

Translation from the Information Bottleneck Perspective: an Efficiency Analysis of Spatial Prepositions in Bitexts

Antoine Taroni, Ludovic Moncla, Frédérique Laforest

INSA Lyon, CNRS, Lyon 1 Université, LIRIS, UMR5205,
69621 Villeurbanne, France
first.last@insa-lyon.fr

Abstract

Efficient communication requires balancing informativity and simplicity when encoding meanings. The Information Bottleneck (IB) framework captures this trade-off formally, predicting that natural language systems cluster near an optimal accuracy-complexity frontier. While supported in visual domains such as colour and motion, linguistic stimuli such as words in sentential context remain unexplored. We address this gap by framing translation as an IB optimisation problem, treating source sentences as stimuli and target sentences as compressed meanings. This allows IB analyses to be performed directly on bitexts rather than controlled naming experiments. We applied this to spatial prepositions across English, German and Serbian translations of a French novel. To estimate informativity, we conducted a pile-sorting pilot-study ($N = 35$) and obtained similarity judgements of pairs of prepositions. We trained a low-rank projection model ($D = 5$) that predicts these judgements (Spearman $\rho = 0.78$). Attested translations of prepositions lie closer to the IB optimal frontier than counterfactual alternatives, offering preliminary evidence that human translators exhibit communicative efficiency pressure in the spatial domain. More broadly, this work suggests that translation can serve as a window into the cognitive efficiency pressures shaping cross-linguistic semantic systems.

Keywords: Information Theory, Spatial Language, Spatial Cognition, Information Bottleneck, Linguistic Efficiency, Translation

1. Introduction

Languages partition semantic domains in remarkably diverse ways. For example, [Levinson, 2003](#) and [Levinson et al., 2003](#) intensively document how spatial relations are named differently by speakers of typologically diverse languages. Yet beneath this diversity, systematic regularities emerge when naming spatial relations, but also within the semantic domains of colour and kinship *inter alia* ([Kemp et al., 2018](#)). One influential standpoint invokes the principle of efficient communication to explain these regularities ([Gibson et al., 2019](#); [Kemp et al., 2018](#); [Zaslavsky et al., 2018](#)). According to this view, when speakers of the same language balance the informational and cognitive load in their utterances, the mapping from meanings to words is optimised with respect to that trade-off between informativity and simplicity. Languages around the world achieve different realisations of this trade-off through evolution ([Kirby et al., 2015](#)).

The Information Bottleneck (IB) framework ([Tishby et al., 2000](#)) was put forward by [Zaslavsky et al., 2018](#) to operationalise this trade-off for natural languages. It states that languages achieve near-optimal compression of meanings from an information-theoretic perspective, and has recently gained substantial empirical support. Usually, an IB analysis focuses on a specific semantic domain, and relies on naming data (e.g. natural descriptions of stimuli such as pictures or videos). The underlying assumption is that a stimulus evokes

a mental representation, which is efficiently compressed into a description. The communicative cost of the description (complexity) is then compared to the relevant information it conveys about the stimuli (informativity), across natural and hypothetical languages. Prior IB applications relied on various stimuli like colour chips ([Zaslavsky et al., 2018](#)), static visual referents from images ([Gualdoni et al., 2023](#)), and kinematic events in videos of human locomotion ([Langlois et al., 2025](#)). However, to our knowledge, no IB analysis has considered linguistic stimuli.

We propose to regard a word in context as a stimulus in its own right and to perform an IB analysis in the semantic domain of spatial relations. To do so, we use a novel type of data: bitexts, e.g. translations of source sentences into target ones (illustrated in [Table 1](#)). We frame the act of translation as a constrained optimisation problem in the IB sense, and assume that human translators aim for exact meaning preservation. A word of interest from a source sentence is then considered a stimulus, which triggers its own mental representation, which is then linguistically encoded in the target language.

We tested this new methodology on a cornerstone of spatial language: spatial prepositions. These terms overtly mark for spatial relations between entities, and are highly prone to engage spatial representations ([Landau and Jackendoff, 1993](#)). We first extracted cross-linguistic variations of these spatial terms in four languages, namely detecting

id	French	English	German	Serbian
n1909	le pilote [...] prit place sur la passerelle	The pilot [...] took his place on the bridge	der Lootse [...] und nahm seinen Platz auf dem Steg	pilot [...] zauze mesto na komandnom mostu
n3688	Mudge [...] se maintenait dans la ligne droite	Mudge [...] kept in a straight line	Mudge [...] hielt sich in gerader Linie	Mody se drzxao prave linije
n3110	[...] il reprit la direction de l'est à travers le massif accidenté des monts Wahsatch.	[...] it took an easterly direction towards the jagged Wahsatch Mountains.	[...] nahm er wieder östliche Richtung durch den unregelmäßigen Hauptstock des Wahsatchgebirges.	[...] krenuo je prema istoku preko masiva planine Vesecy.
n2481	Quelques matelots [...] descendirent le pauvre garçon dans une cabine des secondes	Several sailors [...] carried the poor Frenchman down into the second cabin	Einige Matrosen [...] trugen den armen Jungen in eine Cabine zweiten Ranges	Nekoliko mornara [...] odnesosxe jadnog momka u jednu kabinu druge klase
n1243	le guide dirigerait l'éléphant vers la pagode de Pillaji	the guide should direct the elephant towards the pagoda of Pillaji	der Führer solle den Elephanten nach der Pagode Pillaji leiten	vodicy uputi slona ka pagodi u Piladyi

Table 1: Excerpt of (LIFAT et al., 2024). Spatial prepositions in the French source sentences are emphasised. `id` refers to the original identifier as released in the corpus.

occurrences in source sentences from LIFAT et al., 2024, and aligning them with their target counterparts. This allowed us to assess the complexity of spatial prepositions translations. Furthermore, to estimate the informativity of these translations, we also modelled the psychological space of spatial relations by leveraging pile-sorting results from a pilot study ($N = 35$). We show that a low-dimensional space suffices to predict similarity judgements. Ultimately, we provide preliminary evidence for greater communicative efficiency of natural translations compared to random and counterfactual systems, in the semantic domain of spatial relations.

2. Theoretical framework

This section outlines how the IB theory has been applied to language. Tracing back to the Rate-Distortion theory (Cover, 1999), communication between a speaker and a listener is formalised as a joint search for an encoder and a decoder that enable lossy data compression. In this schematic communicative scenario, the speaker seeks a linguistic encoder q that maps meanings to forms, which is minimally complex, and simultaneously minimises the *distortion* between the intended meaning M , and the listener’s reconstructed meaning \hat{M} . An operationalisation of this trade-off was first proposed by Tishby et al., 2000, and later applied to colour-naming (Zaslavsky et al., 2018), and other semantic domains (Mollica et al., 2021; Chen et al., 2023; Langlois et al., 2025).

In this line of work, the Complexity of the stochastic encoder q is quantified by the mutual informa-

tion¹ $I_q(M; W)$ between two random variables: M and W ². The mental representations the speaker wishes to communicate are denoted by M , and W represents the discrete linguistic forms they are mapped onto, according to the policy $p_q(w|m)$ of the encoder q . Conversely, the *distortion* between the speaker’s intended meaning M , and the listener’s reconstructed meaning \hat{M} , is measured using the expected Kullback-Leibler distance, $\mathbb{E}_q[D(M||\hat{M})]$. Minimising this distortion is mathematically equivalent to maximising the quantity $I_q(W; U)$, which is defined as the Accuracy of the encoder (Zaslavsky, 2019). In this formulation, U represents the possible states of the world that the speaker perceives and describes, and about which the listener forms probabilistic inferences.

The core trade-off between the compression of mental representations (measured by Complexity) and the preservation of meaningful information (measured by Accuracy) translates directly to optimisation. For any fixed amount of Accuracy, the encoder and decoder are optimised to ensure that Complexity is minimised. Introducing a Lagrange multiplier β , this trade-off can be framed as a con-

¹ The mutual information is a general measure of how related two random variables are. It measures the reduction in uncertainty in one variable when observing the other: $I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$, with $H(X)$ being the entropy of X . Note that the joint distribution $p(X, Y)$ is all we need to compute the mutual information: $I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x) \cdot p(y)}$. It is a symmetric measure: $I(X; Y) = I(Y; X)$.

²We refer to random variables using upper-case letters, and realisations with lower-case letters.

strained optimisation problem with the following objective function:

$$\mathcal{F}_\beta[q] = I_q(M; W) - \beta I_q(W; U) \quad (1)$$

where β controls the trade-off. Any incremental 1-bit growth in Accuracy of a given optimal encoder incurs a cost of β bits in Complexity. All natural or artificial encoders can be located in the Accuracy-Complexity information plane. From the Data Processing Inequality it follows that Complexity serves as an upper bound for Accuracy. Consequently, any encoders lying above the optimal frontier defined by the set $\{\mathcal{F}_\beta^* \mid 0 \leq \beta < +\infty\}$ are unachievable under the assumptions of the IB framework.

3. General methodology

Within a specific semantic domain, an IB analysis relies on two types of data to compute the Complexity and Accuracy of an encoder g : naming data and human similarity judgements. Examples of naming data include colour names (when describing colour chips in Zaslavsky et al., 2018), verbs of human gait (when describing a person moving in short videos, in Langlois et al., 2025), object names (when describing real-world images in Gualdoni et al., 2023), or spatial relations (when describing drawings of spatial scenes in Khetarpal et al., 2013). In this work, we apply the IB framework to the translation task. In this section, we detail how naming data and similarity judgements can be recovered from translation data, and how Complexity and Accuracy measures are derived.

Using a bitext, we treat each occurrence of a term of interest within its sentential context U as a stimulus that evokes a meaning representation M . Consider, for example, $u =$ the apple in the bowl. Here, the spatial relation between the apple and the bowl is not only determined by the lexical item in alone, but emerges from the full sentential context in which it is embedded. Crucially, two occurrences of the same lexical item evoke the same meaning M if and only if they appear in identical sentential contexts, that is, when the same phrase is repeated in the bitext corpus. Under this assumption, M is contextually grounded rather than lexically fixed. The evoked representation M is then encoded using a lexical item W in the target language. The schema $U \rightarrow M \rightarrow W$ summarises how translations are generated under these assumptions.

3.1. Naming data

Given a bitext corpus, our goal is to align terms of interest from source sentences with their counterparts from target sentences. For example, the French spatial prepositions from Table 1 ($u_1 =$ le pilote prit place sur la passerelle and

$u_2 =$ Mudge se maintenait dans la ligne droite) map to their English equivalents $w_1 =$ on and $w_2 =$ in. This process yields alignment tables like the one presented in Table 2. From these alignments, we can derive the conditional probability distribution $p_q(w|m)$ where m ranges over each distinct meaning and w ranges over the target lexicon. Assuming a uniform prior distribution $p(m)$ over all meanings in the source text³, we compute the joint distribution: $p_q(w, m) = p_q(w|m)p(m)$. From this, we derive the Complexity $I_q(M; W)$ (see footnote 1 for details).

id	French	Automatic alignment with English				
		above	in	into	on	towards
n1909	et prit place sur la passerelle	0	0	0	1	0
n3688	se maintenait dans la ligne droite	0	1	0	0	0
n3110	la direction de l'est à travers le massif	0	0	0	0	1
n2481	descendirent le pauvre garçon dans une cabine	0	0	1	0	0
n1243	le guide dirigerait l'éléphant vers la pagode de Pillaji	0	0	0	0	1

Table 2: French to English excerpt of computed alignments of spatial prepositions from LIFAT et al., 2024. Each row is a one-hot encoding of the source preposition over the target lexicon. The source preposition is emphasised in bold.

3.2. Similarity judgements

Collecting contextual similarity judgements of source terms grows quadratically in the size of the bitext. Being able to estimate the pairwise similarity between any pair of source terms is therefore crucial. As part of our proposed methodology, we first select a subset of terms in their sentential context $\{u_i \mid 1 \leq i \leq K\}$ of reasonable size K , and collect human pairwise similarity judgements $\{sim(u_i, u_j) \mid 1 \leq i, j \leq K\}$. For ease of exposition, we denote these empirical similarities as $sim_{i,j}$. Based on these pairwise similarity judgements, we then train a model to predict the similarity $\widehat{sim}_{i,j}$ for any new unseen pair of terms. Following Zaslavsky et al., 2018 and Chen et al., 2023, we derive a subjective belief distribution on world states conditional on meanings $p(u|m)$, which is estimated as

$$p(u_j|m_i) \propto \exp\left(\gamma \cdot \widehat{sim}_{i,j}\right) \quad (2)$$

³Zaslavsky et al., 2018 show that replacing an empirically derived prior reflecting actual communicative need with a simple uniform prior has little qualitative influence on the core finding. We use therefore a uniform prior distribution in this first step

Here, γ is a temperature parameter that controls the penalty for mistaking two distinct stimuli⁴. Note that when IB models a subjective belief as a distribution on world states U , it captures the ambiguity a term of interest can have in its context. Using the distributions $p(u|m)$, $p_q(w|m)$, and $p(m)$, we can now compute the joint distribution: $p_q(w, u) = p_q(w|m)p(m)p(u|m)$. From this, we derive the Accuracy $I_q(W; U)$.

4. Experiments

For the remainder of this work, we focus on the specific semantic domain of spatial relations. These relations are typically marked by prepositions as discussed extensively in the literature (for French linguistic aspects we refer to [Vandeloise, 1991](#); [Borillo, 1998](#); [Le Pesant, 2012](#), for a broad discussion on spatial cognition we refer to [Landau and Jackendoff, 1993](#)). We also consider “absolute use” of prepositions ([Le Pesant, 2012](#)), which refers to instances where a preposition appears without its standard complement because the referent is understood from context (e.g., *Je me gare devant*, I park in front, *Ich parke davor* and *Parkiram ispred*). More importantly, we consider a preposition to have a spatial sense when its surrounding context triggers a concrete spatial mental imagery, involving entities that can be physically touched or pointed at. This operational definition is theoretically grounded in the notions of *cline of concreteness* ([Brybaert et al., 2014](#)) and *cline of collocation* ([Pragglejaz, 2007](#)).

Throughout our experiments⁵, we utilised English, German, and Serbian translations of Jules Verne’s French adventure novel, *Le tour du monde en 80 jours*. The sentence-level alignment across these four texts was previously processed by [LIFAT et al., 2024](#).

4.1. Naming data: detection and alignment of spatial prepositions

We first collected occurrences of prepositions with potential spatial sense, namely searching for instances from the lexical inventory of spatial prepositions established by [Le Pesant, 2012](#). An initial disambiguation of each occurrence was then carried out jointly by three annotators until consensus was reached. The lexical and semantic disambiguation steps yielded 1,312 prepositions with a spatial sense. Using the Large Language

⁴We set γ so that $p(u|m)$ is defined by a softmax over similarity scores. Alternative methods for grounding the distribution $p(u|m)$ should be explored in future work.

⁵Data and code are available at <https://github.com/antoineOYO/spatial-terms-translations>

Model `mistral-large-2512` ([Mistral AI, 2025](#)) in a few-shot setting, these French prepositions were subsequently aligned with their English, German and Serbian counterparts. The alignment table for Table 1 is illustrated in Table 2. The few-shot prompt for French-to-English alignment is available in Appendix A. To evaluate the quality of these automatic alignments, we sampled 100 preposition pairs for each target language. Errors of alignment were defined as the alignments of source prepositions to either inadequate target prepositions, or to any distinct other parts of speech, such as nouns or verbs. This evaluation yielded precision scores of 81% for French–English, 91% for French–German, and 94% for French–Serbian. Out of the 1,312 source prepositions, 580 were successfully aligned across all three target languages. The remaining instances were discarded, primarily because the source preposition was not relexified with a preposition in at least one of the translations⁶. The cross-linguistic discrepancy between automatic alignments is discussed in section 5.3.

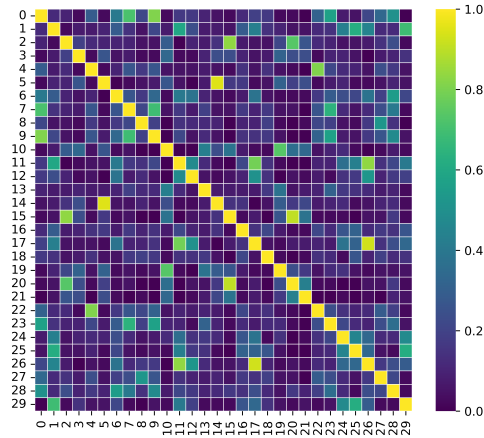
Furthermore, we computed contextual embeddings for each source preposition, using `xlm-roberta-large` ([Conneau et al., 2020](#)). We averaged and stacked the last four hidden layers, yielding embeddings of size $d = 4,096$. We then applied K -means clustering ($K = 30$) to obtain a representative subset of 30 French sentences, each with a spatial preposition, for downstream analysis. We found that $K = 30$ reflected both the distribution of French spatial terms in the corpus, and the embedding space (see Appendix D for a dimension reduction). The clustering and dimension reduction were implemented using `scikit-learn` ([Pedregosa et al., 2011](#)).

4.2. Models of human similarity judgements

Our objective is to predict the human-perceived similarity between two distinct spatial relations expressed by spatial prepositions from any source sentence.

For this purpose, we designed a pile-sorting task based on the $K = 30$ samples obtained from our clustering step. To assess the feasibility of these similarity judgements, we conducted an anonymised pilot study (see section 7) with a cohort of colleagues ($N = 35$), all of whom were native French speakers. The reported patterns should be interpreted as preliminary evidence. The 30 samples were printed on individual cardboard cards. The participants were orally instructed to group the

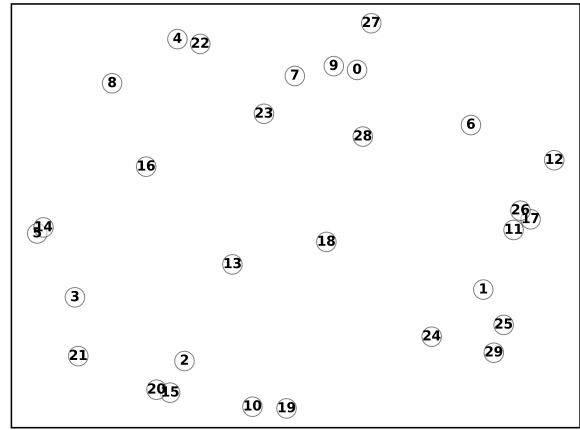
⁶This is also known as *zero-marking*: consider the use of the preposition in *Je suis à la maison* versus *I’m home*. We refer to [Beekhuizen, 2025](#) for related analysis.



(a) Proportion of participants (N=35) who grouped 2 spatial relations into the same pile.

- 0: et prit place **sur** la passerelle
- 1: car **dans** cet immense et admirable port de New York
- 2: le guide dirigerait l'éléphant **vers** la pagode
- 3: ce voyage s'accomplissait régulièrement **autour du** monde
- 4: **sous** ses goélettes, l'Henrietta marcha
- 5: **entre** ces deux importantes cités
- 6: il avait pu s'embarquer **à bord du** Rangoon
- 7: **sur** le bûcher enflammé se détachait alors le corps
- 8: Phileas Fogg prit une chaise et s'assit **près de** la cheminée
- 9: je ne le vois jamais **sur** le pont
- 10: et **par** Rock-Island, il entra dans l'Illinois
- 11: Je suis resté couché **dans** ma cabine
- 12: Étant entré **chez** un barbier chinois pour se faire raser
- 13: la vapeur s'échappait **par** les soupapes
- 14: qui s'étendent **entre** le Japon et le Nouveau Monde

(c) Prepositions (emphasised) and context of the 30 spatial relations sorted by the participants



(b) MDS visualisation of the pairwise similarities.

- 15: un paquebot qui ne partait que le lendemain [...] **pour** Hong-Kong.
- 16: se trouver [...] **au milieu** de voyageurs
- 17: descendirent le pauvre garçon **dans** une cabine des secondes
- 18: un bruit retentit **au – dehors**
- 19: il reprit la direction de l'est **à travers** le massif
- 20: dût-il pousser **jusqu'à** Macao
- 21: un parcours de deux cent vingt milles **depuis** Hong-Kong
- 22: la tête de la procession apparut **sous** les arbres
- 23: couraient **sur** des tuyaux de pipe
- 24: se maintenait **dans** la ligne droite
- 25: La goélette [...] semblait voler **dans** l'air
- 26: descendre **dans** la cabine
- 27: il s'arrêtait **devant** une vaste case
- 28: Au lieu d'arriver **sur** l'Henrietta
- 29: encore en faveur **dans** le Céleste Empire

Figure 1: Empirical similarity data from the pile-sorting task.

samples into as many piles as they deemed necessary based on their perceived similarity; there was no time limit and participants could revise their groupings until satisfied. The full instructions (see Appendix B) were printed and available to each participant throughout the whole task. The distribution of number of piles produced by the participants is displayed in Appendix C.

Following Khetarpal et al., 2013, we define the pairwise similarity $sim_{i,j}$ as the proportion of all participants who sorted u_i and u_j into the same pile. A visualisation of the proportions is shown in Figure 1a. From the list of pairwise similarities, a Multidimensional Scaling (MDS)⁷ was then computed using `scikit-learn` implementation (Figure 1b). Both visualisations suggest that the similarity structure of spatial relations is complex and graded rather than discrete or reducible to surface form (discussed in section 5.2).

We modelled these similarity judgements using three approaches based on the contextual embeddings F_i of size d we computed in the previous step. As a baseline, we computed the Cosine Similarity

between z-scored embeddings, following Timkey and van Schijndel, 2021. We compared this baseline against two learned models: a Ridge regression and a low-rank projection model (replicating Jha et al., 2023). The Ridge regression model is a simple linear regression, where the size of the coefficients is penalised with a parameter α . It is trained on the vertical stack of three features: the Hadamard product $F_i \odot F_j$, their L1 distance, and their cosine similarity. Conversely, the low-rank projection model learns a mapping matrix $P \in \mathbb{R}^{D \times d}$ to project embeddings into a D -dimensional space, predicting similarity as $\widehat{sim}_{i,j} = (PF_i)^\top (PF_j)$. This model minimises the mean squared error loss with L2 regularisation:

$$\min_{\lambda, P} \frac{1}{2} \sum_{i,j} ||sim_{i,j} - \widehat{sim}_{i,j}||^2 + \lambda ||P||^2 \quad (3)$$

Hyperparameters (including the Ridge α , the projection rank D , and the regularisation penalty λ) were optimised within 6-fold inner validation splits, and the generalisation error (measured via Spearman's ρ) was computed across 6 outer test folds using a nested cross-validation procedure. Model

⁷Note that this MDS is for visualisation purpose only.

performance (see Table 3) is reported using Spearman’s rank correlation coefficient ρ , as it is better suited for evaluating monotonic relationships than Pearson’s r_P , which evaluates strictly linear ones (De Winter et al., 2016). Results presented in Table 3 indicate that a simple cosine similarity does not strongly correlate with the sortings we collected, whereas the low-rank projection model achieves a high Spearman correlation of $\rho = 0.7800 \pm 0.0622$, with dimension $D = 5$.

Model	Spearman ρ
Cosine Similarity (Baseline)	0.079
Ridge Regression	0.647 ± 0.056
Low-rank Projection	
$D = 1$	0.263 ± 0.140
$D = 5$	0.780 ± 0.062
$D = 10$	0.762 ± 0.053
$D = 15$	0.751 ± 0.053
$D = 20$	0.749 ± 0.034
$D = 50$	0.648 ± 0.070
$D = 100$	0.578 ± 0.101

Table 3: Spearman correlation between model-predicted pairwise similarities and human pile-sorting judgements. Models all take pairs of contextual embeddings as input. Learned models performance is computed from 6-fold nested cross validation.

4.3. Computation of encoders locations in the information plane

To benchmark the attested translation encoders, we generated $3 \times 10,000$ counterfactual “perturbed encoders” by randomly permuting 1%, 5% and 10% of the rows within the alignment tables. We also generated 300,000 uniform-random translation encoders. We then computed the Accuracy and Complexity for attested, perturbed and fully random translation encoders using the implementation provided by Zaslavsky et al., 2018. The optimal frontier was computed using the reversed deterministic annealing algorithm implementation from Chen et al., 2023, and was traced by iterating over 100 values of β log-linearly spaced between 1 and 2^{20} . Attested, perturbed and random translations, along with the optimal frontier, are displayed in the information plane from Figure 2a. We observe that attested translation encoders lie far from the steepest part of the theoretical bound, which is discussed in section 5.3.

Within the information plane, the distance of any given encoder q to the optimal frontier defines its deviation from optimality, as established by Zaslavsky

et al., 2018:

$$\epsilon_q = \min_{\beta} \frac{1}{\beta} (\mathcal{F}_{\beta}[q] - \mathcal{F}_{\beta}^*), \quad (4)$$

The deviations of translation encoders is shown in Figure 2b. Results indicate that as the degree of perturbation increases, the efficiency of translation encoders decreases.

5. Discussion

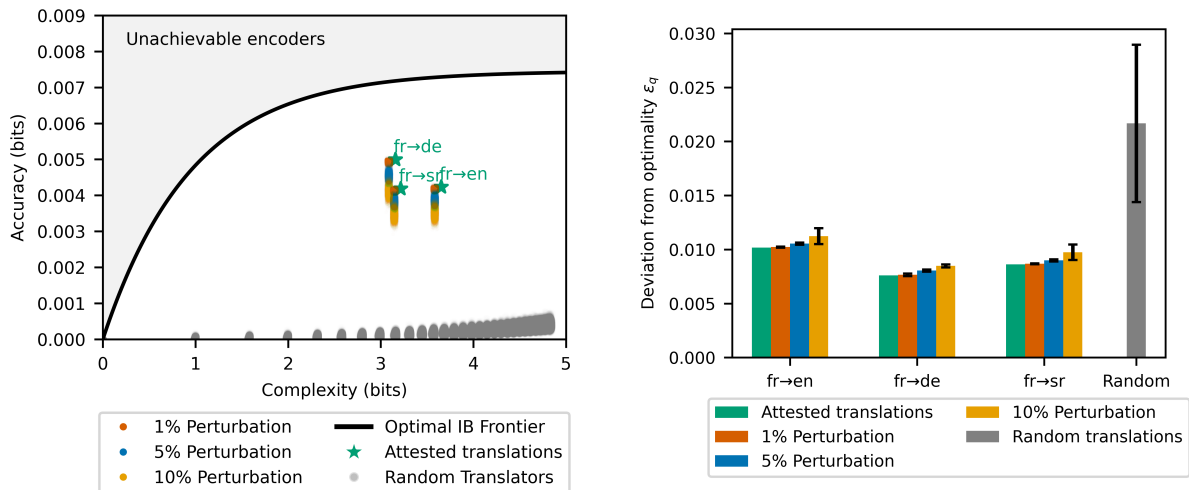
5.1. Characterising cross-linguistic use of spatial prepositions

The automatic detection and alignment of spatial prepositions in bitexts has received limited attention until recent studies such as those by Viechnicki et al., 2024 and Beekhuizen, 2025. For detection, we handled the bitext differently by choosing not to apply a syntactic filter to the source sentences. This choice is based on the observation that spatial terms appear in a variety of syntactic contexts, and range from single terms (sur, dans), to nominal locutions (à bord du, au milieu de). With respect to alignment, Viechnicki et al., 2024 rely on the GIZA++ word alignment tool (Och and Ney, 2003). In contrast, the RUMTE model proposed by Beekhuizen, 2025 avoids traditional word alignment, instead identifying spatial relation markers by calculating statistical associations. Direct performance comparisons between these methods are difficult, as these previous works evaluated highly morphologically complex languages. While our pipeline introduces a dependency on few-shot extraction tools, it generalises to any bitext without requiring morphologically complex language-specific alignment models.

5.2. Estimation of similarity between spatial relations evoked by prepositions

The correlation between judgements of word sense relatedness and computational representations relatedness has been widely investigated using both static embeddings (Hill et al., 2015), contextual embeddings and Large Language Models (Nair et al., 2020; Marjeh et al., 2022). Specifically, how context modulates the perceived meaning similarity between word pairs in written sentences has been graded by human subjects in many experiments (Huang et al., 2012; Pilehvar and Camacho-Collados, 2019; Armendariz et al., 2020; Vulić et al., 2020). However, most of these studies focus on nouns, adjectives or verbs. Our pilot study, conversely, was concerned with relations between entities, and the prepositions that mark them.

Results presented in Table 3 indicate that a simple cosine similarity does not strongly correlate with



(a) Information plane of attested, perturbed and random encoders of spatial prepositions. Dots are jittered slightly along the Complexity axis for visualisation purposes.

(b) Deviation from optimality for attested, perturbed and random encoders. Black error bars indicate standard deviations.

Figure 2: How much do attested and random systems deviate from optimality?

our collected human sortings, whereas the low-rank projection model achieves a high Spearman correlation of $\rho = 0.7800 \pm 0.0622$. Interestingly, low-rank projection models achieve the highest correlation scores with $D = 5$. This suggests that higher-dimensional projections likely lead to overfitting. These findings indicate that low-rank projections may be more effective than standard cosine similarity for predicting contextual similarities in spatial relations.

Aside from modelling, we acknowledge that our subset of 30 spatial prepositions in context does not guarantee cognitive representativeness of the domain of spatial relations. However, results from Figure 1a and 1b suggest that colexification does not explain alone the similarity structure of spatial relations. We assume that modelling the semantic similarity of spatial relation in a continuous D -dimensional space accounts for the graded similarity space we observed, and that a sufficiently large bitext should provide broad coverage of the domain.

However, a lexical bias may arise during the pile sorting task, where shared word forms inflate human judgements of semantic similarity. As a result, similarity ratings may partly reflect surface-form overlap rather than purely relational semantics. This introduces a potential language-specific skew, whereby the lexical structure of French influences perceived relatedness. To mitigate this effect, future work should control for form overlap in the stimuli and compare across source languages, in order to quantify the contribution of lexical structure to similarity judgements.

We also acknowledge that this subset is not fully

representative of all linguistic means available for encoding spatial relations. While we focused on prepositions as a first step, alternative encoding mechanisms must be accounted for in future research, such as interactions between the preposition and the verb (see for example the verb-framing typologies from Talmy, 1983 and Aurnague, 2012).

5.3. Efficiency of attested systems

In Figure 2a we observe that attested translations of spatial prepositions lie closer to optimal systems than random ones. To further investigate whether this is a feature of “reasonable” translations, we generated counterfactual translations with varying degrees of perturbation. The results from Figure 2b indicate that as the degree of perturbation increases, the deviation to optimality increases. We believe that this comparison would benefit from including another baseline, such as a surjective mapping from meanings to terms, as used by Strickland and Chemla, 2018. This is equivalent to removing the non-deterministic assumption on the encoder g (Strouse and Schwab, 2017). Future work should also test how altering the similarity geometry affects Accuracy. Nevertheless, these results suggest that attested systems exhibit a pressure for efficiency under the IB framework.

We conclude by addressing why attested encoders deviate from optimality, and the confounds inherent to translation data. This question lies at the crossroads of research on language efficiency and translation process (Teich et al., 2020). Translating involves genre conventions and notably stylistic incentives. This is illustrated by observations from

the corpus: while German and Serbian translations of *Le Tour du monde en 80 jours* remain stylistically conservative, the English version acts as a “free” translation with added details. This divergence is clear in the original corpus, where LIFAT et al., 2024 found and explicitly annotated changes in meaning, but were found only in the English version. From an information-theoretic perspective, these changes in meaning represent a deviation from strict communicative efficiency, as the translator prioritised semantic elaboration or expressivity over the compact encoding predicted by the IB optimum. Yet, a conservative translator might also imitate some of the lexical choices of the French original despite what would be natural in their language (Volansky et al., 2015; Lim et al., 2024). The alignment of preposition-to-preposition we performed (see section 4.1) is by design too coarse to capture the aforementioned phenomena. Therefore we argue that IB analysis of translation data must narrow the scope to relexification of terms of interest, such as prepositions, in order to capture only one aspect of the translation process.

6. Conclusion

This work proposes a novel methodology for analysing communicative efficiency via translation data. We performed an IB analysis on spatial prepositions across three translations of a French novel. Our preliminary results suggest that the translation of spatial prepositions exhibits pressure toward optimal trade-offs between informativity and complexity, consistent with the IB literature, and more broadly, recent evidence from language efficiency. An important next step is to apply this approach to typologically diverse, genealogically distant languages, allowing us to better separate general efficiency pressures from patterns arising through shared linguistic ancestry. We believe we presented a plausible way to extend to other semantic domains, other linguistic forms than prepositions, and scale beyond fully elicited naming datasets while staying grounded in natural speech.

7. Ethics Statement

Written informed consent was obtained from all participants. Pile-sorting data consisted only of lists of sentences. These data were fully anonymised at the point of collection, such that no personal data were retained; the data from the pile-sorting experiment therefore fall outside the scope of the General Data Protection Regulation (GDPR, Regulation (EU) 2016/679). Given the minimal-risk nature of this feasibility study (involving no sensitive information and no identifiable data) institutional ethics committee approval was not sought.

8. Acknowledgements

We sincerely thank our colleagues at Laboratoire d’Informatique en Image et Systèmes d’information for participating in the pile-sorting experiment. We further thank the participants of the SIGTYP workshop for their stimulating discussions and valuable feedback during the presentation of this work. We are also grateful to the anonymous reviewers for their constructive suggestions, which helped us improve this paper.

9. Bibliographical References

- Carlos Santos Armendariz, Matthew Purver, Matej Ulčar, Senja Pollak, Nikola Ljubešić, and Mark Granroth-Wilding. 2020. *CoSimLex: A resource for evaluating graded word similarity in context*. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 5878–5886, Marseille, France. European Language Resources Association.
- Michel Aurnague. 2012. De l’espace à l’aspect: les bases ontologiques des procès de déplacement. *Corela. Cognition, représentation, langage*, (HS-12).
- 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.
- Andrée Borillo. 1998. *L’espace et son expression en français*. Editions Ophrys.
- Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. 2014. *Concreteness ratings for 40 thousand generally known English word lemmas*. *Behavior Research Methods*, 46(3):904–911.
- 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.

- Thomas M Cover. 1999. *Elements of information theory*. John Wiley & Sons.
- Joost CF De Winter, Samuel D Gosling, and Jeff Potter. 2016. Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological methods*, 21(3):273.
- Edward Gibson, Richard Futrell, Steven P Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen, and Roger Levy. 2019. How efficiency shapes human language. *Trends in cognitive sciences*, 23(5):389–407.
- Eleonora Gualdoni, Charles Kemp, Yang Xu, and Gemma Boleda. 2023. Quantifying informativeness of names in visual space. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 45.
- Felix Hill, Roi Reichart, and Anna Korhonen. 2015. Simlex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics*, 41(4):665–695.
- Eric Huang, Richard Socher, Christopher Manning, and Andrew Ng. 2012. [Improving word representations via global context and multiple word prototypes](#). In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 873–882, Jeju Island, Korea. Association for Computational Linguistics.
- Aditi Jha, Joshua C. Peterson, and Thomas L. Griffiths. 2023. [Extracting Low-Dimensional Psychological Representations from Convolutional Neural Networks](#). *Cognitive Science*, 47(1):e13226.
- 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.
- Simon Kirby, Monica Tamariz, Hannah Cornish, and Kenny Smith. 2015. Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141:87–102.
- 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.
- Stephen C Levinson. 2003. *Space in language and cognition: Explorations in cognitive diversity*, volume 5. Cambridge University Press.
- Stephen C Levinson, Sérgio Meira, et al. 2003. ‘natural concepts’ in the spatial topological domain—adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language*, 79(3):485–516.
- LIFAT, LLL, and Université de Belgrade. 2024. [En80jours](#). ORTOLANG (Open Resources and TOols for LANGuage) —www.ortolang.fr.
- Zheng Wei Lim, Ekaterina Vylomova, Trevor Cohn, and Charles Kemp. 2024. [Simpson’s Paradox and the Accuracy-Fluency Tradeoff in Translation](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 92–103, Bangkok, Thailand. Association for Computational Linguistics.
- Raja Marjeh, Pol Van Rijn, Ilia Sucholutsky, Theodore R Sumers, Harin Lee, Thomas L Griffiths, and Nori Jacoby. 2022. Words are all you need? language as an approximation for human similarity judgments. *arXiv preprint arXiv:2206.04105*.
- Mistral AI. 2025. Introducing mistral 3. Mistral Documentation.
- Francis Mollica, Geoff Bacon, Noga Zaslavsky, Yang Xu, Terry Regier, and Charles Kemp. 2021. The forms and meanings of grammatical markers support efficient communication. *Proceedings of the National Academy of Sciences*, 118(49):e2025993118.
- Sathvik Nair, Mahesh Srinivasan, and Stephan Meylan. 2020. [Contextualized Word Embeddings Encode Aspects of Human-Like Word Sense Knowledge](#). In *Proceedings of the Workshop on the Cognitive Aspects of the Lexicon*, pages 129–141, Online. Association for Computational Linguistics.
- Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational linguistics*, 29(1):19–51.

- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Mohammad Taher Pilehvar and Jose Camacho-Collados. 2019. [WiC: the word-in-context dataset for evaluating context-sensitive meaning representations](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1267–1273, Minneapolis, Minnesota. Association for Computational Linguistics.
- Group Pragglejazz. 2007. [MIP: A Method for Identifying Metaphorically Used Words in Discourse](#). *Metaphor and Symbol*, 22(1):1–39.
- 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).
- Daniel J Strouse and David J Schwab. 2017. The deterministic information bottleneck. *Neural computation*, 29(6):1611–1630.
- Leonard Talmy. 1983. *How Language Structures Space*, pages 225–282. Springer US, Boston, MA.
- Elke Teich, José Martínez Martínez, and Alina Karakanta. 2020. Translation, information theory and cognition. In *The Routledge Handbook of Translation and Cognition*. Routledge. Num Pages: 16.
- 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*.
- Claude Vandeloise. 1991. *Spatial prepositions: A case study from French*. University of Chicago Press.
- 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.
- Vered Volansky, Noam Ordan, and Shuly Wintner. 2015. On the features of translationese. *Digital Scholarship in the Humanities*, 30(1):98–118.
- Ivan Vulić, Simon Baker, Edoardo Maria Ponti, Ulla Petti, Ira Leviant, Kelly Wing, Olga Majewska, Eden Bar, Matt Malone, Thierry Poibeau, et al. 2020. Multi-simlex: A large-scale evaluation of multilingual and crosslingual lexical semantic similarity. *Computational Linguistics*, 46(4):847–897.
- Noga Zaslavsky. 2019. *Information-theoretic principles in the evolution of semantic systems*. Ph.D. thesis, The Hebrew University of Jerusalem Jerusalem, Israel.
- 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.

Appendix

A. Prompt for French to English few-shot alignment

ALIGNMENT OF SPATIAL TERMS

You are an expert linguist and translator specialised in French and English. French is always the source language, and the target language is English. You will be provided with a source sentence in French, and the target sentence in English. In the source sentence, a spatial term is wrapped with <loc> and </loc>. This spatial term marks a concrete spatial relationship. Your task: align the spatial term in the source sentence with its translation in the target sentence, if any.

How: wrap with <loc> and </loc> the corresponding spatial term in the target sentence, if any. Also return the normalized form of the target spatial term (e.g., "on the right of" to "on the right").
Constraints:

1. Spatial terms only include prepositions, adverbs, and locutions.
2. Zero-alignment: if you cannot find an adequate spatial term in the target sentence corresponding to the translation of the source spatial term, don't

wrap anything in <loc> from the target sentence and return "None" for the normalized form.

3. If the meaning of the source spatial term is translated only by a verb or a noun in the target sentence, don't wrap anything in <loc> from the target sentence and return "None" for the normalized form.

Example 1: basic alignment

```
INPUT = { "source_sentence": "Passepartout
demeura seul <loc>dans</loc> la maison de
Saville-row", "target_sentence": "Passepartout
remained alone in the house of Saville-row", }
OUTPUT = { "target_sentence": "Passepartout
remained alone <loc>in</loc> the house of
Saville-row", "target_normalized_term": "in" }
```

Example 2: zero-alignment

```
INPUT = { "source_sentence": "Il rentra
<loc>chez</loc> lui à dix heures.",
"target_sentence": "He returned home at ten
o'clock.", }
OUTPUT = { "target_sentence": "He returned
home at ten o'clock.", "target_normalized_term":
"None" }
```

Example 3: complex locution

```
INPUT = { "source_sentence": "Au moment où
l'Henrietta appareillait,tous quatre étaient <loc>à
bord</loc>.", "target_sentence": "all four were
aboard at the moment when the Henrietta was
setting sail.", }
OUTPUT = { "target_sentence": "all four were
<loc>aboard</loc> at the moment when the
Henrietta was setting sail.",
"target_normalized_term": "aboard" }
```

Example 4: zero-alignment

```
INPUT = { "source_sentence": "Il a marché
<loc>juqu'à</loc> la ville.", "target_sentence": "He
reached the city by foot.", }
OUTPUT = { "target_sentence": "He reached the
city by foot.", "target_normalized_term": "None" }
# Now is your turn. Provide the output in JSON
format as in the examples.
```

B. Instructions for pile sorting

French instructions (as printed for the participants): "Sur chaque billet est exprimée une relation spatiale entre 2 entités. Une préposition en gras vous signale cette relation. Formez des piles regroupant les relations spatiales qui expriment des configurations spatiales similaires. Vous pouvez faire autant de piles que vous le souhaitez (y compris des piles d'un seul billet si une relation vous semble unique). Le numéro sur les billets est aléatoire."

A possible English translation: "On each card, a spatial relationship between 2 entities is expressed. A bolded preposition indicates this relationship. Form piles grouping the spatial relationships that

express similar spatial configurations. You can create as many piles as you like (including stacks with a single card if a relationship seems unique). The numbers on the cards are random."

C. Pairwise similarity modelling

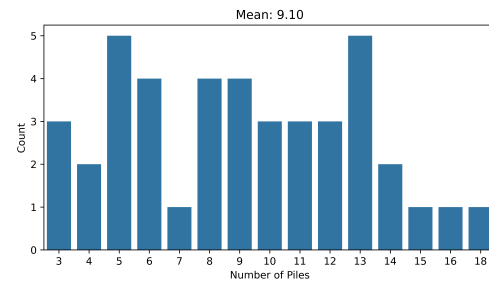


Figure C-1: Distribution of the number of piles across participants (N=35).

D. Dimension reduction of the embedding space

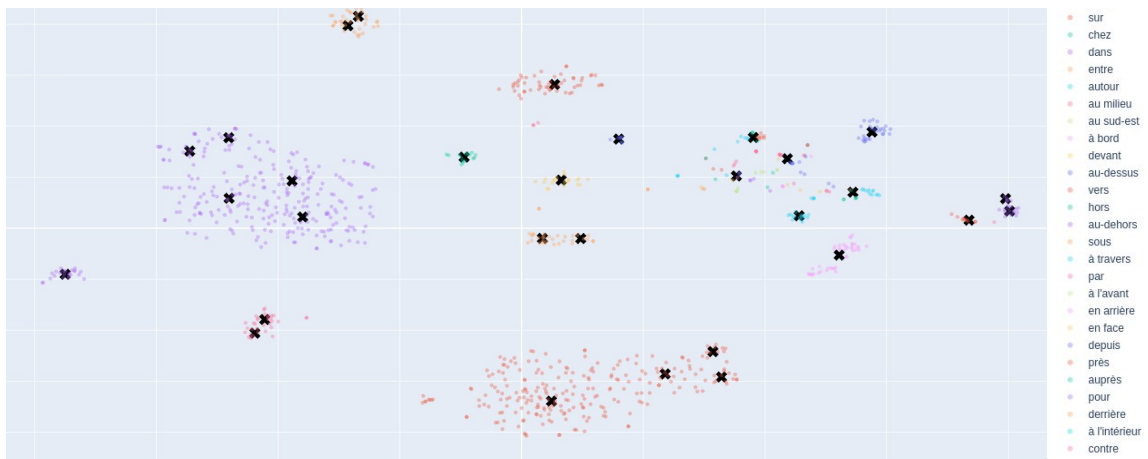


Figure D-1: t-SNE visualisation of the 30 K-means centroids (computed from stacked final-layer representations of `xlm-roberta-large`) and embeddings of the most frequent French prepositions