



How human and AI modulate informativity across genres: Evidence from translation

Jisoo Choi
(Hanyang University)

Choi, Jisoo. 2026. How human and AI modulate informativity across genres: Evidence from translation. *Linguistic Research* 43(Special Edition): 77-97. Whether AI translation departs from human translation in measurable ways remains largely a matter of impression rather than evidence. This study tests that question on Korean source texts rendered into German by human translators, neural machine translation, and a large language model, across news and literary fiction. Surprisal values from a pretrained language model quantify informativity at sentence and word levels, while coded cohesive devices such as recurrence, proforms, and ellipsis anchor the interpretation. Analysis of variance locates significant divergence in two configurations only. News texts diverge at the sentence level between human translation and LLM output ($F = 3.788$, $p = .030$). Literary texts diverge at the word level between human translation and NMT ($F = 3.815$, $p = .022$). Effect sizes stay small to medium throughout. Divergence is therefore selective, bound to genre and structural level rather than uniform. The contribution lies in converting informativity and cohesion from descriptive categories into replicable quantities for a language pair rarely examined this way. Corpus size restricts the findings to an exploratory case study. (Hanyang University)

Keywords informativity, textuality, Cohesion, AI literature translation, surprisal

1. Introduction

With the advancement and widespread availability of machine translation and large language models (LLMs), human translation is increasingly losing its edge in terms of speed. However, despite such progress, AI-generated translations still fall short of human translations in terms of communicative effectiveness and impact. This is particularly evident in literary translation, where AI frequently produces sentences

that are awkward in contrast with those crafted by humans.

This study contributes to clarifying the causes of this difference in effectiveness. To focus more precisely on differences, this study analyzes translations derived from the same text. Recent studies have addressed AI's translation capabilities (Donthi et al. 2024; Elkins 2024; Li 2024; Zhang et al. 2024). However, many of these studies have largely focused on translation error typologies, comparative evaluations of translation quality (Zhang et al. 2024), and improvements in idiomatic translation (Donthi et al. 2024). While Elkins (2024) and Choi (2024, 2026) have addressed stylistic aspects of AI translation, and such research remains scarce. There is a growing need for studies to include stylistic and textual analysis.

In particular, Lee and Zhao (2023, 2024) analyzed the German, English, and Chinese translations of Korean short stories qualitatively with a focus on textual cohesion and emphasized the importance of informativity-oriented translation. In their 2023 study, they examined the Korean source texts alongside their multilingual translations, concentrating on cohesive aspects such as recurrence, pro-forms, temporal markers, and conjunctions. In their 2024 follow-up study, they argued that informativity-oriented translation is the hallmark of high-quality translation.

This study focuses on cohesion and informativity as two standards of textuality (De Beaugrande and Dressler 1981), which are often interrelated, as shifts in cohesion tend to accompany changes in informativity. To examine how these relations differ between human and AI translations, the study identifies instances where cohesion diverges across the two translation types and compares how such divergences correspond to differences in informativity, using both quantitative and qualitative analyses.

2. Theoretical background

2.1 Informativity

Informativity refers to the degree of unpredictability in a text. The less predictable and probable a word is, the higher its informativity; conversely, the more predictable and probable it is, the lower its informativity.

From the perspective of textuality, informativity is concerned with the interaction

between the semantic structure of the text and the reader's cognitive state. Thus, informativity can be upgraded or downgraded as information is transmitted from the producer to the receiver (De Beaugrande and Dressler 1981).

Although translated texts often attempt to preserve the degree of informativity present in the source text, adjustments are sometimes necessary depending on the target culture and readership. If informativity is excessively high, the reader may fail to establish continuity between the text and their prior experience. Conversely, the source text's intended attention or interest may be lost. Strategies such as downgrading the informativity of cultural references or proper nouns can mitigate this risk.

Treatment of informativity also varies according to text type and genre¹. In news texts, if informativity is too high at the beginning, comprehension may be hindered (see Kim 2003: 52). By contrast, in literary text, excessive explicitation (Baker 2000) or simplification in translation may compromise the intended effect of the source text. Therefore, translators should not merely perform word-for-word substitution but should instead attempt to reproduce the text's intended effect through calibrated control of informativity.

2.2 Cohesion

De Beaugrande and Dressler (1981: 57) identified changes in cohesion as emerging through mechanisms such as recurrence, paraphrase, substitution, ellipsis, tense, aspect, modality, and conjunctions. This study particularly focuses on recurrence and ellipsis, which can occur across various textual layers. De Beaugrande and Dressler (1981) emphasize lexical recurrence, which is the recurrence of the same lexical item or phrase. While such recurrence can facilitate cohesion, excessive use may reduce the informativity of a text.

(1) John ran home and John ran home.

The example (1) above employs a predictable syntactic structure, close to default and preferred constructions. Despite the cognitive space available for processing more

¹ Text type refers to internal and functional-linguistic classifications of texts, while genre is reserved for externally defined and socially recognized categories (cf. Lee 2002).

complex information, unnecessary recurrence lowers the informativity to an excessive degree, which may diminish reader interest and attentional engagement. In contrast, recurrence in literary contexts, such as poetry, may serve a rhetorical function and therefore warrants preservation as illustrated in example (2).

(2) break, break, break

In such cases, lexical recurrence enhances rhythm and reinforces the emotional tone of the text. Other cohesion devices such as paraphrase and substitution offer variation in expression while preserving meaning and referential clarity, thereby playing a crucial role in the regulation of informativity. In this study, particular emphasis is placed on the phenomenon of recurrence, as it affects the reader's cognitive load and the retention of key information. A combination of quantitative and qualitative analyses will be employed to investigate this aspect of cohesion.

2.3 Surprisal

The concept of informativity in the context of text linguistics is rooted in Shannon's (1948) information theory (De Beaugrande and Dressler 1981: 155). According to this theory, the more alternatives are available at a given point in a message, the higher the informativity of the selected option. If only one word can logically occur in a given position, its informativity is low. Consequently, the more unpredictable the upcoming word is, the greater the degree of actual surprise when it appears.

In information theory, this surprise is referred to as 'Surprisal', which has been proposed as a proxy for human processing complexity (Hale 2001), corresponding to increased reading time (Demberg and Keller 2008). Wilcox et al. (2020) and Jon and Bojar (2024) have also argued that higher Surprisal values correlate with greater cognitive effort required by the reader.

Surprisal has been widely used to quantify linguistic predictability and complexity across multiple levels of analysis (Hale 2001; Levy 2008; Monsalve et al. 2012; Smith and Levy 2013; Wilcox et al. 2020). In cognitive processing research, surprisal has been validated as a reliable correlate of processing complexity, particularly through its association with sentence- or phrase-level reading times (Monsalve et al. 2012;

Xu et al. 2023) and eye-tracking data (Demberg and Keller 2008; Balling and Kizach 2017; Salicchi et al. 2023). More recently, the validity of Surprisal as a fine-grained measure of linguistic predictability has been further demonstrated by linking it to cortical EEG responses during real-time spoken language comprehension (Weissbart et al. 2020).

The present analysis focuses on translated texts. By comparing genre-based translation patterns across translation agents, it is possible to examine discourse-level characteristics of AI translationese. Rubino et al. (2016) demonstrated that Surprisal, in addition to complexity features based on n-gram frequency, serves as a meaningful indicator for distinguishing between human translations and originally authored texts. This suggests that the presence of translationese can be quantitatively identified and that informational properties, in addition to linguistic features, can serve as classification criteria.

Surprisal has also been used to model the routine of translation in previous research. Martinez and Teich (2017) incorporated Surprisal as an indicator of the predictability of information within a computational-cognitive framework that bridges translator expertise with degrees of automation. Degaetano-Ortlieb and Teich (2017) quantitatively analyzed intra-textual shift, a phenomenon of central concern in translation studies, using entropy and average Surprisal. Their study revealed genre- and style-specific differences in information patterns, aligning closely with the objectives of the present study, which aims to identify human and machine translation differences based on genre.

This study explores genre-sensitive differences using Surprisal values from neural language models and initiates a qualitative discussion on informativity as a textuality feature, alongside cohesion as a regulating factor.

3. Method

3.1 Corpus

The corpus employed in this study consists of translated texts from Korean into German. The translations were produced by three different translators: human translator, Google Translate, and ChatGPT (GPT-4o). The prompt was entered as

follows:

“You are a professional translator who has been translating between German and Korean for more than 10 years. Translate this text from Korean to German, considering the genre and considering the German readers, but do not delete or add content arbitrarily and do not make up words that you do not understand..”

Two source texts were selected for analysis. The first text group is made by three news articles published by *Korea.net*.² The second is a literary text, taken from the monologue section of *The Vegetarian* (Chae-sik-ju-eui-ja) (Oct. 2007) by Han Kang, a writer that won the Nobel Prize in Literature, and translated into German by Ki-Hyang Lee as *Die Vegetarierin* (Aug. 2016).

News articles contain longer, information-dense sentences, so surface cohesion must remain unobtrusive to ensure clarity. In contrast, the dream monologue in *Die Vegetarierin* uses short, repetitive, or elliptical sentences that heighten cognitive load but serve key stylistic functions. Thus, translations should preserve these literary effects rather than uniformly reducing informativity.

This study, designed as an experimental case study bridging text and quantitative linguistics, employs a focused small corpus to enable parallel qualitative analysis. The news text translations comprised 492 words (human), 521 words (NMT, Google Translate), and 539 words (LLM, ChatGPT/GPT-4o, OpenAI 2025). For the literary text, the respective word counts were 704 (human), 703 (NMT), and 709 (LLM).

3.2 Surprisal measurement

To obtain Surprisal values (Shannon 1948; Hale 2001), this study employed the GPT-2 language model to retrieve the log-probability of each token given its left-hand context.³ These log-probabilities were then used to compute Surprisal according to the following formula:⁴

2 The translator remains unidentified, despite the byline naming the Korea.net journalist. This precludes attributing stylistic patterns to either individual idiosyncrasy or general characteristics of human translation.

3 Surprisal values were computed using 117M-Parameter GPT-2 model (Radford et al., 2019), an autoregressive language model publicly available via Hugging Face (<https://huggingface.co/gpt2>). GPT-2 was selected for its accessibility, reproducibility, and ability to return log-probability distributions required for Surprisal computation. Although larger models such as GPT-4 offer improved fluency, their closed-source nature and restricted access to token-level probabilities render them unsuitable for reproducible Surprisal-based analysis.

$$Surprisal(n) = -\log_n P(w_n | (w_1 \dots w_n))$$

Tokenization was performed using the *AutoTokenizer* from the *dbmdz/german-gpt2* model. In word-level Surprisal calculation, all special characters and punctuation marks were removed. For sentence-level Surprisal, the text was segmented into sentences based on terminal punctuation such as periods, exclamation marks, and quotation marks. While German compound nouns may be segmented by subword-level tokenization, no explicit decomposition was applied to preserve semantic integrity.⁵ All steps were implemented using the Python-based Transformers library (Wolf et al. 2020) to ensure compatibility with the pretrained model and reproducible word segmentation.

Sentence-level Surprisal is computed as the negative log of the product of conditional token probabilities, which is mathematically equivalent to the sum of word-level Surprisal within the sentence. These levels differ interpretively. At the word-level, each token is treated as an individual data point, resulting in a large number of observations and higher statistical sensitivity, but without capturing sentence-level dependencies. In contrast, sentence-level Surprisal provides a single value per sentence, allowing for broader comparisons across texts. Accordingly, both levels were examined for three translator types (HT, NMT, LLM) and two genres (news and literature).

3.3 Cohesion analysis

To analyze textual cohesion across translation outputs, a manual, sentence-by-sentence comparative analysis was conducted by a single annotator. The analysis focused on identifying and comparing cohesive devices in translations produced by human translators, an NMT, and an LLM-based system. This analysis was grounded in the text linguistics framework proposed by De Beaugrande and Dressler (1981), with particular attention to two key cohesive mechanisms: recurrence (the repetition of

4 Surprisal (n) measures the probability (P) that a given word (w_n) occurs after the preceding words ($w_1 \dots w_{n-1}$). The lower the probability (P), the higher the surprisal value.

5 Future work could apply morpheme or subword-level tokenization to compound expressions and compare the resulting Surprisal estimates with those obtained under the current approach, thereby isolating the effect of tokenization granularity on the reported results.

lexical items or concepts) and ellipsis (the omission of elements recoverable from the context). Each sentence from the source text was aligned with its three corresponding translations, and the presence, absence, or alteration of these cohesive features was annotated and qualitatively compared. This approach enabled a focused, in-depth discourse-level assessment that complements quantitative analyses, with a single-annotator design ensuring consistency across all samples, though this also constitutes a methodological limitation.

4. Results and discussion

4.1 Informativity

Sentence-level Surprisal values were compared across three translation Genre (HT, NMT, and LLM). Group differences were tested with one-way ANOVA (Lakens 2013), followed by Tukey's HSD for pairwise comparisons.

In news texts (Figure 1), ANOVA revealed a statistically significant difference between groups ($F = 3.788$, $p = .030$). Tukey post-hoc comparisons indicated a difference between HT and LLM ($p = .028$), with HT showing a higher mean surprisal ($\mu = 6.20$) than LLM ($\mu = 5.37$). The difference between HT and NMT ($\mu = 5.64$, $p = .170$) and that between NMT and LLM ($p = .662$) were not significant. In other words, in the news dataset, human-translated sentences were less predictable than LLM outputs under the surprisal measure used here (See Table 1).

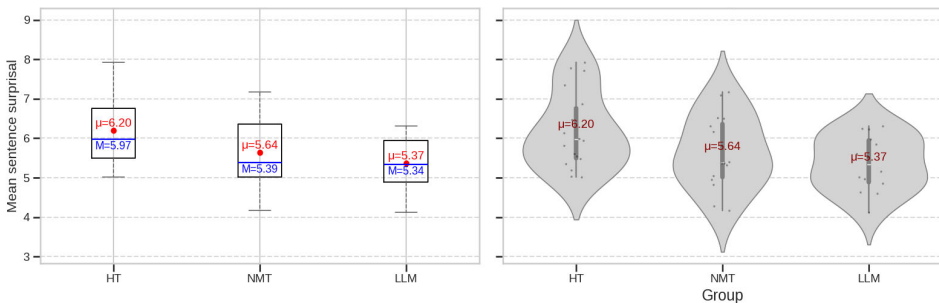


Figure 3. Sentence-level Surprisal – News texts

By contrast, sentence-level surprisal in the literary texts (Figure 2) did not differ significantly across HT, NMT, and LLM ($F = 1.354$, $p = .261$). The mean values were also relatively close, and none of the Tukey pairwise comparisons reached significance. This suggests that, at the sentence level, the literary dataset does not provide strong statistical evidence for systematic differences in predictability across the three translators.

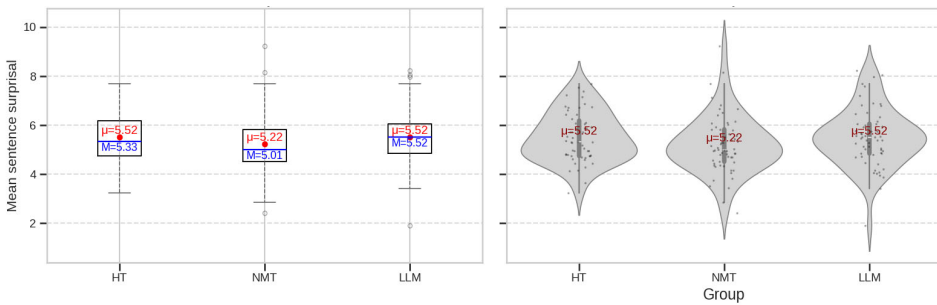


Figure 4. Sentence-level Surprisal – Literary texts

At the word level (Figure 3), the news texts likewise showed no statistically significant difference across translators ($F = 1.913$, $p = .148$). Although HT yielded the highest mean token-level surprisal ($\mu = 5.29$), followed by NMT ($\mu = 5.25$) and LLM ($\mu = 4.75$), none of the pairwise comparisons was significant. Accordingly, the numerical differences observed in the news word-level distributions should not be interpreted as robust evidence of a translator effect.

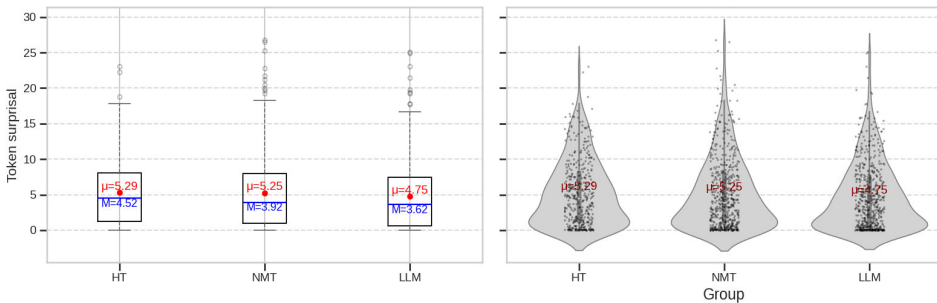


Figure 5. Word-level Surprisal – News texts

A different pattern emerged in the literary texts at the word level (Figure 4). ANOVA revealed a significant overall effect of translator ($F = 3.815$, $p = .022$). Tukey's HSD showed that the only significant pairwise difference was between HT and NMT ($p = .016$), with HT showing the highest mean surprisal ($\mu = 7.64$) and LLM the lowest ($\mu = 6.87$), LLM occupied an intermediate position ($\mu = 7.26$), and did not differ significantly from either HT or NMT. These results indicate that, in the literary dataset, human translations were lexically less predictable than LLM outputs, while LLM occupied an intermediate position without forming a statistically distinct group.

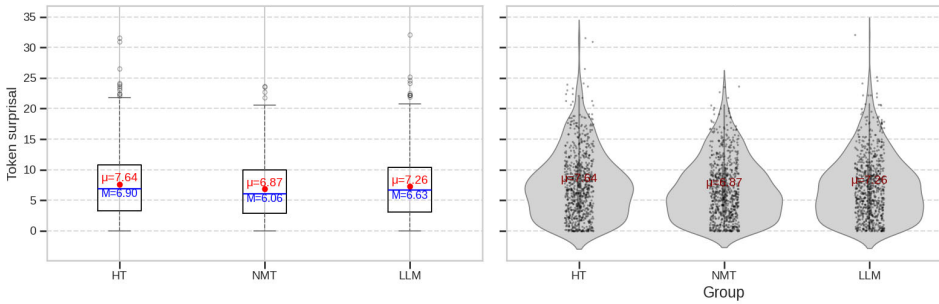


Figure 6. Word-level Surprisal – Literary texts

Taken together, these findings point to a genre-sensitive but non-uniform distribution of informativity across translators. Significant differences emerged in sentence-level surprisal for the news texts and in word-level surprisal for the literary texts, but the significant pairwise contrasts were limited in scope. Rather than supporting a simple human-versus-machine hierarchy, the results suggest that differences in predictability arise at different linguistic levels depending on genre and textual function.

This interpretation also requires caution in relation to non-significant results. The absence of a statistically significant difference does not demonstrate equivalence between human and machine translation. Rather, it indicates that the present data and model did not provide sufficient evidence for a reliable distinction under those specific conditions. This is particularly important in the comparisons involving NMT and LLM, where assumptions about “human-likeness” should not be inferred solely from lower or similar surprisal values.

More broadly, surprisal should not be treated as a direct proxy for overall

translation quality. Lower surprisal reflects greater probabilistic predictability under the language model, whereas higher surprisal reflects lower predictability. Neither value is inherently preferable independent of genre, style, or communicative purpose. In this sense, the findings are best understood as evidence that translation types distribute informational density differently across genres and across sentence-level versus word-level organization.

From the perspective of textuality as discussed by de Beaugrande and Dressler, these results suggest that informativity is not fixed at a single global level but interacts with discourse type and translational choice. The quantitative patterns reported here therefore motivate a more fine-grained qualitative analysis of how cohesion and genre-sensitive discourse structuring are realized across HT, NMT, and LLM outputs. The following section turns to that qualitative dimension.

Table 1. Results of ANOVA and post-hoc tests

Domain	Level	ANOVA F	p-value	Group 1	Group 2	Mean Diff	p-adj	95% CI (Lower, Upper)	Significant
News	Sentence	3.788	.030	HT	LLM	-0.830	.028	(-1.587, -0.074)	Yes
HT	NMT	-0.552	.170	(-1.281, 0.178)	No				
LLM	NMT	0.279	.662	(-0.498, 1.055)	No				
Literature	Sentence	1.354	.261	HT	LLM	-0.087	.934	(-0.673, 0.498)	No
HT	NMT	-0.385	.267	(-0.971, 0.200)	No				
LLM	NMT	-0.298	.335	(-0.796, 0.200)	No				
News	Word	1.913	.148	HT	LLM	-0.538	.216	(-1.293, 0.216)	No
HT	NMT	-0.043	.990	(-0.803, 0.717)	No				
LLM	NMT	0.495	.220	(-0.203, 1.193)	No				
Literature	Word	3.815	.022	HT	LLM	-0.379	.363	(-1.034, 0.276)	No

Domain	Level	ANOVA F	p-value	Group 1	Group 2	Mean Diff	p-adj	95% CI (Lower, Upper)	Significant
HT	NMT	-0.773	.016	(-1.430, -0.117)	Yes				
LLM	NMT	-0.394	.336	(-1.049, 0.262)	No				

The following chapter therefore turns to a qualitative analysis of cohesion, focusing on how human and machine translators handle these devices in genre-sensitive ways.

4.2 Cohesion

While the quantitative Surprisal analysis revealed differences in informativity across translation types and genres, it could not fully explain the underlying linguistic strategies. The next chapter therefore turns to cohesion, in the sense of De Beaugrande and Dressler's (1981) standards of textuality, and discusses representative cases of recurrence and ellipsis as devices that structure information and maintain referential continuity.

4.2.1 Cohesion in news text

The human translations examined in this study predominantly preserved the subject-verb order characteristic of journalistic prose and largely avoided nominalization. Notably, human translators rarely used the German preposition *laut* (according to), which often functions as an expression indicating source in journalistic discourse (Smailagić 2012: 23). Instead, they preferred adjective constructions over adverbial ones. In contrast, AI translations tended to replicate the source text more directly, frequently using prepositions such as *laut* (according to), *wegen* (due to), and *mit* (with) (example 3).

(3) German

ST 식품의약품안전처(식약처)가 28일 발표한 2023년도 의료기기 생산·수입·수출실적에 따르면 지난해 국내 의료기기산업 시장 규모는 10조7270억

원에 달했다.

According to the results on the production, import, and export of medical devices in 2023 announced by the Ministry of Food and Drug Safety on the 28th, the size of the domestic medical device market reached 10.727 trillion KRW last year.

HT *Das Ministerium für L... **teilte** am 28. Mai das Ergebnis für P... der koreanischen medizinischen Geräte im Jahr 2023 **mit**. Demnach **betrug** das koreanische Marktvolumen ...*

The Ministry for L... **shared** on the 28th of May the result for P... of the Korean medical devices in the year 2023. According to this amounted the Korean market volume ...

NMT ***Laut** der am 28. vom Ministerium bekannt gegebenen Produktion...-sleitung **erreichte** die Größe des heimischen Marktes...*

According to the on the 28th by the ministry announced production...performance **reached** the size of the domestic market ...

LLM ***Laut** dem am 28. veröffentlichten Bericht der Behörde über die Produktion ... **belief sich** der Markt für die heimische Medizingeräteindustrie ...*

According to the on the 28th published report of-the agency about the production ... **amounted** the market for the domestic medical device industry ...

It seems that the human translator avoided a literal rendering of the long noun phrase, possibly in an effort to manage informational density. Given that the Korean source text itself employs complex noun phrases, this translation strategy may appear somewhat counterintuitive. However, in light of the universal phenomenon of explication in translation (Baker 2000), it can be assumed that the translator's intention was to enhance clarity as much as possible.

Differences also emerged in the translation of repetition (Example 1). For example, the Korean term for 'domestic' was rendered by human translators as explicit *koreanisch* (Korean) only in the initial part of the text and was omitted entirely in subsequent instances. Machine translations, by contrast, repeatedly used a range of nominal and adjectival variants such as *inländisch* (domestic), *heimisch* (domestic), and *im Inland* (within the country) instead of omitting or varying the term. In cases

of verbs, AI often substituted the repeated term as *betrug* (amounted to)’ with different expressions, such as *belief sich* (amounted to), or *wies auf* (indicated).

These findings partly align with a recent study on informational text translation that compared ChatGPT’s performance with human translators across four language pairs (Yan et al. 2024). Yan et al. (2024) qualitatively identified systematic differences in translation strategies, with GPT-4 producing overly literal renderings and showing lexical inconsistencies, whereas human translators more often added content not present in the source. From a textual perspective, this pattern reflects differing strategies for managing informativity and target-audience orientation.

4.2.2 Cohesion in literary text

While the differences between human and AI translations in news articles were mainly observed in the use of recurrence, in literary texts, the most notable divergence was found in the handling of ellipsis. Humans effectively used ellipsis to eliminate informational redundancy thereby improving textual flow and cohesion. In contrast, AI generally failed to eliminate such redundancy (example 4).

(4) German

ST 어두운 숲이었어. 아무도 없었어.

It was a dark forest. There was no one.

HT *Ein dunkler Wald. Kein Mensch weit und breit.*

A dark forest. No one in sight.

NMT *Es war eindunkler Wald. Es war niemand da.*

It was a dark forest. There was no one there.

LLM *Es war ein dunkler Wald. Es war niemand da.*

It was a dark forest. There was no one there.

In example 4, the source text consists of two short sentences, “It was a dark forest. There was no one.” The human translation renders this as “Ein dunkler Wald. Kein Mensch weit und breit.” whereas both NMT and LLM use structurally parallel sentences, “Es war ein dunkler Wald. Es war niemand da.” From the perspective of Surprisal as an index of unexpectedness and information density, the human version increases Surprisal through lexical variation and pragmatic enrichment in “Kein

Mensch weit und breit,” which adds idiomatic nuance and enhances reader engagement. In contrast, the machine translations repeat sentence structure and lexis with the highly predictable pattern “Es war [X]. Es war [Y].” which lowers Surprisal and yields grammatically correct but stylistically monotonous and pragmatically reduced output.

The human translator’s ellipsis of the copula in the first sentence and idiomatic intensification in the second can be seen as an effort to balance textual economy with reader involvement, whereas the machine versions prioritize syntactic completeness and semantic clarity, optimizing internal consistency rather than communicative impact.

Conversely, example 5 shows the opposite tendency. The source text uses intentional repetition “I ran and ran” to convey emotional intensity and rhythm, and the human translation preserves this effect with “Ich rannte und rannte,” while NMT (“Ich rannte noch einmal”) reduces recurrence and undertranslates the intensity, and the LLM version (“Ich rannte und weiter”) breaks the symmetrical pattern, weakening stylistic cohesion.⁶ Together these cases show that AI systems often mishandle expressive repetition, either overreducing or unnecessarily altering it, which disrupts cohesion and mismanages informativity in literary contexts.

(5) German

ST 달리고 또 달렸어.

I ran and ran.

HT *Ich rannte und rannte.*

I ran and ran.

NMT *Ich rannte noch einmal.*

I ran one more time.

LLM *Ich rannte und weiter.*

I ran and kept going.

Cohesion devices such as recurrence, ellipsis, and paraphrasing appear to shape

6 The tendency to break the symmetrical pattern of “I ran and ran” could stem either from a structural property of the attention mechanism or from a training bias toward standardized, non-repetitive prose. These two explanations are difficult to disentangle from output analysis alone, as both would produce the same smoothing effect.

how informativity is distributed and modulated, and the observed Surprisal patterns can therefore be interpreted, with caution, in relation to translators' cohesion choices. Human translators seemed to adjust cohesion more sensitively by genre: in literary texts, they used it more flexibly to preserve source text intentionality and literary effect rather than simply maximizing processing ease. This may help explain why word level Surprisal differences reached significance in the literary text. By contrast, NMT and LLM outputs tended to favor clarity and semantic transparency over genre-adaptive cohesion, at times producing redundancy, excessive repetition, and less consistent cohesive structure.

5. Conclusion

This study examined differences in informativity and cohesion between human and AI-generated translations (NMT and LLM) of Korean texts into German across two genres, news articles and literary fiction. By combining Surprisal-based quantitative analysis with qualitative examination of cohesive devices such as recurrence and ellipsis, it provided a genre-sensitive account of divergent translation strategies.

The findings showed that human translators displayed genre awareness in modulating informativity and cohesion. In news texts, they used pro-forms, ellipsis, and structurally economical phrasing to reduce Surprisal and support informational clarity, whereas AI translations favored nominalization and paraphrasing, often introducing redundancy and elevating Surprisal. The distinction drawn here between human translators' economical phrasing and AI systems' more elaborated output is broadly consistent with the translation universal of explicitation, although a clearer boundary between explicitation and mere simplification would require further theoretical refinement, particularly with respect to how each relates to Surprisal-based measures.

In literary texts, human translators preserved stylistic features such as deliberate repetition and ellipsis to maintain rhythm and emotional nuance, while AI systems frequently flattened or disrupted these effects, resulting in maladjusted Surprisal patterns and weaker textual coherence. It should be noted, however, that the proportion of variance attributable to translator type was generally small to medium, and this modest effect size warrants a more cautious interpretation of the claim that

AI systems show limited discourse-level modulation.

Rather than supporting a simple human-versus-machine hierarchy, the findings indicate that differences among HT, NMT, and LLM are genre-sensitive and level-specific, emerging in different ways across sentence-level and word-level organization. Surprisal should not be interpreted as a direct index of translation quality, but rather as a measure of how informational density and predictability are distributed within translated texts.

Theoretically, this small corpus case study contributes to textuality-oriented translation research by empirically linking cohesion strategies to indicators and thus operationalizing notions of informativity and cohesion that have often remained descriptive. Practically, the findings underscore the need for future AI translation systems that are sensitive to cohesion and able to adjust informativity dynamically in line with genre-specific discourse norms, minimizing redundancy in informational texts while preserving rhythmic and affective patterns in literary works.

Several methodological limitations should also be acknowledged. The corpus size is small, and this may limit the statistical power of the ANOVA as well as the generalizability of the findings to broader translation patterns. In addition, Surprisal values were calculated using a GPT-2 model, which may not be fully commensurate, in theoretical terms, with the predictive capacities of the considerably more advanced LLMs whose outputs were evaluated. The qualitative cohesion analysis was furthermore conducted by a single annotator, which ensured consistency but did not allow for the calculation of inter-rater reliability. Future studies incorporating multiple annotators would strengthen the validity of such discourse-level assessments. Finally, while this study distinguished news and literary texts as broad genre categories, a more fine-grained specification of their communicative roles and reader expectations, along with clearer criteria for what constitutes successful informativity within each genre, would further sharpen the qualitative benchmarks employed.

Given the limited corpus, the conclusions are tentative, and future research should broaden the range of languages, genres, and translation directions while combining corpus-based analysis with reader-based comprehension or processing measures to further explore how information-theoretic measures and cohesive patterns jointly shape the reception and cognitive accessibility of translated texts.

References

- Baker, Mona. 2000. Towards a methodology for investigating the style of a literary translator. *Target. International Journal of Translation Studies* 12(2): 241–266.
- Balling, Laura Winther and Johannes Kizach. 2017. Effects of surprisal and locality on Danish sentence processing: An eye-tracking investigation. *Journal of Psycholinguistic Research* 46(5): 1119–1136.
- Choi, Jisoo. 2024. Stilometrische Analyse der menschlichen und maschinellen Übersetzung - mit Fokus auf koreanisch-deutsche Reden. *Koreanische Zeitschrift für Germanistik*, 65(2), 263-282.
- Choi, Jisoo. 2026. Computational stylometric analysis of informativity in human and generative AI translation using surprisal. *Language & Information Society*, 58, 299-321.
- De Beaugrande, Robert-Alain and Wolfgang U. Dressler. 1981. *Introduction to text linguistics* (Vol. 1). London: Longman.
- Degaetano-Ortlieb, Stefania and Elke Teich. 2017. Modeling intra-textual variation with entropy and surprisal: Topical vs. stylistic patterns. In Beatrice Alex, Stefania Degaetano-Ortlieb, Anna Feldman, Anna Kazantseva, Nils Reiter, and Stan Szpakowicz (eds.), *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, 68–77. Vancouver: Association for Computational Linguistics.
- Demberg, Vera and Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109(2): 193–210.
- Donthi, Sundesh, Maximilian Spencer, Om Patel, Joon Yong Doh, Eid Rodan, Kevin Zhu, and Sean O'Brien. 2025. Improving LLM abilities in idiomatic translation. In Kohei Arai (ed.), *Advances in Information and Communication: Proceedings of the 2025 Future of Information and Communication Conference (FICC)*, Volume 3, 361–375. Cham: Springer Nature Switzerland AG.
- Elkins, Katherine. 2024. In search of a translator: Using AI to evaluate what's lost in translation. *Frontiers in Computer Science* 6: 1444021.
- Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*, 1–8. Pittsburgh, PA: Association for Computational Linguistics.
- Han, Kang. 2004. *The vegetarian*. Paju: Changbi.
- Han, Kang. 2016. *Die Vegetarierin*. Translated by Ki-Hyang Lee. Berlin: Aufbau Verlag.
- Jon, Josef and Ondřej Bojar. 2024. An analysis of surprisal uniformity in machine and human translations. In Bram Vanroy, Marie-Aude Lefer, Lieve Macken, and Paola Ruffo (eds.), *Proceedings of the 1st Workshop on Creative-text Translation and Technology*, 40–56. Sheffield: European Association for Machine Translation.

- Kim, Sejung. 2003. Informativity of text and translation—With a focus on words in literary texts. *The Journal of Translation Studies* 4(2): 49–69.
- Lakens, Daniël. 2013. Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology* 4: 863.
- Lee, David. 2002. Genres, registers, text types, domains and styles: Clarifying the concepts and navigating a path through the BNC jungle. In Bernhard Kettemann and Georg Marko (eds.), *Teaching and learning by doing corpus analysis: Proceedings of the Fourth International Conference on Teaching and Language Corpora*, 245–292. Amsterdam: Rodopi.
- Lee, Jae-Won and Xin Zhao. 2023. Informativitätsorientierte Übersetzung—Fokussierung auf die englische, deutsche und chinesische Version von 『Menschenwerk』 von Han Kang. *Textlinguistik* 54: 211–235.
- Lee, Jae-Won and Xin Zhao. 2024. Ein Beitrag zu kohäsionsorientierten Übersetzungen—In Bezug auf die englische, deutsche und chinesische Version von Han Kangs Menschenwerk. *Koreanische Zeitschrift für Deutschunterricht* 89(0): 59–83.
- Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3): 1126–1177.
- Li, Qi. 2024. Bridging languages: The potential and limitations of AI in literary translation—A case study of the English translation of A Pair of Peacocks Southeast Fly. *Advances in Humanities Research* 8: 1–7.
- Martínez, José Manuel and Elke Teich. 2017. Modeling routine in translation with entropy and surprisal: A comparison of learner and professional translations. In Larisa Cercel, Marco Agnetta, and María Teresa Amido Lozano (eds.), *Kreativität und Hermeneutik in der Translation*, 403–426. Tübingen: Narr Francke Attempto Verlag GmbH + Co. KG.
- Monsalve, Irene Fernandez, Stefan L. Frank, and Gabriella Vigliocco. 2012. Lexical surprisal as a general predictor of reading time. In Walter Daelemans (ed.), *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 398–408. Avignon: Association for Computational Linguistics.
- OpenAI. 2025. ChatGPT (GPT-4o) [large language model]. Available at <https://chat.openai.com>.
- Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. OpenAI. Available at <https://cdn.openai.com/better-language-models/language-models.pdf>.
- Rubino, Raphael, Ekaterina Lapshinova-Koltunski, and Josef van Genabith. 2016. Information density and quality estimation features as translationese indicators for human translation classification. In Kevin Knight, Ani Nenkova, and Owen Rambow (eds.), *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 960–970. San Diego, CA: Association for Computational Linguistics.
- Salicchi, Lavinia, Emmanuele Chersoni, and Alessandro Lenci. 2023. A study on surprisal and semantic relatedness for eye-tracking data prediction. *Frontiers in Psychology* 14: 1112365.
- Shannon, Claude Elwood. 1948. A mathematical theory of communication. *The Bell System*

- Technical Journal* 27(3): 379–423.
- Smailagić, Vedad. 2012. Die Präpositionen laut, zufolge und gemäß als Intertextualitätsmarker. *Acta Facultatis Philosophicae Universitatis Ostraviensis/Studia Germanistica* 7(10): 21–35.
- Smith, Nathaniel J. and Roger Levy. 2013. The effect of word predictability on reading time is logarithmic. *Cognition* 128(3): 302–319.
- Weissbart, Hugo, Katerina D. Kandylaki, and Tobias Reichenbach. 2020. Cortical tracking of surprisal during continuous speech comprehension. *Journal of Cognitive Neuroscience* 32(1): 155–166.
- Wilcox, Ethan G., Tiago Pimentel, Clara Meister, Ryan Cotterell, and Roger P. Levy. 2023. Testing the predictions of surprisal theory in 11 languages. *Transactions of the Association for Computational Linguistics* 11: 1451–1470.
- Wolf, Thomas, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In Qun Liu and David Schlangen (eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45. Online. Association for Computational Linguistics.
- Xu, Weijie, Jason Chon, Tianran Liu, and Richard Futrell. 2023. The linearity of the effect of surprisal on reading times across languages. In Houda Bouamor, Juan Pino, and Kalika Bali (eds.), *Findings of the Association for Computational Linguistics: EMNLP 2023*, 15711–15721. Singapore: Association for Computational Linguistics.
- Yan, Jianhao, Pingchuan Yan, Yulong Chen, Jing Li, Xianchao Zhu, and Yue Zhang. 2024. Benchmarking GPT-4 against human translators: A comprehensive evaluation across languages, domains, and expertise levels. *arXiv preprint arXiv:2411.13775*.
- Zhang, Ran, Wei Zhao, and Steffen Eger. 2024. How good are LLMs for literary translation, really? Literary translation evaluation with humans and LLMs. *arXiv preprint arXiv:2410.18697*.

Jisoo Choi

Lecturer

Department of German Language and Literature

Hanyang University

222 Wangsimni-ro, Seongdong-gu,

Seoul, 04763, Korea

E-mail: jsuuch@gmail.com

Received: 2026. 04. 16.

Revised: 2026. 07. 26.

Accepted: 2026. 07. 27.