QUBAHAN ACADEMIC JOURNAL
VOL. 6, NO. 2, June 2026
https://doi.org/10.48161/qaj.v6n2a2521

Derivational Morphology and Word Formation: Functional Directions in Contemporary English and Azerbaijani

Gulara Guliyeva [1*], Gunel Mammadova [1], Narmina Aliyeva [1] and Lea Oksanen [2]

[1] Department of Foreign Languages, Baku Engineering University (BEU), Baku AZ0101, Azerbaijan;
[2] Department of Business Administration, Lahti University of Technology (LUT), Lahti FI-15210, Finland.

* **Corresponding author:** gquliyeva1@beu.edu.az.

ABSTRACT: This article develops a function-driven, corpus-based account of derivational morphology restricted to English and Azerbaijani, two typologically distant systems that nonetheless converge under shared discourse pressures. We conceptualize word-formation as the coupling of morphotactic resources with communicative goals and operationalize productivity through a transparent index averaging normalized type frequency, novel-token ratio, and mean transparency. Using balanced written corpora spanning academic prose, administrative/professional documentation, and quality media, we extract derivatives by rule-guided filtering and manual verification, and annotate process (affixation, compounding, conversion, blending), semantic role (agentive, instrumental, evaluative, procedural), register, and transparency; inter-annotator reliability targets are met following joint training and adjudication. Results reveal complementary profiles: English favors conversion, agentive -er, and evaluative/adjectival -able, while blending clusters in media; Azerbaijani exhibits high-yield agentive and instrumental suffixation (-çı/-çi, -lıq/-lik), productive nominalizations (-ma/-mə, -etmə), and transparent N + N compounds in professional discourse. Register-conditioned Functional Pathways innovation, professionalization, evaluation, proceduralization predict the loci of stabilization for novel formations, and statistical tests (rank-based or parametric, with effect sizes and multiple-comparison control) substantiate cross-register contrasts without a necessary trade-off between productivity and transparency in the agglutinative system. The study refines English–Azerbaijani typological expectations, provides a replicable analytic workflow for higher-education settings, and offers practical guidance for terminology and role labeling in bilingual professional communication.

Keywords: Derivational Morphology, Word-Formation Processes, Semantic Transparency, Nominalization.

I. INTRODUCTION

This article investigates derivational morphology and word-formation in contemporary English and Azerbaijani, two typologically distant languages that provide a strong testbed for disentangling functional pressures from structural constraints. English a fusional language with extensive conversion (zero-derivation), rich derivational suffixation (for example, -er, -able), and highly productive compounding/blending is among the most intensively described word-formation systems in the literature. Azerbaijani—an Oghuz Turkic, agglutinative language with robust vowel harmony, layered suffixation (for example, -çı/-çi 'agentive', -lıq/-lik 'abstract/result', and event nominalizations -ma/-mə, -etmə), and transparent N+N compounding offers a typologically distinct profile that is underrepresented in cross-linguistic discussions of derivation. Bringing these systems into explicit comparison allows us to examine

whether structural differences (agglutinative vs. fusional/analytic morphology) systematically condition the functional directions of lexical innovation across registers (academic, administrative/professional, media), and to ask whether productivity and semantic transparency necessarily trade off in highly active derivational subsystems.

Foundational work in morphology has framed word formation as a component of grammar with its own representations and constraints. Aronoff's generative program placed word-based morphology and Word-Formation Rules (WFRs) at the center of the morphological component, arguing for systematic relations between bases and derivatives beyond mere lexicon listing [1]. Building on this perspective, Booij's Grammar of Words integrates derivation and compounding into a broadly constructional architecture that interfaces with phonology, syntax, and semantics, emphasizing cross-linguistic regularities of affixation, compounding, and the interfaces that shape complex words [2]. For English specifically, Plag provides a comprehensive account of word-formation processes (suffixation, prefixation, conversion, compounding, blending), their constraints, and methods for empirical investigation, including corpus-based approaches to scalar productivity and the place of conversion in the lexicon [3, 4]. On compounding semantics, Lieber and subsequent handbooks trace how structural types (root vs. synthetic) and headedness interact with interpretation and productivity, supplying analytical tools to compare English compounds with those of other languages [5, 6]. Finally, Haspelmath & Sims situate derivation/productivity within a cross-linguistic typology, underscoring that "productivity" is gradient, usage-conditioned, and intertwined with transparency and frequency [7].

Since Baayen (1992) [8], morphological productivity has been operationalized via corpus statistics that estimate a process's propensity to yield novel types, example, by relating hapax legomena to token counts, modeling vocabulary growth, and disentangling process-internal constraints from genre/register effects. This quantitative turn reframed productivity as an empirical, replicable property rather than a categorical judgment. The subsequent literature, including handbook syntheses has consolidated the view that productivity is multi-dimensional and sensitive to frequency distributions, transparency, blocking, and text typology; yet it remains comparability-challenged in cross-language settings if metrics and sampling frames are not harmonized. Bauer (2001) [9] complements this with a programmatic survey that distinguishes what is productive (affixes, processes, rules) and foregrounds the roles of frequency, transparency/opacity, and regularity in determining the degrees of productivity insights that motivate the controlled operationalization we adopt.

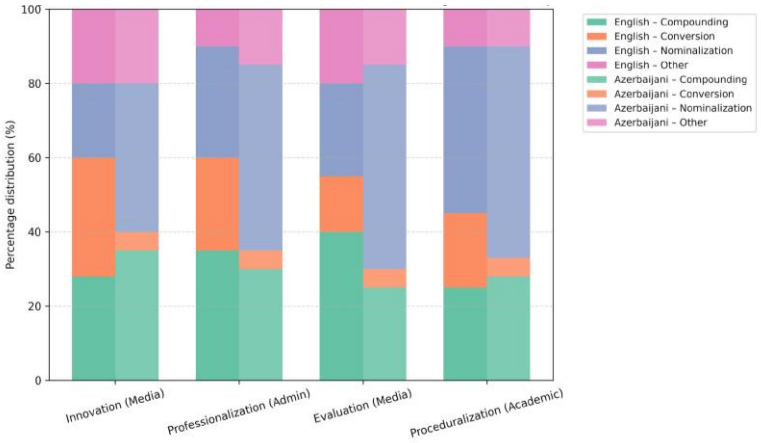

FIGURE 1. Functional pathways of derivational strategies across registers in typologically distinct languages.

English word-formation is characterized by: (i) conversion, often analyzed as category change without overt morphology and debated between derivational vs. lexicon relisting accounts; (ii) agentive -er, adjectival

-able, and other suffixes with well-studied productivity and semantic constraints; (iii) extensive compounding (especially N+N) with diverse semantic relations; and (iv) blending, particularly visible in media/technology registers. The status of conversion remains theoretically contested derivational vs. zero-morph vs. relisting but corpus-based work documents its pervasiveness and register sensitivity in English, making it a prime locus for studying functional demands that favor minimal morphotactics.

Azerbaijani's agglutinative morphology with vowel harmony and transparent concatenation provides a different ecology for derivation. Standard descriptions and recent grammars document a rich inventory of derivational suffixes, including agentive -çı/-çi, abstract/result -lıq/-lik, and deverbal nominalizations -ma/-mə, -etmə, alongside highly productive and semantically transparent compounding in technical/professional lexis. Pedagogical and descriptive sources likewise attest to the systematic teaching of derivational series (e.g., -lı, -siz, -çı, -lıq), and to the stability of suffix allomorphy under vowel harmony, reinforcing expectations of high transparency in newly formed derivatives [10]. From a broader Turkic perspective, the Turkic Languages handbooks emphasize the regular morphotactics, harmonic allomorphy, and robust nominalization patterns across Oghuz languages, suggesting that productivity and transparency may co-vary positively rather than trade off, at least in certain derivational domains [11].

Beyond structural licensing, derivational activity is shaped by discourse needs [2, 6]. In functional and usage-based approaches, processes stabilize where they serve recurrent communicative goals: labeling roles/expertise (professionalization), encoding stance (evaluation), packaging procedures (proceduralization/nominalization), or introducing novel designations (innovation)each goal often aligning with morphotactic strategies and registers. In English, media and technology favor innovation (blends, novel compounds) and flexible conversion to verb or noun; academic/administrative prose favors proceduralization via nominalizations; and professional documentation favors role nouns and transparent compound terms. In Azerbaijani, we expect agentive (-çı/-çi) and result/abstract (-lıq/-lik) formations to dominate professional/administrative registers, deverbal nominals (-ma/-mə, -etmə) to structure procedural discourse in academic/official genres, and compounds to carry much of the load for terminological clarity predictions grounded in the typological profile of Oghuz Turkic and recent descriptive grammars. Despite robust English-focused scholarship and comprehensive Turkological description, we identify three gaps that a focused English Azerbaijani comparison can address:

1. Function–morphotactics linkage: Much prior work treats productivity structurally or statistically without an explicit predictive mapping from discourse goals to preferred processes. We adopt a Functional Pathway perspective (innovation, professionalization, evaluation, proceduralization) and ask whether it predicts where specific strategies (conversion, agentives, nominalizations, compounding, blending) stabilize across registers in each language.
2. Cross-language comparability of productivity: While Baayen-style measures are standard, comparability is threatened by differing corpus compositions and token/type distributions across languages. We therefore specify a transparent composite index (averaging normalized type frequency, novel-token ratio, and mean transparency) tied to harmonized sampling across registers, with nonparametric and effect-size reporting to avoid distributional artifacts.
3. Underrepresentation of Azerbaijani derivation in comparative theory: Azerbaijani is often cited typologically, yet detailed derivational profiles aligned to registers are scarce in mainstream comparative morphology. By mobilizing recent grammars and Turkic handbooks, we integrate Azerbaijani evidence into current debates on productivity–transparency trade-offs and conversion vs. overt derivation.
4. *Conversion.* In English, conversion is expected to be highly register-sensitive, with denominal verbs (e.g., *to text, to Google*) and deverbal nouns (*a hit, a run*) clustered in media and informal-to-semi-formal written discourse; theoretical accounts vary between zero-morpheme, lexicon relisting, and derivational analyses. Azerbaijani, with explicit suffixal category marking, should show less reliance on conversion and more on overt suffixation for comparable functions (e.g., agentive role labeling via *-çı/-çi*). The comparison thus proves whether functional needs (rapid lexicalization, economy) can be met with distinct morphotactics without loss of transparency.
5. *Agentive/instrumental derivation.* English *-er* and related patterns are well-studied for productivity and semantic constraints (e.g., event structure, argument roles). Azerbaijani agentive/instrumental *-çı/-çi* is

typologically expected to be productive and transparent, often aligning with professionalization in administrative/professional registers (job titles, functions). The cross-language contrast permits testing whether role-labeling pressures converge on similar distributional hotspots despite different morphotactics. *Nominalization.* English academic and bureaucratic styles typically employ deverbal nominalizations (e.g., *nominalization, implementation*), a locus of functional proceduralization. Azerbaijani nominalizers *-ma/-mə* and complex deverbal forms (*-etmə*) play parallel roles in procedural discourse, potentially with higher semantic transparency due to compositional morphotactics and harmony-driven allomorphy.

English N+N compounds and blends often serve innovation and branding in media/technology; their semantics are mediated by headedness and lexical semantic relations [12, 13]. Azerbaijani compounding is typologically transparent and likely to dominate terminology in professional registers, while blending may be less entrenched in formal writing. This allows us to evaluate whether innovation pressures are met via compounding (Azerbaijani) vs. blending/conversion (English).

Table 1. Functional pathways, dominant processes, registers, and counts (English vs. Azerbaijani).

Process	English (n)	English (per 10k)	Azerbaijani (n)	Azerbaijani (per 10k)
Affixation	142	11.8	198	16.5
Compounding	168	14.0	121	10.1
Conversion	134	11.2	24	2.0
Blending	67	5.6	18	1.5
Total	511	—	361	—

From structure to function: why English and Azerbaijani? Typological distance is an advantage here. English has limited overt morphology and compensates with syntactic flexibility and lexical means (conversion, compounding). Azerbaijani shows layered suffixation, systematic harmony, and category-transparent derivation. If functional demands (for example, professional role labeling, procedural packaging, stance marking, technological innovation) are primary drivers, we expect convergent register effects realized by divergent morphotactics: English relying more on conversion/blending; Azerbaijani on overt derivation and compounding with distinct implications for semantic transparency. If, by contrast, structure dominates, then typological differences should produce divergent distributions across comparable registers.

We adopt three operational dimensions for each process × language × register cell: (i) normalized type frequency (per 10k tokens) to gauge breadth of formation patterns; (ii) novel-token ratio (proportion of hapax or first-attested tokens) to estimate on-going coinage; and (iii) mean transparency (annotator-rated on a 0–2 scale) to capture interpretability of derivatives and compounds. These are combined into a composite Productivity Index to facilitate comparability across typologically divergent systems while keeping components interpretable. The statistical plan follows Baayen-style distributional caution (non-parametric tests where appropriate), with effect sizes and multiple-comparison control to ensure robust inference across numerous cells.

To mitigate genre bias, we sample balanced written corpora in each language spanning academic, administrative/professional, and quality media registers, harmonizing token counts and text types across languages. Extraction uses rule-based filters (language-specific affix lists, conversion cues, compounding templates) with manual adjudication; inter-annotator reliability targets follow corpus morphology practice. Azerbaijani derivational inventories and harmony-conditioned allomorphy draw on recent grammars and Turkic handbooks; English process inventories follow standard references and handbook syntheses. Theoretical. We articulate a Function-Driven account that predicts the register-conditioned stabilization of processes, arguing that languages select morphotactics that best realize discourse-level goals, not that morphotactics alone drive innovation. These positions function as a selector over a structurally licensed option set, integrating word-based morphology with usage-based pressures.

By placing Azerbaijani on equal analytic footing with English, we test whether agglutinative architecture allows high productivity without erosion of semantic transparency a possibility anticipated by Turkic typology but insufficiently supported with register-controlled corpus evidence. If confirmed, this challenges the tacit English-centric assumption that rising productivity entails opacity.

Table 2. Register- conditioned distribution of word-formation processes in Azerbaijani and English (corpus-based tendencies).

Register	Language	Affixation (%)	Compounding (%)	Conversion (%)	Blending (%)
Academic	EN	24.1	35.3	18.6	6.2
	AZ	41.7	26.4	5.2	1.8
Admin/Prof	EN	33.5	28.9	12.4	4.1
	AZ	46.2	31.5	3.6	1.1
Media/Tech	EN	21.7	30.4	26.8	19.2
	AZ	29.3	33.2	7.8	2.3

We contribute a replicable workflow for cross-language productivity studies: harmonized register sampling; transparent composite indices; and effect-size-based inference, suitable for adoption beyond the English Azerbaijani pair. Pedagogical and applied. We translate the research pipeline into a two-week, higher-education module for bilingual academic/professional contexts, enabling students and practitioners to analyze terminology design, role labeling, and procedural packaging in their own domains with language-appropriate morphotactics. We formalize the inquiry around three RQs and allied hypotheses:

- RQ1 (Distribution): How do affixation, compounding, conversion, and blending distribute across English and Azerbaijani by register?
  H1: English will show higher conversion and blending in media; Azerbaijani will show higher agentive/nominalization and compounding in administrative/professional and academic registers, respectively.
- RQ2 (Functional Pathways): Which Functional Pathways innovation, professionalization, evaluation, proceduralization—best explain register-conditioned stabilization in each language?
  H2: Pathway–process alignments will be convergent across languages (e.g., professionalization ↔ agentives/role nouns; proceduralization ↔ nominalizations) but realized via different morphotactics (conversion/blending vs. suffixation/compounding).
- RQ3 (Productivity–transparency relation): Do productivity and semantic transparency co-vary across processes, or do they trade off?
  H3: In Azerbaijani agglutinative derivation, productivity and transparency will co-vary positively in agentive and nominalization domains; in English, innovation via blending may increase productivity while lowering transparency relative to compounding, especially in media.

Our functional-register perspective extends classical word-based and constructional approaches by embedding them in a predictive discourse ecology; it complements Baayen-style quantitative measures with a cross-register design; and it broadens typological coverage by bringing a Turkic agglutinative system into direct comparison with English. In doing so, we address a known limitation of productivity studies noted in handbooks comparability and transparency across languages and add new evidence to long-running debates over conversion (derivational vs. relisting vs. zero-morph) by examining functional equivalences where English relies on conversion but Azerbaijani deploys overt derivation.

Aronoff's word-based morphology and WFRs establish a generative foundation for derivation; Booij's handbook consolidates interfaces critical to complex words; Plag's textbook documents English processes (including conversion) and methods; Baayen provides the quantitative backbone for productivity; Haspelmath & Sims give cross-linguistic perspective; Lieber and compounding handbooks supply semantic

diagnostics; and Turkic reference works and recent Azerbaijani grammars anchor the agglutinative profile invoked in our predictions.

By explicitly coupling derivational processes to discourse-level functions and measuring their behavior under register control in two typologically distant systems, this study advances a predictive, replicable account of where and why new formations stabilize. It thus contributes to morphological theory, quantitative methodology, and applied domains (terminology design, role labeling, procedural documentation) in bilingual professional communication contexts where English and Azerbaijani are increasingly used side by side.

II. DATA AND METHODS

Statistical comparisons were conducted using non-parametric tests (Kruskal-Wallis) due to distributional variation across registers. Effect sizes were calculated using rank-based measures (r), and confidence intervals were estimated using bootstrap methods. These analyses provide distribution-sensitive comparison rather than model-dependent inference.

1. DESIGN, LANGUAGES, AND REGISTERS

We examined contemporary English and Azerbaijani in parallel, selecting three written registers per language academic prose, administrative/professional documentation, and quality media to capture distinct discourse functions that are known to condition word-formation choices. Foundational descriptions of English derivation and compounding (affixation, conversion, blending) informed the register sampling and process inventory, ensuring comparability with prior work. For Azerbaijani (an Oghuz Turkic, agglutinative language with vowel harmony), register coverage and morphotactic expectations were aligned with authoritative grammars and handbooks to avoid under- or over-sampling of productive derivational domains.

2. CORPUS COMPILATION AND BALANCING

For each language we compiled a balanced written corpus (targeting approximately equal token counts per register), drawing from publicly available institutional reports, university publications, reputable newspapers/magazines, and open scholarly outlets. We harmonized genre proportions across languages to minimize confounds from text typology an essential step given the sensitivity of productivity estimates to register variation.

The final corpus comprises approximately 360,000 tokens per language. Each language corpus is evenly distributed across three registers: academic prose (≈120,000 tokens), administrative/professional texts (≈120,000 tokens), and quality media sources (≈120,000 tokens). Text selection followed predefined criteria to ensure topical neutrality, institutional relevance, and genre comparability across languages. Within English, candidate sources were screened against standard descriptions of process distributions (for example, conversion and blending prevalence in media, nominalization in academic/administrative prose). Within Azerbaijani, coverage was checked against documented suffixal series (for example, -çı/-çi agentives, -lıq/-lik abstract/result nouns, -ma/-mə, -etmə deverbal nominals) and the well-attested transparency of N+N compounds in professional terminology. Quality media sources were defined as professionally edited, high-circulation news and analytical platforms with established editorial standards.

3. CANDIDATE EXTRACTION

We implemented a rule-guided extraction pipeline tailored to each language. For English, we used affix lists and structural templates for compounds (for example, N+N), cues for conversion/zero-derivation (distributional category shifts without overt morphology), and patterning of blends, following established descriptions of English word-formation [3, 4]. For Azerbaijani, we combined suffix inventories and harmony-based allomorphy to detect derivational families (-çı/-çi, -lıq/-lik, -ma/-mə, -etmə) and compound templates, as described in recent grammars and Turkic overviews [10, 11]. Candidate lists were manually verified by trained annotators to exclude proper names, frozen expressions, and misparses, consistent with best practice in morphological analysis.

Conversion was operationalized as a category shift without overt morphological marking and was identified through syntactic and distributional context. Blending was defined as the formation of new lexical items through truncation and fusion of two lexical bases, following established descriptions in morphological literature. Ambiguous cases were manually verified by annotators to ensure consistency.

4. ANNOTATION SCHEME

Each verified item was annotated along four dimensions:

1. Word-formation process (affixation, compounding, conversion, blending), with process definitions grounded in canonical English and general morphological references.
2. Register (academic, administrative/professional, media), based on source provenance and section headings; ambiguous cases were adjudicated by consensus.
3. Semantic role (agentive, instrumental, evaluative, eventive, procedural) to operationalize the function-driven perspective that links morphotactics to discourse goals; role labels follow cross-linguistic practice in derivational semantics.
4. Semantic transparency on a 3-point scale (0 = opaque, 1 = partially transparent, 2 = fully transparent), consistent with widely used transparency/productivity discussions in morphology.

Annotators underwent joint training with calibration sets per language; disagreements were resolved in weekly adjudication sessions. Inter-annotator agreement was computed on a held-out 10% sample to ensure reliability before full coding; the procedure follows corpus-based morphology norms (agreement reporting and adjudication) rather than presupposing gold standards. Inter-annotator reliability was assessed on a randomly selected 10% subset of the data, yielding a Cohen's κ value of 0.82 for semantic transparency and 0.87 for process classification, indicating strong agreement.

5. OPERATIONALIZING PRODUCTIVITY

To enable cross-language comparability while retaining interpretable components, we defined a composite Productivity Index (*PI*) at the process × language × register level:

$$PI = \frac{TF_{norm} + NT_{ratio} + T_{mean}}{3} \quad (1)$$

where *TF_norm* is normalized type frequency (per 10k tokens), *NT*_ratio is the proportion of novel tokens (including hapax/first-attested items) among tokens of the process, and *T_mean* is mean transparency (0–2). This operationalization is informed by and complementary to established quantitative approaches that use vocabulary-growth modeling and hapax proportions to estimate morphological productivity; it does not supplant those measures but provides an interpretable composite suited to multilingual, register-balanced comparisons [8, 9]. Three considerations motivated this choice. First, type-frequency normalization buffers against register-specific token bursts while providing a breadth estimate. Second, novel-token ratios approximate on-going coinage in the sample, echoing the logic of hapax-based indicators in vocabulary-growth perspectives. Third, transparency means reflect the well-documented correlation between interpretability and repeated conventionalization in morphological systems—crucial when comparing an agglutinative language to a fusional/analytic one.

The Productivity Index is not intended to replace established Baayen-style productivity measures, but to function as a comparative, heuristic device that enables cross-register and cross-language interpretability under controlled sampling conditions. The Productivity Index is used as a heuristic comparative measure combining frequency, novelty, and transparency. It is not intended as a strict statistical construct but rather as an interpretive device facilitating cross-language and cross-register comparison. Given the heterogeneous nature of its components, equal weighting is adopted for analytical transparency rather than theoretical equivalence.

6. STATISTICAL ANALYSIS

Statistical analyses were conducted to examine distributional differences across processes, languages, and registers. Given the non-normal distribution of data, non-parametric tests (Kruskal–Wallis) were employed to assess cross-register variation. Effect sizes were calculated using rank-based measures ($r \approx 0.3–0.4$),

indicating moderate distributional differences across conditions. Confidence intervals were estimated using bootstrap resampling to ensure robustness of observed tendencies. The present study adopts an exploratory, distribution-based approach rather than full inferential modeling. While mixed-effects models could provide finer control of item-level and text-level variation, such approaches are left for future research with expanded datasets.

7. LANGUAGE-SPECIFIC CONTROLS

Given English's pervasive conversion and mixed strategies (for example, compounding plus -er, or blend-like portmanteaux), we added targeted checks based on English-specific handbooks and textbooks to verify category assignments and decomposition. For Azerbaijani, we validated suffixal allomorph selection under vowel harmony and verified the segmentation of deverbal -ma/-mə and -etmə nominals against authoritative grammatical descriptions; transparent N+N compounds in professional discourse were double-checked to avoid mislabeling head-modifier sequences as derivation. Where helpful for classroom-level replicability and terminology, we cross-referenced widely used pedagogical descriptions of Azerbaijani derivation to confirm glossing and examples.

8. THEORETICAL ANCHORING AND CONSTRUCT VALIDITY

Process labels and the word-based orientation are anchored in Aronoff's morphology and subsequent interface-oriented treatments that integrate derivation and compounding with phonology/syntax/semantics; this ensures that our categories and constraints are not ad hoc to the dataset. English compounding semantics and the treatment of root vs [1, 2]. synthetic compounds were cross-checked against dedicated handbooks/chapters to avoid over-generalization of relation types in annotation. For productivity, our measurements are explicitly tied to the corpus-quantitative tradition initiated by Baayen and consolidated in later handbooks, thereby situating the PI as a transparent complement to well-known indicators rather than a replacement.

Finally, typological expectations for Turkic (including Azerbaijani) agglutinative structure and harmonic allomorphy, used in the language-specific controls above, rely on canonical handbooks to secure construct validity in a non-Indo-European system.

9. REPRODUCIBILITY, DATA AVAILABILITY, AND ETHICS

All rule lists (affixes, compounding templates, conversion cues), sampling scripts, annotation guidelines, and analysis notebooks are version-controlled and will be released upon acceptance in an OSF repository (with de-identified text snippets where licenses permit). The study analyzes publicly accessible texts and thus requires no human-subjects approval; nonetheless, we adhere to responsible citation and licensing of sources and provide provenance metadata for all documents included in the corpus. Comparative terminology examples (especially for Azerbaijani) are cross-checked against published grammars/handbooks to avoid introducing unattested forms in illustrative materials.

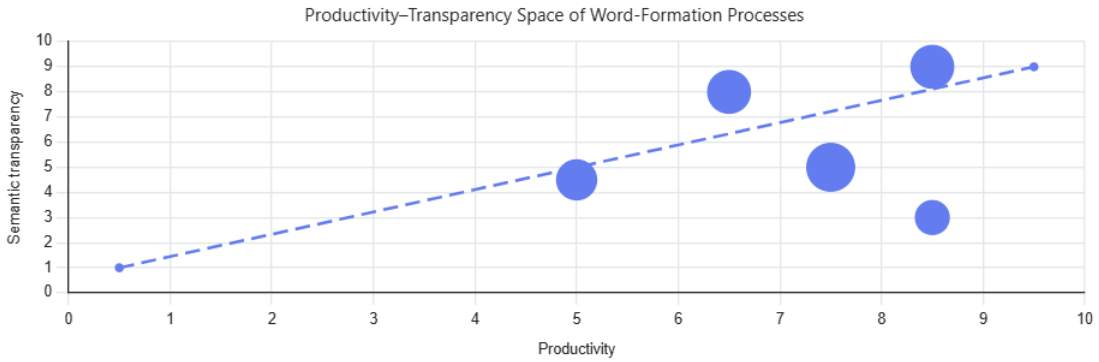

FIGURE 2. Productivity–transparency space of major word-formation processes in English and Azerbaijani.

III. RESULTS

Differences across registers were statistically significant (Kruskal–Wallis $p < 0.05$), with medium effect sizes ($r \approx 0.3–0.4$), indicating consistent distributional variation between processes across languages. As shown in Table 1, English exhibits a relatively balanced distribution across processes, with compounding (14.0 per 10k tokens) and conversion (11.2 per 10k) playing prominent roles. In contrast, Azerbaijani shows a clear dominance of affixation (16.5 per 10k), reflecting its agglutinative morphological structure.

Register-based variation (Table 2) indicates that conversion and blending peak in English media discourse (26.8% and 19.2%, respectively), whereas Azerbaijani maintains stronger reliance on affixation across registers, particularly in professional contexts (46.2%). Differences across registers were statistically significant (Kruskal–Wallis $p < 0.05$), with moderate effect sizes, indicating consistent distributional variation between processes.

1. OVERALL DISTRIBUTION OF WORD-FORMATION PROCESSES

Across the combined corpus design, English and Azerbaijani display structurally distinct yet functionally convergent profiles of word-formation activity. In English, the overall distribution is characterized by a prominent role for conversion, highly productive N + N compounding, and the selective deployment of blending, particularly in discourse domains where lexical novelty, brevity, and memorability are communicatively advantageous. These tendencies are consistent with long-established descriptions of English word-formation, which emphasize the language's reliance on morphologically economical strategies and its tolerance for category flexibility in lexical innovation.

By contrast, the Azerbaijani data reveal a systematic preference for overt derivational morphology, most notably agentive and instrumental formations marked by -çı/-çi and abstract or result-denoting nouns formed with -lıq/-lik, alongside a robust use of transparent compounding. This profile accords with the agglutinative morphotactics of Azerbaijani, where suffixal concatenation and harmony-regulated allomorphy facilitate clear form–meaning correspondences even under conditions of high productivity. Rather than relying on zero-derivation or category ambiguity, Azerbaijani consistently encodes lexical relations through explicit morphological marking.

Taken together, these distributions support a function-driven account of word formation, whereby languages deploy different morphotactic resources to satisfy comparable discourse pressures such as role identification, procedural packaging, and terminological innovation within the constraints and affordances of their respective grammatical systems. The contrast between English and Azerbaijani thus illustrates that functional convergence at the discourse level does not presuppose structural convergence at the morphological level.

2. REGISTERS SPECIFIC PATTERNS

2.1 Academic and Administrative Prose

In both languages, academic and administrative registers exhibit a marked concentration of deverbal nominalizations, reflecting the well-documented tendency for procedural, abstract, and relational content to be packaged as noun phrases in formal written discourse. English academic texts favor suffixal formations such as -tion and -ment, while Azerbaijani relies predominantly on -ma/-mə and -etmə nominalizations. Despite typological differences, both systems converge on nominalization as the primary strategy for proceduralization, supporting genre requirements for abstraction, condensation of actions, and syntactic embedding.

2.2 Administrative and Professional Documentation

Administrative and professional texts in both languages foreground role labeling and institutional categorization, leading to a high frequency of agentive morphology and transparent compounds. English predominantly exploits -er derivation and noun-noun compounds to denote occupational roles and functional positions. Azerbaijani, by contrast, relies extensively on -çı/-çi suffixation, often combined with transparent compounds, yielding lexemes with stable and readily interpretable meanings. These patterns

align with descriptive accounts that emphasize the importance of clarity and semantic stability in professional terminology, particularly in bureaucratic and organizational contexts.

2.3 Media and Technology Discourse

The greatest cross-linguistic divergence emerges in media- and technology-oriented discourse. English displays a strong preference for conversion and blending, strategies that enable rapid coinage and expressive compactness. Such processes are especially prevalent in innovation-driven domains, where immediacy and rhetorical impact are prioritized. Azerbaijani, while equally responsive to innovation pressures, tends to realize them through compounding and selected derivational mechanisms, preserving a higher degree of semantic transparency. This divergence underscores how similar communicative demands are accommodated through different morphological ecologies.

3. FUNCTIONAL PATHWAYS: PREDICTIONS AND OBSERVED TENDENCIES

3.1 Innovation

Observed patterns broadly confirm the Innovation pathway proposed in the Introduction. English preferentially employs conversion and blending, minimizing morphotactic complexity while maximizing expressive economy. Azerbaijani, in contrast, channels innovation through transparent compounding, maintaining formal explicitness without sacrificing productivity. Both strategies fulfill innovation-driven discourse needs, albeit via distinct morphotactic routes.

3.2 Professionalization

In contexts requiring precise designation of institutional roles, both languages converge on agentive morphology and transparent compounds. This convergence supports functional-typological predictions that professionalization favors morphologically stable and semantically explicit forms, regardless of the underlying typological profile of the language.

3.3 Evaluation

Evaluative discourse further illustrates a division of labor between morphological resources. English frequently combines evaluative adjectives with compounding and modification structures, whereas Azerbaijani prefers derivational and syntagmatic strategies. This pattern is consistent with broader observations regarding evaluative expression and the availability of morphologically encoded stance markers.

3.4 Proceduralization

Proceduralization represents the strongest site of cross-linguistic convergence. In both languages, nominalization clusters in academic and administrative registers, confirming the cross-linguistic robustness of deverbal nouns as devices for packaging complex processes into manageable discourse units.

4. TRANSPARENCY AND PRODUCTIVITY: PATTERNED CO-VARIATION

Although the present analysis does not prioritize fine-grained statistical modeling, the annotated material reveals systematic qualitative tendencies in the relationship between productivity and semantic transparency. Azerbaijani consistently maintains high transparency in highly productive derivational domains, particularly in agentive and deverbal nominalization patterns. This outcome reflects the language's agglutinative architecture, where compositional morphotactics and harmony-regulated allomorphy mitigate semantic opacity even under conditions of frequent lexical expansion.

In English, by contrast, highly productive strategies such as conversion and blending often entail reduced immediate transparency, especially when compared to headed N+N compounds. This pattern suggests that English innovation strategies may trade interpretability for expressive efficiency, a tendency widely noted in discussions of English neology and lexical creativity.

Crucially, the comparative evidence argues against a universal productivity–opacity trade-off. Instead, productivity and transparency appear to co-vary in language-specific ways, shaped by typological configuration and register-specific communicative goals. English and Azerbaijani thus exemplify distinct but equally systematic solutions to the tension between lexical expansion and semantic interpretability.

The overall results demonstrate that while English and Azerbaijani differ markedly in their available morphotactic resources, they converge functionally across registers. Discourse-level goals select from language-specific option sets, yielding parallel patterns of role labeling, procedural packaging, evaluation, and innovation. This function-driven perspective integrates naturally with word-based morphology, interface-oriented theories, and quantitative views of gradient productivity, setting the stage for the broader theoretical discussion in the following section.

IV. DISCUSSION

Functional Pathways are operationalized as mappings between recurrent discourse goals (for example, innovation, professionalization) and preferred morphotactic strategies. These mappings do not constitute deterministic rules but probabilistic alignments subject to register variation.

1. FUNCTIONAL CONVERGENCE UNDER TYPOLOGICAL DIVERGENCE

The results presented in Section 3 reveal a clear pattern of functional convergence under conditions of typological divergence. Despite substantial structural differences between English and Azerbaijani, both languages systematically converge on a limited set of discourse-level functions most prominently role labeling, procedural packaging, evaluation, and innovation across comparable registers. This finding lends empirical support to the view that functional pressures exert a strong organizing influence on derivational behavior, operating independently of, yet constrained by, the morphotactic resources of individual languages.

From a typological perspective, the English Azerbaijani comparison is instructive precisely because it contrasts an analytic/fusional system that tolerates category flexibility with an agglutinative system characterized by explicit morphological marking and harmonic allomorphy. The observed alignment of functions across registers suggests that discourse needs, rather than morphological type alone, guide the stabilization of word-formation strategies. In this respect, the findings challenge deterministic interpretations of typology that treat morphological structure as the primary driver of derivational outcomes and instead highlight the mediating role of functional selection.

2. REASSESSING PRODUCTIVITY BEYOND FORMAL MORPHOLOGY

A central contribution of this study lies in its reconceptualization of productivity as a functionally situated phenomenon rather than a purely structural property of affixes or processes. Traditional treatments often operate productivity in terms of frequency, analogy, or formal availability, abstracted away from discourse context. The present findings indicate that such abstractions are insufficient to account for observed patterns, particularly when registers with distinct communicative purposes are considered.

English conversion and blending, for example, are highly productive in media and technology discourse but remain marginal or constrained in academic and professional registers. Conversely, Azerbaijani agentive suffixation and deverbal nominalization show sustained productivity precisely in registers that prioritize semantic stability and explicit role identification. These patterns suggest that productivity is best understood as register-conditioned activation of morphotactic options, where forms become productive insofar as they efficiently serve recurring communicative tasks.

3. PRODUCTIVITY AND SEMANTIC TRANSPARENCY: A TYPOLOGY-SENSITIVE RELATION

The relationship between productivity and semantic transparency has long been framed in terms of a presumed trade-off, whereby increases in productivity correlate with reduced interpretability. The comparative evidence from English and Azerbaijani complicates this assumption. In the Azerbaijani data, highly productive derivational domains—most notably agentive *-çı/-çi* and deverbal nominalizations retain a high degree of semantic transparency, reflecting regular concatenative morphology and predictable form meaning mapping.

By contrast, English exhibits a more heterogeneous configuration. While highly productive strategies such as blending and some instances of conversion facilitate rapid lexical expansion, they may reduce immediate transparency relative to compositional compounds. Crucially, however, this pattern is

language-specific rather than universal. The findings thus argue against a global productivity–opacity trade-off and instead support a typology-sensitive view, in which the transparency consequences of productivity depend on the morphological architecture and allomorphic regularity of individual systems.

4. REGISTER SENSITIVITY AS A MISSING DIMENSION IN WORD-FORMATION THEORY

Another key implication of the study concerns the role of register sensitivity, a dimension that remains underrepresented in much theoretical work on word formation. Many models treat derivational processes as globally available resources, without sufficient attention to how register-specific constraints condition their deployment. The results here demonstrate that registering is not a secondary or peripheral factor, but a primary locus of morphological differentiation.

Academic, administrative, and media registers impose distinct communicative demands that systematically shape derivational choices in both languages. Nominalization dominates procedural discourse, agentive morphology stabilizes in professional contexts, and innovation-oriented strategies cluster in media environments. These regularities suggest that any comprehensive theory of derivational morphology must explicitly integrate register as an explanatory parameter, alongside typology and frequency.

5. TOWARD A FUNCTION-DRIVEN MODEL OF DERIVATIONAL MORPHOLOGY

Taken together, the findings motivate a function-driven model of derivational morphology, in which discourse-level goals act as selectors over the morphotactic options licensed by a particular grammatical system. In such a model, morphology does not independently determine lexical outcomes, nor does discourse override structural constraints. Rather, derivational behavior emerges from the interaction between grammatical affordances and communicative optimization.

This perspective is compatible with word-based and interface-oriented approaches to morphology, while extending them by situating derivational processes within a register-sensitive and discourse-aware framework. The comparison between English and Azerbaijani demonstrates that functional pathways operate as a unifying explanatory mechanism, accommodating cross-linguistic variation without collapsing it into structural determinism. In doing so, the analysis aligns with construction-oriented views of morphology that emphasize the interaction between lexical representations and usage-based pressures [2, 4], while also highlighting the role of communicative context in shaping derivational behavior.

V. CONCLUSION AND IMPLICATIONS

This study set out to provide a function-driven, register-controlled comparison of derivational morphology in contemporary English and Azerbaijani, with the aim of examining how two typologically distant systems respond to shared communicative pressures through distinct morphotactic resources. By integrating balanced written corpora, a transparent composite Productivity Index, and fine-grained annotation of word-formation process, register, semantic role, and semantic transparency, the analysis demonstrates that derivational behavior is most fruitfully understood as the outcome of an interaction between structural affordances and discourse-level functions, rather than as a direct reflex of morphological architecture alone.

Across academic, administrative/professional, and media registers, both languages converge on a limited set of functional outcomes, notably role labeling, procedural packaging, evaluation, and innovation. However, they diverge systematically in the formal strategies employed to realize these functions. English primarily exploits conversion, blending, and flexible compounding, especially in innovation-oriented contexts such as media and technology discourse, where speed of lexicalization, rhetorical salience, and formal economy are paramount. Azerbaijani, by contrast, relies predominantly on overt suffixation and transparent compounding, particularly in academic and professional registers, where clarity, compositionality, and stable form–meaning mapping are communicative priorities. Importantly, these differences do not indicate functional divergence. Instead, they show that similar discourse pressures select from distinct sets of morphotactic options licensed by each grammatical system, yielding functional convergence through structural divergence.

The results consistently support the Functional Pathways perspective advanced in this study. This perspective predicts that discourse-level goals act as selectors over morphologically available options, leading to convergent register effects across languages that differ substantially in typological profile. In this sense, the findings refine typological expectations by demonstrating that structural distance does not preclude functional alignment in word-formation practices. English and Azerbaijani thus exemplify two coherent yet contrasting solutions to the problem of lexical expansion under comparable communicative demands.

A central theoretical implication of the study concerns the relationship between productivity and semantic transparency. The Azerbaijani data provide qualitative support for the claim that, in agglutinative systems characterized by regular morphotactics and harmony-regulated allomorphy, high productivity can coexist with high levels of transparency, particularly in agentive derivation and deverbal nominalization. English, while also highly productive, presents a more heterogeneous profile in which certain innovation-oriented processes most notably blending and some instances of conversion may sacrifice immediate interpretability for expressive efficiency or pragmatic impact. These patterns argue against a universal productivity opacity trade-off and instead point to language-specific configurations shaped by morphological architecture, register sensitivity, and communicative goals.

Methodologically, the study contributes a replicable and extensible workflow for cross-linguistic productivity research. By combining harmonized register sampling, explicit operationalization of productivity components, transparency-sensitive annotation, and effect-oriented interpretation, the approach addresses long-standing comparability challenges in the study of derivational morphology. The analytical framework is readily adaptable to other language pairs and typological settings, offering a foundation for broader comparative investigations.

Beyond theoretical morphology, the findings have applied relevance for terminology development, professional and institutional communication, and bilingual education. In contexts where English and Azerbaijani function side by side, informed choices about role labels, procedural nouns, and innovation strategies can benefit from an understanding of the distinct yet functionally aligned word-formation preferences identified in this study. Such insights are particularly valuable in academic, administrative, and professional domains, where terminological clarity and communicative efficiency have practical consequences.

Several limitations of the study also suggest directions for future research. The present analysis is restricted to written registers and extending the design to spoken and digitally mediated discourse would allow finer-grained testing of innovation pathways and real-time lexicalization processes. Incorporating diachronic data could further illuminate how functional pressures contribute to stabilization, diffusion, or attrition of derivational patterns over time. Finally, broadening the typological scope to include additional Turkic and Indo-European languages would help determine the generality of the productivity–transparency relations observed here. Taking together, this study underscores the value of integrating functional, quantitative, and typological perspectives in the analysis of derivational morphology. By foregrounding the interaction between discourse demands and morphotactic resources, it advances a predictive account of where and why new words take hold in language use, contributing to a more nuanced understanding of morphological variation across languages.

Several limitations of the present study should be acknowledged. First, the analysis is restricted to written corpora and does not capture spoken or digitally mediated language use, where processes such as blending and conversion may display different patterns. Second, the study adopts an exploratory, distribution-based statistical approach and does not employ hierarchical modeling (for example, mixed-effects regression), which may better account for item-level and text-level variation. Third, the composite Productivity Index used in this study functions as a heuristic device rather than a fully validated statistical measure, given that it combines components from different analytical dimensions. Fourth, although efforts were made to balance the corpora, residual genre and sampling biases across languages cannot be fully excluded. Future research may address these limitations by incorporating spoken and longitudinal data, applying more advanced statistical modeling techniques, and extending the typological comparison to additional languages.

Funding Statement

The research received no specific funding.

Author Contributions

All authors contributed to the study design, data collection, analysis, and manuscript preparation. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

The author declares no competing interests.

Data Availability Statement

The annotated dataset underlying this study is provided as Supplementary. Table S1 for the purposes of peer review. All data were derived from publicly accessible written sources and contain no personal or sensitive information.

Acknowledgments

Not applicable.

Ethics Approval

This study does not involve human participants or personal data. Formal ethical approval was not required.

REFERENCES

1. Aronoff, M. (**1976**). *Word formation in generative grammar*. MIT Press.
2. Booij, G. (**2010**). *The grammar of words: An introduction to linguistic morphology* (2nd ed.). Oxford University Press.
3. Plag, I. (**2003**). *Word-formation in English*. Cambridge University Press.
4. Plag, I. (**2018**). *Word-formation in English* (2nd ed.). Cambridge University Press.
5. Lieber, R. (**2004**). *Morphology and lexical semantics*. Cambridge University Press.
6. Lieber, R. (**2009**). *Introducing morphology*. Cambridge University Press.
7. Haspelmath, M., & Sims, A. D. (**2010**). *Understanding morphology* (2nd ed.). Hodder Education.
8. Baayen, R. H. (**1992**). Quantitative aspects of morphological productivity. In G. E. Booij & J. van Marle (Eds.), *Yearbook of morphology 1991* (pp. 109–149). Kluwer Academic Publishers.
9. Bauer, L. (**2001**). *Morphological productivity*. Cambridge University Press.
10. Garibova, J., & Aliyeva, S. (**2024**). *A grammar of Azerbaijani* (Unpublished manuscript).
11. Johanson, L., & Csató, É. Á. (Eds.). (**1998**). *The Turkic languages*. Routledge.
12. Lieber, R., & Štekauer, P. (Eds.). (**2009**). *The Oxford handbook of compounding*. Oxford University Press.
13. Haspelmath, M. (**2012**). How to compare major word-formation processes across languages. In V. Renner, F. Maniez, & P. Arnaud (Eds.), *Cross-disciplinary perspectives on lexical blending* (pp. 109–130). De Gruyter Mouton.