Can information theory unravel the subtext in a Chekhovian short story?

Workshop Slavic NLP 2025, co-located with ACL 2025

J. Nathanael Philipp[1] Olav Mueller-Reichau[2] Matthias Irmer[3] Michael Richter[2] Max Kölbl[4]

[1]Sächsische Akademie der Wissenschaften zu Leipzig [2]Leipzig University [3]Digital Science & Research Solutions [4]Osaka University

Text, aim, theory and motivation

Literary texts convey meaning on two levels:

1. explicit content (linguistic form),
2. implicit subtext (world knowledge and context, Gricean inference).

Aim: Quantify subtextual effects in **Anton Chekhov's** story Ward No. 6.

Theory: Information-theoretic analysis using Surprisal and Information Density.

Motivation: Subtext effects are to be demonstrated mathematically.

Anton Chekhov and the story Palata No 6

- Chekhov's short stories are known for containing a lot of subtext.
- Chekhov's style is marked by brevity and a minimalist use of figurative language, inviting an interpretive effort, making his prose ideal for subtext analysis.
- Palata No 6 is a relatively long story (186 paragraphs and 8398 tokens),
- our study is based on the original Russian version.

Hypotheses

To identify the subtextual structure, we enrich the Russian original with glosses, thus modelling implicit content explicitly.

H1: Adding meaningful glosses to the original text reduces surprisal and leads to a well-balanced flow of information,

H2: Adding fake glosses to the original text leads to greater surprisal fluctuations.

Information indices

Surprisal: Reflects information conveyed by a word in a given context. Surprisal s of a word w depends on its conditional probability in a given context:

$$\mathrm{s}(w_i) = -\log_2 P(w_i \mid w_1, \ldots, w_{i-1}, \mathrm{CONTEXT})$$

Where $w_{<i}$ represents co-occurrences, and $CONTEXT$ extra-sentential context that, in this study, is defined as semantic topics, from which semantic surprisal is derived.

Uniform Information Density (UID): wordwise operationalisation of UID.

$$UID = -\frac{1}{n-1}\sum_{i=2}^{n}\left(I(w_i) - I(w_{i-1})\right)^2$$

Information Density and the **UID principle** represents the Flow of Information in sentences. A uneven flow of information in the sentence (extreme information troughs and peaks) can lead to high cognitive processing effort.

The study

Information models:

M1: LLaMA 3.2B (incremental LLM-based surprisal)

M2: TCM (Topic Context Model) (topic-based Bag of Words-surprisal)

Enrichment types:

- no glosses, original text only (OT)
- meaningful glosses (MG)
  - NLP-based (BabelNet glosses)
  - LLM-based (Bison, Gemini)
- Fake glosses (FG)
  - Random sentences from Russian news corpus

Experimental setup:

- 8 conditions: combinations of OT, MG (LLM-enrichment & NLP enrichment), FG with M1 and M2,
- UID distributions computed per sentence,
- comparison across models and enrichment types.

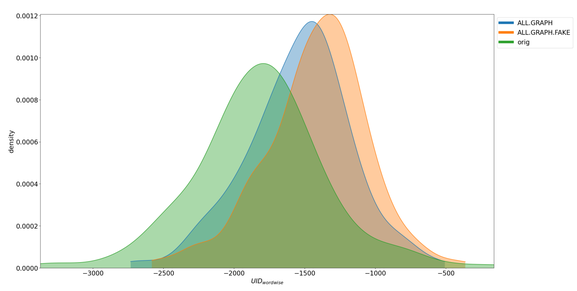
Results

M1 (Llama): shows near-normal UID distributions. MG and fake glosses improve UID compared to OT. H1 is confirmed for both.

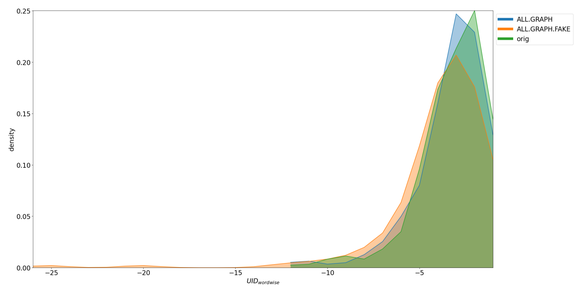
M2 (TCM) : yields peaked, skewed distributions across all conditions. Both MG and fake glosses perform worse than OT. H2 is confirmed. H1 is not.

Gemini + M1: UID improves slightly over OT, positioned between OT and fake glosses.

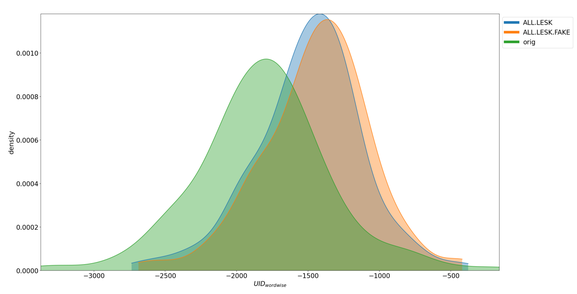
Bison + M2: UID for MG lies between OT and fake glosses, contradicting H2.



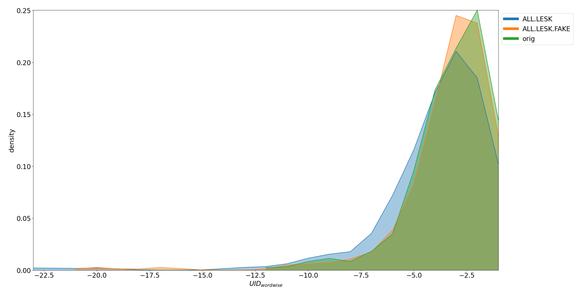
(a) ALL GRAPH LLama (M1)



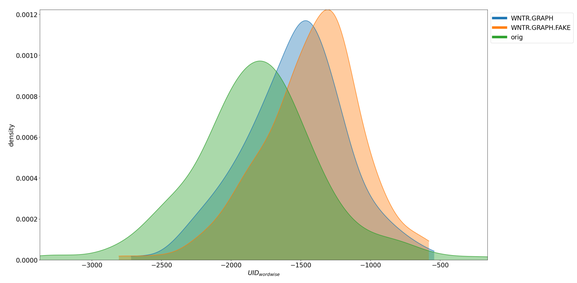
(b) ALL GRAPH TCM (M2)



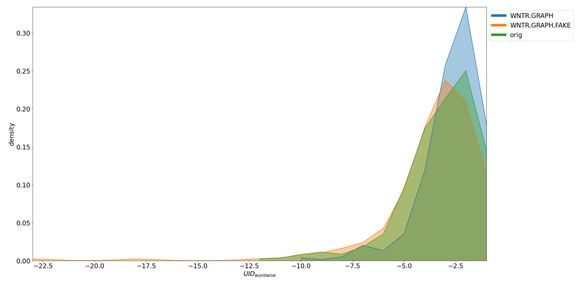
(c) ALL LESK Llama (M1)



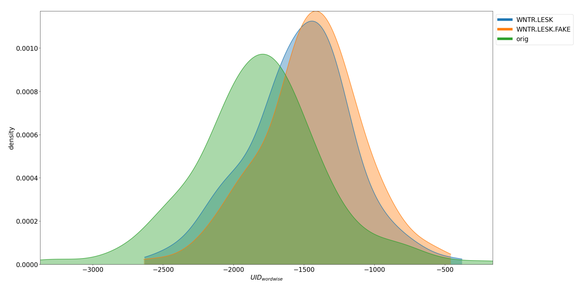
(d) ALL LESK TCM (M2)



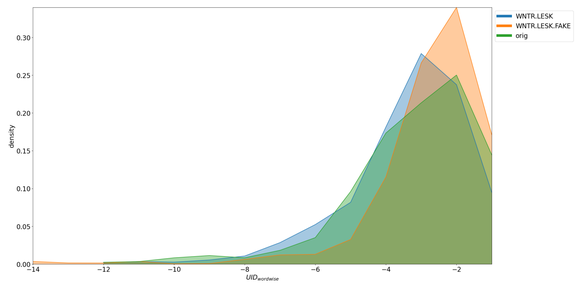
(e) WNTR GRAPH Llama (M1)



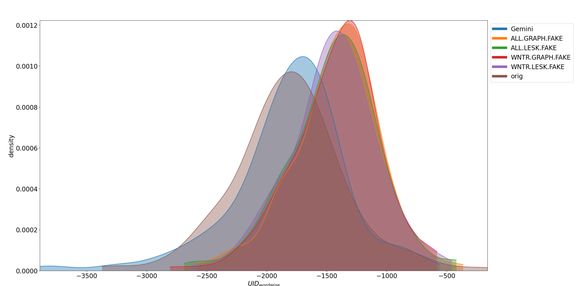
(f) WNTR GRAPH TCM (M2)



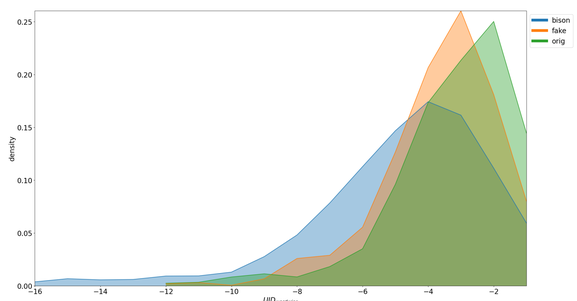
(g) WNTR LESK Llama (M1)



(h) WNTR LESK TCM (M2)



(i) Gemini Llama (M1)



(j) Bison TCM (M2)

Figure: Density plots from UID-distributions.

Conclusion

Under the experimental conditions both with NLP and Bison-glossing and employing M2, our hypotheses could not be confirmed at the same time: H2 turned out to be true, while H1 did not. Meaningful and fake enrichments could be distinguished from OT through UID-distributions, however, both fake-glossing and MG had a lower density of UID than OT.

- Our attempt at adding explicitised subtext does not achieve its intended goal,
- a subtext effect exists but it is overshadowed by the enrichment effect due to (i) a inadequate experimental setup, or (ii) a information density as an adequate measure for quantifying subtext effect.