Indian Institute of Technology – ISM, Dhanbad

Department of Humanities and Social Sciences

TEXT MINING STUDY OF K-DRAMA – SQUID GAME SEASON 1: SUBTITLES

A Subtitle Translation Analysis Using Text Mining Techniques

Submitted by:

GAURISHA PANDEY

24MA0007

Master's in Digital Humanities and Social Sciences

2024-26

Submitted to:

Dr. Shamugapriya

Course: Text Mining

Date of Submission: April 2025, WINTER SEMESTER

CONTENTS

INTRODUCTION

K-dramas and K-pop culture appeared limelight during the COVID-19 pandemic, offering comfort, escapism, and emotional connection during a time of collective uncertainty. What began as localized entertainment in South Korea rapidly evolved into a global cultural phenomenon. This wave, known as *Hallyu* or the Korean Wave, has captivated audiences across various languages, continents, and age groups. Despite the content being in Korean, it was to a surprise that it gained its worldwide popularity that has been largely supported by one silent hero—subtitles. Subtitle have acted as a bridge, allowing stories deeply rooted in Korean culture to be experienced and appreciated by millions across the globe relating to it.

But subtitling is far more than simple translation. It is a delicate art that involves navigating between languages and cultures, making choices that shape how a story is received by the viewers, therefore any small mistake could lead to harvest of misconception or misunderstanding about a different culture. A good subtitle is not just accurate—it's emotionally and culturally resonant. It must preserve the rhythm of dialogue and convey tone, honor social hierarchies (especially important in Korean society). This becomes particularly challenging in Korean, as the language rich in honorifics, layered politeness, and expressions that often lack direct equivalents in English. A subtitle translator, therefore, becomes a kind of cultural mediator, tasked with preserving not just what is said, but how it is said and why it matters.

Interestingly, Korea's storytelling tradition—rooted in folklore, historical epics, and Confucian values—have always emphasized emotional depth, social bonds, and moral dilemmas. These themes remain central in today's K-dramas as well, which continue to explore complex human relationships and societal issues in deeply moving ways not only showcasing a lifestyle but also giving heartfelt moral values one should have possession of. What make this remarkable is how these traditional narrative structures, when paired with modern digital platforms and translated subtitles, manage to resonate universally. A global viewer might not understand the nuance of Korean class systems or familial terms, but through careful subtitling, they can still connect to the character's struggles, hopes, and humanity.

This project investigates how subtitle translation affects meaning through simplification and cultural adaptation, using the internationally acclaimed Korean drama *Squid Game* as a case study. The drama, with its intense emotional arcs and sharp social critique, provides a rich text for analysis. The drama has gained immense popularity worldwide because of its nerve-racking plot and intriguing storyline expressing human dilemmas and choices through children gamely.

This study focuses on two key areas: simplification of language and the translation of politeness and cultural nuances. Through this, we can better understand how translation choices impact the perception and emotional engagement of international audiences. Ultimately, this project aims to shed light on how language, culture, and storytelling interact in the digital age—and how Korea, through its dramas and subtitles, continues to connect hearts across the world.

Review of Literature

Previous studies on audio-visual translation (AVT) have elaborated on subtitling as a bounded type of translation that is subject to spatial, temporal, and linguistic constraints (Díaz-Cintas & Remael, 2007). Translation simplification has been noted to be the inclination towards using less complex and less implicit wording in translated texts (Laviosa, 2002). Other research points to subtitler writer's negotiation of levels of politeness, generally encoded deeply in the hierarchical and culture-based structure of the source language (House, 2015). In Korean, politeness is marked by honorifics, speech levels, and indirectness, all challenging to translate into English, whose politeness structure is less lexically explicit (Kim, 2018).

Further, studies have used text mining techniques such as keyword extraction, KWIC (keyword in context), and word frequency to determine patterns in translated text (Baker, 2006). These tools are particularly handy in determining systematic patterns and changes in large corpora, which allow researchers to transcend anecdotal descriptions. This study builds on such research by coupling close reading with computational methods in examining patterns in Squid Game subtitle corpus.

Research Questions

1. How does subtitle translation in *Squid Game* simplify the original Korean dialogues?
2. How are politeness and cultural nuances, particularly those expressed through honorifics and idioms, translated into English subtitles?
3. How consistent are the English subtitles in rendering Korean cultural expressions and idioms throughout the series?

Methodology

The corpus for this project includes the complete Korean and English subtitle files for all nine episodes of *Squid Game season 1*. The subtitles were collected from publicly available fan-translated and official versions. The selection criteria included:

- Completeness: All episodes included.
- Alignment: Each Korean sentence is aligned with its English counterpart.
- Relevance: The content is rich in cultural idioms and speech level variations.

The subtitle files were **pre-processed using Python**, including **tokenization, stopwords removal,** and **alignment** checks to ensure consistency between source and target texts.

For this project, I focused exclusively on *Squid Game* for three main reasons. First, it is a globally recognized Korean drama that has reached an international audience, making it highly relevant for analysing the impact of translation in cross-cultural media consumption. Second, *Squid Game* includes a rich variety of linguistic and cultural elements — such as honorifics, idioms, and politeness levels — making it an ideal microcosm for studying subtitle translation strategies.

Third, due to system limitations and the scope of this student project, expanding to multiple series was not feasible. However, in Digital Humanities, focused case studies are a valid and often productive method, especially when applying computational text analysis. This study **aims not to generalize across all K-dramas**, but to demonstrate and explore how text mining can reveal meaningful translation patterns in a single, culturally rich corpus.

The primary data for the analysis came from the complete subtitle files of *Squid Game*, both in Korean and in their official English translations. These subtitle files were processed and structured into aligned datasets using Python. Each subtitle line was matched with its translation using timestamp proximity, and organized with metadata such as episode number, start and end time, and the actual dialogue text. While this method does not always guarantee perfect semantic alignment, it was sufficient for the kind of comparative and contextual analysis the project set out to perform.

This study explored three major aspects of subtitle translation. One area of focus was **translation simplification**, particularly how the constraints of subtitling — such as character limits and reading speed — may lead to a reduction in linguistic richness when moving from Korean to English. To investigate this, I used keyword frequency analysis and basic measures of lexical diversity, such as type-token ratio, to compare the complexity and variety of vocabulary across the two languages.

Another significant area of analysis was the translation of **politeness and formality levels**, which are deeply embedded in Korean through its complex system of honorifics and speech levels. English lacks direct equivalents for many of these forms, which makes subtitling a particularly interesting challenge. I compiled a list of common Korean honorific expressions, such as “-nim”, “-ssi”, and respectful verb endings, and examined how they were represented in the English subtitles. This part of the study aimed to reveal patterns of loss, flattening, or creative adaptation in the translation of hierarchical speech.

Finally, I examined the handling of **culturally specific expressions and idioms** in the subtitles. Using a curated list of commonly used Korean phrases — such as “아이구”, “수고하셨습니다”, and “형님” — I applied keyword-in-context (KWIC) analysis to observe how these terms appeared in different scenes and how their meaning was rendered in English. This approach allowed for a more nuanced understanding of the choices translators make when dealing with cultural expressions that may not have direct equivalents in the target language.

The entire analysis was carried out using Python, with libraries like Pandas for data manipulation, NLTK for basic natural language processing, and Matplotlib/Seaborn for generating visualizations. All work was done in Google Colab to ensure easy data handling and reproducibility. The goal throughout was not only to extract patterns from the data but also to interpret them in the context of cultural and linguistic meaning — staying true to the interdisciplinary spirit of Digital Humanities.

While the project is limited in its scope, focusing on a single drama and translation pair, it offers a detailed snapshot of how digital tools can help uncover subtleties in language transfer. Rather than making broad generalizations, this study aims to show how computational methods can enhance our understanding of translation strategies, especially in a cross-cultural media context where subtitling plays a crucial role in shaping global perceptions of language and culture.

Analysis and Interpretation

Alignment of subtitles

The first step I did was the alignment of Korean and English subtitles to find the translation simplification, formality and politeness and translation accuracy.

Korean: "자, 자, 자. 모두 다 나눠 갖는 거야. 알았지?"

English: "All right. Share them with everyone."

➤ **Simplification:** The repeated “자, 자, 자” which shows urgency or emphasis is dropped.

The Korean sentences are found to be longer and richer in detail and emotions whereas when translated in English it tends to vanish the details. Also there are some sentences that not at all matched the source and the target language when translated resulting in inaccurate and inconsistent translations tending towards the loss of meaning.

Korean: "돈도 없으면서 왜 그런 걸 사."

English: "You can’t even afford that."

➤ **Loose Translation:** No mention of “돈도 없으면서” ("even though you have no money").

When comes to politeness and formality shift it is usually seen that Korean language have a lengthy and important formality hierarchy usage on the other hand in English language it is more into neutral form or mostly similar form of response is used irrespective of person spoken to or for. There is one of the examples that I can explain through

Korean: "형, 일단 여기서 나가야 해요."

English: "Hey, we got to get outta here."

➤ **Formality Drop:** "형" is a respectful word for elder brother/friend, and "해요" is polite. But English just says “Hey.

Word frequency analysis

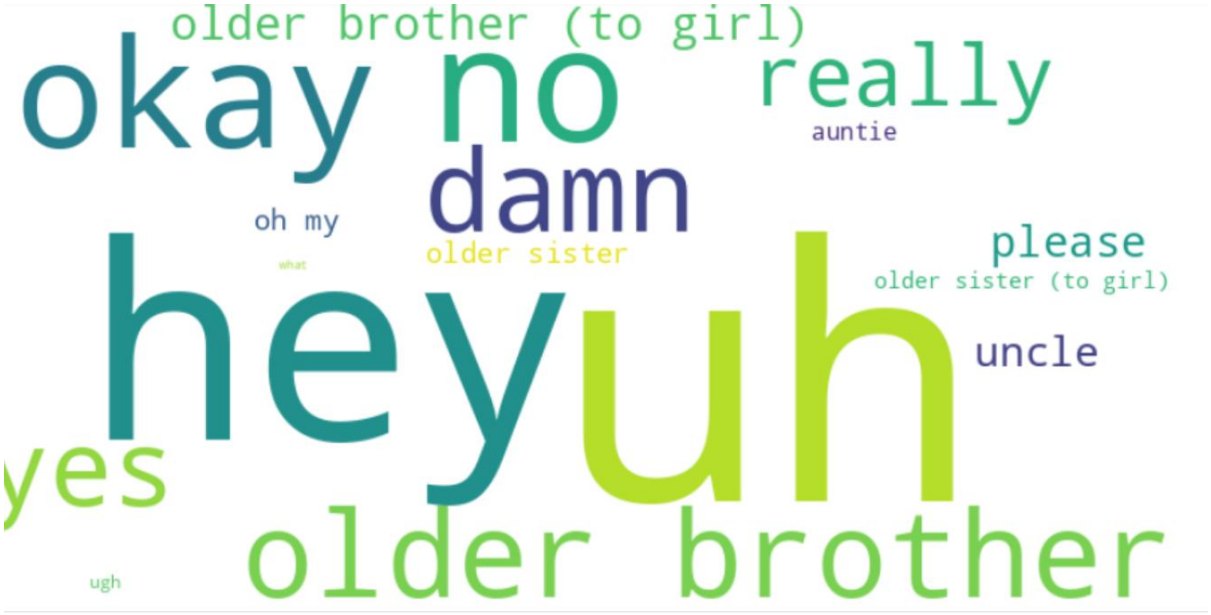
The top 50 words from both the Korean and English subtitles of *Squid Game* highlight key themes like survival, tension, and character relationships. In the Korean subtitles, frequent references to the main characters, 기훈 (Gi-hun) and 상우 (Sang-woo), show the focus on their interactions. Words related to emotions and sensory experiences, like 숨소리 (breathing) and 비명 (scream), emphasize the show's intense atmosphere. The Korean list also reflects themes of family and gender, with words like 엄마 (mom) and 여자 (woman).

The English subtitles highlight similar themes, with frequent occurring words like - **I**, **MAN**, **PLAYER**, and **HUN** (Gi-hun). Words like **MUSIC**, **SIGHS**, and **GASPS** emphasize emotional and physical reactions. Focus on game structure, seen in terms like **PLAYER** and **VIP**, and dialogue markers like "**Hey**" and "**Okay**" brigs out more conversational exchanges.

Overall, both languages emphasize survival and emotional responses suggesting that the drama is inclined more into the action and survival storyline, the Korean version focuses more on character dynamics, while the English version includes more dialogue-driven expressions.

In analysing the translation of cultural expressions between Korean and English in the subtitles of *Squid Game*, we see how the subtitlers faced the challenge of adapting complex cultural nuances into a format that is comprehensible to an international audience, often at the expense of the richness and depth of the original language.

Word cloud: Korean cultural terms word frequency



Word cloud: Korean cultural terms word frequency translated in English

Top 50 high frequency Korean words

Top 50 high frequency English words

Even after adding stop words such as “I”, “do” it was still showing because the name in the drama must be translated in English creating another word in the cloud.

Cultural terms analysis using KWIC

Korean, as a language, is filled with expressions, idioms, and cultural references that carry significant meaning within Korean society. These expressions carry deep-rooted cultural values, familial bonds, and societal hierarchies which plays a signifant role in lifestyle of Korean people. In translating these into English, subtitlers face the challenge of conveying not only the literal meaning but also the cultural weight of the expression. Often, the original expression is simplified to ensure the audience grasps the general idea without necessarily understanding the cultural context.

It is observed that the frequent use of "형님" (hyeongnim), a term that denotes respect for an older male, typically an older brother or a senior figure in a group. The cultural weight of this term is significant, as it encapsulates both familial affection and societal hierarchy. In the English subtitles, "형님" is often rendered as "sir" or "bro," both of which lose the intimate and hierarchical meaning expressed in the Korean version . The translation shifts from a culturally specific expression of familial or group-based respect to a more general term of politeness, thereby losing the connection to the nuanced relationships among characters.

Another notable example is the term "새끼" (sae-ggi), which translates to "bastard" or "son of a bitch" in English, used to express intense anger or insult, often invoking a sense of disrespect tied to Korean values of hierarchy and honor. In English, "새끼" is

translated as "bastard" or simply as "you," which depersonalizes the insult and removes the weight of the cultural insult. Here the subtitlers are deciding the strength of the insult.

Furthermore, we see how Korean expletives like “씨발” (ssibal), a curse word often used to express frustration or aggression, are sometimes omitted or softened in English translations. Instead of a direct translation, the subtitlers may use milder alternatives or even leave the expletive out completely in order to ensure that the content is suitable for international audiences. During the translation the cultural aspect of the sentences are lost, it is connecting the international audience through subtitles but at the same time not.

Politeness and Formality Shifts

Complex system of speech levels in Korean reflects the social hierarchy, age, and status. In Korean, the formality of an expression can significantly change based on the relationship between the speaker and the listener. A person talking to their mother and teacher or a relative would be very much different in all the three cases depending on the relationship type and the closeness with listener.

It has been observed that exchanges where characters use formal speech (존댓말, jondaetmal) like "그럼, 형" (geureom, hyeong), the term "형" (hyeong) is used to refer to an older brother or senior figure with respect by the males. The English subtitles tend to omit the honorifics, translating the phrase simply as "bro" or "okay" without any reference to the underlying cultural respect embedded in the Korean term. The honorifics in Korean are vital in demonstrating the formality of the interaction by omitting these markers, the translation loses much of the subtlety of the original relationship dynamics, rendering the dialogue more universally accessible but culturally less specific.

Similarly, the use of "형님" (hyeongnim) in Korean is a culturally rich expression of both familiarity and respect. While "형님" can be translated to "sir" or "boss" in English, these terms do not capture the familial bond or the specific cultural nuances associated with the word in Korean society. The loss of the honorific term reflects a shift from the hierarchical nature of the Korean social structure to a more egalitarian form of address that is common in English-speaking cultures. This shift alters the perception of the relationship, as the respect conveyed in the original term is diminished in the translation.

Same goes with the Korean explicit expressions, the softened version of these expressions in English tends to make audience take the scene on a more lighter note whereas maybe the particular scene showcasing a very meaningful or deep emotions that can be missed by the international audience.

Therefore it can be said that the cultural nuance, the simplification of translation and the formality drop is somehow seen I the translated version of subtitles which needs to be worked on by the subtitle writers. This maybe can be overcome by the proper learning and teaching of subtitle writing and before that a in depth knowledge of not just a foreign language but the cultural root it comes with as well.

AI Help and Transparency

For this research, I would rate my reliance on AI as **7 out of 10**. AI tools like Chat GPT were used primarily for coding part of the project, drafting code for keyword extraction and context analysis (KWIC), and helping interpret some outputs. However, the decision of analysing which subtitle pair and explaining which particular result was a individual decision.

All code provided by AI was explained in detail using Markdown and the comment (#) method, includes how each function operates, what data it processes, and how the results support the research objectives. The transparency in documentation ensures that the analysis remains reproducible and clearly separates human insight from automated assistance.

References

- **Gottlieb, H. (2001).**
  Screen Translation: Six Studies in Subtitling, Dubbing and Voice-Over.
  University of Copenhagen, Department of English.
- **Chaume, F. (2012).**
  Audiovisual Translation: Dubbing.
  Routledge.
- **Pedersen, J. (2011).**
  Subtitling Norms for Television: An Exploration Focussing on Extralinguistic Cultural References.
  John Benjamins Publishing.
- **Nida, E. A., & Taber, C. R. (1982).**
  The Theory and Practice of Translation.
  Brill Archive.
- **House, J. (2015).**
  Translation Quality Assessment: Past and Present.
  Routledge.
- **Venuti, L. (2012).**
  The Translator's Invisibility: A History of Translation.
  Routledge.
- **Jin, Y. (2016).**
  Translation and Cultural Identity: Korean Television Drama in China.
  The Translator, 22(2), 189–203.
- **Oh, D. C. (2021).**
  Global Netflix: K-dramas, Cultural Proximity, and Platformed Transnationalism.
  Media, Culture & Society, 43(6), 1056–1071.
- **Munday, J. (2016).**
  Introducing Translation Studies: Theories and Applications (4th ed.).
  Routledge.
- **Baker, M. (1992).**
  In Other Words: A Coursebook on Translation.
  Routledge.
- **Piper, A. (2019).**
  Enumerations: Data and Literary Study.
  University of Chicago Press.
- **Bamman, D., & Crane, G. (2011).**
  The Logic and Discovery of Textual Allusion.
  Digital Scholarship in the Humanities, 26(2), 213–229.