A Cognitive Study on the Strongly Condensed and Implicit Features of Chinese-Characteristic Chunked Discourse and Human-Machine Multilingual Translation Among English, French and Spanish

Yuhan Wang, Shaolong Liu

Zhejiang Yuexiu University, Shaoxing 312000, Zhejiang, China

Abstract: This paper aims to systematically explore the strongly condensed and implicit features of Chinese-characteristic chunked discourse and its cognitive processing patterns in AI-based multilingual machine translation (among English, French, and Spanish) from the perspective combining multilingual comparison and artificial intelligence translation technology. The research shows that: 1) In the digital-intelligent era with deepening globalization, AI translation technology has provided new perspectives and exploration paths for multilingual conversion. However, when facing significant differences in language structures—especially Chinese implicit chunked expressions—AI translation exhibits various problems in source language cognition and target language cognitive processing, and there remains a significant gap compared with high-level human translation. 2) Chinese differs from Indo-European languages (such as English, French, and Spanish) in terms of grammatical structures and expression methods in temporal and spatial dimensions. In particular, Chinese-characteristic chunked discourse is usually highly condensed, often showing surface-level concealment of semantics (e.g., agentive meaning or agent), which easily leads to misinterpretation and omission in the understanding and production processes of multilingual translation. 3) The "strongly spatially condensed" chunked discourse focused on in this paper is not only an important manifestation of the characteristics of Chinese expression but also carries rich Chinese concepts and cultural connotations. However, due to the high simplicity of their surface structure, simply adopting strategies of literal and rigid translation in translation tends to cause the loss of deep semantic concepts of the source language or even misunderstanding and mistranslation, thereby weakening the effectiveness of international communication of China's inherent thoughts. Therefore, it is urgent to comprehensively and in-depth analyze the cognitive comprehension and explicit translation processing mechanisms of AI translation systems regarding Chinese implicit chunked expressions from a multilingual perspective in the context of AI, so as to evaluate their semantic restoration ability and ideological communication effectiveness. The relevant explorations in this paper provide new theoretical perspectives and operational paradigms for future human-machine translation research.

Keywords: Chinese-characteristic discourse; Condensed and implicit features; Comparison between human and machine translation; Differences in cognitive processing

1. Chinese-characteristic chunked expressions and their spatial characteristics

Chinese-characteristic chunked expressions usually appear in the form of multi-character structures, featuring highly concise and condensed expression. The unique linguistic structure of Chinese embodies the characteristics of chunkiness and discreteness under strong spatial constraints [1], which is particularly prominent in the process of cross-linguistic translation and easily leads to implicit or ambiguous semantics. In contrast, Western language systems represented by English, French, and Spanish focus more on the characteristics of connectivity and continuity under strong temporal constraints, manifested in the explicit relevance of grammatical structures and semantic concepts, as well as the clarity of temporal clue expression.

In recent years, with the development of artificial intelligence technology, machine translation has gradually entered the practical field of language translation and attracted widespread attention. In particular, the application of artificial intelligence tools in the translation of Chinese-characteristic chunked expressions has become one of the research hotspots in the current linguistic and translation circles. However, existing studies[2] have pointed out that artificial intelligence technology usually fails to accurately reproduce the agentive meaning and its implied meaning in Chinese-characteristic chunked expressions [3]. This translational deficiency not only significantly weakens the expressive effect of the source language but also poses certain obstacles to the international dissemination of "Chinese spirit".

This paper aims to explore the explicit translation status of implicit agentive meaning in Chinese-characteristic chunked expressions in official translations and machine translations of English, French, Spanish, etc., from the perspective of human-machine translation comparison. The study finds that the four-character chunked structure is particularly prominent and widely existing in Chinese-characteristic chunked expressions. Therefore, this study mainly selects chunked expressions with a length of "4±1" (three-character, four-character, and five-character structures) as the main research objects, in order to reveal the cognitive processing characteristics of implicit chunked expressions in their translation and their explicit transformation in the target language. Let's examine the following examples of Chinese implicit chunked expressions:

Table 1. Examples of Chinese implicit chunked expressions

Chunked expressions	Example 1	Example 2
Three-character	If one can be renewed in a day, he should strive to be renewed every day, and keep renewing day after day.	Advocate that major countries take the lead in emphasizing equality, good faith, cooperation and the rule of law, and take the lead in abiding by the UN Charter and international law.
Four-character	In time of peace, think of war; forethought ensures preparation, and preparation prevents calamity.	Aim high and keep feet firmly on the ground.
Five-character	Think more for the youth and less for oneself.	Observe the people to understand governance, and awaken oneself to rectify one's conduct.

The examples listed above demonstrate typical political discourse chunks with different character counts in China's current political discourse. They generally exhibit formal features such as "strong spatial condensation" and "block discreteness," showing a highly concise way of organizing information and a strong ability to compress discourse structures. In specific language practices, such discourse forms often cause certain semantic components, especially the agentive meaning and the ideological connotations it carries, to be in a state of "obscuration" or "latency," presenting the pragmatic feature of "form latent and meaning hidden." This structure has gradually become a key word and core expression in the construction of discourse with Chinese characteristics in the generation and dissemination of current political discourse.

From the perspective of cross-linguistic expression and translation, the implicit semantic information and cultural context implied in such "strongly spatially block-discrete" discourse units constitute difficulties in translation for external communication. Its research not only involves the implicit mapping relationship between linguistic form and meaning

but also relates to the efficiency and communication effect of the external discourse conversion of China's governance concepts and administrative intentions. Accurately interpreting the form-meaning structure of such block expressions and their potential pragmatic functions is of important theoretical and practical significance for improving the translatability and cross-cultural acceptability of political texts. Specifically, in the above block examples, the three-character and four-character sentences (Examples 1 and 2) both omit explicit agentive meanings or agent components, such as "we" or "the government." The omission of agentive meaning is extremely common in Chinese block discourse, with linguistic characteristics such as generality and implicitness, and native speakers often achieve semantic supplementation through contextual inference. In an official context, the English translation of this sentence explicitly processes the agentive meaning through contextual reasoning, for example: The government should care more about the younger generation. It can be seen that manual translation appropriately supplements and reconstructs the implicit information of the original block language through discourse analysis and context restoration.

However, in the context of artificial intelligence, especially in the process of automatic conversion relying on machine translation systems, such agentive components "obscured" by the block context are often difficult to be correctly identified or reasonably restored [4]. Existing machine translation systems mainly rely on superficial grammatical structures and big data-driven corpus matching mechanisms, with limited ability to interpret latent agents, and it is difficult to effectively capture the implicit logical relationships and pragmatic intentions of intra-block and inter-block components in the block context. This not only leads to incomplete translation information but may also cause semantic misunderstandings among target audiences, reflecting the shortcomings of AI translation tools in dealing with current political discourse with Chinese characteristics.

2. Settings of machine translation instructions and comparison between human and machine translation

On the other hand, Western languages such as English, French, and Spanish demonstrate coherent and continuous coding under strong temporal constraints, where the semantic connotations within or between chunk structures remain obscure (i.e., "explicit in form with clear meaning"). The temporal-spatial differences between Chinese and Western languages represented by English, French, and Spanish are likely to cause semantic comprehension errors and translation obstacles for (human or machine) translators.

Source Language 1: Advocate that major countries take the lead in upholding equality, good faith, cooperation, and the rule of law, ... (Xinhua News Agency: Global Security Initiative Concept Paper, 2023)

Official English Translation: The Chinese government will call on major countries to lead by example in honoring equality, and… (BEIJING, Feb. 21 (Xinhua) --"The Global Security Initiative Concept Paper.")

Official French Translation: Le gouvernement chinois(The Chinese government)appellera...

Official Spanish Translation: El gobierno chino (The Chinese government)instará a los…

Observing the above examples, it can be seen that Chinese three-character chunk structures exhibit obvious characteristics of high chunk discreteness, and the agent in the source language is often in an implicit or "obscured" state, meaning that the specific identity of the agent is hidden. Without in-depth analysis of the internal linguistic structure of the source language and understanding of the external linguistic context, it is difficult to accurately identify and discern the potential agent (or governing subject, agentive meaning). This situation directly leads to the difficulty of AI translation tools in correctly and explicitly expressing "the Chinese government" as the specific executor of governance concepts in the translation process into English, French, and Spanish.

To more accurately compare the effect of making implicit meanings explicit in translating Chinese characteristic chunked expressions between human translation and intelligent translation, this study specifically designed three different translation instructions (Prompts). These three instructions show a progressive change in the clarity of content, the depth of semantic prompts, and the application of translation strategies. The specific content and operational requirements are as

follows.

Prompt 1 (P1): The most basic translation instruction: Please translate the sentence into English (for French translation instructions, "English" is replaced with "French"; for Spanish translation, it is replaced with "Spanish"). This instruction only requires AI to directly translate the sentence without putting forward special translation requirements, mainly to examine the performance of machine translation systems in basic translation tasks.

Prompt 2 (P2): On the basis of P1, adding the requirement for making implicit information explicit: Please translate the sentence into English, adding in the implicit meanings in the translated sentence where necessary (the same applies to French and Spanish translation instructions). This instruction emphasizes the excavation and explicit processing of potential implicit information in the translation process, especially regarding the agent role or semantic elements not explicitly stated in the context.

Prompt 3 (P3): Further deepening the requirements of P2, specifically specifying the clear explicit translation of the agent role: Please translate the sentence into English, adding in the implicit personal subject or agent in the translated sentence where necessary (the same applies to French and Spanish translation instructions). This instruction emphasizes the explicit coding of the subject or nominal agentive meaning, that is, requiring AI machine translation to explicitly introduce the potential agent or agentive meaning in the translation process, or to make the implicit subject or agent in the chunked source language visible in the target translation.

By using the above three translation instructions with progressively optimized prompt information, we will deeply observe the effect of making obscured meanings explicit in AI translation of Chinese chunked source texts under different instructions, and then analyze the quality of target language processing generated by machine translation driven by instructions with different optimization levels. This study will adopt a qualitative analysis method to conduct detailed interpretation and evaluation of the results. Theoretically, with the progressive human-machine interaction of the three translation instructions, the depth and complexity of translation should show a linear increasing trend. However, do the actual translation results fully conform to this theoretical expectation? When comparing official translations with ChatGPT-driven translations, in what specific aspects is the effectiveness of machine translation reflected in actual cognitive processing? At the same time, what are its limitations? This paper will explore the above issues, aiming to reveal the advantages and disadvantages of artificial intelligence translation tools in handling the translation tasks of Chinese characteristic chunked adverbials, and provide theoretical support and practical basis for improving translation quality and AI technology in the future.

3. Types of explicitation of agentive meaning in human and machine translation and differences in cognitive processing

This paper adopts a combination of qualitative and quantitative methods to conduct a systematic analysis of translation data. First, a Pre-text Framework is constructed. By benchmarking against official translations and based on the differential characteristics of AI's explicitation effects, translation results are categorized into the following five types:

(1) Zero Explicitation (ZE): The translated text does not explicitly convey any latent meaning of the source language;

(2) Exceptional Explicitation (EE): The translated text contains explicitation that significantly deviates from the expected meaning of the original text;

(3) Partial Explicitation (PE): The translated text explicitly conveys part of the latent meaning of the source language;

(4) Mixed Explicitation (ME): The translated text presents a mixed state of both explicitation and non-explicitation;

(5) Same Explicitation (SE): The translated text completely and systematically explicates the latent information in the original text.

After obtaining machine-translated texts in English, French, and Spanish, the translation texts are coded one by one according to the above classification system. Subsequently, the percentage of each type of explicitation is calculated for quantitative comparative analysis. This method can objectively and effectively reveal the actual performance of different

AI translation systems in explicating the implicit meaning of Chinese chunk structures. To ensure that the collected corpus faithfully reflects the performance of AI translation tools (represented by ChatGPT) in explicating implicit meaning, the research team designed a corpus collection method combining multi-frequency instruction input and random sampling. Under two instructions (P1 and P2), the following translations are obtained:

Source Language 1: 倡导大国带头讲平等、讲诚信、讲合作、讲法治，……。

Prompt 1 English translation: Encourage major powers to lead by promoting equality and….

Prompt 2 English translation: We advocate for major powers to take the lead in promoting equality, and…

Prompt 1 French translation: Nous (we) préconisons que les grandes puissances prennent…

Prompt 2 French translation: Nous (we) préconisons que les grandes puissances prennent…

Prompt 1 Spanish translation: Abogar por que las grandes potencias tomen la iniciativa…

Prompt 2 Spanish translation: Abogar por que las grandes potencias asuman un papel…

The above Prompt 1 and Prompt 2 represent two rounds of translation instructions sent to the ChatGPT system at different times during the research process, both putting forward basic translation task requirements. Among them, Prompt 2 has a certain advancement in semantic level compared to Prompt 1, but it does not put forward clear operational requirements for the explicit expression of "implicit agent". The machine translation results generated based on these two rounds of instructions are as shown above.

By comparing the above machine translation outputs with human translations, it can be found that current AI systems generally tend to adopt processing strategies such as zero explicitness, partial explicitness, and mixed explicitness when dealing with such chunked adverbials. In the absence of clear guidance, even if the system can provide explicit compensation for some potential agentive meanings, it is still difficult to accurately identify and effectively convey the identity of the agent implied in the semantics. This finding suggests that even against the backdrop of significant improvements in the performance of current large language models, machine translation still has obvious cognitive obstacles and generative limitations when dealing with complex semantic phenomena such as implicit meaning and pragmatic gaps in language, and there is an urgent need to further optimize the cognitive processing capabilities of AI language models in implicit meaning inference and reconstruction.

Furthermore, in terms of instruction design, Prompt 3 has achieved a significant strategic shift compared to the previous two rounds of instructions (Prompts 1-2), explicitly putting forward specific requirements for translating the "implicit agent" for the first time. To examine the intervention effect of this guidance strategy on the cognitive processing mechanism of machine translation, the four-character chunk structures listed above are taken as samples, and corresponding translations are generated based on Prompt 3. The processing methods of agentive meaning are analyzed and evaluated to further reveal the performance mechanism and reconstruction characteristics of AI systems in the "implicit → explicit" semantic conversion under specific instruction optimization. The specific machine translation results are as follows:

Source Language 2: 志存高远，脚踏实地。

Official English translation: When we live in stability and peace, we must remain…, we ensure that we are…, and with thorough preparation, we can prevent crises and avoid harm.

Prompt 3 English translation: When we live in…. By…, we ensure…, and…, we can ….

Prompt 3 French translation: Lorsque nous (we) vivons…, nous devons…. En envisageant de manière proactive les défis possibles, nous (we) veillons …, nous (we) pouvons….

Prompt 3 Spanish translation: Cuando vivimos…, podemos prevenir….

It should be particularly noted that under translation instructions emphasizing the explicitation of implicit agentive meaning or agents, the machine translation results in English and French explicitly express "we" (French "nous"), both of which are generic personal pronouns expressing a general "we". However, in the Spanish translation, the agentive meaning of all verbs – "nosotros/nosotras" (we) – remains in a state of non-explicitation: that is, the agentive subject is not explicitly presented in the translation but is indirectly manifested through the inflectional endings of the verbs. Although

these translations have explicated the implicit agentive meaning to a certain extent, there is still a certain gap compared with the results of human explicitation.

Based on the qualitative analysis of the above AI machine translation performance and multi-frequency input sampling results, it is not difficult for us to find the following: First, driven by simple translation instructions (Prompt 1), the agentive meaning in the Chinese source language generally presents an implicit or latent state in the machine translation results. Second, driven by relatively more explicit translation instructions (Prompt 2), although there still occur cases of zero explicitation of part of the agentive meaning, after multiple repeated inputs and operations of the instructions, the agentive meaning gradually shows varying degrees of explicitation characteristics, mainly manifested in the frequent use of personal pronouns or generic pronouns to replace specific agentive meanings. This coding strategy for agent translation is widely used in Western languages. Pronouns have the dual functions of nouns and adjectives, which can effectively avoid the repetition of nouns, improve the conciseness and fluency of language expression, and at the same time enhance the affinity and sense of involvement of the translation for readers.

In addition, driven by translation instructions that explicitly emphasize agentive meaning (Prompt 3), the degree of explicitation of agentive meaning is further strengthened. The examples above even show additional agentive meanings that did not appear in the original translations. This processing strategy of adding agentive subjects expresses the firm attitude and strong stance of the original author. This explicitation strategy not only emphasizes the translation of the surface layer of language but also highlights the author's attitude, emotions, and communicative intentions in the original text, that is, it reveals the potential semantic prosody characteristics of the chunked original text [5]. It indicates that when understanding and translating Chinese implicit chunked adverbials, translators not only need to pay attention to the linguistic level or surface meaning of the chunked source language but also need to construe and convey the attitudinal meaning or communicative intentions inherent in the linguistic level.

4. Conclusion

The block-like discourse coding with Chinese characteristics highlights the highly condensed nature of its form/structural configuration and the highly implicit nature of its semantic concepts. This stands in significant cross-linguistic contrast to the linear sequential representation of "explicit in form and clear in meaning" in highly temporal Western languages[6]. This unique cross-linguistic phenomenon poses cognitive processing challenges to multilingual translation practices in the context of human-machine collaboration. Against the backdrop of globalization, digitalization, and intelligentization, the widespread application of intelligent translation tools has become a significant dimension of innovation in translation practice. The human-machine collaborative translation model not only significantly improves translation efficiency but also effectively identifies and corrects cognitive biases in the translation process through real-time feedback mechanisms. Particularly in the translation of block-like discourse with Chinese characteristics, intelligent translation tools have the advantage of quickly processing massive amounts of text data, which can effectively assist human translators in cross-linguistic text conversion. However, when dealing with complex cultural contexts or block-like discrete languages, existing intelligent translation tools still require further system updates and operational training.

Based on the analysis results of machine translation quality in this study, it can be seen that currently, AI translation tools such as ChatGPT and DeepSeek have not yet reached the ideal level of "full explicit translation" (SE) similar to human translation in handling the explicit translation of Chinese implicit block-like discourse, and mostly manifest as partial explicit translation (PE) or mixed explicit translation (ME). For this reason, further strengthening the design and operation of "multi-dimensional and high-frequency" instructions in human-machine collaborative translation in the context of AI may be a goal to be tackled in future related research, as it is crucial to the translation quality of block-like discourse with Chinese characteristics and the effectiveness of international communication of its inherent ideas.

Based on the analysis results of machine translation quality in this study, it can be seen that currently, AI translation tools such as ChatGPT and DeepSeek have not yet reached the ideal level of "full explicit translation" (SE) similar to

official human translation in handling the explicit translation of Chinese implicit block-like discourse, and mostly manifest as partial explicit translation (PE) or mixed explicit translation (ME). For this reason, further strengthening the design and operation of "multi-dimensional and high-frequency" instructions in human-machine collaborative translation in the context of AI may be a goal to be tackled in future related research, as it is crucial to the translation quality of block-like discourse with Chinese characteristics and the effectiveness of international communication of its inherent ideas.

Disclosure statement

The author declares no conflict of interest.

References

[1]   Wang W, Gao J, 2019, On the Blockiness and Discreteness of Chinese Four-Character Idioms. Journal of Beijing International Studies University, 41(02): 3-19.

[2]   Zhao Q, Huang J, 2018, The Mapping Model and Constraint Mechanism of Synesthesia Metaphor in Modern Chinese. Language Teaching and Research, 2018, Issue 1: 240-253.

[3]   Cao H, Liu S, 2024, The Effectiveness of ChatGPT in Translating Chunky Construction Texts in Chinese Political Discourse. Journal of Electrical Systems, 20-2 (2024): 1684-1698.

[4]   Salvagno, M., Taccone, F. S. & Gerli, A. G., 2023, Can artificial intelligence help for scientific writing? Critical Care, 27(1):75-75.

[5]   Fillmore, C J., 1976, Frame Semantics and the Nature of Language. Annals of the New York Academy of Sciences, 1976: 20-32.

[6]   Liu S, Wang H, Wang L, 2021, A cognitive study on the English translation of Chinese characteristic chunk adverbials: An investigation of translation competence from the perspective of current political texts. Foreign Language Teaching Theory and Practice, 2021(04):137-145+162.