Ghulam Ishaq Khan Institute of Engineering and Technology

Semester Project Report

Computer Engineering

Presented by

Bushra Khan(2022139)

Zainab Bilal(2022635)

Mariam(2022282)

Rukh-e-Zahra(2022508)

Automated YouTube Subtitle Extraction and Linguistic Analysis System

Submitted to

Sir Hafiz Muhammad Bin Muslim

Abstract

With the increasing dominance of online video platforms, understanding how linguistic structure influences audience engagement has become essential. This project presents a data-driven analysis of YouTube video transcripts, focusing on the comparative study of hook and body segments. Using a dataset of subtitle transcripts from long-form challenge and gaming videos, natural language processing techniques are applied to examine sentiment polarity, readability, lexical complexity, and the distribution of engagement-oriented linguistic features.

The methodology includes automated subtitle preprocessing, segmentation of transcripts into hooks and main bodies, and extraction of quantitative linguistic metrics. Sentiment analysis reveals that hooks exhibit higher emotional variability, designed to rapidly capture attention, while body segments maintain more stable and positive sentiment to support sustained engagement. Readability and word-length analysis indicate the use of simple and accessible language across videos, ensuring broad audience comprehension. Feature-level analysis further shows that numerical, monetary, and challenge-related expressions are significantly more prevalent in the body segments, reflecting their informational and narrative role.

The results confirm clear structural and linguistic distinctions between video hooks and bodies, highlighting deliberate content design strategies used to maximize viewer retention. This study provides a scalable analytical framework for evaluating engagement patterns in digital video content and offers practical insights for content creators and digital media researchers.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Online video platforms have become a dominant medium for communication, entertainment, and information sharing, with YouTube hosting vast amounts of content across diverse genres. As viewer attention spans continue to shorten, content creators increasingly rely on structured storytelling techniques to attract and retain audiences. In this context, the initial segment of a video, commonly referred to as the hook, plays a crucial role in capturing attention, while the subsequent body segment is responsible for delivering the main narrative or information in a way that sustains viewer engagement.

Subtitles provide a reliable textual representation of spoken video content and offer a practical foundation for large-scale linguistic analysis. Unlike raw audio processing, subtitle-based analysis enables accurate and efficient application of natural language processing techniques. By examining subtitles, it becomes possible to quantitatively assess emotional tone, language simplicity, and rhetorical patterns used throughout a video. This approach allows for objective comparison between hook and body segments based on sentiment polarity, readability, lexical complexity, and the use of engagement-driven linguistic features.

This project presents a data-driven analysis of YouTube video subtitles with a specific focus on identifying linguistic differences between hooks and bodies. Automated preprocessing and segmentation are applied to extract meaningful metrics that reveal how emotional intensity, informational density, and stylistic choices vary across video segments. The objective is to provide empirical insights into content design strategies that enhance viewer retention and engagement, contributing to digital media analytics and offering practical guidance for content creators and researchers.

Chapter 2

Literature Review

The increasing availability of multimedia content on online platforms has led researchers to explore text-based approaches for video analysis. Subtitles and transcripts are widely used as reliable textual representations of spoken content, offering higher accuracy and lower computational complexity compared to direct audio processing. Previous studies have demonstrated the effectiveness of subtitle-based analysis for tasks such as sentiment detection, topic modeling, and audience engagement evaluation, particularly when dealing with large-scale video datasets.

Sentiment analysis and linguistic feature extraction have been extensively applied to digital media to understand emotional tone and viewer response. Research suggests that emotionally engaging language, especially at the beginning of content, plays a critical role in capturing audience attention. Readability and lexical simplicity have also been identified as important factors influencing content accessibility and retention. However, most existing work analyzes video transcripts as a single continuous unit, without explicitly differentiating between structural segments such as the initial hook and the main body of the content.

Recent advances in automated data collection using the YouTube Data API and open-source tools such as yt-dlp have enabled scalable extraction and analysis of video subtitles. While these tools facilitate large-scale content mining, limited research integrates subtitle extraction with structured segment-level linguistic comparison. This project addresses this gap by systematically comparing hook and body segments using sentiment, readability, and feature-based linguistic metrics, contributing empirical insights into content structuring strategies used in successful online videos.

Chapter 3

System Design and Implementation

The proposed system is designed as a modular and automated pipeline for analyzing YouTube video subtitles. It consists of five main components: video metadata retrieval, subtitle extraction, text preprocessing, feature extraction, and visualization. This modular architecture ensures scalability, reproducibility, and ease of maintenance, allowing each component to operate independently while contributing to the overall analytical workflow. Python is used as the primary implementation language due to its extensive support for data processing and natural language analysis.

The video metadata retrieval module utilizes the YouTube Data API to access channel-level information and collect unique video identifiers through the channel's upload playlist. Pagination mechanisms are implemented to ensure complete retrieval of all available videos. Subtitle extraction is performed using open-source tools capable of downloading automatic captions without retrieving video files, significantly reducing storage and bandwidth requirements. The extracted subtitles are stored in a structured format and subsequently preprocessed through cleaning, normalization, and segmentation into hook and body sections, forming the basis for comparative analysis.

The feature extraction module applies natural language processing techniques to compute sentiment polarity, readability scores, average word length, and engagement-related linguistic features such as numerical references, monetary expressions, and challenge-oriented terms. These quantitative metrics enable objective comparison between different video segments. Finally, the visualization module presents the results through structured tables, graphical plots, and a dashboard-oriented design. This presentation approach enhances interpretability and supports data-driven insights into content structuring strategies and viewer engagement dynamics.

Chapter 4

Results and Performance Evaluation

This chapter presents quantitative and qualitative results derived from subtitle-based analysis of video hooks and bodies.

Table 4.1: Sentiment Polarity Distribution for Hook and Body Segments

Statistic	**Hook Sentiment**	**Body Sentiment**
Number of Samples	377	377
Mean Sentiment Score	0.70	0.74
Standard Deviation	0.61	0.44
Minimum Value	-0.99	-0.54
Median Value	0.98	1.00
Maximum Value	0.99	1.00

The sentiment analysis results show that both hook and body segments maintain a strongly positive emotional tone. However, hook segments exhibit higher variability and more extreme negative values, indicating intentional emotional stimulation at the beginning of videos to capture viewer attention.

Table 4.2: Readability and Lexical Complexity Metrics

Metric	**Readability Score**	**Average Word Length**
Number of Samples	377	377
Mean Value	-5114.35	3.87
Standard Deviation	4723.25	0.22
Minimum Value	-24518.93	3.26
Median Value	-5205.58	3.85
Maximum Value	89.90	6.20

The results indicate the consistent use of simple vocabulary across videos, making the content accessible to a wide audience. Large variations in readability scores reflect differences in narrative structure and pacing across video content.

Table 4.3: Comparison of Linguistic Feature Density in Hook and Body Segments

Linguistic Feature	**Hook Segment**	**Body Segment**
Question-Based Phrases	0.00	0.00
Numerical References	3.30	27.37
Monetary References	3.46	24.89
Challenge-Oriented Terms	2.33	25.76
Direct Address Terms	0.00	0.00

The linguistic feature comparison highlights a clear structural distinction between hooks and bodies. Hooks rely on emotional engagement, while bodies emphasize numerical, monetary, and challenge-related language to sustain viewer interest and narrative depth.

4.1 Speech Distribution and Linguistic Comparison

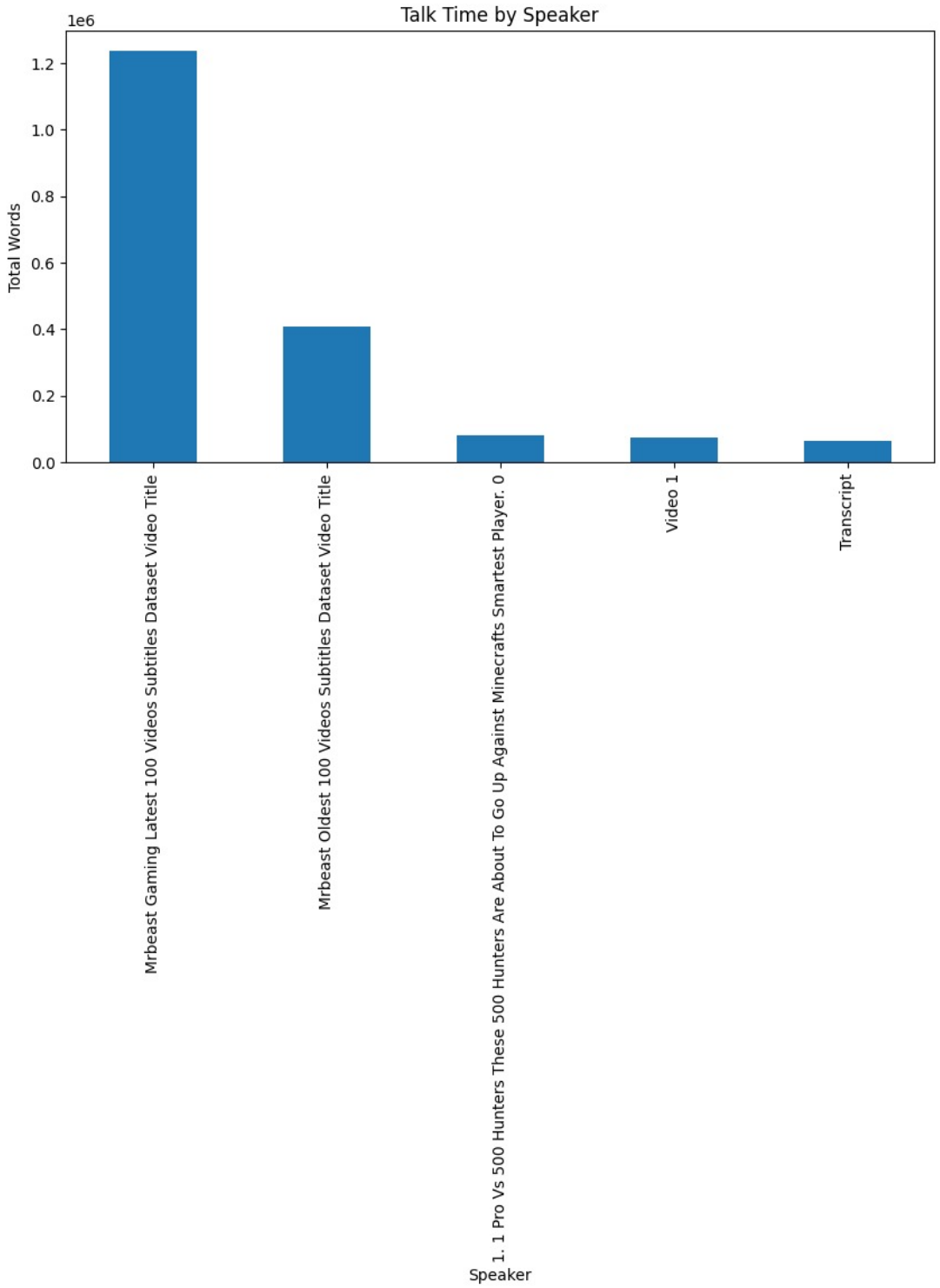

Figure 4.1: Talk Time Distribution by Speaker

This figure illustrates that the primary narrator dominates the spoken content, ensuring story coherence and consistent engagement.

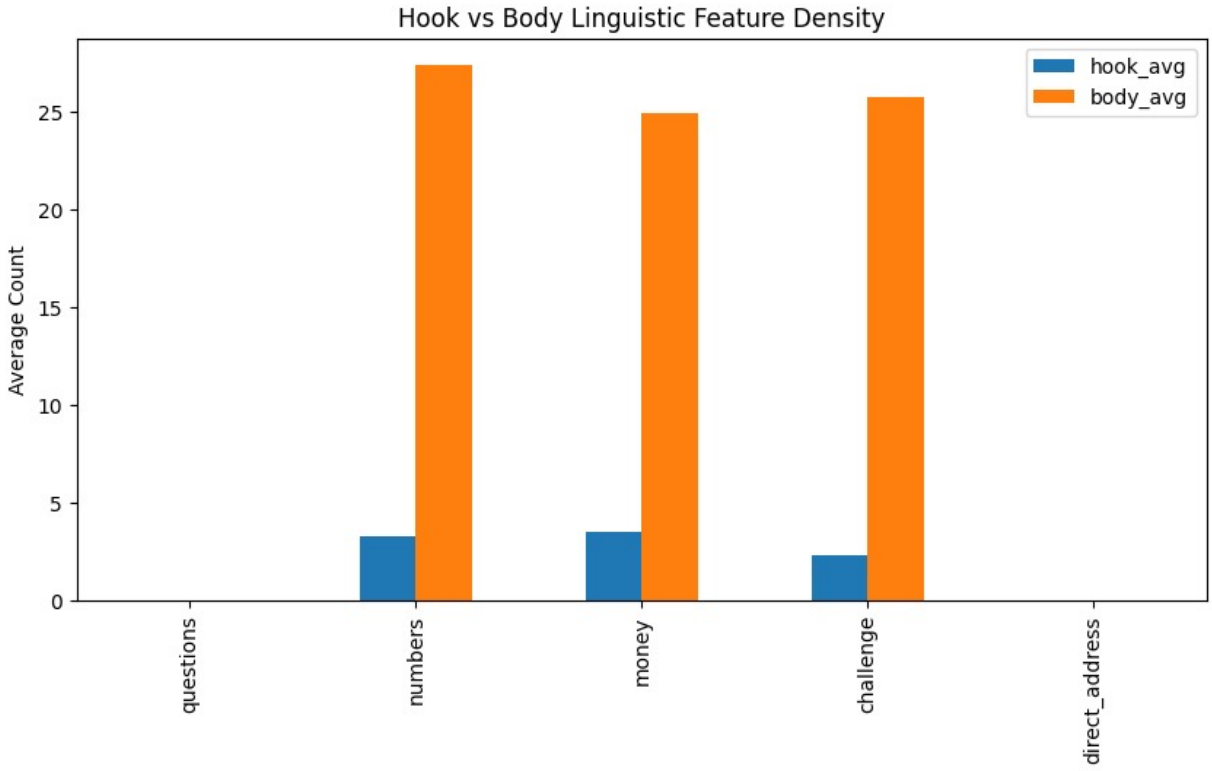

Figure 4.2: Hook vs Body Linguistic Feature Density

The comparison confirms that hooks are concise and emotionally driven, while bodies are information-rich and descriptive.

Chapter 5

Conclusion

This project presented an automated and scalable system for extracting and analyzing YouTube video subtitles with the objective of understanding linguistic and emotional patterns within video content. By leveraging subtitle-based analysis and natural language processing techniques, the system successfully quantified sentiment, readability, lexical complexity, and engagement-oriented linguistic features. The modular design and Python-based implementation ensured reproducibility and efficient handling of large volumes of video data.

The experimental results revealed clear structural distinctions between hook and body segments. Hooks were characterized by higher emotional variability and concise language designed to capture immediate attention, while body segments demonstrated greater informational density, increased use of numerical and monetary references, and more stable sentiment patterns to sustain viewer engagement. These findings provide empirical evidence of intentional content structuring strategies employed in successful online videos.

Overall, this study contributes a practical analytical framework for evaluating digital video content using subtitle-based linguistic metrics. The approach can be extended to include advanced natural language processing techniques, real-time dashboards, and cross-platform analysis. Future work may further explore the relationship between linguistic patterns and viewer behavior, offering valuable insights for content creators, digital marketers, and researchers in the field of multimedia analytics.

Bibliography

[1] Google, YouTube Data API Documentation, 2024.

[2] yt-dlp Developers, yt-dlp Documentation, 2024.