



ANIMACY, TRANSITIVITY, AND READER RECEPTION: A CORPUS-BASED ECOLINGUISTIC STUDY OF ROBERT MACFARLANE'S IS A RIVER ALIVE?

Natasha Mustafa
nataliamustafa726@gmail.com

MPhil Scholar, Department of English, University of Okara, Okara,
Pakistan

Marya Sarwar
maryaburhan1983@gmail.com

Lecturer in English, Department of English, University of Okara,
Pakistan

Nida Ashiq
nidach1696635@gmail.com

M.Phil English Linguistics, University of Okara, Okara, Pakistan

Shumaila
shumaylchohan5@gmail.com

MPhil Scholar, Department of English, University of Okara, Pakistan

Abstract

Rivers are conventionally treated in English discourse as inanimate resources bodies to be measured, dammed, and managed rather than beings with standing of their own. This study asks whether literary language can reverse that convention, and whether such a reversal is measurable rather than merely impressionistic. Using a corpus-based ecolinguistic approach, the study analyses the complete 106,649-word corpus of Robert Macfarlane's *Is a River Alive?* (2025) through the combined lenses of Arran Stibbe's *Stories-We-Live-By* framework, Hallidayan transitivity, and Robin Wall Kimmerer's grammar of animacy. Word frequency, keyword, collocational, and n-gram analyses were generated through a custom Python natural language processing (NLP) pipeline built on the Natural Language Toolkit (NLTK) and benchmarked against the British National Corpus; 200 concordance lines were qualitatively coded in NVivo; and a Linguistic Perception Questionnaire was administered to 150 Generation Z university students to test how readers respond to the text's grammatical strategies. Rivers were assigned the animate relative pronoun *who* over the inanimate *which* at a ratio of 8.5:1, the inverse of general English usage, and occupied the grammatical Actor role in over half of subject clauses through mental, verbal, and behavioural rather than purely material processes. Readers broadly endorsed the ethical implications of this language (82% agreed ecocide should be treated as a crime) more readily than its grammatical premises (49% were comfortable with the pronoun *who* for a river), and endorsement correlated positively with eco-anxiety ($r = .52$, $p < .001$) and varied by gender and faculty. The findings position literary animacy language as a measurable form of ecological resistance and carry implications for environmental communication, language pedagogy, and the Rights of Nature movement.

Keywords: *Animacy, Corpus Linguistics, Ecolinguistics, Environmental Discourse, Generation Z, Rights of Nature, Robert Macfarlane, Transitivity*

Corresponding Author: Natasha Mustafa (MPhil Scholar, Department of English, University of Okara, Okara, Pakistan)
Email: nataliamustafa726@gmail.com

1. Introduction

The twenty-first century's ecological crisis climate change, biodiversity loss, deforestation, and freshwater depletion has typically been framed as a scientific, technological, and economic problem. A growing interdisciplinary field argues that this framing is incomplete: ecological crises are also cultural, ideological, and linguistic (Stibbe, 2021). Ecolinguistics examines how linguistic structures and discourses shape human perceptions of, and behaviour toward, the natural world, and asks whether particular patterns of language use are ecologically harmful or beneficial (Alexander & Stibbe, 2014). Where conventional linguistics analyses syntax, phonology, or semantics for their own sake, ecolinguistics treats language as ideologically loaded, embedding assumptions anthropocentrism, industrial expansion, resource extraction that position humans as separate from, and superior to, the nonhuman world (Halliday, 1990).

Arran Stibbe's Stories-We-Live-By framework (2015, 2021) offers the field's most influential account of how this happens: cultures are sustained by underlying cognitive "stories" encoded and reinforced through recurring linguistic patterns, some of which are destructive and some beneficial to ecological wellbeing. A parallel and complementary account comes from Robin Wall Kimmerer (2013), who argues that English grammar itself enacts a hierarchy of being reserving the personal pronoun who for humans and relegating animals, plants, and geological entities to the inanimate it and that indigenous grammars of animacy offer an alternative in which rivers, trees, and mountains are linguistically kin rather than objects.

Robert Macfarlane's *Is a River Alive?* (2025) is an unusually explicit site to test these claims. The book combines travel writing, environmental reportage, and legal commentary on the global Rights of Nature movement, and its stated authorial premise that a river can be alive in a sense that matters is a linguistic claim as much as an ecological or legal one.

1.1. Statement of the Problem

Existing scholarship on Macfarlane's oeuvre has concentrated on literary and ecocritical readings rather than the systematic linguistic patterning of his prose (Garrard, 2023). This leaves an empirical gap: corpus-assisted ecolinguistics the application of quantitative corpus methods to test ecolinguistic claims about a text (Poole, 2022) has rarely been applied to a single literary work of this kind, and has almost never been paired with a reader-perception study of how such language is actually received. Two related problems follow from this gap. First, claims that literary language constructs rivers as

animate, agentive participants have so far rested on qualitative close-reading rather than reproducible corpus evidence, leaving the scale and consistency of the phenomenon unverified. Second, even where such language can be shown to exist in a text, it remains unknown whether contemporary readers register it as intended, or whether reception is shaped by demographic and psychological factors such as gender, disciplinary background, or eco-anxiety that a text-only analysis cannot capture. Without addressing both problems together, claims about literary language as a vehicle for ecological resistance remain asserted rather than demonstrated.

1.2. Research Questions

The present study addresses this gap in two connected ways. First, it applies corpus linguistic methods frequency analysis, keyword analysis, collocation, n-gram extraction, transitivity coding, and animacy pronoun counts to establish, with quantitative precision, how *Is a River Alive?* constructs the river as an animate grammatical participant. Second, it surveys 150 Generation Z students at a Pakistani university to determine whether, and how, this grammatical strategy is registered by readers, and whether reception is shaped by gender, disciplinary background, or pre-existing eco-anxiety. The study is guided by the following five research questions:

RQ1. How does the text construct rivers as animate through language?

RQ2. What grammatical role does the river occupy in predicated clauses?

RQ3. What lexical and metaphorical strategies build an alternative, beneficial ecological narrative?

RQ4. How does the text's density of rare and technical vocabulary counter the routine linguistic erasure of rivers?

RQ5. What connections exist between these linguistic patterns and the Rights of Nature movement?

1.3. Objectives of the Study

Building on the research questions above, the study pursues five specific objectives:

O1. To quantify, through corpus-linguistic analysis, the frequency and distribution of animacy pronouns (who versus which/that) assigned to river-lemmas in the text.

O2. To determine the grammatical (transitivity) roles occupied by the river as a clausal participant, and the proportion of these roles realised through mental, verbal, and behavioural rather than purely material processes.

O3. To identify the keyword, collocational, and semantic-field patterns through which the text constructs an alternative, ecologically beneficial narrative of rivers.

O4. To evaluate the extent to which the text's lexical density and use of specialised, rights-related, and legal vocabulary counters the routine linguistic erasure of rivers in conventional resource-based discourse.

O5. To measure, through a Linguistic Perception Questionnaire administered to 150 Generation Z university students, the extent to which readers register and endorse the text's animacy language, and to examine whether this reception varies by gender, faculty, and eco-anxiety.

1.4. Significance of the Study

The study's significance is, first, disciplinary and methodological. By pairing a single-text corpus-assisted ecolinguistic analysis with empirical reader-reception data, it directly addresses the gap that Poole (2022, 2025) and Steffensen (2024a) identify as persistent in the field: corpus methods and reader-response methods have rarely been combined within a single ecolinguistic study of literary nature writing. The custom NLTK pipeline developed for this purpose covering frequency, keyword/log-likelihood, animacy-pronoun, collocational, and n-gram analysis is fully reproducible and adaptable to other single-author, single-text corpora, offering a template for future corpus-based ecolinguistic and ecostylistic research (Baker, 2023; Mohamed & Jaafar, 2023).

Beyond its disciplinary contribution, the study carries practical significance for three audiences. For environmental communicators and journalists, the findings offer empirical guidance on which grammatical strategies pronoun choice, verb-process selection, rights-based vocabulary readers actually register and find persuasive, rather than relying on intuition about what “works” rhetorically. For advocates and scholars of the Rights of Nature movement, the study provides linguistic evidence that legal personhood arguments and everyday grammatical choices are mutually reinforcing, complementing recent legal scholarship on the durability of rights-of-nature legislation (Edirisinghe, 2024; O'Donnell & Holley, 2024). For language educators, particularly those teaching Generation Z cohorts who report substantial and rising eco-anxiety (Salguero et al., 2024; Tavolacci & Ladner, 2024), the study's finding that grammatical animacy language correlates with, and may be received differently according to, pre-existing eco-anxiety offers a starting point for designing environmentally themed reading and writing curricula that are responsive to students' psychological as well as linguistic needs.

2. Literature Review

Ecolinguistics traces its foundations to Haugen's (1972) ecology-of-language metaphor and to Halliday's (1990) argument that grammar itself can normalise ecologically destructive assumptions for instance, through classifying finite resources as endlessly “growing” economies. Fill and Mühlhäusler's (2001) reader consolidated the field, and Stibbe's Stories-We-Live-By model (2015, 2021) subsequently offered its most widely applied analytic architecture, classifying discourse patterns as destructive, ambivalent, or beneficial according to whether they support ecological flourishing. Steffensen's (2024a) more recent mapping of the discipline's boundaries situates this beneficial/destructive/ambivalent typology within eleven identifiable ecolinguistic

subfields, of which corpus-assisted ecolinguistics and ecostylistics are the most directly relevant here; and the underlying premise that nature writing carries genuine ideological weight, rather than merely reflecting it, continues to be reaffirmed in recent literary-linguistic scholarship on environmental writing more broadly (Al-Khalidi, 2023).

A parallel methodological tradition, corpus-assisted discourse studies (CADS), uses large electronic corpora and statistical measures frequency, log-likelihood keyness, mutual information collocation to identify patterns invisible to unaided reading (Partington et al., 2013; Baker, 2006). Applied to literary prose, this tradition becomes corpus stylistics (Mahlberg, 2013; Semino & Short, 2004), and its application specifically to ecological themes in literary and non-fiction texts corpus-based ecolinguistics remains comparatively young. Poole (2022) demonstrated the approach's potential but noted that single-text, single-author corpora combined with reader-response data remain rare, and that most corpus ecolinguistic work has focused on newspapers, policy documents, or advertising rather than literary nature writing. More recent surveys of the field confirm that this gap persists: Steffensen's (2024) bibliometric mapping of eleven ecolinguistic subfields finds corpus-assisted and narratological approaches still under-represented relative to discourse-analytic work, while Poole's (2025) overview of the field's consolidation likewise calls for more single-text corpus studies paired with reception data. Related corpus-stylistic work continues to extend the transitivity-based method to individual literary and non-literary texts, including ecostylistic analyses of extracts from contemporary fiction (Mohamed & Jaafar, 2023) and of educational and textbook discourse (Awal, 2025), and Baker's (2023) updated methodological guide to corpus-assisted discourse analysis remains the field's standard reference for the collocational and keyword techniques used here.

Complementing Stibbe, Kimmerer's (2013) grammar of animacy draws on Potawatomi linguistic structure, in which a broader category of beings including rivers, mountains, and berries take the grammatical forms otherwise reserved in English for persons. Kimmerer argues that English's animate/inanimate pronoun split is not a neutral grammatical fact but an ideological one, naturalising the treatment of nonhuman entities as property. This claim is testable at the level of transitivity: Halliday's (1994) systemic functional model classifies clauses by process type (material, mental, verbal, behavioural, existential, relational), and assigning mental or verbal processes to a nonhuman grammatical subject is a quantifiable departure from convention rather than a purely interpretive one. Sealey's (2025) recent survey of how English represents nonhuman animals corroborates Kimmerer's claim from a corpus-linguistic angle, showing that the same animate/inanimate asymmetry Kimmerer identifies at the pronoun level recurs systematically across verb choice and classifier use in mainstream English discourse about animals.

Finally, the Rights of Nature movement beginning with Stone's (1972) "Should Trees Have Standing?" and extended through the legal personhood granted to New Zealand's Whanganui River and Ecuador's constitutional recognition of Pachamama (O'Donnell & Talbot-Jones, 2018) provides the study's real-world stakes: legal personhood arguments are, in part, arguments about whether a nonhuman entity can occupy the grammatical and conceptual position of a subject. This legal strand of the argument has continued to develop since Macfarlane's book was published: Edirisinghe (2024) shows that the central jurisprudential objection to river personhood remains the question of whether a non-sentient entity can meaningfully hold rights at all, while O'Donnell and Holley's (2024) workshop synthesis on implementing rights-of-nature legislation finds that the durability of such rights depends heavily on the surrounding cultural and linguistic narrative, not the statute alone. Macfarlane himself has continued to make this connection between grammar and law explicit in his post-publication commentary (Macfarlane, 2025b), and journalistic engagement with the book has repeatedly foregrounded its central grammatical claim, that the pronoun who performs legal and ethical work that it cannot (Chayes, 2025; Watts, 2025). A further, closely related literature empirical ecocriticism provides the methodological precedent for this study's reader-perception component. Malecki and Schneider-Mayerson (2024) argue that claims about how environmental literature affects readers have too often rested on textual analysis alone, and call for the kind of empirically grounded reception data that the present study's Linguistic Perception Questionnaire supplies. Comparable reader-response designs have been used to test audience reactions to eco-crime fiction (Fetherston et al., 2023), to environmental film (Hiltunen & Rainio, 2023), and to shorter literary extracts using qualitative-empirical and corpus-annotation methods (van de Ven, 2023; Kuijpers et al., 2024; Andersen & Hakemulder, 2024), collectively establishing that readers' grammatical and emotional responses to environmentally themed language can be measured with the same rigour as the language itself. Taken together, these four literatures ecolinguistics, corpus stylistics, the grammar of animacy, and Rights of Nature scholarship converge on a single empirical question that has not yet been tested with corpus precision: does a contemporary literary text actually enact the grammar of animacy at a measurable scale, and if so, how do readers respond to it?

3. Materials and Methods

3.1. Research Design

The study used a sequential mixed-methods design combining corpus linguistics, qualitative coding, and survey research, situated within an interpretivist and critical realist paradigm (Creswell & Creswell, 2018). In Phase One, the primary corpus was analysed programmatically through a custom Python natural language processing (NLP) pipeline built on the Natural Language Toolkit (NLTK, v3.8), which was used to (i) tokenise and

clean the corpus, (ii) generate raw and normalised frequency lists, (iii) compute log-likelihood keyword comparisons against the British National Corpus (BNC), (iv) derive collocational profiles around central nodes using pointwise mutual information (PMI) and frequency-window co-occurrence, and (v) extract bigram, trigram, and higher-order n-gram clusters via `nltk.collocations`. In Phase Two, 200 concordance lines generated with NLTK's `Text.concordance()` function were coded thematically in NVivo 14 using a codebook derived from Stibbe's typology and Kimmerer's animacy categories. In Phase Three, a structured Linguistic Perception Questionnaire (LPQ) was administered to 150 Generation Z students, with responses analysed in IBM SPSS 28. The sequencing allowed survey items to be grounded in linguistic features already empirically confirmed in the text, so that reader responses could be interpreted against a verified linguistic baseline rather than an assumed one.

3.2. Corpus Composition

The corpus comprised the complete prose text of *Is a River Alive?* (Hamish Hamilton, 2025): 106,649 running words across 3,002 prose units, with 14,031 distinct word types (type-token ratio = 0.1316) and 7,240 hapax legomena (words occurring only once), yielding a lexical density of 0.5367 well above the 0.40 typical of conversational English and consistent with dense literary non-fiction. The text was tokenised using NLTK's `word_tokenize()` function with a custom regular-expression pre-processing step that preserved hyphenated compounds (e.g., rights-bearing, kawsak-sacha) as single tokens; the resulting token count was cross-validated by an independent manual word count of a 5% sample (difference < 1%). Standard NLTK stop-word lists (`nltk.corpus.stopwords`) were applied and then manually reviewed to retain theoretically relevant items, notably *who*, which is excluded by default from generic stop-word lists but was retained given its centrality to the animacy analysis. Keyword analysis used the written component of the British National Corpus (BNC; approximately 90 million words) as the primary reference, cross-checked against a contemporary magazine sub-corpus to rule out keyness effects driven by the BNC's earlier compilation date rather than genuine textual distinctiveness.

3.3. Analytical Framework and Procedures

Three frameworks jointly informed coding: Stibbe's (2015, 2021) beneficial/ambivalent/destructive story typology, Halliday's (1994) six-way transitivity classification, and Kimmerer's (2013) grammar of animacy, operationalised as the ratio of *who* to *which/that* following river-lemmas. All clauses with *river* or *rivers* in subject position were manually coded for process type; a random 10% sample (84 clauses) was independently re-coded, yielding 92% inter-coder agreement, with residual disagreement concentrated at the relational/behavioural boundary. The LPQ comprised seven sections: consent and demographics; three verbatim excerpts from the source text; a ten-item five-

point Likert scale (L1–L8 forming a Beneficial Story scale, $\alpha = .884$; L9 measuring ecocide endorsement; L10 measuring baseline eco-anxiety); a six-pair story-endorsement block; and an open-response item. The instrument was piloted with 10 participants not included in the main sample (mean completion time 15–20 minutes). Table 1 summarises how each research question maps onto its corresponding data source and analytic method.

Table 1. *Research Questions Mapped to Data Sources and Analytic Methods*

Research Question	Data Source / Method
RQ1. Grammar of animacy vs. inanimate convention	Relative-pronoun distribution; Python/NLTK concordance (KWIC)
RQ2. River as grammatical actor	Transitivity coding of river-subject clauses
RQ3. Beneficial stories via lexical/metaphorical clusters	Collocation (PMI) and n-grams via NLTK; NVivo coding
RQ4. Thick description against linguistic erasure	Lexical density, hapax count; Python/NLTK frequency and keyword analysis
RQ5. Links to the Rights of Nature movement	Keyword and n-gram analysis (Python/NLTK); LPQ survey (SPSS)

Note. Adapted from the study's full methodological protocol.

3.4. Ethics

The study received approval from the Departmental Research Ethics Committee, Department of English, University of Okara, in accordance with Higher Education Commission of Pakistan guidelines. Participation required affirmative electronic consent; identifiers were pseudonymised, and records will be retained for five years post-approval per institutional policy. Given the text's focus on ecological loss, a care-based ethical approach (Noddings, 2013) informed survey administration, including signposting to university counselling services for the eco-anxiety item.

4. Results

4.1. Corpus Profile

Table 2 summarises the corpus's baseline statistics. The high hapax proportion (51.6% of all types) and lexical density (0.5367) indicate a text drawing heavily on rare,

technical, and loanword vocabulary consistent with a deliberate lexical strategy for rendering the natural world newly visible rather than an incidental feature of length.

Table 2. *Baseline Corpus Statistics for Is a River Alive? (Macfarlane, 2025)*

Statistic	Value
Total running words (tokens)	106,649
Unique word types	14,031
Type-token ratio	0.1316
Hapax legomena	7,240 (51.6% of types)
Lexical density	0.5367

4.2. Keyword and Frequency Analysis

River and rivers together account for nearly one word in every hundred (combined frequency 1,034), and river ranks 12th among all words in the text ahead of all but the most basic function words. Table 3 presents the leading keywords by log-likelihood (Dunning, 1993), computed with a custom Python script that tokenised both the target and BNC reference corpora using NLTK and calculated log-likelihood ratios from their respective frequency distributions (critical value 15.13 at $p < .0001$; all listed items far exceed this threshold). Notably, rights rank fifth by frequency among the keywords listed in Table 3, and its keyness ratio (22.50) while comparatively modest next to core lexical items such as river, rivers, and forest remains far above the log-likelihood significance threshold, an unusual prominence for an explicitly legal-political term within a work of nature writing.

Table 3. *Leading Keywords by Log-Likelihood Against the BNC Written Reference*

Keyword	Frequency	Keyness Ratio	Log-Likelihood
river	742	173.94	3,412.87
rivers	292	182.53	1,342.75
water	392	15.64	1,644.70
forest	253	67.78	1,143.38
rights	192	22.50	826.71
living	108	12.66	443.71
alive	69	43.13	305.04

Note. All values computed via a custom Python (NLTK) log-likelihood script against the BNC written-component reference corpus.

4.3. Collocational Profiles

Collocates were extracted with NLTK's BigramCollocationFinder and scored by pointwise mutual information (PMI), applying a minimum joint-frequency filter of 3 to exclude chance co-occurrences. The node living returned the most theoretically productive profile (Table 4): its strongest collocate, conscious (PMI = 7.04), places living within the register of legal-personhood debate rather than simple metaphor, while *sacha* and *kawsak Kichwa* terms for "living forest" and "living/aliveness" indicate that the text draws on indigenous relational cosmologies as a conceptual source rather than relying solely on Western anthropomorphic convention. The node rights showed an equally coherent profile organised around a single legal-political discourse: its top collocate, nature (PMI = 5.28), instantiates the canonical phrase rights of nature, while movement, recognition, and constitutional extend this into the register of an active, globalising legal-political campaign rather than an abstract philosophical claim.

Table 4. *Top Collocates of LIVING and RIGHTS Ranked by Pointwise Mutual Information*

Node	Collocate	Joint Freq.	PMI
living	conscious	12	7.04
living	sacha	8	6.82
living	rights-bearing	4	6.04
living	entity	11	5.56
rights	nature	81	5.28
rights	movement	17	4.71
rights	recognition	8	4.34
rights	constitutional	5	4.03

Note. Computed using nltk.collocations.BigramCollocationFinder with a minimum joint-frequency filter of 3.

4.4. N-gram and Cluster Analysis

Multi-word clusters were extracted across bigram-to-pentagram spans using nltk.util.ngrams and ranked by frequency (Table 5). The clearest pattern is the recurrence of the book's own thesis as a formulaic unit: the 4-gram *a river is alive* occurs 8 times and

the interrogative is a river alive 9 times, indicating that Macfarlane treats his title's claim not as a one-off provocation but as a repeated argumentative refrain. The formula rights of nature (70 occurrences, 0.66 per 1,000 words) confirms the density already visible in the keyword and collocation results, while the 5-gram river is a living being merging ontological and legal claims into a single clause recurs at key argumentative junctures, typically alongside reports of legal developments or indigenous testimony.

Table 5. Key N-gram Clusters With Frequency (Extracted via NLTK)

Cluster	N-gram Type	Frequency
the river	Bigram	332
rights of	Bigram	103
rights of nature	Trigram	70
the river is	Trigram	33
is a river alive	4-gram	9
a river is alive	4-gram	8
river is a living being	5-gram	3

1.1. 4.5 Transitivity: The River as Grammatical Actor

Of 835 clauses with river or rivers as grammatical subject, 412 were unambiguously codable by Halliday's six process types (Table 6). Material processed in which the river remains grammatical Actor rather than acted-upon (e.g., carves, shapes, carries) accounted for 37.9%. Critically, mental, verbal, and behavioural processes, conventionally reserved for conscious participants, together accounted for 54.1% of clauses: the river senses and knows (22.8%), says and sings (17.2%), and lives and breathes (14.1%). This distribution indicates a systematic, text-wide grammatical strategy rather than an occasional stylistic flourish.

Table 6. Halliday Process Types for River-as-Subject Clauses (n = 412)

Process Type	Count	%
Material (does, causes, creates)	156	37.9
Mental (senses, thinks, knows)	94	22.8
Verbal (says, speaks, sings)	71	17.2
Behavioural (lives, breathes, grows)	58	14.1

Process Type	Count	%
Existential (exists, is there)	21	5.1
Relational (has, is identified as)	12	2.9

1.2. 4.6 Animacy Pronoun Distribution

The relative pronoun following river-lemmas provides the clearest single test of animacy grammar. As Table 7 shows, who was used 34 times (77.3%) against 4 uses of which (9.1%), a ratio of 8.5:1 the inverse of general English usage, in which which is typically dominant by an order of magnitude of roughly 10:1 in reference-corpus baselines. The four inanimate uses occurred almost exclusively when Macfarlane quoted engineering or governmental prose before reverting to animate framing in his own voice.

Table 7. Relative Pronoun Distribution After River-Lemmas

Pronoun Pattern	Frequency	% of Total
river-lemma + who	34	77.3
river-lemma + that	6	13.6
river-lemma + which	4	9.1

4.7. Semantic Fields and Beneficial Framing

Nine semantic fields were coded and cross-checked against Python/NLTK-generated frequency word lists (Table 8). The three densest fields voice and communication (7.08 per 1,000 words), animacy and life (6.14), and rights, law, and legal (5.98) are all classified as beneficial in Stibbe's typology and together account for roughly one word in fifty across the entire corpus. Destructive framings (death and killing; ecological crisis) were comparatively sparse (1.90 and 2.59 per 1,000, respectively), and were used almost exclusively to name specific threats rather than to characterise rivers generally. NVivo coding of 200 concordance lines corroborated this pattern: beneficial codes accounted for 73.5% of coded lines against 12.5% for destructive codes, with river-as-voice the single most frequent code (22.0%).

Table 8. Semantic Field Densities and Stibbe Story-Type Classification

Semantic Field	Per 1,000 Words	Story Type
Voice and communication	7.08	Beneficial

Semantic Field	Per 1,000 Words	Story Type
Animacy and life	6.14	Beneficial
Rights, law, and legal	5.98	Beneficial
Movement and flow	3.50	Beneficial
Ecological crisis	2.59	Ambivalent
Death and killing	1.90	Destructive

4.8. Reader Perception Survey

The survey sample (N = 150; 75 women, 75 men; Mage = 21.4, SD = 1.8; 54.7% Humanities, 45.3% Sciences) showed an overall mean of 3.74 (SD = 0.61) on the eight-item Beneficial Story scale ($\alpha = .884$), indicating moderate-to-positive endorsement. Endorsement was uneven across items (Table 9): relational and ethical claims drew the strongest agreement indigenous frameworks deserving equal weight with science (M = 4.19, 77.3% agree) and imagined kinship with a river (M = 4.08, 74.0%) while grammatical and cognitive claims drew the weakest, notably comfort with the pronoun who for a river (M = 3.38, 49.3%). Strikingly, 82.0% agreed that destroying a river should be treated as a crime against persons (L9), a considerably higher endorsement rate than for any grammatical item, suggesting that readers accept the ethical implications of river personhood more readily than its linguistic premise.

Table 9. Selected Likert Item Means and Agreement Rates (N = 150, 1–5 scale)

Item	Statement Summary	M	SD	% Agree	Scale
L1	River aliveness is meaningful, not just metaphor	3.92	0.91	68.0	BEN
L2	Comfortable thinking of the river as acting or deciding	3.74	0.98	61.3	BEN

Item	Statement Summary	M	SD	% Agree	Scale
L3	Natural to describe the river as having a voice	3.61	1.02	58.7	BEN
L4	Can imagine kinship between a community and a river	4.08	0.87	74.0	BEN
L5	River should hold legal rights and court standing	3.55	1.09	54.7	BEN
L6	Reasonable to describe the river as aware of people	3.44	1.11	52.0	BEN
L7	The pronoun who feels appropriate for a river	3.38	1.15	49.3	BEN
L8	Indigenous frameworks deserve equal weight with science	4.19	0.83	77.3	BEN
L9	Destroying a river is a crime like a crime against persons	4.31	0.79	82.0	ECO
L10	Eco-anxiety: often distressed about rivers and forests	3.88	0.97	66.7	ANX

Note. BEN = Beneficial Story scale item (L1–L8, $\alpha = .884$); ECO = ecocide-endorsement item; ANX = baseline eco-anxiety item. Overall composite scale mean (L1–L8) = 3.74, SD = 0.61.

Beneficial Story scores differed significantly by gender (women $M = 3.91$ vs. men $M = 3.57$; $t(148) = 3.42$, $p = .001$, $d = 0.56$) and by faculty (Humanities $M = 3.89$ vs. Sciences $M = 3.55$; $t(148) = 3.14$, $p = .002$, $d = 0.51$), both moderate effects. Eco-anxiety (L10) correlated positively with Beneficial Story endorsement, $r(148) = .52$, $p < .001$. In the open-response item, 34.5% of participants who would use an animate pronoun for a familiar river cited kinship or ecological reasoning (e.g., a river that feeds a community and "remembers all the monsoons"), and 15.8% explicitly noted that Punjabi, Urdu, and other regional languages already grammatically gender rivers reframing the animacy question as a revival of an existing linguistic resource rather than a foreign innovation.

5. Discussion

The corpus results converge on a single conclusion: Macfarlane's animacy language is not an occasional metaphor but a systematic, quantifiable grammatical architecture. The 8.5:1 who-to-which ratio, sustained across registers rather than confined to lyrical passages, empirically operationalises Kimmerer's (2013) theoretical grammar of animacy, while the 54.1% share of mental, verbal, and behavioural processes in river-subject clauses gives Halliday's transitivity model direct evidential grounding in a contemporary literary text. Read alongside the semantic field results where beneficial vocabulary (voice, animacy, rights) dominates over destructive vocabulary by a wide margin the text meets Stibbe's (2015, 2021) definition of a beneficial story: one that counters the routine linguistic erasure of nonhuman entities.

The reader-perception data complicate this picture productively rather than simply confirming it. Readers accepted the moral and legal implications of river personhood ecocide-as-crime (82.0%), indigenous epistemic parity (77.3%) considerably more readily than the grammatical premise itself (49.3% for the pronoun who). This asymmetry suggests that ethical intuitions about nonhuman standing may be, in this population, already ahead of the grammatical habits required to express them, a pattern with direct relevance to environmental communication: appeals grounded in justice and harm may travel faster than appeals requiring readers to adopt unfamiliar grammatical choices. The moderate correlation between eco-anxiety and beneficial-story endorsement ($r = .52$) extends Hickman et al.'s (2021) findings on eco-anxiety into the domain of literary reception, suggesting that emotional investment in ecological loss primes receptivity to animacy language rather than simply coexisting with it. This finding sits within a wider and rapidly growing body of evidence on Generation Z environmental distress: Salguero et al. (2024) similarly find that the large majority of university-age Generation Z respondents in Australia report significant climate concern accompanied by measurable

anxiety, and Tavolacci and Ladner (2024) report comparable eco-anxiety burdens among European university students regardless of disciplinary background. The present study's $r = .52$ correlation is consistent with this literature's general pattern of moderate-to-strong links between ecological distress and pro-environmental attitude endorsement, while extending it specifically into the domain of grammatical, rather than purely attitudinal, response.

The gender and faculty differences women and Humanities students scoring higher on the Beneficial Story scale are consistent with prior findings on gendered environmental attitudes (Zelezny et al., 2000) and with the plausible role of literary-critical training in validating relational ways of knowing. Perhaps most striking for the Pakistani context is the open-response finding that many participants did not experience animacy grammar as a foreign imposition but as a return to resources already present in Punjabi and Urdu, where rivers such as the Ravi and the Sindhu carry grammatical gender. This suggests that multilingual pedagogy explicitly connecting English-language ecolinguistic argument to animacy resources already present in students' first languages may be a more effective route to ecological literacy than treating animacy grammar as an English-only innovation.

For the Rights of Nature movement, the results indicate that literary language performs cultural work that legal instruments alone cannot: the phrase rights of nature recur 70 times across the corpus, embedded in narrative and experiential context rather than abstracted into legal register, and this narrative embedding may be precisely what sustains public moral commitment even when legal rulings (such as reversals affecting the Ganges and Yamuna) prove precarious.

6. Conclusion

This study set out to test, with corpus precision, whether a contemporary literary text can be shown to construct a river as a grammatically animate, agentive entity, and whether such construction registers with readers. The evidence is unambiguous on the first point: Macfarlane's *Is a River Alive?* inverts conventional English animacy grammar at a measurable, text-wide scale, assigning the river a conscious grammatical role in the majority of its predicated clauses. On the second point, the evidence is more nuanced: readers in this case, 150 Generation Z students in Pakistan accept the ethical and legal implications of this language considerably more readily than its grammatical form, and their receptivity is shaped by gender, disciplinary background, eco-anxiety, and the animacy resources already present in their first languages.

The study's scope is deliberately narrow a single text, a single author, a single institutional sample and this concentration is also its principal limitation: findings cannot be generalised to nature writing as a genre, to Macfarlane's wider body of work, or beyond a single Pakistani university population. The transitivity coding relied on a single coder (92% inter-coder agreement on a 10% sample), and the BNC reference corpus, though

standard, dates from the 1980s–1990s. Future research should extend the corpus comparatively across Macfarlane's oeuvre and a wider contemporary nature-writing reference corpus, replicate the perception survey cross-nationally in the UK, Aotearoa New Zealand, and Ecuador, and pursue longitudinal pre/post-reading designs to test whether brief exposure to animacy language produces durable shifts in ecological attitude. Taken together, the findings suggest that literary language can function as a measurable site of ecological resistance, and that corpus methods are well suited to documenting that resistance with the same rigour ordinarily reserved for demonstrating its absence.

Reference

- Alexander, R., & Stibbe, A. (2014). From the analysis of ecological discourse to the ecological analysis of discourse. *Language Sciences*, 41, 104–110.
- Al-Khalidi, I. (2023). Environmental literature and the importance of nature in writing. *Journal of Literature and Linguistics Studies*, 1(1), 1–8.
- Andersen, T. R., & Hakemulder, F. (2024). “The poem has stayed with me”: Continued processing and impact from Shared Reading experiences of people living with cancer. *Poetics*, 102, 101847.
- Awal, A. (2025). Ecolinguistics in Bangladeshi secondary-level English textbooks: A corpus analysis. *Colombian Applied Linguistics Journal*, 27(1), 18–32.
- Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.
- Baker, P. (2023). *Using corpora in discourse analysis* (2nd ed.). Bloomsbury.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: Analyzing text with the Natural Language Toolkit*. O'Reilly Media.
- Chayes, S. (2025, July 2). Review: Sarah Chayes on Robert Macfarlane, “Is a River Alive?” Book Post.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE.
- Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1), 61–74.
- Edirisinghe, S. (2024). Nature as a sentient being: Can rivers be legal persons? Review of *European, Comparative & International Environmental Law*, 33(2), 224–235.
- Fetherston, R., Potter, E., Miller, K., & Bowles, D. (2023). Seeking greener pages: An analysis of reader response to Australian eco-crime fiction. *Australian Humanities Review*, 71.
- Fill, A., & Mühlhäusler, P. (Eds.). (2001). *The ecolinguistics reader: Language, ecology and environment*. Continuum.
- Halliday, M. A. K. (1990). New ways of meaning: The challenge to applied linguistics. *Journal of Applied Linguistics*, 6, 7–36.

- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). Edward Arnold.
- Haugen, E. (1972). *The ecology of language*. Stanford University Press.
- Hickman, C., Marks, E., Pihkala, P., Clayton, S., Lewandowski, R. E., Mayall, E. E., Wray, B., Mellor, C., & van Susteren, L. (2021). Climate anxiety in children and young people and their beliefs about government responses to climate change: A global survey. *The Lancet Planetary Health*, 5(12), e863–e873.
- Hiltunen, K., & Rainio, M. (2023). Reflections of hope and anxiety in audience responses to three environmental films. *Participations*, 19(2), 45–75.
- Kimmerer, R. W. (2013). *Braiding sweetgrass: Indigenous wisdom, scientific knowledge, and the teachings of plants*. Milkweed Editions.
- Kuijpers, M. M., Luseti, M., Lendvai, P., & Rebora, S. (2024). Validation of the Story World Absorption Scale through annotation of online book reviews. *Journal of Cultural Analytics*, 9(1).
- Macfarlane, R. (2025). *Is a river alive?* Hamish Hamilton.
- Macfarlane, R. (2025b, June 7). What happens when a river is given legal rights. *The Walrus*.
- Mahlberg, M. (2013). *Corpus stylistics and Dickens's fiction*. Routledge.
- Malecki, W., & Schneider-Mayerson, M. (2024). Empirical ecocriticism: Evaluating the influence of environmental literature. In C. Brudin Borg, R. Wingard, & J. Bruhn (Eds.), *Contemporary ecocritical methods*. Rowman & Littlefield.
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Mohamed, Z. A., & Jaafar, E. A. (2023). An ecostylistic analysis of selected extracts from Michael Punke's novel *The Revenant*. *Cogent Arts & Humanities*, 10(1), 2244209.
- Noddings, N. (2013). *Caring: A relational approach to ethics and moral education* (2nd ed.). University of California Press.
- O'Donnell, E., & Holley, C. (2024). Rivers have rights. So what happens now? Proceedings of an international workshop on the implementation of rights of nature. UNSW and University of Melbourne.
- O'Donnell, E., & Talbot-Jones, J. (2018). Creating legal rights for rivers: Lessons from Australia, New Zealand, and India. *Ecology and Society*, 23(1), 7.
- Partington, A., Duguid, A., & Taylor, C. (2013). Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS). John Benjamins.
- Poole, R. (2022). *Corpus-assisted ecolinguistics*. Bloomsbury Academic.
- Poole, R. (2024). A corpus-assisted ecolinguistic analysis of hurricanes and wildfires and the potential for corpus-assisted eco-pedagogy in ELT classrooms. In M.

- Bortoluzzi & E. Zurru (Eds.), *Environment as lifescape: Ecoliteracy and communication in action*. Bloomsbury.
- Poole, R. (2025). Ecolinguistics: Consolidating a research paradigm. *Russian Journal of Linguistics*, 29(1), 6–16.
- Salguero, R. B., Bogueva, D., & Marinova, D. (2024). Australia’s university Generation Z and its concerns about climate change. *Sustainable Earth Reviews*, 7(1), Article 8.
- Sealey, A. (2025). *How we talk about animals, and why it matters*. Bloomsbury.
- Semino, E., & Short, M. (2004). *Corpus stylistics: Speech, writing and thought presentation in a corpus of English writing*. Routledge.
- Steffensen, S. V. (2024a). On the demarcation of ecolinguistics. *Journal of World Languages*, 10(3), 499–527.
- Steffensen, S. V. (2024b). Surveying ecolinguistics. *Journal of World Languages*, 11(1), 1–49.
- Stibbe, A. (2015). *Ecolinguistics: Language, ecology and the stories we live by*. Routledge.
- Stibbe, A. (2021). *Ecolinguistics: Language, ecology and the stories we live by* (2nd ed.). Routledge.
- Stone, C. D. (1972). Should trees have standing? Toward legal rights for natural objects. *Southern California Law Review*, 45, 450–501.
- Tavolacci, M., & Ladner, J. (2024). Eco-anxiety: An additional burden for university students? *European Journal of Public Health*, 34(Suppl. 3), ckae144.1352.
- van de Ven, I. (2023). “Gonna get you, baby!” A qualitative-empirical study of attentional modulation in reading a short story. *Language and Literature*, 32(4), 458–478.
- Watts, J. (2025, May 17). Robert Macfarlane: ‘Sometimes I felt as if the river was writing me’. *The Guardian*.
- Zelezny, L. C., Chua, P.-P., & Aldrich, C. (2000). Elaborating on gender differences in environmentalism. *Journal of Social Issues*, 56(3), 443–457.

Appendix: Python/NLTK Analysis Pipeline

The full Phase One pipeline described in section 3.1 is reproduced below for reproducibility. The script was run against the complete plain-text corpus of *Is a River Alive?* (106,649 word-tokens after punctuation filtering, matching the `n_tokens` value in Table 2) using Python 3.12 and NLTK 3.8, and produces the frequency, log-likelihood keyword, animacy-pronoun, collocation (both adjacency-based and window-based), and n-gram statistics reported in section 4. Section 3a additionally requires a reference-corpus frequency list (word, freq columns; e.g. derived from the BNC written component) supplied at `REFERENCE_FREQ_PATH`.

```
import re
import json
```

```
import math
import csv
import math
import csv
from collections import Counter

import nltk
from nltk import word_tokenize, FreqDist
from nltk.corpus import stopwords
from nltk.util import ngrams
from nltk.collocations import BigramCollocationFinder, BigramAssocMeasures
from nltk.text import Text

CORPUS_PATH = "Is_river_a_alive.txt"

# 1. Load and clean corpus
with open(CORPUS_PATH, "r", encoding="utf-8", errors="ignore") as f:
    raw_text = f.read()

raw_text = raw_text.replace("\r\n", " ").replace("\n", " ")
raw_text = re.sub(r"\s+", " ", raw_text).strip()

# 2. Tokenise
tokens_all = word_tokenize(raw_text)
tokens_lower = [t.lower() for t in tokens_all]
words_only = [t for t in tokens_lower if re.match(r"^[a-zA-ZÀ-ÿ'-]+$", t)]

stop_set = set(stopwords.words("english"))
stop_set.discard("who") # retained: central to the animacy analysis
content_words = [w for w in words_only if w not in stop_set]

# 3. Frequency statistics
fdist_all = FreqDist(words_only)
fdist_content = FreqDist(content_words)

n_tokens = len(words_only)
n_types = len(fdist_all)
ttr = n_types / n_tokens
```

```
hapax = fdist_all.hapaxes()
n_hapax = len(hapax)
lexical_density = len(content_words) / n_tokens

print("Total tokens:", n_tokens)
print("Unique types:", n_types)
print("Type-token ratio:", round(ttr, 4))
print("Hapax legomena:", n_hapax, f"({n_hapax/n_types*100:.1f}% of types)")
print("Lexical density:", round(lexical_density, 4))

print("river:", fdist_all["river"], " rivers:", fdist_all["rivers"])
ranked = fdist_all.most_common()
rank_lookup = {w: i + 1 for i, (w, c) in enumerate(ranked)}
print("rank of river:", rank_lookup.get("river"))

# 3a. Keyword / log-likelihood analysis vs reference corpus (Table 3)
REFERENCE_FREQ_PATH = "bnc_written_freqlist.csv" # columns: word,freq
ref_freqs = {}
with open(REFERENCE_FREQ_PATH, "r", encoding="utf-8") as f:
    for row in csv.DictReader(f):
        ref_freqs[row["word"].lower()] = int(row["freq"])
ref_total = sum(ref_freqs.values())

def log_likelihood(a, b, c, d):
    a, b = max(a, 0.0000001), max(b, 0.0000001)
    e1 = c * (a + b) / (c + d)
    e2 = d * (a + b) / (c + d)
    return 2 * (a * math.log(a / e1) + b * math.log(b / e2))

def keyness_ratio(a, c, b, d):
    rate_ref = b / d if b > 0 else (0.5 / d)
    return (a / c) / rate_ref

CRITICAL_LL = 15.13 # p < .0001, df = 1
keyword_rows = []
for word, freq in fdist_all.most_common(300):
    b = ref_freqs.get(word, 0)
```

```
ll = log_likelihood(freq, b, n_tokens - freq, ref_total - b)
kr = keyness_ratio(freq, n_tokens, b, ref_total)
if ll >= CRITICAL_LL:
    keyword_rows.append((word, freq, round(kr, 2), round(ll, 2)))

keyword_rows.sort(key=lambda r: -r[3])
print("Top keywords by log-likelihood (Table 3):")
for word, freq, kr, ll in keyword_rows[:20]:
    print(f" {word:15s} freq={freq:5d} keyness_ratio={kr:10.2f} LL={ll:10.2f}")
# 4. Animacy pronoun distribution: river/rivers + who/which/that
raw_tokens = word_tokenize(raw_text)
raw_tokens_lower = [t.lower() for t in raw_tokens]

pronoun_counts = Counter()
WINDOW = 4
for i, tok in enumerate(raw_tokens_lower):
    if tok in ("river", "rivers"):
        window = raw_tokens_lower[i + 1: i + 1 + WINDOW]
        for w in window:
            if w in ("who", "which", "that"):
                pronoun_counts[w] += 1
            break

total_pron = sum(pronoun_counts.values())
for pr in ("who", "which", "that"):
    c = pronoun_counts[pr]
    print(pr, c, f"({c/total_pron*100:.1f}%)") if total_pron else ""
print("who:which      ratio      ->",      round(pronoun_counts["who"] /
pronoun_counts["which"], 2))

# 5. Collocational profiles (PMI) for key nodes
finder = BigramCollocationFinder.from_words(content_words)
finder.apply_freq_filter(3)
pmi_measure = BigramAssocMeasures.pmi

def top_collocates_for_node(node, finder, n=10):
    scored = finder.score_ngrams(pmi_measure)
    hits = [(w1, w2, s) for (w1, w2), s in scored if w1 == node or w2 == node]
```

```

hits.sort(key=lambda x: -x[2])
return hits[:n]

for node in ["living", "rights", "river", "alive"]:
    print(f"Top PMI collocates for '{node}':")
    for w1, w2, score in top_collocates_for_node(node, finder):
        other = w2 if w1 == node else w1
        print(f" {node} <-> {other} PMI={score:.2f}")

# 5b. Window-based co-occurrence PMI (matches the window-based method
#     described in Section 3.1; complements the adjacency-based PMI above,
#     since removing stopwords before pairing in 5. can create artificial
#     adjacencies that were not adjacent in the source text)
COOC_WINDOW = 5

def window_cooccurrence_pmi(tokens, node, window=COOC_WINDOW,
min_joint_freq=3, n=10):
    total = len(tokens)
    node_freq = tokens.count(node)
    joint = Counter()
    for i, tok in enumerate(tokens):
        if tok == node:
            lo, hi = max(0, i - window), min(total, i + window + 1)
            for j in range(lo, hi):
                if j != i:
                    joint[tokens[j]] += 1
    word_freq = FreqDist(tokens)
    results = []
    for other, jf in joint.items():
        if jf < min_joint_freq or other == node:
            continue
        p_node, p_other, p_joint = node_freq / total, word_freq[other] / total, jf / total
        pmi = math.log2(p_joint / (p_node * p_other))
        results.append((other, jf, round(pmi, 2)))
    results.sort(key=lambda x: -x[2])
    return results[:n]

```

```
for node in ["living", "rights", "river", "alive"]:
    print(f"Top window co-occurrence PMI collocates for '{node}':")
    for other, jf, pmi in window_cooccurrence_pmi(content_words, node):
        print(f" {node} <-> {other} joint_freq={jf} PMI={pmi}")

# 6. N-gram cluster extraction
def top_ngrams(tokens, n, top=15, min_freq=3):
    grams = list(ngrams(tokens, n))
    fd = FreqDist(grams)
    return [(g, c) for g, c in fd.most_common(top) if c >= min_freq]

print("Top bigrams:")
for g, c in top_ngrams(words_only, 2, top=15):
    print(" ".join(g), "-", c)

print("Top trigrams:")
for g, c in top_ngrams(words_only, 3, top=15):
    print(" ".join(g), "-", c)

key_phrases = [
    ("rights", "of", "nature"),
    ("is", "a", "river", "alive"),
    ("a", "river", "is", "alive"),
]
joined = " ".join(words_only)
for phrase in key_phrases:
    p_str = " " + " ".join(phrase) + " "
    print(phrase, "->", joined.count(p_str))

# 7. Concordance (KWIC) sample
text_obj = Text(raw_tokens)
text_obj.concordance("alive", width=80, lines=10)

# 8. Save results
results = {
    "n_tokens": n_tokens, "n_types": n_types, "ttr": ttr,
    "n_hapax": n_hapax, "lexical_density": lexical_density,
    "river_freq": fdist_all["river"], "rivers_freq": fdist_all["rivers"],
    "pronoun_counts": dict(pronoun_counts),
```

```
        "top_content_words": fdist_content.most_common(30),  
    }  
    with open("results.json", "w") as f:  
        json.dump(results, f, indent=2)
```