ISSN:2828-6235
Impact factor:13.77

International Journal of Literature and Language Studies(IJLLS)

TRANSLATING BEYOND WORDS: PARALINGUISTIC CHALLENGES IN MACHINE TRANSLATION AND AUDIOVISUAL TRANSLATION

Umidaxon Aslamova

Bachelor Student, Faculty of English Philology and Translation Studies (Ingliz Filologiyasi va Tarjimashunoslik Fakulteti), Samarkand State Institute of Foreign Languages, Samarkand, Uzbekistan

Doi: https://doi.org/10.5281/zenodo.23423093

Abstract: Translation is not limited to transferring words and grammatical structures from one language into another, because communication also depends on voice, emotion, intonation, pauses, gestures, facial expressions, and other paralinguistic signals. This article examines the main difficulties that arise when such information is transferred through machine translation and audiovisual translation. Particular attention is given to stress, intonation, tone of voice, hesitation, laughter, sarcasm, gestures, and culturally specific visual signals. The study uses a qualitative comparative approach and analyzes representative examples in order to identify how communicative meaning may be weakened or lost during automated translation. The discussion shows that a translation can be lexically and grammatically correct while still failing to reproduce the speaker's emotional attitude or communicative intention. The article therefore argues that future translation technologies should treat paralinguistic information as an important part of meaning rather than as secondary information.

Keywords: paralanguage, machine translation, speech translation, prosody, intonation, sarcasm, audiovisual translation, subtitling, gestures, post-editing

Introduction:

Translation is often understood as the process of transferring meaning from a source language into a target language. In practice, however, meaning is rarely carried by words alone. When people communicate, they also use the tone of their voice, stress, pauses, rhythm, facial expressions, gestures, laughter, and other signals to express attitudes and emotions. The same sentence can therefore communicate different meanings depending on how it is spoken and in what situation it occurs. This becomes especially important in speech translation, interpreting, subtitling, dubbing, and other forms of audiovisual translation.

For example, the sentence "That was great" may express sincere approval, but it may also express disappointment or sarcasm if it is pronounced with a particular tone. The written words themselves do not explain the speaker's real intention. In face-to-face communication, listeners normally combine the words with the speaker's voice, facial expression, and situation. A translation system that receives only the written sentence, however, may not have access to all of this information. The development of machine translation has significantly changed the way translation is performed. Modern systems are able to produce fluent translations of increasingly complicated linguistic structures. Nevertheless, fluency does not always mean communicative accuracy. Recent research on speech-to-text translation has specifically shown that prosodic features such as stress, intonation, and rhythm can influence semantic interpretation, while current systems do not always use these signals strongly enough when producing translations (Tsiamas et al., 2024). Their study also introduced the ContraProSt benchmark for evaluating prosody-sensitive translation.

The problem becomes even more complicated in audiovisual translation, where spoken language exists together with facial expressions, gestures, music, sound effects, silence, and other

visual or auditory elements. A subtitler, for instance, has to decide what information can be transferred through a limited number of written characters and within a limited amount of screen time. Therefore, translation in such contexts is not simply a matter of replacing words with their equivalents. The purpose of this article is to examine how paralinguistic and nonverbal information influences translation and to identify the main difficulties that machine translation and audiovisual translation systems face in transferring such information. The article also considers how human translators can compensate for information that cannot be transferred directly.

Literature Review

The relationship between verbal and nonverbal communication has been discussed in translation studies for several decades. One of the important contributions to this area is the work of Fernando Poyatos. His edited volume *Nonverbal Communication and Translation* brought together research on nonverbal communication in literary translation, interpreting, film, television, and other areas. The book includes studies of paralanguage, gestures, theatrical communication, poetry, simultaneous interpreting, and dubbing, demonstrating that nonverbal information can form an important part of the translation process. Poyatos (1997) particularly emphasized the difficulties involved in transferring nonverbal communication in literary translation. His work shows that translation should not be restricted to verbal elements when nonverbal signals contribute to characterization, emotion, atmosphere, or interpersonal relationships. Christiane Nord also examined descriptions and transcriptions of paralanguage in literary translation, demonstrating that vocal and nonverbal elements may require specific translation decisions rather than simple lexical substitution. Audiovisual translation provides another important area of research. Frederic Chaume and Patrick Zabalbeascoa examined the nonverbal dimension of dubbing and demonstrated that audiovisual translation operates through several channels simultaneously. Dubbing and subtitling therefore require translators to consider information that comes from images, gestures, sound, and speech rather than from dialogue alone. These studies provide a useful theoretical basis for the present article. They demonstrate that translation quality should be evaluated not only through lexical and grammatical accuracy but also through the preservation of communicative intention.

Methodology:

This article applies a qualitative comparative method. The analysis is based on representative examples of spoken and audiovisual communication in which the same lexical material may receive different interpretations because of paralinguistic or nonverbal information. The examples focus on six main areas: intonation, stress, tone of voice, pauses and hesitation, laughter, and culturally dependent gestures.

The first example illustrates the importance of intonation:

Source: “You’re coming?”

Alternative interpretation: “You’re coming.”

Although the lexical items are almost identical, the first version can function as a question, while the second can function as a statement. In spoken communication, rising or falling intonation helps the listener understand the speaker’s communicative intention. If the original speech is converted into neutral written text, this acoustic distinction can become less visible. A translation system may therefore produce an output that is grammatically acceptable but pragmatically inappropriate. This example demonstrates that the translator needs to consider not only what was said but also how it was said.

Stress provides another important example:

Source: "I didn't say he stole the money."

This sentence can receive different interpretations depending on which word is stressed. If the speaker stresses "I," the implication may be that somebody else said it. If "he" is stressed, the speaker may suggest that another person committed the action. If "stole" is stressed, the speaker may be distinguishing stealing from another action. The words remain unchanged, but the communicative focus changes. Therefore, a speech translation system that does not adequately recognize stress may lose an important part of the speaker's intended meaning.

Tone of voice creates a similar problem:

Source: "That's perfect."

In a friendly context, the sentence can express genuine approval. In another situation, the same words may communicate irritation, disappointment, or sarcasm. A human listener can normally use the speaker's tone, facial expression, and context to interpret the utterance. A text-based machine translation system may not receive these signals. Consequently, the output may preserve the dictionary meaning of the sentence while losing its emotional and pragmatic meaning.

Pauses and hesitation also deserve attention:

Source: "Well… I'm not sure."

The pause after "well" can communicate hesitation, uncertainty, reluctance, or emotional discomfort. If the pause is removed and the sentence becomes "Well, I'm not sure," some of the original communicative effect may be weakened. In translation, punctuation, discourse markers, or changes in sentence structure can sometimes compensate for this loss. This is an area in which human translators can make contextual decisions that are difficult to reproduce through direct automatic substitution.

Results:

The analysis demonstrates several important findings. First, paralinguistic information can change the pragmatic meaning of a sentence without changing its lexical structure. Intonation, stress, and tone can determine whether an utterance is understood as a question, statement, criticism, compliment, or ironic remark. Second, information loss may occur before the actual translation stage. In a speech-to-text-to-translation pipeline, the original acoustic signal is first converted into written language. During this process, features such as pitch, duration, stress, and laughter may be weakened or omitted. The translation system then works with an incomplete representation of the original message. The final translation can consequently be grammatically correct but communicatively incomplete. Third, sarcasm is particularly difficult for automated systems. Consider a student arriving one hour late and a teacher saying, "Great. Just great." The literal meaning appears positive, but the situation and tone indicate the opposite. A human translator can recognize the contradiction between the words and the context. A machine system may interpret the sentence literally unless it has sufficient contextual and prosodic information.

Research supports these observations. Kano et al. demonstrated that selected acoustic properties such as duration and power can be computationally transferred between languages, showing that paralinguistic translation is technically possible. However, Tsiamas et al. found that although speech-translation systems can encode some prosodic information, that information does not consistently influence the translation output strongly enough. More recent PLaST research

indicates that explicitly separating linguistic and paralinguistic representations can improve performance on paralinguistic-sensitive translation tasks. The results therefore suggest that the main problem is not simply the absence of technology. Rather, the problem concerns how effectively a system can interpret the communicative function of paralinguistic information and reproduce it in another language.

Discussion:

The findings support a broader understanding of translation as an interpretive and communicative activity. When a translator works with written language, lexical and grammatical structures naturally receive considerable attention. However, when the source material contains speech, video, or emotionally expressive dialogue, the translator must also consider how the message is delivered. One important distinction is between lexical accuracy and communicative accuracy. A translation can contain the correct words and grammatical structures but still fail to communicate the speaker's intention. For example, translating "That's great" as an expression of approval may be completely correct at the lexical level, but if the speaker says it sarcastically, the communicative result can be misleading.

This issue also explains why human translators remain important even as machine translation becomes more advanced. Human translators can combine linguistic information with world knowledge, cultural knowledge, situational context, and pragmatic inference. They can also decide when an apparently small feature, such as a pause or change in tone, is important enough to reproduce in the target text. For translators, several practical strategies can be useful. One is **explicitation**, in which an implicit emotional or pragmatic meaning is made clearer in the target language. Another is **compensation**, where information lost in one part of the text is recreated through another linguistic feature. **Contextual adaptation** is also important when a literal translation would fail to reproduce the communicative effect. In audiovisual translation, punctuation, lexical choice, timing, and carefully selected expressions can sometimes compensate for information that cannot be reproduced directly.

Conclusion:

Paralinguistic information is an important but sometimes overlooked part of translation. Stress, intonation, rhythm, pauses, hesitation, laughter, tone of voice, facial expressions, gestures, and other signals can influence how a message is understood. As the examples discussed in this article demonstrate, identical words may express different intentions depending on their delivery and context. The study also shows that meaning can be lost before the translation itself begins, especially when spoken communication is first converted into text. In such cases, machine translation may operate on a reduced representation of the original message. This explains why a translation can be grammatically and lexically accurate while still failing to reproduce the speaker's emotional attitude or pragmatic intention.

For this reason, successful translation should be understood as more than the transfer of words between languages. The translator, whether human or machine, must consider the communicative event as a whole. In speech and audiovisual translation particularly, understanding how something is said can be just as important as understanding what is said. Future translation technologies will therefore need to become increasingly multimodal and context-sensitive, while human translators will continue to play an important role in interpreting the subtle relationship between language, voice, emotion, culture, and communicative intention.

References:

1. Birdwhistell, R. L. (1970). *Kinesics and Context: Essays on Body Motion Communication*. University of Pennsylvania Press.
2. Chaume, F. (2012). *Audiovisual Translation: Dubbing*. St. Jerome Publishing.
3. Díaz Cintas, J., & Remael, A. (2007). *Audiovisual Translation: Subtitling*. St. Jerome Publishing.
4. Hatim, B., & Mason, I. (1997). *The Translator as Communicator*. Routledge.
5. Kano, T., Sakti, S., Takamichi, S., Neubig, G., Toda, T., & Nakamura, S. (2012). A method for translation of paralinguistic information. *Proceedings of the 9th International Workshop on Spoken Language Translation*, 158–163.
6. Li, Y., Zhao, R., Zhang, R., Su, J., Wei, D., Zhang, M., & Chen, Y. (2026). PLaST: Towards paralinguistic-aware speech translation. *Proceedings of the AAAI Conference on Artificial Intelligence, 40*(38), 31805–31813.
7. Nord, C. (1997). Alice abroad: Dealing with descriptions and transcriptions of paralanguage in literary translation. In F. Poyatos (Ed.), *Nonverbal Communication and Translation: New Perspectives and Challenges in Literature, Interpretation and the Media* (pp. 107–130). John Benjamins.
8. Poyatos, F. (1997). Aspects, problems and challenges of nonverbal communication in literary translation. In F. Poyatos (Ed.), *Nonverbal Communication and Translation: New Perspectives and Challenges in Literature, Interpretation and the Media* (pp. 17–48). John Benjamins.
9. Poyatos, F. (Ed.). (1997). *Nonverbal Communication and Translation: New Perspectives and Challenges in Literature, Interpretation and the Media*. John Benjamins.
10. Tsiamas, I., Sperber, M., Finch, A., & Garg, S. (2024). Speech is more than words: Do speech-to-text translation systems leverage prosody? *Proceedings of the Ninth Conference on Machine Translation*, 1235–1257.
11. Viaggio, S. (1997). Kinesics and the simultaneous interpreter: The advantages of listening with one's eyes and speaking with one's body. In F. Poyatos (Ed.), *Nonverbal Communication and Translation* (pp. 283–294). John Benjamins.
12. Zabalbeascoa, P. (1997). Dubbing and the nonverbal dimension of translation. In F. Poyatos (Ed.), *Nonverbal Communication and Translation* (pp. 327–342). John Benjamins.
13. Gumperz, J. J. (1982). *Discourse Strategies*. Cambridge University Press.
14. Halliday, M. A. K. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. Edward Arnold.
15. McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press.
16. Mehrabian, A. (1972). *Nonverbal Communication*. Aldine-Atherton.