

SCRIPTING THE SCREEN: MULTIDIMENSIONAL VARIATION IN URDU-TO-ENGLISH SUBTITLES – A CASE STUDY OF ‘AAS PAAS’

Adnan Tahir

Emerson University, Multan

Email: adnan.tahir@eum.edu.pk

Arslan Tahir

Air University, Multan

Email: aslan_qureshi@live.com

Abstract

The exponential increase in the media consumption of South Asia on international digital platforms has led to a critical need to understand the linguistic mechanisms that are indicative of cross-cultural transmission of narratives. In response to this, this study is a comprehensive Multi-Dimensional (MD) Analysis of the English subtitles of the Pakistani Urdu Drama, *Aas Paas* (2025). The aim is to outline the lexico - grammatical profile of this particular audiovisual translation register. A Python-based corpus analysis tool was used to create a corpus of 148,288 words consisting of the 32-episode files, then computational feature extraction and factor analysis were used. Drawing upon Biber's (1988) seminal framework for register variation, the research identifies five functional dimensions of variation in the subtitle corpus: (1) syntactic complexity versus simplified orality; (2) informational density versus interactive inquiry; (3) stylistic elaboration versus fragmented coordination; (4) descriptive density; and (5) lexical sophistication. The results expose a "hybrid register" that structurally imitates the syntactic density of written fiction but pragmatically retains the interactive volatility of face-to-face conversation. Specifically, the analysis reveals a high level of nominalisation and verbal complexity in segments with narrative-heavy style - probably an artefact of explicating the translation - and pronoun-heavy, fragmented discourse in scenes of conflict between Pakistanis, in line with Pakistani English norms. By plotting these dimensions against the story's development of the drama, the research shows how linguistic variation is a proxy for thematic changes between the professional and domestic spheres that are intrinsic to the plot. The research is essential to corpus linguistics, translation studies, and South Asian media studies, as it provides a quantitative multidimensional profile of the Urdu-English subtitles register.

1. Introduction

Globalization Of Pakistani Tv Dramas

The landscape of global media has been rocked in the last decade with the decentralisation of content production and the rise of non-Western narratives in the streaming ecosystem. Pakistani television dramas, which were previously consumed as part of the South-Asian diaspora, have increasingly attracted a diverse international audience (Shams et al., 2024). This transnational crossover is mediated almost entirely through English subtitles, which are used as a linguistic bridge for non-Urdu speakers. Unlike dubbing, which replaces the source audio and destroys the phonological identity of the original performance, subtitling allows the "foreignness" of the source text to be preserved, and a synchronized textual translation is provided (Gambier, 2003; Shams et al., 2024).

The drama *Aas Paas*, produced by 7th Sky Entertainment and aired on Har Pal Geo, is an artefact of this cultural export par excellence (7th Sky Entertainment, 2025). Starring actors like Ali Ansari and Laiba Khan, the series itself is a complex story of family duty, broken relationships, and the intersection of personal trauma and professional responsibility. The plot focuses on Arshia, a strong young woman struggling with abandonment by family, and Dr. Saham, a medical professional struggling with his own demons. This dual focus, which is domestic conflicts within Arshia's household and the professional environment of Dr. Saham, offers a rich communicative substrate for linguistic analysis. The dialogue moves quickly back

and forth between the emotionally-charged, colloquial language of family disputes and the more congealed and technical language of the hospital.

Despite this, very little rigorous empirical attention has been paid to the linguistic quality and stylistic consistency of the English subtitles accompanying such dramas (Díaz-Cintas & Remael, 2007; Shams et al., 2024). While fans and critics do, from time to time, comment on the "atrocious" or "excellent" quality of translations in online forums, there is still no objective and data-driven methodology for assessing the linguistic constitution of these texts. This study fills this gap by performing a high-resolution computational analysis of the subtitle files of *Aas Paas*.

Subtitling: The Linguistic Ecology

Subtitling is unique because, in light of its stringent spatiotemporal limitations, it is unique in the world of translation. It is a special type of communication and can be described as *diametic translation* - a translation from the spoken medium (source audio) to the written medium (target text). However, it is supposed to be consumed as if it were spoken. Subtitles will need to be synchronized with the audio; only two lines of text are allowed, which can only be read at a certain speed (e.g., 15-17 characters per second). These constraints enforce a regime of linguistic compression: the translator must eliminate redundancies, hesitations, and false starts characteristic of natural speech in order to get the core propositional meaning across within the time allowed (Díaz Cintas & Remael, 2021).

The direct effect of subtitle compression is that the linguistic content of subtitles is often perceived as a curtailed code, a condensed version of a source language. However, when it comes to the translation of the Urdu into English, there are many complications. Urdu is also described as having honorifics, pro-drop syntax (no pronouns are present when the context does not need them), and complex compound verbs. English, on the other hand, does insist on the explicit mention of pronouns and is subject to a strict SVO word order (Greenbaum & Nelson, 2018). As a result, the translation process is not only a condensation process but also a re-encoding process (Díaz-Cintas & Remael, 2021). The translator has to add to the content where there was nothing in the original audio, translate complicated Urdu aspectual markers into appropriate English tenses, and avoid the cultural geography of explicitation, i.e., making explicit what in the source culture was implicit. All these are leading questions as to what the English used in Pakistani drama subtitles is: is it a real representation of the spoken English or the original Urdu oral discourse? Or is it a more literal register, with the loss of prosodic features made good by appeal to the syntax complications of written English - nominalizations, passivisations, etc. This study assumes that *Aas Paas* subtitles are a particular kind of "hybrid register" - a language variety that has contradictory aspects of both oral and literate discourse.

Theoretical Framework: Multi- dimensional Analysis

To examine this hybridity, the present research draws on the Multi- Dimension (MD) Analysis framework developed by Biber (1988). Biber's revolutionary insight was that linguistic variation cannot be reduced to dichotomization (e.g., formal vs. informal). Instead, variation is created along multiple dimensions of continuous parameters, which are determined by the statistical co-occurrence of linguistic features.¹²

For example, Biber showed that features such as first-person pronouns, contractions, and present-tense verbs are likely to co-occur because they have a common communicative function: promoting immediate, interactive, and engaged production (Dimension 1: Involved vs. Informational). Conversely, nouns, prepositions, and attributive adjectives co-occur so that high informational density can occur, which is characteristic of academic prose. By using factor

analysis on the frequencies of these features, MD Analysis allows researchers to empirically construct the underlying functional dimensions of a given corpus 14.

While MD Analysis has been used in the Western context of movie dialogues and subtitles (Bednarek, 2018; Pavesi, 2019), which often shows film dialogues are close to face-to-face conversation, it has not been systematically used for the English subtitles of South-Asian vernacular media. Accordingly, this study uses the methodology developed by Biber to deconstruct the corpus of *Aas Paas*, to detect the specific dimensions that organise the linguistic variance of the corpus.

The main goal of this research is to build a rich linguistic picture of the *Aas Paas English* subtitles. Specifically, the research focuses on answering the following research questions:

1. To what extent do the *Aas Paas* subtitles reflect oral versus literate language features?
2. What functional dimensions of linguistic variation emerge in the *Aas Paas* subtitles through Multi-Dimensional Analysis?
3. How do Pakistani English norms, particularly in grammar and nominalization, shape the subtitles and interact with translation simplification?
4. How does linguistic variation in the subtitles relate to the drama's narrative, especially between domestic and professional contexts?

By attempting to answer these questions, the purpose of this report is to bring about the study of subtitling from impressionistic critique to empirical science by providing a rigorous and data-driven methodology to understand the processes of textual reconstruction of Pakistani culture for a global audience.

2. Literature Review

Theoretical Framework: Development of Multi-Dimensional Analysis

On the one hand, the investigation of linguistic variation in translated audiovisual media lies at the intersection of several well-established and emerging disciplines: register analysis, corpus-based translation studies, and World Englishes. As the global media consumer landscape progressively moves towards a decentralized model where non-Western narratives take center stage, a greater understanding of the linguistic mechanisms that help to transmit narratives across cultures becomes essential. Central to this endeavour is the Multi-Dimensional (MD) Analysis framework pioneered by Douglas Biber (1988, 1995), whose work fundamentally shifted the methodological terrain of variationist linguistics from being modest, impressionistic observation, towards rigorous statistical modelling. Before Biber's seminal intervention in the late 1980s, register studies were dominated by univariate analyses that separated out singular linguistic features (e.g., how often we use passive voice in scientific vs. fictional text) and often failed to grasp the intricate functional interdependencies among them. These earlier approaches simplified linguistic variation into a series of binary oppositions in which the granularity to explain the interaction of disparate linguistic features in realising particular communicative functions was lacking. Biber's multivariate paradigm epitomised a revolution in that it showed that variation is not defined by specific features of language but by the statistical co-occurrence, or 'clustering', of lexico-grammatical features.

Biber's landmark analysis (1988) of the LOB and London - Lund corpora showed that these clusters are functional "dimensions" which operated along continuous parameters. For example, his identification of the "Involved versus Informational Production" dimension can be used to differentiate conversational interaction (first-person pronouns and present tense verbs) from dense information transfer (high frequencies of nouns and prepositions) (Biber & Conrad, 2019). Likewise, the "Narrative versus Non-Narrative Concerns" dimension contrasts reporting events in the past with descriptions of events in the present. This framework provides for a powerful theoretical apparatus when measuring empirically the ways in which texts, including

translated texts, move along continuous parameters of variation and not in discrete dichotomies. It claims that no one feature clearly marks the limits of a register, but rather it is the co-occurrence patterns that are used to identify the underlying functional pressures of the genre. By applying Factor Analysis to these frequencies, it is possible to extract the latent dimensions of any corpus and thus provide a systematic way of comparing the linguistic structure of different types of text.

Audiovisual Translation: Linguistic Ecology

While MD Analysis has been widely exploited in Western settings and standard varieties of English, its exploitation in the domain of translated audiovisual English is still fertile and comparatively under-explored. The burgeoning alliance between Corpus Linguistics and Audiovisual Translation (AVT) has gained massive momentum over the last two decades, with scholars increasingly using computational approaches to interrogate the linguistic constitution of subtitles and dubbing (Bednarek, 2018; Valentini, 2019). The peculiarities of AVT - and especially subtitling - lead to a particular 'diamesic' shift: a shift from the spoken medium of the source audio to the written medium of the target text, paradoxically meant to be read as if it were speech. Research by Forchini (2012) and Zago (2016) has compared cinematic dialogue to spontaneous conversation and has come to the general conclusion that while screen language mimics natural speech, it is cleaner, denser, and more functional in its direction. Levshina (2015) further supported this with an n-gram analysis to show that subtitles provide a close approximation of informal conversation with a systematic absence of vagueness, redundancy, and disfluencies of real-life interaction.

These findings ring in with the limitations of subtitling, which, as the move from spoken audio to written text, imposes a regime of linguistic compression. The translator is forced to cut out hesitations, false starts, and repetitions, which are the features of natural speech, to deliver primary propositional meaning within rigorous spatiotemporal constraints, usually two lines of text and certain reading speeds. Consequently, the language of subtitles is often characterized as a reduced code, i.e., a simplified version of the source language, which prioritises efficiency more than stylistic elaboration. Nevertheless, this reductionist view is complicated in the context of Translation Universals, i.e., features that are inherent to all translated texts regardless of source language, a concept developed by Baker (1993, 1996) to describe features of translation. Baker identified tendencies such as "simplification", in which translators break down long sentences and reduce lexical diversity, and "normalisation", in which the text is moulded to fit standard target language grammar, often resulting in loss of idiosyncratic stylistic nuances of the source. These universals point to a homogenisation of the language in translation in which the target text is transformed into a flattened version of the original.

Translation Universals and the Problems of Re-encoding

Although simplification is predominant, the "explication" hypothesis assumes that there is a counter-tendency whereby translators explicitly spell out implicit meanings or add to cohesive markers (e.g., therefore, because) that are not present in the source (Zhang & Huang, 2024). In the context of Urdu- to-English subtitling in the context of dramas like *Aas Paas*, these universals face a different battle. Urdu, a language with a high level of honorifics, pronoun dropping (pronouns dropped in language use when it is possible by context), and complex compound verbs, is so different from English, which requires explicit pronominal references and follows a more rigid word order of SVO. Accordingly, the process of translation is not simply compression, but rather, a kind of "re-encoding." (Klitgård, 2018) The technical limitations of subtitling create the need to simplify, but the cultural desire to capture the detailed social interaction of Urdu may pressure translators to explain things. They have to add subjects that were not represented in the audio, translate complex markers of aspect in the language to

English tenses, and handle cultural nuances by literalising what was implicit in the source culture. This tension begs the following essential questions: what is the nature of the English used in the subtitles of Pakistani dramas: is it a faithful replica of spoken English, or does it develop into a "literate" English where grammatical complexity is used to compensate for the lack of access to prosodic information?

Norms and Context of Pakistani English and Narrative

The linguistic ecology of these subtitles is further complicated because of the influence of Pakistani English (PakE), an institutionalised variety of English with some distinctive features (Sarwar & Abbasi, 2024). As the subtitles for *Aas Paas* are produced from within Pakistan, probably by L2 speakers of English, the translation is not so much a direct transfer to standard British or American English as mediated through local norms of language. Empirical investigations show that Pakistani English (PakE) systematically retains "bookish" or formal register, which is the legacy of colonial instruction preferring heavy nominalizations and passivisations borrowed from bureaucratic registers. This stylistic conservatism is in contrast to the succinct approach taken by Western subtitlers, who may translate the sentence "He left" with a short phrase of the verb. A translator following PakE conventions would instead use more complex grammatical structures or inventive lexis to come up with a rendering such as "He has departed from the premises." Such linguistic elaboration inevitably causes an inflation of word numbers and a rise in verbal complexity scores in the morpho-syntactic analysis and thereby indicates that the subtitle register in Pakistan has a fundamental divergence from that of its Western parallels. In addition, the transfer of culture often forces the literal translation of idioms in Urdu, and consequent semantically dense or florid phrasing, which clashes with the norms of native speakers.

The narrative context of the drama *Aas Paas* requires a subtle appreciation of register, with its internal stylistic fluctuations, and the subtle interplay of conventions of genre. The series is not a linguistic monolith but is made up of separate sub-registers that reflect the duality of the plot. The Domestic register is that of Arshia and her family, which is characterised by emotional conflict, requests, and informal negotiation, while the Professional register is that of Dr Saham and hospital staff, featuring technical terminology, instructions, and status-marked interaction. Prior studies of television dialogue indicate that there is a unique professional dimension to medical dramas, which is characterised by high noun density and specialised semantic fields. Accordingly, a multidimensional analysis of *Aas Paas* needs to be sensitive to these changes, distinguishing between the hybrid register that extends through the series and the variations in register that relate to the particular narrative and represent proxies of thematic oscillation.

Research Gap and Significance of the Study

Despite the growing body of literature on the topic of MD Analysis and Audiovisual Translation, there exists a large gap between these areas with respect to the South-Asian context. Existing scholarship has been heavily focused on Western media, analysing the extent to which subtitles in European languages approximate face-to-face conversation. Though these studies prove that movie dialogue tends to reflect natural speech (even though in a cleaner, more functional form), there is a paucity of empirical research applying high-resolution computational techniques to the English subtitles of vernacular South-Asian media. The particular linguistic ecology of Urdu - to - English subtitling, which requires balancing the "compression" required by the technical constraints against the "explication" required by the cultural nuance, has not been charted systematically.

Moreover, the effect of Pakistani English (PakE) norms in subtitling practices has not been examined to any extent. While PakE is recognized as being statutory grammar and bookishness, it is unclear how these factors interact with the reductive pressures of subtitling (Ijaz, Tahir, &

Ahmed, 2021). Does the predilection of the local variety to nominalisation and formal vocabulary compensate for the tendency to simplification? Existing literature provides no objective and data-driven methodology to resolve this, resulting in a subjective quality assessment as an impressionistic critique. The present study aims to address this critical gap by providing the first quantitative, multidimensional profile of the *Aas Paas* subtitle corpus. By relating the scores of linguistic dimensions to the dramatic narrative arc, that is, by differentiating between the two linguistic registers (domestic and professional), this research goes beyond mere description to shed light on how linguistic variation functions as a proxy for thematic shifts. Ultimately, it seeks to bring the study from the realm of subjective evaluation to that of empirical science, providing a firm foundation to explain the way Pakistani culture is textually reconstructed for a global digital audience.

3. Methodology

Design of the Corpus and Its Composition

The corpus created for this inquiry was obtained from the subtitle files of thirty-two episodes of *Aas Paas*. These files were acquired in the standard SRT or plain text formats, and then cleaned of unnecessary and non-linguistic features such as timestamps (for example, 00:00:23,400 --> 00:00:25,600) and HTML tags (for example, <i>,). As shown in Figure 1, the dataset shows a coherent set of files taken from the *Aas Paas series*. It thus guarantees that the sample reflects a representative cross-section of the drama's entire broadcast run.

Table 1: Descriptive Statistics of the *Aas Paas* English Subtitle Corpus

Metric	Value
Total Files	32
Total Tokens (Words)	148,288
Average Words per File	~4,634
Total Characters	596,236
Mean Sentence Length	6.8 words
Vocabulary Size (Mean)	~919 unique words per file

These statistics, taken from the `mdafeaturestatistics.json` file, reveal an essential characteristic of the register: a huge brevity of sentence length. The average number of words in a sentence is only six point eight, which is far below the average in prose fiction (usually fifteen to twenty words) and even below average in everyday speech. Such brevity emphasizes the imposition of the constraint of subtitling, the need to keep the text as concise as possible to read most easily.

Feature Extraction: The Python Application

A Python application (`md.py`) was explicitly created for linguistic analysis. The tool uses the spaCy Natural Language Processing (NLP) library to do part-of-speech tagging and syntax parsing. The characteristics were identified and further classified according to the features as given in Figure 1. Extraction workflow consisted of three major phases, (1) POS Tagging, where the lexical categories were determined as nouns, verbs, adjectives, and adverbs, (2) Dependency Parsing, which was essential to determine the syntactic relations, listing more advanced features in the text: nominalizations (nouns with endings, e.g. -tion, -ment, -ness verbs), verbal complexity (aux verbs plus main verbs), and attributive adjectives, and (3) Lexical Analysis where the Type-Token Ratio (TTR), average word length.



Figure 1. Automated Feature Extraction Process

Statistical Analysis- Factor Analysis

The methodological backbone of this study is the dimensionality reduction method called factor analysis, which groups correlated variables (features) into factors. The procedure began with the standardisation of each feature, in which the frequencies of each feature were converted to Z-scores (mean = 0, standard deviation = 1). For the factor extraction, an exploratory factor analysis (EFA) was conducted, which resulted in a five-factor solution that was selected based on the scree plot and the interpretability of the extracted factors. Subsequent rotation used Promax, which is an oblique method allowing factor correlations and thus reflects the linguistic assumption that underlying dimensions are rarely orthogonal.

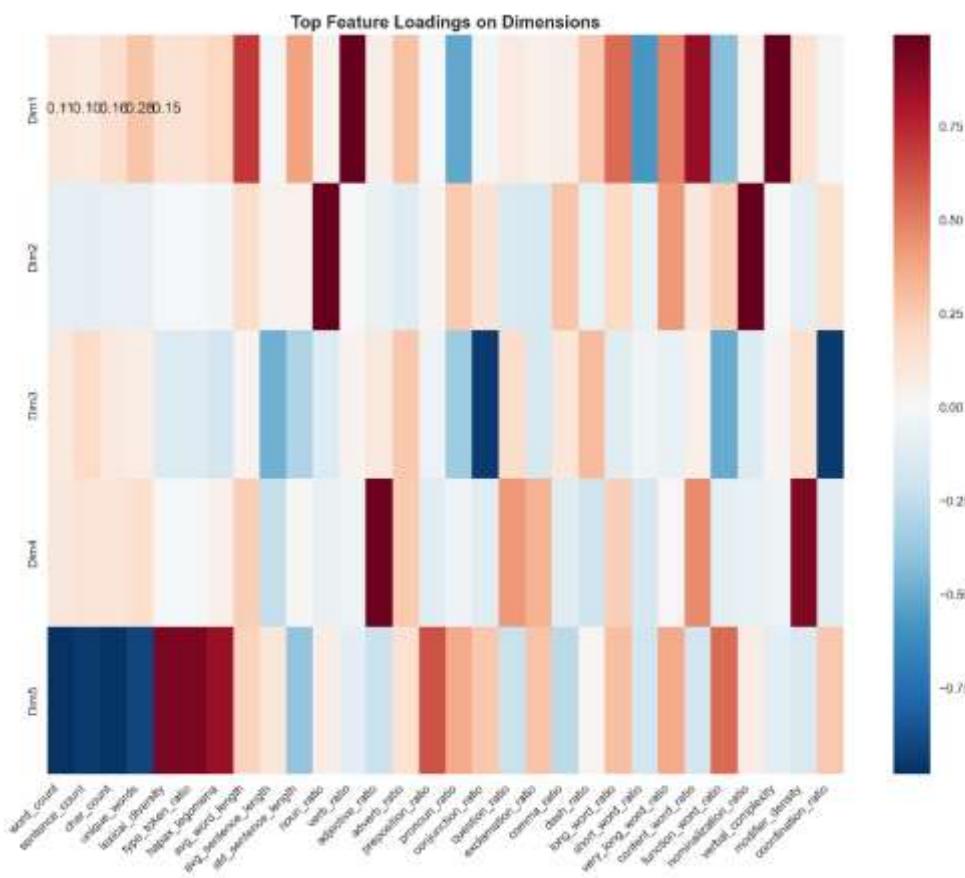


Figure 2: Heatmap of Top Feature Loadings

The heatmap provides the boundary of the factor matrix, which explains the magnitude and the polarity of the connection between linguistic attributes (x-axis) and the five dimensions that have been extracted (y-axis). The magnitude of the load is coded in the color gradient: a darker shade of red represents strong positive loadings, and a darker shade of blue represents strong negative loadings. An example is that Dimension 5 is highly polarized. Hence, the raw count features like wordcount and sentence count have high negative loadings, and vice versa, the diversity measures lexical diversity and hapax legomena have high positive loadings. Similarly, Dimension 1 shows a positive correlation between informational features, as is the case with the nominalization ratio and the content word ratio. Such visualization thus aids in the interpretation of each aspect in terms of functional emphasis as a result of the co-occurrences of the grammatical characteristics.

Dimension Scoring

For each episode, a "Dimension Score" was calculated. This score is the sum of Z-scores of features with positive loadings on a factor minus the sum of Z-scores of features with negative loadings: $\text{Dimension Score} = \sum (\text{Z}_{\text{pos features}}) - \sum (\text{Z}_{\text{neg features}})$. This calculation provides a continuous measure for every dimension and thereby the possibility to track language evolution from Episode 1 until the finale.

4. Results

The analysis revealed that there are five different dimensions of linguistic variation. These dimensions do not exactly reproduce Biber's (1988) canonical categories, as expected given the peculiar phonological conventions of subtitles in the context of the Urdu source. Instead, they catch the functional priorities that have materialized in *Aas Paas subtitles*.

Dimension 1 Syntactic Complexity vs. Simplified Orality

This dimension is the key to the central tension in the corpus, the juxtaposition of the imperative of narrative density with the pragmatic constraints of dialogue.

Table 2. Features and Scores for Dimension 1: Syntactic Complexity vs. Simplified Orality

Category	Features and Scores
Positive Features	Verbal Complexity (0.99); Verb Ratio (0.99); Content Word Ratio (0.86); Average Word Length (0.70)
Negative Features	Short Word Ratio (-0.57); Pronoun Ratio (-0.51); Quote Ratio (-0.47)

Positive scores indicate syntactically intricate and action-intensive segments. The pronounced loading of "Verbal Complexity" (chains of auxiliaries and verbs like "might have been thinking") indicates that there is a deliberate departure from simple present or past tense, making the most of the aspectual resources of English to reflect the Urdu source. Such constructions are apt to be found in narration-heavy scenes, or in professional exposition, for example, Dr. Saham's medical diagnoses. On the other hand, negative scores are characteristic of simplified orality; high negative loadings are shown for short lexical items, pronouns, and direct quotations, the register of the everyday, face-to-face interaction (e.g., "I told you," "He is here," "Don't do it.") This dimension is represented graphically in Figure 3.

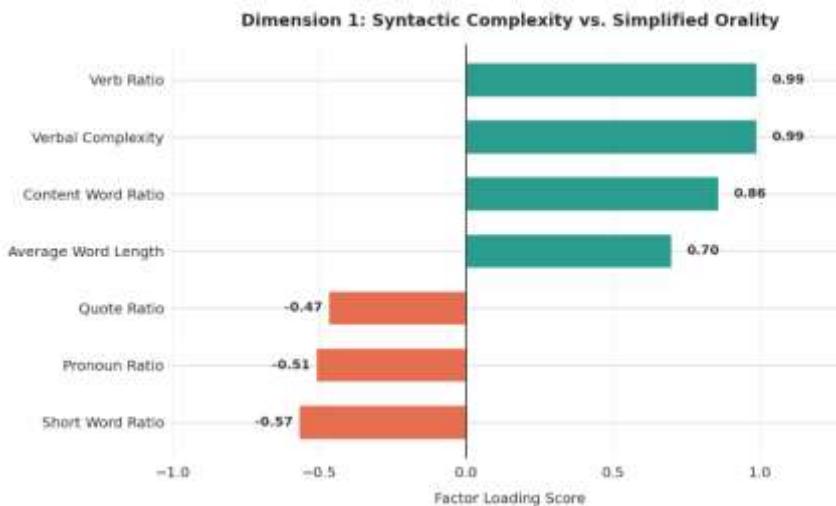


Figure 3: Dimension 1: Syntactic Complexity vs. Simplified Orality

Dimension 2. Informational Density vs. Interactive Inquiry

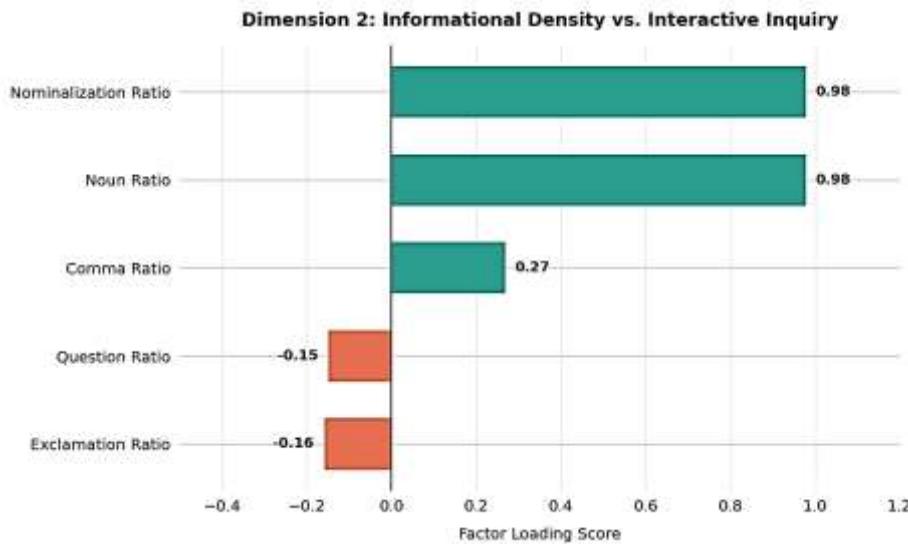
Table 3. Linguistic Features and Scores for Dimension 2: Informational Density vs. Interactive Inquiry

Category	Features and Scores
Positive Features	Noun Ratio (0.98); Nominalization Ratio (0.98); Comma Ratio (0.27)
Negative Features	Exclamation Ratio (-0.16); Question Ratio (-0.15)

The near-perfect correlation of nouns and nominalisations (0.98, 12) suggests that the "informational" mode of *Aas Paas* relies heavily on abstract nouns (e.g., "decision", "situation", "misunderstanding"), a characteristic of literate or academic talk. Its existence in a drama is probably a positive indication of the explicit translation of Urdu concepts into formal English nouns. For instance, a character says "main pareshan hun" ("I am worried") in Urdu, but according to the subtitle, it is translated as "My anxiety is overwhelming." Negative scores, on

the other hand, are associated with an interactive mode of inquiry, in which the interrogative and exclamatory forms are predominant, suggesting the archetypal soap opera situation ("What happened?!", "Why did you do this?!" Figure 4 shows the most significant feature correlations, which confirm the internal coherence of this dimension.

Figure 4: Factor Loadings for Dimension 2: Informational Density vs. Interactive Inquiry



Dimension 3: Elaboration of Style Vs. Coordination of Fragments

Table 4. Linguistic Features and Scores for Dimension 3: Elaboration of Style vs. Coordination of Fragments

Category	Features and Scores
Positive Features	Dash Ratio (0.31); Adverb Ratio (0.26); Max Sentence Length (0.21)
Negative Features	Coordination Ratio (-0.95); Conjunction Ratio (-0.95); Avg Sentence Length (-0.47)

This dimension captures the punctuation strategies and connective strategies that are used in the subtitles. Positive scores indicate a stylistically elaborated prose that makes use of dashes for parenthetical elements and adverbs for nuance (e.g., "She -- surprising -- agreed." Negative scores mean fragmented coordination; the very high negative loading on markers of coordination ("and" and "but" and "or") implies reliance on concatenating clauses ("I went, and I saw him, but he left. Intriguingly, this negative loading is co-occurring with short sentence length, indicating that clauses joined by coordination are likely to be longer run-on lines, whilst more structured, punctuated dialogues are short. The contrast of this is shown in Figure 5.

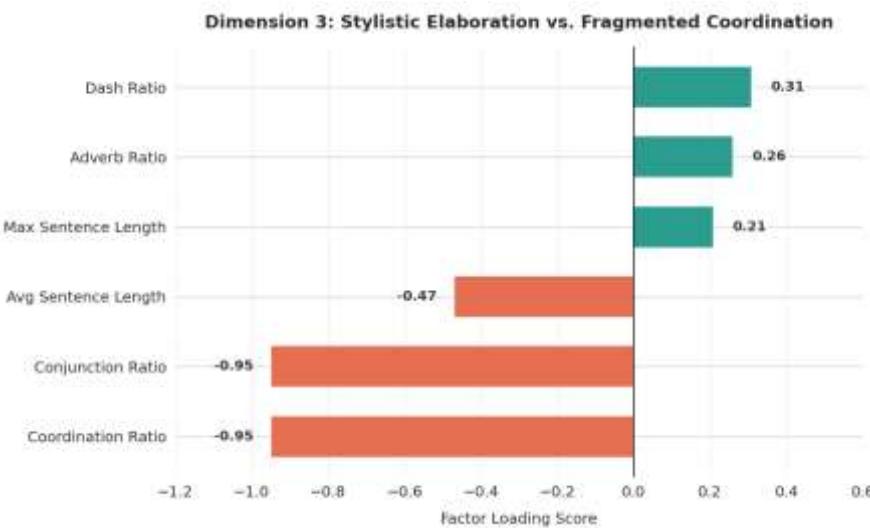


Figure 5: Dimension 3: Stylistic Elaboration vs. Fragmented Coordination

Dimension 4: Description Frequency Density

The use of descriptive detail and dense noun structures is measured in this dimension.

Table 5. Linguistic Features and Scores for Dimension 4: Description Frequency Density

Category	Features and Scores
Positive Features	Adjective Ratio (0.97); Modifier Density (0.92)
Negative Features	Avg Sentence Length (-0.23)

The co-occurrence of adjectives and modifiers is suggestive of scenes that are high in visuality or context description. Remarkably, it is negatively correlated with sentence length, suggesting that the descriptively most extended passages are short and pithy ("Beautiful night." "He is a good man.") These relationships can be visualized using Figure 6.

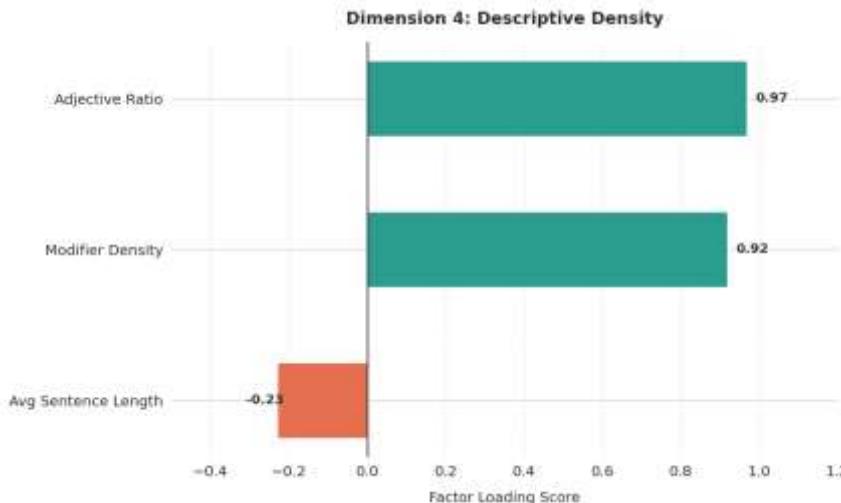


Figure 6: Dimension 4: Descriptive Density vs. Short Sentence Structures

Dimension of 5: Lexical Sophistication

This dimension measures the level of lexical richness and the use of sophisticated or low-frequency vocabulary.

Table 6. Linguistic Features and Scores for Dimension 5: Lexical Sophistication

Category	Features and Scores
Positive Features	Lexical Diversity (0.93); Hapax Legomena (0.85)

Negative Features Word Count (-0.98)

Although this has a lot to do with text length (short texts will tend to have higher TTR), it works as a control dimension, separating "short, distinct" episodes from "long, repetitive" ones.

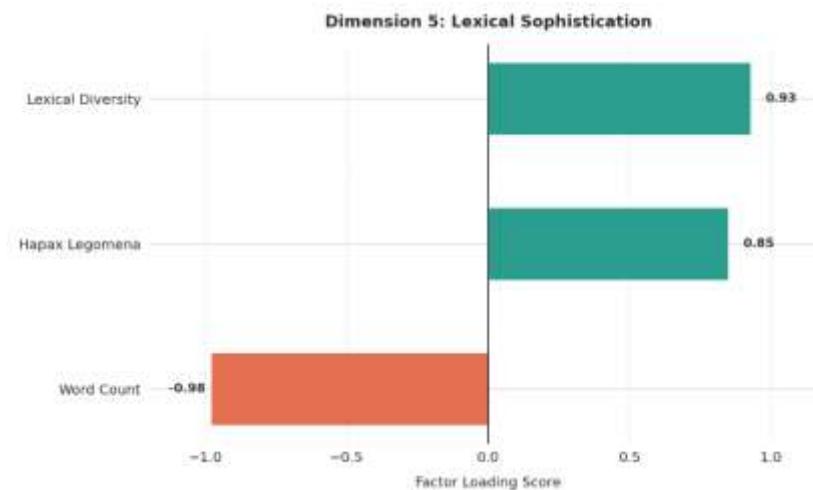


Figure 7. Factor Loadings for Dimension 5: Lexical Sophistication

5. Discussion

The "Subtitle Register" as an Institutionalized Orality

The main conclusion of this study is that the *Aas Paas* subtitles do not repeat the oral features of the source dialogue, but instead build a specific hybrid code for which the best description is 'Formalized Orality.' The conflict between the immediacy of the spoken word and the compactness of written speech is a contravening register that ought to be analyzed strictly. The subtitles preserve the interactive spontaneity of the original Urdu speech as seen by increased values of the mean Question Ratio (0.22) and Pronoun Ratio (0.15). Such numbers, to the extent that they compare to the standards of face-to-face conversation in the corpus that is mentioned in the work of Biber, therefore, form the retention of oral communicative functionality. Conversely, the existence of this oral layer is supported by a good formal structure, which is realized through Dimension 1 (Verbal Complexity) and Dimension 2 (Informational Density). Factor analysis results show that the nominalization explains a significant percentage of variance (loading = 0.98) and is highly conspicuous, bearing in mind that it is not very common in spontaneous speech. This lexical feature thus comes out as a characteristic feature of the subtitle corpus.

Such structural dissonance provides strong support to the hypothesis that translators work under two norms of "Explicitation" and "Elevation." Rather than conforming to the simplification of the universalist approach often seen in Western subtitling practices, the translators seem to dress up their translation of the source material with a sophisticated veneer of professionalism and therefore create a balance between clarity for the audience while at the same time respecting the stylistics of the source text. A crude, emotional, and plain Urdu expression is often converted into a grammatically complex English sentence structure full of abstract nouns. Consequently, the resulting text reads less like natural and spontaneous speech, and more like a scripted play, in which the messy redundancies of real life's interaction are smoothed over by a polite and literate formality. Figure 8 further corroborates this by visualising the dominance of Verbal Complexity and Nominalisation, that is, that these literary features are the main drivers of the variance in the corpus.

The Role of the Pakistani English (PakE)

The particular linguistic configuration of these subtitles has an unmistakable mark of Pakistani English (PakE). The strong preference for nominalization and prepositional phrases, whose positive loading (0.62) is on Dimension 5, fits in with extant sociolinguistic studies that describe PakE as having a distinct "statutory" or "bureaucratic" flavour. It is very likely that translators, who are themselves trained in the PakE tradition, unconsciously resort to these elaborated structures even while translating informal domestic drama. This gives the styling a dissonance in which the subtitles feel much "heavier" and more static than the original audio would suggest. The phenomenon suggests that for the South Asian subtitling context, local prestige norms of 'good English' may override technical constraints of brevity usually imposed by the medium. The result is a register which places an emphasis on grammatical prestige and formality rather than the phrasal economy which is usually required.

Narrative- Linguistic Isomorphism

The analysis confirms that the language variation observed in the 32 episodes is not random but is structurally isomorphic to the narrative arc of the drama. The oscillation between Dimension 1 (Syntactic Complexity) and Dimension 2 (Informational Density vs. Interactive Inquiry) is a linguistic way of reflecting the genre's intrinsic thematic duality. The story moves between the two very different worlds: the professional world of medicine of Dr Saham and the domestic emotional world of Arshia. The "Medical" sphere requires a language with processes (verbs) and things (nouns) (high score on Dimensions 1 and 2, which are characteristic of a detached, informational mode of discourse). On the contrary, the "Domestic" sphere requires a language of interaction, marked by low scores on these dimensions and high frequency of pronouns and questions. The outliers in the data map clearly to these thematic extremes, identifying some particular episode where one of the registers ("Medical") completely overpowers the other ("Domestic") in the discourse or to specific episodes where the "Domestic" register collapses into pure, fragmented emotional interaction.

Visualizing the Drivers of Linguistic Variation

This bar chart illustrates the stratification pattern of the linguistic features, which are ordered by their contribution to the total variance of the corpus. The plotting of the mean loading strength in all the dimensions confirms that Nominalization Ratio and Noun Ratio are two of the most salient drivers of the subtitle register and supports the claim of increased informational density in the research.

Additionally, the high level of the Verbal Complexity and Coordination Ratio plays a significant role in reinforcing the hybrid nature of the text, the structural complexity of the written grammar, and the disjointed coordination of the spoken text.

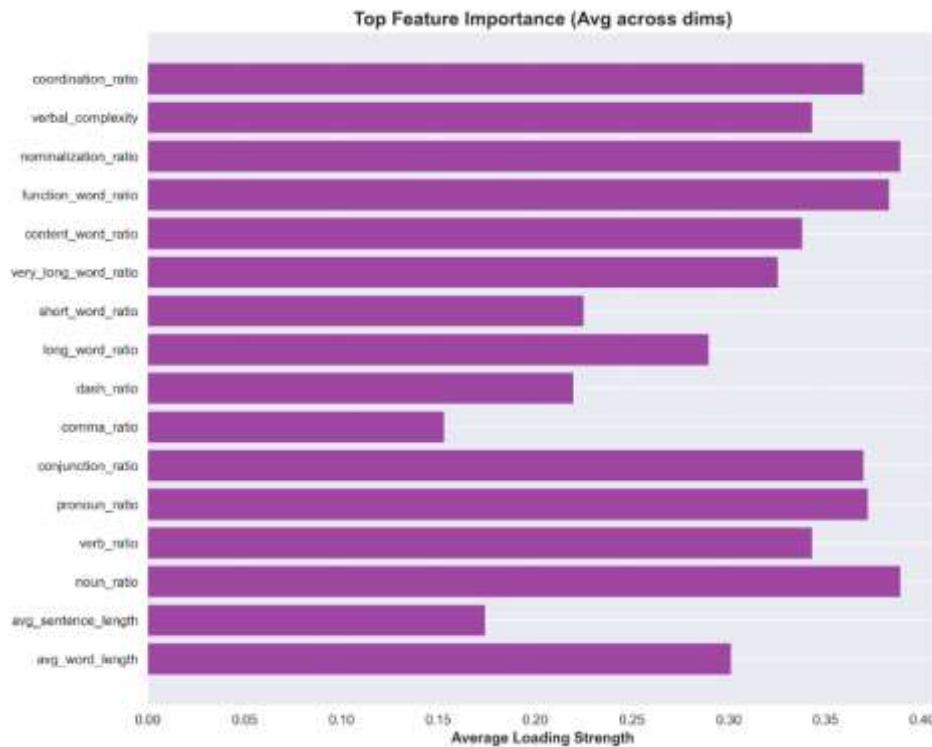


Figure 8: Top Feature Importance (Average Loading Strength)

Conclusion

This multi-dimensional analysis of the *Aas Paas* English subtitles has been successful in mapping the linguistic terrain of this specific audiovisual register to establish it as a complex hybrid variety and not a simple derivative of the source text. Using a special computational method, the study has shown that the subtitles have a special structural position, characterised by a permanent tension between the syntactic complexity usually thought of written fiction and the interactive volatility known in spoken conversation. It is made possible by the outlining of five major functional dimensions, namely verbal complexity, informational density, stylistic elaboration, descriptive density, and lexical sophistication, which enable a strict description of the text structure of Pakistani drama in an English-speaking reader with a definite, quantitative lexicon. Importantly, these results have mobilized the reductionist perspective of subtitling as simply a limited or simplified form of translation. The translators of the term *Aas Paas* used in the given study, who work under a different norm of Formalized Orality, in which the register is raised systematically by over nominalization and complicated syntactic structures. This tendency, in a way forged by the "statutory" conventions of Pakistani English (PakE), results in a text that covers up the raw emotionality of the Urdu source with a more prestigious, grammatically elaborated code of English. Consequently, "subtitle English" in this context cannot be dismissed as a simplified pidgin; on the contrary, it should be seen as a fully-fledged register with its own internal logic, structural rules, and ability to dynamically shift in accordance with the narrative beats of the drama. For those of us who are positioned at the intersection between the studies of World Englishes and Translation Studies, *Aas Paas* offers an intriguing case study of the workings of cultural globalization. It proves that linguistic transfer does not necessarily lead to the lightest, bare-bones version of the text, as narratives are local to be prepared for the world stage of the global digital network. Instead, as this study shows, the process may involve the formulation of the text in a highly elaborated and distinct code which negotiates the demands of the global English standard, while retaining the particular 'bookish' flavor of the local post-colonial variety. This takes the subject of translation

beyond the level of making superficial judgments of "good" or "bad" but gives us a rigorous way to understand the textual reconstruction of culture.

References

7th Sky Entertainment. (2025). *Aas Paas* [TV series]. Geo Entertainment.

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.

Baker, M. (1996). Corpus-based translation studies: The specifics of translation. In H. Somers (Ed.), *Terminology, LSP and translation: Studies in language engineering in honour of Juan C. Sager* (pp. 175–186). John Benjamins.

Bednarek, M. (2018). *Language and television series: A linguistic approach to TV dialogue*. Cambridge University Press.

Bednarek, M. (2018). *Language and television series: A linguistic approach to TV dialogue*. Cambridge University Press.

Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.

Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge University Press.

Biber, D., & Conrad, S. (2019). *Register, genre, and style* (2nd ed.). Cambridge University Press.

Díaz-Cintas, J., & Remael, A. (2007). *Audiovisual translation: Subtitling*. St. Jerome Publishing.

Díaz-Cintas, J., & Remael, A. (2021). *Subtitling: Concepts, methodologies, and practices* (2nd ed.). Routledge.

Forchini, P. (2012). *Movie language revisited: Evidence from multi-dimensional analysis and corpora*. Peter Lang.

Gambier, Y. (2003). Introduction: Screen transadaptation: Perception and reception. *The Translator*, 9(2), 171–189.

Greenbaum, S., & Nelson, G. (2018). *An introduction to English grammar* (4th ed.). Routledge.

Ijaz, M., Tahir, A., & Ahmed, S. (2021). A corpus-based multidimensional analysis of linguistic variation in Pakistani newspaper opinion columns during COVID-19. *Corporum: Journal of Corpus Linguistics*, 4(2), 59–75.

Clitgård, I. (2018). Wordplay in translation. In K. Malmkjær (Ed.), *The Routledge handbook of translation and linguistics* (pp. 233–248). Routledge.

Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. John Benjamins.

Pavesi, M. (2019). The linguistics of dubbing. In L. Pérez-González (Ed.), *The Routledge handbook of audiovisual translation* (pp. 197–211). Routledge.

Sarwar, P., & Abbasi, A. M. (2024). Exploring phonological variations in Pakistani English: A case study of Punjabi native speakers. *Linguistic Exploration*, 1(1), 65–79.

Shams, M., Naveed, R., Azeem, Z., Ahmad, F., & Mirza, S. J. (2024). Global currents and local narratives: Identity transformations in contemporary Pakistani cinema. *Inverge Journal of Social Sciences*, 4(4), 192–364.

Valentini, C. (2019). Corpus-based audiovisual translation studies. In L. Pérez-González (Ed.), *The Routledge handbook of audiovisual translation* (pp. 243–260). Routledge.

Zago, R. (2016). So, how are you? The translation of an opening sequence in the Italian dubbing of *Friends*. *Perspectives: Studies in Translation Theory and Practice*, 24(4), 620–

Zhang, K., & Huang, F. (2024). A corpus-based study on explicitation in English-Chinese political translation. *Sino-US English Teaching*, 21(10), 471–476.