

A Corpus-Based Account of Evidentiality in Hindko and English

Lubna Ellahi *, Dr. Zafar Iqbal Bhatti**

* Scholar of Linguistics, School of English, Minhaj University Lahore, Pakistan

** Professor of Linguistics, School of English, Minhaj University Lahore, Pakistan

KEYWORDS

*Corpus Linguistics,
English,
Epistemic Stance,
Evidentiality,
Hazara Division,
Comparative Linguistics
Hindko,
Language
Documentation,
Mixed Methods,
Systemic Functional
Linguistics,*

ABSTRACT

This article presents a corpus-based comparative study of evidentiality in Hindko and English. Evidentiality refers to the grammatical and lexical encoding of the source of a speaker's information, indicating whether knowledge is derived from direct observation, logical inference, community hearsay, or a specific reported source. While English primarily expresses evidential meanings through optional lexical items such as *apparently*, *reportedly*, or *I heard*, Hindko encodes evidentiality through more grammatically integrated structures involving verbal morphology, aspectual distinctions, modal particles, and discourse-framing constructions. The study draws on a mixed-method, three-strand corpus comprising 2,000 utterances (1,000 per language) from natural conversation, existing narrative and interview discourse, and controlled elicitation tasks. Evidential expressions are classified into four categories: direct, inferential, reportative, and quotative. The findings demonstrate that evidentiality is more grammatically prominent and more discursively frequent in Hindko than in English; that direct evidentiality shows greatest cross-linguistic similarities; and inferential and reportative evidentiality show the largest cross-linguistic differences; and that evidential marking in Hindko carries significant cultural functions related to oral tradition, epistemic authority, and communal knowledge transmission. The article situates these findings within the frameworks of corpus linguistics, typological theory, and Systemic Functional Linguistics, and discusses implications for language documentation, translation practice, and research methodology.

1. Introduction

Communication is not merely about describing the world; speakers do not simply report the facts, they simultaneously signal how they know what they claim about to know. A speaker says *'He left early'* is making a simple statement; one who says *'Apparently he left early'* is describing that the evidence comes from an external source not based on someone's personal reflection. This apparently small difference between observed knowledge and received or indirect information is encoded in a different way across in the world's languages, and in the etymological study, this occurrence is known as evidentiality. Evidentiality has been comprehensively investigated in typological research since the initial work of Chafe (1986), Willett (1988), and Aikhenvald (2004). These studies have established that some languages grammatically involve speakers to gain information source every time they make a statement, while in other languages including English leave a pattern

optional and verbal. Among these two situations lies a variety of transitional systems in which evidential implications are moderately become more grammatical, relating with modality features, and stressed in language-specific techniques. The existing research focuses on Hindko as its primary language of analysis, compared with English. Hindko is an Indo-Aryan language of the northern areas of Punjab gathering spoken online data mainly across the Hazara Division Abbottabad, Mansehra, Haripur, Battagram, and nearby areas as well as in Attock and different parts of Azad Kashmir. Five to seven million valued speaker of Hindko (Rensch, Hallberg, & Radloff, 1992; updated estimates), it is demographically important. Despite its demographic significance, Hindko remains largely un described in the typological and grammatical literature. No systematic grammar description of its types, including evidentiality, aspect-mood contact, and speech structure is largely inattentive in the published record.

Title: *A Corpus-Based Account of Evidentiality in Hindko and English*.....

Author: *Lubna Ellahi, Dr. Zafar Iqbal Bhatti*

This article draws composed academic, analytical, and organizational judgments of the study into a particular combined description. It describes three research questions:

Q1: What evidential categories are grammatically and lexically expressed in Hindko and English languages, and what linguistic structures are used to express each category in the two languages?

Q2: How these evidential categories are represented across the corpus data, and what register-based patterns appear in each language?

Q3: How do register shape evidential designs in each language and what are the comprehensive implications of these findings?

The following sections addressed the theoretical background, methodology, findings, and implications of the study.

2. Theoretical Framework

2.1 Evidentiality: Core Definitions and Typological Scope

Evidentiality, as a linguistic category, argues to the structural or functional coding of information sources. The most comprehensive typological explanation remains Aikhenvald (2004), whose analysis on 500 languages recognized the foundational grouping that used in the existing study. Aikhenvald separates grammatical evidentiality in which devoted morphological indicators obligatorily encode data as part of the verb model and lexical evidentiality, in which central meaning is conveyed through possible free items such as adverbs or recording structures. Languages such as Turkish, Tibetan, Quechua, and Bulgarian belong to the former types; English belongs to the latter. Within grammatical evidentiality methods, group series from modest two-way direct/indirect differences to elaborate arrangements unique visuals from non-visual direct evidence, and result-based on reasoning-based suggestion. The four-category arrangement used in this work direct (DIR), inferential (INF), reportative (REP), and quotative (QUOT), keep

an eye on Aikhenvald's typology at an transitional level of delicacy, suitable for cross-linguistic comparison stuck between an incompletely grammatical system (Hindko) and a lexical one (English).

A critical theoretical change separates evidentiality from epistemic modality. Epistemic modality concerns the speaker's degree of inevitability (how definite they are). Evidentiality concerns with the knowledge where it comes (how they came to know). As De Haan (1999) maintains, conflating the two aspects theoretically disorganized records, since a speaker may be entirely certain of something they only heard from the surrounding, or indefinite about something they saw directly. The present work deals with the two systems as individually but concerning to each other, particularly in Hindko, where epistemic modals and evidential significances often co-occur in the same linguistic features.

2.2 Evidentiality in Systemic Functional Linguistics

The Systemic Functional Linguistics (SFL) presented by Halliday (1994) and Matthiessen (2014) offers discourse-functional elements in the analytical context. In SFL, language works in three immediate meanings: the ideational (encoding knowledge), the textual (organizing dialogue), and the interactive functions that are having most traditional consequence. The interpersonal metafunction talks about how speakers make social interactions, direct attitudes, and position themselves epistemic ally in following relations to the objectives they produce. Evidentiality, in SFL terms, is a resource of the interpersonal metafunction: when a speaker uses a reported evidential, they are not only attaching information to a settings but managing their epistemic responsibility arranging themselves as a receiver rather than an discoverer of knowledge. Martin and White's (2005) Review structure is also significant; particularly do its procedure, which discovers how speakers show the position of their vocabularies in different

Title: *A Corpus-Based Account of Evidentiality in Hindko and English*.....

Author: *Lubna Ellahi, Dr. Zafar Iqbal Bhatti*

situations and assessments. Reportative and quotative evidentiality exemplify forms of heteroglossic arrangement, the categorical obligation that a suggestion comes from outside the speaker's own scope.

2.3 Evidentiality in South Asian Languages

Evidentiality is an area of developing research importance in South Asian Indo-Aryan languages. Hindi-Urdu has established the most devotional, with analysts recording that the perfective verb form in definite dialogue backgrounds carries a non-first-hand evidential analysis the speaker claims that they did not directly observe the event described (Bhatt, 2005; Lazard, 2001). Punjabi, the closest qualified of Hindko, exhibits similar properties (Bhatia, 1993). Languages of the northwest area Khowar, Shina, and Burushaski have been exposed to possess more elaborate evidential morphology (Bashir, 1988; Berger, 1998), proposing that grammatical evidentiality may be a local characteristics of the northwest South Asian linguistic region. The absence of different features like systematic evidentiality explanation for Hindko signifies a gap in this regional representation that the current study begins to complete.

3. Methodology

3.1 Research Design: Mixed Methods

The study proceeds with a mixed-method methodology, collaborating quantitative corpus work with qualitative discourse-functional analysis. The design of the study is not finding the middle ground but just react to the environment of evidentiality in a logical purpose. Quantitative analysis creates distributional preparations: which evidential varieties take place in which situation, how frequently, in which record, and whether cross-linguistic differences are statistically significant or not. Qualitative analysis perceives how evidential markers work together in real dialogue: how they manage speaker's position, build epistemic power, and transfer cultural meaning in the strict background

of Hindko oral discussions. Neither the linguistic elements alone would be satisfactory; their grouping delivers both experimental foundation and informational complication.

The epistemological training of the research is based on critical pragmatism (Bhaskar, 1979): the circumstances that evidentiality is a fixed structural-functional aspect of languages, not only an analytical satiability, but one whose reaction is depending on the specific grammatical and traditional description of each language. This position inspires the qualified design comparing Hindko and English highlights what evidentiality is, accurately because the two languages have linguistically in it changed customs.

3.2 Corpus Design: Three Strands

The corpus comprises 2,000 utterances in total 1,000 per language structured across three equivalent features. Table 1 presents the corpus architecture.

Strand	Data Type	Size (per language)	Primary Purpose
Strand 1	Natural conversation from online sources corpus	330 utterances	Ecological validity; naturalistic evidential frequency
Strand 2	Existing text and discourse samples	430 utterances	Register and genre diversity: narrative, interview, written
Strand 3	Controlled elicitation tasks	240 utterances	Direct probing of grammatical evidential contrasts
Total	Three-strand integrated corpus	1,000 utterances	Quantitative + qualitative analysis

Table 1. Corpus composition by strand and analytical purpose.

Hindko data (Strand 1) were gathered entirely through online and digital sources, three primary digital channels: (i) pre-existing WhatsApp and Telegram voice message chats shared voluntarily by native speakers, in which the researchers were not a participant; (ii) informal YouTube videos including vlogs, family dialogues, and public uploads produced by Hindko narrators in the Hazara Division; and (iii) Facebook Live recordings and social media video content featuring unscripted communication among native speakers. These sources were selected because they represent unmonitored, natural speech produced without researcher presence, so

Title: *A Corpus-Based Account of Evidentiality in Hindko and English*.....

Author: *Lubna Ellahi, Dr. Zafar Iqbal Bhatti*

satisfying natural validity requirements (Hundt, Nesselhauf & Biewer, 2007; Tagliamonte & Denis, 2008; Eisenstein, 2013). The English informal data were drawn from the British National Corpus (BNC) casual sub corpus and the ICE-GB spoken element.

Strand 2 prolonged register analysis by joining oral narrative records, sociolinguistic conversation transcripts, and in written style. Hindko data were collected from SIL Pakistan records, publicly available Hazara University ethnographic recordings, and existing digital discourse samples accessible through online platforms. English data was collected from BNC narrative and interview sub corpora and UK newspaper recording. Strand 3 involved three organized elicitation tasks: a picture-based construction task using visual motivations considered to produce each evidential group; a contentment judgment task in which contributors evaluated evidential forms in context; and a translation task in which participants reduced English evidential vocabularies into Hindko. Tasks were controlled to twelve adult native speakers of each language.

3.3 Annotation and Analysis

Evidential terminologies were identified and explained using a dual-criterion methodology: a communication is coded as evidential merely if it fulfills both a formal standard (it belongs to a morphological, syntactic, or lexical arrangement with evidential prospective) and a practical standard (in its definite discourse setting, it encodes information about the cause of the speaker's knowledge). These dual standard works against both over-annotation and under-annotation. A secondary explanation tier recorded whether each evidential communication was grammatically combined (G) or lexically realized (L), allowing analysis of the important depth of evidential coding in each language. Inter-rater reliability was evaluated through independent re-annotation of a 20% random sample of the corpus by a second proficient annotator. Cohen's kappa values of $\kappa =$

0.81 (Hindko) and $\kappa = 0.84$ (English) indicating near-perfect agreement (Landis&Koch, 1977) confirming the reliability of the annotation system. Quantitative study used frequency totals, proportional distribution per 1,000 utterances, chi-square tests (χ^2), and Cramér's V for effect size. Qualitative analysis used AntConc concordance for English and custom Python scripts for Hindko, added by close speech analysis of complete description and conversational constructions.

4. Evidentiality in English: Structures and Distribution

4.1 The Structural Status of English Evidentiality

English lacks grammatical evidentiality in the plain typological sense: no English morpheme, verb completion, or obligatory syntactic production is committed completely to encoding data source. English evidentiality is lexical that able through words and expressions. This means that an English narrator can make a direct speech, 'The Bridge collapsed' without gesturing everything about how they know, and no grammatical devastation results. If they understand to sign source, they must do so observably through added portion. In the absence of an evidential indicator, English observers understand a statement as inferring direct knowledge by defaulting. The arrangement of an evidential indicator 'Apparently the bridge collapsed' signals a thoughtful disappearance from that avoidance and marks the speaker as division themselves from full epistemic obligation. This irregularity between marked and unmarked is important to how evidentiality meanings in English and contrasts notably with Hindko, where evidential differences are more intensely constructed into the grammar.

4.2 Evidential Resources in English

The most productive evidential possessions in English are epistemic-evidential adverbs: apparently, evidently, reportedly, ostensibly, presumably, supposedly. These encode changed

Title: *A Corpus-Based Account of Evidentiality in Hindko and English*.....

Author: Lubna Ellahi, Dr. Zafar Iqbal Bhatti

combinations of source form and epistemic stand. Apparently typically indicators indirect knowledge suggestion or report with mild narrator distancing. Reportedly and reputedly are definitely reparative, credits to information to indefinite external sources. Evidently rests toward inferential importance. In corpus data, these adverbs are strongly register-stratified: they are noticeably more frequent in reporting and academic records than in casual conversation.

(1) *The minister reportedly flew to Geneva for emergency talks.*

'Attribution to unspecified journalistic or public source (reportative evidential')

(2) *Apparently, the negotiations collapsed over the weekend.*

'Indirect, unspecified-source knowledge (reportative/inferential')

Parenthetical and reporting structures, I heard that, they say, allowing to X, it is said that exemplify a second major means, covering reportative and quotative meanings respectively. These productions trait information to public or specific sources and concurrently signal the speaker's epistemic space from the proposition. The passive production it has been reported that is mainly frequent in formal transcribed English.

Modal verbs work inferential evidential meanings in English: epistemic must translates the speaker's thought from existing evidence to a decision ('the keys must be in the car, I didn't bring them inside') as do might and could for undefined implications. These modal procedures constitute the closest English in the same way to grammaticalised inferential evidentiality, however they simultaneously encode epistemic confidence in a way that pure inferential markers in other languages do not.

4.3 Register Distribution in English

Corpus investigation of the English data approves that evidential pattern is strongly register-dependent. In conversational documents,

categorical evidential words are relatively uncommon: speakers assert directly and rely on shared context rather than evident source acknowledgments. In narrative data, reportative and quotative constructions increase significantly, reflecting the genre's condition to relay others' speech and actions. Interview data shows the highest occurrence of inferential constructions, as speakers estimate and interpret rather than simply explosion. This register-sensitivity means that the complete frequency of evidential marking in English is significantly lower than in Hindko across the corpus, particularly in the conversational strand.

5. Evidentiality in Hindko: Structures and Distribution

5.1 Grammatical Evidentiality in Hindko

Unlike English, Hindko encodes evidential differences through grammatical arrangements that are part of the language's systematic oral structure. The most significant of these operates through the communication of perfective aspect with speech background. In Hindko, as in Hindi-Urdu and Punjabi, the perfective verb form in transitive divisions can carry a non-first-hand evidential impression when used to report events the speaker did not individually witness. This reading stand up from the contact of feature and information source: the same procedure that marks a completed experience in a witnessed environment takes on an incidental evidential suggestion (inference or report) when the dialogue context makes direct observation questionable. This evidential use of the perfective is not grammatically mandatory in all contexts, assigning Hindko at an intermediate place on the grammaticalization cline (more grammaticalized) than English but not as much of so than other languages. However, native speakers consultant to the regular recognition between direct and indirect impressions in elicitation, approving that the evidential differences are part of useful grammatical know-how in Hindko

5.2 Direct Evidentiality

Direct evidentiality in Hindko is determined through the perfective features that perform by first-person or clearly observed in communication contexts. Unlike languages with keen direct evidential morphemes, Hindko marks observed data through the perfect default of the perfective verb form the equal form that converts a completed event in a witnessed situation shifts to a non-first-hand (inferential or reportative) reading when the speech context reduces direct observation incredible. Hindko relies on perception verbs 'dittha' saw, 'sunia' heard, like 'mai barish hondi ditthi' (direct visual evidence), This related understanding of the perfective is a significant feature of the Hindko evidential system and is reliable with the pattern documented for the thoroughly related languages Hindi-Urdu (Bhatt, 2005) and Punjabi (Bhatia, 1993). In the corpus, direct evidential tokens justify for the majority of perfective uses in Strand 1 conversational data, where speakers report events they personally experienced.

5.3 Inferential Evidentiality

Inferential evidentiality marks knowledge that derived reasoning from noticeable conditions is expressed in Hindko through the element *lagda hai / lagdi aay* ('it seems / it appears') and through the reasonable auxiliu`ary "hoga", which codes reasoning about possible states of concerns. These constructions are considerably more grammatically combined and more frequently set up in Hindko discourse than their English functional equals.

(3) *Baarish hoyi lagdi aay.*

'It appears to have rained / It seems that it rained (inferential evidential from visual evidence')

(4) *Ohna'n noo khabar hogi.*

'They would know / they must know (presumptive inference from reasoning)'

Corpus data shows inferential evidential patterns occurring at approximately 2.3 times the amount in Hindko as in English when calculated proportionally through the full corpus, a change

that is statistically significant ($X^2 = 18.4$, $df = 1$, $p < .001$, Cramér's $V = 0.21$). This result is regular across all three corpus strands and is strongest in the conversational data and in the register where English inferential design is most compact.

5.4 Reportative Evidentiality and Oral Tradition

Reportative evidentiality marking information from indefinite common or social sources is among the maximum socially important evidential categories in Hindko. The major reportative markers are the productions *kahndey ne / kahndi aay* ('they say / it is said') and *suniya aay / sunya si* ('it is heard / it was heard'). These structures frame information as mutually received or orally conveyed rather than personally observed, and they carry practical weight beyond their grammatical function.

(5) *Kahndey ne ke pehla'n is darya'n cho paar nahi si.*

'They say that in earlier times there was no crossing across this river (reportative, communal oral tradition)'

(6) *Sunya si ke oh waqt aasman te nishaan aaya si.*

'It was heard that at that time a sign appeared in the sky (reportative, transmitted cultural knowledge)'

The frequency of reportative pattern in Hindko narrative data is significantly higher than in English narrative data. This dissimilarity reflects the oral convention of the Hazara region, in which communal knowledge, local history, and cultural reminiscence are conducted through chains of reporting rather than written documentation. When a Hindko speaker uses "kahndey ne"/*dasdny*, they are not just giving credit to information to an undetermined source; they are indexing a public epistemic power, the combined voice of the community as an information source. This social-indexical meaning of reportative evidentiality in Hindko has no direct corresponding in English. In

Title: *A Corpus-Based Account of Evidentiality in Hindko and English*.....

Author: Lubna Ellahi, Dr. Zafar Iqbal Bhatti

Systemic Functional Linguistics terms, this social-indexical function expresses heteroglossic extension (Martin & White, 2005): the speaker obviously acknowledges that the plan derives from outside their own epistemic field, arranging themselves as a channel of communal knowledge rather than its originator.

5.5 Quotative Evidentiality

Quotative evidentiality in Hindko is carried primarily through the verb *keha* ('said / told') in direct and indirect described speech constructions. What divorces Hindko quotative evidentiality from simple conveyed communication, however, is the availability of additional epistemic view marks in the fixed clause through modal works (*lagda* 'apparently', *shayad* 'perhaps') or the subjunctive mood signifying not only the attribution of data but it based on speaker's evaluation in its consistency. This stance-encoding layer means that Hindko quotative productions do more than report; they also point the speaker's degree of conviction in the reported data.

(7) *Ahmad nay keha ke ohna de nal koi v nahi aaya.*

'Ahmad said that no one came with them — quotative evidential with specific named source'

6. Comparative Analysis: Key Findings

6.1 Grammaticalisation vs. Lexicalisation

The most essential structural difference between the two languages is the point to which evidentiality is grammaticalized. In English, evidential implications are realized through syntactically determined semantically compositional lexical items whose evidential reading get up from the arrangement of their overall lexical meaning and pragmatic situation. In Hindko, evidential distinctions are encoded through spoken morphology, aspect-agreement contacts, and speech particles that form part of the regular grammatical construction of the language. Table 2 summarizes the structural contrast.

Feature	Hindko	English
Grammatical status	Partially grammaticalised — verbal aspect, particles, modals	Lexical — adverbs, reporting constructions, modal verbs
Obligatoriness	Contextually required in indirect-knowledge contexts	Entirely optional; default is direct unmarked assertion
Register sensitivity	Present across all registers including casual conversation	Strongly register-stratified; low in casual conversation
Direct evidentiality	Unmarked default; perfective can imply indirect in context	Unmarked default; no specific form
Inferential marking	<i>lagda hai, hoga</i> — grammatically integrated, frequent	Epistemic must, it seems — optional, register-sensitive
Reportative marking	<i>kahndey ne, sunya si</i> — grammaticalised, culturally weighted	they say, apparently, reportedly — lexical, formal register
Quotative marking	<i>keha ke</i> + stance encoding via modal/subjunctive	X said that — no systematic stance encoding
Epistemic stance layer	Available through morphology and modal particles	Achieved through separate hedging vocabulary or intonation

Table 2. Structural comparison of evidentiality in Hindko and English

6.2 Frequency and Distribution

Across all corpus elements, evidential marking is significantly more regular in Hindko than in English.

Evidential Category	Hindko N (%)	English N (%)	χ^2	p-value	Remarks
Direct (DIR)	413 (41.3%)	253 (25.3%)	72.8	< .001	Significant
Inferential (INF)	273 (27.3%)	167 (16.7%)	42.6	< .001	Significant
Reportative (REP)	213 (21.3%)	347 (34.7%)	66.1	< .001	Significant
Quotative (QUOT)	100 (10.0%)	233 (23.3%)	95.7	< .001	Significant
TOTAL	1000 (100%)	1000 (100%)	—	—	—

Table3. Frequency distribution of evidential categories in Hindko and English (N = 1,000 per language). *** $p < .001$.

Reportative and inferential categories show the largest complete differences; quotative evidentiality shows more equivalent rates in narrative and interview registers, reflecting shared variety agreements for attribution in both languages. The distributional pattern in English is strongly register-stratified: evidential marking rises substantially from conversational to narrative to interview registers, even though in Hindko evidential distinctions are active across all three registers, with the greatest absolute frequency in narrative data reflecting the

prominence of oral tradition in the Hindko-speaking community.

6.3 Epistemic Stance and Speaker Positioning

At the functional level, the most important comparative finding suffers how the two systems manage epistemic attitude of the speaker arranging by themselves in relative to the suggestions that they created. In Hindko, the ironic grammatical possessions for evidentiality enable speakers to sign their epistemic position as a repetitive part of expression structure.

In English, by contrast, the options of evidential marking create a system in which the unmarked report carries a strong outcome on direct, reliable knowledge. When an English speaker produces an evidential marker '*Apparently*', '*I heard that*', '*it seems*' '*lagda aey*' this selection transfer a pragmatic load as sympathetic prevention or separation change. The two systems therefore are changed not only in the regularity of evidential marking but in the retiring meaning of its absence: unmarked statement recommends direct knowledge in English to a position that it does not in Hindko, where indirect evidence commonly goes marked as a position of predictable grammar. This difference has real communicative consequences. In different circumstances it demanding accurate representation of knowledge comes from legal proof, medical history, journalistic attribution, oral history documentation is mismatch between Hindko's grammatically prominent evidential system and English's optional lexical one creates explanation challenges that are not merely formal but epistemic.

7. Implications

7.1 Language Documentation

The findings contribute to the documentary record of Hindko grammar in the field that has not earlier received logical analysis. Hindko is spoken by 5 to 7 million people mainly living in the Hazara division, Pakistan but it is has a great absence of typological records, grammatical

reference works, and the graphic linguistics writings. By providing a cross linguistic analysis, this study begins to document the grammatical features of Hindko that have been largely absent from the typological record. The categories defined here may be imperfectly transmitted to younger speakers with decrease exposure to formal narrative and oral convention records.

7.2 Translation and Cross-Cultural Communication

The structural irregularity between the two systems creates transformation challenges that go out there word-for-word conversion. When explaining from Hindko into English, evidential differences that are grammatically encoded in the source must be reduced through thoughtful lexical choices in the object choices that separate translators may make incompatibly, resulting in versions that change in their representation of the speaker's epistemic stance. In permitted and reporting contexts, this matters: a statement recognized to '*what is said in the community*' (*kahndey ne*) is gradually changed from a direct statement, and that distinction must be preserved in translation. Translator training in cross-linguistic evidential planning is a practical implication of these findings.

7.3 Methodological Contribution

The three-strand corpus design established for this study merging naturalistic conversation, existing speech samples, and controlled elicitation marks addresses a challenge that performs in many studies of under-documented languages: no particular data type offers acceptable analysis of all the evidential phenomena of consideration. The design is convenient to equal studies of other under-described local dialects, including Khowar, Torwali, Shina, and other of the northwest South Asian linguistic areas, where limited evidential structures have been recognized but not fully well-defined. The dual-criterion annotation design, the inter-rater consistency procedures, and the record conventions developed for Hindko

Title: *A Corpus-Based Account of Evidentiality in Hindko and English*.....

Author: *Lubna Ellahi, Dr. Zafar Iqbal Bhatti*

are all documented in sufficient detail for adaptation and replication.

8. Conclusion

This article has presented an integrated analysis of evidentiality in Hindko and English, synthesizing the theoretical, methodological, and empirical dimensions of a corpus-based comparative study. The central finding is that evidentiality operates as a significant communicative system in both languages but is realized through structurally different means with different frequencies, different register distributions, and different functional profiles. This work identified four evidential categories, direct, inferential, reportative, and quotative and revealed clear cross-linguistic differences in how they are expressed in Hindko and English. Hindko expresses evidentiality through grammatically integrated structures; English primarily depend on possible lexical devices. The analysis further revealed that direct evidentiality displayed the extreme cross-linguistic similarity, with both languages preferring firsthand evidence, though Hindko stated it more analytically through grammaticalized productions. Inferential and reportative evidentiality reveals the maximum deviation between the two languages, whereas quotative evidentiality shows comparatively similar patterns in descriptions. Register also shapes evidential appearance irregularly. Hindko works evidential dissimilarities across all communicative settings, including spontaneous conversation, while English focuses on evidential markers in reporting and formal written speeches. These findings have significant effects on language documentation, translation practices, and the typological understanding of grammaticalization.

Direct evidentiality displayed the greatest cross-linguistic similarity, with both languages favoring firsthand evidence, though Hindko expressed it more systematically through grammaticalized constructions

Hindko encodes evidentiality through grammatically embedded verbal and particle constructions that are active across all registers of speech and carry cultural weight related to oral tradition and communal authority. English encodes evidential meanings through optional lexical items whose deployment is register-sensitive and pragmatically marked. The result is a system in which Hindko speakers can signal epistemic positioning as a routine part of utterance construction, while English speakers must make deliberate choices to mark information source choices that carry stronger pragmatic implications precisely because they are not the default.

These differences have implications for language documentation, for translation theory and practice, and for the broader theoretical understanding of how evidentiality functions as a grammatical and discourse category. Hindko is a language of considerable demographic weight whose grammatical richness has been largely overlooked in the linguistics literature. The present study argues that this oversight should be corrected both as a matter of scholarly completeness and as a matter of cultural responsibility toward the communities whose language and knowledge practices the evidential system reflects.

Title: *A Corpus-Based Account of Evidentiality in Hindko and English*.....

Author: *Lubna Ellahi, Dr. Zafar Iqbal Bhatti*

REFERENCES

- Anthony, L. (2022). Ant Conc (Version 4.2.0) [Computer software]. Waseda University.
- Bashir, E. (1988). Inferentiality in Kalasha and Khowar. *Proceedings of the Chicago Linguistic Society*, 24, 25–37.
- Berger, H. (1998). *Die Burushaski-Sprache von Hunza und Nager*. Harrassowitz.
- Bhaskar, R. (1979). *The possibility of naturalism*. Harvester Press.
- Bhatt, R. (2005). Long distance agreement in Hindi-Urdu. *Natural Language and Linguistic Theory*, 23(4), 757–807.
- Bhatia, T. K. (1993). *Punjabi: A cognitive-descriptive grammar*. Routledge.
- 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.
- Chafe, W. (1994). *Discourse, consciousness, and time*. University of Chicago Press.
- De Haan, F. (1999). Evidentiality and epistemic modality: Setting boundaries. *Southwest Journal of Linguistics*, 18(1), 83–101.
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). Edward Arnold.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). Routledge.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Lazard, G. (2001). On the grammaticalization of evidentiality. *Journal of Pragmatics*, 33(3), 359–367.
- Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan.
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Nuyts, J. (2001). *Epistemic modality, language, and conceptualization*. John Benjamins.
- Rensch, C. R., Hallberg, C. F., & Radloff, C. L. (1992). *Sociolinguistic survey of Northern Pakistan: Vol. 2*. National Institute of Pakistan Studies / Summer Institute of Linguistics.
- Willett, T. (1988). A cross-linguistic survey of the grammaticalization of evidentiality. *Studies in Language*, 12(1), 51–97.
- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243–257.

Title: *A Corpus-Based Account of Evidentiality in Hindko and English*.....

Author: *Lubna Ellahi, Dr. Zafar Iqbal Bhatti*

Eisenstein, J. (2013). What to do about bad language on the internet. Proceedings of NAACL-HLT, 359–369.

Hundt, M., Nesselhauf, N., & Biewer, C. (Eds.). (2007). *Corpus linguistics and the web*. Rodopi.

McEnery, T., & Wilson, A. (2001). *Corpus linguistics* (2nd ed.). Edinburgh University Press.

Tagliamonte, S. A., & Denis, D. (2008). Linguistic ruin? LOL! Instant messaging and teen language. *American Speech*, 83(1), 3–34.