

Documenting Lexical Innovation in Azerbaijani Youth Speech: A Methodological Framework

Nahide Heyderova

PhD Candidate, Lecturer, Nakhchivan State University, Nakhchivan, Azerbaijan

Email: nahideheyderova@ndu.edu.az

ORCID: <https://orcid.org/0009-0008-8738-4217>

© 2026 Nahide Heyderova. This is an open-access article published in *Global Spectrum of Research and Humanities (GSRH)*. Licensed under **Creative Commons Attribution 4.0 International (CC BY 4.0)** — free to share, adapt and use commercially, with attribution.

<https://creativecommons.org/licenses/by/4.0/>

DOI: <https://doi.org/10.69760/gsrh.0260304024>

Abstract

Research on youth vocabulary requires a distinction between expressions that attract attention and lexical innovations supported by evidence. This article develops a methodological framework for documenting new forms, meanings, and uses in Azerbaijani youth speech. Drawing on sociolinguistics, word formation, digital discourse research, and computational approaches to lexical change, it separates the discovery of candidate items from the verification of novelty. The proposed procedure combines contextual recording, comparison with earlier sources, multidimensional classification, and documentation of uncertain decisions. It also distinguishes spoken interaction from digitally mediated writing and treats frequency, temporal persistence, and distribution across speakers as different indicators. A standard lexical record and a constructed Azerbaijani example illustrate how original forms, normalised entries, interpretations, and verification status can be maintained without conflating them. The article argues that a useful lexical database must preserve the relationship between an item and its attested uses, including evidence that challenges an initial interpretation. The framework provides a basis for future empirical research; it reports no newly collected corpus or statistical findings.

Keywords: *Azerbaijani youth speech; lexical innovation; neologisms; slang; borrowing; lexical documentation; digital communication*

1. Introduction

An unfamiliar expression in a conversation or social media post can provide a starting point for lexical research. It cannot, by itself, establish that a new word has entered a language. The expression may be an established regional form, a temporary coinage, a borrowed item already familiar outside the observer's social circle, or a familiar word used with a different meaning. Research on Azerbaijani youth speech therefore needs procedures that turn observations into contextualised, verifiable records. The central task is to establish what is new, for whom, and relative to which earlier evidence.

Youth language is particularly suitable for examining the relationship between linguistic resources and social interaction. Sociolinguistic research places variation within the circumstances of language use rather than treating speakers as interchangeable representatives

of a language (Labov, 1972). Tagliamonte's (2016) work on adolescent language also demonstrates the value of systematic investigation in a field often discussed through anecdotes. These perspectives support studying young speakers' vocabulary through actual contexts and differences within groups. They do not justify assuming that every informal expression is a youth-specific innovation.

Digital communication makes many lexical choices visible and searchable. Crystal (2006) discusses the diversity of language associated with internet communication, while Grieve et al. (2017) demonstrate a method for examining lexical emergence in time-stamped online material. Such research offers methodological resources for a future Azerbaijani study. However, searchable material and everyday speech have different conditions of production and preservation. An expression circulating in a highly active online community cannot automatically be described as widespread among Azerbaijani young people.

This article examines collection, contextual documentation, classification, and the relationship between informal vocabulary and the standard language. It addresses three questions. How can researchers distinguish a candidate innovation from an established or insufficiently documented item? Which information must a lexical record preserve to make interpretation open to checking? How can origin, structure, social function, and evidence of diffusion be analysed without forcing each item into a single category?

The contribution is a proposed documentation procedure linking individual usage records with revisable lexical entries. It includes criteria for assessing novelty, a classification matrix, and a worked illustration. The framework is intended for future empirical investigation of Azerbaijani youth speech. It does not present a collected vocabulary list, estimate the prevalence of particular expressions, or claim that the proposed procedure has already been validated.

2. Conceptual Framework

2.1. What counts as lexical innovation

Lexical innovation can involve a new form, an additional meaning of an existing form, or a recurring expression that acquires a recognisable function. Word formation provides resources such as derivation, compounding, and shortening, while semantic processes allow familiar vocabulary to express additional distinctions. Yule (2020) offers an introductory account of these mechanisms. For documentation, the practical implication is that novelty must be investigated at the level of the form and its interpreted use. A search restricted to unfamiliar spellings will miss semantic innovation.

The object of analysis should consequently be an item associated with a meaning or function, supported by identifiable occurrences. Two uses of the same spelling may belong to different senses, while spelling variants may represent the same lexical item. Neither relationship should be decided solely through visual similarity. Researchers need contextual evidence and, where possible, speaker explanations that can be compared with other uses. An interpretation supplied by one participant remains evidence about that participant's understanding until it receives broader support.

Hohenhaus (2005) examines lexicalisation and institutionalisation, directing attention to the processes through which expressions become established. A useful distinction for the present framework is between producing an interpretable form and demonstrating its uptake beyond a single occurrence. A transparent coinage can be understandable without being a conventional expression. Conversely, an established informal word may remain absent from a

particular dictionary. Dictionary inclusion is therefore one source of evidence, not an exhaustive boundary between existing and new vocabulary.

Novelty also requires a stated comparison period and reference variety. A form may be recent within a specific online community while older elsewhere. A new sense may coexist with a longstanding sense. Researchers should make these relationships explicit by recording the earliest evidence they have verified and the limits of the sources searched. The earliest occurrence located by a project should be described in those terms; it should not be presented as the moment when the expression was first created.

2.2. Slang, jargon, and borrowing

Slang concerns informal social use and cannot be equated with chronological novelty. Eble's (1996) account of college slang relates vocabulary to sociability and group interaction. Babayev (2022) likewise discusses slang in relation to neighbouring categories, including jargon, dialectal vocabulary, and idioms. His discussion is relevant to the need for terminological differentiation. The present framework treats these categories as potentially overlapping descriptions. Their application requires contextual evidence because an item's social role can vary between interactions.

Jargon identifies vocabulary associated with a field of activity or specialist knowledge. A gaming term, for example, may function as technical vocabulary in one interaction and become an informal metaphor in another. An idiom can be widespread or socially restricted, and either old or recent. The analyst should therefore record what a label describes and the evidence supporting it. Informality, restricted distribution, figurative meaning, and unfamiliarity to outsiders are distinct observations.

Borrowing concerns the relationship between linguistic resources across languages. Durkin (2014) demonstrates the importance of historical and methodological evidence in investigating loanwords. Applied to Azerbaijani youth vocabulary, this perspective cautions against assigning a direct donor solely because a form resembles an English expression. An item may circulate through an intermediate language, be adapted locally, or share a form with a word already established in another domain. A proposed origin should remain separate from the evidence for its route of transmission.

These distinctions permit multiple annotations. A candidate may combine a borrowed base with Azerbaijani morphology, occur in an informal register, and show a meaning restricted to a particular activity. None of these properties establishes its age. Separating the dimensions makes it possible to revise the novelty assessment without discarding information about structure or social use. It also prevents a large category such as slang from absorbing every unusual expression encountered during collection.

2.3. Social meaning and diffusion

Lexical choices can contribute to affiliation, humour, evaluation, or distance between speakers. Eckert (2012) places social meaning at the centre of an account of sociolinguistic variation. This perspective supports recording how an expression works in an exchange instead of inferring identity directly from a vocabulary label. The same item may be used seriously, quoted ironically, or discussed as an example of someone else's speech. Such occurrences should remain distinguishable in the database.

Online settings add audience and technological conditions to these relationships. Androutsopoulos (2015) examines multilingual practices shaped by digital mediation and networked audiences. His findings encourage attention to language choice and the interactional

setting of an online occurrence. For an Azerbaijani project, this means recording whether a candidate appears in a public comment, a mixed-language exchange, a gaming discussion, or another setting, while avoiding claims that a platform determines one uniform language practice.

Diffusion also has temporal and social dimensions. Grieve et al. (2018) investigate the geographical spread of lexical innovation, while Würschinger (2021) distinguishes usage frequency from temporal patterns and pathways through social networks. Their studies concern English-language online material. They provide analytical distinctions, rather than findings about Azerbaijani. A future study should similarly ask whether repeated appearances involve different speakers and periods or mainly reflect concentrated activity around one event.

3. Scope and Methodological Approach

The article is a methodological synthesis, not a report of fieldwork. It draws on selected publications addressing sociolinguistic variation, youth language, lexical formation, digital communication, and computational documentation. The selection combines foundational accounts of lexical and social variation with studies addressing source provenance, temporal comparison, and digital diffusion. Sources were located and checked through publisher records, institutional or author repositories, and specialist publication databases. The selection is directed by the practical problems of documentation and does not constitute a systematic review of all research on youth vocabulary.

The proposed unit of documentation is an occurrence linked to a candidate lexical entry. The occurrence preserves the material observed in a source; the entry stores the analyst's current account of the item and its possible meanings. Keeping these levels separate allows one entry to link to several occurrences without replacing the original evidence. It also allows an occurrence to receive a revised interpretation when additional context becomes available.

The framework is assessed here through conceptual consistency and a constructed worked example. No participants were recruited, no messages were collected, and no inter-annotator agreement was measured for this article. The illustrative Azerbaijani sentence is an author-constructed example of how a record would be completed. The next stage of research would be to pilot the procedure with ethically collected material, examine ambiguous decisions, and revise the record fields and classification guidance before larger-scale collection.

4. A Documentation Procedure for Future Research

4.1. Defining the population and sources

A project should state its age range, recruitment setting, location, languages of interaction, and collection period before interpreting a vocabulary sample. Youth is not a self-explanatory sampling category. A study of university students aged eighteen to twenty-five, for example, would concern that population and could not represent school-age adolescents or all young adults. This age band is a possible design choice, not the definition adopted for every study of youth speech.

The sampling description should also identify whose language remains difficult to access. University recruitment may favour particular educational experiences, while collection through one platform may favour its active users. A balanced design need not include every setting, but it must describe the boundaries of its evidence. Bender and Friedman's (2018) proposal for data statements is useful here: a dataset should document its composition and the circumstances relevant to interpreting and reusing it.

Source types should remain separate during collection. Recorded conversation, researcher field notes, participant-submitted examples, public posts, and interview explanations provide different kinds of evidence. A note reconstructed after a conversation may preserve an expression but not its exact wording or timing. An interview can clarify a participant's interpretation without reproducing spontaneous usage. Researchers should mark these differences so that later analysis does not treat every record as an equivalent verbatim occurrence.

Digitally mediated writing also needs its own description. Tagliamonte and Denis (2008) examine instant messaging in relation to teenagers' language practices, demonstrating the value of comparing registers through linguistic evidence. For a future Azerbaijani study, online and spoken materials can be compared through shared questions, but should retain separate source labels and denominators. A platform spelling should not be represented as evidence of a spoken pronunciation unless a spoken occurrence supports that analysis.

4.2. Ethical access and contextual preservation

The collection plan should address permission, identifiability, storage, and the conditions under which excerpts may be quoted. Internet Research: Ethical Guidelines 3.0 emphasises contextual judgement across the research process (Franzke et al., 2020). For lexical documentation, this means considering the people and interaction behind a searchable expression. A publicly accessible post may still contain sensitive information or permit identification when quoted alongside other details.

Private messages should enter a research collection only under an appropriate consent and review procedure. Participation in a researcher's class or social circle should not be treated as permission to reuse conversation. If a future project includes minors, its design should address the additional consent and safeguarding arrangements required by the relevant institution. The present article proposes these design considerations; it reports no ethics approval for a study that has not been conducted.

Researchers should preserve enough context to interpret an item while limiting unnecessary personal information. A pseudonymous speaker code, broad age band, and interaction category may be sufficient for analysis. A separate restricted record can hold information needed for consent or source verification. Public outputs should avoid combining details that reveal identity. Where quotation would create a particular risk, the project should document whether it withholds an occurrence, restricts access, or uses a clearly labelled paraphrase for illustration.

4.3. Recording the original form

Each occurrence should receive a stable identifier linked to its source and collection record. The original form must be retained even when a normalised form is added for searching. A transcription, spelling correction, lemma, and English gloss serve different purposes and should occupy different fields. A normalised entry must never silently replace the text from which an interpretation was made. Where audio is available and authorised for retention, the occurrence should link to a time reference within it.

Normalisation requires explicit decisions. Azerbaijani characters and distinctions between dotted and dotless letters should be preserved; replacing them indiscriminately can merge forms that the source keeps distinct. Informal spellings, spacing, abbreviations, and repeated letters may also carry contextual information. A project can use a searchable secondary representation, but should record the rules used to produce it. Possible variants should remain provisional until context supports treating them as one item.

The record also needs an uncertainty field. An unclear transcription, uncertain sense, unidentified language, or doubtful boundary of a multiword expression should be visible to later users. This field prevents a tentative judgement from becoming an apparently settled fact through repeated copying. Table 1 outlines a minimum record that a future project can adapt to its sources and ethical arrangements.

Table 1. Minimum fields for a lexical documentation record

Field	Information to retain	Purpose
Record identifier	Stable code and link to a candidate entry	Connect analysis with individual evidence
Source and time	Source category, source date, collection date	Establish provenance and chronology
Original occurrence	Verbatim form and sufficient context, where permitted	Preserve the basis of interpretation
Search representation	Proposed lemma, segmentation, and normalisation rules	Support consistent retrieval
Meaning and function	Contextual gloss and evidence for the interpretation	Separate interpretation from spelling
Speaker and setting	Permitted pseudonymous metadata and interaction type	Describe the scope of use
Linguistic annotations	Proposed origin, formation process, and register	Permit more than one analytical label
Verification history	Sources checked, earlier evidence, and decision dates	Make the novelty assessment revisable
Uncertainty and access	Unresolved questions and disclosure restrictions	Preserve analytical and ethical limits

4.4. Checking the novelty of a candidate

The initial status of an unfamiliar item should be candidate for investigation. Researchers can then check relevant dictionaries, earlier texts, archived online material, and existing lexical collections. The record should name the sources and editions consulted, the search date, and the variants searched. Absence from one resource establishes only that the item was not found there under those conditions. It does not establish absence from the language.

For a proposed new meaning, earlier and recent contexts should be compared directly. An increased frequency of a familiar word may reflect a changing topic without a new sense. The analyst should explain which meaning distinction the later contexts support and whether alternative readings remain possible. Speaker explanations can contribute to this account, but should be compared with usage rather than treated as an automatic definition for the whole community.

A practical decision record can distinguish a provisionally supported form innovation, a provisionally supported semantic innovation, an established item, and an unresolved candidate. A change in register or community distribution can be documented separately. These statuses describe the evidence available to the project; they are not permanent properties assigned to every future occurrence. When earlier evidence is found, the entry should be revised while retaining the original decision and its rationale.

The method also needs negative decisions. Typographical errors, copied slogans, proper names, and isolated creative formations may attract attention during a search. Excluding them without explanation makes the process difficult to review. A short exclusion log can state the criterion used and retain enough information to reconsider borderline cases. A project can

analyse nonce formations as a separate phenomenon without counting them as established neologisms.

4.5. Classifying items across several dimensions

Classification should describe independent aspects of an item. Table 2 separates the main dimensions proposed here. A single expression can receive several annotations, and different occurrences can support different functions. The scheme therefore avoids forcing a borrowed and informally used derivative to become either a borrowing or slang. It also keeps novelty status separate from origin and social evaluation.

Table 2. Dimensions for classifying candidate lexical items

Dimension	Main analytical question	Possible annotations
Formation	How is the form constructed?	Derivation, compounding, shortening, multiword formation
Meaning	What relationship exists with earlier uses?	Additional sense, figurative extension, unresolved interpretation
Language contact	What evidence supports an origin or transmission route?	Borrowed base, adapted form, mixed formation, uncertain route
Social use	What does the occurrence do in its setting?	Technical reference, affiliation, humour, evaluation, quotation
Evidence status	What can the checked material establish?	Candidate, provisional innovation, established item, unresolved

Analysts should distinguish a context-supported interpretation from an expectation based on the speaker's age or group membership. An expression used by a young participant is not necessarily youth-specific. A technical item can occur in ordinary conversation, and a slang label may vary with the audience. Evidence of social restriction requires comparison with appropriate material outside the focal group. Without that comparison, the conclusion should remain limited to the sampled setting.

Consistency can be improved through a coding guide containing definitions, boundary cases, and worked decisions. A pilot can identify which fields produce disagreement and whether the categories are useful. Where independent coding is feasible, researchers should report the units and categories checked, the amount of material double-coded, the agreement procedure, and how disagreements were resolved. No numerical reliability threshold or agreement result is asserted for the present framework.

4.6. A constructed Azerbaijani illustration

Consider the sentence “Dünən onun paylaşımını laykladım” (“Yesterday I liked their post”). The sentence is constructed for this article. It is not an excerpt from a participant, a platform, or a corpus, and it is not offered as proof that the verb is new or restricted to young speakers.

The original-form field would retain *laykladım*. A separate candidate-entry field could contain *layklamaq*. A proposed morphological segmentation, *layk-la-dı-m*, distinguishes the base, verbal formation, past-tense marking, and first-person ending. This is an illustrative analysis to be checked against actual occurrences in an empirical project. The contextual gloss concerns registering a positive platform reaction, rather than every possible use of the form.

The language-contact field could identify a possible relationship with English *like* while leaving the immediate transmission route unresolved. The provenance field must state constructed example; a participant code, collection date, and frequency would be inapplicable.

The novelty field must remain unverified until earlier sources and authentic contextual evidence have been examined. The example demonstrates why a plausible morphological analysis and a familiar digital meaning cannot substitute for an evidential account of emergence.

5. Analysing Distribution and Digital Evidence

5.1. Frequency, persistence, and independent use

Occurrence counts, speaker counts, and lexical-type counts answer different questions. Repeating an expression ten times in one exchange provides a different distribution from ten independent speakers using it in separate settings. A project should state whether it counts surface forms, normalised entries, or senses. Reposts and duplicated material also need explicit treatment. They may demonstrate circulation, but cannot automatically be treated as independent acts of lexical adoption.

If samples differ in size, researchers can report frequencies relative to a stated amount of comparable material. A rate per ten thousand words, for example, should identify both the occurrences counted and the source word count. Interview responses naming an expression should not be pooled with spontaneous conversation in one frequency rate. Small samples also require the underlying counts so readers can see how much evidence supports an apparent difference.

Eisenstein et al. (2014) show how lexical diffusion in social media can preserve social and geographical differentiation. Würschinger's (2021) findings further demonstrate why similar overall frequencies need not indicate similar diffusion. For an Azerbaijani project, the corresponding methodological implication is to examine persistence across collection periods and distribution across sources and speakers. A temporary surge associated with one event should remain distinguishable from recurring use in different interactions.

Age-related interpretation requires additional care. A difference between two sampled age groups may reflect their activities, education, or platform access. A single observation period cannot establish whether a form will remain in speakers' vocabulary as they age. Repeated observation or appropriately matched comparison material would be needed to investigate those possibilities. A descriptive database can support later work by retaining the information necessary to revisit the initial comparison.

5.2. Computational assistance and human verification

Computational tools can help search variant forms, identify frequency changes, and organise contexts for inspection. Seidova et al. (2026) discuss AI and natural language processing in the modelling of neologisms. This work is relevant to an assisted discovery stage. The present framework requires every candidate selected through automation to be linked to authentic source material and an explicit verification decision. A generated expression is not an observed occurrence.

Semantic-change research offers additional resources. Hamilton et al. (2016) use diachronic word representations to investigate changes in meaning, and Schlechtweg et al. (2020) provide an evaluation framework and manually annotated datasets for lexical semantic change detection. These studies establish the importance of temporal comparison and evaluation. Their methods and results should not be assumed to transfer directly to small, heterogeneous Azerbaijani youth-language collections.

Automated analysis should therefore be piloted against manually examined material from the intended setting. Errors involving spelling variants, segmentation, mixed-language material, or sparse contextual evidence need to be documented. Hovy and Prabhumoye (2021) locate potential bias throughout data selection, annotation, representations, models, and research design. For this proposal, the implication is to inspect how a tool's decisions affect which speakers and linguistic forms become visible. Automation should remain an accountable component of the research procedure.

6. Standard Language and a Revisable Lexical Database

Documenting informal innovation does not require predicting that every item will enter the standard language. Some expressions remain associated with an activity, audience, or moment. Others acquire wider distribution, and their meanings or stylistic associations may change. A database can record these developments without treating standardisation as the inevitable endpoint of lexical change. Evidence of persistence and broader use should be distinguished from a normative judgement about appropriateness.

The relationship between description and language education can also be handled through context. An expression suitable for peer conversation may be unsuitable for a formal administrative text without being linguistically meaningless or evidence of language decline. Research can document these differences and make speakers' interpretations available for analysis. It should avoid assigning value to the young people being studied on the basis of their vocabulary choices.

A reusable database should preserve several linked layers: the restricted source record, permitted occurrence-level evidence, the candidate entry, and the history of analytical decisions. A public version may provide selected examples, glosses, classification labels, and a concise statement of evidential limits. Sensitive material can remain restricted or be omitted. Updates should carry dates and reasons so that a revised gloss or an earlier attestation does not erase the route by which the conclusion was reached.

The first empirical application should be limited enough to examine whether the scheme works. A small pilot across clearly specified source types could test the clarity of record fields, the treatment of variants, and the distinction between occurrence-level and entry-level judgements. Findings from that pilot would justify changes before expansion. Later research could investigate differences between settings, follow individual items over time, or compare spoken and written uses while retaining comparable documentation practices.

The framework has limits. Its literature base is selective and draws substantially on research involving English and other languages. It does not establish the distribution of any Azerbaijani item or demonstrate that the proposed categories exhaust the relevant phenomena. Its value depends on future empirical use, transparent revision, and engagement with the speakers whose practices are represented. These limitations define the next research tasks rather than being resolved through the size of a vocabulary list.

7. Conclusion

Research on lexical innovation in Azerbaijani youth speech needs a clear connection between an expression, its context, and the evidence supporting its interpretation. The framework developed here separates the identification of candidates from the assessment of novelty and treats origin, formation, meaning, and social function as distinct dimensions. This permits overlapping classifications without equating borrowing, slang, or informality with newness.

The proposed procedure preserves original occurrences, records normalisation and uncertainty, checks earlier evidence, and distinguishes frequency from wider diffusion. Its constructed example shows how an apparently familiar digital expression can be analysed without claiming an unverified history or distribution. Applied through a carefully documented pilot, the framework could support a revisable lexical database and future comparisons across speakers, settings, and periods. The next step is empirical evaluation with appropriately collected material.

Declarations

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

Ethics and Data Availability: This methodological article reports no newly collected participant data. The worked sentence is constructed for illustration. No original research dataset accompanies the article.

Use of Generative AI: Generative AI assisted with English rewriting, literature discovery, reference checking, and document formatting; all content was reviewed and verified by the author.

Author Contributions: Conceptualization, analysis, drafting, and revision: N.H.

How to Cite:

Heyderova, N. (2026). Documenting lexical innovation in Azerbaijani youth speech: A methodological framework. *Global Spectrum of Research and Humanities*, 3(4), 238-249. <https://doi.org/10.69760/gsrh.0260304024>

References

- Androutsopoulos, J. (2015). Networked multilingualism: Some language practices on Facebook and their implications. *International Journal of Bilingualism*, 19(2), 185–205. <https://doi.org/10.1177/1367006913489198>
- Babayev, J. (2022). Slang and its characteristic features. In G. Y. Gulyaev (Ed.), *European research: Proceedings of the XXXV international scientific and practical conference* (pp. 152–154). Nauka i Prosveshchenie.
- Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 6, 587–604. https://doi.org/10.1162/tacl_a_00041
- Crystal, D. (2006). *Language and the Internet* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511487002>
- Durkin, P. (2014). *Borrowed words: A history of loanwords in English*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199574995.001.0001>

- Eble, C. (1996). *Slang and sociability: In-group language among college students*. University of North Carolina Press.
- Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, 41, 87–100. <https://doi.org/10.1146/annurev-anthro-092611-145828>
- Eisenstein, J., O'Connor, B., Smith, N. A., & Xing, E. P. (2014). Diffusion of lexical change in social media. *PLOS ONE*, 9(11), Article e113114. <https://doi.org/10.1371/journal.pone.0113114>
- franzke, a. s., Bechmann, A., Zimmer, M., Ess, C., & Association of Internet Researchers. (2020). *Internet research: Ethical guidelines 3.0*. Association of Internet Researchers. <https://aoir.org/reports/ethics3.pdf>
- Grieve, J., Nini, A., & Guo, D. (2017). Analyzing lexical emergence in modern American English online. *English Language & Linguistics*, 21(1), 99–127. <https://doi.org/10.1017/S1360674316000113>
- Grieve, J., Nini, A., & Guo, D. (2018). Mapping lexical innovation on American social media. *Journal of English Linguistics*, 46(4), 293–319. <https://doi.org/10.1177/0075424218793191>
- Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. In K. Erk & N. A. Smith (Eds.), *Proceedings of the 54th annual meeting of the Association for Computational Linguistics (Volume 1: Long papers)* (pp. 1489–1501). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P16-1141>
- Hohenhaus, P. (2005). Lexicalization and institutionalization. In P. Štekauer & R. Lieber (Eds.), *Handbook of word-formation* (pp. 353–373). Springer. https://doi.org/10.1007/1-4020-3596-9_15
- Hovy, D., & Prabhumoye, S. (2021). Five sources of bias in natural language processing. *Language and Linguistics Compass*, 15(8), Article e12432. <https://doi.org/10.1111/lnc3.12432>
- Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.
- Schlechtweg, D., McGillivray, B., Hengchen, S., Dubossarsky, H., & Tahmasebi, N. (2020). SemEval-2020 Task 1: Unsupervised lexical semantic change detection. In A. Herbelot, X. Zhu, A. Palmer, N. Schneider, J. May, & E. Shutova (Eds.), *Proceedings of the fourteenth workshop on semantic evaluation* (pp. 1–23). International Committee for Computational Linguistics. <https://doi.org/10.18653/v1/2020.semeval-1.1>
- Seidova, E., Babayev, J., Sadikhova, S., Aghayeva, G., & Mammadova, Z. (2026). AI-powered lexical innovation: Modeling neologisms with NLP. *Scientific Work*, 20(3), 171–179. <https://doi.org/10.36719/2663-4619/127/171-179>
- Tagliamonte, S. A. (2016). *Teen talk: The language of adolescents*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139583800>
- Tagliamonte, S. A., & Denis, D. (2008). Linguistic ruin? LOL! Instant messaging and teen language. *American Speech*, 83(1), 3–34. <https://doi.org/10.1215/00031283-2008-001>
- Würschinger, Q. (2021). Social networks of lexical innovation: Investigating the social dynamics of diffusion of neologisms on Twitter. *Frontiers in Artificial Intelligence*, 4, Article 648583. <https://doi.org/10.3389/frai.2021.648583>

Yule, G. (2020). *The study of language* (7th ed.). Cambridge University Press.
<https://doi.org/10.1017/9781108582889>

Received: 2 September 2026

Accepted: 20 October 2026

Published: 22 October 2026