

From Subjectivity to Neutrality: A Data-Informed Pedagogical Framework for Teaching Lexicogrammatical Strategies in International English Communication

Kim Tony Dizon [1,2]
[1] College of Arts and Sciences, World Citi Colleges, Philippines
[2] Graduate School, University of Perpetual Help System DALTA, Philippines

Corresponding Author: * Kim Tony Dizon

DOI: https://doi.org/10.5281/zenodo.23334453

Abstract

The pedagogy of international English writing requires more than the mechanical removal of subjective words: writers must consider whether evaluation, stance and evidential commitment are appropriate to their audience and purpose. This study reanalyzed 55,503 paired English Wikipedia revisions from the publicly released Wiki Neutrality Corpus and developed a data-informed instructional framework for teaching lexicogrammatical decision-making. Word Piece-level edit alignment and source part-of-speech tags were analyzed exhaustively. Of the pairs, 29,065 (52.37%) involved deletion, 26,289 (47.37%) substitution, 23 (0.04%) insertion and 126 (0.23%) multiple edit spans. The changed source token was most often tagged as an adjective (41.03%), noun (23.64%), adverb (17.22%) or verb (+14.71%). Deletions were especially frequent among adjective and adverb edits, whereas noun and verb changes predominantly involved substitutions. These patterns inform a five-stage instructional design emphasizing noticing, meaning interpretation, evidence checking, context-sensitive revision and reflective justification. Counterexamples show why an archived neutrality edit should not automatically be treated as semantically safer: removing a marker such as "allegedly" can strengthen an allegation into an assertion. Findings describe the archived edits, not international English usage broadly. The teaching framework and assessment rubric are proposed products, not experimentally validated interventions.
.

Manuscript Information

- **ISSN No:** 3139-9851
- **Received:** 29-08-2026
- **Accepted:** 12-09-2026
- **Published:** 13-10-2026
- **IJCATM:**1(3); 2026: 93-99
- **©2026, All Rights Reserved**
- **Plagiarism Checked:** Yes
- **Peer Review Process:** Yes
- **Paper type:** Research paper

How to Cite

Dizon KT. From Subjectivity to Neutrality: A Data-Informed Pedagogical Framework for Teaching Lexicogrammatical Strategies in International English Communication. Int. J. Current Acad. Trend. Multi. 2026;1(3): 93-99

Access this Article Online

www.ijcatmjournal.com

KEYWORDS: Lexicogrammar; English Language Pedagogy; Stance; Secondary Textual Data; International English Communication

© 2026 Kim Tony Dizon· This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY NC ND).https://creativecommons.org/licenses/by/4.0/

1. INTRODUCTION

English writing for international readers involves a complex relationship between clarity, evidence, interpersonal meaning and rhetorical purpose. In an academic report, the statement that an institution is "prestigious" may be inappropriate without an explicit evidentiary basis; in an appropriately attributed review, evaluative vocabulary can be central to the author's task. Teaching neutrality therefore cannot be reduced to prohibiting adjectives, avoiding opinion or adopting a supposedly universal standard of English.

Systemic Functional Linguistics locates evaluation, engagement and graduation within the interpersonal choices that writers make (Martin & White, 2005). Hyland (2005) likewise shows how stance and engagement help writers position claims relative to readers. For English as an international language, the focus must include audience, genre, cultural plurality and communicative purposes (Matsuda & Friedrich, 2011). Taken together, these approaches support explicit instruction in choosing and justifying wording, rather than automatic bias removal.

Paired revisions from English Wikipedia provide authentic records of editing choices. Pryzant et al. (2020) released an English-language dataset designed for computational bias identification and neutralization. Its 55,503-pair single-word-edit subset offers a manageable basis for investigating which lexical categories are changed, how changes are structurally realized, and what kinds of questions teachers might ask students about stance and semantic fidelity. The source dataset is associated with Wikipedia's neutral-point-of-view conventions; it is not a representative sample of academic, professional or international English across settings.

Prior scholarship has examined bias markers and computational approaches to revising them (Recasens et al., 2013; Pryzant et al., 2020). Reviews of data-driven language learning have meanwhile considered the possibilities and limitations of using authentic linguistic evidence in instruction (Boulton & Cobb, 2017; Chen & Flowerdew, 2018; Casal & Kessler, 2024). The contribution of this paper is a transparent bridge between evidence on sentence revision and a teachable process of noticing, interpretation, contextual evaluation and revision. It does not claim that a proposed teaching sequence has yet improved student performance.

Research questions and analytical orientation

RQ1. How frequently do deletion, substitution, insertion and multi-span revisions occur within the selected English Wikipedia single-word revision dataset?

RQ2. Which source part-of-speech categories account for revisions, and how do edit operations differ across adjective, adverb, noun and verb categories?

RQ3. Which observable examples can support teaching the distinction between unnecessary evaluation, degree/intensity, attribution, lexically loaded naming and legitimate epistemic qualification?

RQ4. How can these observed revisions inform a provisional pedagogical sequence and assessment criteria for international English writing while preserving context, evidential accuracy and writer agency?

Theoretical basis

Appraisal and epistemic stance

Martin and White's (2005) Appraisal perspective distinguishes Attitude (evaluation), Engagement (positioning a proposition among possible voices) and Graduation (scaling intensity or category membership). These distinctions explain why removing "prestigious," replacing "unfortunately" with "however," and removing "allegedly" are not equivalent moves. The first may remove praise, the second changes an affective evaluation into discourse contrast, and the third potentially changes evidential status. Hyland's (2005) stance framework underscores that hedges and boosters may be rhetorically needed in scholarly writing rather than being sources of error by definition.

International English writing as situated practice

International communication includes informational summaries, institutional statements, research writing, correspondence and critical argument. A norm suited to an encyclopedia may be inappropriate for a persuasive essay or an attributed policy critique. An English-as-an-international-language perspective calls for explicit recognition of audiences and acceptable variation (Matsuda & Friedrich, 2011). In this study, "neutrality" is accordingly treated as context-sensitive management of stance, not a ban on strong language or an insistence on value-free communication. The relevance of audience and intercultural context is also reflected in research on virtual exchange among university students, which examines cross-cultural communication and global collaboration rather than treating effective expression as a single context-free standard (Nuryanti et al., 2025).

From archival examples to data-informed teaching

Authentic before-and-after text pairs can make less visible features of meaning noticeable. Reviews of data-driven learning support the pedagogical potential of linguistic data, while emphasizing the need to fit tasks, materials and teacher mediation to specific learners (Boulton & Cobb, 2017; Chen & Flowerdew, 2018; Casal & Kessler, 2024). The teaching design developed below uses curated sentence pairs as prompts for metalinguistic reasoning, not as ready-made rules or as evidence of achieved learning outcomes. Guided inquiry has likewise been investigated as a means of organizing active observation and interpretation in science education (Sugiharti & Osma, 2024). This cross-disciplinary example informs the sequence of noticing, questioning and justification proposed here, without establishing its effectiveness for English writing.

Materials and methods

Design, source and extraction

The study employed quantitative descriptive analysis of a secondary English textual dataset and a theory-informed pedagogical design stage. The archive bias_data.zip was provided for analysis on 9 October 2026. The initial analytical frame was the three supplied files biased. word. train, biased. word. dev and biased. word. test within bias_ data/WNC. These files contain, respectively, 53,803, 700 and 1,000 records, for 55,503 records. The source collection and file

© 2026 Kim Tony Dizon· This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY NC ND).https://creativecommons.org/licenses/by/4.0/

descriptions are documented in the README accompanying the download from https://github.com/rpryzant/neutralizing-bias. The archive SHA-256 checksum was cd9438f1ab9886342823d3f7ffd8f34773bf10559363041363ba577fad22b7f7.

Each tab-separated line includes a revision ID, Word Piece-tokenized pre-edit and post-edit strings, raw pre-edit and post-edit strings, source part-of-speech tags and parse information. Analysis used the original paired strings and supplied linguistic tags without training a language model. The archive also includes a much larger full-edit collection and nearby neutral sentences; neither was incorporated into the present empirical denominator.

Screening and unit of analysis

A paired pre-edit/post-edit record is the unit of analysis. Records were imported across the three single-word files; exact duplication was assessed using the pair of raw sentence strings. Screening for empty sides and identical source–target strings identified no such records. The resulting analytical file comprised all 55,503 unique, non-identical pairs. Unique revision IDs totaled 55,497, confirming that revision identifiers are not perfectly unique and must not be used alone as the primary key. Records were not screened for substantive validity through full manual coding; inclusion therefore reflects data integrity rather than independently verified neutralization.

Computed variables and analytic procedure

Each Word Piece-tokenized source–target pair was aligned using Python's difflib. Sequence Matcher with auto junk disabled. Non-equal spans were classified as deletion, replacement, insertion or multiple edit spans. For a record with one changed span and a nonempty source span, the source part-of-speech category was taken from the first token in that span. Source token counts and POS counts were checked for alignment; all 55,503 records had the same number of sources Word Piece tokens and POS tags. For insertion and multiple-span records, the primary source category was recorded as unavailable for this specific analysis. Edit frequencies and source POS distributions were reported as counts and percentages of the complete single-word dataset; confidence intervals were not needed for these observed archive counts. A restricted 4 × 2 cross-tabulation compared deletion and replacement by adjective, adverb, noun and verb labels. Pearson's chi-square and Cramér's V were calculated as exploratory summaries of association. Independence is not assured because separate Wikipedia edits can come from related pages or editorial routines; conventional inferential claims beyond this archival dataset should therefore be avoided.

Finally, exact source-token strings were used to identify illustrative markers such as "prestigious," "unfortunately," "only" and "allegedly." These form-level counts are descriptive and should not be equated with counts of discourse functions. A 300-pair fixed-seed review worksheet has been generated for potential later independent functional annotation, but it has not been coded or used to assert inter-rater agreement. The examples presented here were inspected as teaching contrasts only.

Quality, ethics and reproducibility

The source pairs record changes made under Wikipedia neutrality editing practices, but an editing decision is not necessarily a linguistically or ethically better choice. Specific revisions can remove warranted evaluation, weaken precision, obscure agency or strengthen unverified allegations. No primary participants were recruited and no classrooms were observed; institutional rules should determine whether ethics exemption documentation is needed. Reproducibility material includes the original archive checksum, exact file names, code used to align and summarize edits, aggregated metrics, and an unannotated fixed-seed worksheet. The analytic scripts should accompany journal submission as supplementary files, subject to the repository's reuse terms.

RESULTS

Imported records and operation profile

All 55,503 records were nonempty, distinct at the exact raw-pair level and non-identical across their before-and-after strings. Deletion was slightly more common than replacement, with the remaining records representing insertions or multiple edit spans (Table 1). The dominance of deletion and replacement is partly expected because the source subset was curated around localized edits rather than extensive rewriting; it is not evidence that all natural editing proceeds through these operations.

Table :1 Edit-operation distribution in the complete analyzed dataset (N = 55,503)

Operation	n	% of records
Deletion	29,065	52.37
Replacement	26,289	47.37
Insertion	23	0.04
Multiple edit spans	126	0.23
Total	55,503	100.00

Note. Percentages are rounded to two decimals and may not sum exactly to 100

.

Lexicogrammatical categories and edit form

Adjectives accounted for 22,775 changed-source labels (41.03%), nouns for 13,120 (23.64%), adverbs for 9,558 (17.22%) and verbs for 8,164 (14.71%). These four groups comprised 96.60% of all records. Other source POS tags occurred less frequently; 149 records involving insertion or multiple spans had no single identifiable source POS using the present operational rule. The prominence of adjectives and adverbs aligns with opportunities to teach degree and

© 2026 Kim Tony Dizon· This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY NC ND).https://creativecommons.org/licenses/by/4.0/

valuation, but category frequencies alone cannot establish which individual edits were appropriate.

Table: 2 Source part-of-speech category by major edit operation

Source POS	Deletion n	Replacement n	Total n
Adjective	16,982	5,793	22,775
Adverb	7,827	1,731	9,558
Noun	2,540	10,580	13,120
Verb	1,295	6,869	8,164
Total	28,644	24,973	53,617

Note. Restricted to four common POS classes and delete/replace records; Word Piece source tags are inherited from the dataset. Totals exclude remaining POS classes and insertion/multi-span edits

The association between source category and edit operation was substantial within these 53,617 selected records (χ^2 [3] = 17,950.37, $p < .001$; Cramér's V = .579). Adjectives and adverbs were usually deleted (74.6% and 81.9% within their respective rows), while nouns and verbs were usually replaced (80.6% and 84.1%). These are archival co-occurrences, not evidence that a specific teaching intervention causes improved writing. Similar word-edit templates and related articles may contribute dependencies and inflated statistical precision.

Recurring observable lexical forms

Exact changed-source marker counts illustrated the range of revision functions (Table 3). For example, "prestigious" appeared as the source changed token in 526 records, with 508 deletions; "unfortunately" occurred in 741 records, with 458 deletions and 283 other changes, including replacements. These counts do not indicate independent communicative events or internationally representative preferences. Repeated editorial conventions and topical clusters can contribute many closely related examples.

Table: 3 Selected exact source markers in localized revisions

Changed source marker	All edits n	Deletions n	Illustrative teaching focus
prestigious	526	508	Evaluative adjective
famous	499	409	Unattributed prominence
controversial	416	383	Evaluative characterization
nfortunately	741	458	Affective stance / contrast
great	508	388	Positive appraisal
only	602	577	Focus and scalar implication
allegedly	64	43	Evidential qualification
very	176	162	Degree / graduation

Teaching contrasts and counterexamples

Table 4 presents archive-grounded contrasts and a purposefully critical reading of their implications. Contextual correctness cannot be assessed conclusively from isolated sentence pairs alone, and the table is not a validity-rated random sample.

Table :4 Selected authentic paired revisions and pedagogical interpretation

Archive revision ID	Before → after (short extract)	Question for learners
182463892	"Prestigious universities" → "universities"	What evidence would justify the adjective?
135358434	"Unfortunately, this is not the case." → "However, this is not the case."	How does an affective judgment differ from a logical contrast?
354411195	"Notorious for" → "known for"	Does a lexical replacement change appraisal while preserving the proposition?
712028496	"Actress and comedienne" → "actress and comedian"	What does occupational naming imply about gender and register?
166423956	"Only 42 years of age" → "42 years of age"	What expectation or emotional framing does "only" add?
178006122	"Allegedly by" → "by"	Why might deleting a hedge dangerously strengthen an unverified allegation?

Note. Quoted contrasts are brief excerpts of records in the uploaded Wiki Neutrality Corpus. Revision IDs enable checks against the original edit; examples are interpretive prompts, not claims of universally preferable wording.

A data-informed pedagogical framework

Design principle: context before mechanical neutrality

The instructional objective is not to train learners to delete evaluative vocabulary automatically. It is to develop evidence-sensitive decisions about interpersonal meaning across tasks. A direct lesson can combine lexical appraisal, degree modifiers, evidential hedges, person reference, semantic fidelity and audience awareness. The teacher must emphasize that stance is a resource to be controlled, not a flaw to be eradicated.

Proposed five-stage learning sequence

© 2026 Kim Tony Dizon· This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY NC ND).https://creativecommons.org/licenses/by/4.0/

Table: 5 Data-informed revision cycle for international English writing

Stage	Teacher–student learning activity	Assessable output
1. Notice	Compare short paired revisions; highlight changed lexicogrammatical form.	Marked linguistic trigger
2.Interpret	Classify its effect on evaluation, intensity, attribution, naming or certainty.	Brief form–meaning explanation
3. Verify	Check whether the original and revision preserve factual content and warranted uncertainty.	Evidence and meaning check
4. Revise	Produce an alternative for a specified audience (report, institutional email, policy brief).	Revised passage with edit log
5. Justify	Defend the choice; discuss an occasion when retaining strong or hedged language is better.	Short metalinguistic rationale

Note. Proposed instructional design derived from the descriptive findings and relevant pedagogy literature; not classroom-tested

Sample instructional application

At the noticing stage, learners compare "prestigious universities" with "universities" and identify the removed evaluative adjective. At interpretation, they explain what would make the adjective defensible in an institutional profile and why it may be out of place in an informational summary. At verification, learners identify evidence that could substantiate an institutional claim. At revision, they create one formulation for a factual description and one for a sourced promotional text. At justification, they explain why the two genres require different decisions rather than treating one as universally correct.

A second, deliberately problematic example uses "allegedly by" versus "by." In an unsourced or disputed context, eliminating the hedge may imply established responsibility that the evidence does not support. Learners should identify the epistemic change and supply a responsible alternative with attribution if supported. This adversarial case prevents the teaching material from confusing archival editor action with normative linguistic correctness.

Proposed analytic assessment rubric

Table: 6 Draft rubric for evaluating justified English revisions

Dimension	0 = Limited	1 = Developing	2 = Proficient
Form–function identification	No relevant feature found	Feature noted with vague function	Relevant form and meaning identified
Semantic fidelity	Claim materially distorted	Core information mostly preserved	Claim and warranted uncertainty preserved
Evidence and attribution	Unwarranted certainty or judgment	Some sourcing or qualification	Evidence, sourcing and stance calibrated
Audience and genre	Register mismatched	Partially adjusted to situation	Choices aligned to purpose and audience
Reflective justification	Mechanical rule invoked	Some contextual rationale	Reasoned defense incl. possible counter case

Note. This 0–10 point instructional rubric is newly proposed; classroom scoring reliability and validity have not been tested

.

DISCUSSIONS

The dominant operation patterns reveal two pedagogically distinguishable challenges. First, an adjective or adverb can contribute stance that is removable without eliminating the core event (as when an unsupported favorable modifier is deleted). Second, nominal and verbal choices may require replacement because deleting a participant label or action often destroys the grammar or alters the propositional content. The observed association between source category and editing operation provides an empirical rationale for teaching deletion and replacement as different decisions, rather than a single generalized strategy of "making language objective."

More fundamentally, the data reveal why language pedagogy must foreground meaning and attribution. Some lexical revisions appear to reduce unnecessary endorsement or condemnation; others risk blurring distinctions. The deletion of "allegedly" is a particularly consequential case: an apparently simple adverb edit can alter whether responsibility is asserted or reported as uncertain. A lesson organized around such contrasts can develop learners' capacity to argue from evidence and to notice how interpersonal meanings are packaged in grammar. This remains a plausible instructional mechanism, not an observed achievement of learners.

The framework should be interpreted as a designed teaching resource rather than a validated treatment. International English use is heterogeneous, and even were informational genres value restrained assessment, writers still need legitimate stance for argument, critique, negotiation and advocacy. A decision-oriented approach permits explicit evaluation when it is sourced and genre-appropriate. Future evaluation could compare learner revisions before and after instruction and examine scoring agreement, delayed retention and transfer across communication contexts, but those questions cannot be answered by this archival analysis alone.

Limitations and implications for subsequent validation

First, the original dataset was created through Wikipedia neutrality-related revision processes, not through multilingual classroom observation. Genre-specific editorial decisions should not be generalized to international communication. Second, a localized Word Piece edit is not always a complete semantic edit, and the first source POS tag may misrepresent constructions or parsing errors. Third, exact nonduplication does not remove near-duplicates or clustering within articles, as repeated forms (including occupational-name edits) demonstrate. Fourth, the functional interpretations in this manuscript are pedagogical readings of selected contrasts: no independent coder has yet adjudicated a random sample, and no agreement statistic is claimed. Fifth, no learners, teachers, classes or outcomes were sampled; the rubric and instructional sequence lack empirical validation. A stronger submission should add an independently double-coded and adjudicated set

© 2026 Kim Tony Dizon· This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY NC ND).https://creativecommons.org/licenses/by/4.0/

of randomly selected pairs and make the annotation data accessible.

CONCLUSIONS

The reanalysis of 55,503 paired English Wikipedia revisions identified a reproducible pattern: deletions were common in adjective and adverb edits, whereas nouns and verbs were more often replaced. Selected revisions show how stance, degree, interpersonal evaluation and evidential commitment can change through localized lexicogrammatical choices. These observations support a provisional data-informed framework in which learners notice a form, interpret its discourse function, verify evidence and meaning, revise for a specific audience, and justify the result. The contribution is the transparent translation of secondary textual evidence into defensible teaching tasks, not a claim of experimentally established effectiveness.

Research declarations

Acknowledgements

The author acknowledges Reid Pryzant, Richard Diehl Martinez, Nathan Dass, Sadao Kurohashi, Dan Jurafsky, and Diyi Yang for making the Wiki Neutrality Corpus publicly available for research. The author also recognizes the English Wikipedia contributors whose archived sentence revisions provided the source material for this secondary-data analysis (Pryzant et al., 2020).

Funding: This research received no funding.

Conflicts of interest: The author declares no conflicts of interest.

Ethics: This research analyzed publicly released archival English-language texts. No primary human participants were recruited, contacted, or subjected to an intervention; no new participant data were collected.

Data availability: The source dataset, the Wiki Neutrality Corpus, is available through the original project repository at https://github.com/rpryzant/neutralizing-bias. The accompanying supplementary ZIP archive contains the analysis scripts, screening and results summaries, and an unannotated 300-record review worksheet. Any redistribution of source excerpts must comply with applicable licensing and attribution requirements.

AI assistance: ChatGPT was used to assist with computational scripting, initial text drafting, language editing, and manuscript review. The author reviewed the analyses, checked the cited references, made the final interpretive and editorial decisions, and accepts full responsibility for the accuracy and integrity of the manuscript. This use is disclosed in accordance with the journal's AI-assistance policy.

REFERENCES

1. Boulton A, Cobb T. Corpus use in language learning: a meta-analysis. Language Learning. 2017;67(2):348-393. doi:10.1111/lang.12224.
2. Casal JE, Kessler M. New theoretical and practical horizons in data-driven learning: Introduction to the special issue. TESOL Quarterly. 2024;58(3):1027-1045. doi:10.1002/tesq.3331.
3. Chen M, Flowerdew J. A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. International Journal of Corpus Linguistics. 2018;23(3):335-369. doi: 10.1075/ijcl.16130.che.
4. Hyland K. Stance and engagement: A model of interaction in academic discourse. Discourse Studies. 2005;7(2):173-192. doi:10.1177/1461445605050365.
5. Martin JR, White PRR. The language of evaluation: Appraisal in English. Palgrave Macmillan; 2005. doi:10.1057/9780230511910.
6. Matsuda A, Friedrich P. English as an international language: A curriculum blueprint. World Englishes. 2011;30(3):332-344. doi:10.1111/j.1467-971X.2011.01717. x.
7. Nuryanti D, Kusumastuty I, Osma AN. Virtual exchange programs for cross cultural communication and global collaboration among university students. International Journal of Educational Technology and Society. 2025;1(3):52-61. http://dx.doi.org/10.61132/ijets.v1i3.413
8. Pryzant R, Diehl Martinez R, Dass N, Kurohashi S, Jurafsky D, Yang D. Automatically neutralizing subjective bias in text. Proceedings of the AAAI Conference on Artificial Intelligence. 2020;34(1):480-489. doi:10.1609/aaai. v34i01.5385.
9. Recasens M, Danescu-Niculescu-Mizil C, Jurafsky D. Linguistic models for analyzing and detecting biased language. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics; 2013. p. 1650-1659. Available from: https://aclanthology.org/P13-1162/
10. Sugiharti T, Osma AN. Guided inquiry learning model to improve scientific literacy on energy and its transformation. International Journal of Science and Mathematics Education. 2024;1(3):34-43. https://doi.org/10.62951/ijsme.v1i3.296

© 2026 Kim Tony Dizon· This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY NC ND).https://creativecommons.org/licenses/by/4.0/

Appendix A. Operational annotation codebook for further validation

Table A1. Proposed functional labels for an independently reviewed sample

Functional label	Coding question	Disconfirming or cautionary condition
Evaluation / attitude	Does wording endorse, praise, condemn or attribute values?	Evaluation may be legitimate if sourced or genre required
Graduation / intensity	Does wording amplify, restrict or scale the proposition?	Removal may erase meaningful degree or expectation
Epistemic stance	Does wording assert, hedge or attribute certainty?	Deleting a hedge may create an unsupported fact claim
Participant reference	Does wording label a group, person or profession?	An updated label can still distort referential identity
Agency / attribution	Does wording assign an actor or source of a claim?	Replacing a verb may change agency or legal meaning
Other / non-target	Is the edit orthographic, formatting or factual updating?	Do not code as neutrality strategy without evidence

Note. Allow primary and secondary codes; separately score whether meaning is preserved, uncertain or altered. Coding and reliability are not yet completed.

Annotation recommendation: Use a fixed random seed (20261009), independently double-code at least 20% of the proposed 300-pair review sample, discuss discrepancies, and report Cohen's κ or an appropriate agreement coefficient for each sufficiently prevalent label. Preserve a record of excluded or disputed cases. The worksheet currently contains blank fields intentionally.

Appendix B. Reproducibility and reporting checklist

- Preserve uploaded bias_data.zip unchanged and retain SHA-256 checksum shown in §4.1.
- Report single-word train/dev/test split counts and do not substitute counts from biased. full or neutral.
- Disclose that the source dataset consists of neutrality-related Wikipedia edits, not representative international English.
- Keep original revision IDs and source–target pairs for audit; avoid reproducing long quotations unnecessarily.
- Inspect both archive source and target before interpreting a change as a linguistic improvement.
- Document operation extraction, tokenization, POS-matching logic and source software versions.
- Complete independent annotation and inter-rater reliability before making prevalence claims about semantic functions.
- Validate any classroom teaching rubric using real learner work before claiming pedagogical effectiveness.

Creative Commons (CC) License

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution–Non-Commercial–No Derivatives 4.0 International (CC BY-NC-ND 4.0) license. This license permits sharing and redistribution of the article in any medium or format for non-commercial purposes only, provided that appropriate credit is given to the original author(s) and source. No modifications, adaptations, or derivative works are permitted under this license.

© 2026 Kim Tony Dizon· This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY NC ND).https://creativecommons.org/licenses/by/4.0/