

Quantitative Analysis of Video Game Localization—A Case Study of Metro: Exodus

Xiujuan Guan*, Junyuan Na

Institute of Russian, Heilongjiang University, Heilongjiang Province, Harbin 150080, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Video games are a multimodal cultural product, but localization research on them is still limited. This study takes Metro: Exodus as a case and examines Russian-to-Chinese localization with quantitative methods. Based on a self-built Russian-Chinese parallel corpus and grounded in Corpus-based Translation Studies and Translation Universals theory, it measures lexical richness, mean sentence length, part-of-speech distribution, sentiment, and dependency relations in the target text. The results show a clear tendency toward simplification in vocabulary and sentence structure, together with selective explicitation and emotional adjustment. Overall, the Chinese version keeps the text readable without weakening immersion.

Keywords: Localization translation; Video game; Quantitative analysis; Metro: Exodus

Online publication: July 15, 2026

1. Introduction

With the growth of the global game industry, localization has become an important factor in market performance and player experience. It is not a simple language transfer, but a multimodal adaptation process involving narrative tone, cultural references, interface constraints, and immersion^[1-2]. Metro: Exodus is a useful case because its Russian post-apocalyptic setting, colloquial dialogue, and slang require the Chinese version to preserve both information and atmosphere.

Most existing studies still focus on strategy discussion or individual cases. Research on Chinese game localization, general localization processes, and Russian-language cases has clarified many practical issues, but it often centers on culture-loaded items, humor, or single dialogues^[3-6]. Broader stylistic tendencies are less often described with reproducible evidence.

Corpus-based translation studies make it possible to move from impressionistic judgment to measurable description^[7-10]. On this basis, this paper builds a Russian-Chinese parallel corpus of Metro: Exodus and examines lexical richness, mean sentence length, part-of-speech distribution, sentiment, and dependency

relations. The goal is to see how the Chinese localization handles information transfer, readability, and atmosphere. Because the corpus includes dialogue, interface prompts, combat instructions, and emotionally marked narration, it also allows us to check whether simplification, explicitation, and emotional adjustment are recurring tendencies rather than isolated choices.

2. Lexical level quantitative analysis

This section examines lexical features of the self-built bilingual game corpus, including text richness, part-of-speech (POS) distribution, and emotional expression.

2.1. Text complexity and player cognitive load

In games, text competes with visual and audio input, so localization tends to reduce textual processing pressure. Lexical richness and lexical density are useful indicators for this tendency ^[11]. Because Chinese segmentation is semantic while Russian tokens largely follow word boundaries, the two corpora are compared mainly through relative tendencies; the Universal Dependencies Syntagrus corpus is used only as a reference baseline. The point is not absolute comparability, but whether both game corpora move away from general written discourse in the same direction and whether the Chinese version does so more strongly.

Table 1 shows that the Chinese translation uses more tokens (308,100) than the Russian source (245,740) but fewer types. TTR, STTR, and Dugast Index all follow the same descending pattern: Reference Corpus > Game Russian Corpus > Game Chinese Corpus. The game corpora are therefore more repetitive and standardized than general texts, and the Chinese localization is the most simplified one ^[12]. Lexical density was then calculated with Halliday's clause-based formula to see whether the same tendency appears at the structural level. Lower diversity alone could come from repeated character names or setting terms; when lower density appears as well, the case for broader simplification is stronger.

Table 1. Lexical richness features

Item	Game Russian Corpus	Game Chinese Corpus	Reference Corpus
Tokens	245,740	308,100	128,728
Types	17,170	16,307	17,355
TTR	6.98%	5.29%	13.4%
STTR	43.09%	42.70%	46.37%
Dugast Index	57.89	54.34	68.24

Figure 1 shows that both game corpora have lower lexical density than the reference corpus. Although the Chinese corpus is larger in volume, its density stays close to the Russian corpus, which suggests that the localization spreads information across more clauses instead of packing it into dense sentences. This kind of segmentation suits Chinese reading habits and reduces processing load during play.

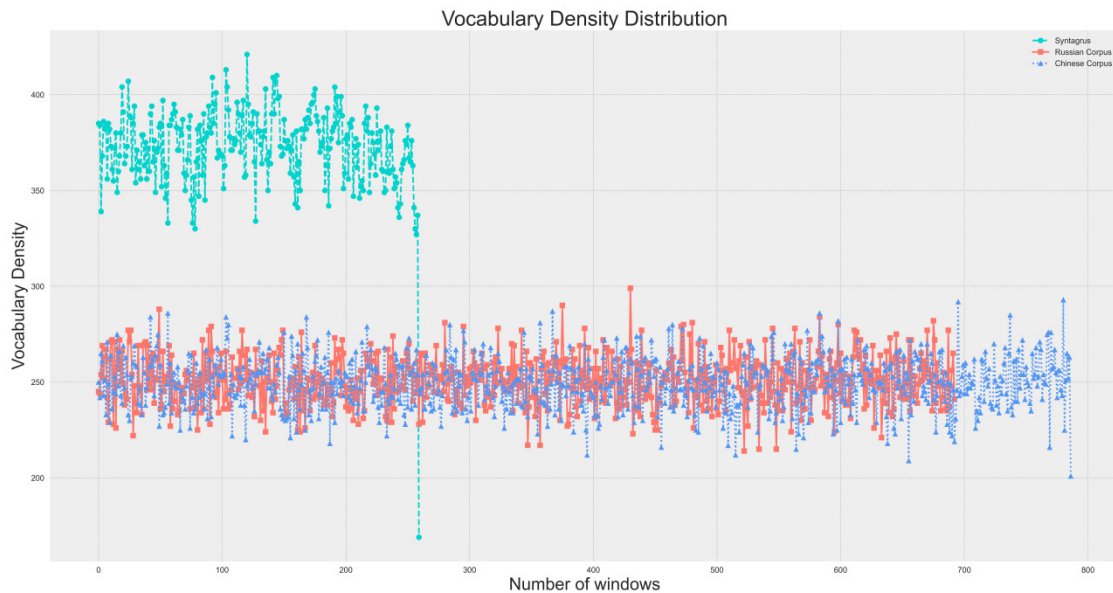


Figure 1. Lexical density distribution

2.2. Comparative quantitative analysis of part-of-speech features

POS distribution provides another view of stylistic transfer. This study used spaCy to tag the bilingual game corpus. The results are shown in **Figure 2**.

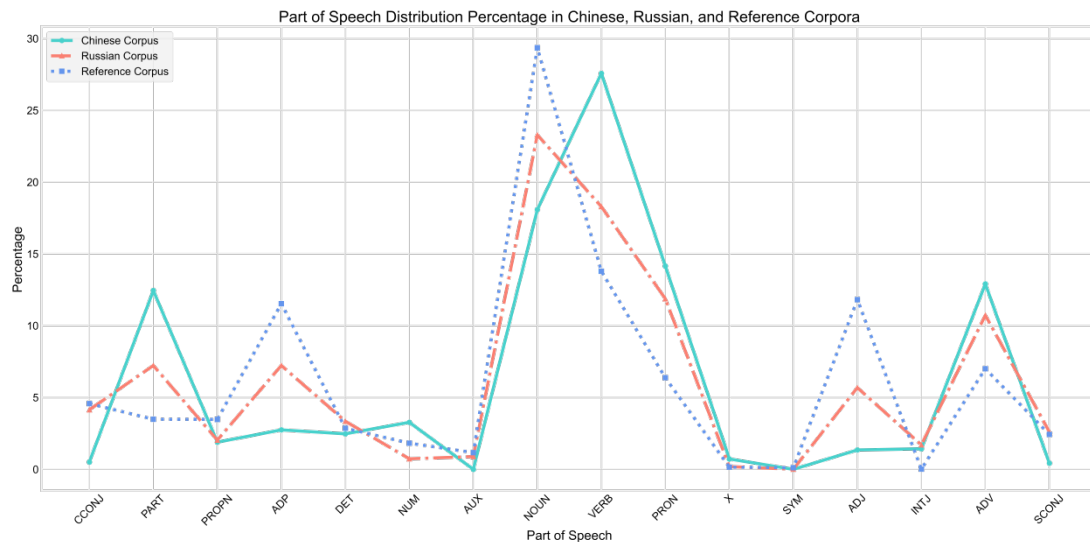


Figure 2. Part of speech distribution

Figure 2 shows clear grammatical differences. Nouns appear most frequently in the Reference corpus (33.19%), followed by the Game Russian corpus (27.89%) and the Game Chinese corpus (20.75%). Conversely, verb usage is highest in the Game Chinese corpus (31.26%), compared with 21.84% in the Game Russian corpus and 15.61% in the Reference corpus. General texts are more nominal, while game discourse is more action-oriented. The Chinese localization pushes this tendency further: adjective proportion drops to 1.55%, while adverbs remain frequent, showing substantial part-of-speech conversion in translation. The

result fits the practical demands of game text, where instructions, reactions, and event progression need to be processed quickly.

Conjunction data points to another shift. Coordinating conjunctions fall from 4.96% in Russian to 0.58% in Chinese, and subordinating conjunctions also decrease. In Blum-Kulka's terms, the Chinese text relies less on overt connectors and more on paratactic relations, which fits implicature in translation^[13]. The "Activity" metric (Verbs / [Verbs + Adjectives]) also peaks in the Chinese corpus (96.34%), confirming the dynamic character of game discourse.

2.3. Comparative quantitative analysis of emotional expression

For a narrative-heavy game such as Metro: Exodus, emotional tone is part of the gameplay experience. Using SnowNLP for the Chinese corpus and Dostoevsky for the Russian corpus, sentiment scores were normalized to [0, 1] for comparison (**Figure 3**).

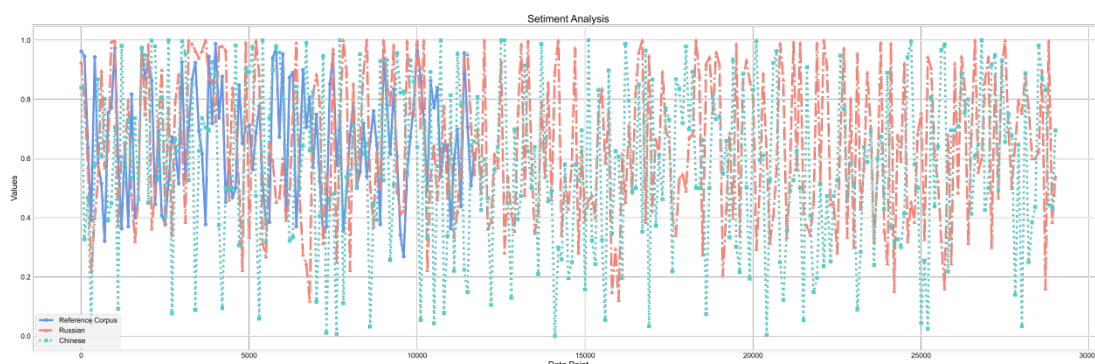


Figure 3. Visualization chart of sentiment distribution

Figure 3 shows that both game corpora contain more emotional extremes than the reference corpus, which fits the conflict-driven plot. The Chinese corpus is noticeably more negative than the Russian original, and **Table 2** confirms that the differences among all three corpus pairs are significant ($P < 0.05$).

Table 2. Analysis of commonality in corpus sentiment results

Items	Reference Corpus	Russian Corpus	Chinese Corpus
Mean	0.649	0.642	0.572
STD	0.0452	0.0638	0.0783
Pairs	Ref — Russian	Ref — Chinese	Russian — Chinese
T-Value	2.74	26.74	31.44
P-Value	<0.05	< 0.05	< 0.05
Significance	√	√	√

The Chinese text's lower mean (0.572) and higher standard deviation indicate emotional amplification. The wording is likely intensified in some places, or modal particles are added to strengthen resonance, but the effect still needs to match characterization. Stronger affect may improve immediacy, yet overuse can flatten speaker differences or turn tension into melodrama.

3. Syntactic level quantitative analysis

Sentence-level metrics were then used to test readability and structural shifts in the Russian and Chinese corpora.

3.1. Text readability and player group acceptance analysis

Readability matters because it affects how quickly players process narrative information. Mean Sentence Length (MSL) metrics were calculated and adjusted for corpus length imbalances. The results are shown in **Table 3**.

Table 3. Mean sentence length

Items	Russian	Chinese	Reference Corpus
Tokens	245,740	308,100	128,728
Total Sentences	41,405	38,433	9,125
Mean Sentence Length	5.93	8.01	14.10
Mean Sentence Length (Chunked)	5.77	7.66	14.67

Table 3 shows that both game corpora use much shorter sentences than the reference corpus. The chunked mean sentence length of the Chinese localization (7.66) is slightly above the Russian original (5.77), which suggests limited explicitation: some implicit links are verbalized to support comprehension. Even so, both remain far below the reference corpus (14.67). The result points to controlled explicitation rather than simple expansion.

Readability formulas further support this result (**Table 4**)^[14]. The Russian game text reaches a Flesch Reading Ease of 83.16, far above the reference corpus (49.5), and the other indices point in the same direction. The text is geared toward rapid uptake rather than academic density. Even though the formulas are language-specific, they still show that the source text is designed for accessible reception, and the Chinese localization keeps that principle instead of shifting to a more literary or academic style.

Table 4. Readability data for the game Russian corpus and the reference corpus

Readability Formula	Russian	Reference Corpus
Flesh Reading Ease	83.16	49.5
Flesh Kincaid Grade	9.7	17.3
SMOG	10.7	16.8
Coleman Liau Index	7.2	15.59
Automated Readability Index	6.9	14.3
Dale Chall	5.83	8.46

3.2. Dependency syntax quantitative analysis

Dependency relations provide a compact way to observe syntactic restructuring across languages^[15]. The statistical results are shown in **Figure 4**.

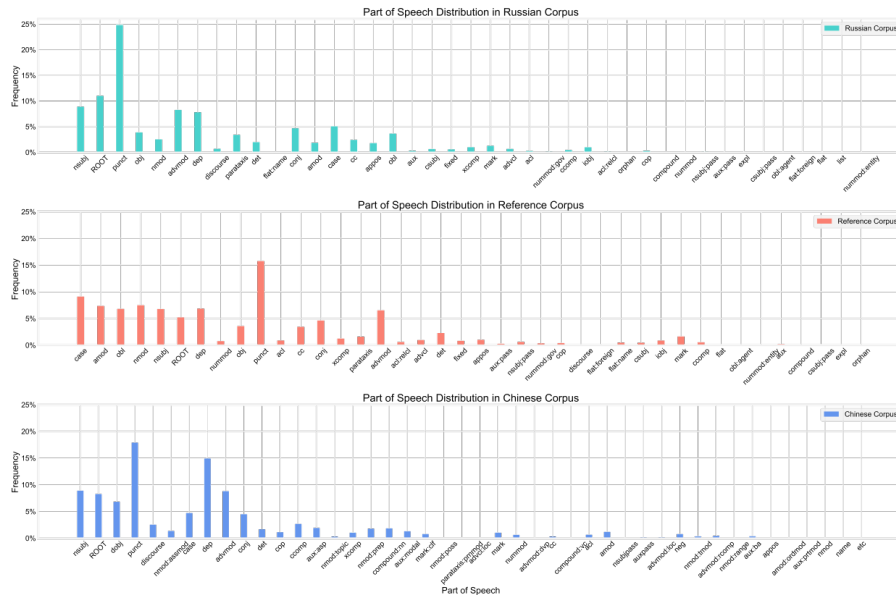


Figure 4 shows clear syntactic shifts across the corpora. Punctuation is more frequent in the Game Russian corpus (24.79%) than in the Chinese text (17.90%), while verb-related dependencies decrease from 11.01% to 8.26%. By contrast, nominal subjects rise in Chinese, with more than 3,500 added subjects relative to the source text, which suggests clause-level explicitation. At the same time, nmod relations almost disappear in the Chinese corpus, discourse markers increase, and parataxis drops from 3.41% to 0.07%. Overall, the shift is from Russian case- and punctuation-based structuring to Chinese reliance on particles and word order. In places where Russian morphology or omitted subjects would slow reading, the localization makes relations clearer.

3.3. Comparative quantitative analysis of theme expression

To test whether these local shifts alter the global theme, high-frequency words were extracted and analyzed via PCA (Principal Component Analysis) to visualize semantic clustering (**Figure 5**).

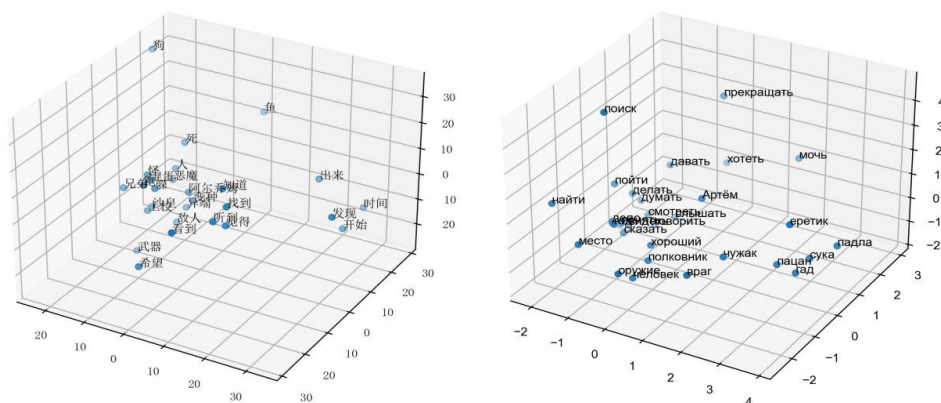


Figure 5 shows similar clustering trends in Russian and Chinese, indicating that the localization preserves the thematic core of the source text. The denser Chinese verb cluster also matches the action-oriented tendency found in the POS analysis. This suggests that the target text can simplify and restructure locally without losing the global semantic focus of the game world.

4. Conclusion

This paper uses a Russian-Chinese parallel corpus of *Metro: Exodus* to describe video game localization quantitatively. The Chinese localization shows lower lexical richness and density than general texts, stronger verbal dynamism, selective explicitation of subjects, and emotional amplification, while still keeping high readability and thematic consistency. The findings support the value of corpus-based translation studies for game localization research and show that quantitative evidence can complement qualitative case analysis. In this case, successful localization is achieved through controlled simplification, moderate explicitation, and emotionally calibrated adaptation under gameplay constraints. Future work can extend the corpus to other genres and language pairs to test whether these patterns hold beyond a single title.

Funding

This work was supported by the National Social Science Fund of China (No. 19BYY207), the Key Project of the Heilongjiang Philosophy and Social Science Fund (No. 25YYA001), the Project of the Institute of Translation Yiguoyimin “Research on Translation and Translation Industry in the New Era” (No. Z202304), and the Key Project “Research on the Multimodal Context Adaptation Mechanism of Russian Translation of Chinese Modernization Documentaries” (No. WY2026012).

Disclosure statement

The authors declare no conflict of interest.

References

- [1] O’Hagan M, Mangiron C, 2013, *Game Localization*. John Benjamins Publishing Company, Amsterdam.
- [2] Zhang W, 2013, Application of Transcreation in Game Localization Translation. *Modern Communication (Journal of Communication University of China)*, 2013(12): 165–166.
- [3] Bai T, Yuan P, 2023, Analysis of Chinese Localization Translation Strategies for Video Games. *Journal of Hubei University of Education*, 2023(3): 92–96.
- [4] Morozov MD, Features of the Computer Game Localization Process. *Bulletin of the Astrakhan State Technical University*, 2019(68): 59–63.
- [5] Kotlyar YV, Popova NV, 2020, The Problem of Localization of Computer Games into Russian, thesis, Peter the Great St. Petersburg Polytechnic University.
- [6] Kondratyeva AA, Seltsova AV, Difficulties in Localization of the Computer Game “The Witcher 3”. *Bulletin of the Chuvash State Pedagogical University named after I.Y. Yakovlev*.
- [7] Baker M, 1995, Corpora in Translation Studies. *TARGET*, 1995(7): 223–242.

- [8] Baker M, 1996, Corpus-based Translation Studies: The Challenges that Lie Ahead. *Terminology, LSP and Translation*, 1996(18): 175–186.
- [9] Guan X, Tao Y, 2017, *Corpora and Translation*. Fudan University Press, Shanghai.
- [10] Dai G, Liu S, 2023, Research Progress on Corpus-based Translation Teaching (2007—2022): A Comparative Analysis Based on Domestic and International Academic Papers. *Foreign Language World*, 2023(1): 40–48.
- [11] Dai G, 2008, A Corpus-based Study on English-Chinese Lexical Translation. *Journal of Xiamen University of Technology*, 2008(3): 94–98.
- [12] Yu L, 2024, A Study on Lexical Diversity and Syntactic Complexity of ChatGPT Translation. *Foreign Language Teaching and Research*, 2024(2): 297–307.
- [13] Blum-Kulka S, 1986, *Shifts of Cohesion and Coherence in Translation*. Narr, Tübingen.
- [14] Wang L, Developmental Stages and Characteristics of Readability Formula Research. *Language Teaching and Linguistic Studies*, 2022(2): 29–40.
- [15] Lei L, Jockers ML, Normalized Dependency Distance: Proposing a New Measure. *Journal of Quantitative Linguistics*, 2018(1): 1–18.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.