Translation as Efficient Meaning Compression: An Information Bottleneck Analysis of Spatial Terms

Antoine Taroni, Ludovic Moncla, Frederique Laforest
`first.last@liris.cnrs.fr`
INSA Lyon, CNRS, Lyon 1 Université, LIRIS, UMR5205,
69621 Villeurbanne, France

1 Introduction

The semantic typology of the lexicon describes how differently languages tend to group and dub concepts. Converging evidence indicates that speakers may face pressure toward efficient communication when describing stimuli from the surrounding world (Kemp et al., 2018). The Information Bottleneck (IB) framework (Zaslavsky et al., 2018) accounts for this optimization problem. It has been tested in various semantic domains (Khetarpal et al., 2013; Langlois et al., 2025). In such experiments, descriptions of pictures or videos are elicited, and human similarity judgments are collected. The communicative cost of the description is compared to the information it conveys about the stimuli, with an information-theoretic approach. But so far, we observe at least one of the following limitations: the meanings of interest are established prior to the experiment (Chen et al., 2023), or lexical items are examined in isolation rather than in natural sentences or discourse (Zaslavsky et al., 2018).

We propose to frame the act of translation as a constrained optimization problem. We assume that human translators aim for an optimal translation, without any change in meaning. To study the compression of meanings in translation, we propose to use bitexts. Each sentence to translate is considered a stimulus, which triggers its own unique mental representation. We define this problem in terms of a trade-off between Accuracy and Complexity, as posited in the IB framework. Aforementioned limitations are addressed because the meaning space is not designed prior to the experiment; instead, it is derived from naturally occurring discourse.

We tested this new method on a cornerstone of spatial language: the prepositions that overtly mark for spatial relations. These terms are indeed highly prone to engage spatial representations (Landau and Jackendoff, 1993). We first extracted cross-linguistic variations of these spatial terms in four languages, building on Viechnicki et al. (2024) and Beekhuizen (2025). Also, we relied on an ersatz for human similarity judgments: the cosine similarity between contextual embeddings (Marjieh et al., 2022). From the perspective of the IB principle, preliminary analyses show a higher efficiency of natural translations compared to random systems.

2 Methodology and Experiments

We use `EN80jours` (LIFAT et al., 2024), a parallel corpus featuring English, German and Serbian translations of the French novel *Le Tour du monde en quatre-vingt jours*. Starting with the source text, we collect occurrences of prepositions from the inventory established by Le Pesant (2012). Initial disambiguation is carried out using the LLM `mistral-large-2512` (Mistral AI, 2025) in a few-shot setting. Three annotators subsequently reviewed the output and retained terms with a spatial sense [1]. We evaluated an automatically produced alignment in a similar experimental setup by sampling 100 aligned term pairs for each target language. The resulting precision scores were 81% (French–English), 91% (French–German), and 94% (French–Serbian), leading to 608 French spatial terms aligned in all three target languages[2] (illustrated in Figure 1.A).

In order to account for the psychological meaning space of spatial relations, we introduce a new substitute for human similarity judgments in IB experiments: the cosine distance between contextual embeddings of the spatial terms. We embedded each spatial term with `xlm-roberta-large` (Conneau et al., 2020), averaged and stacked the last four hidden layers, and z-scored (Timkey and van Schijndel, 2021) these embeddings before comput-

---

[1] Whenever the context triggered concrete spatial mental imagery, involving entities that can be touched or pointed at.

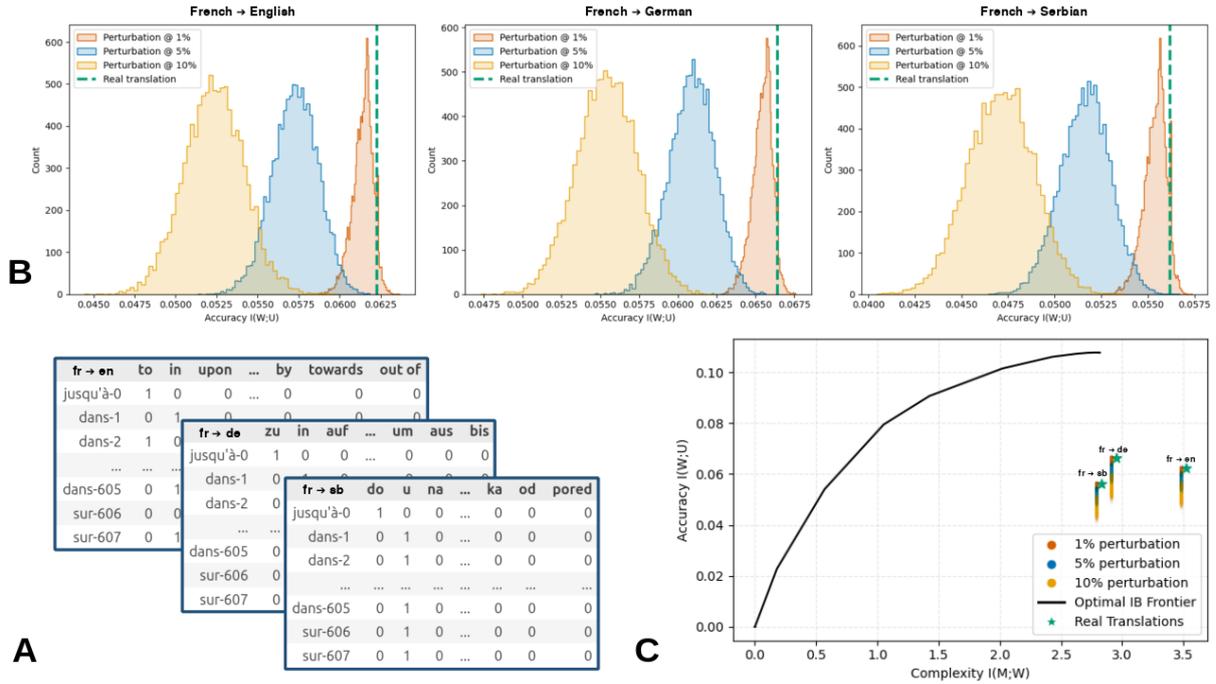[2] Data and code are available at `https://github.com/antoineOYO/spatial-terms-translations`

Figure 1: Efficiency analysis of spatial term translations. **A**: overview of the three contextual alignment tables we obtained. Rows are the samples, columns the target spatial terms. **B**: Accuracy of spatial terms translations (vertical dashed line), compared with 30 000 randomly altered translations. **C**: Accuracy versus Complexity information plane, featuring the theoretical frontier, real and random translations. At a given Complexity, dots are jittered.

ing the cosine similarity. We obtained a symmetric similarity matrix.

Theoretical Framework. Each spatial term from the source text, with its context, is an instance from the world U. The meaning space M of spatial relations has the same dimension as U: 608. W is the lexicon of the target language: $|W_{en}| = 21$, $|W_{de}| = 16$ and $|W_{sb}| = 19$. The pressure for efficiency can be defined in terms of Accuracy and Complexity of the description. Complexity accounts for how unpredictable the linguistic encoder is. A natural measure for this quantity is the Mutual Information between the two distributions of meanings m and words w: $I(M : W)$. Accuracy represents how much ambiguity is removed by a description, it must therefore compare the uncertainty about plausible worlds states u to the uncertainty of the linguistic encoder: $I(U : W)$. The IB framework states that optimal compression of meaning is achieved when the Accuracy is maximized for a given Complexity.

We computed the Accuracy and Complexity for our three pairs of source and target languages, using the implementation of Zaslavsky et al. (2018); Chen et al. (2023). We also generated "random translators" by permuting[3] 1, 5 and 10% of the alignment tables rows (Figure 1.B). The optimal frontier was computed with the Blahut-Arimoto algorithm from Tishby et al. (2000) (Figure 1.C).

3 Discussion

Real translations lie closer to the optimal systems than the majority of their slightly altered variants. These results are in line with the Information Bottleneck principle applied to natural language. Along with the new methodology, they call together for thorough investigations on the interplay between translation, spatial representations and the IB principle. We intend to quantify the deviation from the optimal bound. We also plan on comparing the computational similarity metrics derived from our model against human semantic judgments to quantify the alignment between machine-learned and cognitive representations. Other baselines than randomly altered translations could be considered, like dictionary lookup (Strickland and Chemla, 2018). We aim to replace LLM pre-annotators with dedicated alignment tools, and extend the work to other types of corpora than a single adventure novel.

---

[3]Fully random alignments (not presented here) yielded accuracies close to zero.

References

Barend Beekhuizen. 2025. Spatial relation marking across languages: extraction, evaluation, analysis. In *Proceedings of the 29th Conference on Computational Natural Language Learning*, pages 571–585, Vienna, Austria. Association for Computational Linguistics.

Sihan Chen, Richard Futrell, and Kyle Mahowald. 2023. An information-theoretic approach to the typology of spatial demonstratives. *Cognition*, 240:105505.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised Cross-lingual Representation Learning at Scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Charles Kemp, Yang Xu, and Terry Regier. 2018. Semantic Typology and Efficient Communication. *Annual Review of Linguistics*, 4(1):109–128.

Naveen Khetarpal, Grace Neveu, Asifa Majid, Lev Michael, and Terry Regier. 2013. Spatial terms across languages support near-optimal communication: Evidence from Peruvian Amazonia, and computational analyses. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 35. Issue: 35.

Barbara Landau and Ray Jackendoff. 1993. "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16(2):217–238.

Thomas A. Langlois, Roger Levy, Nidhi Seethapathi, and Noga Zaslavsky. 2025. Efficient compression in locomotion verbs across languages. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 47.

Denis Le Pesant. 2012. Essai de classification des prépositions de localisation. In *3ème Congrès Mondial de Linguistique Française (CMLF 2012)*, pages 921–936.

LIFAT, LLL, and Université de Belgrade. 2024. En80jours. ORTOLANG (Open Resources and TOols for LANGuage) –www.ortolang.fr.

Raja Marjieh, Ilia Sucholutsky, Ted Sumers, Nori Jacoby, and Tom Griffiths. 2022. Predicting human similarity judgments using large language models. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 44.

Mistral AI. 2025. Introducing mistral 3. Mistral Documentation.

Brent Strickland and Emmanuel Chemla. 2018. Cross-linguistic regularities and learner biases reflect "core" mechanics. *PLOS ONE*, 13(1):e0184132. Publisher: Public Library of Science (PLoS).

William Timkey and Marten van Schijndel. 2021. All bark and no bite: Rogue dimensions in transformer language models obscure representational quality. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4527–4546, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Naftali Tishby, Fernando C Pereira, and William Bialek. 2000. The information bottleneck method. *arXiv preprint physics/0004057*.

Peter Viechnicki, Kevin Duh, Anthony Kostacos, and Barbara Landau. 2024. Large-Scale Bitext Corpora Provide New Evidence for Cognitive Representations of Spatial Terms. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1089–1099, St. Julian's, Malta. Association for Computational Linguistics.

Noga Zaslavsky, Charles Kemp, Terry Regier, and Naftali Tishby. 2018. Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, 115(31):7937–7942.