High Efficiency in Interpreting: Evidence from the Memory-Surprisal Trade-Off

Maria Kunilovskaya, Elke Teich,
Ekaterina Lapshinova-Koltunski, Christina Pollkläsener

Using Corpora in Contrastive and Translation Studies
Hildesheim, Germany

September 09, 2025

Outline

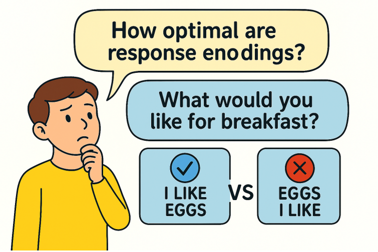
Information theory for language variation studies

Linguistics is concerned with explaining why some options are selected over others.

Do language structures and norms evolve to favour more optimal choices?

Information Theory Hypothesis

Language users are cost-effective:
they optimise their speech by **balancing effort and informativeness**, committing minimum cognitive resources to convey just enough information.



Information-theoretic measures

How can this hypothesis be tested? Which encodings are more optimal?

Experimental psycho- and neurolinguisitc measures:

- N400 Effect, brain response to unexpected words
- acceptability judgement
- reading times and eye-tracking metrics

Information-theoretic measures

How can this hypothesis be tested? Which encodings are more optimal?

Modelling: Measures of information density and distribution

1. **Surprisal**: cognitive processing difficulty (Hale, 2001; Levy, 2008)

> "A word's predictability, as measured by surprisal, is a major determinant of processing difficulty." (Hahn et al., 2022)

Experimental psycho- and neurolinguisitc measures:

- N400 Effect, brain response to unexpected words
- acceptability judgement
- reading times and eye-tracking metrics

Information-theoretic measures

How can this hypothesis be tested? Which encodings are more optimal?

Modelling: Measures of information density and distribution

1. **Surprisal**: cognitive processing difficulty (Hale, 2001; Levy, 2008)

> "A word's predictability, as measured by surprisal, is a major determinant of processing difficulty." (Hahn et al., 2022)

GPT2-based surprisal approximates psycho- and neurolinguisitc measures:

- N400 Effect, brain response to unexpected words (Huber et al., 2024; Michaelov et al., 2024),
- acceptability judgement (Wallbridge et al., 2022),
- reading times and eye-tracking metrics (Jaeger, 2010; Hao et al., 2020).

Information-theoretic measures

How can this hypothesis be tested? Which encodings are more optimal?

Modelling: Measures of information density and distribution

1. **Surprisal**: cognitive processing difficulty (Hale, 2001; Levy, 2008)

   > "A word's predictability, as measured by surprisal, is a major determinant of processing difficulty." (Hahn et al., 2022)

   GPT2-based surprisal approximates psycho- and neurolinguisitc measures:
   - N400 Effect, brain response to unexpected words (Huber et al., 2024; Michaelov et al., 2024),
   - acceptability judgement (Wallbridge et al., 2022),
   - reading times and eye-tracking metrics (Jaeger, 2010; Hao et al., 2020).
2. **Uniform Information Density (UID)**, inc. operationalised as inverse **Average Sentence Surprisal (AvS)**: how efficiently a communication channel is used (Meister et al., 2021).

Information-theoretic measures (cont.)

3. **Mean Dependency Distance (MDD)**, comprehension difficulty

> "the linear distance between syntactically dependent words, ... as an indicator of language processing difficulty and **working memory load** in humans." (Chen et al., 2024)

   trade-off with MHD (Mean Hierarchical Distance, production difficulty) (Jing and Liu, 2015)

4. **Memory-surprisal trade-off (MST)**: how cost-effective the encoding is.

> "A certain level of average surprisal per word can only be achieved at the cost of storing some amount of information about past context." (Hahn et al., 2022)

Application to Translation: Motivation

Mediated language –

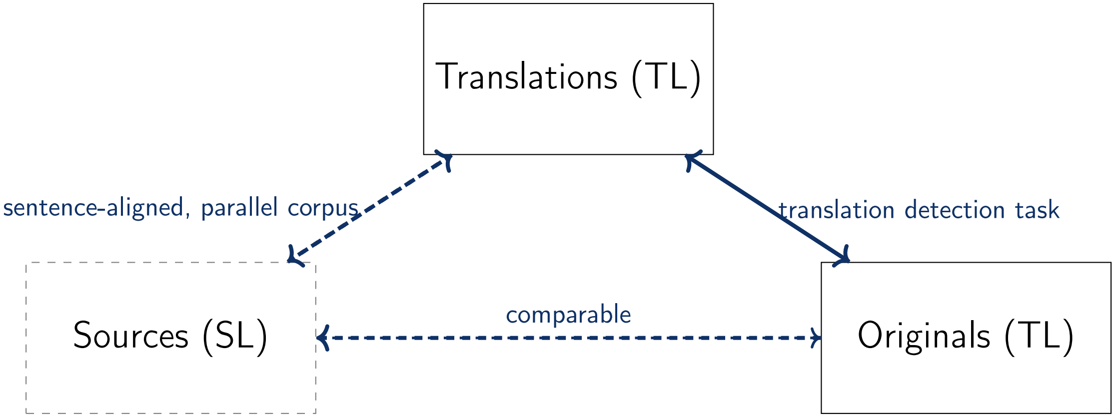
- differs systematically from native TL productions,
- can be regarded as a distinct TL subsystem (Chesterman, 2017).

Information-theoretic assumption: Language production balances *informativeness/clarity* against *cognitive effort/processing cost*.

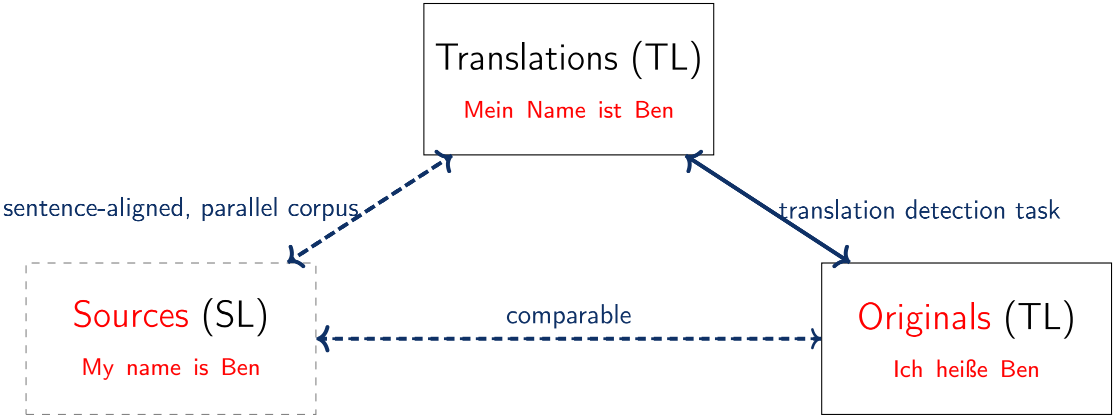
Research Question

Can translated, especially interpreted, language be described as a form of **optimised information management** compared to sources and originals?

Typical comparisons in translation studies



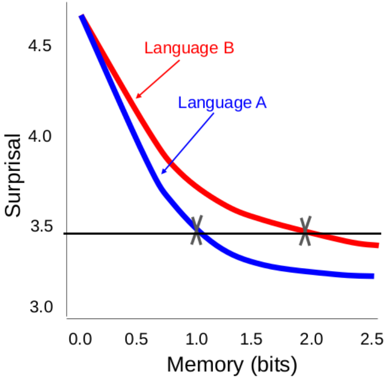
Typical comparisons in translation studies



Outline

Language-level comparisons after Hahn et al. (2021)

MST: A more optimal **Language A** needs less memory to produce the same level of surprisal.



Mathematical model from Hahn et al. (2021)

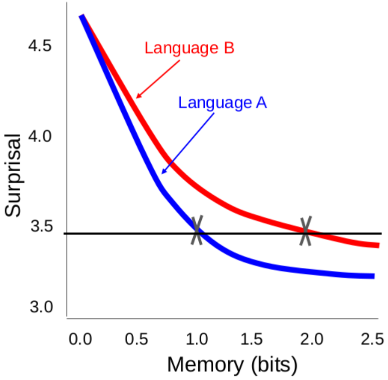
Empirical support:
shorter MDD (for same content) == better MST

1. Humans favour short-term dependencies, i.e. optimised MST,
2. Natural languages have lower MDD and more efficient MST than counterfactual languages with manipulated word order rules.

MST converts context length into bits of information, making it compatible with surprisal.

Language-level comparisons after Hahn et al. (2021)

MST: A more optimal **Language A** needs less memory to produce the same level of surprisal.



Mathematical model from Hahn et al. (2021)

Empirical support:
shorter MDD (for same content) == better MST

1. Humans favour short-term dependencies, i.e. optimised MST,
2. Natural languages have lower MDD and more efficient MST than counterfactual languages with manipulated word order rules.

> MST converts context length into bits of information, making it compatible with surprisal.

Implementation: Steps and Caveats

Goal: Assess the (relative) efficiency of translated/interpreted language against originals.

Language modelling

German and English n-gram models trained with KenLM toolkit (Heafield et al., 2013)
Train data does not intersect with test data.

Training setup decisions

-1- preprocessing: raw vs. lemma vs. lempos (lowercase, tokenised MWE, xxxx for NUMs);

-2- train content: written_orig+tran or written_orig,

-3- type of n-gram boundary: segment vs. document,

-4- testset OOV handling: back-off probability for <unk>,

-5- unigramM=0: Memory for unigram model (N=1) is assumed 0.

Surprisal and Memory Calculations

Surprisal from five n-gram models, one for each $N \in [2, 6]$

"negative log probability of a word in context"

e.g if N=3:

$$S(w_i) = -log2(P(w_i | w_{i-3}, w_{i-2}, w_{i-1})) \tag{1}$$

averaged across all words in the subcorpus/languages variety (AvS).

Memory estimate for each N

$$M_{t+1} = M_t + (t * (S_t - S_{t+1})) \tag{2}$$

e.g if N=3:

$$memory\_rate\_3 = 2 * (AvS_2 - AvS_3)$$
$$memory\_3 = M_2 + memory\_rate\_3$$

Pre-preprocessed text and Resulting data table

Example of preprocessed textual data

Jane Doe visited Paris as well as Rome in 2024.

jane_doe_PROPN visit_VERB paris_PROPN in_ADP xxxx_NUM as_well_as_SCONJ rome_PROPN ._PUNCT

Data table

TL	ttype	ngram_size	surprisal	$memory(St - St + 1)$
de	org	1	9.151	0
de	org	2	7.387	1.764
de	org	3	7.035	2.468
de	org	4	6.966	2.674
de	org	5	6.957	2.712
de	org	6	6.953	2.729
en	tgt	1	9.151	0
en	tgt	2	7.387	1.764
...	...	...	...	...

B7 Data: *German ↔ English*, European Parliament debates

EPIC-UdS (spoken): manually transcribed video recordings of parliamentary speeches and their simultaneous interpreting (Przybyl et al., 2022),

Europarl-UdS (written): transcripts of speeches (adapted for reading) and their written translation from European Parliament website (extracted using Martinez's pipeline),

- L1 originals,
- document- and segment-alignment,
- unified punctuation, spoken features removed,
- balanced for modes and directions,
- lemmatised: Stanza (Qi et al., 2020),

lang	mode	type	docs	segs	tokens
de	wr	org	170	2,796	67,726
		tgt			77,427
	sp	org	165	3,247	56,142
		tgt			49,265
en	wr	org	170	2,790	67,965
		tgt			66,462
	sp	org	137	3,435	64,645
		tgt			46,462

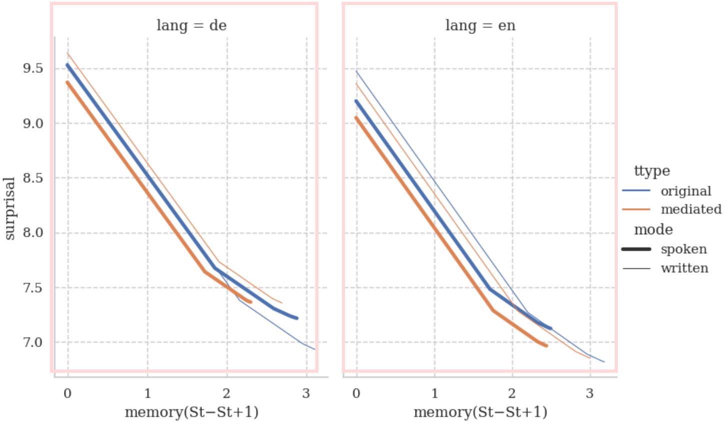
(after filtering and pre-processing)

Written Original Train Data

EN, DE (each): 12 K documents, 3.3 M tokens

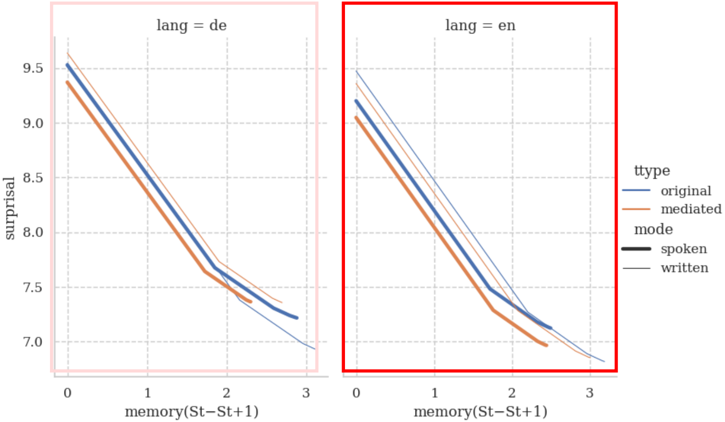
Outline

MST relational plots (a line per variety, org and tgt)



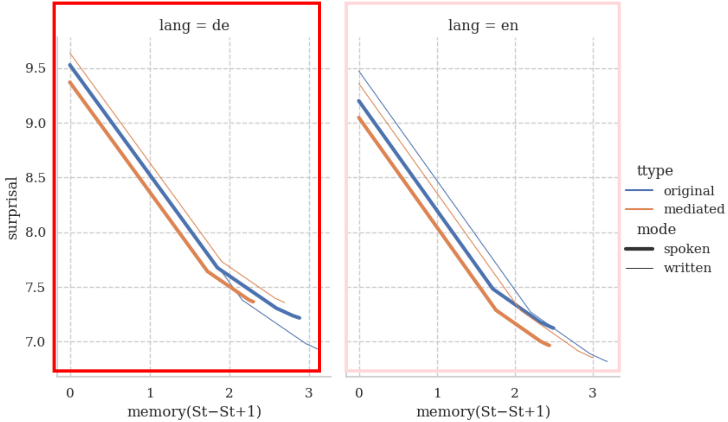
Spoken mediated (**thick brown line**) is the most optimal language variety.

MST relational plots (a line per variety, org and tgt)



EN: mediated (both modes) has better MST (=more optimal) than original, spoken is more optimal than written, both the original or mediated (as expected).

MST relational plots (a line per variety, org and tgt)



DE: interpreting is more optimal than translation and than originals, written translation is least optimal, no difference between original modes (read-out speeches?).

MST: Is mediated language more optimal ...

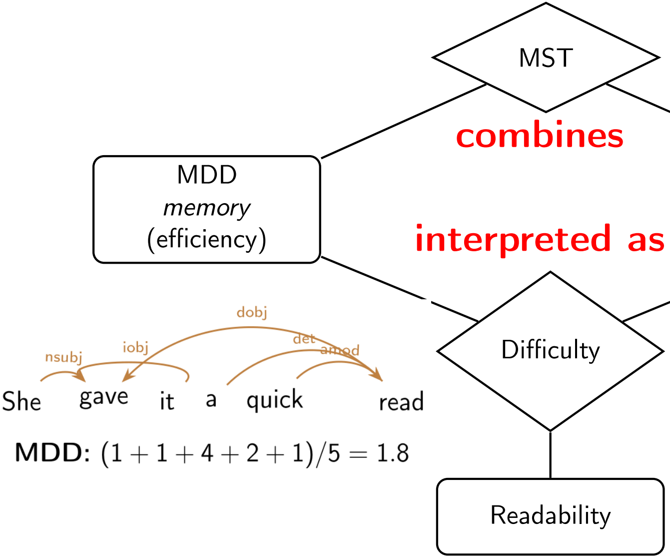
than originals?

TL	mode	measure	outcome
de	sp	MST	✓
	wr		✗
en	sp		✓
	wr		✓

(n-gram KenLM models, trained on 12K original lemposed documents)

Outline

Language efficiency-informativeness descriptive approaches



MDD: $(1+1+4+2+1)/5 = 1.8$

UID as linear function of word surprisal (Meister et al., 2021, p.966)

$$\mathrm{UID}^{-1} = \frac{1}{N}\sum_{n=1}^{N} S(w_i|\mathrm{context})^{k=1}$$

and $\mathrm{AvS} \propto \dfrac{1}{\mathrm{optimality}}$

Cross-checking with alternative approaches (segment level)

Information locality (measured as MDD)

Dependency distance minimisation?

tlang	mode	org	tgt	org-tgt trend
de	sp	3.6	3.4	↘
	wr	3.9	3.9	↗
en	sp	3.0	2.9	↘
	wr	3.2	3.3	↗

Information density (measured as AvS)

Lower AvS, more optimal UID?

tlang	mode	org	tgt	org-tgt trend
de	sp	7.3	7.7	↗
	wr	6.5	6.5	–
en	sp	6.7	6.6	↘
	wr	6.3	6.1	↘

Is mediated language more optimal ...

than originals?

TL	mode	measure	outcome
de	sp	MDD	✓
	wr		✗
en	sp		✓
	wr		✗

TL	mode	measure	outcome
de	sp	AvS / UID	✗
	wr		–
en	sp		✓
	wr		✓

Cross-checking with alternative approaches (segment level)

Information locality (measured as MDD)

Dependency distance minimisation?

tlang	mode	org	tgt	org-tgt trend
de	sp	3.6	3.4	↘
	wr	3.9	3.9	↗
en	sp	3.0	2.9	↘
	wr	3.2	3.3	↗

Information density (measured as AvS)

Lower AvS, more optimal UID?

tlang	mode	org	tgt	org-tgt trend
de	sp	7.3	7.7	↗
	wr	6.5	6.5	–
en	sp	6.7	6.6	↘
	wr	6.3	6.1	↘

Is mediated language more optimal ...

than originals?

TL	mode	measure	outcome
de	sp	MDD	✓
	wr		✗
en	sp		✓
	wr		✗

TL	mode	measure	outcome
de	sp	AvS / UID	✗
	wr		–
en	sp		✓
	wr		✓

Cross-checking with alternative approaches (segment level)

Information locality (measured as MDD)

Dependency distance minimisation?

tlang	mode	org	tgt	org-tgt trend
de	sp	3.6	3.4	$\searrow$
	wr	3.9	3.9	$\nearrow$
en	sp	3.0	2.9	$\searrow$
	wr	3.2	3.3	$\nearrow$

Information density (measured as AvS)

Lower AvS, more optimal UID?

tlang	mode	org	tgt	org-tgt trend
de	sp	7.3	7.7	$\nearrow$
	wr	6.5	6.5	–
en	sp	6.7	6.6	$\searrow$
	wr	6.3	6.1	$\searrow$

Is mediated language more optimal ...

than originals?

TL	mode	measure	outcome
de	sp	MDD	✓
	wr		✗
en	sp		✓
	wr		✗

TL	mode	measure	outcome
de	sp	AvS / UID	✗
	wr		–
en	sp		✓
	wr		✓

Cross-checking with alternative approaches (segment level)

Information locality (measured as MDD)

Dependency distance minimisation?

tlang	mode	org	tgt	org-tgt trend
de	sp	3.6	3.4	↘
	wr	3.9	3.9	↗
en	sp	3.0	2.9	↘
	wr	3.2	3.3	↗

Information density (measured as AvS)

Lower AvS, more optimal UID?

tlang	mode	org	tgt	org-tgt trend
de	sp	7.3	7.7	↗
	wr	6.5	6.5	–
en	sp	6.7	6.6	↘
	wr	6.3	6.1	↘

Is mediated language more optimal ...

than originals?

TL	mode	measure	outcome
de	sp	MDD	✓
	wr		✗
en	sp		✓
	wr		✗

TL	mode	measure	outcome
de	sp	AvS / UID	✗
	wr		–
en	sp		✓
	wr		✓

Cross-checking with alternative approaches (segment level)

Information locality (measured as MDD)

Dependency distance minimisation?

tlang	mode	org	tgt	org-tgt trend
de	sp	3.6	3.4	↘
	wr	3.9	3.9	↗
en	sp	3.0	2.9	↘
	wr	3.2	3.3	↗

Information density (measured as AvS)

Lower AvS, more optimal UID?

tlang	mode	org	tgt	org-tgt trend
de	sp	7.3	7.7	↗
	wr	6.5	6.5	–
en	sp	6.7	6.6	↘
	wr	6.3	6.1	↘

Is mediated language more optimal ...

than originals?

TL	mode	measure	outcome
de	sp	MDD	✓
	wr		✗
en	sp		✓
	wr		✗

TL	mode	measure	outcome
de	sp	AvS / UID	✗
	wr		–
en	sp		✓
	wr		✓

Cross-modality comparison: spoken is more surprising, while having lower MDD than written

Are translations easier to read than originals and sources?

Data and resources:

Pre-trained LM: *xlm-roberta-base*

Readability dataset: Newsela (Xu et al., 2015)

Methods:

- fine-tune NLM to classify sentences based on simplicity
- apply to translational data

than originals?

TL	mode	measure	outcome
de	sp	R-score	✓
	wr		✗
en	sp		✓
	wr		✗

than sources?

TL	mode	measure	outcome
de	sp	R-score	✓
	wr		✗
en	sp		✓
	wr		✓

(Kunilovskaya et al., 2023a)

Are translations easier to read than originals and sources?

Data and resources:

Pre-trained LM: *xlm-roberta-base*

Readability dataset: Newsela (Xu et al., 2015)

Methods:

- fine-tune NLM to classify sentences based on simplicity
- apply to translational data

than originals?

TL	mode	measure	outcome
de	sp	R-score	✓
	wr		✗
en	sp		✓
	wr		✗

than sources?

TL	mode	measure	outcome
de	sp	R-score	✓
	wr		✗
en	sp		✓
	wr		✓

(Kunilovskaya et al., 2023a)

Are translations easier to read than originals and sources?

Data and resources:

Pre-trained LM: *xlm-roberta-base*

Readability dataset: Newsela (Xu et al., 2015)

Methods:

- fine-tune NLM to classify sentences based on simplicity
- apply to translational data

than originals?

TL	mode	measure	outcome
de	sp	R-score	✓
	wr		✗
en	sp		✓
	wr		✗

than sources?

TL	mode	measure	outcome
de	sp	R-score	✓
	wr		✗
en	sp		✓
	wr		✓

(Kunilovskaya et al., 2023a)

Bringing together

Is mediated language more optimal ...

than originals?

TL	mode	MST	MDD	AvS	R-Score
de	spoken	✓	✓	✗	✓
	written	✗	✗	–	✗
en	spoken	✓	✓	✓	✓
	written	✓	✗	✓	✗

Bringing together

Is mediated language more optimal ...

than originals?

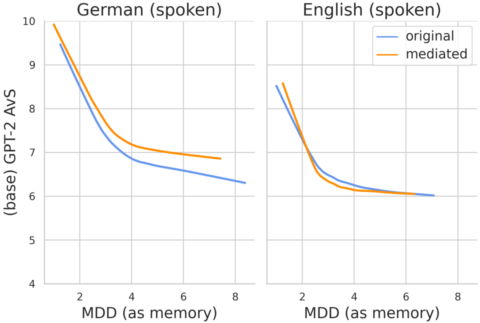
TL	mode	MST	MDD	AvS	R-Score
de	spoken	✓	✓	✗	✓
	written	✗	✗	–	✗
en	spoken	✓	✓	✓	✓
	written	✓	✗	✓	✗

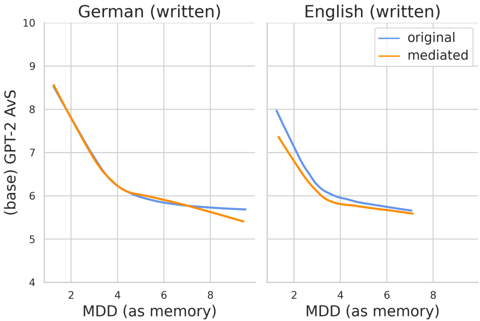
How strong are the claims?

From descriptive approaches
to explanatory and predictive potential.

MST2: MDD∼AvS, Fitted regression lines*

*with locally weighted scatterplot smoothing; AvS from monolingual base GPT-2





TL	ttype	mode	ρ	R^2
de	original	spoken	-0.36	0.14
	mediated	spoken	-0.37	0.15
en	original	spoken	-0.23	0.07
	mediated	spoken	-0.25	0.08

TL	ttype	mode	ρ	R^2
de	original	written	-0.43	0.19
	mediated	written	-0.39	0.18
en	original	written	-0.27	0.08
	mediated	written	-0.26	0.09

Updating

Is mediated language more optimal ...

than originals?

TL	mode	MST	MDD	AvS	R-Score	MST2
de	spoken	✓	✓	✗	✓	✗
	written	✗	✗	–	✗	–
en	spoken	✓	✓	✓	✓	–
	written	✓	✗	✓	✗	✓

Translation detection: SVM classifications results

(best results after feature selection out of 46 predictors, excluding MDD and AvS)

Segment level

TL	mode	F1±SD	feats
de	sp	61.04±1.88	39
	wr	63.05±2.45	23
en	sp	59.00±1.74	22
	wr	57.20±2.20	10

Document level

TL	mode	F1±SD	feats
de	sp	76.13±6.55	21
	wr	86.00±5.54	18
en	sp	76.84±7.08	23
	wr	74.33±7.31	15

Translation detection: SVM classifications results

(best results after feature selection out of 46 predictors, excluding MDD and AvS)

Segment level

TL	mode	F1±SD	feats
de	sp	61.04±1.88	39
	wr	63.05±2.45	23
en	sp	59.00±1.74	22
	wr	57.20±2.20	10

Document level

TL	mode	F1±SD	feats
de	sp	76.13±6.55	21
	wr	86.00±5.54	18
en	sp	76.84±7.08	23
	wr	74.33±7.31	15

Is interpreting more optimal than translation, i.e., blends with originals better?

lower F1 in spoken?

TL	mode	measure	outcome
de	sp∼wr	SVM F1	✓
en	sp∼wr		✗

Translation detection: SVM classifications results

(best results after feature selection out of 46 predictors, excluding MDD and AvS)

Segment level

TL	mode	F1±SD	feats
de	sp	61.04±1.88	39
	wr	63.05±2.45	23
en	sp	59.00±1.74	22
	wr	57.20±2.20	10

Document level

TL	mode	F1±SD	feats
de	sp	76.13±6.55	21
	wr	86.00±5.54	18
en	sp	76.84±7.08	23
	wr	74.33±7.31	15

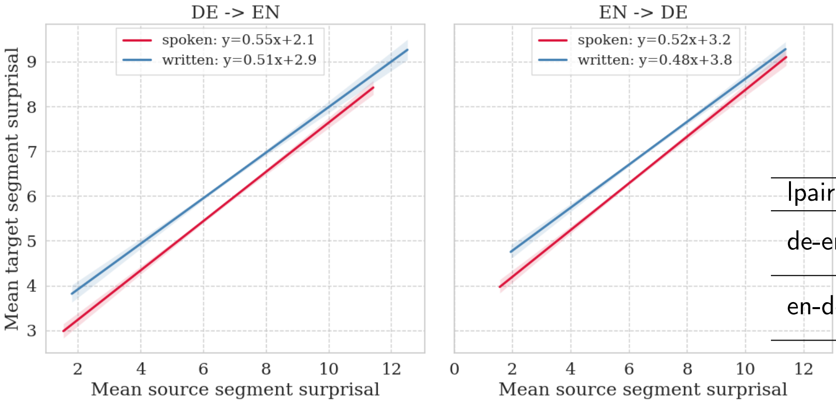
Does interpreting deviate from originals more than translation?

higher F1 in spoken?

TL	mode	measure	outcome
de	sp∼wr	SVM F1	✗
en	sp∼wr	SVM F1	✓

Relation to the source information content: Model's effect

Linear regression lines, predicting mediated AvS from source AvS:
based on in-house n-gram models

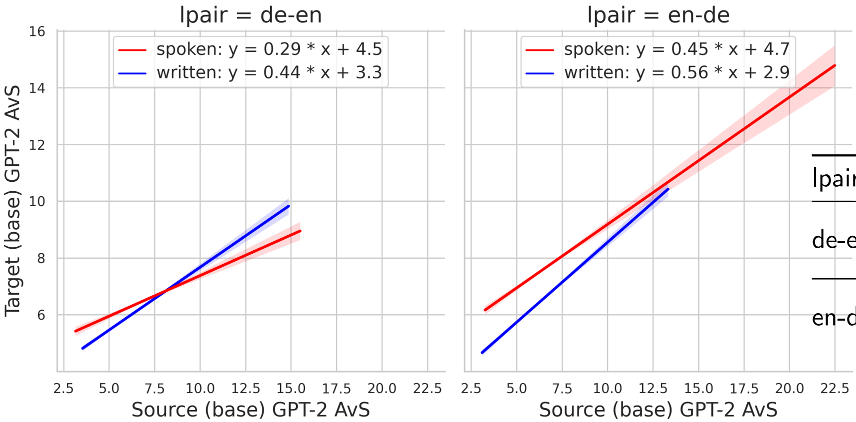


lpair	mode	ρ
de-en	spoken	0.48
	written	0.47
en-de	spoken	0.44
	written	0.51

Information bits are NOT a metric (like kg): a bit from src_GPT2 $\neq$ a bit from tgt_GPT2.

Relation to the source information content: Model's effect

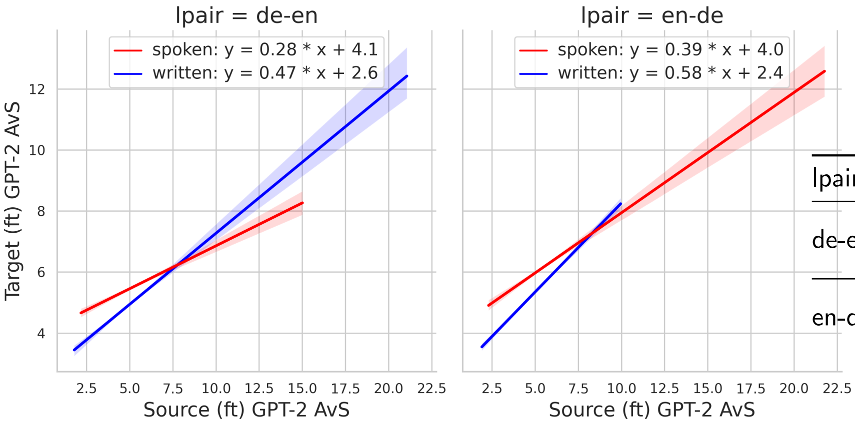
Linear regression lines, predicting mediated AvS from source AvS:
based on monolingual **base** GPT-2 models



lpair	mode	ρ
de-en	spoken	0.33
	written	0.48
en-de	spoken	0.39
	written	0.51

Relation to the **source** information content: Model's effect

Linear regression lines, predicting mediated AvS from source AvS:
based on monolingual **domain-adapted** GPT-2 models



lpair	mode	ρ
de-en	spoken	0.32
	written	0.51
en-de	spoken	0.37
	written	0.52

Outline

Findings about interpreting

1. Interpreting is more efficient: better MST (Hahn et al., 2021) than both originals and translations.
2. Interpreting is more efficient than translation on some attempted alternatives (compared to originals: MDD; compared to originals and sources: readability).
3. Interpreting is more divergent from sources than translation (lower Spearman ρ for src$\sim$tgt AvS), with smaller information gains per unit of information in source (flatter slope for src$\sim$tgt AvS regression).
4. Interpreting is more surprising (AvS), with shorter dependencies (MDD) (directly compared to written).
5. Translation direction asymmetries:
    - mediated German is more source-faithful than mediated English (higher Spearmen ρ src$\sim$tgt),
    - mediated German is more deviant from expected TL than mediated English (F1 scores higher),
    - mediated German shows fewer cross-modality differences than mediated English. Interpreted German is less TL-deviant than translated German, whereas in English the reverse holds (F1 scores).

Findings about interpreting

1. Interpreting is more efficient: better MST (Hahn et al., 2021) than both originals and translations.
2. Interpreting is more efficient than translation on some attempted alternatives (compared to originals: MDD; compared to originals and sources: readability).
3. Interpreting is more divergent from sources than translation (lower Spearman ρ for src$\sim$tgt AvS), with smaller information gains per unit of information in source (flatter slope for src$\sim$tgt AvS regression).
4. Interpreting is more surprising (AvS), with shorter dependencies (MDD) (directly compared to written).
5. Translation direction asymmetries:
    - mediated German is more source-faithful than mediated English (higher Spearmen ρ src$\sim$tgt),
    - mediated German is more deviant from expected TL than mediated English (F1 scores higher),
    - German as a source language triggers problems in English interpreting and leads to more over-polished English translations.

Methodological comments

1. Independent proof of MST concept: Support for inverse relation between information locality and density (MDD∼AvS), the results for MST and MST2 are not the same.
2. Volatility: The information theoretical measures, inc. MST, are sensitive to data balance and preprocessing and underlying models.
3. Interpretability issues: Efficiency is method-dependent. Different operationalisations of *efficiency* (MST, AvS, MDD, R-score) don't always converge. There is no reliable verification for the results, except researchers' expectation and trust in the selected approach. Large room for cherry-picking.
4. There is hardly any experimental evidence of what these measures capture in cross-linguistic production.

Thank you!

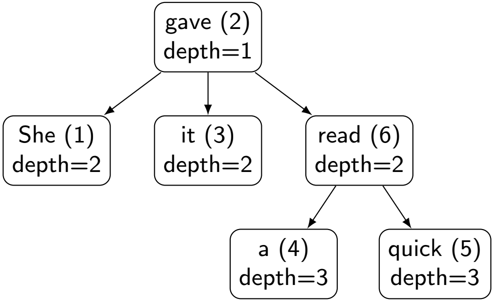
High Efficiency in Interpreting: Evidence from the Memory-Surprisal Trade-Off

SFB 1102 – Information Density and Linguistic Encoding (IDeaL)
funded by the Deutsche Forschungsgemeinschaft, Project ID 232722074

Questions?

Mean Hierarchical Distance Example

She gave it a quick read.



$$\text{Leaf depths} = \{2, 2, 3, 3\}$$

$$\text{MHD} = \frac{2+2+3+3}{4} = \frac{10}{4} = 2.5$$

Outline

References I

Bizzoni, Y., Juzuk, T., Espã na Bonet, C., Dutta Chowdhury, K., van Genabith, J., and Teich, E. (2020). How Human is Machine Translationese? Comparing Human and Machine Translations of Text and Speech. In Proceedings ofthe 17th International Conference on Spoken Language Translation (IWSLT), pages 280–290. Association for Computational Linguistics.

Chen, Y., Zhao, W., Breitbarth, A., Stoeckel, M., Mehler, A., and Eger, S. (2024). Syntactic language change in english and german: Metrics, parsers, and convergences. arXiv preprint arXiv:2402.11549.

Chesterman, A. (2017). Reflections on Translation Theory: Selected papers 1993 – 2014.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., and Stoyanov, V. (2019). Unsupervised cross-lingual representation learning at scale. CoRR, abs/1911.02116.

Degaetano-Ortlieb, S. and Teich, E. (2022). Toward an optimal code for communication: The case of scientific english. Corpus Linguistics and Linguistic Theory, 18(1):175–207.

Hahn, M., Degen, J., and Futrell, R. (2021). Modeling word and morpheme order in natural language as an efficient trade-off of memory and surprisal. Psychological Review, 128(4):726.

References II

Hahn, M., Futrell, R., Levy, R., and Gibson, E. (2022). A resource-rational model of human processing of recursive linguistic structure. Proceedings of the National Academy of Sciences, 119(43):e2122602119.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics, pages 159–166, Stroudsburg, PA. Association for Computational Linguistics.

Hao, Y., Mendelsohn, S., Sterneck, R., Martinez, R., and Frank, R. (2020). Probabilistic predictions of people perusing: Evaluating metrics of language model performance for psycholinguistic modeling. In Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, pages 75–86.

Heafield, K., Pouzyrevsky, I., Clark, J. H., and Koehn, P. (2013). Scalable modified Kneser-Ney language model estimation. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 690–696, Sofia, Bulgaria. Association for Computational Linguistics.

References III

Hu, H. and Kübler, S. (2021). Investigating Translated Chinese and Its Variants Using Machine Learning. Natural Language Engineering, 27(3):339–372.

Huber, E., Sauppe, S., Isasi-Isasmendi, A., Bornkessel-Schlesewsky, I., Merlo, P., and Bickel, B. (2024). Surprisal from language models can predict erps in processing predicate-argument structures only if enriched by an agent preference principle. Neurobiology of Language, 5(1):167–200.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. Cognitive psychology, 61(1):23–62.

Jing, Y. and Liu, H. (2015). Mean hierarchical distance augmenting mean dependency distance. In Nivre, J. and Hajicova, E., editors, Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015), pages 161–170.

Kunilovskaya, M., Lapshinova-Koltunski, E., and Loáiciga, S. (2025a). Information divergence in translation and interpreting: Findings from same-source texts. In Proceedings of the KONVENS 2025 (Konferenz zur Verarbeitung natürlicher Sprache), page in print.

References IV

Kunilovskaya, M., Mitkov, R., and Wandl-Vogt, E. (2023a). Cross-lingual Mediation: Readability Effects. In Štajner, S., Saggio, H., Shardlow, M., and Alva-Manchego, F., editors, Proceedings of the Workshop on Text Simplification, Accessibility, and Readability (TSAR-2023), pages 33–43. INCOMA Ltd., Shoumen, Bulgaria.

Kunilovskaya, M., Przybyl, H., Teich, E., and Lapshinova-Koltunski, E. (2023b). Simultaneous Interpreting as a Noisy Channel: How Much Information Gets Through. In Proceedings of the International Conference on Recent Advances in Natural Language Processing, pages 608–618. INCOMA Ltd.

Kunilovskaya, M., Zaitova, I., Xue, W., Stenger, I., and Avgustinova, T. (2025b). Predictability of microsyntactic units across slavic languages: A translation-based study. In The Joint 25th Nordic Conference on Computational Linguistics and 11th Baltic Conference on Human Language Technologies.

Levy, R. (2008). Expectation-based syntactic comprehension. Cognition, 106(3):1126–1177.

References V

Meister, C., Pimentel, T., Haller, P., Jäger, L., Cotterell, R., and Levy, R. (2021). Revisiting the Uniform Information Density hypothesis. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 963–980, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Michaelov, J. A., Bardolph, M. D., Van Petten, C. K., Bergen, B. K., and Coulson, S. (2024). Strong prediction: Language model surprisal explains multiple n400 effects. Neurobiology of Language, 5(1):107–135.

Nikolaev, D., Karidi, T., Kenneth, N., Mitnik, V., Saeboe, L., and Abend, O. (2020). Morphosyntactic predictability of translationese. Linguistics Vanguard, 6(1):1–12.

Przybyl, H., Lapshinova-Koltunski, E., Menzel, K., Fischer, S., and Teich, E. (2022). EPIC-UdS - creation and applications of a simultaneous interpreting corpus. In Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 1193–1200, Marseille, France. ELDA.

References VI

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., and Manning, C. D. (2020). Stanza: A Python natural language processing toolkit for many human languages. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations.

Rubino, R., Lapshinova-Koltunski, E., and van Genabith, J. (2016). Information density and quality estimation features as translationese indicators for human translation classification. In Proceedings of NAACL HT 2006, pages 960–970, San Diego, California.

Wallbridge, S. C., Lai, C., and Bell, P. (2022). Investigating perception of spoken dialogue acceptability through surprisal. In INTERSPEECH, pages 4506–4510.

Xu, W., Callison-Burch, C., and Napoles, C. (2015). Problems in Current Text Simplification Research: New Data Can Help. Transactions of the Association for Computational Linguistics, 3:283–297.