International Journal of Linguistics, Literature and Translation
ISSN: 2617-0299 (Online); ISSN: 2708-0099 (Print)
DOI: 10.32996/ijllt
Journal Homepage: www.al-kindipublisher.com/index.php/ijllt

IJLLT
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Satirical Humor Translation in YouTube Automatic Indonesian Subtitles

Nafingatul Mustafidah
Master's Student, Applied Linguistics Study Program, Department of Applied Linguistics, Faculty of Languages, Arts, and Culture, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia
Corresponding Author: Nafingatul Mustafidah, **E-mail**: nafingatulmustafidah.2025@student.uny.ac.id
ORCID: https://orcid.org/0009-0009-9993-018X

| ABSTRACT
This study investigates the translation of satirical humor in YouTube's automatic Indonesian subtitles of a stand-up comedy performance by Armando Anto. Using a qualitative descriptive case study, the research compares the English source text with the automatically generated Indonesian subtitles. Eleven satirical segments were identified and analyzed to examine meaning shifts, humor preservation, and the contribution of multimodal elements such as music, gestures, and facial expressions. The findings show that the subtitles generally preserve the literal meaning of the source text but frequently weaken or fail to convey the humorous effect. Irony, wordplay, and culturally embedded satire are often reduced, while humor conveyed through musical and visual elements is largely absent from the subtitles. In addition, several automatic speech recognition (ASR) errors negatively affect humor comprehension. This study concludes that YouTube's automatic subtitles function more effectively as lexical translations than as translations of multimodal comedy performances. The findings highlight the challenges of translating satirical humor through automatic subtitle systems and the importance of multimodal features in humor translation.

| KEYWORDS

Satirical humor, Automatic subtitles, Audiovisual translation, Multimodality, YouTube

| ARTICLE INFORMATION

ACCEPTED: 01 June 2026 **PUBLISHED:** 29 June 2026 **DOI:** 10.32996/ijllt.2026.9.7.1

1. Introduction

The rapid growth of digital video platforms, particularly YouTube, has significantly changed the way people consume audiovisual content worldwide. With more than 500 hours of video uploaded every minute and approximately 2.7 billion monthly active users, YouTube has become one of the largest platforms for audiovisual communication (Georgakopoulou, 2019). For many non-English-speaking viewers, YouTube's automatic subtitle feature, which relies on Automatic Speech Recognition (ASR) and Neural Machine Translation (NMT), provides the main means of accessing English-language content. As a result, the quality of these automatic subtitles plays an important role in shaping how audiences understand and interpret online media.

Comedy, particularly stand-up comedy, is widely recognized as one of the most difficult genres to translate. Humor often depends on linguistic features, cultural references, and contextual meanings that do not easily transfer across languages (Chiaro, 2010; Vandaele, 2010). The challenge becomes even greater in satirical humor, which commonly relies on irony, exaggeration, incongruity, and social criticism. In such cases, a translation may successfully convey the literal meaning of an utterance while failing to preserve the humorous effect that makes it meaningful to the audience.

These difficulties become more complex in audiovisual contexts. Unlike written texts, audiovisual content combines spoken language with visual and auditory elements that contribute to meaning-making (Díaz Cintas & Remael, 2021; Pérez-González, 2014). In comedy performances, humor is often created through timing, intonation, facial expressions, gestures, and audience reactions. Such elements are not merely supportive features but essential parts of the joke itself. Consequently, subtitle systems that focus only on spoken language may capture the verbal message while overlooking other elements that contribute to the humorous effect.

Armando Anto provides an interesting case for examining this issue. As an Iranian-French comedian performing in English for American audiences, he frequently addresses topics related to cultural identity, racism, politics, media representation, and everyday life. His performances are particularly distinctive because he incorporates the violin as part of his comedic style. Rather than serving as background music, the instrument functions as a tool for delivering satire, emphasizing punchlines, and expressing meanings that are not conveyed through words alone. As a result, much of the humor in his performances is constructed through the interaction of verbal and non-verbal elements.

Although previous studies have examined automatic subtitling (Sandrelli, 2021; González-Iglesias & Toda Iglesia, 2023), humor translation in audiovisual texts (Chiaro, 2008; Dore, 2019; Sepielak & Matamala, 2022), and multimodal communication (Kress & van Leeuwen, 2001; Jewitt, 2014; Taylor, 2016), limited attention has been given to the translation of satirical humor in automatically generated subtitles, particularly when humor is distributed across multiple semiotic modes. This study addresses that issue by investigating the translation of satirical humor in YouTube's automatic Indonesian subtitles of Armando Anto's stand-up comedy performance.

The study is guided by four research questions: (1) What forms of satirical humor appear in Armando Anto's comedy performance? (2) What semantic and pragmatic shifts occur in YouTube's automatic Indonesian subtitles? (3) To what extent is the satirical humor preserved or lost in the translation process? and (4) What role do multimodal elements, including violin performance, facial expressions, and gestures, play in constructing humor, and how is this role represented in the automatic subtitles?

2. Literature Review

2.1 Audiovisual Translation and the Subtitle Constraint

Audiovisual translation (AVT) refers to the transfer of multimodal texts that combine language, images, sound, music, and gestures from one language into another (Gambier, 2018; Pérez-González, 2019). Among the various forms of AVT, subtitling is one of the most widely used. It involves presenting spoken dialogue as written text on the screen, either within the same language (intralingual subtitling) or across different languages (interlingual subtitling). Unlike literary translation, subtitling is subject to limitations of space, timing, and synchronization, which often require condensation and adaptation of the original message (Gottlieb, 1994; Díaz Cintas & Remael, 2021). For this reason, Gottlieb (1994) describes subtitling as a form of "diagonal translation" that involves a shift from spoken language to written text.

The increasing use of automatic subtitling has introduced additional challenges to the translation process. Platforms such as YouTube rely on Automatic Speech Recognition (ASR) to transcribe speech and Neural Machine Translation (NMT) to generate subtitles in other languages. Although these technologies have improved considerably in recent years (Sandrelli, 2021), they still face limitations. ASR systems may produce transcription errors when dealing with accented, informal, or fast speech, while NMT systems often struggle to capture pragmatic meanings such as irony, implicature, and register (González-Iglesias & Toda Iglesia, 2023). As a result, automatically generated subtitles may reproduce the literal content of an utterance while failing to convey its intended communicative effect.

2.2 Humor Translation: Theory and Challenges

Humor translation is widely regarded as one of the most challenging areas of translation studies (Chiaro, 2010; Vandaele, 2010). Humor is often shaped by linguistic creativity, cultural knowledge, and contextual interpretation, making it difficult to transfer across languages. Several theories of humor, including Incongruity Theory, Superiority Theory, and Relief Theory, emphasize that humor depends not only on what is said but also on how audiences interpret a situation and recognize unexpected meanings (Vandaele, 2010). Consequently, a successful translation must convey both the semantic content and the humorous effect of the original text.

Chiaro (2008, 2010) distinguishes between language-dependent humor and culture-dependent humor. Language-dependent humor includes puns, wordplay, and phonological jokes, whereas culture-dependent humor relies on shared knowledge, stereotypes, and social references. Satirical humor often combines both forms. It draws on cultural and political contexts while also relying on linguistic and pragmatic cues to communicate criticism through irony and exaggeration. These characteristics create difficulties for automatic subtitle systems, which generally process linguistic patterns but lack the contextual awareness needed to interpret irony or cultural implications.

In audiovisual texts, humor is frequently created through the interaction of verbal and non-verbal elements. Dore (2019) shows that humor in audiovisual media can emerge from the relationship between dialogue, images, music, gestures, and vocal performance. Similarly, Sepielak and Matamala (2022) classify humor in audiovisual texts as monomodal, complementary, or contradictory, depending on how different semiotic modes contribute to the humorous effect. This framework is particularly

relevant to Armando Anto's performances, where humor is often constructed through the interaction between spoken language and violin performance.

2.3 Multimodality and Audiovisual Discourse

Multimodal discourse analysis views communication as a process that involves multiple semiotic resources working together to create meaning. According to Kress and van Leeuwen (2001), meaning is distributed across different modes, including language, images, sound, gesture, music, gaze, and spatial organization. Jewitt (2014) further argues that understanding contemporary communication requires attention to the interaction among these modes rather than focusing exclusively on language. This perspective is particularly relevant to audiovisual comedy, where humorous meaning is often distributed across verbal and non-verbal channels.

Applying multimodal theory to audiovisual translation, Taylor (2016) argues that subtitles should be understood as translations of the entire audiovisual text rather than translations of dialogue alone. From this perspective, an effective subtitle should preserve not only the semantic content of speech but also the relationships between verbal and non-verbal elements that contribute to meaning. Martinec and Salway (2005) likewise highlight the importance of examining image–text relationships, while Antonini (2020) emphasizes the continuing gap between the multimodal richness of audiovisual content and subtitle systems that primarily focus on spoken language. These perspectives provide a useful framework for examining how satirical humor is represented in YouTube's automatic subtitles.

3. Methodology

3.1 Research Design

This study adopts a qualitative descriptive design within the framework of Descriptive Translation Studies (Toury, 2012). The study is product-oriented, focusing on the translation outcomes produced by YouTube's automatic subtitle system rather than on the technical processes underlying subtitle generation. A qualitative approach is appropriate because satirical humor is highly dependent on context, irony, and pragmatic meaning, all of which require close interpretation beyond lexical comparison (Chiaro, 2010; Vandaele, 2010). The study also incorporates a multimodal perspective to examine how verbal and non-verbal elements contribute to humor construction in audiovisual performance (Kress & van Leeuwen, 2001; Jewitt, 2014).

3.2 Data Source

The data were drawn from a stand-up comedy performance by Armando Anto published on YouTube (Armando Anto, 2024). The approximately eight-minute performance is delivered in English before a live American audience and includes both spoken comedy and violin-based musical humor. Anto frequently addresses issues related to cultural identity, racism, politics, and everyday life through satirical commentary.

The source text consists of the comedian's spoken English performance, while the target text consists of the Indonesian subtitles automatically generated by YouTube. Unlike professionally produced subtitles, YouTube's subtitles are generated through a combination of Automatic Speech Recognition (ASR) and Neural Machine Translation (NMT), making them a suitable source for examining the strengths and limitations of automatic subtitle translation in humorous audiovisual content.

The unit of analysis is the satirical humor segment. A segment was included when it met three criteria: (1) the presence of irony, exaggeration, or incongruity; (2) a recognizable satirical target, such as race, nationality, politics, or domestic life; and (3) audience response, including laughter or applause, indicating successful humor delivery in the original performance. Applying these criteria resulted in a corpus of eleven satirical segments.

3.3 Analytical Framework

Three analytical frameworks were applied to the corpus. First, humor was classified using Chiaro's (2008, 2010) distinction between language-dependent and culture-dependent humor, together with Vandaele's (2010) incongruity-based perspective on humor. These frameworks were used to identify the mechanisms through which satirical effects were created in the source text.

Second, translation shifts were examined within Toury's (2012) descriptive translation framework through semantic and pragmatic comparison of the source text and the automatically generated subtitles. Particular attention was given to changes affecting irony, cultural references, and humorous intent.

Third, multimodal analysis was conducted using the semiotic framework proposed by Kress and van Leeuwen (2001) and later developments in multimodal discourse studies (Jewitt, 2014). Sepielak and Matamala's (2022) taxonomy of multimodal humor served as a reference for identifying the contribution of verbal and non-verbal elements to humor construction. This framework enabled the analysis of how violin performance, facial expressions, gestures, and audience reactions interacted with spoken language to create satirical meaning.

Humor preservation was evaluated at three levels: preserved, partially preserved, and not preserved. A segment was categorized as preserved when both the propositional meaning and humorous effect remained accessible in the subtitle. Partial preservation was assigned when the basic meaning was retained but aspects of irony, timing, cultural references, or multimodal support were weakened. A segment was categorized as not preserved when the humorous effect was substantially reduced or lost altogether.

3.4 Analytical Procedures

The analysis was conducted in four stages. First, the performance was transcribed verbatim and aligned with the Indonesian automatic subtitles generated by YouTube. Second, all potential instances of satirical humor were identified and evaluated according to the inclusion criteria. Third, each segment was coded for humor type, translation shift, humor preservation level, and multimodal contribution. Finally, a qualitative close-reading analysis was conducted to examine representative examples and illustrate how automatic subtitle translation affected the preservation of satirical humor across the corpus.

4. Result and Discussion

Table 1. Comparative Analysis of Satirical Humor: Source Text vs. YouTube Automatic Indonesian Subtitles

No.	Source Text (English)	Auto Subtitle (Indonesian)	Humor Type	Meaning Shift	Humor Preserved?
1	"they accepted me as a Mexican"	"mereka menerima sebagai seorang Meksiko" [+ Mexican music on violin]	Racial satire / self-deprecation	Grammatical omission ('me' deleted); multimodal punchline (violin Mexican melody) absent in subtitle	Partially — only with visual support
2	"racism in France... we're better at it"	"lebih baik dalam hal itu"	Irony / political satire	The phrase 'better at racism' loses its ironic register; 'lebih baik' is neutral, not ironic in Indonesian context	Partially — irony weakened
3	"they actually know where the country is"	"mereka sebenarnya tahu di mana negara itu berada"	Geopolitical satire	Acceptable translation; however, the word 'actually' (implying contrast/surprise) is rendered as 'sebenarnya', which is semantically accurate	Yes
4	"Where is America? America is everywhere — Afghanistan, Iraq..."	"di mana Amerika... Amerika ada di mana-mana Afghanistan Irak"	Political satire / anti-imperialism	Core irony preserved; punctuation and pauses lost — run-on subtitle reduces comedic timing of the punchline	Partially — timing lost
5	"I will not use the f word — I will use the F chord" [plays violin aggressively]	"saya tidak akan menggunakan kata f... saya akan menggunakan akor F"	Wordplay / phonological humor	Homophonic wordplay ('word' vs. 'chord') is coincidentally preserved in Indonesian ('kata' vs 'akor'); violin gesture absent in subtitle	Partially — wordplay preserved, gesture not

No.	Source Text (English)	Auto Subtitle (Indonesian)	Humor Type	Meaning Shift	Humor Preserved?
6	"I also have a sound for the C word... like my ex, she was a real [violin screech]"	"saya juga memiliki suara untuk kata se... seperti mantan saya dia adalah seorang yang nyata"	Sexual/crude wordplay	'SE word' is opaque; 'seorang yang nyata' ('a real [person]') loses the expletive implication entirely; violin screech absent	No — humor lost
7	"my parents come from another... romantic country... Iran"	"orang tua saya berasal dari negara romantis lainnya uh Iran"	Irony / political satire	The strategic pause ('uh') and ironic register around 'romantic' are flattened; audience's 'Hey!' reaction not captioned	Partially — irony weakened
8	"I call this the American Iran Concerto" [plays contrasting musical phrases]	"saya menyebut ini konserto Amerika Iran" [no musical description in subtitle]	Satirical musical performance	The violent contrast between peaceful American melody and chaotic Iranian theme — core humor vehicle — is entirely absent from subtitle	No — multimodal humor lost
9	"If you want to download this music, it's on CNN"	"jika Anda ingin mengunduh musik ini ada di CNN kemudian"	Media satire	'Kemudian' (then/next) is an ASR error, corrupting the punchline; the joke targets CNN's conflict-heavy coverage	No — ASR error destroys joke
10	"The lying chord... do you remember when I told you, you were the most important thing in the world?" [plays sour chord]	"akor berbohong... apakah kamu ingat ketika saya memberi tahu kamu bahwa kamu adalah yang paling penting di dunia"	Family satire / dark humor	Translation adequate but ASR misses the violin 'lying chord' cue that signals joke; no [music] notation	Partially — verbal OK, sonic cue absent
11	"And then my wife came — she said: do you remember when I told you he was your son?" [shocked violin chord]	"dan kemudian istri saya datang dia berkata apakah kamu ingat ketika saya memberi tahu kamu bahwa dia adalah anakmu"	Dark irony / plot-twist satire	Plot-twist punchline verbally intact, but the shocked violin chord and Anto's expression (core of the comedic climax) are invisible	Partially — verbal intact, multimodal climax lost

4.1 Overview of Findings

Analysis of the eleven satirical segments reveals a consistent pattern in the performance of YouTube's automatic Indonesian subtitles. Although the system generally succeeds in transferring the literal meaning of spoken utterances, it performs less effectively when humor relies on irony, cultural references, multimodal interaction, or precise comedic timing. Of the eleven segments analyzed, only one segment (Segment 3) can be categorized as fully preserved. In this case, both the propositional meaning and the humorous effect remain accessible in the target language, allowing Indonesian viewers to understand not only the content of the joke but also its intended comedic impact. This finding suggests that automatic subtitle systems are capable of preserving humor when the joke depends primarily on explicit verbal information and does not require extensive contextual or multimodal interpretation.

The majority of the corpus falls into the category of partial preservation. Six of the eleven segments successfully transfer the main verbal message, enabling viewers to follow the topic and general meaning of the joke. However, important elements

contributing to the humorous effect are weakened or omitted during the translation process. In most cases, the loss occurs because humor is distributed across multiple semiotic modes, including violin performance, facial expressions, gestures, and audience reactions. While the subtitles provide access to the linguistic content of the performance, they often fail to represent the non-verbal elements that help construct the joke. Consequently, viewers may understand what is being discussed without fully experiencing the humorous effect intended by the comedian. This pattern supports the argument of Kress and van Leeuwen (2001) that meaning in multimodal communication is distributed across different modes and cannot always be recovered through language alone.

Three segments are categorized as not preserved because the humorous effect is substantially reduced or completely lost in translation. In some cases, errors generated by the Automatic Speech Recognition (ASR) system interfere with the coherence of the subtitle and obscure the intended punchline. In other cases, humor is conveyed primarily through musical performance rather than spoken language, leaving the subtitle system with little material to translate. Segment 2 presents a slightly different challenge. Although the translation accurately reproduces the literal meaning of the source utterance, it fails to preserve the ironic stance that gives the joke its satirical force. As a result, a critical or humorous statement appears neutral in the target language. Taken together, these findings indicate that YouTube's automatic subtitles function more effectively as lexical representations of speech than as complete translations of comedy performances. The system performs reasonably well at transferring semantic information, but it remains limited in its ability to reproduce irony, multimodal meaning, and the broader communicative effect of satirical humor.

4.2 Racial and Self-Deprecating Satire

The performance begins with a joke about identity and racial perception. Anto explains that French people never fully accepted him because of his appearance, whereas people in Texas accepted him, but as Mexican. The statement creates an immediate contrast between how he identifies himself and how others perceive him. Through this contrast, Anto introduces a form of self-deprecating satire that invites the audience to laugh at his experience while simultaneously questioning the assumptions behind such categorizations. Rather than presenting a direct criticism, he uses humor to expose the tendency to associate identity with physical appearance rather than with cultural or national background.

The joke operates on more than one level. On the surface, the audience is invited to laugh at the absurdity of a French-Iranian comedian being mistaken for Mexican. However, beneath the humorous surface lies a commentary on racial stereotyping and simplified perceptions of identity. The joke suggests that people often rely on visual assumptions when identifying others, regardless of their actual background. This incongruity between appearance and identity becomes the primary source of humor in the segment. The audience's laughter indicates that the satirical message is successfully communicated and recognized within the performance context.

The Indonesian subtitle transfers the verbal content of the joke with relatively minor loss. Although the object pronoun *me* is omitted, the overall meaning of the statement remains understandable for Indonesian viewers. Readers can still follow the narrative and recognize the contrast between Anto's French identity and the way he is perceived by others. From a semantic perspective, the translation can therefore be considered largely successful. The main information required to understand the situation is preserved in the subtitle.

However, the subtitle does not represent the Mexican-style violin melody that immediately follows the statement. This omission is significant because the melody functions as the actual punchline and provides the final humorous cue for the audience. The music reinforces the stereotype referenced in the joke and signals the moment at which laughter is expected. Viewers who can hear the performance receive both the verbal setup and the musical payoff. By contrast, viewers who rely primarily on subtitles encounter only the setup, which makes the humorous sequence feel incomplete.

This finding supports Sepielak and Matamala's (2022) concept of complementary multimodal humor. According to this perspective, humor is sometimes created through the interaction of multiple semiotic modes rather than through language alone. In this segment, the spoken statement establishes the context, while the violin melody completes the joke and strengthens its satirical effect. When the musical component is removed, part of the humorous meaning becomes inaccessible. As a result, the subtitle preserves the basic message of the joke but only partially preserves its humorous effect.

4.3 The Problem of Irony in Automatic Translation

Segments 2 and 7 reveal a limitation that extends beyond the issue of multimodality. In these segments, the primary challenge lies in the ability of the automatic translation system to recognize and preserve irony. Unlike lexical meaning, irony is not communicated solely through words. It often depends on tone of voice, contextual knowledge, shared cultural assumptions, and the audience's ability to recognize a contrast between what is said and what is actually meant. Consequently, even a translation that is linguistically accurate may fail to convey the intended humorous effect if the ironic meaning is not preserved.

This problem is clearly illustrated in Segment 2. Anto states, "racism in France—we're better at it," a remark that functions as a satirical criticism rather than a genuine expression of pride. The humor depends on the audience recognizing that the statement is intentionally inverted. Rather than praising racism, Anto uses irony to criticize it by pretending to celebrate something that is socially unacceptable. In the original performance, this interpretation is supported by his tone of voice, the comedic setting, and the audience's understanding that stand-up comedians often employ irony to comment on social issues. The laughter that follows further confirms that the audience interprets the statement as satire rather than as a sincere opinion.

The Indonesian subtitle translates the expression as *lebih baik dalam hal itu* ("better at it"). From a semantic perspective, the translation is accurate because it reproduces the literal meaning of the original statement. However, the subtitle does not contain any linguistic indication that the statement is intended ironically. Viewers who rely primarily on the subtitle may interpret the phrase as a straightforward comparison rather than as a satirical criticism. As a result, the critical stance embedded in the joke becomes less visible. This finding supports Mukherjee's (2021) observation that neural machine translation systems often struggle to preserve irony because irony is constructed through pragmatic inference rather than through vocabulary or grammatical structure alone.

A similar pattern appears in Segment 7. In this segment, Anto introduces Iran by describing it as "another… romantic country." The humor emerges from the contrast between the adjective *romantic* and the audience's likely perception of Iran. The deliberate pause before naming the country creates anticipation and encourages the audience to interpret the statement ironically. Rather than presenting Iran as genuinely romantic, Anto draws attention to the gap between the positive connotations of the word and dominant media representations of the country. The joke therefore relies on incongruity, a mechanism that Vandaele (2010) identifies as central to many forms of humor.

The Indonesian subtitle translates the word *romantic* accurately. Nevertheless, the translation alone cannot reproduce the hesitation, vocal delivery, and cultural assumptions that make the statement humorous in the original performance. Without access to these contextual cues, viewers may simply read the adjective as a literal description. Although the lexical meaning is preserved, the ironic meaning becomes difficult to reconstruct. The segment demonstrates that successful humor translation requires more than semantic equivalence; it also requires the preservation of pragmatic meaning. When irony is translated literally without contextual support, much of its satirical force is weakened or lost.

4.4 Wordplay: A Rare Success and a Clear Failure

Segments 5 and 6 provide contrasting examples of how automatic subtitle translation handles wordplay. Unlike irony, which depends heavily on pragmatic interpretation, wordplay often relies on linguistic form, including sound patterns, ambiguity, and phonological similarity. Such features are widely recognized as difficult to translate because they are closely tied to the structure of a particular language (Chiaro, 2010). The two segments demonstrate that while wordplay occasionally survives automatic translation, successful preservation is largely accidental rather than systematic.

Segment 5 represents the most successful example of humor preservation in the corpus. Anto states that he will not use the "F word" but will instead use the "F chord." The joke depends on the audience recognizing the relationship between the prohibited swear word and the musical alternative provided by the comedian. The humorous effect is strengthened by the phonological similarity between *word* and *chord* and by the unexpected substitution of a musical term for an expletive. Through this substitution, Anto transforms potentially offensive language into a harmless musical reference while maintaining the comedic effect.

The Indonesian subtitle translates the expressions as *kata F* ("F word") and *akor F* ("F chord"). Although the phonological similarity present in English does not exist in Indonesian, the structural relationship between the two expressions remains recognizable. Viewers can still understand that one expression is being substituted for another and that the humor emerges from this substitution. As a result, the basic joke remains accessible despite the loss of the original sound pattern. The preservation of humor in this segment appears to result from the compatibility between the source and target expressions rather than from any deliberate strategy employed by the translation system.

Segment 6 demonstrates the opposite outcome. Here, Anto constructs a joke around the expression "C word," which requires the audience to understand the euphemistic abbreviation and connect it with a high-pitched violin sound that functions as a substitute for the offensive term. The humor depends on shared cultural knowledge, auditory cues, and the audience's ability to infer the omitted word. In the original performance, these elements work together to create a clear punchline and generate audience laughter.

The automatic subtitle, however, renders the expression as *kata se*, a phrase that carries little meaning for Indonesian viewers because there is no comparable abbreviation convention in Indonesian. The subsequent translation, *dia adalah seorang yang nyata* ("she was a real person"), further weakens the joke by removing the implied insult that motivates the humor. Without knowledge of the original expression and without access to the violin cue that substitutes for the omitted word, viewers receive a

statement that appears unrelated to the intended punchline. Consequently, both the linguistic and multimodal dimensions of the joke are lost. This segment illustrates how automatic subtitle systems can struggle when humor depends on language-specific conventions and implicit cultural knowledge that cannot be transferred through literal translation alone.

4.5 The American-Iran Concerto: When the Humor Is Entirely Musical

Segment 8 represents one of the most significant findings in this study because it exposes a limitation of automatic subtitle systems that cannot be explained solely through translation errors. The segment revolves around Anto's "American-Iran Concerto," a short musical performance introduced through a brief verbal announcement. After introducing the piece, Anto performs two contrasting musical styles: a calm, melodic sequence representing America and a sudden shift to chaotic, rapid, and dissonant violin sounds representing Iran. He alternates between these musical patterns several times while maintaining a deliberately neutral facial expression. The audience responds with laughter because the contrast between the two musical representations creates an immediately recognizable satirical message.

Unlike most of the other segments in the corpus, the humor in this sequence is not primarily carried through spoken language. The verbal statement, "I call this the American Iran Concerto," functions only as a setup that prepares the audience for the performance. The actual humor emerges through the contrast between the musical passages and the cultural assumptions attached to them. Through this contrast, Anto satirizes simplified representations of international relations and, more specifically, Western media portrayals of Iran. The joke relies on incongruity, as the audience is invited to recognize the exaggerated difference between the two musical styles and connect it to broader political narratives. As Vandaele (2010) suggests, incongruity often functions as a central mechanism of humor, and this segment provides a particularly clear example of that process.

The automatic subtitle system successfully transcribes the introductory statement but does not represent any of the musical content that follows. As a result, viewers who depend primarily on subtitles receive only the setup while missing the humorous element itself. They can understand that a piece of music is being introduced, but they receive no information about how the music develops, what it represents, or why it generates laughter from the audience. Consequently, the subtitle provides access to the existence of the joke without providing access to the joke's meaning. This creates a substantial gap between the experience of viewers who can hear the performance and those who rely mainly on the translated text.

From a multimodal perspective, this finding highlights the limitations of subtitle systems that focus exclusively on speech. According to Kress and van Leeuwen (2001), meaning in multimodal communication is distributed across multiple semiotic resources rather than being contained in language alone. In this segment, the musical mode carries most of the communicative burden, while speech plays only a supporting role. Because the subtitle system is designed to process verbal language, it cannot adequately represent the semiotic resources through which the satirical meaning is constructed. The result is not merely a reduction in humorous effect but a substantial loss of meaning.

Taylor (2016) argues that audiovisual translation should be understood as the translation of a multimodal text rather than the translation of dialogue in isolation. Viewed from this perspective, the subtitle in Segment 8 captures only one component of the communicative event. The verbal introduction is preserved, but the musical contrast that generates the humor remains inaccessible within the subtitle. This finding suggests that automatic subtitle systems are particularly limited when humor depends heavily on music, gesture, or other non-verbal modes. The American-Iran Concerto therefore serves as strong evidence that current automatic subtitle technologies cannot fully represent multimodal comedy performances and should be viewed as partial rather than complete translations of such content.

4.6 ASR Error and the Destruction of a Punchline

Another important finding concerns the impact of Automatic Speech Recognition (ASR) errors on humor preservation. Although previous sections have demonstrated the importance of multimodality and irony, the data also show that technical transcription errors can significantly affect the delivery of a joke. This problem becomes especially visible in Segment 9, where a relatively simple satirical punchline is weakened by a small but consequential subtitle error. The segment demonstrates that even when humor is primarily verbal, successful translation still depends on accurate speech recognition. Without an accurate transcription, the translation process begins from an already distorted source.

In Segment 9, Anto concludes the American-Iran Concerto sequence by stating, "If you want to download this music, it's on CNN." The joke functions as a form of media satire. By suggesting that the chaotic musical representation of Iran can be found on CNN, Anto humorously implies that media coverage of US-Iran relations resembles the exaggerated disorder represented in the performance. The joke is concise, culturally recognizable, and delivered immediately after the musical sequence, allowing the audience to connect the punchline with the preceding performance. The laughter that follows indicates that the audience successfully interprets the satirical intent.

The Indonesian automatic subtitle reproduces the statement but adds the word *kemudian* ("then" or "next") at the end of the sentence. Although this addition appears minor, it changes the rhythm and clarity of the punchline. Instead of ending with a complete and self-contained joke, the subtitle concludes with a conjunction that implies continuation. As a result, the sentence appears unfinished and may prompt viewers to expect additional information. The humor therefore loses some of its immediacy and impact because the punchline no longer arrives with the same degree of precision.

This finding illustrates how ASR errors can interfere with humor even when the translation itself is relatively accurate. In comedy, timing is often as important as content. A joke may depend on a specific pause, a sudden conclusion, or a carefully positioned punchline. When an ASR system inserts unnecessary words or misidentifies speech, the structure of the joke can be disrupted. In this case, the added word does not completely obscure the meaning of the joke, but it reduces the effectiveness of its delivery. The audience reading the subtitle receives a less coherent version of the original utterance.

The result is consistent with Sandrelli's (2021) observations regarding the limitations of automatic subtitling systems in live-performance contexts. Stand-up comedy frequently includes audience laughter, applause, overlapping sounds, and rapid shifts in vocal delivery, all of which increase the likelihood of transcription errors. Because many ASR models are trained primarily on relatively controlled speech environments, they may struggle to process the acoustic complexity of live performances. Segment 9 demonstrates how such errors can directly affect humor preservation. Even when the satirical message remains recognizable, inaccuracies in transcription can weaken the punchline and reduce the overall comedic effect experienced by subtitle users.

4.7 Dark Family Humor and the Lying Chord

The final section of Anto's performance centers on a recurring joke about dishonesty, introduced through what he calls the "lying chord." This chord consists of a deliberately unpleasant violin note that is played whenever a lie is revealed during the narrative. Throughout the sequence, Anto gradually escalates the humor by moving from relatively harmless lies to increasingly dramatic revelations. He first admits that he lied to his son by claiming that the child was the most important thing in his life. He then admits lying to his wife when he said that he was sorry. The sequence culminates in a final revelation from the wife, who suggests that their son may not actually be Anto's biological child. This progressive escalation creates anticipation and encourages the audience to expect each new confession to be more surprising than the previous one.

From a verbal perspective, this sequence represents one of the most successful examples of humor preservation in the corpus. The Indonesian subtitles reproduce the narrative progression with relatively high accuracy, allowing viewers to follow the development of the joke from beginning to end. The final plot twist remains understandable, and the surprise element that generates laughter is largely maintained in translation. Unlike several earlier segments that rely heavily on cultural references or wordplay, the humor here is based primarily on narrative escalation and unexpected revelation. As a result, the joke remains accessible to viewers even when translated automatically.

Despite this relative success, an important aspect of the humorous structure is absent from the subtitle. Each confession is accompanied by the lying chord, which functions as a recurring comedic signal throughout the sequence. The chord informs the audience that a lie has just been exposed and prepares them for the next stage of the joke. It also helps regulate the rhythm of audience laughter by marking the precise moment at which the humorous revelation occurs. Consequently, the chord performs a structural rather than decorative function within the performance.

The absence of this musical cue illustrates a recurring limitation identified throughout the study. Although the subtitles preserve the verbal content of the sequence, they do not represent the mechanism that organizes the joke and controls its timing. Hu (2016) argues that in stand-up comedy, performative elements such as gesture, movement, and musical accompaniment are often integral parts of the joke itself rather than optional additions. The lying chord exemplifies this principle. It is not simply a sound effect inserted for entertainment; it acts as a recurring punchline device that shapes audience expectations and guides interpretation. When the chord is removed from the translated representation, viewers receive the narrative content of the joke but not its complete comedic structure.

The findings from this segment further demonstrate the distinction between understanding a joke and experiencing a joke. Subtitle users can understand what the characters are saying and why the final revelation is surprising. However, they do not experience the same rhythmic pattern that structures the audience's response in the original performance. The sequence therefore remains humorous in translation, but its comedic effect is reduced because one of its central multimodal components is unavailable in the subtitle text.

4.8 Multimodal Loss as a Systemic Problem

Across all eleven segments analyzed in this study, the most consistent pattern is not mistranslation, lexical inaccuracy, or even ASR error. Instead, the dominant issue is multimodal loss, namely the systematic absence of information conveyed through music, gesture, facial expression, and other non-verbal resources. This pattern appears repeatedly throughout the corpus regardless of the type of humor being used. Whether the joke involves satire, irony, wordplay, or narrative surprise, important

aspects of meaning are frequently communicated through modes that automatic subtitle systems are unable to represent. Consequently, the loss observed in the data is not primarily the result of individual translation mistakes but of broader technological limitations.

The current ASR–NMT pipeline is designed to process spoken language. It converts speech into text and then translates that text into another language. While this process can effectively transfer verbal information, it does not analyze facial expressions, annotate gestures, or interpret musical performance. As a result, any meaning conveyed through these channels remains outside the scope of the system. The subtitles therefore provide a representation of speech rather than a representation of the complete audiovisual event. This distinction becomes particularly important in performances where humor depends heavily on the interaction between verbal and non-verbal modes.

Armando Anto's comedy illustrates this issue especially clearly because the violin functions as an extension of his verbal performance. Throughout the routine, musical phrases repeatedly complete, reinforce, or even replace spoken punchlines. In several instances, the joke remains unfinished until a violin chord is played. The instrument therefore serves not merely as background accompaniment but as a communicative resource that carries meaning in its own right. When the subtitle omits these musical elements, a substantial portion of the humorous message disappears. The audience receives the words but not the full communicative event through which the humor is constructed.

This observation supports the multimodal perspective proposed by Kress and van Leeuwen (2001), who argue that meaning is distributed across different semiotic modes and cannot always be recovered from a single channel. The findings of the present study provide concrete evidence for this claim. Several jokes in Anto's performance depend on the interaction between speech and violin performance, while others rely on facial expression, timing, or gesture. When only the verbal channel is translated, the relationship between these modes becomes invisible. Consequently, viewers relying on subtitles encounter a reduced version of the performance rather than the complete communicative experience available to the original audience.

One possible response to this limitation can be found in the practice of Subtitling for the Deaf and Hard of Hearing (SDH). Unlike conventional subtitles, SDH subtitles often include descriptions of relevant non-speech sounds such as music, laughter, applause, or environmental noise. Annotations such as "[violin plays]" or "[audience laughs]" provide viewers with information that would otherwise be inaccessible through text alone. Although such annotations cannot fully reproduce the humor of a musical performance, they can at least signal the presence of meaningful non-verbal elements. Previous studies have suggested that integrating this type of annotation may improve the accessibility and communicative completeness of audiovisual translation (Antonini, 2020; Orero & Tor-Carroggio, 2021).

Nevertheless, current automatic subtitle systems generally do not incorporate SDH-style descriptions of musical or gestural information. Until such capabilities become standard, automatic subtitles will remain fundamentally speech-centered technologies applied to inherently multimodal forms of communication. The findings of this study therefore suggest that future developments in automatic subtitling should move beyond speech recognition and machine translation alone. Greater attention to multimodal meaning may be necessary if automatic subtitles are expected to function not only as transcripts of speech but also as effective translations of audiovisual humor.

5. Conclusion

This study examined the translation of satirical humor in YouTube's automatically generated Indonesian subtitles using Armando Anto's stand-up comedy performance as a case study. The analysis focused on eleven satirical segments and investigated the types of humor employed, the translation shifts introduced by the automatic subtitle system, the extent to which humor was preserved, and the role of multimodal elements in constructing comedic meaning. The findings demonstrate that YouTube's automatic subtitles are generally effective in transferring the literal meaning of spoken utterances but considerably less successful in preserving the pragmatic and multimodal dimensions of humor.

The analysis identified several forms of satirical humor, including racial satire, self-deprecating humor, irony, wordplay, media satire, political satire, and dark family humor. While some verbal jokes remained understandable in Indonesian, many humorous effects were weakened or lost during translation. Irony proved particularly difficult for the system to preserve because its interpretation depends on contextual and pragmatic cues rather than on lexical meaning alone. Similarly, wordplay produced mixed results: one joke survived largely by coincidence, whereas another became incomprehensible after translation. These findings suggest that semantic accuracy does not necessarily guarantee the preservation of humorous meaning.

The study also revealed that multimodality constitutes the greatest challenge for automatic subtitle systems. Throughout the performance, humor was frequently constructed through the interaction of speech, violin performance, facial expression, gesture, and audience response. In several segments, the violin functioned as an essential component of the punchline rather than as background accompaniment. However, these non-verbal elements were consistently absent from the automatically generated subtitles. As a result, viewers relying on subtitles often received only a partial representation of the original performance. The

findings support multimodal theories of communication that view meaning as distributed across multiple semiotic modes rather than contained in language alone (Kress & van Leeuwen, 2001).

Overall, the study concludes that YouTube's automatic Indonesian subtitles function more effectively as speech transcripts than as complete audiovisual translations of comedy performances. Although advances in ASR and NMT technology have improved subtitle accessibility, current systems remain limited in their ability to represent irony, cultural nuance, and multimodal meaning. Future developments in automatic subtitling may benefit from incorporating multimodal annotation practices, particularly for performances in which humor depends heavily on music, gesture, and other non-verbal resources. Such improvements would contribute to a more comprehensive and culturally sensitive translation of audiovisual humor across languages and audiences.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

Acknowledgments: The author would like to express heartfelt gratitude to her family for their unconditional support, patience, and encouragement throughout this research. Sincere thanks are also extended to Dr. Majedah Alaiyed for her continued support during the preparation of this manuscript.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Antonini, R. (2020). The multimodal richness of audiovisual texts and the subtitle constraint. Routledge.
[2] Armando Anto. (2024). Armando Anto live stand-up comedy performance [Video]. YouTube. https://www.youtube.com/
[3] Chiaro, D. (2008). Verbally expressed humor and translation. In V. Raskin (Ed.), The primer of humor research (pp. 569–608). Mouton de Gruyter.
[4] Chiaro, D. (2010). Translation and humour. Continuum.
[5] Díaz Cintas, J., & Remael, A. (2021). Subtitling: Concepts and practices (2nd ed.). Routledge.
[6] Dore, M. (2019). Humour in audiovisual translation: Theories and applications. Routledge. https://doi.org/10.4324/9781003001928
[7] Gambier, Y. (2018). Translation studies: Reshaping the discipline. Target, 30(1), 19–33. https://doi.org/10.1075/target.30.1.02gam
[8] Georgakopoulou, P. (2019). Audiovisual translation and the digital age. In M. O'Hagan (Ed.), The Routledge handbook of translation and technology (pp. 105–121). Routledge. https://doi.org/10.4324/9781315311258-7
[9] González-Iglesias, J. D., & Toda Iglesia, F. (2023). Automatic subtitling and machine translation: Evaluating user comprehension in online platforms. Perspectives: Studies in Translation Theory and Practice, 31(2), 245–262. https://doi.org/10.1080/0907676X.2021.1997973
[10] Gottlieb, H. (1994). Subtitling: Diagonal translation. Perspectives: Studies in Translatology, 2(1), 101–121. https://doi.org/10.1080/0907676X.1994.9961225
[11] Hu, S. (2016). Performative elements and rhythmic structures in stand-up comedy translation. Journal of Humor Studies, 4(2), 89–104.
[12] Jewitt, C. (Ed.). (2014). The Routledge handbook of multimodal analysis (2nd ed.). Routledge.
[13] Kress, G., & van Leeuwen, T. (2001). Multimodal discourse: The modes and media of contemporary communication. Arnold.
[14] Martinec, R., & Salway, A. (2005). A system for image–text relations in new (and old) media. Visual Communication, 4(3), 337–371. https://doi.org/10.1177/1470357205055928
[15] Mukherjee, S. (2021). Irony and pragmatic failures in neural machine translation systems. Translation and Interpreting Studies, 16(3), 412–433. https://doi.org/10.1075/tis.20002.muk
[16] Orero, P., & Tor-Carroggio, I. (2021). User-centric approaches to Subtitling for the Deaf and Hard of Hearing (SDH) in automated environments. Translation Spaces, 10(1), 120–141. https://doi.org/10.1075/ts.21004.ore
[17] Pérez-González, L. (2014). Audiovisual translation: Theories, methods and issues. Routledge. https://doi.org/10.4324/9781315762975
[18] Pérez-González, L. (2019). Audiovisual translation. In M. Baker & G. Saldanha (Eds.), Routledge encyclopedia of translation studies (3rd ed., pp. 30–35). Routledge.
[19] Sandrelli, A. (2021). Automatic subtitling in live performances: Challenges of automatic speech recognition (ASR) in comedy. The Journal of Audiovisual Translation, 4(1), 62–85. https://doi.org/10.47476/jat.v4i1.2021.151
[20] Sepielak, K., & Matamala, A. (2022). Multimodal humor in audiovisual texts: A new taxonomy for translation studies. Meta: Translators' Journal, 67(1), 143–165. https://doi.org/10.7202/1092196ar
[21] Taylor, C. (2016). Multimodality and audiovisual translation. In Y. Gambier & L. van Doorslaer (Eds.), Border crossings: Translation studies and other disciplines (pp. 223–244). John Benjamins. https://doi.org/10.1075/btl.126.11tay
[22] Toury, G. (2012). Descriptive translation studies and beyond (Revised ed.). John Benjamins. https://doi.org/10.1075/btl.100
[23] Vandaele, J. (2010). Humor in translation. In Y. Gambier & L. van Doorslaer (Eds.), Handbook of translation studies (Vol. 1, pp. 147–152). John Benjamins. https://doi.org/10.1075/hts.1.hum1