Фараджова
Доктор філологічних наук
Азербайджанська Державна Морська Академія
AZ1000, вул. Заріфи Алієвої, 18, м. Баку, Азербайджан
https://orcid.org/0009-0000-2006-1706

Анотація. У цій дослідницькій роботі представлено розширене та глибоке порівняльне дослідження механізмів словотворення в азербайджанській мові у зв'язку з кількома основними мовами світу, зокрема англійською, російською, німецькою, турецькою та перською. Розміщуючи азербайджанську мову в ширших міжмовних рамках, дослідження мало на меті з'ясувати як типологічні особливості, так і універсальні закономірності в морфологічних процесах. Методологія інтегрувала когнітивно-морфологічний аналіз, корпусне дослідження, обчислювальне та цифрове моделювання ресурсів, а також міжмовне типологічне порівняння для дослідження морфологічних структур азербайджанської мови. Дослідження показало, що азербайджанська мова зберегла свою основну аглютинативну структуру, розвиваючи гібридні утворення через інтеграцію запозичень та афіксів, демонструючи збільшення частоти змішаних морфологічних ланцюгів у цифрових корпусах та розширення продуктивних афіксаційних моделей у відповідь на лексичний приплив, зумовлений контактами. Емпіричний аналіз показав, що азербайджанська морфологія була одночасно гнучкою та стійкою, здатною генерувати нові лексичні одиниці та адаптуватися до семантичних зрушень у відповідь на соціальний, технологічний та міжкультурний розвиток. Ці висновки підкреслили подвійний характер морфологічної еволюції: вона виявилася універсальною у своїх структурних тенденціях, водночас унікально формуючись культурним та лінгвістичним контекстом азербайджанських носіїв. Розміщуючи азербайджанську морфологію в порівняльному ландшафті світових мов, це дослідження сприяло глибшому розумінню міжмовної креативності, типологічної варіації та взаємодії між морфологією та соціолінгвістичною динамікою, пропонуючи розуміння, актуальне для теоретичної лінгвістики, викладання мов та прикладної лексикографії. Практична цінність цього дослідження полягає в тому, що воно надає лінгвістам, лексикографам, викладачам та розробникам цифрових мовних технологій емпірично обґрунтовані моделі азербайджанського словотворення, які можна безпосередньо застосовувати для складання словників, розробки навчальних програм, автоматизованої обробки морфології та розробки інструментів вивчення природного мови, таких як морфологічні аналізатори та засоби перевірки орфографії

Ключові слова: аглютинативна структура; типологічний контраст; лексичні інновації; контактно-індуковані зміни; афіксальна продуктивність; цифрові корпуси; міжмовна креативність