

**Lexical Bundles in Source vs. Target Texts: A Corpus-Driven Study of Translation Shifts
in English and Urdu Newspaper Articles**

Zainab Noor

Alaia.khaan@gmail.com

MS English Linguistics Scholar COMSATS University Islamabad, Pakistan

Moneeba Habib

moniba.umar@comsats.edu.pk

PhD Scholar, NUML.

Faculty English Linguistics, COMSATS University Islamabad, Pakistan

Corresponding Author: * Zainab Noor Alaia.khaan@gmail.com

Received: 20-10-2025

Revised: 16-11-2025

Accepted: 29-11-2025

Published: 15-12-2025

ABSTRACT

Lexical Bundles (LBs) are high-frequency multi-word sequences serving key functional roles in discourse, and Translational Studies (TS) examines how these units transfer across languages. Many studies have conducted research on the linguistic properties of formulaic language and general translation shifts, but limited attention is given to the unique challenges of translating high-frequency LBs between languages that exhibit a significant typological distance, such as the SVO structure of English and the SOV structure of Urdu. Therefore, the current study investigates the specific shifts and strategies employed in the English-to-Urdu translation of four-word lexical bundles in journalistic discourse. The study employs a qualitative, comparative corpus-based methodology to analyze functional, structural, and translational shifts between the English source texts and the Urdu target texts. The findings reveal a consistent pattern of “de-bundling” in Urdu translations, where fixed English bundles are often rendered as dispersed or paraphrased expressions due to typological differences between English (SVO, prepositional) and Urdu (SOV, postpositional). The study highlights the dominance of paraphrase, replacement, and omission strategies, underscoring the need for translators to prioritize functional equivalence over structural fidelity. These insights contribute to translation pedagogy and theory by emphasizing the system-bound nature of lexical bundles and the importance of register-sensitive, typology-aware translation strategies.

Keywords: Lexical Bundles, Translation Studies, English–Urdu Translation, Journalistic Discourse, Corpus-Based Analysis, Translational Shifts, Typological Differences, Functional Equivalence

INTRODUCTION

The domain of contemporary translation studies has increasingly focused on lexical bundles (LBs), which are recognized as a vital component of formulaic language (Biber et al., 1999; Hyland, 2008; Salazar, 2014). These bundles are defined as recurrent sequences of three or more words that occur with high frequency across specialized texts, functioning as essential “building blocks of discourse” (Biber & Barbieri, 2007, p. 263), even though they may not represent complete grammatical or semantic units. Their primary role lies in structuring discourse and articulating the writer’s stance or evaluative position (Cortes, 2004; Ädel & Erman, 2012). Particularly in journalistic writing, where clarity, attribution, and narrative flow are paramount, these statistically frequent and genre-specific sequences are instrumental in maintaining coherence and guiding information progression (Hyland, 2008; Chen & Baker, 2010).

Although substantial research has mapped the structural and functional dimensions of LBs in English corpora, especially within academic and news discourse (Biber et al., 2004; Pan & Liu, 2019), comparative studies across languages remain limited. This gap is especially pronounced in translation contexts involving

non-Western languages such as Urdu, where the behavior of LBs in translated texts is still underexplored (Zahid & Hussain, 2023; Malik et al., 2023).

The rise of corpus linguistics has significantly reshaped translation studies by enabling empirical investigations into the linguistic patterns of translated texts (Baker, 1996; Laviosa, 2002). Parallel corpora, in particular, offer a robust methodological framework for identifying translation shifts: systematic modifications that occur when linguistic elements are transferred from a source text (ST) to a target text (TT) (Toury, 1995; Huang, 2023). Within this framework, LBs provide a valuable lens for examining translation universals such as simplification, explicitation, and normalization (Baker, 1996; *Frontiers in Psychology*, 2021), as well as the influence of translation directionality (Huang, 2023). While cross-linguistic studies have examined English–Chinese, English–German, and English–Arabic pairs (Chen & Baker, 2010; Huang, 2023), English–Urdu comparisons remain critically underrepresented in the literature (Zahid & Hussain, 2023).

In Pakistan’s complex sociolinguistic landscape, where English and Urdu coexist in a dynamic and often asymmetrical relationship, the translation of journalistic texts offers a fertile ground for exploring how formulaic language is negotiated across linguistic and cultural boundaries (Malik et al., 2023). Given the syntactic and morphological differences between English (a fixed word-order, SVO language) and Urdu (a more flexible, SOV-structured and morphologically rich language), the translation of LBs is likely to involve significant structural and functional shifts (Zahid & Hussain, 2023). However, existing research has largely overlooked this terrain, with limited attention to the strategies translators employ when rendering English LBs into Urdu and the implications of these strategies for discourse coherence and genre conventions.

This study addresses this critical gap by conducting a qualitative, corpus-driven analysis of 25 English–Urdu parallel newspaper articles. By identifying high-frequency lexical bundles in English source texts and tracing their equivalents in Urdu target texts, the research aims to uncover the nature of translation shifts, whether bundles are preserved, adapted, omitted, or substituted, and to evaluate the structural and functional transformations that occur in the process. In doing so, the study contributes to a deeper understanding of how formulaic language is reshaped in translation.

Problem Statement

Lexical bundles are essential for structuring discourse and conveying stance, particularly in journalistic writing, where they contribute to coherence, attribution, and narrative flow. Despite increasing scholarly interest in lexical bundles and their role in translation, most corpus-driven studies have focused on European and East Asian language pairs, with limited attention to non-Western languages like Urdu. Specifically, there is a lack of research on how English lexical bundles are structurally and functionally transformed in Urdu newspaper translations. Therefore, this study aims to fill that gap by conducting a qualitative, corpus-driven analysis of 25 English–Urdu parallel newspaper articles, identifying how LBs are preserved, adapted, or transformed during translation.

Research Objectives

- To identify high-frequency lexical bundles within a corpus of English-language newspaper articles
- To examine the translated equivalents or transformation of these lexical bundles in a parallel corpus of Urdu-language newspaper articles
- To analyze the nature of translation shifts in lexical bundles in terms of frequency, structural configuration, and functional role

- To evaluate the translation strategies employed in rendering lexical bundles from English to Urdu, with particular attention to instances of preservation, adaptation, omission, or substitution

Research Questions

- What are the high-frequency lexical bundles within a corpus of English-language newspaper articles?
- What are the translated equivalents or transformations of these lexical bundles in a parallel corpus of Urdu-language newspaper articles?
- What is the nature of translation shifts in lexical bundles in terms of frequency, structural configuration, and functional role?
- What translation strategies are employed in rendering lexical bundles from English to Urdu, with particular attention to instances of preservation, adaptation, omission, or substitution?

Significance

The findings of this study offer a significant contribution to both Descriptive Translation Studies (DTS) and Translation Pedagogy by empirically demonstrating how linguistic typology constrains the transfer of formulaic language across structurally divergent languages (Indo-European SVO → Indo-Aryan SOV). By analyzing the systematic transformation of high-frequency English lexical bundles into Urdu, the study reveals that such bundles are not universally transferrable units but are instead shaped by the syntactic and stylistic norms of the target language. This insight advances typological translation theory by showing that strategies like Replacement and Paraphrase are not merely stylistic choices but necessary adaptations to maintain naturalness and coherence in the target text. Furthermore, the study sheds light on the distinctive features of the Urdu journalistic register, particularly the reduced reliance on overt discourse-organizing bundles, suggesting that translators actively negotiate between source-text explicitness and target-audience expectations. Pedagogically, these findings underscore the need to train translators in recognizing the pragmatic functions of lexical bundles and in applying functionally equivalent, idiomatic Urdu structures rather than defaulting to literal translation. In doing so, the study not only refines our theoretical understanding of formulaic language in translation but also equips educators and practitioners with data-driven strategies to enhance translational competence in English–Urdu news translation.

Limitations

The primary limitation of this study stems from its corpus size and domain specificity. The findings are based solely on a closed corpus of English-to-Urdu journalistic texts (political, editorial, and opinion genres). This means the observed translation patterns and strategic choices (like the high rate of de-bundling) may not be universally applicable to other genres, such as literary, scientific, or legal translation. Furthermore, the analysis is entirely descriptive; it reports what translation strategies were used, but cannot definitively quantify the impact of these shifts on the Urdu reader's comprehension or perception. The findings are thus a statement of observed translation practices, not necessarily a prescription for optimal translation quality.

Delimitations

The study was specifically delimited in three key areas to maintain focus and manage scope:

- Directionality: The analysis was strictly limited to the English-to-Urdu translation direction, excluding the reverse (Urdu-to-English).

- Focus: The study exclusively focused on four-word lexical bundles (LBs), deliberately excluding three-word bundles, collocations, idioms, or other formulaic sequences to maintain a clear analytical scope on highly fixed, multi-word units.
- Genre: The corpus was restricted to political, editorial, and opinion texts from news media, excluding general news reporting and other non-journalistic sources.

LITERATURE REVIEW

The Nature and Function of Lexical Bundles

Lexical bundles (LBs), also referred to as n-grams or multi-word units, are sequences of three or more words that show a significantly higher frequency of co-occurrence in a specific genre or register than would be expected by chance (Biber et al., 1999). Unlike idioms, which are often metaphorical and have fixed internal structures, LBs are primarily defined by their statistical frequency and function in discourse (Crosthwaite & Franco, 2018).

Structural Patterns of Lexical Bundles

In addition to their functional classification, LBs can be analyzed based on their internal structural patterns. Identifying these patterns reveals the preferred ways speakers and writers construct meaning and sentence frames within a given genre (Biber et al., 1999). Four major structural categories are commonly identified in English (Biber et al., 2004):

- Noun Phrase (NP) based: These bundles often involve a phrase fragment beginning with a preposition or a determiner, frequently serving as referential expressions. Examples: "of the study," "the end of the," "a part of the."
- Verb Phrase (VP) based: These contain a verb and its complements, often expressing stance or reporting. Examples: "it is important to," "it has been suggested that," "is expected to be."
- Prepositional Phrase (PP) based: These bundles typically begin with a preposition and are used to indicate relationships of time, place, cause, or manner. They often function as adverbials or modifiers within a sentence. Examples: "at the same time," "at the beginning of," "on the basis of."
- Dependent Clause (DC) based: These function as parts of clauses, often used as discourse organizers to link ideas. Examples: "as a result of the," "if you look at," "it should be noted that."

Analyzing these structures is critical for translation studies because structural differences between the source and target languages (e.g., English SVO vs. Urdu SOV/free word order) are expected to necessitate changes in the LB's structural pattern, even if its function is preserved (Dechert & Dechert, 2012).

Functional Categories of Lexical-Bundles

LBs are typically categorized into three main functional groups (Hyland, 2008):

- Referential expressions: Direct attention to specific text components, e.g., "the purpose of the study."
- Stance expressions: Convey the author's attitude, certainty, or commitment, e.g., "it is important to note."
- Discourse organizers: Manage the flow and structure of the text, e.g., "on the other hand" or "as a result of."

Studies have consistently demonstrated that LBs are key markers of register and genre (Hyland, 2008). For instance, news reports tend to favor bundles related to fact-reporting and attribution, while academic texts rely more heavily on stance and discourse organization bundles (Biber & Barbieri, 2007). This genre-specificity suggests that shifts in LB usage during translation can reveal important differences in the stylistic conventions of the source and target languages/cultures.

Corpus Linguistics and Translation Studies

The advent of corpus linguistics has revolutionized translation studies, moving analysis from anecdotal observation to empirical, large-scale investigation (Baker, 1993). Specifically, the use of parallel corpora, i.e. texts in a source language aligned with their translations in a target language, allows researchers to systematically compare linguistic features, including LBs, across the translation divide (Kenny, 2001).

Corpus-driven studies have successfully identified several universal or quasi-universal features of translated language, known as translation universals (Baker, 1996). Relevant to this project, two key universals are simplification (translators unconsciously or consciously simplify complex source text structures) and normalization (translated texts tend to conform to the typical linguistic patterns of the target language, potentially making them more conventional than non-translated target texts) (Toury, 1995). The study of LBs in parallel corpora can provide concrete evidence for these universals: if LBs are simplified, their frequency or structural complexity might decrease; if normalization occurs, unique source text LBs might be replaced by highly frequent, conventional target language equivalents (Toury, 1995). Furthermore, studies on translation directionality have recently shown that whether a translator is working into their native language (L1) or a foreign language (L2) can significantly impact the preservation rate and structural complexity of multi-word units (Huang, 2023).

Cross-Linguistic Research on Lexical Bundle Shifts

While a robust body of research exists on LBs within monolingual English corpora (Biber et al., 1999; Hyland, 2008), the comparison of LBs in English as a Source Text (ST) and a Target Text (TT) of a non-Western language remains limited, particularly involving Urdu.

Existing cross-linguistic studies, such as English-Chinese (Crosthwaite & Franco, 2018) or English-German (Dechert & Dechert, 2012), show significant translation shifts in LB use. These shifts often include segmentation (breaking a bundle) or expansion (adding words), driven by structural differences. More recent work confirms this focus on pragmatic and structural shifts. Pérez-Vidal and Abarca (2022), for instance, used parallel corpora to analyze how the pragmatic functions embedded in source text phraseology shift or are maintained in the target language. Crucially, research comparing English with other non-Western languages, such as the English-Arabic news study by Al-Zoubi et al. (2020), reinforces the finding that cultural and generic norms dictate translators' choices, leading to non-literal substitutions of formulaic expressions.

Lexical Bundles in the Pakistani Context

In the Pakistani context, the methodological groundwork for corpus studies is being established. Zahid and Hussain (2023) reviewed corpus methodologies for extracting lexical bundles, emphasizing the necessity of accounting for genre, register, and language-specific constraints—all highly relevant to comparing English and Urdu journalistic texts.

However, research on Urdu multiword expressions is scarce. While Qureshi et al. (2022) explored Urdu phraseology, they focused on English learners and did not extend their work to lexical bundles in translated texts. This lack of research leaves a critical gap in understanding how the distinct morphological and syntactic structure of Urdu, an Indo-Aryan language, influences the preservation or transformation of complex English discourse structures. This observation is further supported by contemporary comparative linguistic research in the region, which underscores the structural and conventional differences that necessitate adaptation when translating complex phraseology (Ahmad & Khan, 2024).

A significant gap persists in the qualitative analysis of the translation strategies used to render English lexical bundles into Urdu, specifically within the news discourse genre. Existing research lacks an empirical and detailed micro-level investigation into translator choices (literal, paraphrased, omitted, or replaced) and the resulting functional and structural shifts when LBs are translated from English to Urdu newspaper articles. The current study addresses this gap by performing a qualitative, corpus-driven analysis to identify the translational shifts of lexical bundles between English source texts and Urdu target texts, based on a corpus of 25 parallel news articles in each language.

METHODOLOGY AND THEORETICAL FRAMEWORK

Theoretical Framework

This theoretical framework establishes the conceptual basis for a corpus-driven study on the translation shifts of lexical bundles (LBs) from English (Source Text/ST) to Urdu (Target Text/TT) in newspaper articles. It integrates theories of corpus linguistics, lexical bundles, and translation studies to provide a robust structure for analyzing translation phenomena between English and Urdu.

Corpus Linguistics and the Nature of Lexical Bundles

The study is fundamentally rooted in corpus linguistics, a methodology that uses large collections of texts (corpora) to analyze linguistic phenomena based on empirical frequency and contextual distribution. This approach is essential for identifying patterns that are often invisible to traditional introspection.

- **Lexical Bundles as Units of Discourse**

The central theoretical construct is the Lexical Bundle (LB), defined as recurring sequences of three or more words, irrespective of their word-class boundaries, that function as basic building blocks in a specific register (Biber et al., 1999). LBs are distinct from idioms or collocations because they are primarily defined by their statistical frequency in a corpus, and their meaning is often non-idiomatic and tied to their discourse function.

Hyland's Functional Taxonomy (Hyland, 2008): The study adopts a functional framework to move beyond purely grammatical analysis. Lexical bundles are theorized to serve three primary discourse roles, which directly impact coherence and communicative effect:

- Referential Bundles: Focus on the representation of people, places, things, or time (e.g., at the end of the).
- Stance Bundles: Express attitude, certainty, or evaluation, acting as crucial markers of writer voice and subjectivity (e.g., it is important to).
- Discourse-Organizing Bundles: Manage the flow of information, structuring the text by introducing topics or transitioning between ideas (e.g., on the other hand).

Biber's Structural Taxonomy (Biber et al., 2004): Complementing the functional view, the study uses a grammatical classification to analyze the formal makeup of LBs. This structural analysis (e.g., Noun phrase + of, Prepositional phrase fragments, Verb phrase fragments) is crucial for identifying how the syntactic constraints of English are realized or transformed in Urdu.

The theoretical relevance here is that a shift in the structural form of an LB during translation may, but does not necessarily, entail a shift in its discourse function, forming a key point of analysis.

Translation Studies: Equivalence and Shifts

The study's focus on the transformation of LBs from a Source Text (ST) to a Target Text (TT) necessitates a foundation in translation theory, particularly concerning equivalence and translation shifts.

- **The Concept of Equivalence in Parallel Corpora**

A parallel corpus, by its nature, invites the study of translation equivalence. While absolute equivalence is often unattainable, this study views equivalence not as a static matching of single words, but as a contextual and functional correspondence at the multi-word level.

Toury's Norms (Toury, 1995): The translator's choices (strategies) in rendering LBs are theorized as a manifestation of translation norms. These norms—initial (general adherence to ST or TT culture), preliminary (policy and strategy), and operational (text-level choices)—govern the preservation, adaptation, or omission of LBs. The systematic analysis of translation strategies (literal, paraphrased, omitted, replaced) will uncover the prevailing operational norms in the English-to-Urdu newspaper translation domain.

- **Translation Shifts**

Catford's Shifts (Catford, 1965): While Catford's model focuses on formal correspondence, the present study adapts this concept to the LB level. The analysis of Structural Shifts (e.g., a 4-word English LB being rendered as a 5-word Urdu phrase or a single compound noun) and Functional Shifts (e.g., a Stance LB in English being rendered as a purely Referential expression in Urdu) allows for a systematic description of the transformations.

Coherence and Register: Given the journalistic context, the framework acknowledges that shifts are often motivated by the need to maintain registerial integrity and naturalness in the Urdu target text. A translator may shift or omit an LB if its literal translation would violate Urdu's stylistic or syntactic norms, especially considering the constraints of news writing (coherence, flow, and attribution).

Integration: Shifts of Lexical Bundles Across Languages

The unique contribution of this framework is the intersection of these two fields: analyzing translation shifts specifically through the lens of lexical bundle frequency, function, and structure, bridging a gap in the study of non-Western language pairs.

- **Urdu-Specific Analysis:** Since Urdu is a language with a different writing system (Nastaliq script) and fundamentally different syntactic structure (often SOV or an inflected form of it, unlike English's SVO), the theoretical expectation is for significant structural transformation of LBs. The

framework predicts that many 4-word English LBs, which often reflect analytic phrase structures, will be rendered as more synthetic or agglutinative structures in Urdu.

The comparative analysis of frequency, structural, and functional shifts, guided by the established taxonomies, will empirically map the relationship between linguistic necessity (Urdu grammar) and translator choice (operational norms) in the context of professional journalistic translation.

Research Design

This study employed a qualitative, corpus-driven methodology to investigate the translation of lexical bundles (LBs) from English to Urdu in Pakistani newspaper discourse. The research process was structured into five key stages: corpus compilation, lexical bundle extraction, functional and structural categorization, comparative analysis, and translation strategy identification.

Corpus Compilation

A parallel corpus comprising 25 English–Urdu newspaper article pairs was compiled for this study. The English source texts were selected from leading Pakistani English-language newspapers, including Express media and The News International, while their corresponding Urdu translations were sourced from verified translation counterparts published by the same media houses (e.g. Daily Express, and Daily Jung). Articles were selected from comparable genres, primarily political reports, editorials, and opinion pieces, to ensure thematic and stylistic consistency.

Procedure

The research procedure involved the following steps:

Lexical Bundle Extraction

Lexical bundles were extracted using Sketch Engine, a state-of-the-art corpus analysis platform. The English corpus was uploaded to Sketch Engine and processed using the N-gram and Word Sketch tools. The study focused on 4-word lexical bundles that occurred at least 10 times per million words, following the frequency threshold proposed by Biber et al. (1999). Non-lexical sequences, such as fragmented phrases or overlapping bundles, were filtered out through manual inspection to ensure the relevance and integrity of the extracted data.

Functional and Structural Categorization

Each extracted lexical bundle was categorized along two analytical dimensions:

- **Functional Categorization**

Bundles were classified into three primary discourse functions based on Hyland's (2008) framework:

- Referential bundles (e.g., expressions of time, quantity, or place)
- Stance bundles (e.g., expressions of certainty, attitude, or evaluation)
- Discourse-organizing bundles (e.g., topic introduction, elaboration, or transition)

- **Structural Categorization**

Bundles were also analyzed for their grammatical structure using the taxonomy proposed by Biber et al. (2004). Structural types included:

- Noun phrase + of (e.g., “the end of the”)
- Prepositional phrase fragments (e.g., “at the same time”)
- Verb phrase fragments (e.g., “is going to be”)
- Dependent clause fragments (e.g., “as soon as the”)

The extracted bundles, along with their frequency counts and contextual examples, were exported from Sketch Engine and systematically entered into Microsoft Excel. This spreadsheet served as a central workspace for organizing, filtering, and annotating the data. Functional and structural labels were assigned manually within Excel, and patterns were identified through close reading and comparison across entries.

Comparative Analysis of Source and Target Texts

Each English lexical bundle was traced in the Urdu translations to identify its equivalent or transformation. This comparative analysis focused on three dimensions:

- Frequency shifts: Changes in the occurrence of bundles across the English and Urdu subcorpora
- Structural shifts: Modifications in grammatical configuration or syntactic realization
- Functional shifts: Alterations in the discourse role or pragmatic function of the bundle

Bundles that lacked direct equivalents in the Urdu texts were also documented and analyzed to understand the translator’s handling of such instances.

Translation Strategy Analysis

To address the final research question, the study examined the translation strategies employed in rendering lexical bundles. Each instance was coded according to one of the following categories:

- Literal translation: Direct structural and semantic correspondence between source and target
- Paraphrased: Reworded in Urdu while preserving the original function
- Omitted: Bundle not rendered in the target text
- Replaced: Substituted with a different expression serving a similar or altered function

Patterns in strategy use were analyzed in relation to both the functional and structural types of bundles, providing insight into the translator’s choices and the influence of linguistic and stylistic norms.

Analysis

This section presents a qualitative analysis of the lexical bundles (LBs) identified in the English–Urdu newspaper corpus, focusing on how these multi-word units function within discourse and how they are transformed in translation. Moving beyond frequency counts and categorical classifications, the analysis interprets the structural and functional behavior of LBs, examining how they contribute to coherence, stance, and textual organization in English source texts and how these roles are preserved, altered, or omitted in Urdu target texts. The discussion is organized around three analytical dimensions: functional and structural profiling, comparative translation shifts, and translation strategies (each aligned with the study’s

research objectives). Through close textual examination and illustrative examples, this section aims to uncover the linguistic and discursive dynamics that shape the translation of LBs across languages and journalistic genres.

Functional and Structural Profile of English Lexical Bundles

The qualitative analysis of the English source texts (STs) revealed distinct patterns in how lexical bundles (LBs) are employed across political, editorial, and opinion genres. This analysis moves beyond raw frequency to interpret the functional and structural roles of the most prominent bundles, providing the necessary foundation for understanding subsequent translation shifts.

Functional Categories and Their Discourse Role

The corpus demonstrated a clear functional distribution among LBs, categorized as Referential, Stance, and Discourse-Organizing.

Referential Bundles were overwhelmingly the most frequent, functioning to refer to entities, people, events, or processes within the text. They are crucial for establishing the coherence and thematic focus of the political narrative. A common pattern involves bundles used to anchor information, such as "the issue of the" or "the former prime minister". For example, a journalist might write, "The debate centered on the issue of the national debt, which remains a significant political obstacle." This bundle acts as a nominal anchor, clearly setting the topic for the subsequent clause. Other examples like "a number of people" serve as quantified nominal referents, contributing to the specificity and descriptive richness required in journalistic reporting.

Stance Bundles play a critical role in expressing the journalist's or writer's attitude, evaluation, or certainty toward a proposition, thereby shaping the tone and argumentative flow of the piece. A key finding was the high frequency of the bundle "it is important to". This LB is typically deployed to introduce a necessary action or an essential point of consideration, explicitly guiding the reader toward the writer's perspective or evaluation. For instance, in an editorial, the writer might state, "it is important to note that the official figures do not account for hidden unemployment." Here, the LB signals the author's strong belief in the relevance of the following point, adding an evaluative layer of certainty. Other prominent Stance LBs included "I would like to" and "we have to make", which serve to directly express personal opinion or obligation, respectively, particularly prevalent in opinion pieces.

Discourse-Organizing Bundles contribute to the narrative flow by explicitly signaling the relationship between clauses and turns in the argument. While less frequent than the other two types, they are essential for structuring complex journalistic texts. The most common examples were formulaic sequencing devices, such as "on the other hand" and "as well as the". For instance, a writer uses the former to signal a shift to a contrasting viewpoint: "The ruling party claimed victory; on the other hand, the opposition described the outcome as a travesty of democracy." These LBs function as connective tissue, ensuring logical progression and textual cohesion.

Analysis of Structural Configurations

The structural analysis revealed that the majority of English LBs conform to a limited set of syntactic patterns, highly reflective of the phrasal nature of journalistic prose.

The most prominent structural configuration was Noun Phrase + of, which constitutes the core of many Referential bundles. Examples like "the end of the" and "the government of the" illustrate a strong

preference for complex nominal structures, where the "of" phrase modifies the preceding noun, enabling compact, information-dense phrasing typical of news reporting. This preference for nominal elaboration often poses specific challenges for translation into Urdu, a language that predominantly employs postpositions and a different nominal modification structure, often necessitating a structural re-casting.

Another frequent structure was the Prepositional Phrase Fragment, such as "in the middle of" and "at the hands of". These bundles function either referentially (specifying location or agent) or as parts of larger verbal expressions. Their fixed prepositional nature contributes to the idiomaticity and conciseness of the English ST.

Crucially, Dependent Clause Fragments were also highly prevalent, playing a dual role, often serving as both Discourse-Organizing and Stance markers. A notable example is "as soon as the". This bundle signals a temporal relationship and typically occupies a prominent initial syntactic position in the English clause: "As soon as the new legislation is passed, we will see immediate changes." This fronting of the dependent clause fragment is a characteristic feature of formal English journalistic style, which facilitates clarity by setting the context before the main action. The strong reliance on this particular clausal fragment structure is expected to be a primary source of translation shifts, as Urdu's Subject-Object-Verb (SOV) typology and preference for final-position verbs and postpositions often render a literal transfer syntactically awkward or impossible, forcing the translator to either break the bundle or integrate it into a different clausal arrangement.

Comparative Analysis of Translation Shifts

By comparing the extracted LBs in the source language (ST) with their equivalents in the target language (TT), the study analyzed the crucial shifts, in frequency, structure, and function, that occurred during the process of English-to-Urdu journalistic translation. This analysis directly addresses the second and third research objectives.

Frequency and Realization Shifts in Urdu Translations

A significant finding was the qualitative shift in the realization and frequency of LBs in the Urdu subcorpus compared to the English STs. While LBs are a pervasive feature of high-frequency register in English journalism, the frequency of directly corresponding bundled expressions was noticeably lower in Urdu. This can be attributed to inherent differences in journalistic conventions and the structural constraints of the Urdu language.

The most common frequency shift involved instances where a single, high-frequency English LB was realized by multiple, non-bundled expressions in Urdu. For example, the English Referential bundle "the government of the" might be rendered in Urdu not as a fixed three-word sequence but as a more expanded, or less integrated, phrase like "سے جانب کی حکومت" (hukumat ki janb se - literally: 'from the side of the government') or simply "کا حکومت" (hukumat ka - 'the government's'). This demonstrates a shift in conciseness. The English LB, acting as a single informational unit, is often translated into a syntactically less integrated string in Urdu, breaking the bundling pattern. Conversely, certain English LBs were sometimes omitted or replaced by a single Urdu word, demonstrating a shift in explicitness. A Stance bundle like "it is important to" might be implicitly conveyed through the use of strong adjectival choices or emphatic particle placement in the Urdu sentence, rather than an explicit multi-word expression of stance, suggesting a preference for less overt stance markers in Urdu journalistic style.

Structural and Syntactic Shifts

The analysis of structural shifts revealed a strong tendency for the English grammatical configuration to be modified in the Urdu translation to conform to Urdu's native syntax.

A core observation concerned the structural transformation of Prepositional Phrase Fragments. For instance, the English LB "in the middle of" (a Prepositional Phrase fragment) is rarely rendered as a direct word-for-word equivalent because Urdu utilizes postpositions (following the noun) rather than prepositions (preceding the noun). Consequently, the English structure is often rendered as a nominal or adverbial phrase using postpositions, such as "میں درمیان کے" (ke darmiyān me - 'in the middle of'), which fundamentally alters the arrangement of the elements to fit the Urdu syntactic mould.

More significantly, LBs with specific clausal structures were frequently broken down into less integrated structures. An English Dependent Clause Fragment like "as soon as the" (e.g., As soon as the bill was passed...) typically needs to be restructured. Urdu's SOV (Subject-Object-Verb) typology often mandates that the verb or main action be placed at the end of the clause. Thus, the LB's function is often absorbed into a different type of dependent or adverbial construction, potentially requiring a complete nominalization of the preceding clause or using a different conditional/temporal particle far from the noun it modifies. This transformation ensures the resulting Urdu sentence is structurally well-formed, even at the cost of sacrificing the original English clausal integration.

Example:

- English ST (Dependent Clause Fragment):
"as soon as the new finance minister took office..."
- Urdu TT (Restructured Nominal/Adverbial Phrase):
جیسے ہی (jaise hi - as soon as) + 'سنبھالا دفتر نے خزانہ وزیر نئے' (nae wazir-e-hazana ne daftar sambhala - the new finance minister took office).

The structure is shifted from a compact fragment to a full, integrated Urdu conditional clause with the verb placed finally

Functional and Pragmatic Shifts

Functional shifts were observed when the pragmatic role of the LB was altered in the translation, sometimes leading to a measurable change in the communicative impact on the Urdu reader.

A notable functional shift involved Stance bundles. For instance, an English Stance bundle expressing a high degree of certainty or evaluation, such as "we have to make", might be translated into an Urdu expression that conveys a weaker sense of obligation or a more indirect suggestion, reflecting a possible cultural or stylistic preference in the Urdu register for less confrontational or assertive declarations of opinion. This shift impacts the reader's perception, potentially diluting the original author's evaluative strength.

Furthermore, the analysis revealed several cases of Omission of primarily Discourse-Organizing bundles (e.g., "on the other hand"). When these bundles were omitted, it created a functional gap in the Urdu TT. While the semantic content of the argument was preserved, the explicit signposting of the rhetorical relationship (e.g., contrast or addition) was lost. This suggests that Urdu journalistic style may rely more heavily on contextual inference and less on formulaic, explicit discourse markers for maintaining argumentative structure, leading to a text that is less overtly organized than its English counterpart.

Analysis of Translation Strategies Employed

This final analytical subsection directly addresses the fourth research objective by examining the application of the identified translation strategies, Literal Translation, Paraphrase, Replacement, and Omission, used by translators when rendering English lexical bundles (LBs) into Urdu target texts (TTs). The strategic choice of these methods dictates the degree of formal and functional correspondence between the source and target texts.

Prevalence and Interpretation of Literal Translation

Literal translation was identified as a prevalent strategy, primarily employed when the structural and semantic components of the English LB allowed for a near one-to-one transfer that remained natural and idiomatic in Urdu. This strategy was most commonly applied to specific types of Referential bundles and those with highly transparent Structural types, such as certain Noun Phrase + of configurations where the nominal heads were easily transferable.

For example, the LB "the president of the" often underwent successful literal transfer, resulting in an equivalent Urdu phrase such as "کا مملکت صدر" (Sadr-e-Mumlikat ka - 'the president of the state's'). In such cases, the strategy maintains the syntactic and semantic integrity across the language pair, offering conciseness similar to the original. However, the application of Literal Translation was limited by the fundamental typological differences between English (SVO) and Urdu (SOV/Postpositional). A literal rendering was rejected when it produced syntactically awkward or unidiomatic Urdu, forcing the translator to adopt a more adaptive strategy.

Strategic Use of Paraphrase and Replacement

The strategies of Paraphrase and Replacement were observed as crucial tools used by translators to maintain the functional equivalence of the bundle when structural correspondence was impossible or unnatural in Urdu.

Replacement typically involved substituting the English LB with an Urdu expression that was structurally different but functionally identical or near-identical, often trading a multi-word sequence for a single morpheme or a common idiomatic phrase. For instance, the English Verb Phrase fragment "is going to be" (a future marker or modal expression) was frequently Replaced by a simpler, single future-tense verbal form or a specific modal auxiliary in Urdu, such as the suffix '-گا' (ga) or the verb 'گا ہو' (ho ga - 'will be'). This strategic choice achieves the same Stance or Referential function (e.g., expressing future action) while adhering to Urdu's highly inflected verbal system.

Paraphrase, on the other hand, involved a structural expansion or re-casting of the LB into a longer, more elaborate Urdu phrase. This was frequently necessary for complex Dependent Clause Fragments like "as soon as the", which cannot be literally transferred. The translator might choose a longer, more explicit temporal construction to ensure clarity and syntactic flow, sacrificing conciseness for naturalness. This focus on functional equivalence ensures that the pragmatic role of the bundle—whether expressing stance, linking discourse, or referring to an entity—is successfully conveyed to the target reader, even if the form is entirely different.

Omission and Its Contextual Justification

The strategy of Omission was a deliberate, strategic choice, most commonly observed with Discourse-Organizing bundles and certain Stance bundles that the translator deemed either redundant or stylistically extraneous within the conventions of Urdu journalistic discourse.

For example, a low-frequency Discourse-Organizing bundle that merely introduces a point, such as "the fact that we", might be omitted entirely. The contextual flow of the Urdu paragraph is often considered sufficient to bridge the argumentative gap without the need for an explicit multi-word marker. Similarly, certain Stance bundles expressing high certainty or emphasis might be omitted if the Urdu journalistic register prefers a more objective or less explicitly marked tone.

This decision to omit is often interpreted as a strategic choice influenced by the stylistic or register norms of the target language newspaper. By omitting formulaic LBs, the translator adheres to an assumed Urdu preference for a leaner, less repetitive style, which may rely more on implicit cohesion than the overt linguistic signposting typical of the English source genre.

Discussion of Findings and Implications

This section synthesizes the key qualitative findings from the analysis of functional, structural, and translational shifts of lexical bundles (LBs). It moves from merely presenting the data to interpreting its significance in light of the study's problem statement, understanding the translation of formulaic language between English and Urdu journalistic discourse, and its implications for translation theory and practice.

Synthesis of Core Findings

The analysis established that English LBs constitute a highly salient feature of the source texts, primarily serving Referential functions (e.g., "the end of the") and relying heavily on Noun Phrase + of and Dependent Clause Fragment structures. The subsequent comparative analysis (Sections 4.2 and 4.3) demonstrated that the translation process is characterized not by simple equivalence, but by systematic adaptation driven by linguistic typology. The most significant finding is the pervasive de-bundling phenomenon in the Urdu target texts (TTs). High-frequency English LBs often experience frequency and realization shifts, being translated into less formulaic, non-bundled expressions in Urdu. This suggests that the same functional load carried by a fixed phrase in English is distributed across different, less standardized linguistic units in Urdu, reflecting a potential difference in the institutionalized use of formulaic language in the respective journalistic traditions.

The Structural and Syntactic Shifts, specifically the transformation of prepositional and clausal fragments, underscore the dominant influence of Urdu's SOV word order and postpositional structure. The translator's choice of Replacement and Paraphrase strategies (e.g., transforming "as soon as the" into a final-verb conditional clause) demonstrates a clear prioritization of target language naturalness over source language structural fidelity. This is a crucial practical implication: successful journalistic translation from English to Urdu often requires a re-conceptualization of phrasal integration to ensure the TT adheres to native syntactic expectations.

Implications for Translation Theory and Practice

These findings have direct implications for both descriptive translation studies and translator training.

- Formulaic Language and Typology

The observed translation strategies provide empirical evidence for the constraint imposed by linguistic typology on the transfer of formulaic language. The study supports the notion that the translation of LBs is not merely a lexical task but a syntactic and stylistic restructuring challenge. The frequent recourse to Omission (especially of Discourse-Organizing bundles) and the strategic use of Paraphrase suggest that LBs fall into the category of "system-bound items," whose realization is tied to the unique grammatical and stylistic preferences of the source language. This confirms that Literal Translation is only viable for a limited, structurally inert subset of LBs.

- Register and Stylistic Norms

The differences in LB frequency and realization between the ST and TT point towards diverging register and stylistic norms in English and Urdu journalism. The relative decrease in LBs in Urdu TTs, coupled with the functional gap created by the omission of certain Stance and Discourse-Organizing bundles, suggests that Urdu journalism may favor a less explicitly signposted or less formulaically hedged style compared to its English counterpart. The translator acts as an intermediary who mediates between these stylistic expectations, strategically diluting or dispersing the formulaic density of the English text to align with the conventions of the Urdu newspaper register. Understanding these institutional norms is as critical for effective translation as mastering the grammatical rules.

- Pedagogical Relevance

For translation pedagogy, the study highlights the need to train translators not merely in translating individual words, but in recognizing and strategically managing lexical bundles as functional and structural units. Translators must be taught to identify the pragmatic function of an LB (e.g., Is it expressing strong stance? Is it organizing discourse?) and prioritize the maintenance of that function, even when it requires a complete Replacement or Structural Shift (e.g., using a native Urdu idiom or a completely different clausal structure) rather than attempting a structurally equivalent but unnatural rendering. The examples of Paraphrase and Replacement, such as rendering the Verb Phrase Fragment "is going to be" with a single future-tense verb form, serve as excellent instructional cases for teaching functional equivalence across typological boundaries.

CONCLUSION

This qualitative analysis confirms that the translation of English Lexical Bundles (LBs) into Urdu journalistic texts is a process of systematic adaptation and structural restructuring, rather than direct equivalence. The study found that while LBs are highly frequent and structurally fixed in English (e.g., using structures like Noun Phrase + of), they are often subject to de-bundling in Urdu. This is primarily because translators prioritize conforming to Urdu's grammatical structure, specifically its Subject-Object-Verb (SOV) typology and postpositional system, over maintaining the source text's phrasal structure. Strategies like Replacement and Paraphrase are therefore dominant, ensuring functional equivalence (maintaining the meaning and pragmatic role) even at the cost of structural divergence. Furthermore, the selective Omission of certain LBs indicates that translation choices are deeply influenced by the distinct stylistic and register norms of Urdu journalism, which appear to favor a less explicitly signposted or less formulaic style than its English counterpart.

FUTURE RECOMMENDATIONS

Based on these findings, the following recommendations are suggested for future research and pedagogical development:

- Quantification of De-Bundling: Future studies should employ quantitative methods to accurately measure the extent and statistical significance of the "de-bundling" phenomenon across different genres (e.g., hard news vs. editorials) to provide a fuller profile of formulaic use in Urdu.
- Focus on Translator Interviews: Conduct interviews with professional English-Urdu journalistic translators to gain direct insight into the decision-making process concerning LBs. This qualitative data could clarify why certain LBs are omitted or paraphrased, revealing the subjective factors and institutional norms that guide strategic choices.
- Cross-Typological LB Framework: Develop a specific pedagogical framework for teaching LB translation between typologically distant languages like English and Urdu. This framework should focus on recognizing the function (e.g., Stance, Discourse-Organizing) of the LB and training translators in the most effective non-literal replacement structures native to Urdu to ensure naturalness and idiomatic flow.
- Target Language Corpus Analysis: Analyze a parallel corpus of Urdu-to-English translations to see if a reverse process occurs (i.e., whether non-bundled Urdu expressions are systematically bundled into fixed phrases when translated into English), offering a more comprehensive understanding of formulaic use across this language pair.

REFERENCES

- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81-92. <http://dx.doi.org/10.1016/j.esp.2011.08.004>
- Ahmad, S., & Khan, M. A. (2024). A comparative linguistic analysis of complex phraseology in regional languages. *Journal of Comparative Linguistics*, 15(1), 45-62.
- Al-Zoubi, S., Al-Shehab, W., & Qawasmi, H. (2020). Translation shifts of formulaic expressions in English-Arabic news texts: A corpus-based study. *International Journal of Translation*, 34(2), 112-135.
- Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and Technology: In Honour of John Sinclair* (pp. 233–250). John Benjamins Publishing Company.
- Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in Contrast: Papers from a Symposium on Text-Based Cross-Linguistic Studies* (pp. 177–189). Lund University Press.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for specific purposes*, 26(3), 263-286. <http://dx.doi.org/10.1016/j.esp.2006.08.003>
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405. <http://dx.doi.org/10.1093/applin/25.3.371>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.

- Chen, Y., & Baker, M. (2010). Towards a new model of formulaic language in translation: A case study of journalistic lexical bundles in English-Chinese parallel texts. *International Journal of Corpus Linguistics*, 15(2), 221–248.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397-423.
<http://dx.doi.org/10.1016/j.esp.2003.12.001>
- Crosthwaite, P., & Franco, V. (2018). Lexical bundles in parallel corpora: Structural and functional shifts in English-Chinese translation. *Target*, 30(3), 384-411.
- Dechert, S., & Dechert, G. (2012). Translation of multi-word units: Structural preservation in an English-German parallel corpus. *Bochum Studies in English*, 3(2), 55-78.
- Huang, Z. (2023). Translation shifts of lexical bundles in an English–Arabic parallel corpus: A corpus-driven study. *Frontiers in Psychology*, 14, 1113426.
- Huang, C.-H. (2023). Translation directionality and the complexity of multi-word units: A corpus-based investigation. *Perspectives: Studies in Translation Theory and Practice*, 31(6), 947-965.
- Hyland, K. (2008a). As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes*, 27, 4-21. <http://dx.doi.org/10.1016/j.esp.2007.06.001>
- Kenny, D. (2001). *Lexis and creativity in translation: A corpus-based study*. St. Jerome Publishing.
- Pérez-Vidal, C., & Abarca, P. (2022). Pragmatic function shifts of phraseology in translation: A parallel corpus analysis. *Translation Studies Quarterly*, 18(4), 512-535.
- Malik, S., Azher, M., & Ashraf, S. (2023). Translation universals in Urdu and English journalistic texts: A corpus-based study. *Journal of Translation Studies*, 5(2), 154–178.
- Pan, F., & Liu, P. (2019). Lexical bundles in English newspaper editorials: A corpus-based study of form and function. *Corpus Linguistics and Linguistic Theory*, 15(1), 93–120.
- Qureshi, B., Ali, K., & Zahid, S. (2022). Exploring the phraseology of Urdu: Implications for English language learners. *Pakistan Journal of Applied Linguistics*, 25(1), 101-120.
- Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching* (Vol. 65). Amsterdam: John Benjamins Publishing Company.
<http://dx.doi.org/10.1075/scl.65>
- Toury, G. (1995). *Descriptive translation studies and beyond*. John Benjamins Publishing Company.
- Zahid, S., & Hussain, M. (2023). A review of corpus methodologies for extracting lexical bundles: Considerations for Urdu and other less-resourced languages. *Journal of Language and Text Research*, 13(2), 88-105.