XVII DAMSS • 26–28 Nov 2026
Druskininkai, Lithuania | Research poster

Human-Language Model Alignment in Native and Foreign Language Processing: An Eye-Tracking Study of Cognitive Effort

Błażej Białczak, Kamil Sterniuk, Bożena Kostek

Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Poland
Multimedia Systems Department and Audio Acoustics Laboratory
Email: bozena.kostek@pg.edu.pl

Background

Understanding how cognitive effort differs between native-language (L1) and foreign-language (L2) reading is important for adaptive language-learning systems and cognitively informed NLP. Eye-tracking can capture human reading effort, while Transformer-based models provide computational indicators such as surprisal, attention-based measures, token importance, and attribution.

Objective

Understanding how cognitive effort differs between native (L1) and foreign (L2) language processing is important for developing adaptive language-learning technologies and cognitively informed natural language processing (NLP) systems. This study investigates the relationship between eye-tracking measures of cognitive effort and computational indicators derived from NLP models, focusing on reading in Polish (L1) and English (L2).

Method

- **Participants:** 26 Polish-speaking students aged 19–25; session lengths differed.
- **Materials:** Polish (L1) and English (L2) reading passages. The final gaze–model correlation synthesis included 8 texts (4 per language).
- **Procedure:** participants read passages aloud and answered questions verbally in selected QA trials; gaze recorded with Tobii and a Python application.
- **Eye tracking:** word-level fixation counts, durations, dwell times, regressions and gazeplots.
- **Models:** HerBERT / XLM-R, mDeBERTa, Polish and English GPT-2 variants, and BERT / RoBERTa QA models.
- **Analysis:** token scores aligned to words; matched gaze–model measures compared via Spearman ρ. QA attribution analysed separately.

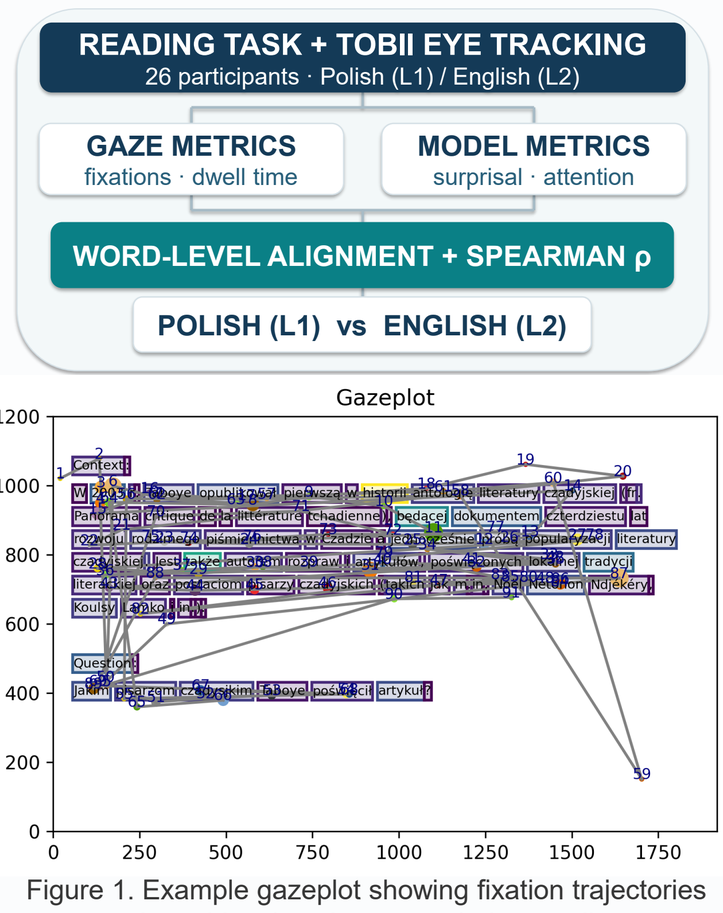

Figure 1. Example gazeplot showing fixation trajectories and word-level assignments during reading.

Results

- **Surprisal showed the strongest agreement** with eye-tracking measures, indicating that less predictable words tended to receive more visual attention.
- **The alignment was markedly stronger for Polish (L1)** than for English (L2).
- **Attention-based predictors** showed meaningful but language-dependent associations.
- **Token masking** produced less stable associations, while NLI and QA attribution showed limited correspondence with gaze behavior.

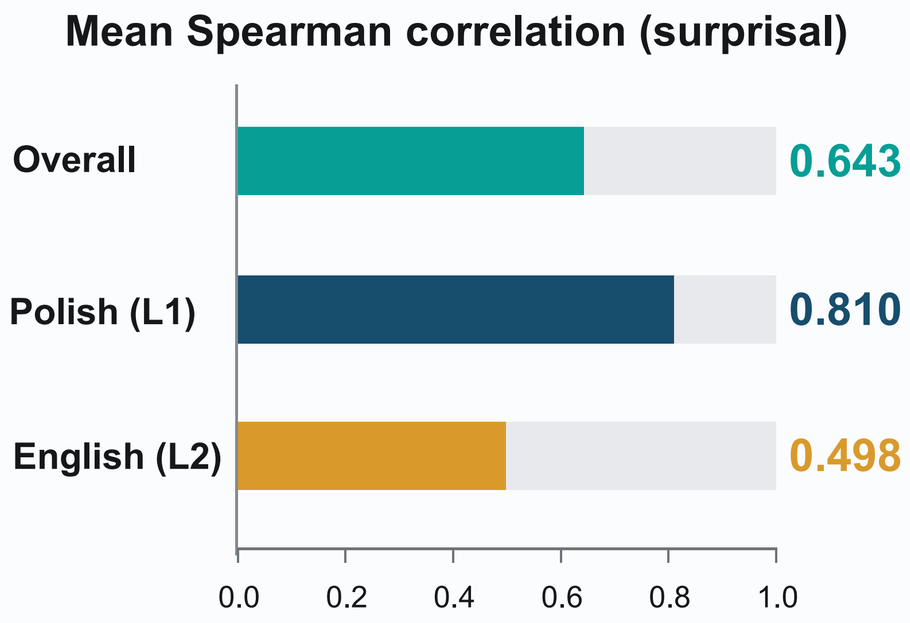

Figure 2. Mean Spearman correlations between language-model surprisal and eye-tracking measures.

Polish texts (n=4)

	Δentr	EMD-R	Manh.-R	Mask	NLI	Surpr.
fix. count	0.63	0.69	**0.78**	0.34	0.21	**0.85**
fix. time	0.58	0.68	**0.73**	0.34	0.21	**0.82**
dwell time	0.60	0.68	**0.77**	0.32	0.18	**0.85**
dwell nonzero	0.53	0.46	0.64	0.17	0.22	**0.70**

English texts (n=4)

	Δentr	EMD-R	Manh.-R	Mask	NLI	Surpr.
fix. count	0.17	0.14	0.17	0.40	-0.15	0.51
fix. time	0.16	0.15	0.16	0.41	-0.14	0.50
dwell time	0.17	0.14	0.17	0.42	-0.15	0.53
dwell nonzero	0.08	0.02	0.09	0.35	-0.13	0.40

Figure 3. Selected gaze–model Spearman ρ values (4 gaze measures shown; Figure 2 summarizes the wider set).
Δentr = change in attention entropy; EMD-R and Manh.-R = distances between model attention distributions (R variant).

Discussion

In this dataset, language-model surprisal provided the most useful approximation of selected aspects of human reading effort. The stronger alignment observed for Polish may reflect language-specific properties, text composition, tokenization, or model differences. Model attention should not be interpreted as equivalent to human cognitive attention.

Study Scope and Limitations

- Small, homogeneous sample (26 students).
- Final correlation synthesis based on 8 texts.
- Reading aloud and verbal QA responses can influence eye-movement patterns.
- Language-specific GPT-2 variants, text selection and tokenization may confound L1/L2 comparisons.
- Correlations indicate association, not identical processing mechanisms in humans and models.

Conclusions

Computational measures of linguistic predictability, especially surprisal, can serve as promising proxies for selected aspects of reading effort.

The integrated eye-tracking and NLP framework may help identify difficult text fragments, support adaptive learning environments, and guide the selection of educational materials. Further validation on larger and more diverse groups is needed.

References

[1] Vaswani et al. (2017). Attention Is All You Need. NeurIPS.

[2] Devlin et al. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL.

[3] Oh & Schuler (2022). Entropy- and Distance-Based Predictors From GPT-2 Attention Patterns Predict Reading Times Over and Above GPT-2 Surprisal. EMNLP, 9324–9334.

[4] Holmqvist et al. (2011). Eye Tracking: A Comprehensive Guide to Methods and Measures. Oxford University Press.

Ethics approval: 1 July 2026 · Informed consent.