Cross-Lingual Account of Memory and Surprisal for Interpreting

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

Keywords: interpreting, cognitive load, cross-lingual models, information theory (IT), MST

1. Motivation and Aims

Which parameters of input-output make interpreter struggle the most? Explore performance in interpreting, focusing on:

- cross-lingual modelling: input-output unity is key for an adequate estimate of cognitive load,
- MST approach: limited-resource-based component is particularly relevant for SI studies

2. Research question

🤔 Do IT indices from a cross-lingual model approximate cognitive load (vs monolingual and corpus-based approaches)?

3. Measures of cognitive load in SI and predictors

1. Target text (TT) production time per source word
   ↳ word translation entropy and TT corpus suprisal [1, 2]
2. Number of filled pauses
   ↳ delivery rate, lexical density, numbers, MWE, clauses [3]
3. Length of filled and silent pauses
   ↳ SL problem triggers: (non-)cognates by frequency [4]
4. Types of interpreting (SI, CI)
   ↳ mean dependency distance (MDD) [5]

4. Methodology

Compare the association trends/strength (r, MAE) and explanatory power (R^2) of various features with cognitive load.

IT indices

- monoling. GPT2 surprisal (for source and target),
- MarianMT cross-lingual surprisal,
- cross-lingual surprisal + various definitions of memory

Corpus-based predictors from source and target

- lexical density,
- frequency of numbers,
- frequency of MWE,
- MDD,
- subordinate clauses,
- word length,
- hierarchical distance,
- frequency of unique PROPN,
- branching factor,
- tree depth,
- TTR

Cognitive load

the difficulty that is posed by a task, measured as N of annotated disfluencies (midword breaks, filled pauses, stutters, truncations).

and finally, / hum / I'm [1#I am] seeking to / euh take out / the s/ [s:] the ad/ dition [2#addition] of split and hm separate [s:eperate] v/ ow/ votes [v:otes] [3#] / to [to:] the procedure that will permit / the President to refer euh / back to a [a:] euh / committee, / a r/ f/ f/ f/ f/ f/ report / which has attracted m/ ow/ m/ more [4#] than euh f/ fifty [f:ifty] [2#] / substantive a/ a/ a/ am/ m/ mendments [6#amendments].

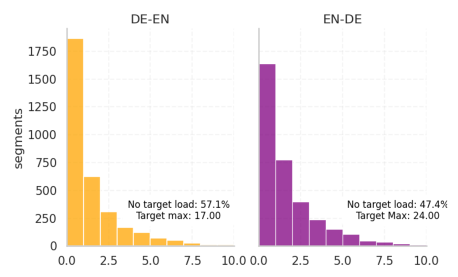
5. Data: EPIC-UdS [6], EN↔DE

	docs	*segs	tok	breaks	filler	stutt	trunc	total
deen_de	165	2,901	64K	376	604	249	134	1,363
deen_en			63K	100	2,340	612	292	3,344
ende_en	137	3,097	71K	92	1,196	612	279	2,179
ende_de			64K	568	3,324	476	195	4,563

*Only ≈50% of segments have disfluencies; NONE excluded

6. Regression setup

- SVR, linear kernel
- Feature selection: RFE
- 10-fold cv



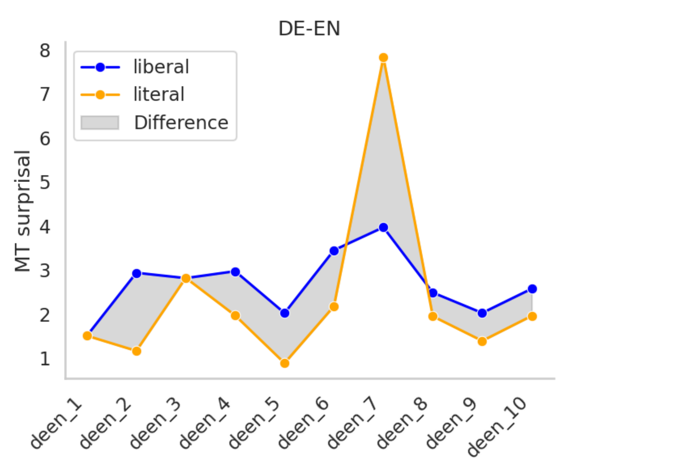
7. SVR Results

	approach	Pearson	MAE	R^2	support
deen	*corpus_src-tgt	0.32±.07	1.29±.14	0.01±.05	1384
ende		0.38±.07	1.34±.15	0.05±.06	1788
deen	srp_gpt2_src-tgt	0.07±.05	1.34±.16	-0.08±.03	1384
ende		0.15±.08	1.42±.14	-0.07±.04	1788
deen	srp_mt	0.14±.05	1.33±.17	-0.07±.04	1384
ende		0.14±.06	1.42±.14	-0.07±.04	1788

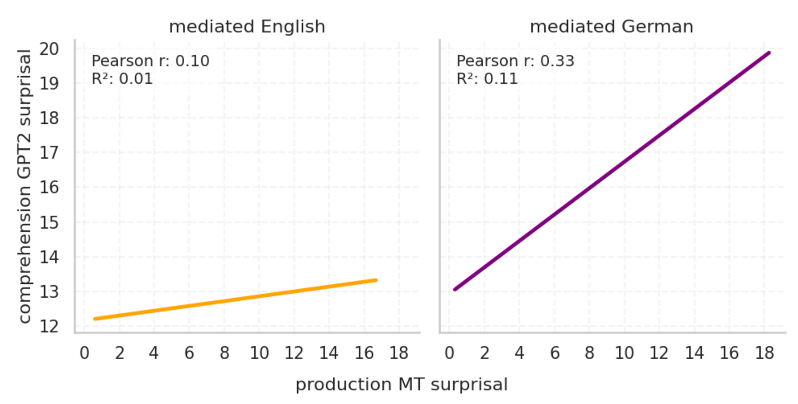
*On the top 5 features (out of 22); all feature perform insignificantly better

8. Notable Insights

MT likes literality (expected)



MT $\propto$ GPT2 surprisal (counter-intuitive)



9. Key Findings

1. The explored properties of input-output are weakly correlated with the cognitive load indexed as frequency of disfluencies.
2. Corpus-based complexity features approximate cognitive load better than IT features.
3. Source text features (esp. mdd, mhd) are more associated with cognitive load than target text features.
4. Cross-lingual approach is the same/better than monolingual.
5. Document level results are better than at segment level.

10. WIP: Memory definitions

Context size [7]

Calculate word-level surprisal (and respective memory) for every N:

$$Mem = \\ = It(1 \to 2) + 2It(2 \to 3) + 3It(3 \to 4),$$
where N=4

N = [0:4] words in ear-voice span

Optimised resource [8]

use attention weights to structure "lossy memory":
retain N context words that are

- most important (highlights),
- most recent (recency),
- most important weighted by distance to node (highlights+)

11. Acknowledgements & References

This research is funded by German Research Foundation (DFG), Project-ID 232722074 – SFB 1102.

[1] Michael Carl and Moritz Jonas Schaeffer. Why translation is diffcult: A corpus-based study of non-literality in post-editing and from-scratch translation. *Hermes (Denmark)*, (56):43–57, 2017.

[2] Yuxiang Wei. *Ambiguity and Entropy in the Process of Translation and Post-editing*. Phd, Dublin City University, 2023.

[3] Koen Plevoets and Bart Defrancq. Imported load in simultaneous interpreting. *Multilingual Mediated Communication and Cognition*, (October 2020):18–43, 2020.

[4] Agnieszka Chmiel, Przemysław Janikowski, Danijel Koržinek, Agnieszka Lijewska, Marta Kajzer-Wietrzny, Dariusz Jakubowski, and Koen Plevoets. Lexical frequency modulates current cognitive load, but triggers no spillover effect in interpreting. *Perspectives*, 0(0):1–19, 2023.

[5] Junying Liang, Yuanyuan Fang, Qianxi Lv, and Haitao Liu. Dependency distance differences across interpreting types: Implications for cognitive demand. *Frontiers in Psychology*, 8(DEC):1–10, 2017.

[6] Heike Przybyl, Ekaterina Lapshinova-Koltunski, Katrin Menzel, Stefan Fischer, and Elke Teich. 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, 20-25 June 2022. ELDA.

[7] Michael Hahn, Judith Degen, and Richard Futrell. Modeling word and morpheme order in natural language as an efficient trade-off of memory and surprisal. *Psychological Review*, 128(4):726, 2021.

[8] Michael Hahn, Richard Futrell, Roger Levy, and Edward Gibson. A resource-rational model of human processing of recursive linguistic structure. *Proceedings of the National Academy of Sciences*, 119(43):e2122602119, 2022.