Multimodal Challenges in Subtitling Climate Documentaries: An Analysis of Arabic-English Transfer

Maha Ashraf Mostafa

Supervised by:

Prof. Nagwa Younis
Professor of Linguistics
Faculty of Education
Ain Shams University
Department of English

Dr. Passant Hamed
Lecturer of Translation Studies
Faculty of Al-Alsun
Ain Shams University
Department of English

Abstract

The climate crisis presents an existential threat to humanity, making its discussion paramount across all fields of study to enhance public awareness. This study contributes to this effort by examining the audiovisual translation (AVT) of climate change documentaries. AVT involves translating multimodal texts, which are texts that convey meaning through the interaction of different modes (e.g., image, sound and text). For audiences unfamiliar with the source language, subtitles are crucial for meaning construction. This research specifically investigates the challenges in transferring key syntactic structures, such as adjective phrases and compounding, during the subtitling process between English (Source Language) and Arabic (Target Language). The fundamental structural differences between these two languages present significant syntactic obstacles for subtitlers. The study employs Hartmut Stöckl's (2004) categorization of core modes and draws upon Kress and van Leeuwen's Multimodal Discourse Analysis (2006) approach. Two Netflix climate documentaries, *Brave Blue World (2020)* and *Our Planet* (2019), serve as case studies. The analysis bridges academic translation theory with the practical strategies utilized by professional subtitlers. The findings indicate that subtitlers frequently employ strategies like omission and modifying "cutting" (text segmentation/line breaks) to effectively manage the syntactic transfer while maintaining multimodal balance which is the harmonious interaction between the on-screen audiovisual elements and the subtitled text.

Keywords: Climate Change, Multimodality, Subtitling Challenges, Compounding, Adjective Phrases, Core Modes, Visual Grammar, Climate Documentaries

الملخص

تُعد أزمة المناخ مصدر قلق عالمي وتستلزم وعيًا واسع النطاق للحد من عواقبها. ويُعد النشر الفعال للمعلومات عن أزمة المناخ أمرًا بالغ الأهمية، مما يجعل الترجمة مجالًا محوريًا للدراسة لقدرتها على نقل المعلومات من لغة إلى أخرى مما يسهم في نشر الوعي عن الأزمات المجتمعية. يركز هذا البحث على تحليل الترجمة السمعية البصرية للبرامج الوثائقية المناخية والتي هي نصوص متعددة الوسائط تنقل المعنى من خلال التفاعل بين وسائط مختلفة مثل الصوت والصورة واللغة المكتوبة. يواجه مترجمو الشاشة مهمة صعبة تتمثل في نقل المعنى بدقة مع الحفاظ على توازن دقيق بين الوسائط. تتحرى هذه الدراسة على وجه التحديد التحديات التي يواجهها مترجمو الشاشة عند نقل تركيبين نحويين وهما العبارات الوصفية والصيغ المركبة من اللغة الإنجليزية إلى اللغة العربية. تُشكل هذه التراكيب صعوبات خاصة بسبب الفروقات النحوية بين اللغتين وتزداد تعقيدًا بسبب القيود التقنية للترجمة السمعية البصرية. لدراسة هذه التحديات تتبنى الدراسة تصنيف هارتومت شتوكل للوسائط (2004) ونموذج القواعد البصرية لكريز وفان ليوين (2006). يحلل البحث ترجمة برنامجين وثائقين عن المناخ على منصة نتفليكس. من خلال دراسة الاستراتيجيات التي يستخدمها مترجمو الشاشة للحفاظ على التوازن بين الوسائط على الرغم من الاختلافات النحوية، تهدف هذه الدراسة إلى سد الفجوة بين النظريات الأكاديمية والممارسات الواقعية للترجمة السمعية البصرية. وقد كشفت نتائج الدراسة أن مترجمي الشاشة يستخدمون استراتيجيات متنوعة تحافظ على التوازن بين الوسائط مثل الحذف وتغيير التوقيت. تقدم النتائج رؤى قيمة لتدريب مترجمي الشاشة المستقبليين، وتزودهم بفهم للعوامل التي تنطوي عليها عملية الترجمة السمعية البصرية متعددة الوسائط.

Multimodal Challenges in Subtitling Climate Documentaries: An Analysis of Arabic-English Transfer

Maha Ashraf Mostafa

Supervised by:

Prof. Nagwa Younis
Professor of Linguistics
Faculty of Education
Ain Shams University
Department of English

Dr. Passant Hamed
Lecturer of Translation Studies
Faculty of Al-Alsun
Ain Shams University
Department of English

Introduction

The climate crisis is a pressing global concern with the potential for profound societal consequences, making its discussion a staple in modern media productions. The central goal of these productions, often taking the form of documentaries, is to raise public awareness about environmental protection. The international dissemination of this vital knowledge requires audiovisual translation (AVT) to bridge linguistic and cultural divides.

Audiovisual translation is the process of transferring a multimodal text, which is a combination of various modes like sound, dialogue, images, movement, and written signs, from one language to another. Multimodality is thus integral to translation studies, investigating how these various modes are accurately transferred to a target audience.

The process of subtitling these multimodal texts presents significant challenges. Translators must meticulously match different modes and maintain meaning while adhering to strict technical constraints, notably the limited character count (typically 36-37 characters per line). This limitation complicates the task of accurately reflecting all layers of the source text.

Compounding this difficulty are the syntactic differences between the source and target languages. For instance, the distinctive arrangement of lexemes in structures like compounding and adjective phrases between English and Arabic creates particular difficulties for subtitlers.

This study focuses on analyzing the translation techniques used to render two specific linguistic structures: compounds and adjective phrases. Compounding is a word-formation process (as described by Hacken, 2021) where two lexemes, each with its own meaning, are combined to form a new word. The analysis encompasses all forms of compounding found in the source texts.

Adjective Phrases are structures headed by an adjective, which may also include one or more modifiers that refine its meaning. Adjective phrases, in particular, prove challenging due to differences in structure arrangement. For example, the English structure "storm-struck area" which includes a modifier followed by the head must be rendered in Arabic as "المنطقة التي ضربتها العاصفة" which constitutes a head followed by modifier, back-translating to "area struck by storm". If the subtitler, constrained by character limits, splits this phrase across two lines, the multimodal balance can be disrupted: the spoken audio might end with the head ("area"), while the screen still displays the Arabic equivalent of the modifier ("storm-struck"), creating a noticeable incoherence for the audience. While these structures are generally unproblematic in written translation, the technical constraints of subtitling make them difficult to manage.

This research aims to investigate the specific strategies subtitlers use to transfer these challenging structures from English into Arabic. The study analyzes the Arabic subtitles of two climate documentaries available on Netflix: *Brave Blue World (2020)* and *Our Planet* (2019). The analysis is guided by Kress and van Leeuwen's (2006) Visual Grammar model and Hartmut Stöckl's (2004) categorization of modes. The study will specifically address the following questions: To what extent do linguistic structures like compounding and adjective phrases pose a challenge in subtitling climate documentaries? What are the different strategies used by subtitlers in dealing with syntactic differences between English and Arabic? How can the study of audiovisual translation shed light on the issue of climate change? The study's importance lies in its dual contribution: highlighting the urgency of the climate crisis by examining the transfer of its awareness messages, and linking academic translation theorization with professional subtitling practices. The goal is to understand the techniques subtitlers employ to ensure multimodal coherence and harmony between the spoken (audio) and written (on-screen text) modes, further emphasizing the crucial integration of multimodality into translation studies.

Theoretical Framework

This study draws on two theoretical frameworks to analyze the Arabic subtitles of selected English multimodal productions and to examine the strategies used in their translation into Arabic. The first framework is Hartmut Stöckl's (2004) multimodal categorization, which outlines core modes, their medial variants, peripheral modes, submodes,

and features. The second is Kress and van Leeuwen's (2006) Multimodal Discourse Analysis theory. Stöckl's model is employed for its detailed classification system that explains the function of each mode and how they interact. Kress and van Leeuwen's framework is used as a foundational theory in the field of multimodality and supports the analytical aims of this study. It is also used as a tool for analyzing the images accompanying the subtitles in the said documentaries, reflecting their impact on the successful delivery of meaning.

Hartmut Stöckl's Multimodal Categorization (2004)

In his chapter, Hartmut Stöckl (2004) defines a mode as a medium embedded with signs through which people communicate. This definition establishes the foundation for understanding the role of modes in meaning-making. By identifying and analyzing the different modes present in a multimodal text, we can uncover the specific functions each mode performs and how they contribute to the overall communicative purpose. In any multimodal product, these modes operate in an interconnected way to construct meaning. Stöckl (2004) stresses the importance of theoretically unpacking these modes, noting that:

> When 'reading' a multimodal text, average recipients will normally become only dimly aware of the fact that they are processing information encoded in different modes. The manifold inter-modal connections that need to be made in order to understand a complex message distributed across various semiotics will go largely unnoticed. All modes, then, have become a single unified gestalt in perception, and it is our neurological and cognitive disposition for multimodal information processing that is responsible for this kind of ease in our handling of multimodal artefacts. (Stöckl 2004, 16)

This highlights the often-invisible cognitive effort involved in decoding multimodal content. Analyzing the characteristics and functions of each mode enables us to better understand their roles, the meanings they imply, and the mental processes involved in interpreting them. In audiovisual media, modes work in such synchrony that the human brain frequently fails to register their separateness. To illustrate this complex interaction, Stöckl presents a visual diagram (figure 1) categorizing what he terms

"core modes" along with their medial variants and submodes. The four primary core modes identified are: image, language, sound and music.

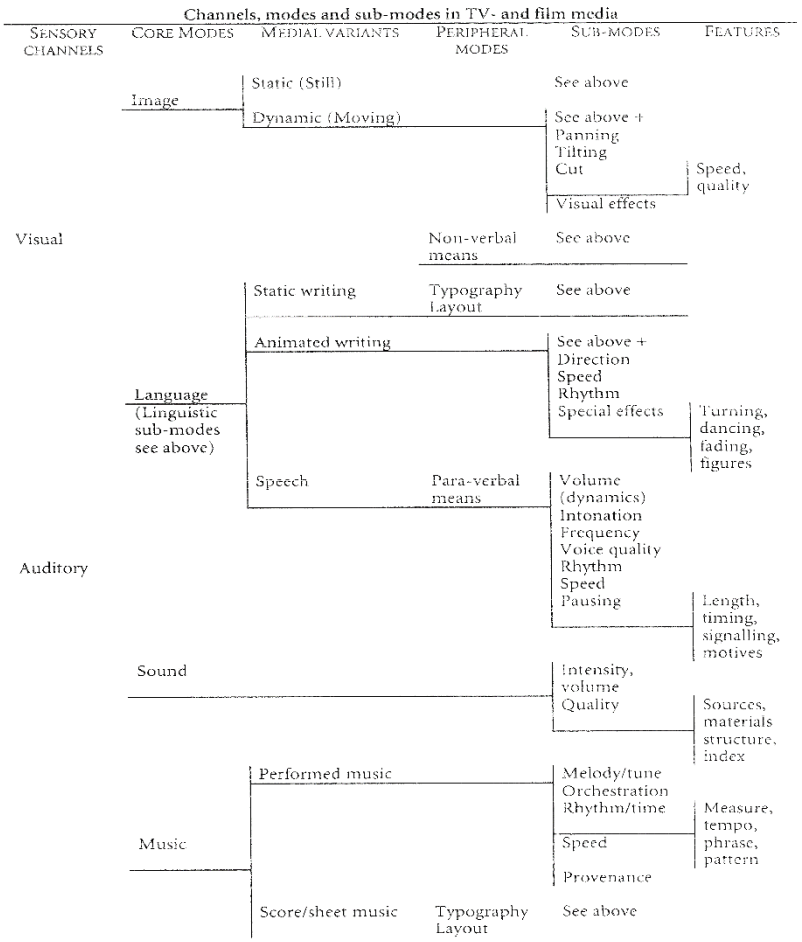

Figure 1: Hartmut Stöckl's Categorization of Modes (2004)

Stöckl begins by discussing image as a core mode, which includes two medial variants: static and dynamic. He then moves on to the language mode, breaking it down into three medial variants: static writing, animated writing and speech. The submodes of static writing include elements such as font size, color, spacing, and lexical choice. For animated writing, submodes include direction, speed, and rhythm. The speech variant includes submodes like volume, intonation, speed, frequency and pauses. The third core mode is sound, which is analyzed through submodes such as intensity, volume, and quality. Finally, the music mode is divided into two medial variants: performed music and score/sheet music.

Stöckl (2004) further emphasizes a key point: these modes do not exist in isolation. Instead, they often overlap and intertwine, with no strict boundaries separating them. For instance, the mode of speech (a variant of language) cannot exist without the sound mode. In audiovisual texts, the interdependence of modes is so deep-rooted that separating them with precision is often unfeasible. However, for analytical purposes, particularly in this study, it is essential to investigate each mode's contribution to meaning-making and examine whether this semiotic coherence is maintained in the translation process. While Stöckl does not directly address translation in his categorization, Luiz Pérez González (2014) extends this model to the field of audiovisual translation.

González (2014) presents multimodality as a useful framework for examining audiovisual translation, critiquing the fact that much of the field tends to concentrate on linguistic transfer alone, often overlooking the semiotic richness of audiovisual texts. Through the lens of multimodality, it becomes possible to explore how different modes interact and how meaning is constructed beyond just language. He adopts Stöckl's categorization to examine the role of each mode, its variants and submodes, proposing that translators must engage with these multiple semiotic resources when adapting audiovisual content for a target audience. In this sense, audiovisual translators are not just transferring text; they are reconfiguring a multimodal message. This study applies Stöckl's framework to investigate how core modes and their submodes are considered in the translation process, particularly how they influence the transfer and interpretation of meaning.

González's work is especially relevant here as it offers a practical model for applying Stöckl's categorization in the analysis of subtitles which is the aim of this study. His approach demonstrates how theoretical categorization can be operationalized in real translation scenarios, making it a foundational resource for this research. Furthermore, González (2014) introduces a taxonomy of subtitling strategies, which this study also adopts. He argues that subtitlers often employ shifts to navigate the technical and temporal constraints inherent in the subtitling process. These shifts typically manifest as either "reformulation" or "omission." Reformulation involves altering the structure of the original message, such as converting active to passive voice, changing sentence polarity, or replacing nouns with pronouns. Omission, on the other hand, can occur at both word and sentence levels, where content is entirely removed due to spatial or temporal limitations. These strategies will be used in the current

analysis to examine how subtitlers of selected climate documentaries respond to the challenges posed by specific linguistic and semiotic structures.

Kress and van Leeuwen's Visual Grammar (2006)

This study employs the Visual Grammar model proposed by Kress and van Leeuwen (2006) to analyze the core mode of image and investigate how subtitlers utilize its semiotic functions to address the challenges posed by the complex linguistic structures under examination. Kress and van Leeuwen's framework is designed to interpret and explain the communicative functions of images, framing them not simply as visual elements but as active participants in meaning-making. According to their theory, images, like language, perform multiple communicative functions and can be systematically analyzed to uncover these roles.

> Every image fulfills both an 'ideational' function, a function of representing 'the world around and inside us' and an 'interpersonal' function, a function of enacting social interactions as social relations. All message entities – texts – also attempt to present a coherent 'world of the text', what Halliday calls the 'textual' function – a world in which all the elements of the text cohere internally, and which itself coheres with its relevant environment. (Kress & van Leeuwen 2006, 15)

The first function, the representational function, refers to the content or concept conveyed by the image or the message or idea it aims to express. Analyzing this function involves exploring the image's intended meaning and communicative goal. The second function, the interactive function, addresses the relationship between the viewer and the image. It considers how the image is constructed to engage its audience, and how visual choices influence viewer perception and interaction. The third function is the compositional function, which relates to the layout, arrangement and structure of visual elements. This function is a key in understanding how the organization of visual content shapes interpretation and supports meaning.

Kress and van Leeuwen's Visual Grammar is presented as a practical tool for image analysis, with a portion of their work dedicated to its applicability to moving images. This study applies their model specifically to examine the visual elements within audiovisual content and to assess how these elements assist subtitlers in communicating meaning across languages. The goal is to analyze how subtitlers engage

with the representational, interactive and compositional aspects of images when dealing with the linguistic challenges posed by adjective phrases and compounding. Additionally, the study incorporates Stöckl's (2004) multimodal categorization to further dissect the audiovisual material into its constituent layers and examine the specific features of each mode. These two models are used in tandem because they complement one another: while Stöckl provides a structured framework for identifying and categorizing the modes present in audiovisual texts, Kress and van Leeuwen offer a deeper lens for interpreting the semiotic functions of those modes. Together, they offer a comprehensive analytical approach for investigating the role of image in the subtitling process.

Review of Literature

In recent years, translation studies have seen a growing shift toward integrating multimodal theory into the analysis of translated texts, particularly in the context of audiovisual translation. Traditional approaches have tended to prioritize linguistic or cultural factors, often overlooking the multimodal makeup of modern texts. However, researchers are now acknowledging that communication in audiovisual media involves a complex interplay of multiple semiotic modes: text, sound, image, and movement, which all contribute to meaning-making. This section reviews the literature that informs the present study, with particular focus on subtitling, multimodal discourse and translation between English and Arabic.

Taylor (2016), in his influential work "The Multimodal Approach in Audiovisual Translation", argues that the study of multimodality has largely remained within the domain of linguistics, and has not been fully applied to translation studies. He advocates for a comprehensive approach that accounts for the interaction of semiotic resources in audiovisual content, emphasizing that these elements do not function independently but contribute jointly to meaning. For subtitlers, especially, understanding the multimodal nature of the source material is crucial for making informed translation choices, particularly when dealing with time and space limitations. Taylor's model is central to the present study's approach, as it frames subtitling as a process that involves managing and reconfiguring several modes simultaneously.

In a similar vein, O'Sullivan (2013) challenges the traditional mono-modal perspective of translation by emphasizing the translator's reliance on various non-verbal cues present in the source text. She suggests that viewing translation through a multimodal lens not only enhances the accuracy of the transfer but also supports the translator's decision-making process. According to O'Sullivan, subtitlers can draw

upon the full range of semiotic resources in the audiovisual text to support or even compensate for linguistic reductions that occur due to temporal constraints.

Suyudi, Nababan, Santosa and Djatmika (2019) offer a practical application of multimodal theory in their analysis of the subtitles in Forrest Gump. Their study focuses on the use of amplification, a strategy that involves adding or paraphrasing content to enhance clarity and comprehension for the target audience. They argue that subtitlers must be aware of the various modes at play and the ways they can be harnessed to preserve or recover meaning across languages and cultures. The relevance of their work lies in its methodological approach, which mirrors the aims of the current research: to examine the effectiveness of subtitling strategies in managing the challenges posed by the multimodal complexity of audiovisual texts.

González (2019) contributes a comprehensive overview of the evolution of audiovisual translation, mapping its transition from being labelled "media translation" to the now widely accepted term "multimodal translation." His study provides a foundational understanding of AVT, not just from a theoretical standpoint but also in terms of its historical trajectory. This overview is particularly useful in contextualizing the current research within the broader development of translation theory and practice.

Gamal (2009), in his chapter "Foreign Movies in Egypt: Subtitling, Dubbing and Adaptation" addresses the regional specificities of audiovisual translation in the Arab world, particularly Egypt. He offers a historical perspective on the emergence of subtitling and dubbing in the region and discusses the linguistic and cultural challenges faced by subtitlers. Importantly, Gamal also reflects on audience reception and the expectations of Arabic-speaking viewers, making his work highly relevant to this study, which focuses on English-Arabic translation in climate documentaries.

Khuddro's (2018) work *Linguistic Issues and Quality Assessment of English-Arabic Audiovisual Translation* directly informs the linguistic aspect of this study. The book delves into the syntactic and semantic differences between English and Arabic, exploring how these differences, such as collective nouns, dual forms, and diacritics, pose challenges in subtitling. He also examines the techniques subtitlers use to overcome these hurdles under time and space constraints. His insights are valuable in understanding the practical implications of translating between

structurally different languages and how these implications may affect the multimodal coherence of the translated product.

González (2014) further emphasizes the need to consider the multimodal makeup of audiovisual texts in translation. He notes that research has long focused on linguistic transfer, often to the exclusion of other semiotic resources. By drawing attention to non-verbal elements, González supports the idea that multimodal analysis is essential for fully understanding how meaning is conveyed in audiovisual media. This perspective strengthens the rationale for using multimodal discourse analysis as a key analytical framework in this study.

The role of translation in communicating climate change messages is addressed by Tchoukaleyska et al. (2021) in their article "Special Issue Introduction: Climate Change Knowledge Translation". They argue that translation plays a crucial role in disseminating climate change discourse to global audiences and fostering cross-cultural understanding. The authors emphasize that academics and translators must work together to ensure that climate change knowledge is accessible and meaningful in different linguistic and cultural contexts. This work supports the purpose of the present study, which focuses on the subtitling of climate documentaries and the role of translation in spreading awareness about environmental issues.

Altogether, the reviewed literature illustrates that translation, particularly in the audiovisual domain, cannot be reduced to mere linguistic transfer. Instead, it involves navigating a web of interacting modes and overcoming structural and technical constraints to produce a cohesive and meaningful target text. Although existing research offers valuable insights into the multimodal nature of audiovisual translation, a noticeable gap remains in the analysis of how subtitlers manage structural linguistic differences, such as adjective phrases and compound forms, within the constraints of timing and space and how these decisions affect multimodal coherence. This study aims to address this gap by examining subtitling strategies in English-Arabic translations of climate documentaries, thereby contributing to the broader discourse on multimodality in translation studies.

Methodology

This research employs a qualitative methodology to examine the Arabic subtitles of selected English-language climate documentaries. The primary objective is to analyze the challenges that arise when translating specific linguistic structures, namely compounding and adjective phrases, and to assess the strategies subtitlers adopt to address these challenges. Furthermore, the study investigates how effectively these strategies

maintain multimodal coherence between the verbal and non-verbal elements of the audiovisual product.

Data Selection

The data for this study consists of two English-language climate change documentaries available on the streaming platform Netflix: Our Planet (2019) and Brave Blue World (2020). Brave Blue World is a documentary that has a runtime of approximately 90 minutes while Our Planet is a series that consists of two seasons and eight episodes each. Both documentaries include official Arabic subtitles. The selection of these particular titles was not influenced by their release dates but rather by their accessibility and relevance to the study's linguistic focus. As Netflix offers globally available, on-demand audiovisual content, it was chosen as the source platform due to its ease of access in comparison to more restricted television channels.

In addition, Netflix is known for implementing a standardized and institutional approach to subtitling. Subtitles on the platform are created by professional translators who follow specific stylistic guidelines and are subject to quality assurance protocols. This ensures that the subtitles examined in this study were produced by skilled subtitlers who are likely to be familiar with the principles of multimodal coherence, thereby making their work suitable for analysis of strategic decision-making in translation.

The focus on climate change documentaries was guided by two primary motivations. First, climate change is a pressing global issue that demands interdisciplinary attention, including from the field of translation studies. Documentaries in this genre serve a critical role in public education and awareness, making their translation particularly significant. Second, these documentaries are rich in complex linguistic constructions and domain-specific terminology, which provide ample material for the analysis of compound structures and adjective phrases. Scientific terminology, which often involves compound forms, appears frequently in the more technical segments, while adjective phrases are commonly used in descriptive sections relating to daily life and environmental practices. Both of which are relevant to the structures selected for investigation in this study.

Procedures of Data Analysis

The data analysis follows a structured, multi-step process. Initially, the source text is examined to identify occurrences of the targeted linguistic structures: compound constructions and adjective phrases. In the second step, the Arabic subtitle equivalents of these structures are

identified and extracted for comparative analysis. The third phase involves categorizing the subtitlers' strategies using Stöckl's (2004) framework of modes. This step includes examining the core mode involved, followed by its medial variant and any applicable sub-modes. The fourth stage evaluates whether the translation choices contribute to maintaining or disrupting the multimodal coherence of the text. This involves examining how well the verbal mode integrates with other semiotic elements such as sound and image. Finally, Kress and van Leeuwen's (2006) Visual Grammar Model is applied to analyze the visual dimension of the selected scenes. This step investigates how images contribute to meaning-making and whether the subtitlers utilized these visual elements to support the translation, particularly in cases where linguistic reduction or reformulation occurred. All Arabic transliterations in the analysis adhere to the Hans Wehr Transliteration System to ensure consistency and academic clarity throughout the study.

Example (1): From Brave Blue World (2020)		
Time	**Source Language**	**Target Language**
29:56	We hit the geological jackpot by having this large natural underground aquafer	حالفنا الحظ من الناحية الجيولوجية بوجود خزان المياه الجوفية الطبيعي هذا

Analysis

In example (1), this subtitle contains an adjective phrase, where the head noun is "aquafer," modified by three descriptors: "large," "natural" and "underground." Each modifier conveys a distinct attribute. The first indicates size, the second denotes the type or origin and the third specifies location. In the Arabic translation, the subtitler rendered the phrase using an equivalent adjective structure. The head noun becomes "خزان المياه" with "الطبيعي" and "الجوفية" serving as modifiers. Notably, the modifier "large" which refers to size has been omitted. By removing this element, the subtitler managed to fit the phrase within the spatial constraints of the subtitle, preserving essential information while adhering to screen-time limitations. As a result, the static writing which is the medial variant of the core mode of language as per Stockl's (2004) categorization aligns with the speech it accompanies, achieving a sense of multimodal coherence.

This omission strategy is supported by the core mode of image. The visual content of the scene depicts a large tank, clearly representing the object being discussed. The image itself conveys the idea of size, making the verbal expression of "large" redundant. The translator, recognizing that viewers can visually perceive the tank's scale, opted to exclude the size descriptor from the translation. This decision illustrates how the representational function of the image as per Kress & van

Leeuwen (2016) model can guide subtitling choices. If the image had instead emphasized the location of the aquafer, the subtitler might have chosen to omit “underground” instead. This interaction between visual and verbal modes underscores the significance of maintaining multimodal harmony, where meaning is distributed across semiotic resources and translation strategies respond accordingly.

Example (2): From Brave Blue World (2020)		
Time	**Source Language**	**Target Language**
37:48	Recently there was a water tanker strike	كان هناك إضراب لسائقي الخزانات

In the source text of example (2), the phrase *water tanker* functions as a compound structure, where "water" and "tanker" combine to form a new lexical item referring to a specific object. Additionally, the expression includes an adjective phrase, where *strike* serves as the head, modified by *water tanker*. In the Arabic translation, both structures are rendered, with the adjective phrase translated as “إضراب سائقي الخزانات,” where “إضراب” is the head and “سائقي الخزانات” functions as the modifier. However, in translating the compound structure, the subtitler retained only “tanker,” rendered as “خزانات,” and chose to omit the word “water.” This use of omission allowed the subtitler to manage spatial limitations while preserving the intended meaning.

From a multimodal perspective, the omission strategy supports coherence across modes. According to Stöckl’s (2004) categorization, the medial variant of speech aligns with the medial variant of static writing in the subtitle, ensuring that the modes are synchronized at the textual level. This cohesion is further supported by the core mode of image. The visual accompanying the subtitle clearly shows a water tanker, fulfilling the representational function of the image explained by Kress and van Leeuwen (2006). Because the concept of "water" is visually conveyed, its verbal repetition becomes unnecessary, allowing for its omission without a loss in meaning.

This example demonstrates how subtitlers can leverage the interplay between audio, text, and visuals to make concise and effective translational decisions. Meaning in audiovisual media is co-constructed through various semiotic modes, and this interdependence allows subtitlers to use strategies like omission while maintaining communicative clarity. In this case, the translator’s choice to omit "water" succeeded in preserving the message and ensuring multimodal harmony.

Example (3): From Our Planet (2019) - Season 2 - Episode 4		
Time	**Source Language**	**Target Language**
4:24	For the past weeks, her lives have been driven about a 30 kilometer daily commute	طوال الأسابيع القليلة الماضية تمحورت حيواتها حول رحلة يومية طولها 30 كيلومتراً

In example (3), the original English text includes an adjective phrase appears with commute as the head noun, modified by 30 kilometer and daily. The subtitler opted to translate all components of this phrase into Arabic without omitting any elements. To achieve this, the subtitler employed the strategy of adjusting subtitle timing, allocating a full subtitle solely for the adjective phrase. This decision was crucial for maintaining alignment between spoken and written modes.

Had the timing not been altered, the subtitle would have been split mid-phrase, likely after one of the modifiers. Given that the head noun commute appears at the end of the spoken phrase, and Arabic word order places the modifiers before the head, the written subtitle would have displayed "30 kilometer" while the audio was delivering the word "commute". This mismatch would have disrupted the synchronization between audio and text, causing a breakdown in multimodal coherence.

By assigning a separate subtitle to the entire adjective phrase, the subtitler successfully avoided this misalignment. This timing strategy accommodates the structural differences between English and Arabic and ensures that both the spoken and written modes remain in harmony throughout the audiovisual experience.

Example (4): From Our Planet (2019) - Season 2 - Episode 4		
Time	**Source Language**	**Target Language**
05:14	So the home straight is often an uphill waddle towards the windblown clifftops	لذلك فإن المرحلة الأخيرة كثيراً ما تكون تبختراً شاقاً نحو المنحدرات التي تهب عليها الريح

In example (4), at the end of the source sentence in English, there appears an adjective phrase with *clifftops* as the head noun. This noun is itself a compound formed from *cliff* and *tops*, referring to the uppermost parts of cliffs. The modifier *windblown* is also a compound structure, combining *wind* and *blown* to describe a place that is exposed to strong winds. In translating this phrase into Arabic, the subtitler employed two key strategies: timing adjustment and omission.

Firstly, the subtitler altered the timing by dividing the sentence across two subtitles, allocating an entire subtitle specifically for the adjective phrase. This ensured that the Arabic written subtitle aligned closely with the corresponding segment in the spoken English, avoiding a misalignment that would occur if the sentence had been split elsewhere. Without this timing adjustment, the head noun in the spoken variant could have overlapped with the Arabic modifier, disrupting coherence between the two modes.

The second strategy used was omission. The subtitler chose not to render the full compound *clifftops* in Arabic, instead translating only *cliffs* as منحدرات, thereby omitting the element *tops*. This reduction enhances subtitle readability, a common objective in audiovisual translation where time and space are limited. Condensation allows viewers to process subtitles more efficiently, particularly during fast-paced scenes.

The effectiveness of this omission was supported by the accompanying visual content. The image displayed with the subtitle featured the upper portions of mountains or cliffs, visually implying the notion of *tops*. As a result, viewers could infer from the image that منحدرات referred specifically to the highest parts, even though the Arabic subtitle did not explicitly state this. This seamless interaction between visual and verbal elements enabled the subtitler to omit part of the compound without compromising the overall meaning or disrupting multimodal coherence.

Discussion

The analysis conducted in this study highlights the strategies adopted by subtitlers when handling complex linguistic structures, particularly in the context of translation between English and Arabic. The examples examined demonstrate that subtitlers preserve multimodal coherence. They succeed to synchronize the auditory and visual modes, namely, sound and written language, so that the spoken and written messages remain aligned. It is evident that subtitlers avoid splitting utterances in a way that might result in the text on screen diverging from the sequence of the speaker's words.

One of the key strategies used to achieve this balance is omission. In several instances, subtitlers omit part of an adjective phrase or compound structure to avoid breaking up the grammatical unit across subtitle lines. Importantly, the omission does not hinder the target audience's understanding. This is due to the presence of supporting modes, particularly the visual mode. In many of the analyzed examples, the image displayed on screen serves to convey the omitted information.

This synergy between image and text allows viewers to grasp the full meaning even when certain linguistic components are left out. As such, omission emerges as an effective strategy for resolving the challenge of translating syntactically complex structures between two languages with differing grammatical systems, while still preserving multimodal cohesion.

Another notable strategy observed is the reallocation of subtitle timing, where subtitlers assign an entire subtitle line to the complete adjective phrase. This prevents splitting the phrase across multiple subtitles, which could lead to confusion or disrupt the flow of reading. In these cases, subtitlers manipulate the segmentation and timing of subtitles to ensure that the phrase appears as a unified structure. This reflects an implicit understanding of the significance of aligning the spoken and written modes. Adjusting subtitle timing in this manner allows for the retention of meaning, accurate synchronization and the maintenance of multimodal integrity. It suggests that restructuring subtitle segmentation is a successful and conscious technique employed by professionals to manage structural differences between source and target languages.

The identification and analysis of these strategies offer valuable insights for subtitler training and the development of audiovisual translation curricula. Emphasizing the importance of multimodal coherence can help aspiring subtitlers recognize how subtitling decisions impact both comprehension and language learning. Since audiovisual content is often used as a tool for second-language acquisition, inconsistencies between spoken and written language can lead to incorrect language learning. Highlighting this aspect during training would encourage subtitlers to avoid multimodal disjunctions and prioritize clarity and alignment across modes.

This study focuses on the challenges of translating content between English and Arabic, where meaning is conveyed not only verbally but also semiotically. The translator's proficiency influences the strategies employed, and this research underscores the significance of understanding the functions of various modes. By becoming aware of how modes interact, novice subtitlers can better navigate the syntactic differences between the two languages and choose more effective translation strategies.

However, one limitation of this study lies in the limited body of research specifically addressing the intersection of audiovisual translation and multimodal analysis within the English-Arabic language pair. The syntactic disparity between these two languages poses unique challenges when rendering meaning within the spatial and temporal constraints of

subtitles. Further research is needed in this area, particularly studies that adopt a multimodal approach to the subtitling process. Future research could also include quantitative analyses to determine the frequency and effectiveness of different strategies used to achieve multimodal balance. Such data could contribute to a more comprehensive understanding of best practices and support the creation of evidence-based guidelines for subtitler training.

Beyond its linguistic contribution, the study also brings attention to the broader issue of climate change. Integrating environmental topics into academic research across disciplines is essential. By analyzing the translation of climate-related content, this study demonstrates how the field of translation studies can play a role in global awareness and education. Making such material accessible in various languages allows for the dissemination of critical information, encouraging behavioral change and informed action. Therefore, accurate translation in this context becomes a matter of urgency and social responsibility.

Additionally, the findings have implications for the production of audiovisual content itself. If translation is considered during the production phase, it may be possible to allocate more time and space within the text for subtitles. This proactive approach could ease the subtitling process and better support the alignment of modes, ultimately resulting in clearer, more effective communication across languages and cultures.

Conclusion

The escalating crisis of climate change poses a serious threat to global human survival, prompting growing concern and becoming a central focus of scholarly inquiry across various fields. Raising awareness of this pressing issue through academic research is essential, especially within disciplines that contribute to the communication and dissemination of knowledge. This study contributes to that effort by examining the subtitling of climate-related media content, focusing specifically on audiovisual translation.

The primary aim of this research has been to explore the strategies employed by subtitlers to maintain multimodal coherence while transferring structurally complex linguistic elements, such as adjective phrases and compounding, from English into Arabic. The findings reveal that subtitlers frequently rely on omission and adjustment of subtitle segmentation as techniques to preserve harmony between modes, particularly between the spoken and written elements. These strategies

help prevent misalignment and ensure the audience receives a coherent and intelligible message without undermining meaning.

The practical implications of these findings are significant for the training of subtitlers. By demonstrating how theoretical frameworks can be applied in professional practice, this study bridges the gap between academic research and real-world subtitling work. It highlights the importance of teaching subtitlers to recognize multimodal interactions and make informed decisions to ensure that different modes work together effectively.

Future research at the intersection of multimodality and translation could expand on this foundation by investigating the cognitive and psycholinguistic dimensions of the translation process. Such studies could explore how subtitlers make strategic decisions in real time and what mental processes guide their handling of multimodal materials. In the context of this study, that would involve examining how subtitlers rationalize their choices when rendering complex linguistic structures within time and space constraints.

Moreover, future investigations could incorporate quality assessment models to evaluate the success and efficiency of the strategies applied in subtitling multimodal texts. This would provide a more robust framework for assessing not only linguistic fidelity but also the degree of multimodal integration. Ultimately, continued research in this area can enhance the quality of audiovisual translations and contribute to the broader effort of disseminating crucial information, such as climate change awareness, across languages and cultures with greater clarity and impact.

References

Davies, A. (Executive Producer). (2019). *Our Planet* [TV series]. Netflix; Silverback Films.

Gamal, M. (2009). Foreign movies in Egypt: Subtitling, dubbing and adaptation. In A. Goldstein & B. Golubovic (Eds.), *Foreign language movies – Dubbing vs subtitling* (pp. 1–16). Verlag Dr. Kovač.

Hacken, P. T. (2021). The nature of compounding. *Cadernos de Linguística*, *2*(1), 1-21.

Khuddro, A. (2018). *Linguistic issues and quality assessment of English-Arabic audiovisual translation*. Cambridge Scholars Publishing.

Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design.* Routledge.

O'Sullivan, C. (2013). Multimodality as challenge and resource for translation. *The Journal of Specialised Translation*, *20*, 240–256. https://doi.org/10.26034/cm.jostrans.2013.398

Pellatt, A. (Director). (2020). *Brave blue world* [Film]. Brave Blue World Foundation.

Pérez-González, L. (2014). *Audiovisual translation: Theories, methods and issues.* London & New York: Routledge.

Pérez-González, L. (2019). Audiovisual translation. In M. Baker & G. Saldanha (Eds.), *Routledge encyclopedia of translation studies* (pp. 30–34). Routledge. https://doi.org/10.4324/9781315678627

Stöckl, H. (2004). In between Modes: Language and Image in Printed Media. In E. Ventola & C. Charles (Eds.), *Perspectives on Multimodality* (pp. 9-30).

Suyudi, I., Nababan, M. R., Santosa, R., & Djatmika**.** (2019). Analysing the multimodal-based amplification technique in Indonesian subtitle translations of *Forrest Gump*. *International Journal of Innovation, Creativity and Change, 7*(8), 232–250. https://api.semanticscholar.org/CorpusID:209411565

Taylor, C. (2016). The multimodal approach in audiovisual translation. *Target: International Journal of Translation Studies, 28*(2), 222–236. https://doi.org/10.1075/target.28.2.04tay

Tchoukaleyska, R., Richards, G., Vasseur, L., Manuel, P., Breen, S.-P., Olson, K., Curtis, J. C. C., & Vodden, K. (2021). Special issue introduction: Climate change knowledge translation. *Journal of Community Engagement and Scholarship, 13*(3), 7–10. https://doi.org/10.54656/MHUN3219