Çankaya University Journal of Humanities and Social Sciences
2026, pp. 411-434
https://doi.org/10.47777/cankujhss.1902350

Linguistics | research article

Ghosts of L1 Evidentiality: Constructional Transfer in L2 English Hedging

Fatih Ünal Bozdağ 0000-0002-9959-4704
Osmaniye Korkut Ata University

ABSTRACT
This study tests the ghost construction hypothesis (that obligatory L1 evidential categories persist as functional patterns in L2 hedging) by analyzing 1,011 English argumentative essays (752,967 words) by Turkish, Bulgarian, and Spanish learners (ICLE) and native English speakers (LOCNESS). Hedging devices were detected through a hybrid rule-based and LLM-assisted pipeline mapped onto Aikhenvald's (2004) typology, and cross-group differences were assessed at the categorical, constructional, and propositional layers within a hierarchical Bayesian framework with L1 and text as random effects. At the categorical layer, the four groups produced similar evidential distributions, and eight directional predictions derived from the L1 typology were either null or directionally reversed. At the constructional layer, however, reportative subtype distributions diverged sharply. Turkish and Bulgarian writers allocated approximately 4% of reportative hedges to named attribution, compared to approximately 22% for Spanish and native writers, while preferring different source-nonspecific alternatives that paralleled their respective L1 evidential structures. At the propositional layer, a consistent L2–native distancing gap survived covariate adjustment for lexical diversity, syntactic complexity, and L2 exposure, and proved robust under an alternative distributional model. The results suggest that L1 evidentiality influences L2 hedging at the constructional rather than the categorical level.

ARTICLE HISTORY
Received 04 Mar 2026
Accepted 14 Aug 2026

KEYWORDS
L2 academic writing, reportative attribution, evidentiality, hedging, cross-linguistic influence, learner corpus research, constructional transfer

Introduction

Hedging is a defining feature of academic discourse. Through devices such as epistemic modals, lexical verbs, adverbs, and attribution constructions, writers modulate commitment to propositions, encoding epistemic stance and managing credibility. Hedges constitute over half of all interactional metadiscourse features in research articles across disciplines (Hyland, 2005). L2 writers deploy hedging devices in patterns that diverge from native-speaker norms, relying on a restricted range of devices, exhibiting stronger epistemic commitment, and favoring simpler constructions (Hinkel, 2005; Hyland & Milton, 1997). These divergences are L1-specific, yet most L2 hedging research treats hedging as a monolithic skill, comparing overall frequencies without examining how the typological structure of the L1 predicts which patterns transfer.

If a writer's L1 obligatorily grammaticalizes the encoding of information source, the cognitive routines associated with those distinctions may transfer to L2 hedging, not as morphological transfer but as functional transfer of the underlying epistemic distinctions that the L1 system makes obligatory. Mushin (2001) showed that speakers of languages with grammaticalized evidentiality maintained explicit source coding in narrative retelling at substantially higher rates than English speakers. We term this the "ghost construction" hypothesis. Obligatory L1 evidential categories without L2 equivalents may persist as functional patterns, shaping which

Fatih Ünal Bozdağ, Asst. Prof. Dr., English Translation and Interpretation, Osmaniye Korkut Ata University, Turkiye | fatihbozdag@osmaniye.edu.tr; ORCID# 0000-0002-9959-4704; doi.org/10.47777/cankujhss.1902350

constructions writers select even when the target language imposes no such requirement.

The study compares L2 English hedging across three L1 groups selected for their position on the evidentiality typology: Turkish (grammaticalized suffixal evidentiality), Bulgarian (mood-based evidential morphology), and Spanish (no grammaticalized evidentiality), with a native English baseline drawn from LOCNESS. Previous studies have compared hedging across L1 groups without deriving directional predictions from the structure of L1 evidential systems. They documented L1-related hedging differences (Demir, 2018; Hyland & Milton, 1997; Oh, 2007; Šinkūnienė, 2008) without connecting them to the evidential architecture of the source languages, and no study has examined whether transfer effects invisible at the category level emerge in how writers constructionally realize reportative hedging.

This study addresses four research questions about cross-linguistic influence in L2 academic hedging.

RQ1. Do L1 groups differ in the distribution of evidential hedge categories?
RQ2. Are the eight directional predictions derived from the L1 evidentiality typology supported in L2 hedging?
RQ3. Does L1 evidentiality shape the constructional realization of reportative hedges in L2 English?
RQ4. Does propositional distancing in L2 hedging reflect L1-specific evidentiality, or is it a shared L2 phenomenon?

The study contributes to learner corpus research and SLA by testing how a cross-linguistic evidentiality typology constrains the constructional choices learners make in academic writing.

Literature Review

Evidentiality as a Typological Variable

In many of the world's languages, the source of evidence underlying a statement must be grammatically specified, whether the speaker witnessed, inferred, or was told about an event (Aikhenvald & Dixon, 2003). This obligatory category, evidentiality, is distinct from epistemic modality, which marks degree of certainty rather than type of evidence, and the two do not necessarily co-occur (Aikhenvald & Dixon, 2003; Faller, 2006). Aikhenvald and Dixon (2003) distinguished Type I systems, which mark the existence of indirect evidence without specifying its kind (termed "indirectivity" by Johanson, 2000), from Type II systems, which specify the nature of the evidence. English and Spanish, by contrast, rely on optional lexical strategies (Aikhenvald & Dixon, 2003).

On the other hand, Turkish exemplifies obligatory evidential marking. Speakers must choose between the direct evidential suffix *-DI* (witnessed experience) and the indirect evidential *-mış* (inference or hearsay) for every past-tense utterance. Şener (2011) showed that -mış covers two semantically distinct subtypes, reportative and inferential, with different temporal and pragmatic properties. The cognitive entrenchment of this distinction begins early, though mapping evidential concepts onto the morphemes remains a protracted developmental challenge (Öztürk & Papafragou, 2016).

Bulgarian presents a typologically distinct case. Rather than a binary suffix system, Bulgarian employs morphologically related verbal forms, notably the renarrative and conclusive, whose evidential-like meanings emerge from the interaction of the l-participle with auxiliary variation. Sonnenhauser (2013) argued that auxiliary presence or omission encodes a "point of view" (narrator versus non-narrator) with evidential interpretations derived pragmatically from context. The system also encodes mirativity, a related but distinct category (Karagjosova, 2021).

These typological contrasts have implications for habitual epistemic construal. Mushin (2001) proposed that speakers adopt an "epistemological stance," a conceptual structure mediating between information source and linguistic expression. In a narrative retelling study, speakers of Macedonian and Japanese, both with grammaticalized reportive marking, maintained explicit reportive coding in roughly half of their narrative clauses (density indices .49 and .53), while

English speakers reached only .07, defaulting to an "imaginative" stance that reconstructed events without explicit source attribution. This pattern suggests, as Mushin (2001) concluded, "a much higher degree of conventional mapping between actual source of information and adoption of epistemological stance in languages with grammatical evidentiality" (p. 42). This mapping may persist when speakers write in a second language.

Hedging in L2 Academic Writing: L1-Specific Profiles

Hedging, the deployment of devices through which writers modulate commitment to propositions (Hyland, 2005), serves the dual functions of credibility and courtesy and signals membership in an academic discourse community. Learners consistently deploy hedging devices in patterns that diverge systematically from target-language norms. Hyland and Milton (1997) compared two corpora of approximately one million words each, texts by Hong Kong student writers and by British school leavers, and found that L2 writers used simpler constructions, a narrower range of devices, and stronger epistemic commitments. The L2 essays contained 60% more certainty markers, and the L1 essays contained 73% more probability markers. Hinkel (2005), examining 745 essays across six L1 backgrounds, confirmed that L2 writers draw on a restricted, conversationally oriented set of hedging devices, with all non-native groups using emphatics at two to four times native-speaker rates.

Studies focusing on Turkish L1 writers reveal a distinctive hedging profile. Demir (2018), analyzing 200 ELT research articles (100 Anglophone, 100 Turkish), found different rank orderings: native writers favored epistemic verbs, whereas Turkish writers favored modals, preferring *can* over *may*. Turkish writers exhibited significantly lower hedge diversity (1,489 vs. 2,424 distinct types). Çandarlı et al. (2015), combining corpus analysis with interviews, found that Turkish EFL students used *cannot* as a booster, attributed to L1 transfer from negative *can* constructions. Eight of ten interviewees reported being taught to avoid strong modals, yet their Turkish essays contained more boosters than their English essays, suggesting L2 instruction incompletely mediates L1 preferences.

Similar L1-specific patterns emerge across other backgrounds. Oh (2007) found Korean college students significantly underused *would* and *may* while overusing *think* (four times as frequent as in the native corpus). Korean learners used approximately 55% certainty devices versus 46% for native speakers, with the top 10 devices accounting for 61% of learner tokens versus 53% in native data. Šinkūnienė (2008) found English researchers hedged their claims substantially more frequently than Lithuanian researchers, differing in preferred type. Lithuanian researchers favored adverbials, while English researchers relied more on lexical verbs and modals. These asymmetries suggest L1 epistemic systems systematically influence L2 hedging repertoires.

Cross-Linguistic Influence and the Evidentiality–Hedging Connection

Cross-linguistic influence (CLI) theory provides the theoretical bridge between evidentiality typology and L2 hedging behavior. Odlin (2003) established that "language transfer affects all linguistic subsystems including pragmatics and rhetoric, semantics, syntax, morphology, phonology, phonetics, and orthography" (p. 436), and that CLI covers both formal structural transfer and conceptual transfer involving the semantics and pragmatics of the native language influencing L2 production. Zhang and Luo (2017) demonstrated this mechanism for spatial cognition, showing that Chinese secondary school students extended the broader conceptual scope of Chinese spatial expressions (zai, shang) onto English prepositions (at, in, on), producing errors where the L1 category was broader than the L2 category. L1 conceptual categories shaping L2 lexical choices has direct implications for the epistemic domain.

The hedging profiles documented above align with what CLI theory would predict. Turkish writers, whose L1 obligatorily marks the source and directness of evidence through the *-DI/-mış* distinction (Şener, 2011), exhibit a hedging repertoire in English that diverges from native-speaker norms in specific ways: a preference for *can* over *may*, restricted hedge diversity, and the transfer of L1 assertion patterns such as negative can constructions (Çandarlı et al., 2015; Demir, 2018). Korean writers, whose L1 also possesses grammaticalized evidential and

epistemic systems, show a strikingly parallel profile: underuse of *would* and *may*, overreliance on a narrow set of lexical verbs, and stronger epistemic commitment (Oh, 2007). These convergent patterns across two typologically unrelated but evidentially similar languages suggest that obligatory L1 encoding of information source may condition how writers use the optional epistemic resources of English academic prose.

Methodology

Learner and Native Speaker Corpora

The learner data were drawn from the International Corpus of Learner English (ICLE; Granger et al., 2020), a multi-L1 collection of argumentative essays written by upper-intermediate to advanced EFL learners at university level. Three L1 subcorpora were selected: Turkish (ICLE-TR; 286 texts, 203,636 words), Bulgarian (ICLE-BG; 300 texts, 199,654 words), and Spanish (ICLE-SP; 250 texts, 198,305 words). Mean text lengths ranged from 665.5 words for the Bulgarian subcorpus (*SD* = 290.2) to 793.2 for the Spanish subcorpus (*SD* = 416.4), with the Turkish subcorpus falling between them (*M* = 712.0, *SD* = 153.1). All texts are argumentative essays produced under examination or classroom conditions, with metadata covering writer age, gender, years of English study, and time spent in English-speaking countries.

The native English control data were drawn from the Louvain Corpus of Native English Essays (LOCNESS), which comprises argumentative essays written by British and American university students. The LOCNESS subcorpus used in this study contained 175 texts totaling 151,372 words (*M* = 865.0, *SD* = 492.4). Because LOCNESS shares the argumentative essay genre with ICLE, it provides a genre-matched native baseline free of confounding register differences.

The three L1 groups were selected to create a principled typological comparison based on the presence, absence, and structural type of grammaticalized evidentiality in the source languages. Turkish possesses a suffixal evidential system in which the direct evidential -DI encodes witnessed experience while the indirect evidential -mış encodes reportative and inferential information sources in past-tense contexts (Aikhenvald, 2004; Johanson, 2003). Bulgarian has a mood-based evidential system in which a renarrative mood and evidential verb morphology distinguish direct from indirect information sources (Friedman, 1986; Leafgren, 2011). Spanish, by contrast, lacks grammaticalized evidentiality entirely, relying instead on lexical and periphrastic means to express source-of-information distinctions. This design yields a three-way comparison of two languages with grammaticalized evidentiality (suffixal vs. mood-based) and one without, with LOCNESS as the native baseline.

The full corpus comprised 1,011 texts and 752,967 words. Table 1 summarizes the corpus composition by L1 group.

Table 1 Descriptive Statistics for Corpus Subcorpora by L1 Group

L1 Group	N texts	Total words	M words/text	SD
Native (LOCNESS)	175	151,372	865.0	492.4
Turkish (ICLE-TR)	286	203,636	712.0	153.1
Bulgarian (ICLE-BG)	300	199,654	665.5	290.2
Spanish (ICLE-SP)	250	198,305	793.2	416.4
Total	1,011	752,967	—	—

Note. ICLE = International Corpus of Learner English; LOCNESS = Louvain Corpus of Native English Essays. Word counts exclude punctuation tokens.

The four subcorpora are of approximately comparable size in number of texts (175–286) and in total word count, so that subsequent cross-group comparisons are not driven by sampling-size asymmetries; the comparatively smaller LOCNESS corpus is discussed in the Limitations.

Analytical Framework

The analytical framework adapted Aikhenvald's (2004) cross-linguistic evidentiality typology to English hedging devices. Chafe (1986) showed that English possesses extensive lexical evidential resources spanning source of knowledge, mode of knowing, and reliability judgments, and Boye (2012) unified evidentiality and epistemic modality under a single epistemicity domain. Neither framework, however, generates the cross-linguistic predictions needed to test L1-specific transfer. Aikhenvald's (2004) typology was selected because its categorical distinctions map directly onto the evidential architectures of the three source languages, enabling directional predictions. Five categories were used: DIRECT (first-person epistemic stance, e.g., "I believe"); INFERENTIAL (reasoning or evidence-based, e.g., "perhaps," "it seems"); REPORTATIVE (attribution to external sources, e.g., "according to Smith"); ASSUMED (general or background knowledge, e.g., "generally"); and NON_EVIDENTIAL (scalar commitment reduction without encoding information source, e.g., "somewhat"). These categories extend Aikhenvald's grammatical evidentiality parameters to English lexical and constructional hedging devices.

Six REPORTATIVE subtypes captured constructional variation in how writers attribute propositions to external sources. Impersonal passive constructions (e.g., "it is argued that," "it has been suggested that") attribute without naming a source. Named attribution constructions cite a specific source by name (e.g., "Smith argues that"). Unnamed attribution constructions cite generic or collective sources (e.g., "researchers claim that," "people believe that"). Prepositional attribution constructions use prepositional phrases for source marking (e.g., *according to,* "based on," "in the view of"). Reportative adverbs convey hearsay via a single adverb (e.g., "reportedly," "allegedly"). Source-citing constructions reference a source document or passage (e.g., "as noted in," "as Smith found"). These subtypes were central to Research Question 3.

Degree adverbs (e.g., somewhat, relatively, rather, quite, slightly, partly, largely, mostly) were classified as NON_EVIDENTIAL because they modulate the scalar strength of a commitment without encoding information source. They were detected and counted but excluded from all evidential proportion calculations, whose denominator throughout is the sum of DIRECT, INFERENTIAL, REPORTATIVE, and ASSUMED hedges.

Hedge Detection System[1]

Architecture Overview

Hedging devices were detected and classified using a hybrid three-tier architecture. Tier 1, a rule-based system built on spaCy's en_core_web_trf model (Honnibal et al., 2020), handled approximately 85–90% of detections through pattern matching and syntactic disambiguation. Tier 2 used selective LLM processing (Claude Sonnet 4.6; Anthropic, 2025) for ambiguous cases and naked claim extraction. Tier 3 validated Tier 1 against an independent LLM reference on a stratified sample. Full technical specifications are in the OSF package.

Tier 1: Rule-Based Detection

Texts were processed sentence by sentence in two phases, the first detecting multi-word patterns and the second single-word hedges. A token lockout mechanism ensured that tokens claimed by Phase 1 could not be re-detected in Phase 2.

Phase 1 detected six types of multi-word constructions in priority order: generic-knowledge patterns (ASSUMED); impersonal passive evidential constructions (REPORTATIVE/ impersonal_passive); prepositional attribution and source-citing constructions (REPORTATIVE); impersonal inferential constructions (INFERENTIAL); and first-person epistemic fixed phrases (DIRECT). Phase 2a detected single-word hedges from a fixed-category lexicon across eight evidential subcategories. Phase 2b applied a context-dependent lexicon (epistemic modals, reporting verbs, cognitive verbs, data-speaking verbs, and inferential adjectives) requiring

[1] Anonymized view only package is accessible on OSF:
https://osf.io/8jq6z/overview?view_only=6707c20b6f6343a5b44436c55d808e18

syntactic disambiguation. Full lexicon and template specifications are in the OSF package.

The core disambiguation mechanism was subject-type classification via dependency parsing. For each detected hedge, the system identified the clause root's grammatical subject and classified it as: first_person (I, we), expletive (impersonal it), inanimate (results, data, etc.), named_human (proper nouns), unnamed_human (researchers, scholars, etc.), passive_impersonal, or unknown. Verb type and subject type together assigned an evidential category and reportative subtype via a lookup table. For example, "I believe X" was DIRECT (first_person + cognitive verb), while "researchers argue that X" was REPORTATIVE/unnamed_attribution (unnamed_human + reporting verb).

Tier 2: LLM Processing

When Tier 1 flagged a hedge but could not confidently classify it, the case was sent to Claude Sonnet 4.6 (Anthropic, 2025) with the full sentence context and the detected hedge span. Prompted with Aikhenvald's (2004) five-category typology and the six reportative subtypes, the model returned an evidential category, a reportative subtype where applicable, and a one-sentence rationale. These cases represented a small fraction of overall detections.

For the propositional distancing analysis (RQ4), a naked claim (the bare assertion without the hedging device) was extracted from every hedged sentence via Claude Haiku 4.5 (Anthropic, 2025), which removed epistemic modals, attribution frames, and hedging adverbs while preserving core propositional content and any L2 grammatical errors.

Propositional Distancing

Propositional distancing was measured using an NLI cross-encoder (He et al., 2021). Each hedged sentence was paired with its naked claim; the naked claim served as premise and the hedged sentence as hypothesis. The distancing score was 1 − P(entailment). Scores near zero indicated proposition-preserving hedges (e.g., adding "according to" without altering the claim), while scores near 1.0 indicated substantial propositional reframing.

The cross-encoder was trained on standard English text, and its behavior on L2 learner English with non-standard syntax has not been independently validated, which is acknowledged as a limitation. The covariate-adjusted RQ4 analysis provides indirect evidence that the metric is not systematically confounded by general L2 proficiency.

Detection Validation

Detection reliability was assessed via Tier 3 validation, comparing Tier 1 detections against an independent LLM reference (Claude Sonnet 4.6; Anthropic, 2025) on 198 stratified sentences (66 per L1 group; native texts excluded). The LLM received the same typology and subtypes with instructions to identify all hedging devices at the functional level. As shown in Table 2a, precision was .682, so approximately two-thirds of Tier 1 detections were LLM-confirmed. Raw recall was .438 (F1 = .534); after excluding LLM over-detections of items outside Aikhenvald's framework (quantifiers, discourse markers, rhetorical questions), adjusted recall was .58–.65. Category agreement for the 103 co-detected hedges was substantial (κ = .627). Only one LLM hallucination was detected via manual verification.

The 131 lexicon gaps (LLM-detected hedges missed by Tier 1; Table 2b) were unevenly distributed across L1 groups, $\chi^2(2) = 8.49$, $p = .014$. Bulgarian texts were over-represented (59 gaps, 45.0%; z = 2.32), suggesting more corpus-atypical hedging forms in Bulgarian learner English. Reportative subtype classification is deterministic. Once a hedge is detected and its subject identified, subtype assignment follows mechanically from the disambiguation rules.

Table 2a Tier 3 Validation: Detection Performance Against LLM Reference (Sonnet 4.6)

Metric	Value
Sentences sampled	198
Tier 1 hedges detected	151
LLM hedges detected	236
True positives (both)	103
Precision	.682
Recall	.438
Adjusted recall (est.)	.58–.65
***F*1**	.534
Category κ	.627
Lexicon gaps	131
LLM hallucinations	1

Note. Precision and recall use LLM detections as reference. Adjusted recall excludes LLM over-detections of items outside Aikhenvald's evidentiality framework. Category κ = .627 indicates substantial agreement (Landis & Koch, 1977).

Table 2b LLM-Detected Hedges Missed by Tier 1 Rules, by L1 Group

L1 Group	Gaps	%	Expected
Bulgarian	59	45.0	43.7
Turkish	39	29.8	43.7
Spanish	33	25.2	43.7

Note. Lexicon gaps = hedges detected by LLM but missed by Tier 1. Expected values assume equal distribution given stratified sampling (66 sentences per group). $\chi^2(2) = 8.49$, $p = .014$. Bulgarian over-represented (std. residual = 2.32).

Tier 3 thus indicates a usable but imperfect detection layer, and the slight Bulgarian under-detection is revisited in the Limitations as a factor that conservatively attenuates rather than inflates the differences reported below.

Statistical Analysis

All four research questions were addressed within a unified hierarchical Bayesian regression framework, implemented in PyMC with NUTS sampling (Hoffman & Gelman, 2014; Salvatier et al., 2016). This framework was selected over univariate non-parametric tests because the data are intrinsically nested (hedge tokens within texts, texts within L1 groups), and treating observations as independent overstates the effective sample size (Gelman & Hill, 2007). It also permits direct quantification of how much L1 evidentiality contributes to variance in hedging behavior.

In all models, the L1 group factor was specified as a fully random effect across the four groups (Native, Turkish, Bulgarian, Spanish). Treating L1 as random rather than fixed aligns with the theoretical claim, since the ghost construction hypothesis concerns not any specific language but the category of "having grammaticalized evidentiality" generating variance in L2 hedging. The L1 random-effect standard deviation, σL1, therefore quantifies cross-linguistic influence at each level of analysis. All models also included text-level random intercepts to absorb writer-specific variation and propagate that uncertainty into the L1 contrasts.

For Research Question 1 (evidential category proportions across L1 groups), a hierarchical multinomial-logit regression was fit at the hedge level with evidential category (DIRECT, INFERENTIAL, REPORTATIVE, ASSUMED; NON_EVIDENTIAL excluded by the framework rule) as a four-level categorical outcome, INFERENTIAL as the reference, L1 random effects per non-reference category, and a shared text-level random intercept. For Research Question 2 (eight directional typological predictions), no separate model was fit; each prediction was tested as a contrast on the joint posterior from the category and subtype multinomial models, reported as

the posterior probability of the predicted direction with the 95% HDI on the contrast.

For Research Question 3 (reportative subtype distributions and the named-attribution split), two complementary models were fit. The primary analysis was a hierarchical Bernoulli regression of named-attribution use among reportative hedges, testing the headline finding directly. The secondary analysis was a hierarchical multinomial-logit regression on the full six-way subtype profile (unnamed attribution as reference).

For Research Question 4 (propositional distancing across L1 groups and evidential categories, with proficiency-confound adjustment), a hierarchical Beta regression was fit at the sentence level. Distancing scores in (0, 1) were the outcome, with boundary observations squashed via Smithson and Verkuilen's (2006) transformation to retain Beta-family support. L1 and evidential category were treated as crossed random effects with their interaction also estimated as random. Two model versions were compared. The unadjusted model contained only the random-effect structure; the adjusted model added standardized covariates for lexical diversity (MTLD; McCarthy & Jarvis, 2010), syntactic complexity (mean length of clause, MLC; Lu, 2010), and L2 exposure (Years of English at university; Months in an English-speaking country). The remaining L2SCA indices (mean length of T-unit, clauses per T-unit, dependent clauses per T-unit) were pruned for severe pairwise collinearity ($r \geq .93$); their shared signal is retained through MLC. Models were compared on out-of-sample predictive accuracy via PSIS-LOO cross-validation (Vehtari et al., 2017). As a sensitivity analysis, the adjusted model was re-estimated as a zero-or-one-inflated Beta (ZOIB) with L1-specific inflation logits, addressing possible point masses at 0 or 1.

Priors throughout were weakly informative (McElreath, 2020). Posterior inference used NUTS sampling, with convergence confirmed via $\hat{R} \leq 1.01$ (Vehtari et al., 2021). Posterior contrasts are summarized as the posterior mean of the contrast, the 95% highest density interval, and P(direction | data), the posterior probability that the contrast lies on the predicted side of zero. No dichotomous significance thresholds are imposed; contrasts whose 95% HDI excludes zero are described as credibly different. Full prior specifications, implementation versions, and replication code are on OSF.

Results

Overall Hedging Patterns

Hedging rates per 1,000 words were broadly comparable across the four L1 groups, ranging from 19.5 (Spanish) to 22.9 (Native), as shown in Table 3. Within-group variation was substantial (SD = 8.6–10.4, CV = 40–48%), so much of the variance in hedging frequency reflects individual writing style rather than L1 background.

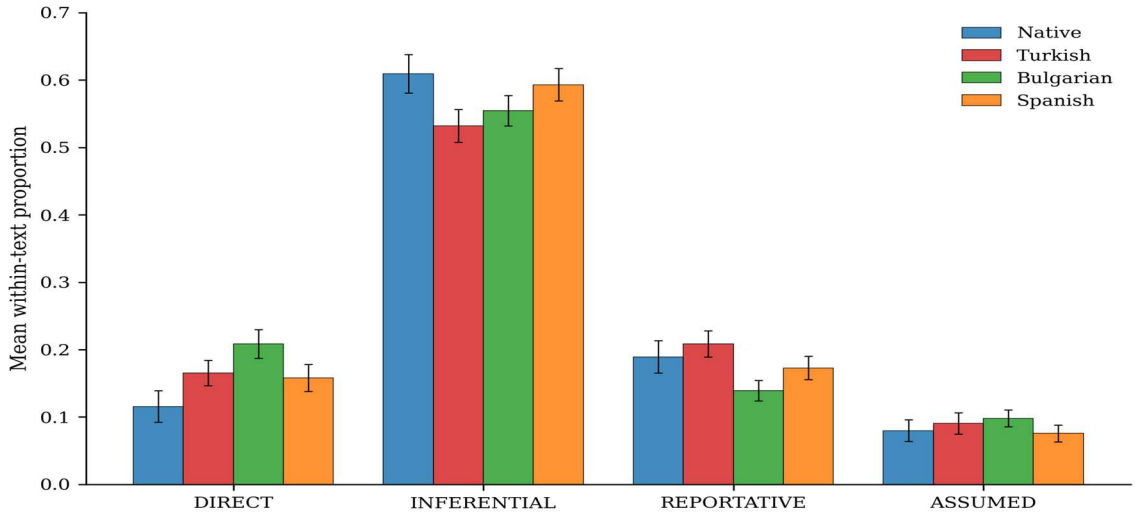
Table 3 Descriptive Statistics for Hedging Frequency and Evidential Category Distribution by L1 Group

L1 Group	N	Total Words	Hedges/1000w *M* (*SD*)	DIRECT *M* (*SD*)	INFERENTIAL *M* (*SD*)	REPORTATIVE *M* (*SD*)	ASSUMED *M* (*SD*)	NON_EV Prop.
Native	175	151,372	22.9 (10.4)	.116 (.158)	.610 (.193)	.189 (.163)	.080 (.110)	.033
Turkish	286	203,636	20.8 (10.0)	.165 (.164)	.532 (.209)	.209 (.168)	.090 (.137)	.026
Bulgarian	300	199,654	21.7 (8.7)	.208 (.188)	.555 (.199)	.139 (.135)	.098 (.110)	.078
Spanish	250	198,305	19.5 (8.6)	.158 (.162)	.593 (.192)	.173 (.141)	.076 (.100)	.033

Note. DIRECT through ASSUMED columns show mean within-text proportions of evidential hedges (excluding NON_EVIDENTIAL). NON_EV Prop. = corpus-level proportion of all detected hedges classified as NON_EVIDENTIAL (degree adverbs). Hedges/1000w = normalized hedging rate per 1,000 words. Leading zeros omitted for proportions bounded by 1.

The distribution of hedges across evidential categories followed a consistent pattern (see Figure 1). INFERENTIAL hedges dominated (53.2–61.0% of evidential hedges), consistent with argumentative register. DIRECT hedges constituted the second-largest category for L2 groups (15.8–20.8%), but native speakers produced more REPORTATIVE (18.9%) than DIRECT hedges (11.6%). Bulgarian writers showed the highest DIRECT proportion (20.8%), nearly double the native rate. ASSUMED hedges were the smallest category (7.6–9.8%). All groups shared a broadly similar profile, with INFERENTIAL dominance and modest variation elsewhere.

Figure 1 Evidential Category Proportions by L1 Group



Note. Bars show the raw within-text proportion of evidential hedges in each category, averaged across the texts in each L1 group; whiskers are 95% confidence intervals around the L1 mean. NON_EVIDENTIAL is excluded from the denominator. Native = LOCNESS corpus. Posterior-predicted proportions with 95% HDI whiskers appear in Figure S1 of the OSF supplement, where model-predicted means shrink each L1 toward the population mean by 1–3 percentage points.

The NON_EVIDENTIAL category, comprising scalar degree modifiers outside Aikhenvald's framework, accounted for only 4.3% of detected hedges, so the framework captured approximately 96% of the hedging inventory. Bulgarian writers showed an elevated NON_EVIDENTIAL rate (.078) compared to the other three groups (.026–.033), driven by more frequent use of scalar degree modifiers. NON_EVIDENTIAL hedges were excluded from all subsequent evidential category analyses.

Evidential Category Comparisons

A hierarchical multinomial-logit regression with L1 random effects and text random intercepts was fit to the hedge-level evidential category outcomes (DIRECT, INFERENTIAL, REPORTATIVE, ASSUMED; NON_EVIDENTIAL excluded by the framework rule). The L1 random-effect standard deviation σ_{L1} quantifies how much L1 evidentiality drives variance in each non-reference category's log-odds: σ_{L1} = 0.63 [0.21, 1.27] for DIRECT, 0.37 [0.10, 0.84] for REPORTATIVE, and 0.39 [0.09, 0.90] for ASSUMED (no σ for the reference category INFERENTIAL by construction). The 95% HDIs for all three hyperparameters exclude zero, so L1 is a credible source of variance in category proportions, though the magnitudes are small to moderate rather than large.

Pairwise L1 contrasts reported in Table 4 reveal a consistent pattern. For DIRECT, Bulgarian writers showed credibly higher use than each other group (Bulgarian − Native = 1.02 [0.83, 1.21]; Bulgarian − Spanish = 0.32 [0.17, 0.47]; Bulgarian − Turkish = 0.22 [0.08, 0.38]; all P ≥ .998). For REPORTATIVE, Native writers exceeded Bulgarian (0.47 [0.30, 0.64]) and Spanish (0.28 [0.12, 0.44]), while Turkish exceeded Bulgarian (0.43 [0.26, 0.58]) and Spanish (0.24 [0.10, 0.40]). For ASSUMED, Bulgarian and Turkish writers exceeded Spanish and Native, with all

Bulgarian contrasts excluding zero and the Turkish–Native contrast credibly different. Figure 1 shows the descriptive within-text proportions per L1 group with 95% confidence intervals, the corresponding posterior marginal probabilities appearing in Figure S1 of the OSF supplement. The substantial overlap across groups for INFERENTIAL and the Bulgarian-DIRECT shift are the most salient features. These category-level effects are credible but substantively modest.

Table 4 Pairwise L1 Contrasts on Evidential Category Log-Odds (Hierarchical Multinomial-Logit Posterior)

Category	Contrast	Posterior mean	95% HDI	P(direction)
DIRECT	Native − Turkish	−0.80	[−0.97, −0.60]	< .001
DIRECT	Native − Bulgarian	−1.02	[−1.21, −0.83]	< .001
DIRECT	Native − Spanish	−0.70	[−0.89, −0.52]	< .001
DIRECT	Turkish − Bulgarian	−0.22	[−0.38, −0.08]	0.002
DIRECT	Turkish − Spanish	0.10	[−0.05, 0.25]	0.884
DIRECT	Bulgarian − Spanish	0.32	[0.17, 0.47]	> .999
REPORTATIVE	Native − Turkish	0.04	[−0.11, 0.19]	0.694
REPORTATIVE	Native − Bulgarian	0.47	[0.30, 0.64]	> .999
REPORTATIVE	Native − Spanish	0.28	[0.12, 0.44]	> .999
REPORTATIVE	Turkish − Bulgarian	0.43	[0.26, 0.58]	> .999
REPORTATIVE	Turkish − Spanish	0.24	[0.10, 0.40]	0.999
REPORTATIVE	Bulgarian − Spanish	−0.19	[−0.34, −0.02]	0.012
ASSUMED	Native − Turkish	−0.23	[−0.45, −0.03]	0.013
ASSUMED	Native − Bulgarian	−0.51	[−0.72, −0.31]	< .001
ASSUMED	Native − Spanish	−0.07	[−0.28, 0.14]	0.263
ASSUMED	Turkish − Bulgarian	−0.28	[−0.46, −0.09]	0.002
ASSUMED	Turkish − Spanish	0.16	[−0.03, 0.36]	0.954
ASSUMED	Bulgarian − Spanish	0.44	[0.25, 0.64]	> .999

Note. Posterior contrasts on the log-odds scale. 95% HDI = 95% highest density interval. P(direction) = posterior probability that the first-listed group's value exceeds the second's; reported as < .001 when below .001 and > .999 when above .999.

These category-level posterior contrasts make the Bulgarian-DIRECT shift and the Native-REPORTATIVE / Turkish-REPORTATIVE pairing visually salient, but the moderate σ_{L1} magnitudes indicate that categorical-level transfer is a secondary phenomenon to the constructional-level effects examined below.

Language-Specific Transfer Predictions

The eight typological predictions derived from Aikhenvald's (2004) framework were tested as posterior contrasts on the joint posterior of the category and subtype multinomial models, marginalizing over writer effects to express each contrast on the proportion scale the predictions assume. For each prediction, Table 5 reports the posterior mean of the contrast, the 95% HDI on the contrast, the posterior probability of the predicted direction, and a categorical verdict (supported, null, or actively reversed).

None of the eight predictions receives credible posterior support. Four (P1-TR, P3a-TR, P4-TR, P2-ES) are null. Their 95% HDIs cover zero and the posterior probabilities of the predicted direction cluster around .2–.4, indicating contrast magnitudes too small to be credibly nonzero.

Four predictions (P2-TR, P3b-TR, P1-BG, P2-BG) are actively reversed. Their 95% HDIs lie entirely on the opposite side of zero from the predicted direction (P(predicted direction) < .05). This pattern indicates that the categorical-level predictions derived from L1 evidential typology do not survive contact with the data. The disconfirmation is not a Type II result from underpowered tests, since the reversed predictions exclude zero by wide margins and the null predictions show narrow HDIs around zero. That L1 evidentiality does not transfer through a direct mapping of categorical proportions motivates the constructional analysis that follows.

Table 5 Posterior Contrasts for Eight Typological Predictions on the Proportion Scale

ID	Claim	Predicted dir.	Δ proportion	95% HDI	P(predicted dir)
P1-TR	Turkish P(REP)+P(INF) > Spanish P(REP)+P(INF)	TR > SP	−0.009	[−0.031, 0.014]	0.210
P2-TR	Within-Turkish P(REP) > P(INF)	REP > INF	−0.353	[−0.398, −0.306]	0.000
P3a-TR	Turkish P(imp_passive|REP) > Bulgarian P(imp_passive|REP)	TR > BG	−0.010	[−0.059, 0.039]	0.350
P3b-TR	Turkish P(imp_passive|REP) > Spanish P(imp_passive|REP)	TR > SP	−0.054	[−0.105, −0.004]	0.018
P4-TR	Turkish P(DIRECT) < Spanish P(DIRECT)	TR < SP	+0.003	[−0.016, 0.020]	0.379
P1-BG	Bulgarian P(REP) > Spanish P(REP)	BG > SP	−0.039	[−0.057, −0.022]	0.000
P2-BG	Bulgarian P(REP)/P(INF) > Turkish P(REP)/P(INF)	BG > TR	−0.127	[−0.173, −0.080]	0.000
P2-ES	Spanish P(DIRECT) > Turkish P(DIRECT)	SP > TR	−0.003	[−0.020, 0.016]	0.380

Note. Posterior contrasts on the proportion scale, computed by marginalizing the category and subtype multinomial posterior draws over writer effects via Monte Carlo. P(predicted direction) is the posterior probability that the contrast lies on the predicted side of zero.

None of the eight predictions receives credible posterior support. The substantive implication is that L1 evidential typology does not map onto L2 hedging through a one-to-one correspondence of categorical proportions.

Reportative Subtype Analysis

The named-attribution Bernoulli model and the six-way subtype multinomial, specified in the Methodology, returned an L1 random-effect standard deviation on named-attribution log-odds of σL1 = 1.14 [0.48, 1.94], a large effect with the 95% HDI well above zero. This is the largest robustly estimated L1 effect in this study, larger than σ_{L1} on any of the four category proportions.

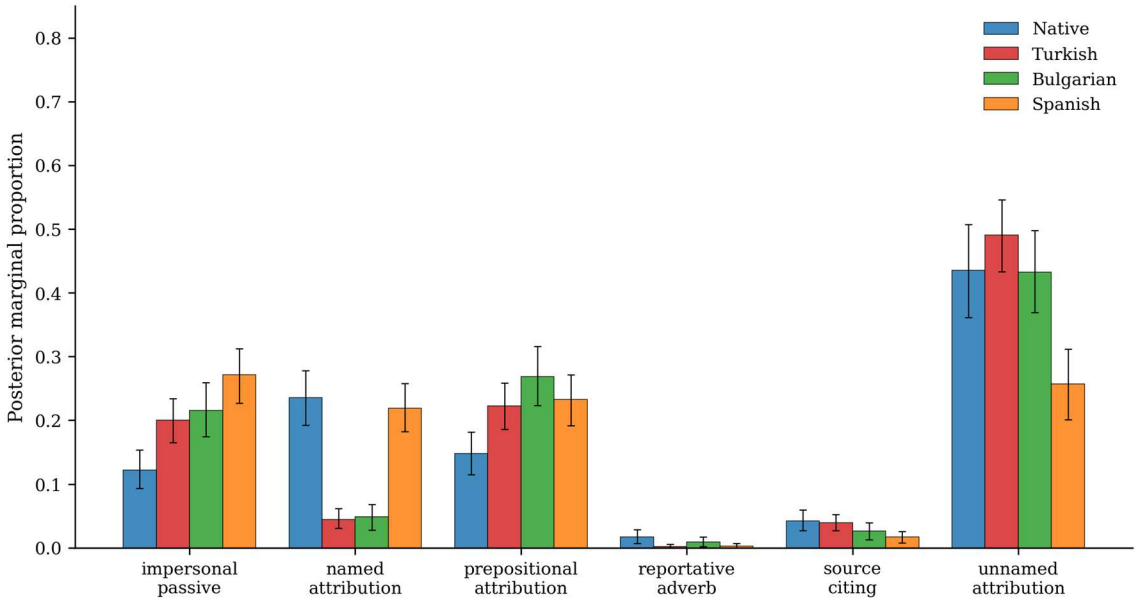
Table 6 reports the L1 pairwise contrasts on the named-attribution log-odds. The Native–Turkish contrast (1.87, 95% HDI [1.25, 2.47], P > .999), Native–Bulgarian contrast (1.79, [1.14, 2.43], P > .999), Turkish–Spanish contrast (−1.80, [−2.37, −1.20], P < .001), and Bulgarian–Spanish contrast (−1.72, [−2.35, −1.12], P < .001) all exclude zero by wide margins. The Native–Spanish and Turkish–Bulgarian contrasts cover zero, confirming that the four L1 groups cluster into two pairs (Native/Spanish high; Turkish/Bulgarian low) with no credible difference within either pair.

Table 6 Pairwise L1 Contrasts on Named-Attribution Log-Odds (Hierarchical Bernoulli Posterior)

Contrast (named-attribution log-odds)	Posterior mean	95% HDI	P(direction)
Native – Turkish	1.87	[1.25, 2.47]	> .999
Native – Bulgarian	1.79	[1.14, 2.43]	> .999
Native – Spanish	0.07	[−0.37, 0.60]	0.614
Turkish – Bulgarian	−0.07	[−0.69, 0.66]	0.405
Turkish – Spanish	−1.80	[−2.37, −1.20]	< .001
Bulgarian – Spanish	−1.72	[−2.35, −1.12]	< .001

Note. Posterior log-odds contrasts on named-attribution from the hierarchical Bernoulli regression. σ_{L1} = 1.14 [0.48, 1.94], indicating large cross-L1 variance in this constructional choice. 95% HDI = 95% highest density interval.

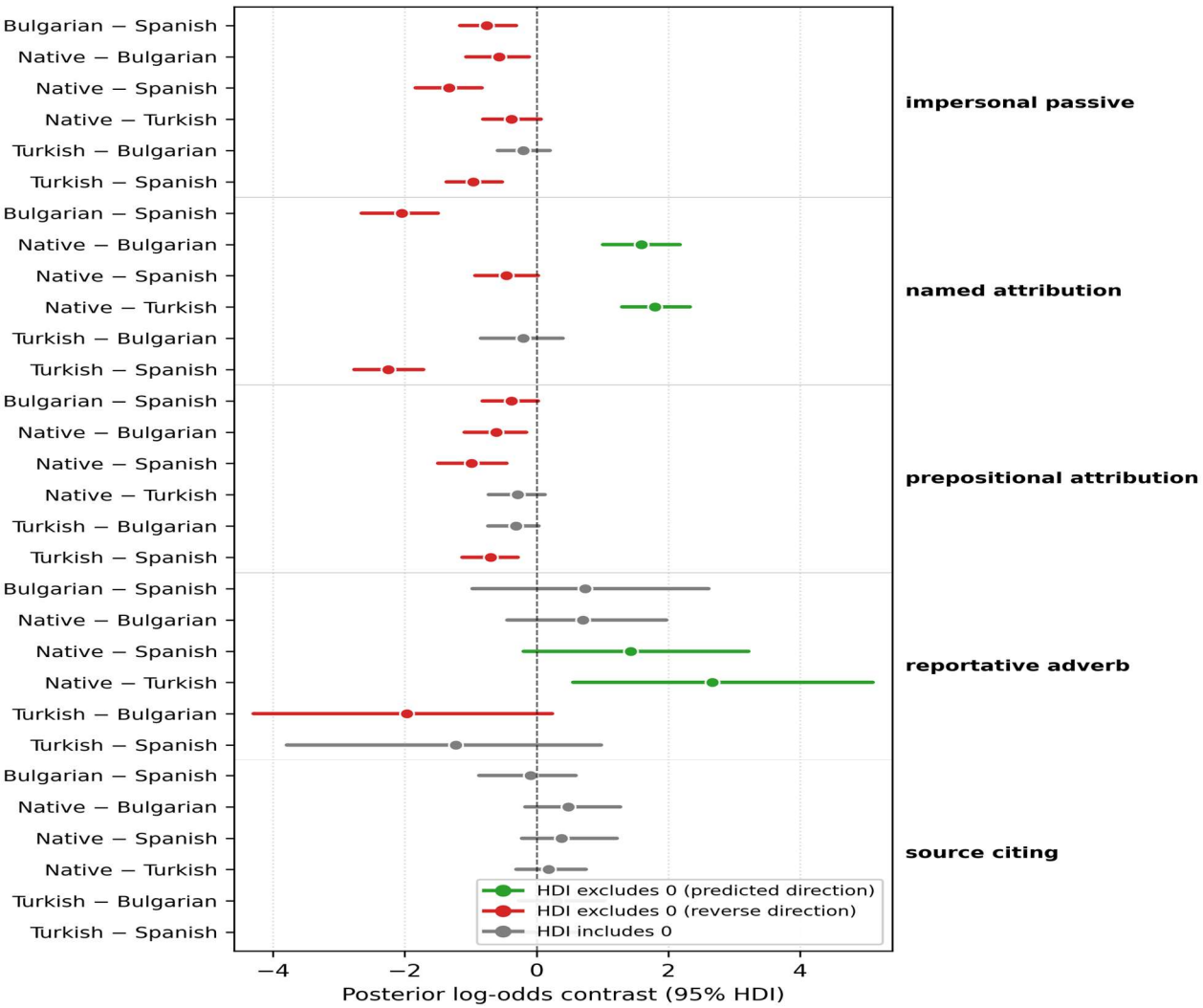
Figure 2 Reportative Subtype Proportions by L1 Group



Note. Proportions within the REPORTATIVE category. Named attribution and unnamed attribution show the largest cross-linguistic variation.

The constructional fragmentation of reportative hedging across L1 backgrounds is the clearest cross-linguistic transfer signal in this study, and the underlying pattern (Turkish/Bulgarian writers replacing named attribution with source-nonspecific impersonal and prepositional constructions) directly mirrors the source-anonymous default of their respective L1 reportative systems.

Figure 3 Reportative Subtype Posterior Log-Odds Contrasts (95% HDIs)



Note. Posterior log-odds contrasts on each non-reference subtype (reference = unnamed_attribution). Points show posterior means; bars show 95% highest density intervals.

Color marks whether the HDI excludes zero in the predicted or reverse direction, or includes zero. σ_{L1} by subtype: impersonal_passive 0.76 [0.26, 1.47]; named_attribution 1.20 [0.56, 2.01]; prepositional_attribution 0.63 [0.17, 1.32] (Bulgarian-driven); reportative_adverb 1.25 [0.34, 2.33]; source_citing 0.43 [0.00, 1.09]. 95% HDI = 95% highest density interval.

Across the full six-way distribution, the L1 contrasts on the impersonal passive run in the opposite direction to named-attribution (Spanish > Bulgarian > Turkish > Native), the prepositional attribution contrasts identify Bulgarian as exceptional (preferring prepositional sources more than all other groups), and the reportative adverb contrasts show Native writers using reportative adverbs far more than any L2 group.

The reportative subtype analysis thus revealed a split between languages with grammaticalized evidentiality (Turkish, Bulgarian) and those without (Spanish, Native) in the dimension of source attribution specificity. Turkish and Bulgarian writers avoided naming sources when constructing reportative hedges, while Spanish and Native writers preferred explicit named attribution. The Turkish–Bulgarian difference in preferred alternatives (unnamed attribution for Turkish, prepositional attribution for Bulgarian) suggests that the specific structure of the L1 evidential system, not merely its presence or absence, influenced constructional preferences. These patterns emerged despite the absence of category-level transfer effects.

Propositional Distancing

The RQ4 Beta regression was fit in unadjusted and covariate-adjusted versions as specified in the Methodology. Table 7 provides example sentences at each distancing level.

Table 7 Example Sentences at Different Propositional Distancing Levels (REPORTATIVE Category)

Score range	Score	L1	Hedged → Naked	Label
Near zero (0.00–0.10)	0.003	Native	Hedged: The government claim that it is a free country, but it is not. Naked: The government claim that it is a free country, but it is not.	Near-paraphrastic
Low-medium (0.20–0.35)	0.298	Native	Hedged: This possible consequence causes people to think twice before accepting and supporting Dr Kevorkian and his methods. Naked: This consequence causes people to think twice before accepting and supporting Dr Kevorkian and his methods.	Mild modification
Medium-high (0.40–0.60)	0.407	Native	Hedged: The sick people that could be helped with doctor prescribed pot should be the first priority foe the DEA. Naked: The sick people that can be helped with doctor prescribed pot are the first priority foe the DEA.	Moderate departure
High (0.80–1.00)	1.000	Native	Hedged: The Democrats propose that a two-year time limit should be placed on what they call a second chance or being able to receive welfare. Naked: A two-year time limit is placed on what is called a second chance or being able to receive welfare.	Full reframing

Note. Score = 1 − P(entailment) from the NLI cross-encoder. A reportative construction can score near zero when the hedge adds a source marker but preserves the proposition nearly verbatim. Corpus excerpts are reproduced verbatim, including L2 grammatical surface variation (e.g., *government claim*, *foe for for*, *prescribed pot*).

The unadjusted model identified a credibly positive log-odds contrast between Native and each L2 group on distancing: Native − Turkish = 0.47 [0.28, 0.65], P = .999; Native − Bulgarian = 0.39 [0.21, 0.58], P = .999; Native − Spanish = 0.55 [0.35, 0.74], P > .999. Within-L2 contrasts had 95% HDIs covering zero. The adjusted model returned essentially the same L1 contrasts (Table 8): Native − Turkish = 0.43 [0.23, 0.62], Native − Bulgarian = 0.46 [0.26, 0.66], Native − Spanish = 0.57 [0.36, 0.77], all with P > .999. PSIS-LOO model comparison favored the adjusted model (ΔELPD = −9.0, SE 3.5, model weight = 1.0), indicating that the complexity battery carries genuine signal, but as the surviving contrasts demonstrate, that signal operates within rather than between L1 groups. MTLD has a credibly positive slope on distancing (γ_{MTLD} = 0.074 [0.04, 0.11], P > .999), while MLC, Years at university, and Months abroad show null effects. The substantive implication is that proficiency-related variance in distancing exists but is not the source of the L2–Native gap.

Table 8 Pairwise L1 Contrasts on Distancing Log-Odds: Unadjusted vs Covariate-Adjusted Models

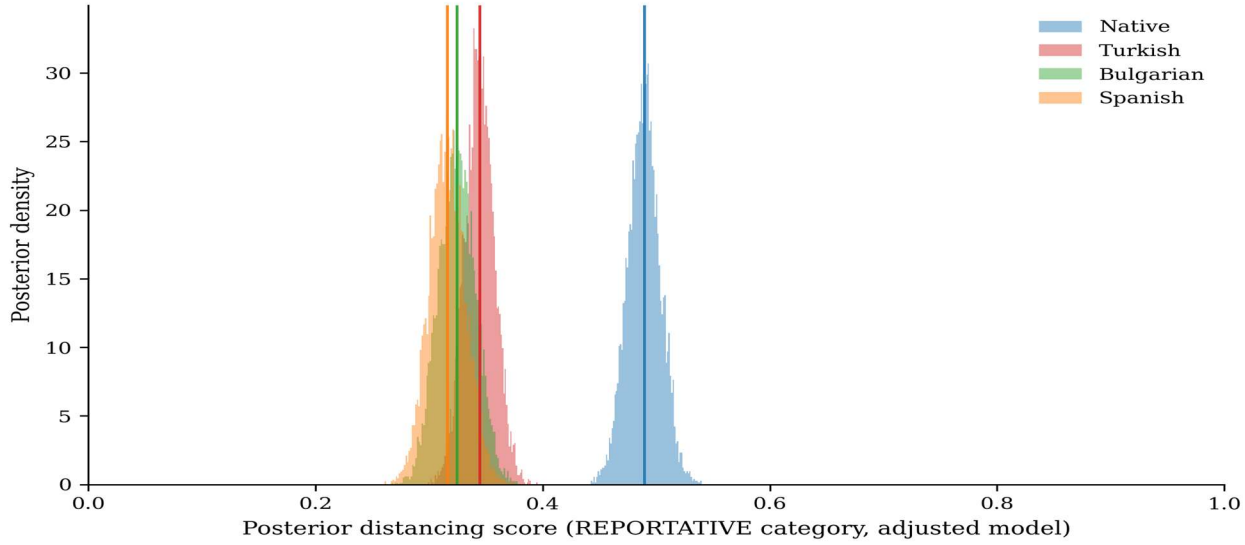
Contrast		**Unadj. mean**	**Unadj. 95% HDI**	**Unadj. P(dir)**	**Adj. mean**	**Adj. 95% HDI**	**Adj. P(dir)**
Native	− Turkish	0.465	[0.279, 0.648]	0.999	0.430	[0.229, 0.625]	> .999
Native	− Bulgarian	0.387	[0.207, 0.577]	0.999	0.462	[0.259, 0.662]	> .999
Native	− Spanish	0.549	[0.351, 0.738]	> .999	0.567	[0.356, 0.768]	> .999
Turkish	− Bulgarian	−0.078	[−0.253, 0.095]	0.173	0.032	[−0.169, 0.217]	0.638
Turkish	−	0.083	[−0.094, 0.259]	0.831	0.138	[−0.064, 0.316]	0.927

Contrast	Unadj. mean	Unadj. 95% HDI	Unadj. P(dir)	Adj. mean	Adj. 95% HDI	Adj. P(dir)
Spanish						
Bulgarian – Spanish	0.161	[−0.030, 0.337]	0.961	0.105	[−0.075, 0.294]	0.877

Note. Pairwise L1 log-odds contrasts on the distancing outcome from the hierarchical Beta regression. The unadjusted model contains only the random-effect structure (L1, evidential category, L1×category interaction, text). The adjusted model adds standardized covariates for MTLD, MLC, Years at university, and Months abroad. PSIS-LOO favored the adjusted model (ΔELPD = −9.0, SE 3.5, weight = 1.0).

In the adjusted Beta model (Figure 4), the leftward shift of the three L2 distributions relative to Native is the empirical signature of the L2–Native gap, and the substantial within-group dispersion visible in each posterior density indicates that this is a group-level shift rather than a uniform per-writer effect.

Figure 4 Reportative Propositional Distancing Scores by L1 Group



Note. Distancing score = 1 − P(entailment) from the NLI cross-encoder. Higher values indicate greater propositional distancing. Posterior distributions show the Native distribution shifted rightward relative to the three L2 distributions; the three Native–L2 contrasts on the latent log-odds scale have 95% HDIs excluding zero in both the unadjusted and the covariate-adjusted hierarchical Beta models.

A sensitivity analysis re-estimated the adjusted RQ4 model as a zero-or-one-inflated Beta (ZOIB) with L1-specific inflation logits at both endpoints. ZOIB and Beta-adjusted L1 contrasts agreed to two decimal places (max |Δ| = 0.004; HDI widths within 0.014), preserving the L2–native gap under both distributional models. Full ZOIB specifications and posterior comparison are in the OSF package.

Cross-Analysis L1 Effect Summary

The σL1 hyperparameters in Table 9 and Figure 5 quantify cross-linguistic influence at each level of analysis. The defining pattern is the ordering. σL1 is small to moderate at the category level (0.37–0.63), large at the constructional level (1.14 for named-attribution as a binary outcome, 1.20 for the named-attribution dimension in the full multinomial, 1.25 for the reportative adverb subtype), and small at the propositional-distancing level (0.42). Within the constructional layer, the largest σL1 values correspond to the subtypes whose source-specificity properties differ across L1 evidential systems. This is the quantitative form of the ghost construction argument, in which L1 evidentiality enters L2 hedging as a constructional template (a preference for one realization of reportative source-marking over another) rather than as a

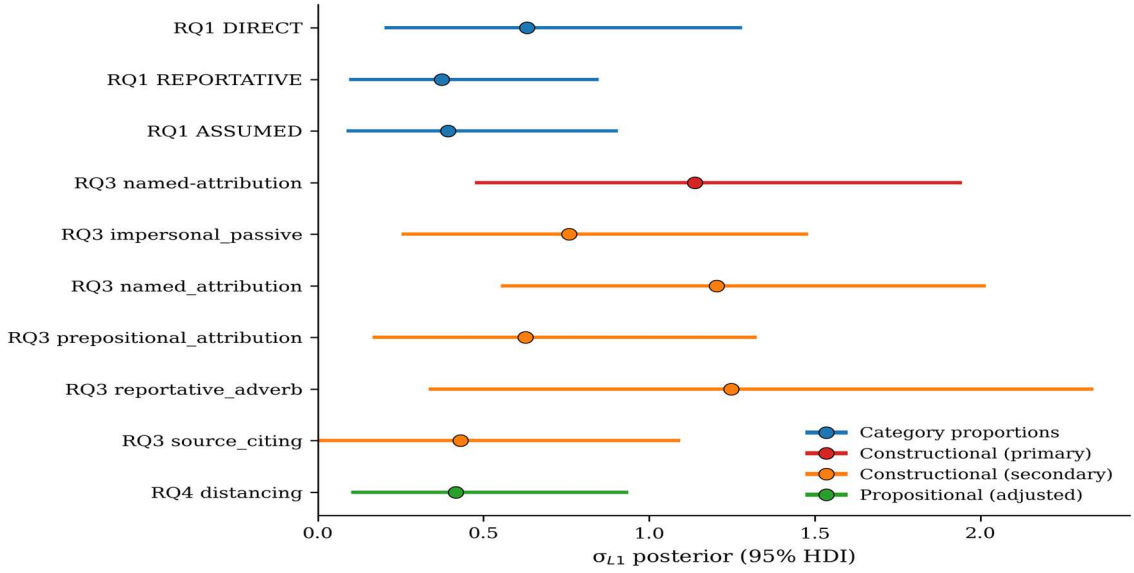
categorical re-weighting. The cross-L1 variance hyperparameter is 1.8 to 3.4 times larger across the constructional subtypes than at the categorical level, a quantitative expression of where L1 evidentiality leaves its trace.

Table 9 σ_{L1} Standard Deviation Decomposition Across Bayesian Analyses (Hierarchical Random-Effect σ)

Analysis	Section	σ_{L1} mean	95% HDI	Magnitude
RQ1 DIRECT	Category proportions	0.63	[0.21, 1.27]	Moderate
RQ1 REPORTATIVE	Category proportions	0.37	[0.10, 0.84]	Small
RQ1 ASSUMED	Category proportions	0.39	[0.09, 0.90]	Small
RQ3 named-attribution	Constructional (primary)	1.14	[0.48, 1.94]	Large
RQ3 impersonal_passive	Constructional (secondary)	0.76	[0.26, 1.47]	Moderate
RQ3 named_attribution	Constructional (secondary)	1.20	[0.56, 2.01]	Large
RQ3 prepositional_attribution	Constructional (secondary)	0.63	[0.17, 1.32]	Moderate
RQ3 reportative_adverb	Constructional (secondary)	1.25	[0.34, 2.33]	Large
RQ3 source_citing	Constructional (secondary)	0.43	[0.00, 1.09]	Small
RQ4 distancing	Propositional (adjusted)	0.42	[0.11, 0.93]	Small

Note. σ_{L1} = posterior mean of the L1 random-effect standard deviation on each outcome's natural scale (log-odds for RQ1 and RQ3; logit-scale for RQ4 distancing). 95% HDI = 95% highest density interval.

Figure 5 σ_{L1} Forest Plot Across Bayesian Analyses (95% HDIs)



Note. Across the four analyses, σL1 is largest for the named-attribution contrast and smaller for the category-level (0.37–0.63) and distancing (0.42) contrasts, whose ranges overlap, locating the strongest cross-linguistic signal at the constructional layer without credibly separating the categorical and propositional layers. Read together, the σ_{L1} estimates support the inference that L1 evidentiality influences L2 hedging chiefly through the constructional realization of reportative meaning rather than through scalar shifts in category proportions or sentence-level commitment.

Discussion

The ghost construction hypothesis predicted that obligatory L1 evidential categories would persist as functional traces in L2 hedging. The results partially support this, but at an unexpected level. At the evidential category level, the four groups were more similar than different, and none of the eight typological predictions were confirmed. At the constructional level, reportative subtype distributions varied substantially, with the named-versus-unnamed attribution split the primary locus of L1 difference. Propositional distancing showed a consistent L2–native gap but no L1-specific effects.

Category-Level Similarity and the Bulgarian DIRECT Anomaly (RQ1)

The finding that the three non-reference evidential categories produced small-to-moderate L1 random-effect standard deviations (σ_{L1} = 0.37–0.63, with 95% HDIs above zero but well below the σ_{L1} = 1.14 observed at the constructional level) indicates that L1 background contributes credibly but modestly to variance in category-level hedging proportions. The substantial within-group dispersion and extensive overlap visible in Figure 1 imply that the four groups are predominantly drawn from a shared distribution of hedging behavior, with L1-typological influence a secondary modulating factor rather than a primary driver.

The pattern that emerges is one of a universal hedging scaffold with L1-specific marginal adjustments. INFERENTIAL hedges accounted for the majority of evidential hedges in every L1 group (53–61% descriptively), indicating that the dense, grammatically flexible inferential register of academic English (perhaps, it seems, the results suggest) is reached for by writers across L1 backgrounds. The L1 evidentiality effect did not redistribute this scaffold; rather, it shaped the marginal choices around it: how much first-person commitment (DIRECT), how much external-source attribution (REPORTATIVE), and how much generic-knowledge framing (ASSUMED). Read this way, the category-level similarity is not a null result but a substantive finding. The L1 effect on hedging acts on the non-default category choices, not on the universal inferential frame within which those choices sit. This reading is conservative. The model regularizes each L1's estimate toward the population mean in proportion to that group's standard error, so per-L1 posterior proportions are more compressed than their descriptive counterparts (Figure S1 in the OSF supplement makes this shrinkage visible). The contrasts whose 95% HDIs exclude zero in Table 4, most notably the Bulgarian DIRECT contrasts, therefore survive a regularization stricter than a frequentist test of group means would impose.

The one category-level finding that deserves closer examination is the elevated DIRECT proportion among Bulgarian writers, with all three Bulgarian contrasts on DIRECT excluding zero by substantial margins (Bulgarian − Native = 1.02 [0.83, 1.21]; Bulgarian − Spanish = 0.32 [0.17, 0.47]; Bulgarian − Turkish = 0.22 [0.08, 0.38]). This Bulgarian-DIRECT bias is striking because Bulgarian has grammaticalized evidentiality (the renarrative mood), which ought, on a naive typological reading, to suppress rather than amplify direct claims. We read this anomaly as evidence that grammatical encoding of evidentiality in the L1 does not transfer straightforwardly into a corresponding categorical distribution in L2 hedging.

The Disconfirmation of Typological Predictions (RQ2)

The blanket failure of the eight typological predictions (four null, four actively reversed) calls into question the categorical reading of L1 evidentiality on which they rest. The predictions assumed that a grammaticalized evidential distinction in the L1 would translate into a proportional bias in the corresponding L2 hedging category. The data show no such mapping, and the reversals indicate that any relationship operates in the opposite direction. The remaining question is therefore not whether L1 evidentiality leaves a trace in L2 hedging, but at what level of linguistic organization that trace appears.

This result demands careful interpretation. Yıldız and Turan (2021) found that Turkish writers of English overused induction markers (*seem, appear, must*) relative to native speakers (1.0 vs. 0.6 occurrences per 1,000 words; LL = 23.68). In their dataset, this was the only evidential

category in which Turkish writers deviated from the native baseline in the direction of overuse. The authors attributed this to Turkish writers discussing findings "more elaborately in consideration of the available evidence," a formulation consistent with the thinking-for-speaking hypothesis (Slobin, 2016). Habitual attention to inferential evidence source in Turkish, encoded by the inferential function of -mış (Aksu-Koç et al., 2009), may transfer to English as an orientation toward inference-marking constructions. Critically, the inferential function of Turkish -mış, encoding conclusions drawn from physical evidence, maps more directly onto English inferential hedges (*seem, appear, probably*) than onto English reportative hedges (*according to, Smith argues*). The massive within-subjects INFERENTIAL dominance may therefore reflect the dominance of inferential over reportative evidential transfer, rather than its absence. Çandarlı et al. (2015) documented a complementary pattern, in which Turkish EFL students used the negative modal cannot as a booster, transferring from Turkish assertion patterns and suggesting that Turkish writers channel epistemic force through modal and inferential devices rather than reportative source attribution.

The broader implication is that Aikhenvald's (2004) typological framework, developed for cross-linguistic comparison of grammatical evidential systems, does not generate predictions that survive operationalization at the level of L2 lexical hedging. It maps categorical boundaries between languages with obligatory morphological evidentiality and was not designed to predict how those categories are realized through optional lexical resources in a typologically different target language, so its predictive scope may be limited to grammatical transfer contexts. Slobin's (2016) thinking-for-speaking framework suggests habitual attention to information source transfers as cognitive orientations finding novel constructional expression in the receiving language. Jarvis (2016) similarly argued that conceptual transfer involves meaning rather than form. The disconfirmation of category-level predictions reflects a disconfirmation of the operationalization rather than of the underlying transfer mechanism.

Constructional Transfer: The Named Attribution Split (RQ3)

The constructional analysis produced the largest single L1 effect in this study. The named-attribution random-effect standard deviation, σL1 = 1.14 [0.48, 1.94], is 1.8 to 3.1 times the magnitude of the category-level σL1 values reported above. L1 evidentiality thus appears to enter L2 hedging not as a rebalancing of which evidential categories writers select, but as a constructional rearrangement of how they realize the reportative category. Native and Spanish writers cluster together at high named-attribution use (approximately 22%); Turkish and Bulgarian writers cluster together at low named-attribution use (approximately 4%). A LOCNESS token illustrates the construction Turkish and Bulgarian writers avoid, "Pattullo feels that homosexuals in the military would destroy the traditional close knit relationships formed between the soldiers," in which the source is identified by a proper-name subject and a finite reporting verb. The clustering is striking because Spanish lacks grammaticalized evidentiality whereas Turkish and Bulgarian have it, so the pattern aligns not with the presence of L1 evidential morphology but with the source-specificity properties those morphologies encode. The Turkish/Bulgarian pattern of attributing claims without naming sources, realized through impersonal passives and prepositional attribution (Figure 3), parallels the source-anonymous default of their respective L1 reportative systems.

For Turkish writers, the preference for unnamed attribution (M = .507) over named attribution (M = .042) is interpretable through multiple converging mechanisms. Tosun and Filipović (2022) identified a one-to-many mapping problem in Turkish–English evidential translation, where the single morpheme -mış must be mapped to multiple English constructions (apparently, it seems, reportedly, it is said that), creating difficulty at the constructional level. Uzum et al. (2025) documented this mapping in practice, showing that Turkish users of English drew on lexical evidential strategies to project the reportive -ImIş in written intercultural exchanges. Turkish -mış encodes reportative meaning without specifying the source, predisposing writers toward unnamed or impersonal constructions that preserve this source-nonspecific quality. ICLE-TR essays illustrate this with constructions such as "Some people think that ignorance is the root of

all evil and believe that providing right education can deter criminal activities," in which a generic collective subject carries the reporting verb and the source remains unindividuated, mapping onto the source-anonymous default of -mış. Akbaş (2014) found that Turkish-background writers, in both L1 and L2 English, constructed less personal academic prose, preferring implicit authorial references (passive constructions, inanimate subjects). Çandarlı et al. (2015) documented the same pattern. In their data, 35 of 48 Turkish students' English essays contained no first-person pronoun I, and all ten interviewees reported being taught to avoid it.

Bulgarian writers showed a different constructional profile despite similarly low named attribution (.046), distinctively overusing prepositional attribution (.263). Vassileva (2001) noted that Bulgarian writers' linguistic devices served different pragmatic functions than English equivalents, with pragmatic failure to hedge being culturally conditioned. The prepositional preference may reflect the renarrative system's perspective-encoding function (Sonnenhauser, 2013), with non-firsthand information expressed through perspective constructions (according to, based on) rather than source-naming ones. An ICLE-BG essay exemplifies this pattern with "According to the system of Bulgarian educating, the first seven years are mostly essential" (institutional source attribution through a prepositional phrase rather than a finite reporting clause), and "in the words of Oscar Wilde who once said …," presenting a named source through a prepositional frame. The preference extends beyond strict source attribution, since the corpus also shows non-attributive uses of according to (e.g., "according to their interests") in which the preposition functions as a general adverbial, a breadth itself consistent with the perspective-encoding interpretation. Because Bulgarian texts were over-represented among missed detections (45.0% of gaps), these constructional patterns are likely conservative estimates. Spanish writers, by contrast, resembled native speakers in named attribution (.213) while overusing impersonal passives (.266) and underusing unnamed attribution (.276), consistent with oscillating between personal and institutional authority.

The convergence of Turkish and Bulgarian writers on low named attribution, despite typologically distinct evidential systems, suggests the shared driver is not the specific evidential structure but the habitual orientation toward information-source marking. Both groups have lifelong routines for marking firsthand versus non-firsthand information, which Tosun et al. (2013) showed create measurable effects including privileging firsthand information and discounting non-firsthand sources. When constructing reportative hedges in English, these writers may default to patterns approximating their L1's source-nonspecific marking: unnamed attribution for Turkish (paralleling -mış), prepositional attribution for Bulgarian (paralleling the renarrative). The "ghost" of L1 evidentiality appears not in hedging frequency but in constructional choices.

An alternative explanation for the named attribution split is that it reflects differences in citation pedagogy across educational systems rather than L1 evidential transfer. Turkish and Bulgarian university writing instruction may not emphasize named source citation to the same degree as Anglophone and Spanish-medium programs, and Uysal (2008) showed that Turkish writers' rhetorical patterns were shaped more by educational context than by broad cultural norms. However, the pedagogical explanation cannot account for the convergence of Turkish and Bulgarian writers on low named attribution despite different educational systems, writing cultures, and evidential structures. If pedagogical context were the primary driver, we would expect the two groups to diverge in their constructional alternatives, yet both independently avoided named attribution while preferring different source-nonspecific strategies (unnamed attribution for Turkish, prepositional attribution for Bulgarian) that parallel the functional properties of their respective L1 evidential systems. This patterned divergence within a shared avoidance is more consistent with L1 evidential transfer than with a purely pedagogical account, though the mechanisms are not mutually exclusive.

Propositional Distancing: A Shared L2 Effect (RQ4)

The distancing analysis found a consistent L2–native gap that survived covariate adjustment for the validated proficiency battery (MTLD for lexical diversity, MLC for syntactic complexity) and the available L2 exposure indicators (Years at university, Months abroad). The posterior contrast between Native and each L2 group on distancing log-odds remained credibly positive after adjustment (Native − Turkish = 0.43 [0.23, 0.62]; Native − Bulgarian = 0.46 [0.26, 0.66]; Native − Spanish = 0.57 [0.36, 0.77]; all P > .999). The complexity battery is not inert, since MTLD has a credibly positive slope on distancing, but the variance it carries operates within rather than between L1 groups. A ZOIB refit confirmed the conclusion is robust to the distributional model. The L2–native distancing gap is therefore best understood as a shared property of L2 writing (perhaps a developmental signature of incomplete commitment-management) rather than as an L1-specific effect, and it is not attributable to proficiency differences as captured by validated text-internal indices.

This converges with Hyland and Milton's (1997) finding that Cantonese L2 writers relied on simpler constructions and stronger commitments, and Hinkel's (2005) documentation of restricted L2 hedging. The restricted formulaic repertoire (Wray, 2002) may explain the gap. Writers lacking formulaic competence default to formulas modifying the surface of claims (adding perhaps or it seems) without transforming the proposition, whereas native speakers access richer conventionalized constructions enabling deeper transformation. Consistent with this, Leclercq and Mélac (2021) found that evidential markers emerged late in the production of L2 learners of French and English, tying the development of evidential expression to proficiency. L2 writers hedge frequently but shallowly. The absence of L1-specific effects provides negative evidence for the ghost construction hypothesis at this level. Constructional traces do not extend to propositional modification depth, which appears governed by L2 proficiency and formulaic competence.

The Ghost Construction Hypothesis Revisited

The σL1 comparison in Figure 5, where constructional-band values dwarf the small-to-moderate categorical and propositional effects, matches the ghost construction prediction of a mechanism that works through the choice of reportative construction rather than evidential category.

This is consistent with Slobin's (2016) account, where evidential categories transfer through creative adaptation of receiving-language constructions, and with Jarvis's (2016) conceptual-versus-formal distinction. What transfers is habitual attention to epistemic dimensions, finding novel constructional expression in L2. Mushin's (2001) framework likewise suggests that speakers of evidential languages adopt a reportive stance that may transfer to L2 writing rather than any specific grammatical form.

Limitations

This study has several limitations. First, the rule-based detection system achieved a precision of .682 and an estimated adjusted recall of .58–.65 against the LLM reference, indicating that approximately one-third of detected hedges may be false positives and one-third to two-fifths of hedges in the texts may have been missed. These detection limits counsel caution in interpreting the constructional contrasts, although the named-attribution split is unlikely to be an artifact of lexicon gaps, since subtype assignment is deterministic once a hedge is detected. The uneven distribution of lexicon gaps across L1 groups (Bulgarian 45.0% of gaps, $\chi^2(2) = 8.49$, $p = .014$) suggests that Bulgarian hedging rates may be slightly underestimated, attenuating cross-group differences involving Bulgarian. Second, ICLE and LOCNESS argumentative essays may elicit different hedging patterns from the research articles studied in Demir (2018) and Vassileva (2001) and the doctoral dissertations examined by Yıldız and Turan (2021), with reportative subtypes likely differing in citation-heavy genres. Third, the LOCNESS native baseline, while genre-matched to ICLE, is relatively small (175 texts) and was collected under different institutional conditions; results involving the native group should be interpreted with this asymmetry in mind. Partial pooling buffers against extreme estimates from the smallest group

but does not eliminate the sampling-asymmetry concern. Fourth, the cross-sectional design precludes causal inference; observed differences may reflect developmental, instructional, or proficiency factors, though the Turkish–Bulgarian convergence despite different educational systems suggests the typological mechanism does work distinct from any single educational tradition. Finally, the proficiency-confound analysis operationalized proficiency through validated text-internal indices (MTLD, McCarthy & Jarvis, 2010; MLC, Lu, 2010) and the two ICLE metadata fields with usable within-group variance. External measures such as standardized placement-test scores are absent from ICLE, a constraint inherent to corpus-based L2 research. The covariate-adjusted regression and the ZOIB refit provide the strongest available tests under these constraints, and both support the conclusion that the L2–native distancing gap is not explained by proficiency variation as captured by these indicators.

Conclusion

This study tested whether L1 evidentiality systems leave functional traces in L2 English hedging. The traces reside in constructional choices rather than categorical proportions. Turkish and Bulgarian writers converged on a distinctive profile, avoiding named attribution in favor of unnamed, impersonal, or prepositional strategies, consistent with thinking-for-speaking and conceptual transfer operating at a finer grain than category-level predictions suggest. The ghost of L1 evidentiality appears not as a global shift in hedging quantity but as a specific imprint on how writers construct reportative hedges.

Theoretically, the thinking-for-speaking hypothesis (Slobin, 1996, 2016) predicts constructional rather than categorical effects in L2 written discourse, and the present results support that reading. Tanış Şapcı's (2025) study of evidentiality in deception suggests such variation extends beyond academic discourse.

Pedagogically, that all L2 groups produce propositionally shallower hedges than native speakers regardless of L1 suggests EAP instruction should attend to the depth and complexity of hedging constructions, beyond frequency counts. Turkish and Bulgarian writers may benefit from instruction targeting source attribution constructions, deploying named citations alongside the unnamed and impersonal constructions they already favor. The avoidance of named attribution may reflect L1 evidential systems that encode source without specifying identity, so instruction should expand constructional repertoire rather than correct deficiency, consistent with Uzum et al.'s (2025) argument that Turkish English's evidential features represent a legitimate variety.

Future work could extend this constructional reading by testing L1–L2 pairings with other reportative-evidential languages (for example, Quechua, Japanese) and by using production experiments with the Tosun and Filipović (2022) translation paradigm to probe whether the one-to-many mapping problem extends to academic discourse.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Data Availability Statement

The learner corpus data are drawn from the International Corpus of Learner English (ICLE Version 3; Granger et al., 2020) and the Louvain Corpus of Native English Essays (LOCNESS), both licensed from the Center for English Corpus Linguistics, Université catholique de Louvain. Detection pipeline code, configuration files, and aggregated output data are on OSF at https://osf.io/8jq6z/?view_only=6707c20b6f6343a5b44436c55d808e18. Raw corpus texts cannot be redistributed due to licensing restrictions.

AI Disclosure Statement

Large language models were used at three points in the pipeline. Claude Sonnet 4.6 (Anthropic, 2025) classified ambiguous hedging cases Tier 1 could not resolve, a small fraction of detections, and served as the independent reference in Tier 3 validation. Claude Haiku 4.5 (Anthropic, 2025) extracted naked claims from hedged sentences for the propositional distancing analysis.

All LLM outputs were constrained by Aikhenvald's (2004) typology and the six reportative subtypes. No LLM was used in statistical analysis, interpretation, or writing.

References

Aikhenvald, A. Y. (2004). *Evidentiality*. Oxford University Press.

Aikhenvald, A. Y., & Dixon, R. M. W. (Eds.). (2003). *Studies in evidentiality*. John Benjamins.

Akbaş, E. (2014). Are they discussing in the same way? Interactional metadiscourse in Turkish writers' texts. In A. Łyda & K. Warchał (Eds.), *Occupying niches: Interculturality, cross-culturality and aculturality in academic research* (pp. 119-133). Springer. https://doi.org/10.1007/978-3-319-02526-1_8

Aksu-Koç, A., Ögel-Balaban, H., & Alp, I. E. (2009). Evidentials and source knowledge in Turkish. *New Directions for Child and Adolescent Development*, *2009*(125), 13-28. https://doi.org/10.1002/cd.247

Anthropic. (2025). *Claude* (Sonnet 4.6 and Haiku 4.5) [Large language model]. https://www.anthropic.com

Boye, K. (2012). *Epistemic meaning: A crosslinguistic and functional-cognitive study*. De Gruyter Mouton. https://doi.org/10.1515/9783110219036

Çandarlı, D., Bayyurt, Y., & Marti, L. (2015). Authorial presence in L1 and L2 novice academic writing: Cross-linguistic and cross-cultural perspectives. *Journal of English for Academic Purposes*, *20*, 192-202. https://doi.org/10.1016/j.jeap.2015.10.001

Chafe, W. (1986). Evidentiality in English conversation and academic writing. In W. Chafe & J. Nichols (Eds.), *Evidentiality: The linguistic coding of epistemology* (pp. 261-272). Ablex.

Demir, C. (2018). Hedging and academic writing: An analysis of lexical hedges. *Journal of Language and Linguistic Studies*, *14*(4), 74-92.

Faller, M. (2006). *Evidentiality and epistemic modality at the semantics/pragmatics interface* [Conference paper]. Workshop in Philosophy and Linguistics, University of Michigan, Ann Arbor, MI, United States.

Friedman, V. A. (1986). Evidentiality in the Balkans: Bulgarian, Macedonian, and Albanian. In W. Chafe & J. Nichols (Eds.), *Evidentiality: The linguistic coding of epistemology* (pp. 168-187). Ablex.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press. https://doi.org/10.1017/cbo9780511790942

Granger, S., Dupont, M., Meunier, F., Naets, H., & Paquot, M. (2020). *International Corpus of Learner English* (Version 3). Presses Universitaires de Louvain.

He, P., Liu, X., Gao, J., & Chen, W. (2021). DeBERTa: Decoding-enhanced BERT with disentangled attention. In *Proceedings of the 9th International Conference on Learning Representations* (ICLR 2021). https://doi.org/10.48550/arXiv.2006.03654

Hinkel, E. (2005). Hedging, inflating, and persuading in L2 academic writing. *Applied Language Learning*, *15*(1-2), 29-53.

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*(1), 1593-1623.

Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). *spaCy: Industrial-strength natural language processing in Python*. Zenodo. https://doi.org/10.5281/zenodo.1212303

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. Continuum.

Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing*, *6*(2), 183-205. https://doi.org/10.1016/S1060-3743(97)90033-3

Jarvis, S. (2016). Clarifying the scope of conceptual transfer. *Language Learning*, *66*(3), 608-635.

https://doi.org/10.1111/lang.12154

Johanson, L. (2000). Turkic indirectives. In L. Johanson & B. Utas (Eds.), *Evidentials: Turkic, Iranian and neighbouring languages* (pp. 61-87). Mouton de Gruyter. https://doi.org/10.1515/9783110805284.61

Johanson, L. (2003). Evidentiality in Turkic. In A. Y. Aikhenvald & R. M. W. Dixon (Eds.), *Studies in evidentiality* (pp. 273-290). John Benjamins. https://doi.org/10.1075/tsl.54.15joh

Karagjosova, E. (2021). Mirativity and the Bulgarian evidential system. In A. Blümel, J. Gajić, L. Geist, U. Junghanns, & H. Pitsch (Eds.), *Advances in formal Slavic linguistics 2018*. (pp. 133-168). Language Science Press.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174. https://doi.org/10.2307/2529310

Leafgren, J. (2011). *A concise Bulgarian grammar*. SEELRC, Duke University. http://www.seelrc.org:8080/grammar/pdf/stand_alone_bulgarian.pdf

Leclercq, P., & Mélac, E. (2021). Second language acquisition of evidentiality in French and English in a narrative task. *Language, Interaction and Acquisition*, *12*(2), 251-283. https://doi.org/10.1075/lia.20025.lec

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474-496. https://doi.org/10.1075/ijcl.15.4.02lu

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381-392. https://doi.org/10.3758/BRM.42.2.381

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). CRC Press. https://doi.org/10.1201/9780429029608

Mushin, I. (2001). *Evidentiality and epistemological stance: Narrative retelling*. John Benjamins. https://doi.org/10.1075/pbns.87

Odlin, T. (2003). Cross-linguistic influence. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 436-486). Blackwell. https://doi.org/10.1002/9780470756492.ch15

Oh, S. Y. (2007). A corpus-based study of epistemic modality in Korean college students' writings in English. *English Teaching*, *62*(2), 147-175. https://doi.org/10.15858/engtea.62.2.200706.147

Öztürk, O., & Papafragou, A. (2016). The acquisition of evidentiality and source monitoring. *Language Learning and Development*, *12*(2), 199-230. https://doi.org/10.1080/15475441.2015.1024834

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, *2*, e55. https://doi.org/10.7717/peerj-cs.55

Şener, N. (2011). *Semantics and pragmatics of evidentials in Turkish* [Doctoral dissertation, University of Connecticut]. https://digitalcommons.lib.uconn.edu/dissertations/AAI3485418

Šinkūnienė, J. (2008). Autoriaus pozicijos švelninimas: Tarpdalykiniai ir tarpkalbiniai raiškos priemonių ypatumai [Hedging in Lithuanian and English research articles: A cross-disciplinary and cross-linguistic study]. *Kalbotyra*, *58*(3), 97-108. https://doi.org/10.15388/Klbt.2008.7586

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking." In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70-96). Cambridge University Press.

Slobin, D. I. (2016). Thinking for speaking and the construction of evidentiality in language contact. In M. Güven, D. Akar, B. Öztürk, & M. Kelepir (Eds.), *Exploring the Turkish linguistic landscape: Essays in honor of Eser Erguvanlı-Taylan* (pp. 105-120). John Benjamins. https://doi.org/10.1075/slcs.175.07slo

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, *11*(1), 54-71. https://doi.org/10.1037/1082-989X.11.1.54

Sonnenhauser, B. (2013). 'Evidentiality' and point of view in Bulgarian. *Contrastive Linguistics (Съпоставително езикознание), XXXVIII*(2-3), 110-130.

Tanış Şapcı, S. B. (2025). *Evidentiality as a deceptive function: A cross-linguistic study in English, French, Turkish, and Japanese* [Master's thesis, Sabancı University]. https://research.sabanciuniv.edu/id/eprint/53112

Tosun, S., & Filipović, L. (2022). Lost in translation, apparently: Bilingual language processing of evidentiality in a Turkish–English translation and judgment task. *Bilingualism: Language and Cognition*, *25*(5), 739-754. https://doi.org/10.1017/S1366728922000141

Tosun, S., Vaid, J., & Geraci, L. (2013). Does obligatory linguistic marking of source of evidence affect source memory? A Turkish/English investigation. *Journal of Memory and Language*, *69*(2), 121-134. https://doi.org/10.1016/j.jml.2013.03.004

Uysal, H. H. (2008). Tracing the culture behind writing: Rhetorical patterns and bidirectional transfer in L1 and L2 essays of Turkish writers in relation to educational context. *Journal of Second Language Writing*, *17*(3), 183-207. https://doi.org/10.1016/j.jslw.2007.11.003

Uzum, M., Lindahl, K., Uzum, B., Yazan, B., & Akayoglu, S. (2025). Epistemic modality and evidentiality in virtual intercultural exchanges between Turkish and Texan users of English. *World Englishes*, *44*(4), 606-623. https://doi.org/10.1111/weng.12712

Vassileva, I. (2001). Commitment and detachment in English and Bulgarian academic writing. *English for Specific Purposes*, *20*(1), 83-102. https://doi.org/10.1016/S0889-4906(99)00029-0

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413-1432. https://doi.org/10.1007/s11222-016-9696-4

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved R̂ for assessing convergence of MCMC (with discussion). *Bayesian Analysis*, *16*(2), 667-718. https://doi.org/10.1214/20-BA1221

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press. https://doi.org/10.1017/cbo9780511519772

Yıldız, M., & Turan, Ü. D. (2021). Contrastive interlanguage analysis of evidentiality in PhD dissertations. *Discourse and Interaction*, *14*(1), 124-152. https://doi.org/10.5817/DI2021-1-124

Zhang, S., & Luo, S. (2017). A study on conceptual transfer in the use of prepositions in English writing by Chinese secondary school students. *Crossroads: A Journal of English Studies*, *17*(2), 62-75. https://doi.org/10.15290/cr.2017.17.2.04