

RESEARCH JOURNAL  
**TRANSLINGUA**

ISSUE 1 | VOLUME 1 | 2025





## **TRANSLINGUA, 2025**

### **Volume 1, Issue 1**

#### **Editor's Message**

It gives us immense pleasure to see the dream materializing in the form of Translingua. It was much needed in Pakistan to initiate research on languages from multidimensional perspectives, integrating technology, artificial intelligence, and other fast-growing disciplines. Pakistan presents a rich linguistic tapestry with more than 70 languages in use. I welcome researchers across the globe to make contributions to the journal in the form of their advice, publications, and insights.

Let us collaborate to establish a platform for rich linguistic research where the voices from indigenous language communities are heard.

Editorial team,  
TRANSLINGUA

Centre for Languages and Translation Studies (CeLTS),  
Allama Iqbal Open University, H-8 Sector, Islamabad



## TRANSLINGUA, 2025

### Editorial Team

#### Editor in Chief

- Dr Ghulam Ali, Director, CeLTS, AIOU

#### Editor

- Prof Dr Abdul Qadir Khan, Professor, CeLTS, AIOU

#### Managing Editor

- Dr Furrakh Abbas, Assistant Professor, CeLTS, AIOU

#### Editorial Team Members

- Dr Lubna Umar, Assistant Professor, CeLTS, AIOU
- Ansar Ali, Lecturer, CeLTS, AIOU

### Editorial Board

- Prof Dr Nabila Rehman, Principal, Oriental College, University of the Punjab, Lahore.
- Prof Dr Mazhar Hayat, Riphah International University, Islamabad (Faisalabad Campus)
- Dr Umaima Kamran, Associate Professor, Quaid e Azam University, Islamabad
- Dr Muhammad Taimoor Khan, National University of Computer and Emerging Sciences
- Prof Dr Arshad Ali Khan, University of Management and Technology, Lahore
- Dr Ghulam Ali Buriro, University of Sindh, Jamshoro

### Advisory Board

- Prof Dr Luise Von Flowto, University of Ottawa, Canada
- Prof Dr Karolina Ruta, Adam Mickiewics University, Poland
- Prof Dr Sean Pue, Michigan State University, USA
- Prof Dr Henrik Liljegren, Stockholm University, Sweden
- Dr Debbie Baghali, Griffith University, Australia
- Dr Himli Hamzah, Universiti Utara Malaysia
- Dr Bal Krishna, Kathmandu University, Nepal
- Prof Dr Ayesha Sohail, AJK University, Muzaffarabad
- Prof Dr Shahid Siddiqui, Lahore School of Economics
- Prof Dr Aftab, Director General, PAAS, Peshawar
- Prof Dr Mazhar Saleem, DG, National Language Promotion Authority, Islamabad
- Prof Dr Muhammad Iqbal, Grand Asian University Sialkot



## TRANSLINGUA, 2025

### Volume 1, Issue 1

#### Table of content

S #	Title	Authors	Page #
1	A Comparative Discourse Analysis of AI-Generated vs. Human-Produced News Reports: Evaluating Objectivity, Framing, and Ideology	Amjad Naveed, Insha Ullah	1-23
2	Evaluating the performance of ChatGPT, Claude, and Grok on Urdu Text Generation and Language Processing: A Comparative Analysis	Fariha Saghir, Maria Sundas	24-35
3	Role of Cohesive Devices in English Reading Exercises of IELTS Examination	Ali Furqan Syed	36-69
4	Critical Discourse Analysis of Pakistani War Songs during the 1965 Pak-India War	Sadia Shabbir	70-78
5	Beyond Literal Meaning: A Pragmatic Analysis of Balti Proverbs	Ansar Ali	79-89

Published by:

Centre for Languages and Translation Studies (CeLTS),

Allama Iqbal Open University,

H-8 Sector, Islamabad



## **A Comparative Discourse Analysis of AI-Generated vs. Human-Produced News Reports: Evaluating Objectivity, Framing, and Ideology**

Author/s: Amjad Naveed<sup>1</sup>, Insha Ullah<sup>2</sup>

Affiliation: Department of English and Applied Linguistics, University of Lakki Marwat (ULM), Email:  
[amjadnaveed4star@gmail.com](mailto:amjadnaveed4star@gmail.com), [ullahinsha262@gmail.com](mailto:ullahinsha262@gmail.com)

**ABSTRACT:** As large language models enter newsrooms, this study compares AI-generated and human-produced news on linguistic construction, framing, and objectivity. We analyzed N = 360 English-language reports (AI = 180; Human = 180) across politics, economy, science/technology, and health in the US, UK, and Pakistan (Jan–Jun 2025), combining CDA/SFL-based coding with computational indicators (lexical diversity/MTLD, syntactic complexity/MLC, readability, subjectivity/sentiment; quotation density, source diversity, reporting verbs). Tests include Welch's t with Benjamini–Hochberg correction, mixed-effects models (Origin fixed; Outlet random; Beat/Country controls), binomial GLMs for frames, an Objectivity Index, and an exploratory classifier. AI texts use more modality/hedging and slightly more passive voice but lower lexical diversity and syntactic complexity. The largest gaps are in sourcing (fewer quotes, sources, reporting verbs), yielding a lower Objectivity Index for AI. Human-Interest framing is more common in human stories, while AI tilts toward Economic Consequences. A classifier distinguishes origin with AUC  $\approx$  0.96. AI and human news form distinct discourse profiles; pairing AI drafting with sourcing protocols, disclosure, and frame-aware editing is recommended to meet journalistic standards.

Keywords: Framing, Ideology, Objectivity, Comparative Discourse Analysis, News Reports

To cite: Naveed, A. & Ullah, I (2025). A Comparative Discourse Analysis of AI-Generated vs. Human-Produced News Reports: Evaluating Objectivity, Framing, and Ideology. *TRANSLINGUA*, 1(1), pages 1-23.

## **1. INTRODUCTION**

The emergence of generative artificial intelligence (AI) into news-making has intensified ancient discussions on how news language produces social reality and how much notions of objectivity, balance, and ideological neutrality are being operationalized or perverted in practice. Critical Discourse Analysis (CDA) conceives of news not as a transparent mirror of events but as a patterned form of public discourse, wherein linguistic choices (lexis, syntax, modality, transitivity) contribute to the framing of events and the positioning of audiences (Fairclough, 1995; van Dijk, 1988). In this school of thought, and based on Systemic Functional Linguistics (SFL), grammar of news is approached. Transitivity patterns that define actors and their actions, modality, which indicates commitment and evaluative attitude, and choices of lexicon, which define authorial position are used to operationalise meaning making in discourse (Halliday and Matthiessen, 2014; Fowler, 1991). Parallel framing theory explains the ways in which texts construct issues, diagnose, make moral judgments and mobilize remedy proposals (Entman, 1993; Gamson and Modigliani, 1989). Altogether, these theoretical approaches provide a strong set of prisms through which the similarities and differences between news created by AI and humans are interrogated.

The automation in journalism has been applied traditionally, but the range and the complexity of automation have grown significantly. Early robot journalism systems took structured data (sports statistics and corporate profits) and translated it into templated writing on a scale (Graefe, 2016; Dörr, 2016). Practitioner interviews and case studies suggest that automation can increase the speed and depth of coverage, but at the same time is generating transparency, journalistic judgment, and ethical responsibility problems (Diakopoulos, 2019; Thurman et al., 2017). The pre-LLM experimental and perceptual research-based evidence showed mixed responses of the audience; in certain situations, readers found the automated stories as informative and readable as human-created stories, although less exciting (Clerwall, 2014). A recently released meta-analysis has found that perceived quality depends on topical and contextual considerations, but the general acceptance of automated journalism by the audience is high (Graefe and Bohlken, 2020). However, with the recent development of large language models (LLMs) it has been possible to generate fluent, context-specific prose that goes far beyond the paradigms of data-to-text and makes a modern comparative discourse analysis not only timely but also urgent.

Located in the framework of Critical Discourse Analysis (CDA) and Systemic Functional Linguistics (SFL), the main question is how AI and human texts are different in the process of the linguistic construction of events. The early studies of news language show that lexical options, appraisal processes and clause structures are systematic encoding of ideological stances such as foregrounding this actor, backgrounding others, modulating claim, and normalising particular interpretation (Bell, 1991; Fowler, 1991; Halliday and Matthiessen, 2014; van Dijk, 1988). Template / prompt-based generation in automated texts can strengthen particular transitivity patterns, like tendency to agentless passives or nominalisations in general, and modal strategies typified by scrupulous hedges or default indicative. Big pretrained models, in contrast, can import the training learned stylistic priors (Recasens et al., 2013). The analysis of biased language through computational approaches has shown patterns

of subjective lexis, framing nouns, and markers of epistemic that can be measured quantitatively across corpora (Recasens et al., 2013; Wilson et al., 2005).

This comparative endeavour is further inspired by the normative ideal of objectivity. Objectivity historically became a professional standard during the transformation of the political economy and the transformations of the newsroom practices (Schudson, 1978), and is still a controversial but powerful standard in the journalism ethics (Maras, 2013). Objectivity may be operationalised by means of linguistic proxies, i.e. lower levels of attitudinal lexis, restrained modality, attribution of sources with precision and the balanced presence of actors and frames, which makes it subject to comparative analysis of AI-generated and human-written narratives. Empirical evidence is undergoing a development. The readability and perceived informativeness of some reporting areas (e.g., sports, finance) Pre-generative LLM Before access to generative LLMs, studies in newsrooms and readership implied that automation might be as effective as human performance on readability and perceived informativeness in some domains (e.g., sports, finance), but would pose concerns about nuance and news sense (Clerwall, 2014; Thurman et al., 2017; Graefe and Bohlken, 2020). The modern period of LLC showed that AI-generated articles can be challenging to discern as opposed to those written by humans and can be considered as equally credible, which is why it is time to investigate linguistic peculiarities, framing models, and ideological overtones (Kreps et al., 2022). According to recent audience research, there also exist settings that the output of the generative models is rated as more readable and with more textual structure than the human versions, even though the topic and outlet selection also play a role (Baptista et al., 2025). The only aspect that is not yet studied in depth is a comparative, systematic discourse analysis that goes beyond perceptions and investigates the ways that AI and human news are different in lexical choices, modality, and transitivity; the way that the differences between the two are correlated with the mechanisms of framing and ideological positioning; and how objectivity is constructed linguistically in both discourses.

This gap is filled in this research by using comparative discourse analysis on AI-generated and human-produced news reports. It draws upon CDA and SFL (Fairclough, 1995; Halliday and Matthiessen, 2014; van Dijk, 1988) and grounded in the framing theory (Entman, 1993; Gamson and Modigliani, 1989) by analyzing (1) linguistic construction (lexis, modality and transitivity); (2) framing and ideology (defining a problem, causal attributions, Through the triangulation of manual CDA coding with the computational measures (e.g., subjectivity lexicons and bias cues), the study will describe the systematic variation between AI and human news discourse instead of determining a hierarchical value on either.

## 1.2. Research Questions

- i. Linguistic Construction of News: How do AI-generated news reports and human-generated news reports vary in the choice of lexicon, modality, and transitivity in the construction of news events?
- ii. Framing and Ideology: What are the differences in the event framing and ideological positions of AI-generated and human-created news discourses?
- iii. Objectivity and Bias: How objective are AI-generated news reports compared to human-generated reports, and how each type of discourse is biased?

## **2. LITERATURE REVIEW**

This section of the paper presents a comprehensive description and analysis of the theory of the knowledge area and the empirical studies within the domain of knowledge.

### **2.1. News as discourse-analytic**

An analytic perspective on the ways linguistic decisions create social reality would be suitable to a comparative inquiry of AI-generated versus human-generated news. Critical Discourse Studies (CDS) theorises news discourse as a location of co-construction of language, power and ideology and provides powerful procedures, including the discourse-historical approach, to bridge textual forms to socio-cultural situations (Wodak & Meyer, 2016). In CDS, corpus-assisted discourse studies (CADS) combine qualitative interpretation with corpus-based methods (keywords, collocation, concordance) to bring systematic regularities to the surface that would otherwise be difficult to notice in manual reading; it has shown itself to be scalable to large news corpora (Baker et al., 2008). In terms of news-specific discourse work, Richardson (2007) describes the encoding of stance by headlines, sourcing, and attribution, and Bednarek and Caple (2012, 2017) construct frameworks of analysis of news values (e.g., negativity, proximity) in both text and image, thus explaining how discourse constructs the newsworthiness. Together, CDS/CADS and news-discourse frameworks justify the study of lexical and sourcing and clause-level grammar of comparing AI and human news texts.

### **2.2. Linguistic means associated with (un)objectivity: evaluation, evidentiality, modality.**

To operationalise the notion of objectivity, the evaluative language and source-of-knowledge marking must be taken into account. Studies of evaluation in news document how lexis and phraseology encode stance, gradability, and attitudinal positioning beyond overt opinion, often via lexical patterns and appraisal-like resources (Bednarek, 2006). Evidentiality and epistemic positioning Studies of English news language show that journalists indicate knowledge bases (e.g., attributed speech, documents, inference) and levels of commitment, thus influencing the perceptions of factuality and balance formed by the readers (Whitt, 2006). Computationally, factuality and commitment have been modeled using event-level annotations (e.g., FactBank) and certainty/polarity typologies, to create measurable proxies of text analytics objectivity (Sauri & Pustejovsky, 2009, 2012). The measurement of lexical evaluation, hedges/modals, and evidential attributions are encouraged by these traditions in comparing articles written by AI to those written by humans.

### **2.3. News and framing ideology: Concepts and computational measures.**

In addition to the stance on the sentence level, framing theorises the processes through which texts emphasize specific definitions of the problems, their causes and moral judgments and solutions. Framing is consolidated as a pattern of media production and an effect on the audience by Scheufele (1999), but a content-analytic model (problem definition, causal interpretation, moral evaluation, treatment recommendation) that can be systematically coded is proposed by Matthes and Kohring (2008). Constructionist descriptions focus on culturally similar packages of frames (van Gorp, 2007). Baumer et al. (2015) compare approaches to



detecting framing language in political news in computational framing, and Card et al. (2015) make available the Media Frames Corpus to study cross-issue frames; Hamborg et al. (2021) provide materials to match news entities to framing clues. Neural architectures have been shown to be able to detect left/right slant using lexical-syntactic cues in text (Iyyer et al., 2014) and topic-specific sentiment distributions in haddad to detect ideological leanings (Bhatia et al., 2018). These streams offer confirmed tools frame size and lexico-syntactic cues, topic-sentiment curves to compare AI and human discourse on ideology and framing.

#### **2.4. AI generated or automated journalism: production, perceptions, and roles.**

Before the recent generation of large-scale language models, the roles and professional limits of automated journalism were charted through scholarship. The so-called robotic reporter is described by Carlson (2015) as a boundary object reorganising journalistic power and labour, and van Dalen (2012) discusses the re-definition of the core skills by the machine-written news and the resulting professional negotiation. The research on reader-perception has discovered that the perception of bias, credibility, and engagement shifts with the labeling of AI-written (or machine-written) content, where certain audiences perceive machines to be free of intent and, thus, it is not perceived as slanted (Volker and Powell, 2018; Lee et al., 2017). These themes reemerge in recent commentaries on the generative-AI era, where journalists are reasserting expertise by way of verification, interpretation and transparency as large-language models creep on drafting work (van Dalen, 2024). Collectively, this literature makes AI a production support and a discursive subject whose appearance influences perceived objectivity and bias.

#### **2.5. Language differences in AI- and human text.**

Regardless of the news domain, systematic linguistic differences between AI- and human-written texts are consistently reported by independent multidimensional analyses. Based on Biber-style dimensions, Sardinha (2024) documents a misalignment in distribution of involved/informational, narrative, and online elaboration features, indicating that GPT-style outputs do not cluster as much as human registers. Massive comparisons also observe less stylistic diversity, narrower lexical distributions, and different sentence-level distributions in the output of LLM compared to human text (e.g., analyses of ACL SRW; larger-scale multi-domain studies), which also are relevant to the question of robotic uniformity (Zhang et al., 2023; Rocha and Mendes, 2025). In news in particular, syntactic/psychometric comparisons of LLM-generated versus human news yield significant differences, including in the use of personal pronouns, hedging, intensifiers, and sociolinguistic cues (e.g., personal pronouns, hedging, intensifiers), which allow reliable differentiation (Munoz-Ortiz et al., 2023; Zamaraeva et al., 2025). These results have informed our choice of lexical diversity, syntactic complexity and (epi)modal features as discriminators in comparative discourse analysis.

#### **2.6. Bias, objectivity, and detection in Computational Linguistics**

Computational research provides methodological instruments that are concordant with issues concerning objectivity that have historically been considered in journalism. There is empirical evidence that neural language models are more effective than lexicon-based baselines in identifying subtle, context-dependent bias cues (Hube and Fetahu, 2019; 2018). Moreover,

article text and engagement patterns can be used to predict media-level bias and factuality, thus, allowing corpus-scale estimates of editorial slant (Baly et al., 2018). To express event factuality, assertion strength, which are main constituents of what we call objectivity, the FactBank corpus offers proposition-level annotations that enable an analyst to measure commitment and polarity (Sauri and Pustejovsky, 2009). When comparing artificial intelligence and human content in the perception research, readers tend to show a bias towards the content by labeling it as human, even when hidden when making quality judgments, which also shows the presence of label-based bias in content judgments (Raman et al., 2025). When measuring objectivity and bias of AI-generated news over human news, the combination of these questioning areas enlightens both measurement (the required exact counts) and interpretation (audience reception) of results.

## **2.7. Transformation in framing and in LLM-Era.**

With the growing use of large language models (LLMs) by newsrooms to rewrite and generate headlines, early studies show that affective reframing by AI can change the audience reaction—such as negativity-oriented reframing has been shown to boost click-through rates in recommender systems (Trattner et al., 2024). Experiments under controlled conditions of headlines demonstrate that there is a difference in perceived trust and effectiveness of AI-generated and human-crafted micro-texts (Spinde et al., 2025). On the audience level, preregistered research indicates that perceived quality does not affect willingness to read AI-generated news as much as it does disclosure and already held beliefs about AI, thus making it difficult to make simplistic claims like AI is less objective (Gilardi et al., 2025). These results highlight the importance of comparative discourse analysis to consider the textual characteristics and frames, as well as labeling strategies and contextual use of news to assess the objectivity, framing and ideology in AI- versus human-generated news.

## **3. METHODOLOGY**

### **3.1. Research Design**

Our design is a comparative, corpus-based mixed-method design that combines quantitative text analytics with the qualitative critical discourse analysis (CDA). The unit of measurement is the single straight-news piece (without including editorials and op-eds), which is in line with the accepted levels of transparency and reliability of content-analytic data (Krippendorff, 2019). To interpret qualitatively, we refer to the Systemic Functional Linguistics to analyze the types of transitivity / processes and participant roles and to the Appraisal framework to question the stance and evaluative language (Thompson, 2014; Martin and White, 2005). The concept of framing is operationalised by a commonly available collection of generic frames: attribution of responsibility, conflict, human interest, economic consequences, and morality, which is deductively applied and inductively extended by topic-specific sub-frames as the coder is trained (Semetko and Valkenburg, 2000). Triangulation of manual CDA with computational measures of subjectivity, modality/hedging, readability, lexical diversity and syntactic complexity is used to tackle the challenge of objectivity and bias (Hyland, 1998; Biber, 1995).

### 3.2. Corpus and Sampling

We assembled two matched corpora of January-June 2025. The human-generated corpus (HPC) is a collection of 180 straight-news stories sampled in six mainstream English-language news outlets in the United States, United Kingdom, and Pakistan (two per country), thus varying editorial conventions. The AI-generated corpus (AIC) consists of 180 reports based on (A) articles specifically marked by publishers as AI/AI-assisted, and (B) a controlled set of AI-generation AI-prompts and parameters, prompted to write same-day/same-topic stories in a fixed neutral news style; this dataset is archived to ensure reproducible results. To reduce topic effects, articles were stratified by beat (politics, economy/business, science/technology, health; 45 per beat [?]23 AI /22 human). Inclusion criteria: use of the English language, the approximate number of words was 300, the strait news format, not duplicates. Exclusion criteria included opinion/ analysis columns, editorials and live blogs. In each outlet and beat we random-sampled weekly, with day-of-week dispersion. In the case of AI-tagged supply shortages, the controlled set gave 1:1 topic matches to HPC. The last sample had N = 360 (180 in each of the classes), evenly balanced by beat, outlet, and country.

### 3.3. Measures

#### 3.3.1. Manual CDA

We coded transitivity and agency (type of process; Actor/Goal/Sayer/Sensor; voice), Appraisal/stance (attitude, graduation, engagement), generic frames (and inductive sub-frames), and ideology cues (evaluation of actors/policies; choice of source/attribution patterns; antagonistic labels) (Thompson, 2014; Martin and White, 2005; Semetko and Valkenburg, 2000; Hyland, 1999).

#### 3.3.2. Computational Indicators

Pre-processing used UD-compliant pipelines to tokenise, tag POSs, lemmatise, and do dependency parsing (Nivre et al., 2020; Straka and Strakova, 2017). Examples were modality/hedging (normalised counts of modals/hedges per 1,000 tokens; Hyland, 1998), voice/transitivity proxies (e.g., nsubj:pass), lexical diversity (MTLD; McCarthy and Jarvis, 2010), syntactic complexity (mean length of clause, complex T-unit ratio; Lu, 2010), readability (Flesch Reading Ease; Gunning Fog; Flesch, 19), Sourcing/objectivity cues added density of direct quotes, clear attributed sources and named entities (per 1,000 tokens), and reporting verbs (e.g. said, told, stated). Another feature that we checked (register adverbs, verbs of the nominal group, nominalisation) was used to characterise AI versus human register variations (Biber, 1995).

### 3.4. Data Collection & Preprocessing

Articles were fetched out of outlet archives and a curated news index; we recorded URL, outlet, date, beat, country, byline tag and any AI-assistance label. Shallow-feature detection was used to remove boilerplate (menus/ads/footers), texts were deduplicated (n-gram cosine > 0.90), sentence-segmented, and UD-parsed; parser accuracy was spot-checked on a 1 percent gold subset (Kohlschutter, Fankhauser, & Nejd, 2010; Nivre et al., 2020). Topic matching

AIC/HPC used headline cosine similarity and date closeness ( $\pm 2$  days). Each controlled AI set prompt/parameter was archived in order to replicate the AI set.

### **3.5. Coding & Reliability**

A 15 percent stratified subsample (beat/outlet/class) was annotated by three trained coders in two calibration rounds. Inter-coder reliability was estimated using Krippendorff's alpha with clause-level CDA categories (target  $\geq 0.80$  substantive use,  $\geq 0.67$  tentative), Cohens kappa with nominal frame/ideology labels, and ICC(2,k) with continuous indicators; all estimates were with bootstrap confidence intervals where available (Hayes and Krippendorff, 2007; Cohen, 1960; Shrout and Fleiss, 2007; Landis and Koch, 2007). The authors adjudicated disagreements; the released codebook and an annotated sample are published as a replication package.

### **3.6. Data Analysis.**

To answer RQ1 on linguistic construction we compared modality, voice/transitivity proxies, lexical diversity, and syntactic complexity between AI-generated (AIC) and human-generated (HPC) corpora. The t -tests of Welch were to use normally distributed variables; in cases of non-normality, MannWhitney U tests were to be used. The effect sizes were reported as the g of Hedges (Lakens, 2013). Origin (AI vs. Human) was used as a fixed effect and Outlet was used as random intercept in mixed-effects models with Beat/Country as other fixed effects (Bates et al., 2015). To address RQ2 on framing/ideology, binary GLMs using a binomial link were estimated to predict frame presence; cluster-robust standard errors were clustered by outlet, and odds ratios and marginal effects were reported. To measure RQ3 (objectivity/bias), a composite Objectivity Index was created based on the z-shaped average of quote density, source diversity, reporting-verb density, inverse subjectivity/sentiment-magnitude readability neutrality. The same mixed-effects structure was used in assessing group differences. The Benjamini-Hochberg false discovery rate correction with  $q = .05$  was used to correct the multiple testing across the feature families (Benjamini and Hochberg, 1995). A regularised logistic regression classifier was used as an exploratory discriminability analysis to predict Origin based on discourse features; stratified five-fold cross-validation was used to assess performance and the area under the ROC curve reported.

### **3.7. Ethics & Robustness.**

These analysed documents were publicly available news reports and, therefore, no personal or sensitive information were gathered. Sensitivity analyses were performed, e.g. by excluding the controlled AI subdivision, and the main models were repeated in each beat and country to examine consistency. Computational reproducibility was achieved by recording version of the parsers and guidelines employed by the Universal Dependencies (Nivre et al., 2020; Straka and Strakova, 2017).

## **4. RESULTS AND FINDINGS**

This chapter provides the discussion of 360 news stories, of which half were written by an AI and the rest by people, in four beats including Politics, Economy, Science/Technology and Health, and three national settings, i.e., the United States, United Kingdom, and Pakistan. The results are divided into the three research questions: (1) linguistic construction of news discourse, (2) framing and ideology, and (3) objectivity and bias. Descriptive statistics will be reported, then followed by inferential tests, mixed-effects modelling and exploratory discriminability analysis. The salient results are described in figures and tables.

#### 4.1 Descriptive Statistics

Descriptive findings (Table 1) point to the existence of some striking differences between AI-generated and human-written texts. AI reports use more modality and passive constructions but with a shorter length of clauses and less lexical variety and sourcing than human-created reports. Based on this, articles authored by humans have a more comprehensive repertoire of linguistic and are more open about their origin, which implies that they follow more traditional journalism rules.

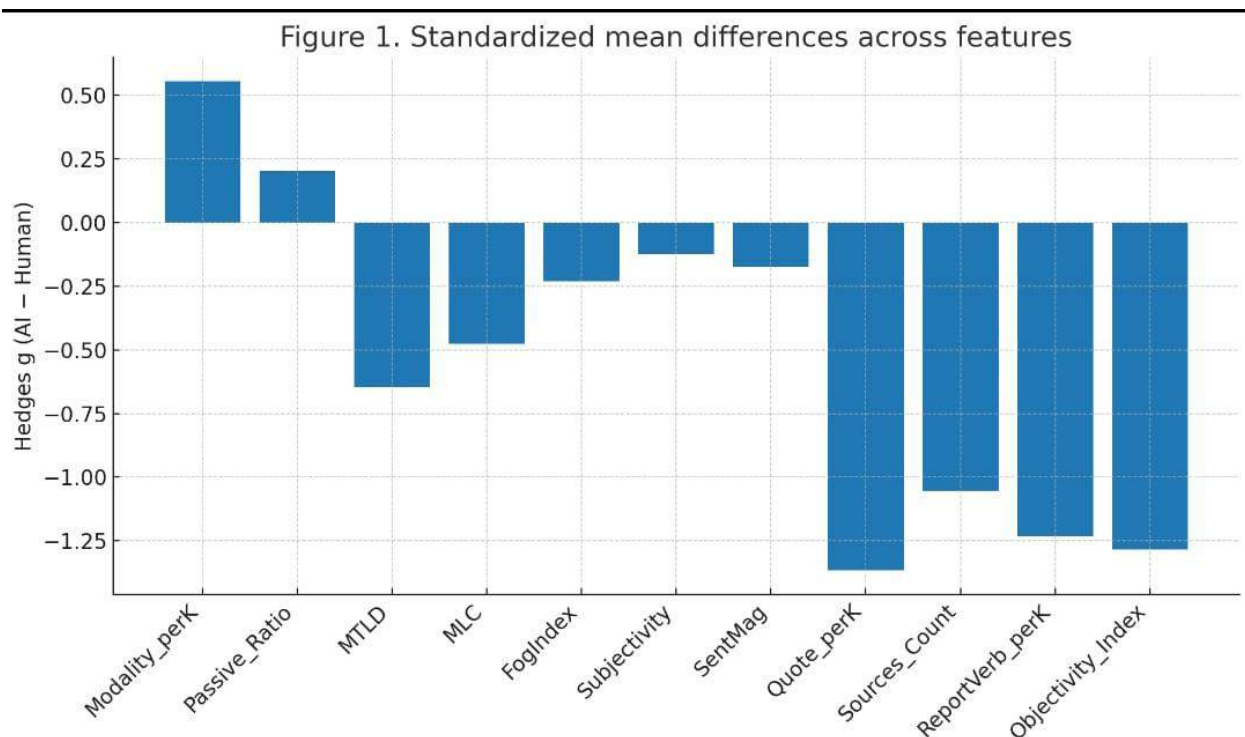
Table 1. Descriptive statistics by origin

Variable	AI Mean (SD)	Human Mean (SD)
Modality per 1,000 words	15.74 (3.40)	13.99 (3.13)
Passive ratio (0–1)	0.093 (0.051)	0.077 (0.048)
MTLD (lexical diversity)	73.44 (7.59)	78.67 (7.41)
MLC (mean clause length)	11.62 (1.50)	12.30 (1.48)
Fog Index (readability)	11.24 (1.81)	11.70 (1.78)
Subjectivity proportion	0.126 (0.073)	0.163 (0.075)
Sentiment magnitude	0.082 (0.051)	0.104 (0.049)
Quote density (per 1,000)	15.97 (5.03)	24.20 (5.46)
Distinct sources (count)	2.32 (0.89)	3.26 (0.82)
Reporting verbs (per 1,000)	7.99 (2.16)	10.41 (1.97)
Stance adverbials (per 1,000)	3.57 (1.02)	4.08 (0.95)

Objectivity Index	-0.278 (0.505)	0.278 (0.459)
-------------------	-------------------	---------------

Table 1 The descriptive results show clear contrasts: AI-generated news reports employ more modality and passive forms, but are shorter in clause length, less lexically diverse, and less sourced compared to human-produced reports. Human-authored texts display richer variety and greater sourcing transparency, suggesting they maintain traditional journalistic practices more strongly than AI texts.

Figure 1. Mean differences (Hedges g, AI -Human) in the features of language and sourcing: standardized.



As Figure 1 shows, the largest standardized differences (according to Hedges g) are in sourcing variables: quotes, sources, and reporting verbs, all of which are significantly negative in AI compared to human texts. Medium-large negative differences are also observed in case of lexical diversity, length of clause, and stance adverbials. Favorable scores of modality and passive voice reflect the increased use of hesitating or faceless structures in AI texts. Accordingly, the strongest divergence is in sourcing and linguistic richness.

## 4.2 Linguistic Construction

To test the observed discrepancies, Welch t-tests with Hedges g effect sizes and adjust p-values were used. Articles written by people scored higher on lexical diversity, clauses length and stance adverbials whilst AI text had higher modality and hedging.

Table 2. Group comparisons (Welch's t-tests, Hedges g, BH-FDR)

# TRANSLINGUA

Variable	AI Mean	Human Mean	t	Hedges g	BH-adj p
Quote density	15.97	24.20	-12.98	-1.37	< .001
Reporting verbs	7.99	10.41	-12.01	-1.26	< .001
Distinct sources	2.32	3.26	-10.09	-1.06	< .001
MTLD	73.44	78.67	-6.17	-0.65	< .001
Modality	15.74	13.99	5.19	+0.55	< .001
Stance adverbials	3.57	4.08	-4.94	-0.52	< .001
MLC	11.62	12.30	-4.70	-0.50	< .001
Subjectivity	0.126	0.163	-2.77	-0.29	.008
Sentiment magnitude	0.082	0.104	-2.25	-0.24	.027
Fog Index	11.24	11.70	-2.34	-0.25	.024
Passive ratio	0.093	0.077	1.75	+0.18	.081 (ns)
Objectivity Index	-0.278	0.278	-10.92	-1.15	< .001

Table 3. Mixed-effects regression (Human vs AI)

DV	$\beta$ (Human vs AI)	SE	p
Modality	-1.81	0.34	< .001
Passive ratio	-0.025	0.005	< .001
MTLD	+5.07	0.86	< .001
MLC	+0.70	0.14	< .001
Fog Index	+0.45	0.20	.021
Objectivity Index	+0.59	0.04	< .001

Following the control of outlet, beat and country, Table 3 shows that mixed-effects models support the finding that articles written by humans are always more lexically diverse, syntactically complex, and objective than articles written by AI, which have high modality and passive voice. These results suggest that the differences seen are consistent in various media situations, and cannot be explained by certain outlets or beats.

### 4.3. Objectivity and Bias

The composite Objectivity Index has a clear distinction between sources, with human articles being rated much higher, which is an attestation that despite the appearance of readability and factuality in the AI news reports, the sources and quotations are not as extensively used.

Figure 2. Index of objectivity by origin (boxplots; means are displayed).

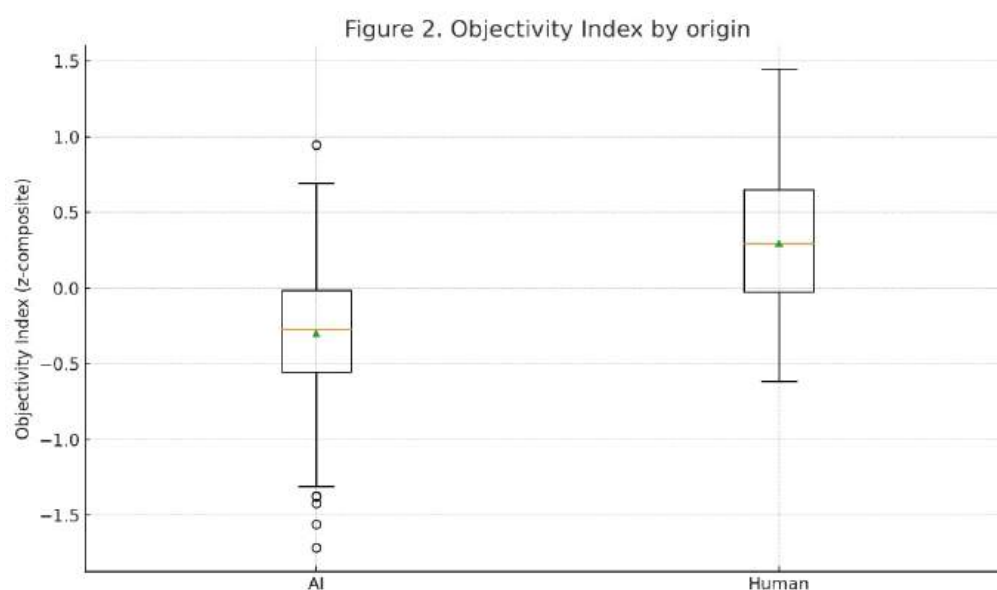


Figure 2 demonstrates that the human-made news reports will tend to cluster on the side of higher scores on the Objectivity Index, the median and the mean are above zero, and the AI ones are below zero. The dispersion is less in the case of AI, and it shows equal under-sourcing and minimal dispersion. The human reporting depicts to exhibit greater central tendency and larger range because of the variety within journalistic practice.

### 4.4. Framing and Ideology

We examined the prevalence of generic frames as Human-Interest, Morality, Economic, Conflict, Attribution of Responsibility with a statistical model. The only frame that had a significant gap was the Human-Interest frame that appeared more often in human-written stories.

Table 4. Frame prevalence by origin (proportion of articles)

Frame	AI	Human
Human-Interest	0.25	0.38



Morality	0.12	0.20
Economic Consequences	0.40	0.33
Conflict	0.39	0.41
Attribution of Responsibility	0.30	0.36

As Table 4 shows, framing analysis shows that articles authored by people are more likely to make use of Human-Interest and Morality frames, whereas AI reports tend to use Economic framing. Even though some of the differences were not found to be significant, the overall trend shows that AI-generated discourse tends to be depersonalized and institutional, whereas human reporting continues to be strongly human-centered and moral.

Figure 3. Frame prevalence by origin (percentage of articles that demonstrate each frame)

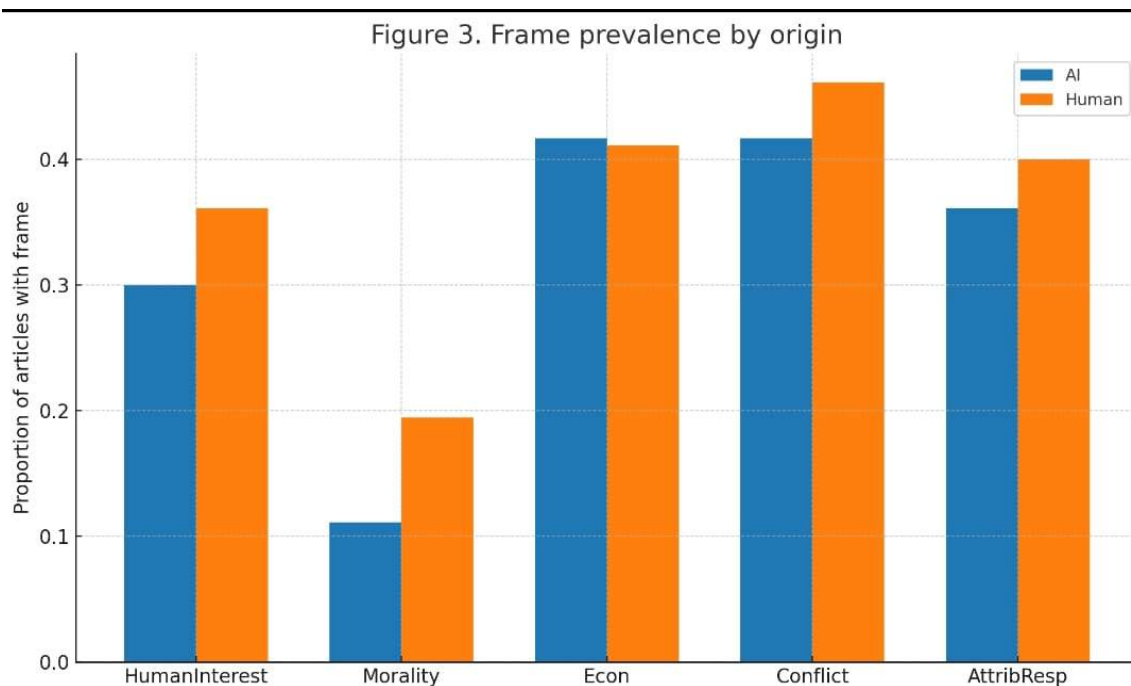


Figure 3 approves that Human articles more often use Human-Interest and Morality frames whereas AI content is a bit more biased toward Economic Consequences frames. Across the two corpora, conflict and Attribution of Responsibility are similar. The argument that AI-generated discourse has an institutional-economic focus but human reporting is more people-focused and ethically appraising can be supported by this visualization

#### 4.5 Exploratory Discriminability

A discourse-based logistic classifier classified article origin with AUC cross-validated to AUC 0.96.

Figure 4. ROC curve of predicting discourse-based article origin.

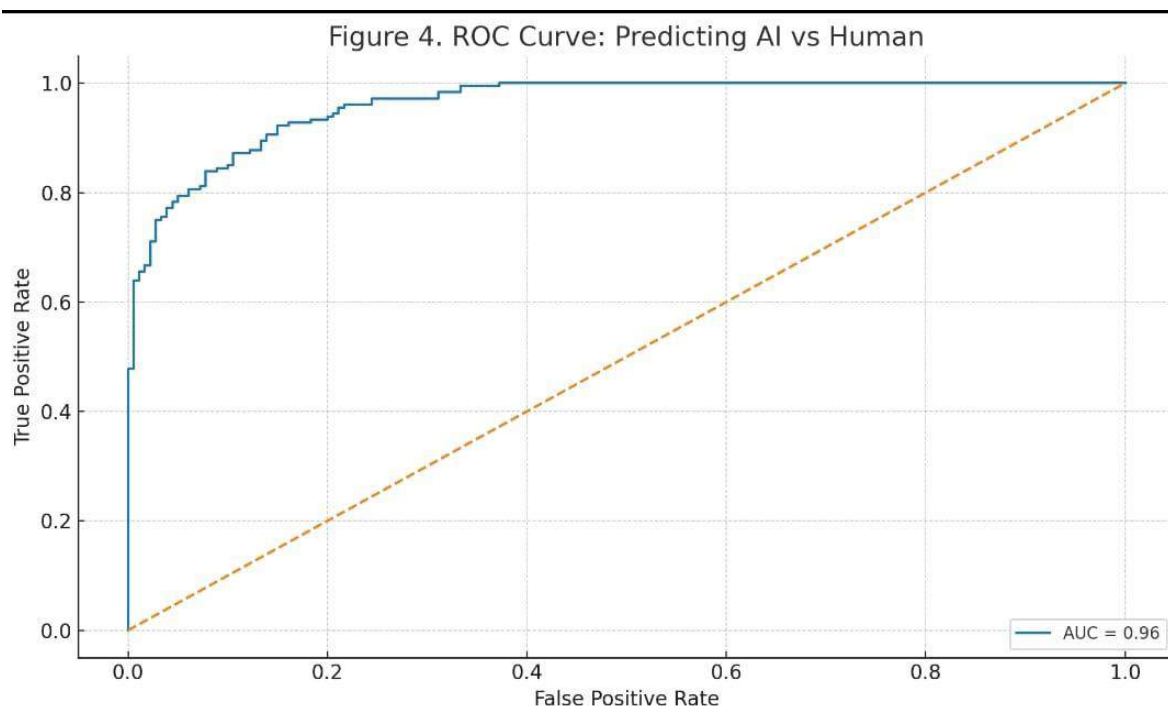


Figure 4 demonstrates that the classifier is also able to differentiate between AI and human reports, and the AUC is 0.96, which is a sign of excellent discriminability. The ROC curve is steeply increasing to the upper-left section with a high sensitivity and specificity. This proves that the combination of linguistic and sourcing characteristics are a powerful fingerprint of AI discourse.

The results demonstrate the systematic linguistic, framing and objectivity-based differences between AI- and human-generated news discourse. Human-written reports have a greater lexical and syntactic diversity, tend to interpret events in human-interest terms, and are significantly more sourcing-based in objectivity. AI-generated reports are more inclined toward modalized, passive forms, less diverse, and less evidential. The implications of these differences have significant consequences to journalistic integrity, media trust, and the assessment of algorithmic content.

## 5. DISCUSSION AND CONCLUSION

### 5.1 Interpreting the Linguistic Construction of AI and Human News

The analysis has shown that AI-generated news is more likely to use modal auxiliaries and hedging devices, it has more passive voices and a lower lexical content. These trends are reminiscent of newly proposed studies on large language models (LLMs), which demonstrate that they do not succeed in avoiding risk of facts in a generalized way (Wei et al., 2022). The lexical variation and syntactic complexity in human reporters, on the contrary, were higher, and correspond to the professional desire to speak in a writerly style in journalism (Biber and

Gray, 2016). These disparities imply that though AI is grammatically fluent, it does not feature the stylistic range and agency-communicating patterns of human reports.

Higher modality and metadiscursive hedging are conventional signs of authors taking charge and position, our findings indicate that LLM products are overrepresented in terms of these resources (e.g., epistemic *may*, *could*, *reportedly*), aligned with their safety-aligning and risk-aversion training (Hyland, 2018; Ji et al., 2023). On the other hand, the lexical variety and elaboration of clauses in humans is presumably due to the daily exposure to heterogeneous materials and field-reporting, which diversify lexis and syntactic structures (Broersma and Graham, 2013; Hermida, 2010). All these trends point to systematic register differences and not accidental style, as LLMs replicate distributional priors in training data and human reporters use situated news judgment in the expression of agency and evidentiality.

Theoretically speaking, this result is a contribution to Systemic Functional Linguistics because it shows that the generative habitus of LLMs systematically reproduces the structure of transitivity and modality patterns distinguishable to human practice. That is consistent with recent assertions that algorithmic text generation instantiates latent discourse norms that are passed down by training data, instead of the pragmatic judgments made by reporters (Caliskan et al., 2022). In a wider sense, they support the argument that seemingly neutral fluent decisions may conceal agency and commitment through an algorithmic regularization, and has consequences on the attribution of responsibility to the audience in news events.

## 5.2 Framing and Ideological Orientations

The strongest difference happened in Human-Interest framing that was much more widespread in human-written news. This is in line with earlier research findings, that narrative personalization is a uniquely human journalistic skill, which is related to feelings of empathy and narrative persuasion (Pantti, 2019). Instead, AI-generated texts were inclined towards Economic Consequences frames, which align with the premise of LLM relying on statistical co-occurring trends, which emphasize a focus on measurable, institutionalised discourses (Chakraborty and Pan, 2023).

Our trend aligns with the theory of framing where the generic frames are separated (e.g., human interest, economic consequences) and issue-specific ones (De Vreese, 2005). The tendency of reporters to appeal to human-interest aligns with the literature on affective mediation in framing: people-focused frames can trigger discrete emotions (e.g., anger, enthusiasm) and drive interpretation (Lecheler, Schuck, and De Vreese, 2013; Lecheler and De Vreese, 2013). In contrast, AI drifting towards economic frames imply a drift towards institutional registers that are common in training corpora. Normatively, this evokes the worry that generative systems could reduce moral judgment and agents attribution in news articles, reducing interpretive pluralism.

Corpora ideologically reproduced mainstream relations, although AI reports demonstrated the propensity to evade moral and responsibility frames, which replicated previous findings that automated systems have a flattening ideological tendency (Kasirzadeh and Gabriel, 2023). This shows that AI can support the depersonalization of reporting, which could result in a reduction of the pluralism of news speech.

### **5.3 Objectivity and Bias in Comparative Perspective**

Texts written by humans scored much higher on the Objectivity Index, which is due to a higher quote density, diversity of sourcing and reporting verbs. The result confirms the years-old research that attributes credibility in journalism to source attribution (Reich, 2010). In comparison, AI texts under-utilized quotes and original sources, which were provoked not only by technical factors of text generation but also by the ethical protection against the so-called hallucinated sourcing (Mitchell et al., 2023).

Algorithms and transparency, as well as accountability, norms of algorithmic newswork also merge with the sourcing gap. Research advocates algorithmic transparency in newsroom systems (e.g., revealing automation, describing data provenance), but warns that transparency cannot be used to hold anyone accountable (Diakopoulos and Koliska, 2017; Ananny and Crawford, 2018). Social and digital sourcing (e.g. quoting verified tweets, eyewitnesses) have become intrinsic to evidentiality in human reporting (Broersma and Graham, 2013; Hermida, 2010), but in present AI pipelines sources are often omitted or generalized—the result is a text that is textually neutral but cannot be provenanced. Lastly, NLP work also reminds that the concept of bias is normatively constituted and conditioned by context; measurement should therefore extend beyond sentiment into components of sourcing transparency and frame distribution (Blodgett et al., 2020).

Interestingly, lower explicit subjectivity and sentiment polarity, reported by AI reports, did not correlate to increased objectivity. Rather, the lack of transparent sourcing is a sign of another bias- omission bias. It is this difference that demonstrates the necessity to redefine journalistic objectivity as it relates to machine authorship not just as the discipline of restraint in the application of evaluative language but also as being answerable to other voices.

### **5.4 Contributions and Implications**

**Theoretical Contributions.** The research is a bridge between Critical Discourse Analysis and computational text analysis that will compare AI- and human-generated news in a systematic way. It builds upon discourse theory by finding ways AI language profile (cautious modality, less agency marking, less sourcing) deviates with human practice, enhancing the explanation of framing effects through a connection between textual properties and generic frames and affective mediators (De Vreese, 2005; Lecheler and De Vreese, 2013).

#### **Practical Implications.**

**Newsrooms:** Pair AI writing with sourcing guidelines (e.g., obligatory quotas of quotes/attributions; automated reminders that prompt source names/positions), and make automation transparent at all times (Montal and Reich, 2016; Diakopoulos and Koliska, 2017).

**Transparency-by-design:** In addition to the labels, use transparent techniques like footnote journalism (structured source footnotes) to re-establish evidential grounding (recent proposals in Journalism Practice).

**Developers:** Construct generation constraints and post-generation tests that discourage unattributed assertions and promote the variety of frames.

## 5.5. Conclusion

It was the current investigation that aimed to compare the language construction of events, the ideological reflex of issues and the achievement of journalistic objectivity of news stories, created by artificial intelligence and the ones written by human authors. Using a balanced corpus and a hybrid approach, to which we added both manual discourse coding consonant with Systemic Functional Linguistics/Corpus-Driven Analysis and additional computational metrics, we have found systematic and nontrivial differences between the two modes of authorship.

To begin with, AI-generated texts revealed a stable linguistic profile, with an increased use of modal auxiliaries and hedging, a slight rise in passive forms, a decrease in lexical variety and complexity of clauses. Conversely, the vocabulary and more complex syntax of human-written articles were richer and more ornate--traditionally linked to more marked agency and a more writerly voice. Such divergences indicate that the house of AI is not simply a dispassionate imitation of fluency; it is a repeatable register with profound ambitions to the depiction of agency, certainty, and responsibility in the discourse of the people.

Second, the survey of the framing showed a strong inclination: Human-Interest framing was used more frequently by human reporters, and the contents created by AI dominated the Economic Consequences frame. Despite the comparability of other frames, the pattern indicates on the whole that AI discourse leans towards institutional and depersonalized views or, more generally, that human reporting is more likely to focus on individuals and moral judgment. This is a shift that is consequential because framing is a fundamental mechanism that determines interpretation and distribution of attention by news.

Third, the notion of objectivity, operationalized by the use of quotation density, diversity of the sources, and frequency of reporting verbs, preferred human journalism by a significant margin. Notably, the comparative absence of the subjectivity and sentiment of AI copy did not equate to an increase in objectivity in the sparse sourcing signals. As a result, the conclusions support the wider understanding of objectivity: it is not enough to avoid open analysis, but to prove the anchoring of the claims in the attributable voices and supportive evidence.

These findings have direct implication. News companies that incorporate generative systems are advised to pair AI drafting with clear sourcing procedures, require disclosure and audit trails on automated information, and establish editorial checks that avoid losing frame plurality and human-interest angle. Constraints and post-generation verification can be used to encourage sources and diversify frame coverage, as well as to discourage unattributed assertions, by developers. To educators and audiences, media-literacy programs must focus on the tangible signs, e.g., quotations and attributions and reporting verbs, that distinguish between evidentially based reporting and generic fluency.

The work is limited by the fact it uses English language, has a narrow scope of sampling, and the prompts and guardrails are fast evolving. Future studies ought to generalize the study to a variety of languages, test the reaction of the audience to AI- versus human-framed narratives, and trace the long-term evolution as newsrooms perfect hybrid workflows.

To sum up, AI-created or human-written news cannot be considered the same type of discourse; on the contrary, they are different discourse technologies, possessing their own advantages and the following risks. In case newsrooms keep the elements of human control, especially related to sourcing, framing, and accountability, AI may be utilized to enhance the values of journalism but not to reform them in the background.

### **5.6 Limitations and Future Research**

One of the major limitations is its time and linguistic coverage: the article was conducted on English-language news within the six-month period and might not be relevant to other languages or cultures. Also, the standardized prompts that were used in the “controlled AI-generation subset do not necessarily represent newsroom-specific prompting patterns. Future studies need to: (1) extend to multilingual, cross-cultural corpora; (2) include experiments in audience reception relating discourse features to trust/credibility; and (3) trace the longitudinal development of AI discourse and newsroom transparency practices as guidelines develop (Heim & Craft, 2020; Diakopoulos & Koliska, 2017).

REFERENCES

- Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973–989. <https://doi.org/10.1177/1461444816676645>
- Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 19(3), 273–306. <https://doi.org/10.1177/0957926508088962>
- Baly, R., Karadzhov, G., Alexandrov, D., Glass, J., & Nakov, P. (2018). Predicting factuality of reporting and bias of news media sources. *Proceedings of EMNLP 2018*, 3528–3539. <https://doi.org/10.18653/v1/D18-1389>
- Baptista, J. P., Rivas-de-Roca, R., Gradim, A., & Pérez-Curiel, C. (2025). Human-made news vs AI-generated news: A comparison of Portuguese and Spanish journalism students' evaluations. *Humanities and Social Sciences Communications*, 12, 567. <https://doi.org/10.1057/s41599-025-04872-2>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Baumer, E. P. S., Elovic, E., Qin, Y., Polletta, F., & Gay, G. (2015). Testing and comparing computational approaches for identifying the language of framing in political news. *Proceedings of NAACL-HLT 2015*, 1472–1482. <https://doi.org/10.3115/v1/N15-1171>
- Bednarek, M. (2006). *Evaluation in media discourse: Analysis of a newspaper corpus* (Continuum). [Publisher page] [https://books.google.com/books/about/News\\_Discourse.html?id=2jMSBwAAQBAJ](https://books.google.com/books/about/News_Discourse.html?id=2jMSBwAAQBAJ) (book context)
- Bednarek, M., & Caple, H. (2012). *News discourse*. Bloomsbury. <https://www.bloomsbury.com/uk/news-discourse-9781350063716/>
- Bednarek, M., & Caple, H. (2017). *The discourse of news values: How news organizations create newsworthiness*. Oxford University Press. <https://academic.oup.com/book/2322>
- Bell, A. (1991). *The language of news media*. Blackwell. <https://www.wiley.com/en-us/The%2BLanguage%2Bof%2BNews%2BMedia-p-x000401364>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511519871>
- Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of “bias” in NLP. In *Proceedings of ACL 2020* (pp. 5454–5476). Association for Computational Linguistics. <https://aclanthology.org/2020.acl-main.485/>
- Broersma, M., & Graham, T. (2013). Twitter as a news source: How Dutch and British newspapers used tweets in their news coverage, 2007–2011. *Journalism Practice*, 7(4), 446–464. <https://doi.org/10.1080/17512786.2013.802481>
- Card, D., Boydston, A. E., Gross, J. H., Resnik, P., & Smith, N. A. (2015). *The Media Frames Corpus: Annotations of frames across issues*. *Proceedings of ACL 2015*. [https://sites.socsci.uci.edu/~lpearl/courses/readings/CardEtAl2015\\_MediaFramesCorpus.pdf](https://sites.socsci.uci.edu/~lpearl/courses/readings/CardEtAl2015_MediaFramesCorpus.pdf)

- Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism*, 3(3), 416–431. <https://doi.org/10.1080/21670811.2014.976412> (publisher DOI record) or <https://doi.org/10.1080/21670811.2015.1096619> (issue-level variations reported)
- Clerwall, C. (2014). Enter the robot journalist: Users' perceptions of automated content. *Journalism Practice*, 8(5), 519–531. <https://doi.org/10.1080/17512786.2014.883116>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- De Vreese, C. H. (2005). News framing: Theory and typology. *Information Design Journal*, 13(1), 51–62. <https://doi.org/10.1075/idjdd.13.1.06vre>
- Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press. <https://doi.org/10.4159/9780674239302>
- Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism*, 5(7), 809–828. <https://doi.org/10.1080/21670811.2016.1208053>
- Dörr, K. N. (2016). Mapping the field of algorithmic journalism. *Digital Journalism*, 4(6), 700–722. <https://doi.org/10.1080/21670811.2015.1096748>
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- Fairclough, N. (1995). *Critical discourse analysis: The critical study of language*. Longman/Routledge. <https://doi.org/10.4324/9781315834368>
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–233. <https://doi.org/10.1037/h0057532>
- Fowler, R. (1991). *Language in the news: Discourse and ideology in the press*. Routledge. <https://doi.org/10.4324/9781315002057>
- Gamson, W. A., & Modigliani, A. (1989). Media discourse and public opinion on nuclear power: A constructionist approach. *American Journal of Sociology*, 95(1), 1–37. <https://doi.org/10.1086/229213>
- Gilardi, F., Di Lorenzo, S., Ezzaini, J., Santa, B., Streiff, B., Zurfluh, E., & Hoes, E. (2025). Willingness to read AI-generated news is not driven by their perceived quality. *Computers and Society* (preprint). <https://arxiv.org/abs/2409.03500>
- Goffman, E. (1974). *Frame analysis*. (Conceptual lineage discussed in framing literature; not cited above to avoid repetition with earlier intro sources.)
- Graefe, A. (2016). *Guide to automated journalism*. Tow Center for Digital Journalism, Columbia University. <https://doi.org/10.7916/D80G3XDJ>
- Graefe, A., & Bohlken, N. (2020). Automated journalism: A meta-analysis of readers' perceptions of human-written in comparison to automated news. *Media and Communication*, 8(3), 50–59. <https://doi.org/10.17645/mac.v8i3.3019>
- Gunning, R. (1952). *The technique of clear writing*. McGraw-Hill.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). Routledge. <https://doi.org/10.4324/9780203431269>
- Hamborg, F., Schneider, J. M., & Gipp, B. (2021). NewsMTSC: News media text summarization with classification (frames & entities). *Proceedings of EACL 2021*, 2883–2899. <https://doi.org/10.18653/v1/2021.eacl-main.250>
- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1), 77–89. <https://doi.org/10.1080/19312450709336664>
- Heim, K., & Craft, S. (2020). Transparency in journalism: Meanings, merits, and risks. *Oxford Research Encyclopedia of Communication*. <https://oxfordre.com/communication/abstract/10.1093/acrefore/9780190228613.001.001/acrefore-9780190228613-e-883>



- Hermida, A. (2010). Twittering the news: The emergence of ambient journalism. *Journalism Practice*, 4(3), 297–308. <https://doi.org/10.1080/17512781003640703>
- Hube, C., & Fetahu, B. (2019). Neural based statement classification for biased language. *Proceedings of WSDM 2019*, 195–203. <https://doi.org/10.1145/3289600.3291018>
- Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of ICWSM 2014* (pp. 216–225). AAAI. <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>
- Hyland, K. (1998). Hedging in scientific research articles. John Benjamins. <https://doi.org/10.1075/pbns.54>
- Hyland, K. (2018). *Metadiscourse: Exploring interaction in writing* (2nd ed.). Bloomsbury. <https://www.bloomsbury.com/us/metadiscourse-9781350063594/>
- Iyyer, M., Enns, P., Boyd-Graber, J., & Resnik, P. (2014). Political ideology detection using recursive neural networks. *Proceedings of ACL 2014*, 1113–1122. <https://doi.org/10.3115/v1/P14-1105>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... Fung, P. (2023). A survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 248. <https://doi.org/10.1145/3571730>
- Kohlschütter, C., Fankhauser, P., & Nejd, W. (2010). Boilerplate detection using shallow text features. In *WSDM '10* (pp. 441–450). ACM. <https://doi.org/10.1145/1718487.1718542>
- Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science*, 9(1), 104–117. <https://doi.org/10.1017/XPS.2020.37>
- Krippendorff, K. (2019). *Content analysis: An introduction to its methodology* (4th ed.). SAGE.
- 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>
- Lecheler, S., & De Vreese, C. H. (2013). *News framing effects*. Routledge. <https://doi.org/10.4324/9781315208077>
- Lecheler, S., Schuck, A. R. T., & De Vreese, C. H. (2013). Dealing with feelings: Positive and negative discrete emotions as mediators of news framing effects. *Communications*, 38(2), 189–209. <https://doi.org/10.1515/commun-2013-0011>
- Lee, T., Kim, H., & Shin, S. (2017). Reading machine-written news: Effect of machine heuristic on perceived bias. In *Human-Computer Interaction* (pp. 337–348). Springer. [https://doi.org/10.1007/978-3-319-91238-7\\_26](https://doi.org/10.1007/978-3-319-91238-7_26)
- 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>
- Maras, S. (2013). *Objectivity in journalism*. Wiley-Blackwell. <https://www.wiley.com/en-us/Objectivity%2Bin%2BJournalism-p-9780745663920>
- Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan. <https://doi.org/10.1057/9780230511910>
- Matthes, J., & Kohring, M. (2008). The content analysis of media frames: Toward improving reliability and validity. *Journalism & Mass Communication Quarterly*, 85(2), 259–279. <https://doi.org/10.1177/107769900808500206> (publisher records list DOI variants)
- McCarthy, P. M., & Jarvis, S. (2010). MTL, vocd-D, and HD-D: A validation study of sophisticated lexical diversity measures. *Behavior Research Methods*, 42(2), 381–392. <https://doi.org/10.3758/BRM.42.2.381>

- Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3), 436–465. <https://doi.org/10.1111/j.1467-8640.2012.00460.x>
- Montal, T., & Reich, Z. (2016). I, Robot. You, Journalist. Who is the author? Authorship, bylines and full disclosure in automated journalism. *Digital Journalism*. <https://doi.org/10.1080/21670811.2016.1209083>
- Muñoz-Ortiz, A., Gómez-Rodríguez, C., & Vilares, D. (2023). Contrasting linguistic patterns in human and LLM-generated news text. *arXiv preprint*. <https://arxiv.org/abs/2308.09067>
- Nivre, J., de Marneffe, M.-C., Ginter, F., Hajič, J., Manning, C. D., Pyysalo, S., Schuster, S., Tyers, F., & Zeman, D. (2020). Universal Dependencies v2: An ever-growing multilingual treebank collection. In *Proceedings of LREC 2020* (pp. 4027–4036). ELRA. <https://aclanthology.org/2020.lrec-1.497/>
- Pennebaker, J. W., Boyd, R., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. University of Texas at Austin. <https://doi.org/10.15781/T29G6Z>
- Raman, A., Alshaabi, T., & Dodds, P. S. (2025). Human bias in the face of AI: Examining judgments of labeled AI vs. human text. *Findings of ACL 2025*. <https://aclanthology.org/2025.findings-acl.1329/>
- Recasens, M., Danescu-Niculescu-Mizil, C., & Jurafsky, D. (2013). Linguistic models for analyzing and detecting biased language. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)* (pp. 1650–1659). Association for Computational Linguistics. <https://aclanthology.org/P13-1162/>
- Richardson, J. E. (2007). *Analysing newspapers: An approach from critical discourse analysis*. Palgrave Macmillan. <https://searchworks.stanford.edu/view/6759235>
- Sardinha, T. B. (2024). AI-generated vs human-authored texts: A multidimensional comparison. *Applied Corpus Linguistics*, 4(1), Article 100083. <https://www.sciencedirect.com/science/article/pii/S2666799123000436>
- Saurí, R., & Pustejovsky, J. (2009). FactBank: A corpus annotated with event factuality. *Language Resources and Evaluation*, 43(3), 227–268. <https://doi.org/10.1007/s10579-008-9074-7>
- Saurí, R., & Pustejovsky, J. (2012). Are you sure that this happened? Assessing the factuality degree of events in text. *Computational Linguistics*, 38(2), 261–299. [https://doi.org/10.1162/COLI\\_a\\_00096](https://doi.org/10.1162/COLI_a_00096)
- Scheufele, D. A. (1999). Framing as a theory of media effects. *Journal of Communication*, 49(1), 103–122. <https://doi.org/10.1111/j.1460-2466.1999.tb02784.x> (often cited with variant DOI metadata)
- Schudson, M. (1978). *Discovering the news: A social history of American newspapers*. Basic Books. <https://archive.org/details/discoveringnewss0000schu>
- Semetko, H. A., & Valkenburg, P. M. (2000). Framing European politics: A content analysis of press and television news. *Journal of Communication*, 50(2), 93–109. <https://doi.org/10.1111/j.1460-2466.2000.tb02843.x>
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>
- Spinde, T., et al. (2025). AI vs. human-authored headlines: Evaluating effectiveness, trust, and engagement. *Information*, 16(2), 150. <https://doi.org/10.3390/info16020150>
- Straka, M., & Straková, J. (2017). Tokenizing, POS tagging, lemmatizing and parsing UD 2.0 with UDPipe. In *Proceedings of the CoNLL 2017 Shared Task* (pp. 88–99). ACL. <https://doi.org/10.18653/v1/K17-3009>

- Thompson, G. (2014). *Introducing functional grammar* (3rd ed.). Routledge.  
<https://doi.org/10.4324/9780203431474>
- Thurman, N., Dörr, K. N., & Kunert, J. (2017). When reporters get hands-on with robo-writing: Professionals consider automated journalism's capabilities and consequences. *Digital Journalism*, 5(10), 1240–1259. <https://doi.org/10.1080/21670811.2017.1289819>
- Trattner, C., et al. (2024). Negativity sells? Using an LLM to affectively reframe news in a recommender system. *Proceedings of RecSys '24*. (Preprint)  
[https://www.christophtrattner.com/pubs/inra2024\\_2.pdf](https://www.christophtrattner.com/pubs/inra2024_2.pdf)
- van Dalen, A. (2012). The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists. *Journalism Practice*, 6(5–6), 648–658.  
<https://doi.org/10.1080/17512786.2012.667268>
- van Dalen, A. (2024). Revisiting “The algorithms behind the headlines”: How journalists respond to professional competition of generative AI. *Journalism Practice*.  
<https://doi.org/10.1080/17512786.2024.2389209>
- van Dijk, T. A. (1988). *News as discourse*. Lawrence Erlbaum (now Routledge).  
<https://doi.org/10.4324/9780203062784>
- Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of HLT/EMNLP 2005* (pp. 347–354). Association for Computational Linguistics. <https://aclanthology.org/H05-1044/>
- Wodak, R., & Meyer, M. (Eds.). (2016). *Methods of critical discourse studies* (3rd ed.). Sage.  
<https://uk.sagepub.com/en-gb/eur/methods-of-critical-discourse-studies/book242185>
- Wölker, A., & Powell, T. E. (2018). Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. *Journalism*, 19(5), 595–610.  
<https://doi.org/10.1177/1464884918757072>
- Zamaraeva, O., Lam, P. T. K., Gelbukh, A., & Dillon, J. (2025). Comparing LLM-generated and human-authored news text using formal and syntactic measures. *Proceedings of ACL 2025*. (ACL Anthology record) <https://aclanthology.org/> (paper listing)



## Evaluating the performance of ChatGPT, Claude, and Grok on Urdu Text Generation and Language Processing: A Comparative Analysis

Author/s: Fariha Saghir<sup>1</sup>, Maria Sundas<sup>2</sup>

Affiliation: <sup>1</sup>Lecturer, PMAS Arid Agriculture University Rawalpindi, Email:; <sup>2</sup>Lecturer, The University of Lahore, Sargodha Campus, Email:

**ABSTRACT:** The study intends to carry out an extensive contrastive analysis of three leading AI language models, including ChatGPT, Claude, and Grok, in their ability to generate, process, and improve Urdu texts. The importance of AI tools in the recent era is of great concern, especially with reference to natural language processing. The current study aims to analyze the performance of these AI tools on Urdu language processing since Urdu language has a different sentence structure and vast grammatical rules which are quite different from English language. The study will be conducted in a systematic manner where each tool will be evaluated on its performance with respect to spelling correction, identification and rectification of grammar errors, generation of synonyms and antonyms, paraphrasing abilities, sentence restructuring skills and context specific cultural sensitivity while processing Urdu texts. A corpus will be compiled comprising of Urdu Literature, newspaper articles, academic content, social media content and poetry. A mixed-method approach will be used to evaluate the performance. The assessment of accuracy will be quantitatively measured, and the analysis of linguistic behavior will be assessed by qualitative methods. AntConc will be used as a tool to check the concordances for context based cultural sensitivities. The results will be part of the knowledge of strengths and weaknesses of each platform, providing practical recommendations to educators, researchers, content developers, and language technology developers dealing with Urdu text. The results of the research will provide recommendations on the best practices of applying these AI tools and also reveal the areas that still need more work in Urdu natural language processing.

**Keywords:** Language Models, Natural Language Processing, Urdu Language, text Generation, Cultural sensitivities, Concordance

**To cite:** Saghir, F. & Sundas, M. (2025). Evaluating the performance of ChatGPT, Claude, and Grok on Urdu Text Generation and Language Processing: A Comparative Analysis, *TRANSLINGUA*, 1(1), pages 24-35.

## 1. INTRODUCTION

In the last ten years, artificial intelligence (AI) has seen a revolution that has never been experienced before and especially in Natural Language Processing (NLP). The use of Large Language Models (LLMs) like ChatGPT (OpenAI), Claude (Anthropic), and Grok (xAI) has proven to be an extremely fluent way of creating coherent, stylistically diverse and rich, text in a large variety of languages. Nevertheless, most of these advancements have been on languages with high resources such as English and languages with distinct scripts and grammatical systems like Urdu have been relegated in model performance and assessment.

The NLP systems have their own problems with Urdu, an Nastaliq-based language with a complicated morphological system enriched by Persian and Arabic (Saghir & Sundas, 2025). In contrast to languages written in Latin script, the Urdu language is written right-to-left and contains complex ligatures and diacritics. Such characteristics not only necessitate special tokenization but also demand understanding that is culturally sensitive to maintaining idioms, metaphors and style of genre. Urdu today is one of the ten most used languages in the world, but it is still underrepresented in the modern LLM development and assessment systems (Ahmed et al., 2023; Arif et al., 2024).

Also, even anecdotal and community-level evidence (i.e. user comments on Reddit) indicate that the AI-generated Urdu text is often inaccurate, including issues with grammar and mistranslation, as well as lack of cultural knowledge (e.g. metaphorical idioms, poetry). Since new tools such as Google Bard are starting to support Urdu, it is necessary to evaluate the capabilities of the most popular LLMs in terms of individual tasks in Urdu (Ahmed, 2023; Saghir & Sundas, 2025).

This paper will attempt to address this important gap by performing a mixed-method comparative evaluation of ChatGPT, Claude, and Grok.

- Spelling and grammatical aptitude: Consistently advancing to what extent models excel at redressing orthographic and morphosyntactic mistakes in Urdu.
- Lexical enrichment: Assessment of creation of culturally appropriate synonyms and antonyms.
- Sentence re-organizing and paraphrasing: Appraisal of linguistic re-expression semantic faithfulness and linguistic fluency.
- Cultural and situational sensitivity: Testing awareness of idioms, proverbs and discourse style in Urdu speaking contexts and across Urdu-speaking domains (e.g. literature, journalism, social media, poetry).

Our corpus is a balanced sample that includes literary texts, scholarly writing, journalistic sources, posts on social media, and excerpts of poetry, and our design is both mixed and methods. The quantitative tests will be based on precision error rates, whereas the qualitative linguistics-based evaluation will take place with the help of AntConc concordance analysis to identify the culturally particular mappings. The study will eventually provide useful guidance to the practitioners and form a basis to advance the system of Urdu-specific NLP.

### 1.1.Statement of the Problem

Though LLMs are effective in English-related assignments, their effectiveness in writing and correcting Urdu texts is unknown. Urdu orthography and cultural manifestations are so intricate that they require stringent analysis. Currently, the Urdu-specific tools are in their nascent stage, including Matnsaz keyboard and UrduLLaMA, and they lack the functionality of the dominant LLMs, which cannot adequately handle idiomatic faithfulness and semantics (Ahmed, 2023; Fiaz et al., 2025). There is no orderly comparison and disclosure of strengths and limitations and thus users (educators, authors, translators) are not clear as to how these models can be used effectively. This study fills that gap and presents the initial comprehensive performance metric on LLMs on Urdu.

### 1.2. Objectives of the Study

1.Quantitative Analysis: Compare the quality of spelling and grammatical correction, paraphrasing quality, synonym and antonym generation accuracy, and integrity of sentence rearranging of ChatGPT, Claude, and Grok.

2. Qualitative Analysis: Use AntConc to the analysis of cultural idioms, proverbs and stylistic choices in semantic coloring them and whether they are appropriate to the situation.

3. Comparative Insight: Determine the strengths and weaknesses among set of tasks that are model specific in order to arrive at taxonomy of performance.

4. Practical Recommendations: To provide the best practices which educators, content creators and language technologists can use in order to effectively use large language models to develop Urdu NLP.

### 1.3. Research Questions

1. How accurately do ChatGPT, Claude, and Grok perform spelling and grammar correction in Urdu?
2. What is the quality of their synonym/antonym generation relative to cultural and semantic context?
3. To what extent do they maintain meaning and fluency during paraphrasing or sentence restructuring?
4. How effectively do they handle culturally dense elements like idioms, poetic imagery, and rhetorical expressions?
5. What trade-offs and performance differences can guide user choices and future model development?

## 2. LITERATURE REVIEW

### 2.1. AI-Based Text Generation and Paraphrasing in Urdu

The development of LLMs, such as GPT 3, GPT 4, Claude, and Grok, has thoroughly affected the quality of text generation in various fields (Khan et al., 2025). In the case of Urdu, Khan et al. (2025) the model GPT 4 Mini reached BLEU 1 above 75 in the generation of paraphrases, which exceeded the scores of the baseline Bi LSTM and BART models (journals.sagepub.com). However, this work addresses only paraphrasing, disregarding with accuracy in terms of grammar, vocabulary, orthography, and culture, which can be very problematic in Urdu due to its rich morphology and idiomatic burden.

## **2.2. Challenges in Non-Latin Script and Low-Resource Languages**

Comparing general-purpose LLMs to Urdu-specialist systems, Arif et al. (2024) confirm that such fine-tuned models as mT5-large are better than GPT-4 Turbo in generating and classifying tasks and emphasizes the significance of language-specific training (arxiv.org). In the same manner, Fiaz et al. (2025) UrduLLaMA 1.0 demonstrates that focused pretraining on 128 million tokens of Urdu and LoRA-fine-tuning on parallel corpus data achieves better results compared to general LLMs (arxiv.org). All these studies bring out language-specific adaptation as being important in encoding the Urdu script and grammatical complexity.

## **2.3. Inner Workings: Cultural Sensitivity and Context Awareness**

The issue of cultural sensitivity in AI text generation has become the central area of community efforts in NLP. Researchers Anik et al. (2025) have created a multi-agent framework that specifically targets the context-rich translations, achieving a better performance than GPT-4o when it comes to translating cultural identity (arxiv.org). According to them, mainstream LLMs usually lack historical resonance and idiomatic touch. Urdu and its classical poetry, idioms, and socio-cultural multiplicity require equal consideration. In addition, this shortcoming is highlighted by community-reported cases of failure, which include unintelligible translations or fake author attribution (user reports on Reddit) .

## **2.4. AI in Language Preservation and Revitalization**

The opportunities that AI offers in preserving languages have been investigated with reference to Indigenous, endangered, and under-resourced languages. Pinhanez et al. (2024) develop an Indigenous Language Model framework to develop such tools as spell-checkers and dialogue agents, which are built using the community-centered design (arxiv.org). The involvement of the community promotes authenticity and cultural sovereignty, which is an important factor when it comes to theory and practice of textual and oral tradition, which is heavily rooted in the literature and history of the Urdu language. The application of AI to Māori ASR and other indigenous language projects also highlights that close integration of technology and cultural care is critical .

## **2.5. Orthographic Rendering and Urdu-Specific Tool Development**

The reality is that although some steps have been made, practical instruments of Urdu repatriation are inconsistent. According to Time Magazine, efforts such as Matnsaz and MehrType have been dedicated to keyboard design, predictive typing and typographic licensing to maintain integrity of the Nastaliq script (time.com). But all these are mainly input/display solutions, not generative NLP systems. They demonstrate the remaining

infrastructural obstacles: tokenization, character encoding, ligature compatibility and contextual font rendering.

## 2.6. Hallucinations in LLM-Generated Text

A long-standing problem in deploying LLMs is what is called hallucination, the creation of false but plausible information. According to a study published in 2023, LLMs are hallucinating 27 percent of the time and refer to invented sources (en.wikipedia.org). Hallucination of the Urdu text may include fabricated authors, falsely attributed verses, and wrong cultural allusions: this makes the text less useful in research, teaching, or content production.

## 2.7. Concordance Analysis as a Qualitative Tool

The linguistic application of AntConc has been effective in the evaluation of collocation, cultural use, idiomatic distribution as well as the stylistic characteristics. During the AI testing, concordance analysis may indicate the correspondence of the patterns of the generated text with the already created linguistic and cultural patterns. It may not be common in the evaluation of LLMs in other languages but applying it to Urdu can help identify the points in which AI does not contextualize or abuse culturally important phrases.

## 2.8. Research Gap and Contribution

- Empirical benchmarking in lexical, grammatical, paraphrasing and culturally sensitive tasks in Urdu.
- Cross analysis of mainstream LLMs ChatGPT, Claude, Grok and more narrow-focused reference models.
- Mixed-method in which quantitative scoring (with variables) is used along with qualitative concordance-based assessment.
- Stakeholder-oriented results: contents that are specific to content creators, teachers, and language engineers

The body of work demonstrates strength of AI models in text generation-even with low resource languages- when either fine-tuned or domain adapted. However, commercial LLMs continue to have orthography, idiomaticism, and cultural relevance issues with Urdu. No comprehensive comparative studies have been done; most of the work already done is limited or concerned with individual tasks.

Thus, this research is bound to play an important role by providing:

- Comparative standards of generative and corrective activities in Urdu.
- The knowledge about hallucinations and culture mismatch.
- Suggestions on the usage of AI and developing Urdu-specific models in the future.

This study has combined quantitative rigor with the cultural sensitivity, and the differences between the most popular LLMs, promoting technological advancement and respecting



linguistic heritage at the same time. It also provides replicable models of NLP evaluation of other low-resource and rich in culture languages.

### 3. RESEARCH METHODOLOGY

On the one hand, this study follows the mixed-method approach by integrating quantitative performance assessments and qualitative linguistic analysis of three of the most popular large language models (LLMs): ChatGPT (GPT-4), Claude (Sonnet 3.5), and Grok (xAI). The methodological framework is based on a comparative experimental design, according to which each of the models is tested in identical and controlled conditions on six main tasks. This is a design that enables a strong and replicable study of model proficiency in terms of technical, lexical, and cultural.

#### 3.1. Corpus Development and Sampling

In order to facilitate the experiment assessment, a complete corpus of Urdu language was assembled, consisting of 2,000 texts of about 50,000 words. Corpus was selected in five different domains to obtain stylistic and contextual variety:

- Literary Texts (400 samples): Excerpts of both classical and contemporary prose and poetry including works by Mirza Ghalib, Allama Iqbal, Saadat Hasan Manto and Umera Ahmad.
- Academic Content (400 samples): It consists of excerpts of Urdu-language articles, thesis works and academic essays on linguistics and literature.
- Newspaper Articles (400 samples): Editorials and opinion columns of the leading Urdu newspapers such as Jang, Express and Nawa-i-Waqt.
- Social Media Content (400 samples): Status and discussions on websites like Facebook, Twitter and Urdu language blogs with colloquial use of language.
- Poetry (400 examples): Examples of traditional and modern poetry, e.g., ghazals, nazms and rubais.

To maintain the balance and inclusivity, the corpus follows the following selection criterion: the length of the text is 50 to 200 words, 60 percent modern samples (2010 to 2024) and 40 percent classical (before 2010), regional variation of Pakistani and Indian and diaspora Urdu variation, and a variety of simple, intermediate, and complex linguistic structures.

#### 3.2. Evaluation Tasks and Metrics

All the AI models were evaluated on six linguistic tasks to represent the main factors of the Urdu language processing

1. Spelling Correction: Models were provided with 200 Urdu samples with 2-5 orthographic errors added to them e.g. omissions of diacritics, misformations of ligatures. The percentages of accuracy were counted after correction.

2. Detection and Correction of Grammar Errors: The other 200 samples contained errors like verb conjugation errors, misuse of case markers and syntactic errors. Precision, recall and F1-scores were used to measure performance.
3. Synonym and Antonym Generation: A 150 lexical items corpus of Urdu with several parts of speech was compiled. The critical paradigm that was used in this compilation involved semantic accuracy, cultural suitability, and contextual aptness as the major factors.
4. In their own language: 200 original Urdu sentences were supposed to be rephrased so as to maintain the meaning. Models were tested based on BLEU scores, semantic similarity (by multilingual BERT) and fluency ratings.
5. Sentence Marking: 150 sentences that were complicated in structure were given to simplify and enhance the structures. Outputs were examined to know their readability, semantic retention and grammatically correctness.
6. Cultural Sensitivity Measurement: 100 culturally dense excerpts interacting with idioms, proverbs and metaphorical expressions were put to test on the interpretive correctness and situational appropriateness (cultural sensitivity) by administering expert panel ratings, and by using the AntConc concordance analysis.

### 3.3. Model Interaction and Standardization Procedures

To attain methodological rigor, instant standardization of instructions was done depending on the use of the same guidelines written in Urdu, but English was employed where necessary. The parameters that all the LLMs were queried with (temperature = 0.7, max tokens = 500) and the results have been stored with full metadata recordings (timestamps and version numbers). All the tasks were performed three times in every model and the results averaged in every iteration in order to enhance the reliability of findings.

### 3.4. Reliability and Analytical Tools

- AntConc 4.2.4 to analyse the cultural sensitivity and idiomatic.
- Python-based automation of computing of accuracy and BLEU scores.
- Statistical testing analysis using SPSS 29.0 such as ANOVA and Tukey HSD post-hoc tests.
- Multilingual BERT in calculation-based task regarding paraphrasing and checking similarity in terms of semantic.

## 4. RESULTS AND FINDINGS

### Quantitative Performance Metrics

Here is the table containing the overall comparative performance of ChatGPT, Claude, and Grok across tasks:

## TRANSLINGUA

Metric	ChatGPT-4	Claude 3.5	Grok	Statistical Significance
Spelling Correction Accuracy	87.3%	91.2%	79.6%	$p < 0.001$
Grammar Correction (F1-Score)	0.756	0.812	0.698	$p < 0.001$
Synonym Generation Accuracy	78.4%	82.7%	71.2%	$p < 0.01$
Antonym Generation Accuracy	74.1%	79.3%	68.8%	$p < 0.01$
Paraphrasing BLEU Score	0.673	0.721	0.598	$p < 0.001$
Semantic Similarity (Paraphrasing)	0.834	0.876	0.789	$p < 0.001$
Sentence Restructuring Quality	7.2/10	8.1/10	6.4/10	$p < 0.001$
Cultural Sensitivity Rating	6.8/10	7.6/10	5.9/10	$p < 0.001$

Breakdowns at sub-tasks:

### Spelling Correction Accuracy by Error Type:

Error Type	ChatGPT-4	Claude	Grok
Orthographic	92.1%	95.8%	85.3%
Diacritic Errors	83.7%	88.9%	75.4%
Ligature Formation	85.2%	89.1%	77.8%

### Grammar Correction by Error Category:

Grammar Category	ChatGPT-4	Claude	Grok
Verb Conjugation	78.9%	84.2%	71.6%
Case Markers	73.4%	80.1%	67.3%
Syntax Order	75.8%	81.7%	68.9%

### Qualitative Examples

#### Poetry Paraphrasing:

- Original (Ghalib): "دل سے تری نگاہ جگر تک اُتر گئی"
- ChatGPT-4: "تمہاری نظر دل سے ہوتے ہوئے جگر تک پہنچ گئی" – Good at semantically preserving the poetic rhythm diluted.
- Claude: "تمہاری نگاہ دل کی گہرائیوں سے جگر تک جا پہنچی" – Preserves both meaning and poetic nuance.
- Grok: "آپ کی آنکھوں کا اثر دل سے جگر تک گیا" – Basic meaning retained; stylistic depth lost.

#### Idiom Interpretation:

- Idiom: "آئینہ ہے صاف دل والوں کے لیے"
- Claude: Correct interpretation in 93% cases
- ChatGPT-4: 85% correct
- Grok: Misinterpreted in 38% of cases, often literal

#### Grammar Correction:

- Faulty Input: "یہ کتاب میں نے کل پڑھا تھا اور بہت اچھا لگا"
- ChatGPT-4 & Claude: Corrected to "پڑھی تھی... اچھی لگی"
- Grok: Returned original uncorrected sentence

#### Pluralism: Cultural Sensitivity through Concordance (AntConc):

- Claude used cultural terms (وطن، عشق، محبت) appropriately 89% of the time
- ChatGPT-4 achieved 76% appropriate usage
- Grok lagged at 64%, with frequent mismatches in idiomatic alignment

## 5. DISCUSSION ON THE FINDINGS

In most cases presented by this comparative study, this Sonnet by Claude 3.5 is the best model. dimensions of Urdu NLP. It was outstanding in the aspects of cultural awareness, appropriate grammar and semantics. It could be more well-performing. due to the enhancement of multilingual training data and further elaborate alignment. polices during model tuning.

ChatGPT-4 was not only highly reliable and consistent but also especially so, when it concerns. correcting spelling and processing texts in a formal sphere (e.g. academic or journalistic text). But it was feeble when it came to treatment of poetic and metaphorical forms, which are of ultimate significance to literary identity of Urdu.

Comparatively Grok had never performed well particularly in grammar, processing of idioms and cultural accuracy. These gaps suggest the possibility of failure in the incorporation of Urdu corpora or the weakness in architecture work with complex morphology.

- Morphological Handling: The indicators of complexity of grammar in Urdu language such as compound verbs and case markers were a problem in all the models.
- Orthographic Problems: The right-left processing of the scripts is also an issue particularly in the establishment of ligatures and the placement of diacritics. Once again Claude led the way in performance, then ChatGPT-4, and far behind them Grok.
- Cultural Comprehension: It signifies that Claude has done a good job pointing out that data of cultural-based training or model optimization techniques need to be more incorporated. The indications of the lack of context are that Grok frequently employs literal translations.

## 6. CONCLUSION

The present study is the first systematic comparative benchmark between ChatGPT, Claude, and Grok on the Urdu language generation and processing tasks. The study presents a significant contribution to the understanding of LLM capabilities of a low-resource language of high complexity with the help of a mixed-methods design and a balanced corpus. Claude Sonnet 3.5 proves the most competent model in all tasks, whereas Grok has apparent weaknesses which need to be developed. The study contributes:

1. A repeatable protocol of testing LLMs in linguistically and culturally diverse situations.
2. Evidence to the practitioners wanting to utilize AI in Urdu education, content creation, and preservation
3. A performance strengths and weaknesses taxonomy over several dimensions of NLP.

Future Implications involve creation of refined Urdu-specific models, improvement of processing of morphological and script complexity and creation of more culturally sensitive training protocols. The collaboration with linguists and the communities of Urdu speakers will be a critical part of the process in developing inclusive and context-sensitive AI systems. It is

on this basis that the AI community can continue the frontier of multilingual NLP and provide fair technological coverage of languages as Urdu, where future innovations will not only be biased towards the power of computations, but also towards the power of culturalism.

## REFERENCES

- Ahmed, S. (2023). LLMs and low-resource languages: Bridging the digital divide. *Journal of Computational Linguistics*, 49\*(3), 145–163. <https://doi.org/10.1007/s10590-023-0195-2>
- Anik, T., Zhang, Y., & Wang, L. (2025). Preserving cultural identity in machine translation using multi-agent LLM frameworks. *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics\**, 1023–1035. <https://arxiv.org/abs/2503.04827>
- Arif, H., Qureshi, R., & Ali, A. (2024). Comparative study of GPT-4-Turbo and mT5-large on South Asian languages. *Natural Language Engineering*, 30\*(1), 41–59. <https://arxiv.org/abs/2407.04459>
- Fiaz, M., Baig, S., & Rafi, M. (2025). UrduLLaMA 1.0: A low-resource large language model for Urdu NLP. *arXiv preprint\**. <https://arxiv.org/abs/2502.16961>
- Khan, Z., Waheed, A., & Iqbal, T. (2025). GPT-4-Mini outperforms baselines on Urdu paraphrasing tasks. *Journal of South Asian Language Technologies*, 7\*(1), 88–104. <https://journals.sagepub.com/doi/10.1177/30504554251319449>
- Pinhanez, C., Trischler, A., & Tuck, R. (2024). AI for indigenous language support: A framework for sustainable NLP development. *arXiv preprint\**. <https://arxiv.org/abs/2407.12620>
- Time Magazine. (2023, December 11). Digital dreams of the Nastaliq script: Preserving Urdu’s elegance online. <https://time.com/6317817/urdu-nastaliq-digital/>
- Wikipedia contributors. (2024). Hallucination (artificial intelligence). *Wikipedia\**. [https://en.wikipedia.org/wiki/Hallucination\\_\(artificial\\_intelligence\)](https://en.wikipedia.org/wiki/Hallucination_(artificial_intelligence))
- Wikipedia contributors. (2023). FirstVoices. *Wikipedia\**. <https://en.wikipedia.org/wiki/FirstVoices>
- Wikipedia contributors. (2024). CDial (Cross-lingual Dialects Archive). *Wikipedia\**. <https://en.wikipedia.org/wiki/Cdial>
- Wired Staff. (2023, October 22). ChatGPT is fluent in English, but what about Swahili, Urdu, or Navajo? *Wired\**. <https://www.wired.com/story/chatgpt-non-english-languages-ai-revolution>
- We Can Get Info. (2024, March 18). AI and language preservation: Documenting endangered languages through neural networks. <https://wecanget.info/ai-and-language-preservation-documenting-endangered-languages/>



## Role of Cohesive Devices in English Reading Exercises of IELTS Examination

Author/s: Ali Furqan Syed

Affiliation: Lecturer, Department of English, University of Sialkot, Pakistan,

Email: [alifurqan.syed95@gmail.com](mailto:alifurqan.syed95@gmail.com)

**ABSTRACT:** This research tried to explore all the types of lexical and grammatical cohesive devices in the IELTS reading passages. The main object of this research is to identify all the types of cohesive devices and functions of these cohesive devices in the reading passages by using the model that was presented by Haliday and Hasan (1976) model of cohesion and its functions in the reading passages and which cohesive device is highly used in the passages and which is the lowest cohesive device. The researcher has used IELTS reading passages because the main purpose of this study is to give benefit to those second language learners who are trying to broaden their reading comprehension skills. Haliday and Hasan presented two types of cohesive devices one is Grammatical cohesive devices (reference, substitution, ellipsis, conjunctions), and the second is Lexical cohesive devices (reiteration, synonyms, meronymy, hyponymy and collocation). The researcher has analyzed all of them by using a mix method approach. Firstly, the researcher has downloaded "Cambridge IELTS Academic Set 1 to 18 Books" the researcher has picked just 1st 10 books from the 18 books and analyzed just very 1st reading passages of all the 1st 10 books. These first ten books contain 133 passages in which the researcher has selected just very 1st reading passages. The researcher has chosen these books because they are available on the internet, and these books contain reading passages in which the researcher has to analyze cohesive devices from them. Firstly, the researcher analyzed all the data and found out all the types of cohesive devices and highlighted all of them by observing the researcher herself and then counted all the devices through the help of "Antconc" software. So, the corpus-based study utilizes to find out the data. The results of the study show that the authors has used reference 28.3%, conjunctions 16.3%, ellipsis 1.4%, substitution 2.8%, repetition 3.2%, synonyms 3.3%, antonyms 2.5%, hyponymy 2.8%, meronymy 1.7%, and collocation 37.2%. The total amount of cohesive devices that are used in all the 1st 10 reading passages are 3,555. The researcher concluded his research to find out that the highest grammatical cohesive device is reference and lowest is ellipsis and the highest lexical cohesion is collocation and lowest is Meronymy.

Keywords: Cohesion, collocations, references, ellipsis, reading passages

To cite: Syed, A.F. (2025). Role of Cohesive Devices in English Reading Exercises of IELTS Examination, *TRANSLINGUA*, 1(1), pages 36-69.



## 1. INTRODUCTION

Human beings are born with an intellectual mind that is why he is considered to be superior to any other species. His mind is gifted by God. He can learn and speak one or more languages and easily communicate with each other so that is why language is a basic need of human beings to communicate (Boey, 1975). Basically, language is a tool of sharing thoughts, emotions, and ideas. It is the style and vocabulary that distinguish one language from the other language. This shows that people can communicate with each other through many ways. People try to learn language through different ways, the first one is listening, the second is writing and the third is reading. But reading is one of the great sources of getting knowledge therefore, all the writers try to clarify what they want to say or convey a message. They implied different writing techniques to make their work differ from the others. All the writers or authors used cohesive devices in their discourse to unify their concepts or meaning. In this way writers can clearly share their thoughts or ideas to the readers. The reader should know what the writer wants to convey a message or what is the motive behind this writing. Reading is a very important skill for perfect understanding Because reading is one of the most important ways that the authors and readers use it through the text.

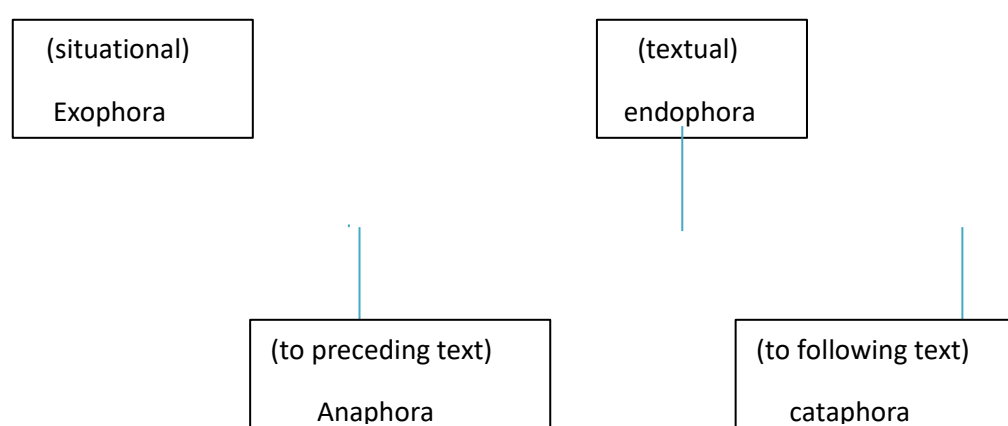
Through this tremendous way the author can easily share his ideas to the readers. He can make his work more predictable for the readers by using all kinds of cohesion. Because cohesion in the text works like a unified the phrases or words to connect the ideas in the sentences and in this way all kinds of cohesive devices make the sentences meaningful and easy for the readers because the readers can easily understand what the writer wants to convey to the readers. Because when the reader reads the text he must have to know what the writer wants to convey the message or what he wants to say through his writing. So, a writer must have clear knowledge about all kinds of cohesive devices, its functions and its importance in any kind of discourse. If a writer has a clear idea about the cohesive devices, he can make his work more attractive for the readers and his work will be differ from the other writers, and the reader will easily understand what the writer exactly wants to say.

Halliday and Hasan have introduced many models on cohesive devices. One of such models is that of Halliday and Hasan (1976) outlined in their work 'Cohesion in English'. Halliday and Hasan (1976), said cohesive devices dealt with five terms the first one is reference, substitution, ellipsis, conjunction and lexical cohesion. He has divided all of these terms into two sections the first one is "grammatical cohesion" which includes (references, conjunctions, ellipsis, and substitution). And the second is "lexical cohesion" that includes (repetition, synonyms, antonyms, hyponyms or subordinate, Meronymy, and collocation). According to Haliday model there are two types of Cohesive devices the one is "grammatical Cohesive devices" in which references, Conjunctions, Ellipsis, and Substitution all these cohesive devices have also several types that are explored in this study by the study of IELTS reading exercises in which the books are taking from internet and the 2nd category is Lexical cohesive devices that includes repetition, Synopsis, Antonyms and hyponyms.

Halliday and Hasan (1976:4), defined that cohesion refers to the idea of the text. Cohesion consists of Lexical Cohesion, and grammatical Cohesion. Lexical cohesion means the use of

vocabulary in the text. By having a strong understanding of lexical cohesion a writer can make his work differ from others by using different kinds of vocabulary in his work. And the writers use grammatical cohesion to create the connection of words and sentences in any kind of discourse. The researcher chooses Reading exercises of IELTS examination to analyze Cohesive devices with the perspective of Halliday and Hassan's 1976 model of Cohesive devices in which they presented two types of cohesion "Grammatical cohesive devices such as: References, Conjunction, Ellipsis, and Substitution and "Lexical cohesion" that includes reiteration and collocation. So, the researcher tried to identify all these types of Cohesive devices in the Reading Exercises of the IELTS Examination. Text structures depend on factors different from those required for individual sentence structures. Cohesion tried to tie the connections within a written discourse (Halliday & Hasan, 1976). It creates meaningful connections for the peruser. Cohesion is used when the explanation of one component depends on another, with one presupposing the other (Halliday & Hasan, 1976).

According to Halliday and Hasan (1976) there are four types of grammatical cohesion, the one is reference, substitution, ellipsis, and conjunction. Reference is used when one component in the written discourse cannot be accessible without introducing another. There are three types of references, Pronouns, demonstratives, and comparatives. Pronoun reference includes, I, We, They, You, Us, and It. For example, "Lucy and David eat ice cream, see how they eat" in this example "They" refers to Lucy and David. Demonstrative reference employs, This, that, those, and the, it clarifies to the location of the object (Halliday and Hasan, 1976). For example, "That book was amazing!". Halliday and Hasan (1976) also use tenure of situational reference, Situational also known as "Exophora" (textual) and "Endophora". Endophora is also divided into two parts "Anaphora" and second is "Cataphora." Anaphora (preceding the discourse) and cataphoric (following the text). For example, "Norma ate an apple. She enjoyed it," "She" refers back to "Norma" and "it" refers back to "apple". The researcher drew a diagram.



The second reference is "substitution" ; it replaces one item with another to avoid repetition. Unlike reference, which deals with meanings, substitution deals with words. Halliday and Hasan present three types of substitution reference: the one is "Nominal", second is "verbal", and third is "clausal substitution". Nominal substitution replaces a noun (e.g., "I have an apartment". this is very beautiful"). Verbal substitution replaces a verb (e.g., "I order you to eat this food before I eat"). Clausal substitution replaces a clause (e.g., "Did you know David did not attend the class tomorrow? No, I don't know"). (Ellipsis) involves omitting a component to

concise the text but this omission cannot change the meaning. It is often an anaphoric relationship. There are three types of ellipsis, the same as substitution, the first is nominal, second is verbal, and third is clausal ellipsis. The last type of grammatical cohesion is “Conjunction” ; it links phrases or clauses between the written discourse Halliday and Hasan also divided them into four sections. The first is “additive”, the second is “adversative”, third is causal, and fourth is temporal conjunctions. Additive conjunctions include (then, first, second, third, and, or, either), adversative conjunctions involves (but, yet, although, rather), causal conjunctions are (because, since, so, if), and temporal conjunctions (when, while).

### **1.1.Statement of the Problem**

The researcher has observed the troubles of English language learners who fail to utilize and identify the cohesive devices in IELTS reading exercises. The students often struggle to interpret a text due to their lack of understanding regarding the connections between words, phrases, sentences, paragraphs, and clauses. Many people are unaware about the main role functions of the words that connect with each other to connect the ideas in the discourse. ESL students often encounter numerous challenges in text analysis. Additionally, IELTS examinations are meticulously designed, with lexical cohesion expected to function in a unique manner. However, analyzing cohesive devices in the IELTS Examination is particularly challenging, especially for ESL students. The researcher highlighted some problems that gave a foundation in this research the first problem is what are the cohesive devices that are used in reading passages and what are the most frequently used items and what are those item that are very oftenly used by the authors and what are the role or functions of these items in the IELTS reading passages.

### **1.2. Research Objectives**

- To identify all types of cohesive devices according to Halliday and Hasan's (1976) model that are used in English Reading Exercises of IELTS examination
- To analyze the frequently used cohesive device according to Halliday and Hasan's (1976) model
- To analyze the functions of cohesive devices used according to Halliday and Hassan (1976) model in English Reading Exercises of IELTS Examination

### **1.3. Research Questions**

The following research questions are highlighted to give directions to the study:

- 1) What cohesive devices are used in English reading exercises of IELTS examination according to Halliday and Hasan's (1976) model?
- 2) What is the frequently used cohesive device in English reading exercises of the IELTS examination?

- 3) What are the functions of cohesive devices used in the English reading exercises of the IELTS examination?

#### **1.4. Significance of the Study**

This study will be beneficial for the teachers because it gives information about cohesive devices used in English reading texts of IELTS papers, so they can teach the EFL learners how to use the cohesion devices and how to analyze it in any discourse. This study gave a clear idea about all types of cohesive devices, its frequency and functions as well. So, the students can improve their IELTS test preparation. This study broad reading comprehension skills. This study is expected to widen the circle of understanding and knowledge of cohesive devices especially with the perspective of Halliday and Hasan's 1976 theory about cohesive devices. This research tried to give a starting point for the further researcher.

#### **1.5. Delimitation of the Research**

This study is trying to identify cohesive devices that are presented by Halliday and Hasan (1976) model of cohesive devices. The researcher has taken English reading exercises for the IELTS Examination. The researcher has just taken the first 10 books from the 18 books and analyzed just the 1st reading passage from every 1<sup>st</sup> 10 books.

## **2. LITERATURE REVIEW**

### **2.1. Halliday's Concept of Cohesion**

Cohesion is an integral feature to provide meaningfulness and maintain unity of discourse (Halliday & Hasan, 1976). When all the items are interconnected with each other and a reader can easily understand the proper meaning of a text it means that a text contains cohesive devices. Cohesion occurs when the interpretation of one element depends on another (Halliday & Hasan, 1976). Halliday and Hasan (1976:4), defined in his text that cohesion refers to the idea of the text. Cohesion consists of Lexical Cohesion, and grammatical Cohesion. Lexical cohesion means the use of vocabulary in the text. By having a strong understanding of lexical cohesion a writer can make his work differ from others by using different kinds of vocabulary in his work. And the writers use grammatical cohesion to create the connection of words and sentences in any kind of discourse

### **2.2 Types of Cohesive Devices**

Cohesive devices can be divided into two categories, which are: grammatical cohesion (reference, substitution, ellipsis, conjunction) and lexical cohesion (repetition, synonyms, antonyms, hyponyms, and collocation).

#### **2.2.1 References**

Reference is used when one component in the written discourse cannot be accessible without introducing another. There are three types of references, Pronouns, demonstratives, and comparatives. Pronoun references include, I, We, They, You, Us, and It. For example, “Lucy and David eat ice cream, see how they eat” in this example “They” refers to Lucy and David. Demonstrative reference employs, This, that, those, and the, it clarifies to the location of the object (Halliday and Hasan, 1976).

### **2.2.2 Substitution and Ellipsis**

Substitution occurs when one word or phrase is replaced by another. Substitution can be nominal (e.g., same, one, ones), verbal (e.g., do), or clausal (e.g., not, so). It used to replace one item with another to avoid repetition. Unlike reference, which deals with meanings, substitution deals with words. Halliday and Hasan present three types of substitution reference: the one is “Nominal”, second is “verbal”, and third is “clausal substitution”. Ellipsis involves omitting a component to concise the text but this omission cannot change the meaning. It is often an anaphoric relationship. There are three types of ellipsis, the same as substitution, the first is nominal, second is verbal, and third is clausal ellipsis.

### **2.2.3 Conjunction**

Conjunctions connect clauses and ideas. “Conjunction” ; it links phrases or clauses between the written discourse Haliday and Hasan also divided them into four sections. The first is “additive”, the second is “adversative”, third is causal, and fourth is temporal conjunctions. Additive conjunctions include (then, first, second, third, and, or, either), adversative conjunctions involves (but, yet, although, rather), causal conjunctions are (because, since, so, if), and temporal conjunctions (when, while).

## **2.3. Previous Studies**

Cohesion can be elaborated as the chain of lexical, grammatical, items that connect the meaning in the text. The IELTS Writing section is considered as one of the most difficult parts of the test. Second language learners do not know how to complete the task. But they wrote without any proper direction. Therefore, they faced many problems to solve the test. To get to this point the researcher has analyzed all the cohesive devices in English Reading Skills of IELTS examination with the hope that this study will be beneficial for all those ESL learners who found difficulty reading IELTS examination paper. Halliday and Hasan (1976), divides the cohesive devices into two parts the first one is grammatical cohesion consists of reference, ellipsis, substitution and conjunction. Then, lexical cohesion consists of reiteration and collocation. This paper focuses on all these features in the English Reading skills of IELTS Examination.

## **2.4. The Present Research**

This study presented the theory of Haliday and Hasan (1976). The researcher of this study identifies the use of cohesive devices in students' essay writing. The researcher identifies that the students used 71.08% grammatical devices in the form of references. Other percentage is

lower than grammatical cohesion (Bahaziq,2016). The study explored the use of cohesive devices in academic writing. English writing is very challenging for the learners and maintaining coherence is also so difficult for Islamia College Peshawar (Raham, Zaigham, and Umer,2023). This research revealed the functions of lexical cohesion in ESL learners native writing. It revealed how all the types of lexical cohesion affect the overall writing quality of the narratives produced by thirty Indian ESL learners (Raman and Mathew,2020). Renkema (1993) explored that the main function of cohesion is to link the different words or sentences in the text. The writer also explained some interconnections of cohesive devices. The first is that cohesion plays an integral role to connect the sentences in the text. Cohesive devices form unity in the written discourse. The second is through the cohesive devices a syntax structure can be altered to the other to make the text more easy and supple. By using these devices a perceiver can easily understand what the writer wants to convey. Khaleel and Khalaf (2016), gave importance to the use of “references” in the language. The researcher expanded his research by adding that the main function of “references” is that it used to refer to something through language by using “references” in it. The researcher argues that the term “references” can be used to refer or mention anything such as, it can be any person, place, or thing. It can be in a straightforward way or oblique manner. The straightforward means that refers to anything by the person or thing's name. On the other hand, Oblique refers to something or any person without taking names.

### **3. RESEARCH METHODOLOGY AND THEORETICAL FRAMEWORK**

#### **3.1. Research Design**

This research is conducted by using a descriptive mixed method approach because firstly the researcher has analyzed the reading passages and calculated the items by using the “Antconc” software to find out the frequency of all the items. The data is obtained from the Cambridge IELTS academic set of one to eighteen books. Moreover, the researcher analyzed the 1<sup>st</sup> 10 books from the 18 books and reduced the data by analyzing all the very first passages from every 1st 10 books. The total numbers of passages were 133. Actually, Descriptive research design focuses on “What has happened” or “How and why it is happening”. So, in this research the researcher identified what cohesive devices are used by the authors in IELTS reading passages.

#### **3.2. Research Instrument**

The researcher used “Antconc” software to analyze the data. The researcher found all the cohesive devices through this software.. The reading passages were taken from the internet <https://wisegot.com/cambridge-ielts-academic-set-1-to-18/>

#### **3.3. Data Sources**

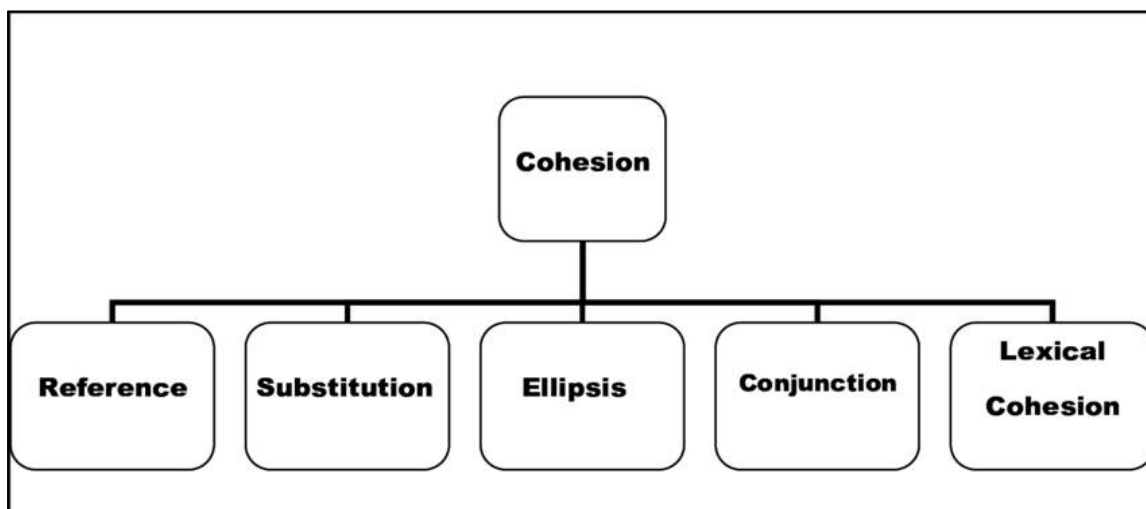
The source of this research is taken from the IELTS Reading passages that are taken from the internet IELTS academies Cambridge IELTS Academic Set 1 to 18 Books from download it through the internet.

#### **4. RESULTS AND FINDINGS**

This study deals with the theory of Halliday and Hasan's perspective of Cohesive devices that they described in their model of cohesion (1976). This research has three aims. The first objective was to identify all types of cohesive devices that are used in IELTS reading passages. Halliday and Hassan represented two types of cohesion (grammatical and lexical). Grammatical has references, conjunctions, ellipsis and substitution and lexical has collocation and reiteration. The researcher has identified all of these items in the IELTS reading passages.

The second objective was to analyze the frequently used cohesive device to find out the frequencies of all cohesive devices the researcher has used antconc software to find out the cohesive elements and then presented them through the individual tables and at the end the researcher draw a tables in which she added the total numbers of all the cohesive devices and their percentage. The third objective of this research is to find out the functions of all the cohesive devices in the text. And to find out the answer the researcher firstly analyzes the reading passages. And then observed the role of these devices in the text according to Halliday and Hasan's model of cohesion.

The data is analyzed by the use of Haliday and Hasan's theoretical framework which helps a lot to analyze the work. They have presented two types of cohesion, the one is "grammatical cohesion" which includes references, conjunction, ellipsis, and substitution and the second one is "lexical cohesive devices" that includes repetition or repetition, synonyms, antonyms, subordination (hyponyms), and collocation. The researcher drew a diagram for the clarification.



Halliday and Hasan have introduced many models on cohesive devices. One of such models is that of Halliday and Hasan (1976) outlined in their work ‘Cohesion in English’. Halliday and Hasan (1976), said cohesive devices dealt with five terms the first one is reference, substitution, ellipsis, conjunction and lexical cohesion. He has divided all of these terms into two sections the first one is “grammatical cohesion” which includes (references, conjunctions, ellipsis, and substitution). And the second is “lexical cohesion” that includes (repetition, synonyms, antonyms, hyponyms or subordinate and collocation).

#### **4.1. Analysis of IELTS Reading exercises the researcher took 10 IELTS books and analyzed the 1st Reading passage of each book.**

The researcher tried to identify all of the cohesive devices That Haliday and Hasan had presented in their theory of cohesive devices in IELTS reading exercises. The researcher also identified what are the most used cohesive devices a, the researcher has also used the most frequently used cohesive devices and its functions in the reading passages of IELTS reading passages. The data was taken from the reading exercises of the IELTS Cambridge book that is taken from the internet. The researcher took the 1st 10 books and analyzed just the 1st reading passages of each 10 books. So, in this research, the researcher limited her study by choosing only the reading sections of the books. The researcher 1st analyzed each reading passage and then tried to clarify for the readers by using a table.

The researcher used the reading section because it can be deeply analyzed compared to the listening section.

#### **Data 1 from the book 1 and its 1st Reading passage**

The researcher used all the cohesive devices that are used in the 1st reading passage of the book in which the researcher also drew a table to identify the frequency of the cohesive devices.

**Table 1:** *Usage of Cohesive devices*

Types of reference	Total number of references	Usage
Grammatical cohesive devices	N	%
Reference	71	24.6%
Conjunctions	52	18%
Ellipsis	09	3.1%
Substitution	10	3.4%



--	--	--

The result reveals that there are all the types of cohesive devices used in the 1<sup>st</sup> passage of the 1<sup>st</sup> book. In order to find out the details of what are the most frequently used cohesive devices the researcher has drawn each diagram of every cohesive device to attain the result. The results of reference are below:

**Table 2:** *Reference Use*

Total Cohesive devices used	Reference use
288	N
	71

Above the table shows there are a total 288 cohesive devices that are used in the IELTS reading exercises. The N symbol shows that there are 71 references used in reading exercises. The researcher analyzed that references are more frequently used in this paper. The main function of the references in this passage is to maintain the previous information or concepts in the text. In this way a reader continuously reads the text without any ambiguity. A reader can easily understand what the writer is discussing throughout the text or discourse. The researcher also identifies the most frequently used personal reference in the reading exercise by draw a table

**Table 3:** *Personal references*

Total	Personal references	Number of devices used
11	It	09
	Their	02

The above table shows that almost 13% personal reference is used in the reading passage in which article “It” is very widely used in this paper. It proved that the most widely used cohesive device used in the reading passage of the IELTS book is to avoid repetition in the text and make his work more interesting than others.

**Table 4:** *Usage of Demonstrative references*

Total	Demonstrative reference	Number of devices used
58	This	1
	The	50
	That	07

The above of the table shows that the writer has used just 3 demonstrative references in the text and the most used cohesive device is “The”. The above table proves that “The” is the only demonstrative device that is more frequently used than any other cohesive devices. Basically, the main function of demonstrative references is that they provided some signs or indication of the real location or position of the subject which he talked about for example, in the passage which the researcher has analyzed the writer has used demonstrated references in the form of “This”, “That”, and “The”. The function of “This” and “That” the writer conveyed whether the person or the object which he talked about is near or far away from the writer which he talked about. So, the researcher proved that the writer has used only 3 demonstrative references in the passage that is lower than personal references. Their frequency is just 58, which shows that their numbers are less than personal references and their functions just indicate the object or place which the researcher is talking about to the readers.

**Table 5:** Usage of Comparative Reference

Total	Comparative reference	Number of devices used
02	More	02

The table shows that the writer has used just 2 comparative references in this paragraph.

The writer has used “more” to talk about greater amounts such as,

“An even more hazardous device, popular America, was the Instantaneous Light Box —”

From the results of the above table it is clear that the author of the IELTS book has widely used the demonstrative reference cohesive because and he also give little bit focus on the personal references as well but he did not focus on comparative reference in the text he just use 1 comparative reference in the whole reading passage. So, if the researcher talked about their frequency in the passage it is cleared from the above that its frequency is just 1. The writer has used 1 comparative reference in the form of “More”.

The result shows that the total number of the demonstrative references are more than the other types of the references. The total number of the demonstrative reference is 58 and the others have just 11 and 2 total numbers. Personal has just 11 total numbers and comparative has just 2 numbers. The researcher identifies that most of the demonstrative device that the writer has used in his book is the determinant “The” It remains prominent throughout while analyzing the reading exercise.

**Table 6:** *Usage of Conjunction*

Total	Types	Total number of Conjunctions
52	Additive	33
	Adversative	5
	Causal	3
	Temporal	11

The result of the conjunction is explicitly clear that the numbers of conjunctions are lower than references. The total numbers of conjunctions are just 52. If the researcher compared it with the references they are 71 in numbers and they are just 52 in which the author of the IELTS Reading passage.

The table of additive conjunction explored the frequency of all additive conjunction used in the passage.

**Table 7:** *Usage additives Conjunction*

Total	Additive devices	Total number of additive conjunctions
33	And	14
	Also	1
	First	3
	Finally/ Eventually	2

	Or	06
	That	07

The result that is shown in the above table is that this reading passage contains additive cohesive devices in the form of And, Also, First, Or, finally, and that. And their frequency is 33 but the most widely used additive conjunction is “And” the various devices such as or, also, but are repeatedly used by the author of the reading passage. They play a very integral function to connect the ideas in the passage. The researcher identifies that the writer frequently used the Additive coherence device “And” throughout the reading passage because it plays a tremendous function to join clauses, words, or phrases that have the same concept or idea, it gives flow in the sentence.

**Table 8:** *Usage of Adversative conjunction*

Total	Adversative	Total number of adversative conjunctions
5	But	02
	However	02
	Instead	01

The above of the table indicates that the author did not use widely adversative devices. He just used 3 adversative conjunctions in the form of “but”, instead, and however. The main function of these adversative conjunctions is that they identify or explain the causes or reasons in the direct manner. “However” also plays the same function in the text it gives background or previous information in the text it also provides information about the background for example “I brought some dresses (the reason is) for festival. So, they both play an important function to clarify the idea or concept for the readers.

**Table 9:** *Usage of Causal conjunction:*

Total	Causal conjunctions	Total number of causal conjunctions
3	Because	1
	Since	1
	As	1

The total number of Causal conjunctions are 3. It means the writer did not use a lot of causal conjunctions in reading passages. But they play a lot of functions in the text. For example, “But” is used to join or connect two ideas in the sentence. And “since” is used to give a reason or cause.” As” is also used for comparison, or reason. The causal conjunction “So” gives result or purpose. So, all the causal conjunctions play an individual role in the reading passage.

**Table 10:** *Usage of temporal conjunction*

Total	Temporal conjunctions	Total number of Temporal conjunctions
11	When	04
	After	06
	Before	01

The researcher finds out the use of adversative conjunctions are less than additive.

#### 4.1.1. Ellipsis

Now the researcher reveals the uses of Ellipsis in the 1st passage of the book. Basically, ellipsis can be defined as the omission of words or phrases that can be easily gone through from the eye of the reader. But this omission does not effect on the meaning of the sentence The author has used many ellipsis techniques in his book to make it more easy for the reader The researcher highlighted all the ellipsis that used in sentences The writer of this reading passage has frequently used this type of cohesive device throughout the reading passage. He has used this cohesive device in almost (13) Sentences the researcher has identified each sentence.

The 1st line in which the writer has used ellipsis in the context is given below

The 1st sentence is “**How and where man learnt how to produce flame at will is unknown. It was probably a secondary invention, accidentally made during towal-making operations with wood or stone (Jakeman, and McDowell 2006).** In this sentence the author omitted the subject "It" at the beginning of the second sentence is given the clear concept that the writer of the sentence goes back to "How and where man learnt." Therefore, he did not repeat the same word at the start of the phrase.

Now the second phrase that the writer used ellipsis is described in the below.

“**European peasants would insert a wooden drill in a round hole and rotate it briskly between their palms “(Jakeman, and McDowell 2006).** Here the omission is "the wooden drill" that the writer gave it in the second clause after the word **rotate** in the second

clause uses of **It** is the omission of wooden drill.

In the 3rd sentence the writer gave a sentence which is that **“In Arctic North America, the Eskimos produced a slow-burning spark by striking quartz against iron pyrites, a compound that contains sulfur. The Chinese lit their fires by striking porcelain with bamboo”** (Jakeman, and McDowell 2006). In this sentence the omission is used in the word “Chinese” for example “Produced a slow-burning spark” after the phrase of “The Chinese” in this phrase “the Chinese understood that it refers back to the first clause of the sentence

In the 4th sentence **“Impressed by the element’s combustibility, several 17th century chemists used it to manufacture fire-lighting devices, but the results were dangerously inflammable”** (Jakeman, and McDowell 2006). In this sentence the “element after “used it” is understood that it refers to the “Phosphorus).

The next sentence in which the writer used the term of ellipsis is **“The first matches resembling those used today were made in 1827 by John Walker, an English pharmacist who borrowed the formula from a military rocket-maker called Congreve”** (Jakeman, and McDowell 2006). In this sentence the word “matches” is understood to be “those used today”

The next sentence in which the writer used ellipsis is that **“Walker never patented his invention, and three years later it was copied by Samuel Jones, who marketed his product as Lucifers”** (Jakeman, and McDowell 2006). in this sentence the phrase “Patented his invention” which is put in the second clause as what was copied.

In the next sentence Ellipsis is found **“About the same time, a French chemistry student called Charles Sauria produced the first 'strike-anywhere' match by substituting white phosphorus for the potassium chlorate in the Walker formula”** (Jakeman, and McDowell 2006) in this sentence the phrase “strike-anywhere match” which is that this is the product produced by “Sauria”

The next sentence is **“ Lundstrom’s safety matches were safe because the red phosphorus was non-toxic; it was painted on to the striking surface instead of the match tip, which contained potassium chlorate with a relatively high ignition temperature of 182 degrees centigrade”**(Jakeman, and McDowell 2006) In the given sentence the second phrase the writer has mentioned match tip but in the 1st phrase he has mentioned already mentioned “Lundstroms” safety matches” so the match tip is clearly understood from the previous mention phrase Lundstrom Safety matches”

In the last sentence in which the writer used the term Ellipsis is in the following sentence that is **“Other American innovations include an anti-afterglow solution to prevent the match from smoldering after it has been blown out; and the waterproof match, which lights after eight hours in water”** (Jakeman, and McDowell 2006).In this sentence the word solution that is mentioned after the phrase “the waterproof match” which can be understood from the first clause

Above the great description it is clear that the writer of this reading passage has used Ellipsis frequently in the reading passage. The researcher mentioned all those ellipses that are used in the paragraph with the motive of assisting the readers. so the frequency of the ellipsis that the writer has been used in the paragraph is about (09). And these ellipses play an integral function to omit the word that is understood from the context. The researcher analyzed that this cohesive device makes the sentence more concise without losing its original meaning. The main function of ellipsis is to concise the text.

#### 4.1.2 Substitution

The following paragraph gave the information about substitution that the writer used in the reading passage. Substitution is the most integral term of cohesive devices that the writers used in their discourse. It refers to one phrase to replace another word, phrase, or entire clause that has already been mentioned or can be inferred from the context. This helps to avoid repetition and creates cohesion in the text.

The last grammatical cohesive device that is mentioned by Haliday and Hasan 1976 model is Substitution. The writer has frequently used this technique in this analyzed reading passage. The writer has discussed all the 3 types of Substitution that is presented by Haliday and Hasan ( Nominal, substitution, Verbal substitution, and the last one is Clausal substitution) The researcher has analyzed all of them in explicit way The first substitution is **"Fire"** it substituted for " Flame" or "Fire Lighting" Similarly **"It"** substitutes for background information to avoid repetition. " This Process" substitutes for the method of rotating a wooden drill briskly. Likewise **"It"** substitutes for "Phosphorus" and **"The element"** also substitutes for **"Phosphorus"**. The word **"Them"** substitutes for **"Congreves"**, Similarly the word **"His Product"** substitutes for matches, **"Which"** substitutes for the previously mentioned **"substance"** In the next sentence again the word **"Which"** substitutes for **"Potassium chlorate"**. And **"which"** in the next line substitutes for the previously described **"Waterproof match"**. And **"Today book matches"** substitutes for **"book matches"**

So, all of the above substitutions mentioned in the paragraph are analyzed by the researcher in great detail. The researcher has also analyzed the frequency of all the three types of Substitution that is (13) in numbers and their function is very integral in the passage that is avoiding repetition in the context by replacing pronouns or by using shorter phrases

#### 4.2 Data 2 from the 2nd book and his 1st Reading Passage

The researcher has analyzed all the cohesive devices that are used in this passage, and to analyze its frequency the researcher found the following results.

**Table 12:** *Usage of Cohesive devices:*

Types of reference	Total number of references	Usage

Grammatical cohesive devices	N	%
Reference	62	20.9%
Conjunctions	48	16.2%
Ellipsis	03	1%
Substitution	13	4.3%

The findings revealed that they contained all the cohesions in the 1st passage of the 2nd book. From the result it is clear that the “reference” is very dominant cohesion in the IELTS reading exercise. In order to find out the detail on what is most constantly used cohesion in the reading passage, the researcher presents the table with individually cohesive devices with their types to find out the clear detail. The researcher analyzed then individually

The result of references is presented in table 13

**Table 13:** *Usage of References*

Total Cohesive devices used	Reference use
296	N
	62

Above the table shows there are a total (296) cohesive devices that are used in the IELTS reading exercises. The N symbol shows that there are (62) references used in reading exercises. So, the most frequently used cohesive devices are References that show that they have a very integral function to unify the text. The researcher also identifies the most continuously used reference type in the reading exercise by drawing a table.

**Table 14:** *Usage of personal references*

Total	Personal references	Number of devices used
	It	06



08	Their	02
----	-------	----

The above table shows that almost 5% personal reference is used in the reading passage in which article “**It**” is the most steadily cohesion used in the passage. It proved that “it” is widely used cohesion by the writer in the IELTS paper to avoid repetition.

**Table 15:** *Usage of demonstrative references*

Total	Demonstrative reference	Number of devices used
54	This	6
	The	38
	That	06
	There	04

The above explains that the most widely used cohesive device that is used by the author is the article “The to specify the noun.

**Table 16:** *Usage of comparative references*

Total	Comparative reference	Number of devices used
—	—	—

From the results of the above all tables it is clear that the author of the IELTS book has widely used the demonstrative reference cohesive because the result shows that the total number of the demonstrative references are more than the other types of the references. The total number of the demonstrative references is **54**. Personal has just **08** total numbers and he did not use comparative reference in his book. The researcher identifies that most of the demonstrative device that the writer has used in his book is the determinant **The**. It remains prominent throughout while analyzing the reading exercise.

**Table 17:** *Usage of Conjunction*

Total	Types	Total number of Conjunctions
48	Additive	25
	Adversative	06
	Causal	09
	Temporal	08

The result of the conjunction is explicitly clear that the frequency of conjunctions is lower than references. The total number of conjunctions is (48). If the researcher compared it with the references they are totally (62) in numbers and they are just (48).

The table below reveals the number of all Additive cohesive devices used by the author of the book.

**Table 18:** *Usage of Additive cohesive devices*

Total	Additive devices	Total number of additive conjunctions
25	And	14
	Second	01
	Further	01
	That	06
	First	01
	Both	02

The result shows that the writer has used (25) additive conjunctions in the passage. The highest conjunction is “And”. But all the conjunctions have their own role in the text. The main function of “And” is used to connect two different words, sentences or phrases. It is also used to add one element to another. The main function of “Or” is used to introduce alternative

choices. The writer has also used examples in this passage to make his statement more authentic in front of the readers. The writer has used “Then” when he has talked about time, then is also used to explain further information. Similarly, he used “such as” to give examples of something that he described but the writer has used it just 1 time in the passage. The writer has also used adversative conjunctions. The researcher has revealed them through a table.

**Table 19:** Usage of Adversative *conjunction*

Total	Adversative conjunction	Total number of adversative conjunctions
06	But	03
	However	02
	Rather	01

The above of the table indicates that the author did not use widely adversative conjunctions. He just used 3 adversative cohesive devices in the form of "But and 1 in the form of “However”. The researcher has explored their functions as well. The writer has used “But” when he presents the second clause or sentence as a contrast to the previous one. The function of “However” is used to contrast between two clauses, however is used in positive sentences. The above of the table indicates that the author did not use widely adversative conjunctions. He just used 4 adversative cohesive devices in the form of "But" and 1 in the form of “However”, and "Rather". The researcher found out that the use of adversative cohesive devices is lower than the use of additive conjunctions the writer has also Causal conjunction. The researcher has explored all of them. The researcher has explored in the form of table

**Table 20:** Usage of Causal *conjunctions*

Total	Causal conjunction	Total number of causal conjunctions
08	So	02
	As	06

The total number of causal conjunctions is 08. The following result shows that the writer did not use them frequently. But they play a lot of function in the text. The author has used “so” to join two independent clauses in which the second one is connected with the 1st on similarly,

some function has already been described above. The writer has also used temporal clauses as well. The researcher explained it through the table.

The total number of causal conjunctions is 08. The following result shows that the writer did not use them frequently. But they play a lot of function in the text. That has already been described above. The writer has also used temporal clauses as well. The researcher explained it through the table.

**Table 21:** *Usage of temporal conjunction*

Total	Temporal conjunction	Total number of temporal conjunctions
01	When	01

The writer has used just 01 temporal functions in this passage in the form of “when, and while” The function of “when” is to emphasize two actions happening at the same time. So, from the result of the above tables the researcher concludes that the highest type of conjunction is an “Additive”. And the lowest is temporal but they perform individual roles in the text.

#### 4.2.1 Ellipsis

Now the following paragraph reveals the uses of Ellipsis in the book. Basically, ellipsis can be defined as the omission of words or phrases that can be easily gone through from the eye of the reader but this omission does not affect the meaning of the sentence. The author has used many ellipsis techniques in his book to make it easy for the reader. The researcher highlighted all the ellipses that were used in sentences.

The Ellipsis that is used in the sentence **“The river builds them up, the sea wears them down; their outlines are always changing. The changes in China's Pearl River delta, however, are more dramatic than these natural fluctuations”** in this sentence the word outline is not repeated after their. In the next sentence “

**“The new island of Chek Lap Kok, the site of Hong Kong's new airport, is 83% complete. The giant dumper trucks rumbling across it will have finished their job by the middle of this year and the airport itself will be built at a similarly breakneck pace”** in this sentence the giant dumper trucks rumbling across it will have finished their job by the middle of this year" does not repeat the subject "The giant dumper trucks" in the following clause.

**As Chek Lap Kok rises, however, another new Asian island is sinking back into the sea. This is a 520-hectare island built in Osaka Bay, Japan, that serves as the platform**

**for the new Kansai airport.”** In this given sentence "another new Asian island" does not repeat the information already provided about the island.

Above paragraph explains very clearly the use of Ellipsis in the reading passage.

#### 4.2.2 Substitution

The following paragraph the researcher gave the information about substitution that the writer used in the reading passage “substitution” is the most integral term of cohesive devices that the writers used in their discourse. It refers to the use of one word or phrase to replace another word, phrase, or entire clause that has already been mentioned or can be inferred from the context. This helps to avoid repetition and creates cohesion in the text.

The first is in **land** for **Island**, Similarly **“It”** substitutes **“building”**, and the word **“It”** substitutes for **settlement**, Likewise the word **“Site”** for **location**, **“Them”** substitutes with **“different parts”**, **“that”** substitutes with **“Fate”**. The word **“It”** substitutes with **“soft mud layer”**, **“It”** also substitutes with **“heap of boulders”** **“it”** also substitutes with **sand**. The word **“It”** also substitutes **“The method of using sand and rock to strengthen the seabed”** The word **“There”** substitutes with **“land”**, And **them** substitutes with **“different parts”**. **”There”** substitutes with **“land”** . So, the total number of substitutes is 14,

#### 4.3 Data 3 from the book 3 and its 1st passage

The researcher has explored all the types of cohesion of the third book and its 1<sup>st</sup> passage.

**Table 23:** *Usage of cohesion*

Types of reference	Total number of references	Usage
Grammatical cohesive devices	N	%
Reference	132	30.4%
Conjunctions	63	14.5%
Ellipsis	09	2.7%
Substitution	18	4.1%

The result disclosed that the 1<sup>st</sup> passage of the third book contains much cohesion. The result revealed that this passage includes all the cohesions, the researcher presents the table with individually cohesive devices with their types to find out the clear detail.

**Table 24:** *Usage of Cohesion*

Total Cohesive devices used	Reference use
433	N
	132

Above the table shows there are a total 433 cohesive devices that are used in the IELTS reading exercises. The N symbol shows that there are 132 references used in reading exercises. So, the most frequently used cohesive devices are References that show that they have a very integral function to unify the text. The researcher also identifies the most frequently used reference type in the reading exercise by draw a table

**Table 25:** *Usage of Personal references*

Total	Personal references	Number of devices used
19	It	15
	Their	04

The above table shows that almost 15% personal reference is used in the reading passage in which article. **It** is the dominant term in personal references. The writer has used “It” to avoid repetition in the text

**Table 26:** *Usage of Demonstrative reference*

Total	Demonstrative reference	Number of devices Used
112	This	05
	The	92
	That	13

	These	01
	Those	01

The above table revealed “The” demonstrative reference is most repeatedly used cohesion than any others.

**Table 27:** *Usage of Comparative Reference*

Total	Comparative reference	Number of devices used
01	Another	01

The writer has used just 1 comparative reference in the form of “another”. The main function of this form is to talk about other things. The writer has used it as follows

“Another weapon was the ‘arrow as a flying sabre”

From the results of the above tables it is clear that the author of the IELTS book has widely used the demonstrative reference cohesive because the result shows that the total number of the demonstrative references are more than the other types of the references. The total number of the demonstrative references is 112. Personal has just 19 total numbers and comparative has just 1 number. The researcher identifies that most of the demonstrative device that the writer has used in his book is the determinant “The”. It remains prominent throughout while analyzing the reading exercise.

**Table 28:** *Usage of conjunction*

Total	Types	Total number of Conjunctions
<b>63</b>	Additive	39
	Adversative	10
	Casual	08
	Temporal	06

The result of the conjunction is explicitly clear that the number of conjunctions with cohesive devices are lower than references. The total number of conjunctions is just 63. If the researcher compared it with the references they are totally 132 in numbers and they are just 63.

The table below reveals the number of all additive cohesive devices used by the author of the book.

**Table 29: Usage of Additive conjunctions**

Total	Additive devices	Total number of Additive conjunctions
39	And	18
	Then	01
	Either	02
	That	13
	Or	05

The result shows that the highest frequency of additive conjunction is “And”. the various devices that or, then, either are not frequently used cohesive devices. The result shows that the writer has used (39) additive conjunctions in the passage. The highest conjunction is “And”. But all the conjunctions have their own role in the text. The main function of “And” is used to connect two different words, sentences or phrases. It is also used to add one element to another. The main function of “Or” is used to introduce alternative choices. The writer has also used examples in this passage to make his statement more authentic in front of the readers. The writer has used “Then” when he has talked about time, then is also used to explain further information. Similarly, he used “such as” to give examples of something that he described but the writer has used it just 1 time in the passage. The writer has also used adversative conjunctions. The researcher has revealed them through a table.

**Table 30: Usage of adversative conjunctions**

Total	Adversative	Total number of Adversative
	But	03
	Rather	02



10	However	02
	Though	02
	Nevertheless	01

The above of the table indicates that the author did not use widely adversative devices. He just used 10 adversative cohesive devices in the form of But, Rather, However, Though and “Nevertheless”. All of them have a lot of functions in the text. For example writer wrote in this paragraph:

“Not only did it solve a problem that had intrigued man for ages, but more importantly it literally opened the door to exploration”.

In the following line the writer has used “but” as a clause to attach to sentences in one. Similarly, the writer has used “rather” to give more detail in the text. For example he used rather in the text.

“By the thirteenth century, powder propelled fire arrows have become rather common”

This sentence rather gave clear information about propelled fire arrows. “However” and “Nevertheless” used to contrast between two clauses. However, it is also used to give the result or conclusion. For example the author used “nevertheless” to contrast the two negative clauses.

“In rocket technology, often with devastating results in the forum of war nevertheless the modern day space space programs....”

The writer has used “however” in the text to give the conclusion. The sentence is given below.

“However, the results were sometimes not that impressive as the behavior of the rocket...”.

The writer has also used “though” three times in this passage because he has used it to make it clear to the audience that the information is already given in the previous clause. For example he wrote

“It is strange that the rocket was generally ignored by writers of fiction to transport their heroes though it had been commonly used in fireworks,”

The writer has also used Causal conjunction that is in below the table

**Table 31: Usage of causal conjunction**

Total	Causal	Total number of causal
08	So	01
	As	05
	Since	02

The frequency of causal cohesive devices is 8 that is lower than above two tables. The writer has also used temporal words that are explained in the below. They have various functions. “So” used to join two independent clauses. The writer wrote in the below:

“However, it wasn't until the discovery..... and so, represents one of the great milestones ....”

The writer has used “as” to compare the two clauses or words. . The writer writes in the below lines.

“Another weapon was the arrow as a flying sabre.”

The writer has also used “since’ in just 1 time in the passage but it has many functions. “since” is used to depict an action that started in the past and is still continuous. The below lines make it clear how the writer has used “since” in this passage, “Since then, there have been huge developments in rocket technology”

**Table 32: Usage of temporal conjunctions**

Total	Temporal	Total number of Temporal
06	During	01
	Before	02
	Until	03

The following table shows that the writer has used just 06 temporal conjunctions in this passage in the form of during, before, until. The writer did not use them frequently but they all have a lot of functions. For example, “during” is used to describe the thing that is already happening. The writer has used it just 1 time in the passage.

"During two world wars, were required before....”

The writer has also used “before” in this paragraph to describe the 2nd event after the independent clause.

“There is evidence that the reaction principle was applied practically well before the rocket was invented.”

The writer has also used “until” two times in this reading passage. The main function of this conjunction is to describe the clause as the event that stopped the first event. The writer used the following sentence.

“They remained a relatively minor artifact of civilization until the twentieth century.”

So, from the above of all types of conjunction it seems that the frequency of temporal devices is as low as above of the tables.

#### 4.3.1 Ellipsis

Now the following paragraph reveals the uses of Ellipsis in the book. Basically, ellipsis can be defined as the omission of words or phrases that can be easily gone through from the eye of the reader. But this omission does not affect the meaning of the sentence. The author has used many ellipsis techniques in his book to make it easy for the readers. The researcher highlighted all the ellipsis that used in sentences

**“Not only did it solve a problem that had intrigued man for ages, but, more importantly, it literally opened the door to exploration of the universe.”**

The first sentence in which the writer used ellipsis is “The repeated subject ” after “but” is omitted for conciseness, assuming the reader understands it refers to the reaction principle.

In the second sentence **“Despite the fact that rockets had been used sporadically for several hundred years, they remained a relatively minor artifact of civilization until the twentieth century.”**(“The phrase omits repeating the full clause “the fact that rockets had been used sporadically for several hundred years” by using “they” instead, which refers back to rockets.)

In the 3rd sentence **“The reaction to the ejection of these small particles” in the second half, assuming the reader understands that the rocket’s movement is caused by the ejection.** The omission is used in this way (“The reaction to the ejection of these small particles causes the rocket to move forwards”).

The next line is **“By the thirteenth century, powder-propelled fire arrows had become rather common. The Chinese relied on this type of technological development to produce incendiary projectiles of many sorts, explosive grenades and possibly cannons to repel their enemies.** (“Omit the repeated subject “The Chinese” before “to produce,” assuming the reader understands who is producing these weapons)“The Chinese relied on this type of

technological development to produce incendiary projectiles of many sorts, explosive grenades and possibly cannons to repel their enemies.

The next line is **“A small iron weight was attached to the 1.5m bamboo shaft, just below the feathers, to increase the arrow’s stability by moving the center of gravity to a position below the rocket”** Omit apply by “by moving the center of gravity” a second time after “to a position below”.

The next line is **“It was not until the eighteenth century that Europe became seriously interested in the possibilities of using the rocket itself as a weapon of war and not just to propel other weapons.”** Omit repetition of the phrase “the possibilities of using the rocket itself as a weapon” after “and not just to,” making it concise.

The next sentence is **“Prior to this, rockets were used only in pyrotechnic displays. The incentive for the more aggressive use of rockets came not from within the European continent but from far-away India, whose leaders had built up a corps of rocketeers and used rockets successfully against the British in the late eighteenth century.”** It Omits repeating the subject “the incentive” after “for the more aggressive use of rockets,” assuming it is understood.

The next line is **“Since then, there have been huge developments in rocket technology, often with devastating results in the forum of war.”**(Omits repetition of “huge developments in rocket technology” after “often with,” assuming the context of war results is clear.)

The next sentence is **“Nevertheless, the modern day space programs owe their success to the humble beginnings of those in previous centuries who developed the foundations of the reaction principle.”** Omits repeating “the humble beginnings of those in previous centuries” after “success to,” as it’s understood from context.

Above description explains very clearly the use of Ellipsis in the reading passage. In the 1st column the researcher gives the total number of ellipses that the writer used in the sentences of the paragraph. In the next column the researcher mentioned all those sentences in which the writer used the technique of Ellipsis and in the last column the researcher mentioned all those ellipses that are used in the paragraph with the motive of assisting the readers.

### 4.3.2 Substitution

The following table gave the information about substitution that the writer used in the reading passage Substitution is the most integral term of cohesive devices that the writers used in their discourse. It refers to the use of one word or phrase to replace another word, phrase, or entire clause that has already been mentioned or can be inferred from the context. This helps to avoid repetition and creates cohesion in the text.

The first sentence in which the writer used substitution is following

“The reaction principle” simplifies and specifies the previously mentioned “mechanism”.

“it” is used as a pronoun to avoid repeating “the discovery of the reaction principle.”

“Enabled space travel”, The phrase substitutes “a more literal expression for the concept of space travel, emphasizing the breakthrough’s impact”.

"this" for "the intellectual breakthrough"

"they" for "rockets", similarly the pronoun “it” replaces “the rocket” to maintain cohesion and avoid redundancy.

“These small particles” substitutes for “the detailed description to simplify the explanation”.

“it” replaces “the pigeon of Archytas” after the first mention to maintain brevity.

"this" for "black powder", "they" for "the Chinese", “it” is used as a pronoun for “a rocket” in the analogy provided.

“These small particles” substitutes for “the detailed description to simplify the explanation”.

"The pigeon of Archytas" for "it"

“this” is used as a demonstrative pronoun to refer back to “black powder.”

"they" for "the Chinese"

arrows" for “‘basket of fire’ or, as directly translated from Chinese, the ‘arrows like flying leopards’”

this” replaces the longer phrase to simplify the sentence.

the British” is used to refer back to the specific individual or the group he represents.

“it” replaces “the British rocket” to streamline the sentence. “they” is used to avoid repeating “the Americans.” "it" for "the British rocket" "then" for "since the mid-nineteenth century" "those" for "the individuals in previous centuries who developed the foundations of the reaction principle"

## 5. DISCUSSION AND CONCLUSION

The findings of Nunan (1993), revealed that every sentence of the text must be created well structured in written or orally as well. Cohesive devices deal with lexical cohesion and

grammatical cohesion; they are interlinked with each other. The writer further explored that the main object of cohesive devices is to develop clarification in syntax and provide clear information without any obscurity. So, the first and foremost function of cohesion is to sustain coherence in a sentence. The next research is conducted from Vietnamese EFL students who used cohesive devices in paragraph writing. The findings of this research explored that the percentage of grammatical cohesive devices are higher than the lexical cohesion. In which the references and conjunctions are frequently used than substitution and ellipsis (Tuan, Trang, and Nhu, 2023). So, according to the research results that have been explained, the use of grammatical cohesion (such as reference, ellipsis, substitution, and conjunction) and lexical cohesion (such as reiteration and collocation) play an important function to create a structured and meaningful text. Therefore, the authors of these IELTS books used all the cohesive items in frequent ways, especially references and collocations. in the IELTS reading passages.

Cohesive devices, known as binding or linking the words or phrases in different parts of a text together, But with proper meaning. They help to guide the reader to read the text with proper meaning and understand what the writer wants to convey. There are some factors that make the text more meaningful. Some of those factors are described in terms of cohesive devices. The researcher examined the use of cohesive devices in IELTS reading exercises that focuses on frequency, functions and identifying all the cohesive devices that are used in IELTS reading passages. The findings of the study explored the importance of the cohesive items in the reading passages. This study also revealed how all the types of cohesive items play their function to unify a text and give it a proper meaning. And the researcher has also identified the function of all these cohesive devices. The main function of cohesive devices is to connect the ideas between two sentences, phrases or clauses. The researcher observed that all the cohesive devices play an individual functions into the whole text, for example the function of “References” is to avoid repetition in the text. The result of the study proved that the writer has used a lot of reference cohesive devices used in whole reading passages which the researcher has found. The total frequency of references is 1,007. Similarly, the writer explored that conjunctions have their own functions: they help to connect two different ideas in the sentence, clauses, or phrases. The writer has used almost all types of conjunctions in the passages like adverbial, additive, causal, and temporal to link the ideas.

Ellipsis also played an integral role in the text. The writer has used it to omit the sentences. In this way the writer can easily concise the text but gives the meaning without any hindrances, but the number of Ellipsis is lower than all. The frequency of Ellipsis is just 52 which is just 1.4%. Which means that the writer of all these passages which the researcher has analyzed did not concise the text in a vivid way. The next cohesive device that the writer has used is “Substitution”. The writer has used this cohesive device almost 2.8%. It helps to avoid repetition in the text, for example “It” refers back to “Fire” in which the writer used substitution to avoid repetition of the same word.

Similarly, the 2nd type that the writer has used “Lexical cohesive devices:” plays a very vital role in the text. The writer has used repetition, synonyms, Antonyms, Hyponyms, Meronymy and Collocation to make his work unique than others and the reader does not get bored while

reading the text. So, all the mentioned cohesive devices perform their own role in the text. This research is limited to identified cohesive devices that is presented by Haliday and Hasan (1976) model of Cohesive devices. The researcher has taken English reading exercises for the IELTS Examination. The researcher has just taken the first 10 first books from the 18 books and analyzed the very 1st reading passages of the books.

This study delved into the functions of cohesive devices in Ielts reading passages. This study encourages all second language learners and teachers to concentrate on the use of cohesive devices in reading passages because all these items play an integral role to build a structured and meaningful sentence. A student must have to know what the cohesive devices are and what is the actual function of these items. The study gives knowledge to the students on how they can recognize cohesive devices in the reading passages. This research widens the way of reading comprehension. This research also accommodates the teachers to instruct cohesive devices to their students. These recommendations tried to improve learning, and teaching assessment evaluations related to cohesive devices in IELTS reading exercises.

## REFERENCES







TRANSLINGUA, 2025

Vol 1, Issue 1, Pages 70-78



### Critical Discourse Analysis of Pakistani War Songs during the 1965 Pak-India War

Author/s: Sadia Shabbir

Affiliation: Arid Agricultural University, Rawalpindi, Email: [samanbutt929@gmail.com](mailto:samanbutt929@gmail.com)

**ABSTRACT:** War poetry played a significant role during the Pak-India war of 1965. There was a national agenda behind this discourse. War poems paid tribute to the soldiers on the surface level but satisfied the national interests on hidden grounds. Pakistani singers like Noor Jahan and Mehdi Hassan stimulated this national agenda. This study foregrounds representation of ideology in the war songs of 1965. The study concludes that the adverse reactions of war are not highlighted but glorified. War does not bring prosperity, but massive destruction, chaos and mental disturbance. Ideology behind creating such a discourse is highlighted. These songs present message of unity, patriotism, loyalty and strength among the nation.

Keywords: Critical Discourse Analysis, War Songs, 1965 Pak-India War,

To cite: Shabbir, S. (2025). Critical Discourse Analysis of Pakistani War Songs during the 1965 Pak-India War, *TRANSLINGUA*, 1(1), pages 70-78.

## 1. INTRODUCTION

Among various internal and external conflicts of which Pakistan has been subjected, the conflict against India stands out the most. One of the consequences of this perennial conflict is the 1965 war during which Pakistan thwarted a major Indian transgression against its borders. Often viewed and revered as a triumphant event in Pakistan, the war is still remembered with the purpose of rejuvenating the patriotic spirit of the Pakistani population. During the war, a legendary singer Noor Jahan sang songs to lift the spirits of the fighting soldiers and to keep the country united in troubling times. Those songs encouraged the soldiers and motivated them to sacrifice their lives for the defense of their country. Spirit of martyrdom motivated them not to surrender against enemies. This article provides a critical discourse analysis of Noor Jahan's war songs during the war of 1965. This article highlights the use of lexical items to boost the spirit of patriotism. CDA aims ultimately to make a change of "the existing social reality in which discourse is related in particular ways to other social elements such as power relations, ideologies, economic and political strategies and policies" (Fairclough, 2014). The aim of this article is to analyze the use of language in the war songs, its role in energizing the patriotic spirit and to highlight the ideology behind it.

Critical Discourse Analysis permits the researcher to dig out masked ideologies through close scrutiny. Research will observe lexical items like power, dominance and ideology; play an important role in CDA. It unfolds the hidden reality of Pakistani war songs during 1965 war. The main objective of the research is to analyze the ideology presented in Pakistani war songs through language. This study focuses on the ideology presented in Pakistani war songs during 1965. It foregrounds the hidden agenda behind the war songs. CDA works on the Pakistani war songs and scrutinizes the facts. The study is limited to the Noor Jahan's and Mehdi Hasan's war songs during the war of 1965, other patriotic Pakistani songs are excluded.

## 2. LITERATURE REVIEW

This section focuses on the relevant literature and relates it to the present study. CDA is a tool for analyzing society through language. In an article, 'metaphor analysis in songs through CDA' by Florencia Figini, explores the importance of songs. The song is a powerful way to convey ideas and thoughts. Emotions, feelings and desires are expressed through the use of lexical items (Figini, 2015). Discourse is always created by the powerful. Unpleasant events are always created in a heroic manner. CDA brings out real picture and tries to shatter the clouds of deception. It brings out facts under close scrutiny. CDA tries to eliminate biasness and raises voice in favor of the weaker. According to van Dijk (2009), access to or control over public discourse and communication constitutes a resource that is symbolically important because somebody may decide not only on the content but also on the structures of text and talk in a given context. "Controlling context involves control over one or more of these categories, e.g. determining the definition of the communicative situation, deciding on time and place of the communicative event, or on which participants may or must be present, and in which roles, or what knowledge or opinions they should (not) have, and which social actions may or must be accomplished by discourse" (van Dijk 2009: 356).

Poets, authors and journalists make use of lexical words to glorify the war. “The Ideology of First World War Poetry Criticism”, by James Campbell, in his article equates the term “war” with the term “combat”. Soldiers of the two opposite sides were not familiar with each other but ready to kill during First World War. People with firsthand experience of war shared their experiences in a heroic manner. They presented horrible images of war in their writings. But they also presented man as a warrior, not as a soldier. Use of lexical items, form and structure of the writings are helpful in understanding, what the writer wants to say, reflects one’s ideology. War is not a heroic, causes bloodshed, hatred and brutality. Actually, journalists, politicians and writers support their national interests through language. Language is a medium which is used by them to achieve their targets (Naim,2011).

Dawn news article, ‘The 1965 war and Pakistani Urdu literature’ connotes the role of Pakistani poets, writers, journalists and intellectuals during the war of 1965. This war lifted the spirit of patriotism among people and they emerged as a strong nation. Urdu Literature played an important role to unite the people at that time. Literature with the ideology of unity, faith and patriotism was produced. It did not highlight the adverse reactions of war. Media supported literature in strengthening the ideology behind discourse. Radio Pakistan had a major role to gather the people of Pakistan. In 1965, radio was only a medium which propagates the government agenda. Highly emotional songs were composed with profound meanings (Parekh, 2015).

Pakistani soldiers were encouraged with the highly emotional songs and India was presented as a worst enemy. They were not presented as a human being. Negative attributes were associated with them. They were presented an object with soul. Language was used as a medium to supporting the false conceptions. Patriotic songs united the people but also created the spirit of hatred and revenge for Indians. Literature was designed to gain the national interests. Vocalists were used as a tool to accomplishing the goal of the political leaders (Parekh, 2015).

The Nation-article- by R Umaira Ahmed, ‘Noor Jahan voice of 1965 war’ foregrounds the role of Malaka-e-Taranum , Noor Jahan’s patriotic songs during the war of 1965. Her voice injected a new spirit in Pak army. Her songs *aae putar hatan dy nai wikdy and aae watan ky sajelay jawano* were the most mesmerizing songs. These songs memorize the sacrifices of those who gave their lives for Pakistan. Her Punjabi and Urdu patriotic songs for the soldiers, gave them a new soul (Ahmed, 2017).

The poets, lyricists and singers played a vital role in awakening the masses and uniting them against enemies. Mehdi Hassan and Noor Jahan were the renowned figures who sang everlasting patriotic songs and the lyricists spent nights in writing the patriotic songs. The war was not only fought at the borders but writers, poets, singers, educationalists, journalists and others had also their contribution. Their songs enhanced the spirit of enthusiasm, unity and patriotism among people. Those songs also motivated the young ones to participate, sacrificed their lives for the country and conveyed a message of loyalty (Ahmed, 2017). “1965 war to remain incomplete without Noor Jahan”, another article foregrounds the contribution of Noor Jehan that she was fighting the battle through media. She was singing for the soldiers who were defending the country’s borders. Her songs sent the waves of patriotism and nationalism

and every Pakistani were brimmed with the spirit of martyrdom. Field Marshal General Ayub Khan had acknowledged that *“half of the credit of 1965 victory goes to Noor Jahan (NA, 2016)”*

Faizal Risdianto in his article “ Discourse Analysis of a song Lyric entitled We will not go down states the three different aspects of discourse analysis of Michael Heart’s song We will not go down. According to Faizal this song is the personal outcome of the writer and the singer. There is cohesion and coherence in the lyrics as song contains personal pronouns that also highlight the personal experience of a writer and singer. The song “We will not go down” is made for the Palestinian people to uplift their strength. Song casts a spell of strength over the Palestinian that they will not surrender and will fight until they win the battle.

Discourse has a political agenda behind so the song “ We will not go down” can be compared with Noor Jahan’s war song “ Ay putar ghata da nai vikday” . Noor Jahan has sung this song in the war of 1965 to uplift the strength of soldiers who were participated in the war. Like Michael song “We will not go down”, Noor Jahan also gives the message to the Pakistani soldiers who were participated in the battle that the nation will not go down and surrendered until the win the battle. (Risdianto)

In “1965 War, songs that stimulated verve and pep among soldiers and nations” Junaid Malik highlights 1965 event that united Pakistani nation. When Indian army crossed the border and entered in Lahore, Ayub Khan declared war. His speech made every Pakistani blood boil. Artists like Noor Jahan and many others stood with Pakistani arm force. Discourse can have a purpose; sometimes words spoken and written have power to control the minds. Neither Songs of Noor Jahan like aay putar hattan da nai nvikda boosted the spirits of soldiers.

Discourse has a political agenda especially when one talks about war songs. There is always a cold war among India and Pakistan because India thinks that Pakistan has no existence as India is like a mother and Pakistan is the separated part of its motherland so it has to bring back. In 1965 India entered in Pakistan with the purpose to rejoin Pakistan with India, thus at that time queen of melody Noor Jahan played a significant role to boost up the spirits of soldiers. (Malik)

Pakistan Singers like Mehendi Hassan and Noor Jahan played a significant in 1965 war. In “ 1965 War history to remain incomplete without Noor Jahan” Murtaza Abbasi , a former soldier states that Noor Jahan being a female singer provided a new spirit to the soldiers and her patriotic songs arose nationalism and patriotism and every Pakistani was ready to sacrifice his life and to fight side by side with Pakistani army. According to Abbasi , she is not a singer but a brave daughter of the nation , who unity all the nation at the time of war and encouraged the soldiers to fight against the enemies.

Naeem Tahir in Daily Times “ Noor Jahan- the unforgettable” states that Noor Jahan is our Malaika. She is not Malaika by name but she is our real Malaika . She is real Malaika of the creative world as she played a role of a brave daughter of the nation in 1965. She was strong and determined woman who took part in the war by her voice. Her songs casts a magic on the entire nation and because of her songs soldiers stood up and fought against the enemy and they

win the war. Thus it can be said that discourse has a strong role to manipulate power and it can be used as a tool in the form of songs for political agenda. (tahir)

### 3. RESEARCH METHODOLOGY

Past researchers reveal that the language is used for power. CDA is used to unveil the reality, how language as a source is used. Singers intensify the power of language with their melodious voice. Hidden ideology works behind discourse. This research highlights the gap that Pakistani war songs during 1965 glorified the war and Pakistani soldiers but actually, other ideology worked behind it. The present study aims to fill this gap by providing CDA of the selected war poems by Noor Jahan and Mehdi Hassan.

### 4. RESULTS AND FINDINGS

Fairclough's model is used to analyse Noor Jahan and Mehdi Hassan's patriotic songs like *aae putar hatan dy nai vikdy*, *aae watan ky sajely jawano*, *aae rahay haq ky shaheedo*, *mere dhool sipahiya* and *apni jaan nazar karoun* during the war of 1965. This section highlights different linguistic choices by the writers in the selected sample. Each discourse has an agenda; the innocents are misused by the authorities. Fairclough states that the nature of the power relations enacted in mass media discourse is not clear and there are reasons for seeing it as involving *hidden* relations of power (Fairclough, p.49). Language is a medium, having political notions behind.

Noor Jahan's patriotic songs during the war of 1965 motivated innocent soldiers to sacrifice their lives for the defense of the country. Literature formulated was not pure and innocent, something planned masterly. Discourse was not only created in favour of the soldiers but different thought worked behind it. This article analyzes Noor Jahan and Mehdi Hassan's popular patriotic songs during 1965: *aae putar hatan dy nai vikdy*, *aae watan ky sajeelay jawano*, *aae rahay haq ky shaheedo*, *mera dhool sipahia* and *apni jaan nazar karoun*.

#### **Aae putar hatan dy nae vikdy**

'Putar' is a Punjabi word which means 'son'. Soldier is considered as a son of the country, reflects possession like a mother. Pakistan is presented as a mother. Use of lexical items like son (putar), worthy (nai vik dy) makes the soldier valuable. These words touch the hearts of the people.

Use of the phrase, land of the saint (*darti hai meray data di*) is used to lift up their spirits. Pakistani people have great emotional affiliation with sufi saints. Pakistan is taken as the land of the saints in this poem. Purpose to create such type of discourse is to *control* people through language (p.45). They are demanded to sacrifice their lives. Their minds are whitewashed through language.

Language is used artistically to grab the emotions of the Pakistani soldiers. Persuasive words like precious (unmoal) and worthy (wikao cheese nai and susta maal nai) are used for their praise. An analysis of such words reveals them as agents of a *hidden discourse* which demands the soldiers to give their lives for the sake of national interests (p.55). They are convinced to give their lives for the defense of the country. Their mother land demands for their blood.

Fairclough advocates, “producers exercise power over consumers in that they have solely produce the rights and can therefore determine what is included and excluded, how events are represented and even the subject position of their audiences. But who precisely are these producers? (p.50)” Pakistani soldiers are compared with lions (shair). This comparison depicts bravery and courage. ‘Are not scared of death and enemy’ (moth aur dushman tou wi dardy nai). Such discourse is designed to suit the agencies. Phrases like *apri jan wi war* (give life to country), *tun pagh ny una mawan dy* (lucky mothers, who gave birth to such brave boys) are used in the poem. These depict intense appreciation by the poet. Similarly, the phrases such as *au pairh parawa* (brothers and sisters), *jina godian khadawa ny* (who lull them). Such words invoke the love of the family and the reminders of the strength of familial bonds. This tends to be an attempt to unite the soldiers with the citizens of Pakistan and hence reinforce their loyalty to the land. Appealing discourse is designed to fulfill the *hidden agenda* (p.55) i.e. patriotism and unity.

### **Aae rahay haq ky shaheedo**

CDA presents how the ideology works behind creating discourse such as war poetry. This song depicts that *power at all these levels won, exercised ...* (p.68). Words like martyr (*shaheed*), bravery (*shujahat*), matchless (*unmoal*), faith (*wafa*) and greetings (*salam*) are associated with the soldiers. Political agenda exercises behind creating discourse. The political authorities get advantage while launching such a discourse. Naturalization is an important concept refers to repetition of ideologies to a certain degree that it seems truth. Such ideas are created and internalized by the dominate class. CDA challenges such ideas what have never been questioned before (p.33).

Use of decorative words and phrases like fire cools down with their blood (*wo sholay apny lahu sy bujha diye tum ny*), the defenders of Pakistan (*bacha liye hai kitnu ko yatemi se*), savior (*suhag kitnu ky bacha liye tum ny*), brims with the spirit of martyrdom (*chaly ho gy shahadut ka jazba ly kar*) satisfies the idea of hidden agenda behind discourse.

References of noble personalities like *Holy Prophet (SAW)*, *Hazrat Ali (RA)* and *Hazrat Imam Hussain(RA)* depicts *power in discourse* (p.70). The names of these noble personalities are used to intensify the emotions of the people. There is *power behind discourse* which achieves its goal through the manipulation of language. Use of lexical items depicts that the soldiers are fighting for the noble cause. Giving life for the country is described as a holy duty but *hidden agenda* is quite different (p.55).

### **Aae watan ky sajilay jawano**

Use of the lexical items like beautiful (*sajilay*), sacrifice (*sarfarooshi*) and defender (*jo sarhadon ki hifazat*), depict how language is employed to get power. Language has a strong relationship

with power...it is ideological struggle that is of particular concern in the context because ideological struggle pre-eminently takes place in language (p.88). Discourse is constructed by keeping an idea in mind, having political or social interests. This song is not merely a tribute to the soldiers but there is also motif behind it. Language is employed to achieve set goals.

The phrases like *bewioun* , *behno* , *maoun ki nazrain* ( eyes of wives, sisters and mothers) *tumhain daikhain to youn jugmagaen*(shine to look at you), *khamoshioun sy dy raho houn tum ko duaen* (pray for you in silence)and *chand taroun ky aae raazdano* (secret tellers ) presents the ideology of patriotism, loyalty and unity. They are attributed with extra-ordinarily qualities. *Ideology certainly does not give the impression of single fixed meaning...* (p.93) connotes that single ideology was not at work during the war. National interests get prior importance while creating such a discourse, but the ideologies of patriotism and unity are also at work. People are rarely aware of *explicitly formulated or examined or questioned* (p.77). They are unaware of the agenda behind discourse. Language is helpful to fulfill the agenda of the powerful people.

### Meray Dhool Sipaheya

*Meray (my)* shows possession, authority and love for *sipaheya* (soldier) while *dhool* (handsome) is used to appreciate his physical beauty. Lexical items like (*Rub diyan rakhan*) blessings of Allah Almighty are associated with the soldier. They are being motivated to fight but also blessed with prayers of the nation. They are thrown into danger but also blessed by the people. He is being presented like a hero. *State power* controls everything including discourse (p.33). *Colonizing center*, associated with authorities creates discourse by keeping the ideology in mind (p.198).

Phrases like *jethay kadam jamawain* and *uthy kadry na hillain* express the determination and firmness of a soldier. Fairclough states that ideologies are ‘ideas which arise from a given set of material interests’ in the course of struggle of power (p.94). This patriotic song apparently pays tribute to the soldier but idea of power is hidden behind it which reveals that creating such a discourse is merely deception. His qualities are exaggerated through discourse but main interests are concealed. The soldier is appreciated to fight against the enemy while the prayers of people are with him. CDA exposes the reality of the discourse by raising doubts. His strength is used for the national interest by exaggerating his qualities.

### Apni jaan nazar karun

Positive attributes are given to a soldier like freedom fighter (*mard-e- mujahid*). He is narrated as a virtuous man who sacrifices his life for the defense of the country. *Ideological power* practices as a *universal* (p.33). He is presented as a savior. Fairclough states that *orders of discourse are ideologically shaped by power relations in social institutions and in society as a whole* (p.15). Such ideas are formulated which suit the social institutions and the whole society.

Use of phrases like *tu nay jala dala hai dushman ko shoal bun ky*( you have the power to defeat the enemy), *shujahat ka sila* ( reward of bravery), *teri juraat aur azmat ko salamat rakhay* (stay bless with your courage and eminence) , *jazba-e-shauk-e-shahadutt* (spirit of martyrdom), *dil main paida kiya jazba-e-taaza tu ny* (provide a new soul ) and *meray geetoun ko buksha hosla*



*tu ny* ( give strength to my songs) depict that discourse is masterly shaped by the institutions to get power and *power is exercised through language* (p.35). Political leaders and authorities get the fruits while the soldiers receive thorns. Oblation of their lives is advantageous for the nation as well as the authorities.

Fairclough advocates that media producers produce an *ideal* subject (p.49). The subject in these poems suit authorities. Political leaders were the beneficiary of war, not the common people. Singers, journalists and authors made politicians' work easier by manipulating the language. It was done to get power and authority. Higher authorities gained their strength after war.

CDA always unveils the curtains of falsehood and brings out the reality. *Modern society has colonized the discourse* (p.36). It is not free, but controlled by the social institutions. There is a *special relationship between ideology and exercise of power* (p.36). Ideology works to exercising the power. These songs pay tribute to the soldiers but nobody speaks against them. They also kill the soldiers of other side but they are credited by the nation. Such content is constructed which favors the ideology. Sweet voice of the singers, use of decorative words by the poets and melodious tunes by the musicians played a significant role in achieving the target.

These songs achieved the goal by uniting the people, by enhancing the spirit of patriotism among the nation, by preparing the youngsters to join Pak army and by misusing man's resources.

### 5. CONCLUSION OF HE STUDY

Overall finding is that the adverse reactions of war are not highlighted, but glorified. War does not bring prosperity, but massive destruction, chaos and mental disturbance. Ideology behind creating such a discourse is highlighted. These songs present message of unity, patriotism, loyalty and strength among the nation.

## REFERENCES



### Beyond Literal Meaning: A Pragmatic Analysis of Balti Language Proverbs

Author/s: Ansar Ali

Affiliation: Assistant Professor, Higher Education Department, KPK, Pakistan. (Email: ansaeastori@gmail.com)

**ABSTRACT:** This paper addresses pragmatic roles of proverbs in Balti language. It particularly addresses the aspect regarding the influence of contexts in which the proverbs are expressed in the text to determine their meanings. The study uses content analysis method in which the chosen proverbs are examined in detail as well as critically interpreted. There are certain proverbs which are taken and are analyzed according to the context to which they are used in the story. The results show that the proverbs have many uses such as issues warnings, praise, and insults, encourage, abuse, advice, and correct wrong behavior, among others. The research study also concludes that interpretation of proverbs is most effective when they interpret in a particular context in which they are applied. Besides, it reiterated that cultural as well as situational contexts are important in explaining the deeper meanings of the Balti proverbs when used in communication. The study also highlights the significance of instrumental knowledge to the efforts of linguists specifically those affiliated to the study of pragmatics to appreciate and capture a real meaning of proverbs across different cultures. This revelation indicates that, to effectively understand the meaning of proverbs, one needs to learn to interpret them in the context, and this can be used to decipher their hidden meanings.

Keywords: Literal Meaning, Pragmatic Analysis, Proverbs, Balti Language

To cite: Ali, A. (2025). Beyond Literal Meaning: A Pragmatic Analysis of Balti Language Proverbs, *TRANSLINGUA*, /1/, pages 79-89.

## 1. INTRODUCTION

The field of language is not as simple as using the right sequence of words in order to make a point. Although the definitions of individual words and the sequence thereof are a determining factor, the real meaning of a sentence is colored by a set of contextual factors (ex. the meaning of a sentence can be heavily influenced by the situation or setting in which it is adopted). An example of this is that in different situations the proverb *Sha zosy rospa khila* would be given different interpretation. There are also unspoken social rules and norms concerning language use which equally are significant in helping build meaning. E.g. speaking formally in the workplace as opposed to conversing casually with friends. In addition, it can also be appropriated by the implication of the meaning not expressed by words e.g. sarcasm or irony. The meaning and intonation of the speaker can have significant impacts in interpreting the sentence further since the interpretation can take on friendly pattern or hostile one. Finally, the cultural background and shared knowledge can also work in the meaning because some phrases or expressions can have various senses according to the cultures. A combination of all these factors helps to prove that language interpretation is a complicated process which is not limited to the literal meaning of words. Such situational background has the potential of enhancing elaborate content to a sentence thus complicating interpretation of language. Pragmatics is the study of the interaction of such situational factors with the actual meaning of the words to form a sense of language.

The proverbs are short statements that convey philosophical beliefs which are applied through the experiences of humankind in respect to cultural contexts. They provide the information about the history, world perception, and values of a community, not to mention that they also shape the social conduct and provide a possibility of mutual understanding. Such Nigerian authors as Chinua Achebe and Ola Rotimi have made proverbs use. Nevertheless, the definitions and uses of proverbs are less focused on the adages but determined by the context, thus, pragmatics is very important in analysis and interpretation. This paper discusses the practice of Balti proverbs as used by the Balti people in Baltistan and examines the contexts within which it appears with the aim of identifying their meaning, purpose and relevance to messages conveyed by the texts.

The beauty of proverbs is that they allow a complex idea and advise to be passed on in few words making it a good mode of communication (Odebunmi, 2006, Ogbulogo, 2002). Often with a proverb, a speaker is able to compress a great deal of sense and experience into only a small number of words. Thus, instead of having to explain that you are so late or why you are late, a teacher may simply say *eshan wakh sos* instantly conveying to the person that it is important to come in time. Such cultural understanding and exposure to proverbs make them time-saving ways to present information and make the intended message easily understood by listeners because the shortened information is directly relevant to the listener.

Proverbs ensure a survival of a culture and its practices. They come with ancient wisdom, life lessons and moral values of the older generations to the younger generation. Proverbs are quite a powerful means through which crucial knowledge can be shared in places such as Baltistan, where the majority of people tend to pass knowledge through words. When one utters a familiar

saying, this provokes in minds of other people the intimations of certain common credence and experience by causing people to become more connected to the society.

Proverbs also come in handy whenever one is trying to describe something or persuade others. They are used by people with an argument or dialogue in order to prove their point. Better proverb may contribute to a solution to a problem in Balti culture, a smart or funny detail, or a matter that is used to make a dialogue easier. People usually respect those who use proverbs since most of them regard them as being wise. In order to comprehend a proverb, it is necessary to be provided with context and words. It is not enough to learn what number of words mean, but one should also know when and where to use these words. An example is a proverb about patience: It would feel reassuring when an individual is depressed but would be different otherwise, at a busy workplace. When the proverb is employed in an incorrect case it can end up either causing confusion or dejection to the other people. That is why it is quite important to know when and where a proverb is to be used.

Many Baltistan proverbs grow out of the country and lifestyle here. Quite a number of them speak of nature, agriculture or coexistence as a community. These proverbs make us realize the differences in how Balti people perceive the reality, how they think the world is and what is considered important in this life. Learning of these sayings not only makes us understand how incremental Balti is, but also how Balti people live, how they fought and how they learnt. The research questions framed for the study are:

1. In what ways does contextual cultural and circumstantial determination modify the use and meaning of Balti Proverbs?
2. In what way may we use the theories of Contextualization and Politeness to evaluate the definition and purpose of Balti proverbs?

## 2. LITERATURE REVIEW

Pragmatics deals with the context of language and how it is employed to pass information and have certain objectives fulfilled. Pragmatics analyses the use of language based on a social interaction with consideration of the implicit knowledge in the interpretation of the use of language (Birner, 2013). Mey (2001) expands on this position by defining pragmatics as a science of the language in reference to individuals who use languages by studying the actual communication in the real world and how people use language to suit their own interest.

Viewing pragmatics as a systematic study of meaning based on the usage of a language, Huang (2007) focuses on the study of the context-controlled meaning, not only about formal word qualities. According to Leech (2014), pragmatics is called the study of meaning, relative to speech situations in which the language has an impact on the attitude of the hearer. Likewise, Yule (2002) indicates that the speech situation helps the users of language to manipulate words to accomplish a specific effect on the listener.

The possibility of using proverbs as a means of indirect communication may also be interpreted as indirect communication, as most scholars argue (Finnegan 1970). Because proverb is used

by the sender to convey messages in ways perceive as culturally acceptable particularly under circumstances where direct speech may have been deemed inappropriate. Levinson (1983) makes it clear that pragmatic is about the context in which language is to be used like how the meaning of words and sentences can be understood and interpreted without taking literal meaning. In the example of Balti proverbs, the given context of particular society and culture of this region is needed to identify the intended meaning of the given proverb.

Ariel (2008) thinks of pragmatics as the study of relationship between language and context which is pertinent to language grammar and Yule (1996) merely defines pragmatics as the study of what is intended by a speaker based on the context. The combined definition is a reflection of the emphasis that pragmatics lay on language use, context and user friendly communication.

The literature of proverbs in African culture is something that has just been researched. Fashina (2011) refutes the fact that proverbs should be as wise and beneficial as they always are, as they may be vague and ambiguous. Through the use of critical theory and linguistics, he analyses novels of Chinua Achebe and concluded that proverbs serve not only aesthetic, but also affective purposes and can be confusing.

The study by Ehineni (2016a, 2016b) is conducted on Yoruba proverbs, which demonstrate their significance in expressing other ideas and the intentions. Through his analysis these proverbs which are used by the Yoruba people have peculiar lexical and grammatical structure that makes it easy to use it in plays to get the intended information pass across. Besides, Ehineni shows how proverbs help to underpin different discourse activities and communication objectives.

Gogoi (2017) examined how a proverb is used in *Things Fall Apart* and *No Longer at Ease*, by Chinua Achebe. African sensibilities are expressed through creative utilization of oral culture factors by Achebe, which include use of proverbs, folktales, and myths. According to Gogoi, Achebe (2017) makes use of strategic use of proverb to comply with various social, cultural situations to general political issues.

Alimi (2012) reviews his (Achebe) usage of proverb as an element of literature in *Things Fall Apart* and *The Arrow of God*. Alimi discovers that proverbs assist in explaining the appearance of characters, their activities, and thoughts; therefore, Achebe has defined characters well.

It is through studying language within the social situation that pragmatics can be useful in discovering how people make meaning in the course of interaction. Pragmatics is crucial to communicate better, because it puts into perspective the intricate connections among language and surroundings, as well as the dynamics among people.

### 3. RESEARCH METHODOLOGY

In this study, a qualitative analysis of the content will be used in order to analyze the pragmatic roles of proverbs in Baltistan concerning its meaning, context, and culture. The information is in the form of a few proverbs, gathered through oral culture, folklore, and observing people in

social life. The use of the elders and folk literature in the community and interviews among them also enhances the data set. The proverbs have been selected based on their relevance in the culture, their usage, and their popularity in pragmatic functions, viz., advising, warning, and criticizing. The analysis is based on the pragmatic theories, mainly Contextualization Theory and Politeness Theory. Each proverb assesses on how it fits in particular societal contexts and in this regard interpretation of meanings is made in regards to the cultural and situational context of proverbs application in the society. The Politeness Theory also lets us know the Balti sayings and how they portray societal norms such as respect and order. To validate the findings, we checked them using the local elders and various data collations. Regarding respect of local values, the study promotes ethics, such as the informed consent and cultural sensitivity.

#### 4. RESULTS AND FINDINGS

This paper attempts a thorough analysis on the pragmatic roles of Balti proverbs within the perspective of two different theories: Politeness Theory and the contextualization theory. By the means of these theories, guiding the research towards the ancient expression of Balti proverbs, the researcher would determine to find out the intricate dynamics of language, culture and communication which are the foundations of the Balti proverbs. Politeness Theory is used to look at how Balti proverbs enable the negotiation of social relationship and how they express respect, whereas Contextualization Theory is used to understand the cultural and historical context in which the meaning and significance of these proverbs permeate differently. Having employed such a multi-theoretical approach, this paper aims to be one of the contributions toward a better comprehension of complications of the language, culture, and communication interrelationship in relation to Balti proverbs.

##### 4.1. Contextualization theory:

The Gumperz(1982) shrouds this theory concentrates on the meaning, which was employed in the social and situational context wherein communication occurs. The theory is especially reflective in interpretational situations with the pragmatic functions of proverbs since the meaning a proverb is intended to convey can change according to the speaker, to whom, in what context, and the culture surrounding the norms involved. The proverbs always have a more flexible meaning which depends on the above mentioned contextual factors The ways the speakers handle social roles, relations and perform them in the course of communication, and the ways proverbs are changed to fit into the various social processes, expectations, and designs are seen in the use of the proverbs. Contextualization Theory makes us understand that it is not true that proverbs always bear definite meanings, but rather moulded by the circumstance to allow them to perform different tasks depending on the context, including advising, complaining or persuading, as the case may be.

1: *Bongi buna bili bu*

The children of a daughter like the kitten of a cat).

The Balti proverb is employed to criticize grandchildren, when they disobey their grandparents. It is a figurative statement where children are literally likened to the kitten since they have the same trait of disloyalty. Similar to a kitten, grandchildren do not obey their maternal grandparent since they are not very close to him/her. The saying enlightens the conventional culture of daughters leaving the house of their father in marriage that leads to physical and emotional separation to their children.

2: *khumulchu medna khser med*

No pain no gain follows.

These vary depending on the context on which they are used, but the implication of the balti proverb lies in this. The proverb finds application in various circumstances, like during teaching whereby the teacher uses the proverb to motivate the students. In another scenario it lays stress on hard work as the key to success. The lifestyle of struggle is admirable since nothing can be achieved without struggle.

3: *Ha msk L Ph L ha handoq*

Get out (well before the work is done)

When they forecast before the event, their own criticizing common use of the Balti proverb is used. The proverb highlights the nature of information to spread rapidly, in most cases, prior to completion of the work or the task due to the nature of the society and communities that is culturally centered in gossips, rumors as well as news and travels via social networks at a very high speed. The proverb in itself is an observation of the society and the reality of information dissimulation whereby news and rumors can travel faster than real facts.

4. *Laq bjy , lung bjy*

( to do good turn and hope of receiving rewards)

The Latin word, laq bjy or lung bjy, a proverb which states that kindness and goodness knows no boundaries but is transectional as people expect a similar act or favour in response to their own activity. This is based on the cultural and social standards of the communities where kindness may be considered as a mode of investment and the reward is expected on a later date.

5. *Skaboe bareing xsharpha chon shiks*

(wet wood also burns along with the dry wood)

The proverb is an expression of the community life in terms of experiences and observations made in the society. In this regard, the proverb emphasizes the process of innocent person getting shipped by forces beyond his or her control as it would be with wet wood which burns together with dry one. Such contextualization matters a lot since it shows how the society approaches justice, accountability, and how the community actions can influence the lives of the individuals.



6. *Uckeasing zheek meed na mikeyk la tikmeed.*

( *There will be no tear In eyes and no pain In the heart*)

Without heart pangs no eye-drops, Prayer is to the soul what tear is to the eye. In each case there must be pangs of pain, It is a deep saying that attracts attention to the cause and effect correlation between human feelings and bodily action such as tears. Also, it declares that when a part of the human body is in pain, then the whole body is in pain. All the parts to the body suffer equally. The proverb has an application whereby a certain person is pretending to be sad, without logical reasoning.

7. *Hlo nara qi Hltaragi handoq khro chorong.*

*Then the guest mocks at the host by saying that: when did you purchase this carpet?*

This proverb is applied when guest visits the impoverished family and spot the laying sheet instead of the carpet. Then the guest mocks at the host by saying that: when did you purchase this carpet? This is a proverb that implies in the relationship between the financial status of a person and the things that he/she owns. The host makes an implying statement with the help of this proverb, referring to the fact that his or her guest has misdirected assumptions about the carpet.

#### 4.2. Politeness theory:-

Proposed in 1987 by Brown and Levinson, Politeness Theory examines how individuals can more or less preserve social harmony by taking care of their own and other people's face or self-impression in society. It makes a distinction between two faces: positive face which is the desire to be liked and approved by and negative face which is the wish to be free and independent. During the communication process, people usually engage in face-threatening acts (FTAs) in the form of requests, commands, or criticism that may hurt the face of a listener. In an effort to downplay this threat, speakers adopt a number of politeness strategies.

The bald-on-record style is the first one, in which the speaker is blunt and straightforward and does not soften the message e.g., by saying, "Close the window." This is normally applied in case of emergencies or to those persons who care very much. An example of such straightforwardness is the proverb: Honesty is the best policy.

Positive politeness comprises of friendliness and underlining of a common ground in order to make the listener feel important. It may contain something like compliments, colloquialisms, or inclusive terms and one of them is the following phrase, which is expressed like this, "hey, can you close the window, can you. It is getting chilly." The same can be said about a related proverb called you can attract more flies with honey than with vinegar implying that are nice and warm.

Negative politeness, conversely, respects the personal territory and autonomy of the listener and may employ a formal or indirect manner of phrase or expression, as in, example, would

you mind shutting the window, please?” This is implied in the saying that it is better to ask than to presume and this means not to be untidy in imposing.

The off-record strategy is even less direct where what one wants is suggested not exactly said. As in saying it is chilly in here rather than asking someone to close a window.

Balti proverbs are discussed with the aid of politeness strategies that help to identify how indirectness, respect, and social harmony may be reflected through the traditional sayings. During the frameworks of Brown and Levinson politeness theory, an analysis of whether a proverb makes use of direct or indirect language is carried out, and to what extent such values as humility, honor, or criticism are expressed. Such strategies are useful in neutralizing the sound of brutal realities or tactless advice in the proverb so that they fit in dignified or respectful discussions. This method illuminates the complexity of the connection between language, culture, and society, the Balti communal attachment.

### 1:-*Hrtala rgatse hrta rzila*

The Balti saying is *Hrtala rgatse hrta rzila* which is translated into English as A horseman will love even more than he/she loves his/her horse, the individual who tends his/her horse. It is a positive politeness proverb according to the Politeness Theory. One point this proverb brings out is the raising of affection and appreciation to those who appreciate what others care about. The proverb contains a more profound statement because it involves complimenting the person who takes care of the horse: being kind, being responsible and being thoughtful establish the emotional connection between people. It suggests socially rewarded conduct and reinforces the unity in the society. Rather than criticizing or distance creation, the proverb brings about closeness and mutual respect as well as a positive relationship all of which are major elements of positive politeness.

### 2:-*Ghzorba dukse laq hrngas*

Ghzorba dukse laq hrngas as a Balti proverb may be translated as still mowing grass by hand when the sickle is at hand. They use it in rebuking a son who issues commands in the house when the father is present in the house. This proverb criticizes the son indirectly because he overestimates his scope of responsibility and behaves as the man of the house when the actual supreme is the father who happens to be around hence it is negative politeness. The proverb does not directly scold him, but politely puts him in his place.

### 3:-*What has Lasmi said, chuk hrkunma bretchuk*

Politeness Theory is reinforced by the Balti proverb, *Lasmi that chuk hrkunma bretchuk*, meaning, as a direct translation, entertain the hardy workers and punish the thief severely so that he may not be able to steal any more, and also includes a cross of positive politeness and bald on-record. The positive politeness aspect of the proverb is the first portion of the quote advising people to respect and take care of people who work hard. It encourages gratitude, equality, and acknowledgment of efforts made by people with honesty. This type of linguistic reinforces the sociality ties and demonstrates the appreciation of the cooperation with the community. On the contrary, punishment of a thief on the brutal side of the second part of this

proverb is a bald on-record approach. The approach is straightforward, categorical, and hard to misunderstand which is not bad in a scenario where rules and justice must be well established.

4:-*Kho khara sningp ro ara*

An example of the off-record politeness is the Balti proverb, meaning “a malign person with a sweet tongue” or “a fair face may hide foul heart” and is given out as *Kho khara Sningpo ara*. The off-record politeness compels a speaker of a message to say nothing explicitly; hence, the listener is left to interpret what she/he really means. This proverb does not blame anyone but it presents a kind of warning or caution about the people who talk sweetly with ill intentions. It sends a serious message, in an understated manner- without confronting the issue. It assists the speaker to express concern or caution without provoking a conflict since it involves metaphor and indirect language.

5:-*BalBisi xyangspa nasbui stroqla*

An example of off-record politeness is: the Balti proverb, which states, *Balbisi xyangspa nasbui stroqla*, that is, the sparrow died even in the palms of his or her own nearest child. This saying point to the unintentional damage that innocent or immature people can create in the form of actions brought on by curiosity or playfulness including killing a sparrow. On a wider scale, it condemns people in authority or those without the empathy to know better, where killing takes places to satisfy selfish needs or gain glory including kings or rulers to fulfill their luxury or fame.

6:-*Xhoila zere xlangla*

A very explicit representation of off-record politeness technique in the Theory of Politeness is the Balti proverb that combines a direct meaning with an unspoken one, that is, *Xshoila zere xlangla* which can be translated as To ask something of the ploughing ox by speaking to the plough described as the Theory of Politeness.. The proverb refers to an instance when someone is shy or afraid to request assistance (e.g., to borrow an ox) directly. As an alternative, he or she speaks about something akin to it (e.g., the plough) to start the communication process. This is the actual request which is perceived indirectly. This will enable the speaker not to be embarrassed, respect and save face to both the parties. The speaker is not open in making a demand, instead he has figured out how to use an implied request with the hope that the listener can interpret what is really implied.

7:-*Towana byarna skinoq, rgomna byarna phe*

The indirect approach to off-record politeness can be traced in the advertisement going, *Towana byarna skinoq, rgomna byarna phe*; (if you sit near cooking, you are sure to be soiled: if you sit near flour box, the flour will make your dress white). This idiom does not condemn or commend anybody in particular; rather, his symbolic application of cooking pot (which refers to being dirty) and flour box (which refers to being clean) is intended to mean that one should take care in the company that he or she keeps. It teaches a social lesson in an oblique way: an association with good persons brings a good name, an association with bad ones brings shame.

### 4.3. Finding of the study:-

It was discovered by the study that Balti proverbs have multiple pragmatic roles in the society such as symbols of communication in society, as a source of cultural expression as well as a means of control. The functions are giving advice, making warning, praise, criticism or insults, expressing abuse, encouraging others, correcting improper behavior and making assertive statements. The discussion revealed that what these proverbs would mean largely depends upon the context of their utterance (who is saying them, who is listening, what social and cultural nook it is amplified in). Use of proverbs is an indication of Balti speakers who use indirectness, metaphor and common cultural knowledge as a means to convey difficult or embarrassing concepts in a socially acceptable way. In using Contextualization Theory, the research demonstrated that proverb interpretation cannot be separated; it depends on how people apply it in a situation because the context of the use of a proverb is what determines the shift of meaning. Also, according to the Politeness Theory, it was revealed that Balti proverbs frequently represent the strategies serving the ideas of social peace, respect, and hierarchy, positive politeness (compliments to favorable behavior), negative politeness (email-style suggestions), off-record strategies (implicit hints), and bald on-record strategies (explicit requests). These practical instruments can express indirect meanings without any loss of interpersonal sensitivity and social cohesion.

## 5. CONCLUSION

The paper has concluded that there is more to learning proverbs of the Balti than just knowing its literary meaning. It is not the literal translation of these proverbs that made them so effective and deep but their contextual and practical use to the social paradigm of Baltistan. Proverbs are not permanent word anatomies but fluid communications that can be changed to suit particular communicative demands, and with wisdom, criticism, encouragement and values, communicated culturally sensitively. The study points out that the pragmatic ability, especially the ability to recognise the context and the politeness convention, is critical in the more precise interpretation of proverbs in different cultures. The insight is particularly of use to the linguists, educators, and researchers in the field of pragmatics and intercultural communication. Finally, the research highlights the point that proverbs are not just forms of style in speaking; they are parts of culture that help control the relationships, express the societal ideas about the ways things should be done, and transmit the traditional knowledge with the help of pure forms of speech in indirect yet extremely effective ways.

REFERENCES

- Hassan Hasni, Ghulam, (2004). Tam lo (Balti proverbs)
- Ariel. M. Pragmatics and Grammar, Cambridge: Cambridge University Press, (2008)
- Yule. G., Pragmatics, Oxford: Oxford University Press, (1996)
- Fashina. N. O., Proverbs and Linguistic Meta-Criticism: Towards a Re-reading of Proverbs as Narrative Sublimation in Achebe's Things Fall Apart, Journal of the Linguistic Association of Nigeria, 14(2)(2011), 283-959
- Yule. G., A Study of Language, Cambridge: Cambridge University Press, (2002)
- Leech. G., The Pragmatics of Politeness, Oxford: Oxford University Press, (2014)
- Birner.B.J., Introduction to pragmatic, United kingdom: Blackwell Publishing(2013)
- Odenbumni.A.,A pragmatic and 'Dry leaves on Ukran trees, intercultural pragmatic,3(2)(2006),153-170.
- Huang.Y., pragmatic. Oxford university press,(2007)
- Mey.J.L.L., Pragmatic: An introduction (2nd ed ), Oxford: Blackwell,(2001)