From Foreignization to Domestication: Quantifying Syntactic Shift in Beowulf Translations

Marten de Jong

s4557182

Abstract

This study aims to show the shift from a synthetic language to an analytic language using Beowulf translations. Comparing translations of William Morris (1895) and Francis Gummere (1909) we analyse SVO frequency, dependency parsing accuracy and pronominal usage in "The watchman's challenge". We hypothesize Morris will exhibit low parseability and non-SVO structures, matching Old English. On the other hand, Gummere is predicted to show higher SVO frequency and parseability, prioritizing readability for more modern readers. These metrics validate the transition in translation strategy from "foreignization" to "domestication".

1 Introduction

As English changed throughout the years, it evolved from a synthetic to an analytic language, where it originally relied on fixed SVO word order it later needed a fixed word order. This resulted in the replacement of inflectional morphology as the primary determinant of grammatical relations by an SVO word order (West, 1973). By the 15th century, non-SVO patterns had largely disappeared from standard prose (Ingham, 2000). This shift poses challenges for translators translating pieces from a foreign source into a domesticated version. Also for translators of *Beowulf*, an Old English poem characterized by the "ornate language" of heroic poetry (Niles, 1993). We speak of "foreignization" when translators choose to preserve archaic structures, staying true to the original source, and "domestication" when they modernize the syntax to align more with the modern reader. Niles (1993) states that translators have to oscillate between these two poles, fidelity and

vigor, constantly.

Translations of Beowulf, due its long history, vary a lot in their approach. Modern translations often prioritize aligning more with the modern reader to improve accessibility. Seamus Heaney (1999) famously sought a "directness of utterance". As Chickering (2002) notes, Heaney described his translation as "about one-third Heaney, two thirds 'duty to the text'", showing a willingness to balance philological fidelity with the creation of an independent work.

19th century translators like William Morris (1895) favoured foreignization, employing obscure vocabulary and inverted syntax often deemed unreadable (Niles, 1993). Gummere (1886) sought a middle ground, maintaining alliterative meter while aiming for greater narrative flow (Crane, 1970).

This study investigates how word order (SVO frequency) and case usage vary between these pre-1900 and post-1900 translations. We hypothesize that Morris will feature a significantly higher frequency of non-SVO order and archaic case usage compared to Gummere, who is expected to "domesticate" while retaining inversions needed to keep the alliterative meter intact.

2 Related Work

2.1 Syntactic shift in English

The transition from Old English to Modern English represents a big shift in syntactic typology. West (1973) demonstrates that Old English, a synthetic language, relied on a rich system of inflectional morphology to determine grammatical relations. This allowed for a lot of flexibility in word order, with Object-Verb (OV) constructions being common, particularly in subordinate clauses.

As English started to lose the inflectional system, nouns and pronouns having distinct case end-

ings, it started to use a word order that was more predictable. Without a fixed SVO order, sentences would become too ambiguous. Ingham (2000) notes that while residual non-SVO patterns, such as negated clauses persisted in specific contexts into the 15th century. The dominant trend was an irreversible move toward the analytic structure of Modern English.

2.2 The translation history of Beowulf

Translating *Beowulf* involves choosing between an archaic, flexible source syntax and the rigid requirements of the target language. Niles (1993) argues that "the task of translation" has historically moved between the two poles: "fully idiomatic current English" and "archaic words". We deem the first to be "domestication" and the latter to be "foreignization".

The late 19th-century approach by William Morris, heavily favoured foreignization. Crane (1970) categorized Morris's work as part of a "literalist" tradition that sought to replicate the strangeness of the original. To achieve this, Morris employed a "word-for-word" translation strategy that retained archaic vocabulary and forced Modern English into Germanic word orders (e.g. OVS). Niles (1993) suggests that while this approach honours the the philological distance of the poem, it often results in a text that is "nearly unreadable" to a non-specialist audience.

In contrast, the early 20th century work of Francis Gummere showed a move towards full modernization. Gummere (1886) argued explicitly against the use of modern blank verse, advocating instead for the preservation of the "original Anglo-Saxon meter" and its "metrical peculiarities". While Crane (1970) evaluates Gummere as a "conscientious" middle ground compared to Morris, Gummere's commitment to the alliterative meter often required syntactic inversions that violate standard modern SVO syntax. Gummere can therefore not be seen as fully domesticated, but as a translator trying to balance metrical fidelity with narrative flow.

2.3 Automated Syntactic Analysis and Stylometry

Researches increasingly use NLP techniques to quantify linguistic shifts in historical texts. Martín Arista et al. (2025) found that neural parsers achieve lower accuracy (LAS 74.23%) on Old English compared to Modern English (∼95%)

due to flexible word order and non-projective dependencies. This performance gap suggests that "parseability" can serve as a metric for modernity. Translations trying to mimic archaic syntax will yield lower accuracy scores.

Parallel to automatic parsing is the field of computational stylometry, which focuses on the statistical frequency of linguistic features to identify stylistic signatures. Yao et al. (2025) demonstrate that quantitative metrics, such as distribution of function words and sentence length, can effectively distinguish between different translation agents (e.g., human and AI). In the context of *Beowulf*, counting specific pronominal case forms (e.g., the ratio of thou/thee to you) functions as a similar stylometric marker. By treating these archaic pronouns as countable features, one can objectively measure the degree of "foreignization" in a translation. Quantifying what Niles (1993) describes qualitatively as the translator's "bias" or "spin"

3 Data

The dataset consists of two translations of *Beowulf*, both available at Project Gutenberg:

- pre-1900 text by William Morris and A.J. Wyatt (1895)

- post-1900 text by Francis B. Gummere (1909)

To ensure comparability, a parallel sample of the "watchman's challenge" was extracted:

- Morris 230-300

- Gummere lines 229-300

Variables We assess the shift from archaic to modern syntax using one independent variable (translation era) and two dependent variables representing syntactic and lexical modernization. Table 1 provides a summary of the data that will be used in this study.

Independent	dependent	dependent
pre-1900	LAS score	ratio archaic to modern
post-1900	LAS score	ratio archaic to modern

Table 1: Overview of the variables that will be measured.

Pre-processing The raw text files were cleaned to remove Project Gutenberg metadata. To ensure the dependency parser functions optimally, the texts were manually segmented into sentences prior to processing. This to address the sentence boundary detection issues common in historical text, as noted by Martín Arista et al. (2025).

To ensure reproducibility of the preprocessing steps and statistical analysis, the full code and datasets used in this study are available at: Beowulf Syntactic Shift"

3.1 Dependency Parsing

To quantify the syntactic "foreignization" of the translations, this study adopts the metric of parseability, defined here as the Labelled Attachment Score (LAS) achieved by a standard Modern English dependency parser. This approach uses the findings of Martín Arista et al. (2025), who established that the flexible word order and non-projective dependencies of Old English significantly degrade the performance of neural parsers.

The parsing pipeline was implemented using the spaCy NLP library. We utilized a transformer-based pipeline (en_core_web_trf), which offers the highest accuracy for Modern English. For the specific text samples extracted (lines 230–300 for Morris and 229–300 for Gummere), a Gold Standard corpus was created by manually correcting the parser's output to reflect the intended syntactic relations of the translators. The final LAS was calculated by evaluating the raw model predictions against the manually annotated ground truth.

4 Predicted Results

We predict significant quantitative differences in the syntactic profiles of the Morris and Gummere translations. A summary is shown in Table 2

Feature	Pre-1900	Post-1900
LAS score	<60%	60-80%
Ratio Pronouns	Archaic	Mixed/Archaic

Table 2: Predicted results

Discussion The results of this study give a measuring tool with which we can view the oscillation of translation strategies described by Niles (1993). By measuring the Parseability (LAS score) and SVO frequency, we can observe the effects of the translator's "bias" or "spin".

The hypothesis that Morris (1895) would show a lower LAS score and non-SVO word orders consistent with Old English syntax is in line with the related work. If Morris's translations achieves low accuracy scores that is in alignment with the results of Martín Arista et al. (2025)'s performance on Old English, this confirms that his syntax follows the "flexible word order" and "non-projective dependencies" of the source language.

While our hypothesis anticipated a high degree of modernization in Gummere, the predicted results likely show a moderate increase in SVO frequency and LAS rather than a total alignment with modern prose. This reflects the goal identified by Crane (1970), where Gummere attempts to maintain the "original metre", a constraint that forces syntactic inversions even when the vocabulary is modernized.

5 Conclusion

This study used syntactic dependency parsing and stylometric analysis to measure the shift from "foreignization" to "domestication" in *Beowulf* translations. The metrics validate the distinction between Morris (1895), whose low parseability scores align with the archaic, synthetic structure of Old English. Gummere (1909), whose higher SVO frequency reflects a movement toward analytic modern syntax.

5.1 Limitations and Future Research

A significant limitation of this study was the reliance of texts available in the public domain via Project Gutenberg. This constraint restricted the analysis to translations, whose copyright has ended, preventing a comparison with fully domesticated modern versions such as those by Chickering (1977) or Heaney (1999).

Furthermore, while our hypothesis framed Gummere as a modernizer, future research must keep his own theoretical stance in mind. In his essay on translation (1886), Gummere explicitly argued against modern blank verse and advocated for the preservation of the "original metre" and its "peculiarities". This suggests that any increase in parseability found in Gummere's translation may be incidental rather than a deliberate attempt at domestication. Future research should expand the dataset to include copyright-protected modern translations and incorporate "gold-standard" human baselines, as suggested by Yao et al. (2025),

to better show the impact of the translator's "bias".

References

Chickering, H. (2002). Beowulf and 'Heaney-wulf'. *The Kenyon Review 24*(1), 160–178. Publisher: Kenyon College.

Crane, J. K. (1970). To Thwack or Be Thwacked: An Evaluation of Available Translations and Editions of Beowulf. *College English 32*(3), 321–340. Publisher: National Council of Teachers of English.

Gummere, F. B. (1886). The Translation of Beowulf, and the Relations of Ancient and Modern English Verse. *The American Journal of Philology 7*(1), 46–78. Publisher: The Johns Hopkins University Press.

Ingham, R. (2000). Negation and OV Order in Late Middle English. *Journal of Linguistics 36*(1), 13–38. Publisher: Cambridge University Press.

Martín Arista, J., A. E. Ojanguren López, and S. Domínguez Barragán (2025, August). Parsing Old English with Universal Dependencies—The Impacts of Model Architectures and Dataset Sizes. *Big Data and Cognitive Computing 9*(8), 199. Publisher: Multidisciplinary Digital Publishing Institute.

Niles, J. D. (1993). Rewriting Beowulf: The Task of Translation. *College English 55*(8), 858–878. Publisher: National Council of Teachers of English.

West, F. (1973). Some Notes on Word Order in Old and Middle English. *Modern Philology 71*(1), 48–53. Publisher: University of Chicago Press.

Yao, X., Y.-B. Kang, and A. McCosker (2025, December). Missing the human touch? A computational stylometry analysis of GPT-4 translations of online Chinese literature. *Translation Spaces 14*(2), 303–330. arXiv:2506.13013 [cs].