Language Dominance in Multilingual Large Language Models

Nadav Shani
Centre for Language Technology
University of Copenhagen
zlq284@alumni.ku.dk

Ali Basirat
Centre for Language Technology
University of Copenhagen
alib@hum.ku.dk

Abstract

This paper investigates the language dominance hypothesis in multilingual large language models (LLMs), which posits that cross-lingual understanding is facilitated by an implicit translation into a dominant language seen more frequently during pretraining. We propose a novel approach to quantify how languages influence one another in a language model. By analyzing the hidden states across intermediate layers of language models, we model interactions between language-specific embedding spaces using Gaussian Mixture Models. Our results reveal only weak signs of language dominance in middle layers, affecting only a fraction of tokens. Our findings suggest that multilingual processing in LLMs is better explained by language-specific and shared representational spaces rather than internal translation into a single dominant language.

1 Introduction

Large Language Models (LLMs) built on the Transformer architecture (Vaswani et al., 2017) have demonstrated remarkable capabilities in acquiring linguistic knowledge from raw text (Devlin et al., 2019; Brown et al., 2020). When exposed to multilingual corpora during training, these models develop the ability to process and generate text across multiple languages (Conneau and Lample, 2019; Conneau et al., 2020; Xue et al., 2021; Shi et al., 2023). Intriguingly, they appear to internalize structural and syntactic differences between languages even without direct supervision or explicit instruction in grammar (Lin et al., 2022).

Recent interpretability studies suggest that this multilingual proficiency is highly influenced by the composition of languages in the pretraining data (Papadimitriou et al., 2023; Üstün et al., 2024; Yue et al., 2025). Languages that are heavily represented in pretraining data, often referred to as dominant languages, appear to play a mediating role in an implicit translation mechanism that may underlie multilingual capabilities (Wendler et al., 2024; Zhang et al., 2023). According to this dominance hypothesis, LLMs internally translate inputs from less-represented languages into a more dominant or familiar language, perform intermediate computations in this shared representation space, and then translate the result back into the original language.

We introduce a novel probabilistic approach to assess how the embedding space of a dominant language may become activated when other languages are processed within an LLM. Our approach employs Gaussian Mixture Models to capture language interactions based on token embeddings from intermediate layers. By analyzing posterior probabilities and cross-language activation patterns within this probabilistic framework, we compute both global dominance scores at the language level and token-level likelihood ratios that a token from a source language might be processed within the embedding space of another language.

Our investigation of two publicly available multilingual language models (mGPT (Shliazhko et al., 2022) and BLOOM (Scao et al., 2023)) with different language coverage shows that language dominance, when present, primarily emerges in the middle layers of the models and only for a relatively small fraction of tokens. At stricter dominance thresholds, these tokens primarily comprise function words and cross-lingually aligned lexical items such as cognates and shared vocabulary. As the threshold relaxes, more content-bearing words begin to emerge in the set of dominated tokens. Overall, our empirical results across a diverse set of languages do not provide clear evidence that any particular language exerts substantial influence over the internal processing of others. Instead, we observe that the models tend to process languages in isolation at their lower and upper layers, while forming shared, language-agnostic representations in their middle layers.

Proceedings of the 8th BlackboxNLP Workshop: Analyzing and Interpreting Neural Networks for NLP, pages 137–148

2 Related Work

Recent work has sought to uncover the internal mechanisms underlying the multilingual capabilities of LLMs. These studies explore a range of factors, including the influence of training data composition and model accent (Guo et al., 2025; Papadimitriou et al., 2023), the role of training time (Blevins et al., 2022) and language diversity (Üstün et al., 2024) in cross-lingual generalization, and the semantic modeling in multimodal settings (Yue et al., 2025). In this section, we focus on those studies that particularly examine whether multilingual models employ an internal translation strategy across languages.

Zhang et al. (2023) explore the internal dynamics of multilingual understanding in a decoder-only GPT-3 architecture. Their behavioral analysis examines how the model performs on tasks requiring reasoning, knowledge retrieval, and pragmatic language use across different languages. Their findings indicate that English often serves as a mediating language, with other languages processed through representations grounded in English to varying degrees. However, these conclusions are drawn solely from model outputs, leaving the underlying internal mechanisms that give rise to such behavior unexplored. Furthermore, the proprietary nature of the model's training data and parameters (OpenAI, 2024) limits insights into the true extent of its multilingual or English-dominated bias.

To address such limitations, Wendler et al. (2024) apply the logit lens method (nostalgebraist, 2020) to investigate whether LLaMA-2 exhibits English dominance in its internal processing. Instead of relying on outputs taken from the final layer, this approach analyzes intermediate activations by transforming internal logits into token probability distributions. Their results suggest that English is predominantly used as an internal "working language" in intermediate layers, with information being translated back into the source language at deeper layers during output generation.

Similarly, by applying the logit lens to bilingual English-Japanese LLMs, Zhong et al. (2025) show that internal representations are influenced by both languages, and that the extent of reliance on one language depends on the model's training balance and the input language. The more balanced bilingual model tends to represent Japanese text more faithfully in Japanese, whereas the English-dominated model relies more heavily on English. This study provides evidence that LLMs may flexibly use multiple languages in their internal processing rather than rigidly translating all content into a single dominant language.

While the logit lens offers a useful tool for interpreting intermediate representations by projecting them into the model's output vocabulary space, it comes with notable limitations. Crucially, the final vocabulary projection layer is optimized during training to map only the final hidden states to token probabilities. Applying the same projection to earlier layers assumes that intermediate representations are already aligned with the output space, which is not necessarily the case and can lead to misleading interpretations. Furthermore, using token probabilities as a proxy for a "working language" can oversimplify the distributed and non-symbolic nature of internal processing.

The logit lens limitations are specifically studied by Belrose et al. (2025), who show that the technique fails to produce meaningful results for modern language models, including BLOOM (Scao et al., 2023). They show that the top prediction of the logit lens is often identical to the input token and that it tends to allocate higher probability mass to tokens that differ from those emphasized by the true output distribution.

To address the limitations of behavioral evaluations and logit lens, we shift focus from projecting intermediate representations into the output vocabulary space toward analyzing the structural relationships among the internal activations themselves. Specifically, we examine the similarity and divergence of layer activations across languages to assess the extent to which the representation of a language is influenced by other languages in the model. This approach allows us to investigate whether multilingual models rely on dominant-language representations or develop shared and language-specific processing at different layers.

3 Method

To investigate the language dominance hypothesis, we propose a framework for modeling language interactions in LLMs based on the geometry of their embedding spaces at each layer. Specifically, we examine whether the embedding space of a dominant language contributes to processing another language. If so, we interpret it as a translation-like mechanism in which representations are internally mapped into the dominant language's space.

3.1 Language Space

Let L denote a language in a language pool $\mathcal{L}$ and M a multilingual language model. For a sample text $\mathcal{S}_L = \{x_1, \ldots, x_n\}$ in L, we input $\mathcal{S}_L$ into M and extract its intermediate layer representations. For each layer $i \in \{1, \ldots, m\}$, we collect the token-level embeddings $\mathbf{H}_i^L = \{\mathbf{h}_{i,1}^L, \ldots, \mathbf{h}_{i,n}^L\}$ from Transformer layers, where $\mathbf{h}_{i,j}^L$ is the embedding of token $x_j \in \mathcal{S}_L$ at layer i.

Given the high dimensionality of the token embeddings, we apply Principal Component Analysis (PCA) independently at each layer to make the analysis tractable. Specifically, for a given layer i, we aggregate the token embeddings $\mathbf{H}_i^L$ across all languages $L \in \mathcal{L}$ into a single matrix, and compute a PCA transformation over this multilingual embedding set. The resulting projections $\tilde{\mathbf{H}}_i^L$ map all languages at layer i into a shared lower-dimensional space, enabling direct cross-lingual comparison.

Finally, we fit a Gaussian Mixture Model (GMM) to the PCA-reduced embeddings at each layer to model the embedding spaces across languages in a unified manner. In the GMM associated with layer i, each component corresponds to a language $L \in \mathcal{L}$ and is parameterized by a uniform prior weight $\pi_i^L = \frac{1}{|\mathcal{L}|}$ over all languages, a mean vector $\boldsymbol{\mu}_i^L$, and a covariance matrix $\boldsymbol{\Sigma}_i^L$, estimated via maximum likelihood from the embeddings in $\tilde{\mathbf{H}}_i^L$ (i.e., the PCA-reduced token embeddings).

In this way, each language is represented as a Gaussian distribution in the latent space of a layer i, from which its hidden states are drawn:

$$\tilde{\mathbf{H}}_i^L \sim \mathcal{N}(\boldsymbol{\mu}_i^L, \boldsymbol{\Sigma}_i^L).$$

We refer to this Gaussian distribution as the *language space* of L at layer i. The GMM probabilistically models interactions between language spaces at each layer, enabling analysis of how one language influences the representation of another.

3.2 Language Dominance

We define the dominance degree of a target language T over a source language S as the expected probability that the language space of T is activated when the model processes input from S. Intuitively, if the language model internally represents or interprets S through structures in T, then the likelihood of activation within T's embedding space should be comparatively high when handling inputs from S. This measure serves as a proxy for identifying whether T functions as an internal mediating language in the model's multilingual representation.

In a continuous formulation, where token embeddings are drawn from the language spaces formed at layer i, this expectation is expressed as:

$$\begin{aligned} P_i(T \mid S) &= \mathbb{E}_{\tilde{\mathbf{h}} \sim \mathcal{N}(\boldsymbol{\mu}_i^S, \boldsymbol{\Sigma}_i^S)} \left[P(T \mid \tilde{\mathbf{h}}) \right] \\ &= \int P(T \mid \tilde{\mathbf{h}}) \cdot P(\tilde{\mathbf{h}} \mid S)\, d\tilde{\mathbf{h}} \end{aligned} \tag{1}$$

Here, $P(T \mid \tilde{\mathbf{h}})$ denotes the GMM posterior probability that a PCA-reduced vector $\tilde{\mathbf{h}}$ belongs to the component associated with T, and $P(\tilde{\mathbf{h}} \mid S)$ is the distribution of the language space for S.

Since the integral in Equation 1 is intractable in practice, we approximate it using a Monte Carlo estimate. Given a set of PCA-reduced token embeddings $\{\tilde{\mathbf{h}}_{i,1}, \ldots, \tilde{\mathbf{h}}_{i,n}\}$ extracted from layer i while processing a sample $\{x_1, \ldots, x_n\}$ from language S, the dominance score is computed as:

$$\hat{P}_i(T \mid S) \approx \frac{1}{n} \sum_{j=1}^{n} P(T \mid \tilde{\mathbf{h}}_{i,j}) \tag{2}$$

A higher score indicates that the internal representations of S are well-aligned with the representational geometry of T, suggesting that T may function as a mediating language within the model's internal processing of S.

Accordingly, we approximate the dominance score for a language T at a given layer i as the average dominance of T over all other languages:

$$\hat{D}_i(T) = \frac{1}{|\mathcal{L}| - 1} \sum_{L \in \mathcal{L} \setminus \{T\}} \hat{P}_i(T \mid L) \tag{3}$$

This score reflects how strongly language T dominates the internal representations of other languages within the model. A higher value indicates that representations from other languages tend to align closely with the embedding space of T, suggesting its role as an internal mediating language.

4 Experiment Setup

This section outlines the experimental setup used to evaluate the language dominance hypothesis. We begin by detailing the selection of the models, followed by a description of the data collection process, and our method for extracting and processing hidden representations from each model layer.

4.1 Model Selection

Since the language dominance hypothesis has primarily been proposed in the context of autoregressive decoder-only architectures, we evaluate it using two publicly available multilingual language models: mGPT (Shliazhko et al., 2022), containing 1.3 billion parameters, and BLOOM (Scao et al., 2023), with 1.7 billion parameters. Both models are built on decoder-only Transformer architectures and consist of 24 layers. They are specifically designed to process multilingual text and are trained on linguistically diverse corpora.

The composition of the training data differs substantially between the two models. mGPT is trained on a large multilingual Common Crawl corpus mainly covering languages from the CIS region, along with several curated text corpora, where English and Russian are particularly prominent. In contrast, BLOOM is trained on the ROOTS corpus, a more diverse multilingual dataset in which English accounts for approximately 30% of the total tokens. Other high-resource languages in BLOOM's training data include French, Spanish, and Arabic. This wider linguistic coverage makes BLOOM's training data substantially more diverse than that of mGPT, although mGPT includes a larger number of languages in its pre-training data (61 vs. 46).

4.2 Data Collection

Our analysis is based on a subset of languages present in the Parallel Universal Dependencies (PUD) treebanks (Zeman et al., 2017). The PUD treebanks offer aligned sentences in various languages from news sources and Wikipedia, annotated for both morphological and syntactic structures. The cross-lingual alignment of sentences ensures that our findings are not skewed by domain-specific variations or differences in syntactic and semantic structures in certain languages. Additionally, the availability of syntactic annotations allows us to effectively assess the syntactic aspects of possible overlaps between the language spaces.

Among the 21 languages available in the PUD treebanks, we select 7 languages that are included in the pretraining data of both mGPT and BLOOM. In addition, we include two languages — Czech and Icelandic, which are not present in the training data of either model. These out-of-distribution languages allow us to investigate the generalization capabilities of the models beyond their supervised multilingual scope. For each selected language, our analysis is conducted on the first 100 sentences from its corresponding PUD treebank, providing a controlled and parallel dataset for cross-lingual comparison. A summary of the selected treebanks is provided in Table 1.

Language	Family	Genus	ISO	Train
Arabic	Afro-Asiatic	Semitic	ar	✓
Czech	Indo-European	Slavic	cs	–
English	Indo-European	Germanic	en	✓
French	Indo-European	Romance	fr	✓
Hindi	Indo-European	Indo-Aryan	hi	✓
Icelandic	Indo-European	Germanic	is	–
Indonesian	Austronesian	Malayo-Polynesian	id	✓
Portuguese	Indo-European	Romance	pt	✓
Spanish	Indo-European	Romance	es	✓

Table 1: Languages used to evaluate mGPT and BLOOM. ✓ indicates presence in the training data.

4.3 Extraction of Hidden Representations

We begin by extracting raw sentences from each of the selected PUD treebanks and input them into the language models under investigation. Each sentence is first tokenized using the model-specific tokenizer. The resulting tokenized sequences are then fed independently into the language models. During processing, we collect the hidden representations produced at each intermediate layer.

For each token in the original sentence, we compute its representation by averaging the embeddings of its corresponding sub-tokens at a given layer. This yields a real-valued tensor of shape $n \times m \times d$, where n is the number of tokens in the sentence, m is the number of layers in the language model, and d is the dimensionality of the token embeddings. For each token, the embeddings averaged over its sub-tokens produce a single token representation, which is subsequently aligned to the token's syntactic properties, such as part-of-speech (POS) tags, using a single token representation per layer.

We collect the token embeddings for all selected sentences across all languages. For each Transformer layer, we aggregate the token embeddings from all languages into a single matrix. We then apply PCA to this aggregated set of embeddings to obtain a shared low-dimensional projection space. The number of principal components is selected such that the PCA-projected embeddings retain 98% of the total explained variance.

Finally, we fit a GMM to the PCA-projected embeddings for each layer, using one mixture component per language, resulting in 9 clusters. Each

cluster represents a language space parameterized by a mean vector, computed as the average of all PCA-projected token embeddings from that language, and a full covariance matrix estimated from the same set of embeddings.

5 Results

We first examine the overall structure and distribution of language spaces modeled by the GMM. Then, we turn to the dominance scores at the token level to assess how internal representations of a language may be influenced by another.

5.1 Language Dominance Analysis

At the distributional level, if the language dominance hypothesis holds, we would expect substantial overlap between the embedding space of a given language and that of the most frequent or dominant language in the model's pretraining corpus, leading to less distinct language-specific clusters. To evaluate this, we compute the Normalized Mutual Information (NMI) between the GMM components (i.e., language spaces) and the true language labels across all layers of the test models. Low NMI values would indicate that the clusters do not align well with language identities, suggesting a shared representational space or some levels of language space overlap. Conversely, higher NMI scores would reflect more distinct language-specific structure, weakening strong dominance effects.

Figure 1 presents the NMI values across all layers of the models for three clustering tasks: languages included in the models' pretraining data (left), languages absent from pretraining (middle), and the full set of languages (right). The results show a consistent pattern across both models where NMI values start relatively high in the initial layers, drop sharply in the middle layers, and rise again in the upper layers. This pattern holds regardless of whether a language was seen during pretraining and suggests that the middle layers (3-14 in mGPT and 5-18 in BLOOM) are more prone to cross-lingual overlap or shared representation, while both early and late layers encode more language-specific structure.

These results are consistent with research on neuron activation patterns in LLMs (e.g., Kojima et al., 2024; Tang et al., 2024), which report that language-specific neurons are mainly found at the top and bottom internal layers of LLMs, and language-agnostic neurons are found in the middle layers.

A comparison between the two models shows that the intermediate representations of BLOOM are generally more informative about the language labels. This could be attributed to the composition of BLOOM's pretraining data, which includes a more diverse and balanced set of languages compared to mGPT, whose training data is biased toward CIS languages (see Section 4.1).

We further analyze the clustering performance of GMMs with low NMI to investigate whether a dominant language might be responsible for reduced alignment with language identities. Figure 2 presents confusion matrices for the GMM clustering of token embeddings across all languages, for mGPT and BLOOM. The matrices are aggregated over layers 3-14 for mGPT and 5-18 for BLOOM, corresponding to the ranges where the GMMs yield the lowest NMI with the true language labels.

Under the language dominance hypothesis, we would expect a dominant language space to span those of other languages, leading to systematic misclustering toward that language in the confusion matrix. Such a pattern would manifest as disproportionately high off-diagonal values concentrated in the dominant language's column. However, our results show no such consistent misclustering. Instead, for both models, strong diagonal dominance indicates that most tokens are correctly assigned to their respective language clusters.

The results further indicate that misclustered tokens are distributed across all languages, providing little evidence for a single dominant language mediating others. A slightly higher rate of misclustering into Hindi across language pairs might be due to distinct statistical regularities in the Devanagari script, particularly in punctuation usage, or to biases in the composition of the pretraining data. Further work, involving a detailed examination of training data distribution and cross-lingual representational similarities, is needed to better understand the factors driving this pattern. However, such analysis is beyond the scope of the present work.

Next, we analyze the dominance scores for each language, as described in Section 3.2. Figure 3 presents the average dominance scores across layers for both models. We ran a two-sided Mann-Whitney U test, finding significant layer-wise differences in dominance scores for most cases ($p < .05$ in 97% of the cases for mGPT, 100% for BLOOM).[1] Consistent with the previous cluster analysis, Fig-

[1]We used this non-parametric test as it does not assume normality in the distribution of the average dominance scores.

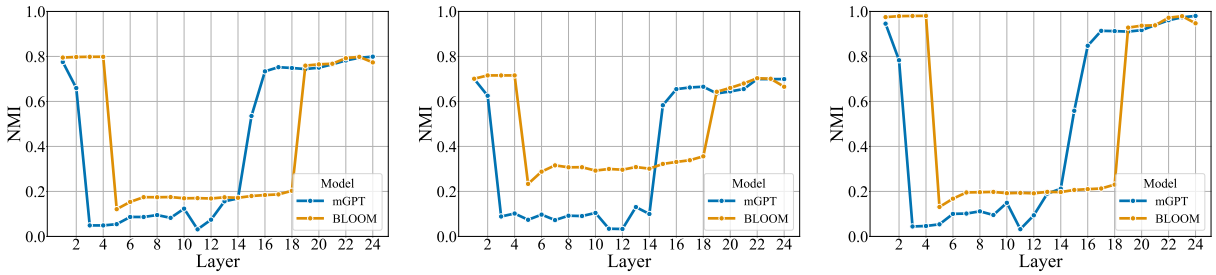

Figure 1: Normalized Mutual Information (NMI) for three language sets: (left) languages present in the models' pretraining data, (middle) languages absent from pretraining, and (right) all languages.

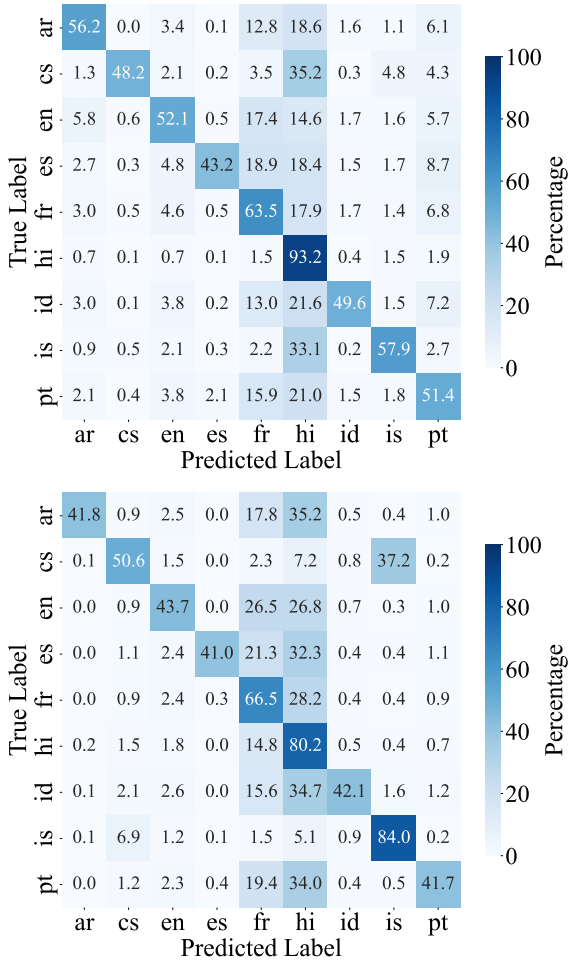

Figure 2: Confusion matrices of the GMM clustering. Top: mGPT; Bottom: BLOOM.

ure 3 shows that language dominance is minimal in the initial and final layers, suggesting that at these stages, the representations are largely language-specific. In parallel, the middle layers, specifically layers 3 to 14 in mGPT and 5 to 18 in BLOOM, form a region where the models internally align or translate representations across languages.

The results do not point to a single dominant language exerting significant influence over all others.[2] Instead, the dominance scores are relatively balanced across the languages, suggesting the formation of a shared representational space in the middle layers rather than reliance on a specific dominant language to mediate cross-lingual processing. Furthermore, the overall dominance scores in the middle layers remain modest, around 0.1, implying that each language dominates only about 10% of tokens from other languages. This score is even lower for the languages outside the training data of the LLMs (between 0.05 and 0.09 for Icelandic and Czech), indicating that these unseen languages are less likely to impose their structure on others.

5.2 Language Dominance per Token

To further investigate this trend at a more granular level, we conduct a token-level analysis based on the likelihood ratio that a token extracted from a sentence in a source language S is internally processed within its own language space versus that of another language T. Specifically, for a contextualized embedding $\tilde{\mathbf{h}}$ originated from an intermediate layer when processing a token in S, we evaluate:

$$\Lambda_{S,T}(\tilde{\mathbf{h}}) = \frac{P(S \mid \tilde{\mathbf{h}})}{P(T \mid \tilde{\mathbf{h}})} \quad (4)$$

A value of $\Lambda_{S,T}(\tilde{\mathbf{h}}) < 1$ indicates that the $\tilde{\mathbf{h}}$'s token, belonging to S, is more likely to be processed within the language space of T rather than S, according to the GMM posterior probabilities. The

[2]The fact that Hindi dominates slightly more tokens than other languages does not stem from any apparent linguistic or computational reason. Further investigation is needed to determine the underlying cause of this discrepancy. One possible explanation is the influence of Hindi's use of the Devanagari script, which may introduce distinct statistical regularities. For instance, the script includes the Danda symbol (।) as a punctuation mark, which might be more predictable within Hindi and, by extension, influence the representation of similar punctuation tokens in other languages. Further, Petrov et al. (2023) have shown that suboptimal tokenization of Hindi can result in longer token sequences than those of Latin-script languages, which in turn may contribute to higher dominance rates.

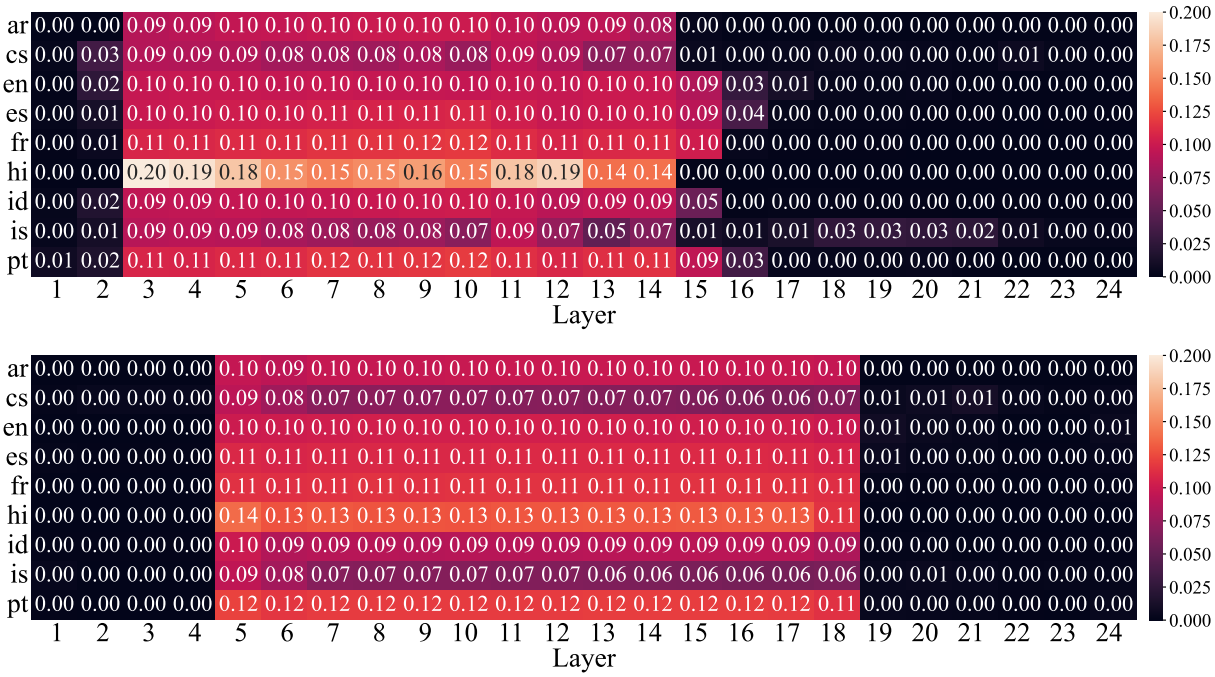

Figure 3: The language dominance scores across layers. Top: mGPT; Bottom: BLOOM.

smaller the value of $\Lambda_{S,T}(\tilde{\mathbf{h}})$, the stronger the evidence that T is acting as an intermediary language for processing the token.

Figure 4 represents the proportion of tokens with $\Lambda < 1$ (i.e., more likely to be processed in a foreign language space) and $\Lambda > 1$ (i.e., more likely to be processed in their own language space) for each layer of the models. Consistent with our previous observations, in both models, although the majority of tokens are ultimately processed within their own language space ($\Lambda > 1$), the early to middle layers (3-14 in mGPT and 5-18 in BLOOM) show a substantial proportion of tokens with $\Lambda < 1$.

Figure 5 presents the aggregated likelihood ratio distributions for tokens with $\Lambda < 1$ across all layers. The strong concentration of mass near zero in both models suggests that a significant number of tokens processed in a language are strongly anchored to a different target language. Notably, mGPT shows more tokens with very low likelihood ratios, suggesting a stronger reliance on shared or dominant language spaces. This could be due to the lower language diversity in the mGPT's training data.

The upward trend in token frequencies shows that, although some tokens are more likely to be processed in the target language space, token representations are still strongly influenced by the source language space. This influence grows as we move toward the source side of the likelihood ratio range, where most tokens are concentrated.[3]

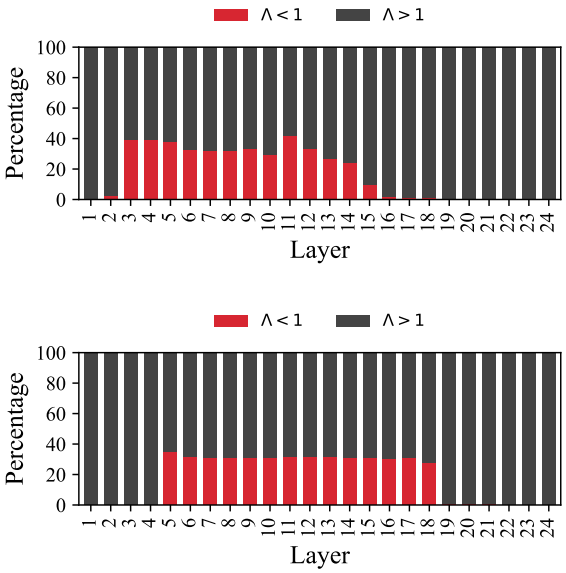

Figure 4: Distribution of tokens with $\Lambda < 1$ and $\Lambda > 1$ per layer (horizontal axis). Top: mGPT; Bottom: BLOOM.

[3]It is also worth mentioning that we conducted an attempt to investigate whether token frequency may correlate with dominance scores. A linear regression model was fitted to analyze the relationship between token frequency and its dominance score (in the form of a likelihood ratio). However, no clear pattern was found.

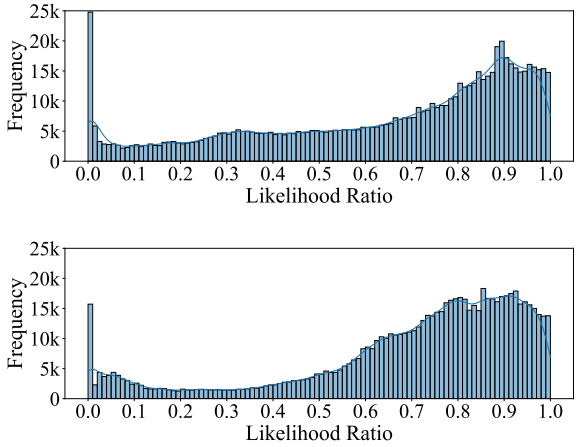

Figure 5: Histograms of likelihood ratios restricted to tokens with $\Lambda_{S,T}(\mathbf{h}) < 1$. The results are aggregated over all layers. Top: mGPT; Bottom: BLOOM.

5.3 Dominated Syntax

To better understand which syntactic categories are most susceptible to cross-lingual dominance, we analyze the part-of-speech (POS) distribution of dominated tokens, i.e., those whose likelihood ratios $\Lambda_{S,T}(\mathbf{h})$ fall below selected thresholds in $\{0.1, 0.2, 0.3, 0.4, 0.5, 1.0\}$.

Figure 6 presents this distribution for both mGPT and BLOOM, aggregated across all layers and target languages. The results show that at lower thresholds (e.g., 0.1 and 0.2), the majority of dominated tokens belong to the categories PUNCT (punctuation) and PROPN (proper nouns). However, as the threshold increases, additional content-bearing categories appear. In particular, grammatical categories such as NOUN, VERB, ADJ, and ADP show a clear upward trend, indicating that more semantically rich tokens are increasingly represented among the dominated words at higher thresholds. Conversely, other semantically-light categories such as DET, PRON, AUX, CCONJ, SCONJ, and PART, which comprise all of the function words except for ADP, follow either a downward or a stable trend (see Appendix A for examples of dominated words). This indicates that the dominated words become increasingly semantically rich as the dominance threshold Λ increases.

To investigate the nature of tokens dominated by another language at the strict threshold of $\Lambda < 0.1$, we conduct a qualitative analysis of English and French as a case study language pair, in which one of the authors is proficient. We focus on POS tags with non-negligible values in Figure 6, namely ADP, DET, NOUN, and PRON. We exclude PUNCT and PROPN, which also appear with notable frequency,

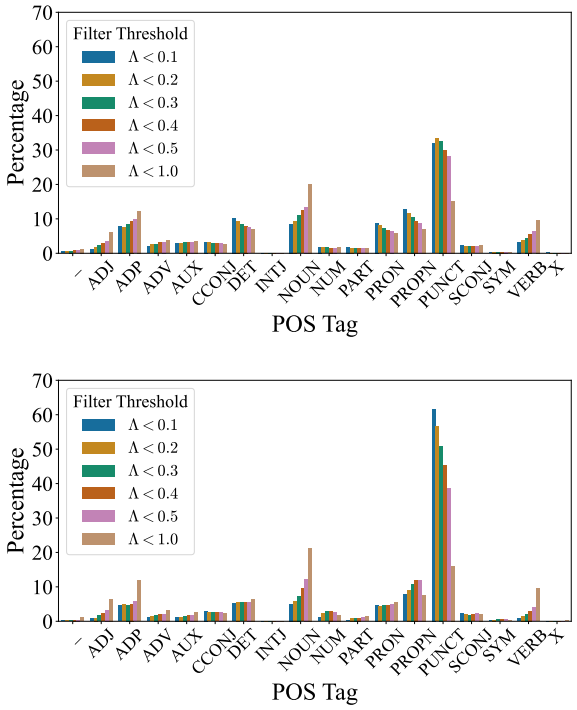

Figure 6: Distribution of dominated words by POS tags. Top: mGPT; Bottom: BLOOM.

but are less linguistically informative for our purposes. Table 2 presents examples of tokens from each selected category that are more likely to be processed within the language space of a different language in mGPT. Similar tendencies to those found in mGPT are observed in BLOOM as well.

en → fr				**fr → en**			
ADP	DET	PRON	NOUN	ADP	DET	PRON	NOUN
to	a	what	Environment	En	La	Il	dollars
by			unit	A	Les	On	M.
in			comparison	Si	Un	Elle	scanner
			decade	De	Des	Ca	Club
			data	Pour	Le	Qui	assemblée
			concert	Malgré	De	Je	expert
			department	Pendant	Cette	Ce	lobbyistes
			contraception	par	Une	Ils	Wifi
			causes	afin	Ce	s'	HFC
			interests	en	un	lui	instituteurs

Table 2: Top 10 tokens per POS for en→fr (left) and fr→en (right) at $\Lambda < 0.1$ taken from mGPT (Source → Target).

Our analysis reveals that, at strict dominance levels ($\Lambda < 0.1$), English and French nouns that are strongly anchored to the other language are predominantly cognates (e.g., *decade* and *décennie*) and borrowed words (e.g., *club* borrowed to French from English). This indicates that in the case of

French and English, tokens from language S that are highly anchored to language T tend to be processed similarly due to shared linguistic form, rather than because language T acts as a dominant intermediary for processing S. Further analysis is needed to confirm whether this observation holds for other language pairs. Nevertheless, this trend is consistent with the observed dominance for the most frequent tags `PUNCT` and `PROPN` in Figure 6, which likewise reflect surface alignment.

A similar pattern is observed for adpositions, which show a higher proportion of cross-lingual anchors with shared orthographic form. For example, the French contraction *d'* of the preposition *de* is orthographically identical to the English reduction *d'* in colloquial forms such as *d'you* (do you). However, we do not find any straightforward pattern for determiners and pronouns as for the other POS tags. Nevertheless, given the extensive borrowing of words and expressions between English and French, it is not surprising that the models exhibit dominance in these categories. Borrowed expressions such as *à la carte* or *c'est la vie* enter English corpora with their original French orthography and grammatical structure, meaning that adpositions and determiners like *à* and *la* appear in both languages in nearly identical form and context. This overlap likely leads the model to process such tokens in a similar space across languages, resulting in the observed cross-lingual dominance.

6 Conclusion

In this work, we investigated the language dominance hypothesis, which suggests that multilingual capabilities of large language models stem from an implicit internal translation process, where input in a source language is transformed into the embedding space of a more dominant language that is heavily represented in the model's pretraining data. To examine this, we proposed a quantitative framework for measuring language dominance based on the interactions of language embedding spaces across intermediate layers of multilingual language models.

Our empirical analysis, conducted on two publicly available multilingual models (mGPT and BLOOM) provides insufficient evidence to support the dominance hypothesis. The results show that, while some degree of cross-lingual influence exists, no single language consistently dominates the internal representations of others. Instead, we observe relatively balanced dominance scores, typically around 10-15%, with no language exerting overwhelming influence. Our findings show that multilingual representation in these models is more likely the result of shared, distributed representations formed in their middle layers rather than a centralized mediation through a dominant language.

Furthermore, our analysis of dominated tokens reveals a clear pattern that tokens strictly processed in a different language space are predominantly function words or items with similar orthographic forms across the two languages, whereas content-bearing words tend to remain within the representational space of their own language.

In future work, we plan to expand our linguistic analysis of dominated tokens to a wider range of languages and language models, enabling a more comprehensive cross-linguistic comparison. We also aim to investigate how additional linguistic factors, such as semantic similarity, morphological complexity, and broader typological relationships, may influence the extent to which one language dominates another within a model's internal representations.

Limitations

Since our data is based on parallel sentences from the PUD treebank, we were unable to collect data for many languages that are shared in the training data for both mGPT and BLOOM. Additionally, due to constraints of computational power, we limited our data to 100 sentences per language.

Future research, therefore, should aim to investigate dominance using high-quality parallel datasets that include more languages and greater linguistic diversity. Additionally, having access to greater computational resources would enable researchers to consider more data and larger LLMs. This would support better validation and generalization across model variants.

Our linguistic analysis of strict dominance is restricted to the English–French pair due to our limited proficiency in other languages. Further investigation is needed to better characterize the nature of dominated words in other language pairs.

Acknowledgments

We thank the anonymous reviewers for their valuable feedback on this paper. We are also grateful to Bolette Sandford Pedersen for her insightful comments. Additionally, we acknowledge the

Danish e-Infrastructure Consortium (DeiC) for providing computational resources through UCloud, supported under the Linguistic Universals in Language Models project.

References

Nora Belrose, Igor Ostrovsky, Lev McKinney, Zach Furman, Logan Smith, Danny Halawi, Stella Biderman, and Jacob Steinhardt. 2025. Eliciting latent predictions from transformers with the tuned lens. *Preprint*, arXiv:2303.08112.

Terra Blevins, Hila Gonen, and Luke Zettlemoyer. 2022. Analyzing the mono- and cross-lingual pretraining dynamics of multilingual language models. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 3575--3590, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, and 12 others. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877--1901. Curran Associates, Inc.

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.

Alexis Conneau and Guillaume Lample. 2019. Cross-lingual language model pretraining. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. 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 4171--4186. Association for Computational Linguistics.

Yanzhu Guo, Simone Conia, Zelin Zhou, Min Li, Saloni Potdar, and Henry Xiao. 2025. Do large language models have an English accent? evaluating and improving the naturalness of multilingual LLMs. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3823--3838, Vienna, Austria. Association for Computational Linguistics.

Takeshi Kojima, Itsuki Okimura, Yusuke Iwasawa, Hitomi Yanaka, and Yutaka Matsuo. 2024. On the multilingual ability of decoder-based pre-trained language models: Finding and controlling language-specific neurons. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 6919--6971. Association for Computational Linguistics.

Xi Victoria Lin, Todor Mihaylov, Mikel Artetxe, Tianlu Wang, Shuohui Chen, Daniel Simig, Myle Ott, Naman Goyal, Shruti Bhosale, Jingfei Du, Ramakanth Pasunuru, Sam Shleifer, Punit Singh Koura, Vishrav Chaudhary, Brian O'Horo, Jeff Wang, Luke Zettlemoyer, Zornitsa Kozareva, Mona Diab, and 2 others. 2022. Few-shot learning with multilingual generative language models. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9019--9052, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

nostalgebraist. 2020. Interpreting *GPT*: *The Logit Lens*. *LessWrong* blog post. `https://www.lesswrong.com/posts/AcKRB8wDpdaN6v6ru/interpreting-gpt-the-logit-lens`.

OpenAI. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.

Isabel Papadimitriou, Kezia Lopez, and Dan Jurafsky. 2023. Multilingual BERT has an accent: Evaluating English influences on fluency in multilingual models. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 1194--1200, Dubrovnik, Croatia. Association for Computational Linguistics.

Aleksandar Petrov, Emanuele La Malfa, Philip Torr, and Adel Bibi. 2023. Language model tokenizers introduce unfairness between languages. In *Advances in Neural Information Processing Systems*, volume 36, pages 36963--36990. Curran Associates, Inc.

Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, Matthias Gallé, Jonathan Tow, Alexander M. Rush, Stella Biderman, Albert Webson, Pawan Sasanka Ammanamanchi, Thomas Wang, Benoît Sagot, Niklas Muennighoff, Albert Villanova del Moral, and 373 others. 2023. Bloom: A 176b-parameter open-access multilingual language model. *Preprint*, arXiv:2211.05100.

Freda Shi, Mirac Suzgun, Markus Freitag, Xuezhi Wang, Suraj Srivats, Soroush Vosoughi, Hyung Won Chung, Yi Tay, Sebastian Ruder, Denny Zhou, Dipanjan Das, and Jason Wei. 2023. Language models are multilingual chain-of-thought reasoners. In *The Eleventh International Conference on Learning Representations*.

Oleh Shliazhko, Alena Fenogenova, Maria Tikhonova, Vladislav Mikhailov, Anastasia Kozlova, and Tatiana

Shavrina. 2022. mgpt: Few-shot learners go multilingual. *arXiv preprint arXiv:2204.07580*.

Tianyi Tang, Wenyang Luo, Haoyang Huang, Dongdong Zhang, Xiaolei Wang, Xin Zhao, Furu Wei, and Ji-Rong Wen. 2024. Language-specific neurons: The key to multilingual capabilities in large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5701--5715, Bangkok, Thailand. Association for Computational Linguistics.

Ahmet Üstün, Viraat Aryabumi, Zheng Yong, Wei-Yin Ko, Daniel D'souza, Gbemileke Onilude, Neel Bhandari, Shivalika Singh, Hui-Lee Ooi, Amr Kayid, Freddie Vargus, Phil Blunsom, Shayne Longpre, Niklas Muennighoff, Marzieh Fadaee, Julia Kreutzer, and Sara Hooker. 2024. Aya model: An instruction finetuned open-access multilingual language model. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15894--15939, Bangkok, Thailand. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Chris Wendler, Veniamin Veselovsky, Giovanni Monea, and Robert West. 2024. Do llamas work in English? on the latent language of multilingual transformers. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15366--15394, Bangkok, Thailand. Association for Computational Linguistics.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483--498, Online. Association for Computational Linguistics.

Xiang Yue, Yueqi Song, Akari Asai, Seungone Kim, Jean de Dieu Nyandwi, Simran Khanuja, Anjali Kantharuban, Lintang Sutawika, Sathyanarayanan Ramamoorthy, and Graham Neubig. 2025. Pangea: A fully open multilingual multimodal LLM for 39 languages. In *The Thirteenth International Conference on Learning Representations*.

Daniel Zeman, Martin Popel, Milan Straka, Jan Hajič, Joakim Nivre, Filip Ginter, Juhani Luotolahti, Sampo Pyysalo, Slav Petrov, Martin Potthast, Francis Tyers, Elena Badmaeva, Memduh Gokirmak, Anna Nedoluzhko, Silvie Cinková, Jan Hajič jr., Jaroslava Hlaváčová, Václava Kettnerová, Zdeňka Urešová, and 43 others. 2017. CoNLL 2017 shared task: Multilingual parsing from raw text to Universal Dependencies. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 1--19, Vancouver, Canada. Association for Computational Linguistics.

Xiang Zhang, Senyu Li, Bradley Hauer, Ning Shi, and Grzegorz Kondrak. 2023. Don`t trust ChatGPT when your question is not in English: A study of multilingual abilities and types of LLMs. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 7915--7927. Association for Computational Linguistics.

Chengzhi Zhong, Qianying Liu, Fei Cheng, Junfeng Jiang, Zhen Wan, Chenhui Chu, Yugo Murawaki, and Sadao Kurohashi. 2025. What language do non-English-centric large language models think in? In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 26333--26346, Vienna, Austria. Association for Computational Linguistics.

A Dominated Tokens

Tables 3 and 4 present the most frequently dominated tokens at various dominance thresholds ($\Lambda < \{0.1, 0.2, 0.3, 0.4, 0.5, 1.0\}$) across all layers of mGPT and BLOOM, respectively. Each cell lists tokens that appear newly at the given threshold, excluding those already observed at lower thresholds. Alongside each token, its part-of-speech (POS) tag and the language pair indicating the direction of dominance (*source language* $\rightarrow$ *target language*) are provided. The table highlights how function words and punctuation dominate at lower thresholds, while more content-bearing words, such as nouns and verbs, emerge as the threshold increases.

	$\Lambda < 0.1$	$\Lambda < 0.2$	$\Lambda < 0.3$	$\Lambda < 0.4$	$\Lambda < 0.5$	$\Lambda < 1.0$
1	. (PUNCT, en → fr)	on (ADP, en → hi)	že (SCONJ, cs → hi)	قالت (VERB, ar → en)	C' (PRON, fr → en)	م (ADP, ar → fr)
2	, (PUNCT, en → pt)	Také (ADV, cs → en)	debaty (NOUN, cs → ar)	því (PRON, is → hi)	D' (ADP, fr → en)	Vladivostoku (PROPN, cs → ar)
3	„ (PUNCT, cs → is)	Sebuah (DET, id → en)	untuk (SCONJ, id → hi)	með (ADP, is → hi)	Naturellement (ADV, fr → en)	Několik (DET, cs → ar)
4	" (PUNCT, en → id)	Seorang (DET, id → en)	dengan (SCONJ, id → hi)	árið (NOUN, is → hi)	Mnoho (DET, cs → en)	lázně (NOUN, cs → ar)
5	इस (DET, hi → ar)	segir (VERB, is → ar)	která (DET, cs → es)	jedna (NUM, cs → hi)	Début (NOUN, fr → en)	terletak (VERB, id → ar)
6	a (DET, en → fr)	उत्तरी (ADJ, hi → ar)	jeho (DET, cs → hi)	zástupci (NOUN, cs → hi)	Contudo (CCONJ, pt → en)	plánů (NOUN, cs → ar)
7	“ (PUNCT, en → hi)	Polisi (NOUN, id → en)	bylo (AUX, cs → hi)	frá (ADP, is → ar)	Primeiro (ADV, pt → en)	Þeir (PRON, is → es)
8	A (CCONJ, cs → pt)	लेकिन (CCONJ, hi → is)	investice (NOUN, cs → ar)	pravděpodobně (ADV, cs → hi)	výše (NOUN, cs → hi)	případu (NOUN, cs → ar)
9	En (ADP, fr → is)	उस (DET, hi → ar)	al (_, es → hi)	jejich (DET, cs → hi)	vypukla (VERB, cs → hi)	Vísindamenn (NOUN, is → ar)
10	Ia (PRON, id → ar)	अध्यक्ष (NOUN, hi → ar)	si (SCONJ, es → hi)	þegar (SCONJ, is → hi)	Snemma (ADV, is → en)	hal (NOUN, id → ar)
11	La (DET, fr → es)	जिसे (PRON, hi → ar)	u (ADP, cs → hi)	þau (PRON, is → hi)	leyst (VERB, is → hi)	terampil (ADJ, id → ar)
12	و (PART, ar → en)	Trumpová (PROPN, cs → ar)	sur (ADP, fr → hi)	draga (VERB, is → hi)	áratug (NOUN, is → hi)	digunakan (VERB, id → ar)
13	Í (ADP, is → ar)	ذات (NOUN, ar → hi)	su (PRON, es → hi)	udělat (VERB, cs → hi)	rozhodnutí (NOUN, cs → hi)	panjang (ADJ, id → ar)
14	यह (PRON, hi → ar)	Nové (ADJ, cs → en)	ऑफ (ADP, hi → ar)	část (NOUN, cs → hi)	cílů (NOUN, cs → hi)	behind (ADP, en → ar)
15	Michael (PROPN, cs → is)	Daripada (SCONJ, id → en)	það (PRON, is → hi)	fyrir (ADP, is → hi)	Pařížská (ADJ, cs → ar)	mengikuti (VERB, id → ar)
16	वे (PRON, hi → ar)	Sepanjang (NOUN, id → en)	ze (ADP, cs → hi)	borð (NOUN, is → hi)	zjevně (ADV, cs → hi)	Paříže (PROPN, cs → ar)
17	और (CCONJ, hi → ar)	يحملون (VERB, ar → hi)	dolarů (NOUN, cs → fr)	součástku (NOUN, cs → hi)	lýðveldisins (NOUN, is → hi)	berbagai (DET, id → ar)
18	एक (DET, hi → ar)	يوم (ADV, ar → hi)	charakter (NOUN, cs → es)	hjá (ADP, is → hi)	kvikmyndasamtökunum (NOUN, is → hi)	ditutup (VERB, id → ar)
19	वह (PRON, hi → ar)	Bagi (ADP, id → en)	Georgetownské (ADJ, cs → ar)	بدأت (VERB, ar → en)	Samkvæmt (ADP, is → es)	menentang (VERB, id → ar)
20	सीगल (PROPN, hi → ar)	Mungkin (ADV, id → en)	वर्षों (NOUN, hi → ar)	كتب (VERB, ar → en)	spíra (VERB, is → hi)	námořnictva (NOUN, cs → ar)
21	sem (ADP, pt → is)	Banyak (DET, id → en)	número (NOUN, es → hi)	hřišti (NOUN, cs → hi)	kynni (VERB, is → hi)	přímou (ADJ, cs → ar)
22	Trump (PROPN, en → is)	za (ADP, cs → fr)	prezidentem (NOUN, cs → es)	učí (VERB, cs → hi)	upplýsingar (NOUN, is → hi)	بعيدة (ADJ, ar → fr)
23	आज (NOUN, hi → ar)	procent (NOUN, cs → ar)	filmem (NOUN, cs → ar)	řeku (NOUN, cs → hi)	fasteignasölum (NOUN, is → hi)	průměru (NOUN, cs → ar)
24	Di (ADP, id → ar)	Kdo (PRON, cs → en)	typicky (ADV, cs → ar)	někoho (PRON, cs → hi)	auknar (ADJ, is → hi)	sel (NOUN, id → ar)
25	قال (VERB, ar → en)	Frekar (ADV, is → en)	milljarðar (NOUN, is → ar)	měl (VERB, cs → hi)	Pedimos (VERB, pt → en)	podezřelého (ADJ, cs → ar)
26	Um (ADP, is → pt)	Apa (PRON, id → en)	sebuah (DET, id → hi)	shledáno (ADJ, cs → hi)	Státech (NOUN, cs → ar)	qualquer (DET, pt → ar)
27	Para (ADP, pt → id)	Studi (NOUN, id → en)	bahwa (SCONJ, id → hi)	celosvětové (ADJ, cs → hi)	průmysl (NOUN, cs → hi)	उसकी (PRON, hi → ar)
28	कई (DET, hi → ar)	Hari (NOUN, id → en)	sus (PRON, es → hi)	omezování (NOUN, cs → hi)	vlastní (ADJ, cs → hi)	العمل (NOUN, ar → es)
29	उसने (PRON, hi → ar)	Satu (NUM, id → en)	abychom (_, cs → es)	Breska (ADJ, is → es)	Občas (ADV, cs → en)	Jakožto (SCONJ, cs → en)
30	RECO (PROPN, en → id)	Janji (NOUN, id → en)	उसे (PRON, hi → pt)	تصدر (VERB, ar → en)	ředitel (NOUN, cs → hi)	deviam (AUX, pt → ar)
31	Mais (ADV, pt → fr)	Beberapa (DET, id → en)	مجدداً (ADV, ar → hi)	يقابل (VERB, ar → en)	svůj (DET, cs → hi)	própria (ADJ, pt → ar)
32	O (PRON, pt → en)	Kadang (ADV, id → en)	biasa (ADJ, id → hi)	opustit (VERB, cs → hi)	mínútur (NOUN, is → hi)	septiembre (NOUN, es → ar)
33	Na (_, pt → cs)	Hanya (ADV, id → en)	harus (AUX, id → hi)	skotským (ADJ, cs → hi)	grunaða (ADJ, is → hi)	वास्तव (NOUN, hi → ar)
34	Seagal (PROPN, en → cs)	Saya (PRON, id → en)	hingga (ADP, id → hi)	nezávislosti (NOUN, cs → hi)	bakslagi (NOUN, is → hi)	conduziram (VERB, pt → ar)
35	Je (PRON, fr → cs)	Sementara (ADV, id → en)	lainnya (_, id → hi)	záměr (NOUN, cs → hi)	námskeiðinu (NOUN, is → hi)	habitual (ADJ, es → fr)
36	Klein (PROPN, fr → en)	Bukan (PART, id → en)	až (PART, cs → hi)	neblíží (VERB, cs → hi)	þyrlum (NOUN, is → hi)	Nokkrir (PRON, is → ar)
37	Os (DET, pt → en)	आया (VERB, hi → ar)	sociálních (ADJ, cs → ar)	června (NOUN, cs → hi)	truflaninni (NOUN, is → hi)	substituiu (VERB, pt → ar)
38	V (ADP, cs → ar)	Realitní (ADJ, cs → ar)	einu (PRON, is → fr)	jehož (DET, cs → hi)	koltvísýring (NOUN, is → hi)	pública (ADJ, pt → ar)
39	Fallon (PROPN, cs → is)	technologie (NOUN, cs → ar)	dans (ADP, fr → hi)	měchýř (NOUN, cs → hi)	hvergi (ADV, is → hi)	प्रमुख (ADJ, hi → fr)
40	Ontario (X, fr → en)	Paling (ADV, id → en)	byla (AUX, cs → hi)	vidět (VERB, cs → hi)	tortímdu (NOUN, is → hi)	45वें (ADJ, hi → ar)

Table 3: Top dominated tokens at increasing dominance thresholds, excluding duplicates from earlier thresholds. The cells represent Token | POS | Source → Target) taken from mGPT.

	$\Lambda < 0.1$	$\Lambda < 0.2$	$\Lambda < 0.3$	$\Lambda < 0.4$	$\Lambda < 0.5$	$\Lambda < 1.0$
1	. (PUNCT, en → pt)	و (PART, ar → en)	Perusahaan (NOUN, id → en)	Það (PRON, is → ar)	nya (PRON, id → hi)	sobre (ADP, pt → hi)
2	, (PUNCT, cs → is)	لكن (ADP, ar → en)	Sebuah (DET, id → en)	कि (SCONJ, hi → en)	Kami (PRON, id → hi)	ujar (VERB, id → ar)
3	„ (PUNCT, cs → is)	Los (DET, es → en)	ट्यूमर (NOUN, hi → is)	में (ADP, hi → fr)	mungkin (ADV, id → hi)	that (SCONJ, en → es)
4	“ (PUNCT, en → id)	El (DET, es → en)	Bagi (ADP, id → en)	Agreement (NOUN, en → hi)	termasuk (VERB, id → hi)	oleh (ADP, id → cs)
5	En (ADP, fr → is)	قال (VERB, ar → en)	Apa (PRON, id → en)	Estate (NOUN, en → hi)	Británica (ADJ, es → hi)	tumor (NOUN, id → ar)
6	A (DET, en → cs)	Hann (PRON, is → ar)	Mungkin (ADV, id → en)	right (ADJ, en → is)	adolescente (NOUN, fr → hi)	नहीं (PART, hi → en)
7	लेकिन (CCONJ, hi → ar)	من (ADP, ar → en)	Mereka (PRON, id → en)	Council (NOUN, en → hi)	grade (NOUN, en → hi)	že (SCONJ, cs → is)
8	इस (DET, hi → ar)	También (ADV, es → en)	Hari (NOUN, id → en)	Metropolitan (ADJ, en → hi)	émissions (NOUN, fr → hi)	العديد (NOUN, ar → es)
9	Seagal (PROPN, en → cs)	س (PART, ar → en)	Satu (NUM, id → en)	cut (VERB, en → hi)	llamados (VERB, es → hi)	dijo (VERB, es → fr)
10	Í (ADP, is → ar)	Þetta (PRON, is → ar)	Sangat (ADV, id → en)	counsellors (NOUN, en → hi)	نهر (PROPN, ar → is)	with (ADP, en → es)
11	Michael (PROPN, cs → is)	قالت (VERB, ar → en)	Ini (PRON, id → en)	Korean (ADJ, en → hi)	transit (NOUN, en → hi)	new (ADJ, en → ar)
12	Trump (PROPN, fr → es)	يقول (VERB, ar → en)	Hanya (ADV, id → en)	Russia (PROPN, en → hi)	plans (NOUN, en → hi)	al (_, es → fr)
13	Na (_, pt → cs)	Es (AUX, es → en)	Hasil (NOUN, id → en)	Mailis (PROPN, en → hi)	Democratic (ADJ, en → hi)	et (CCONJ, fr → es)
14	वह (PRON, hi → ar)	श्रीमान (PART, hi → ar)	Sebagian (NOUN, id → en)	وجود (NOUN, ar → fr)	visage (NOUN, fr → hi)	will (AUX, en → ar)
15	वे (PRON, hi → ar)	على (ADP, ar → en)	Saya (PRON, id → en)	pressing (VERB, en → hi)	batterie (NOUN, fr → hi)	esta (PRON, pt → hi)
16	एक (DET, hi → ar)	Če (PRON, fr → en)	Menurut (ADP, id → en)	able (ADJ, en → hi)	area (NOUN, id → fr)	transisi (NOUN, id → es)
17	यह (PRON, hi → ar)	Nous (PRON, fr → en)	Saat (NOUN, id → en)	Luckily (ADV, en → hi)	كولومبيا (PROPN, ar → hi)	media (NOUN, en → ar)
18	Trudeau (PROPN, cs → is)	Esto (DET, es → en)	GCHQ (PROPN, en → hi)	girl (NOUN, en → hi)	tournage (NOUN, fr → hi)	شرطة (NOUN, ar → es)
19	) (PUNCT, pt → fr)	Según (ADP, es → en)	grossit (VERB, fr → is)	do (AUX, en → is)	Sayangnya (ADV, id → hi)	nama (NOUN, id → ar)
20	a (CCONJ, cs → ar)	जुलाई (NOUN, hi → ar)	untuk (SCONJ, id → fr)	airline (NOUN, en → hi)	porc (NOUN, fr → hi)	कहा (VERB, hi → es)
21	Je (PRON, fr → cs)	Polisi (NOUN, id → en)	germinate (VERB, en → hi)	punishments (NOUN, en → hi)	matériel (NOUN, fr → hi)	العام (ADJ, ar → es)
22	V (ADP, cs → ar)	lung (NOUN, en → cs)	मार्शल (NOUN, hi → cs)	namun (CCONJ, id → fr)	says (VERB, en → hi)	ترامب (PROPN, ar → es)
23	" (PUNCT, en → id)	قد (PART, ar → is)	bahwa (SCONJ, id → fr)	lobbyists (NOUN, en → hi)	Crimée (PROPN, fr → hi)	أكبر (ADJ, ar → es)
24	sem (SCONJ, is → ar)	1 (PUNCT, hi → en)	tumour (NOUN, en → hi)	niveaux (NOUN, fr → hi)	data (NOUN, id → fr)	sa (DET, fr → es)
25	Um (DET, pt → is)	सुश्री (PART, hi → ar)	titres (NOUN, fr → cs)	écoliers (NOUN, fr → hi)	يوم (PROPN, ar → hi)	agenda (NOUN, en → ar)
26	Po (ADP, cs → ar)	Ontario (PROPN, is → ar)	पोटोमैक (PROPN, hi → cs)	collège (NOUN, fr → hi)	hanya (ADV, id → hi)	اسم (NOUN, ar → es)
27	आज (NOUN, hi → ar)	Snemma (ADV, is → ar)	wi-fi (NOUN, en → hi)	immigration (NOUN, en → hi)	cutting (VERB, en → hi)	122 (NUM, ar → es)
28	पहले (ADV, hi → ar)	अधिकतम (ADJ, hi → en)	ruin (VERB, en → hi)	rocket (NOUN, en → hi)	compagnie (NOUN, fr → hi)	other (ADJ, en → ar)
29	और (CCONJ, hi → ar)	17 (NUM, is → ar)	2C (NOUN, en → hi)	out (ADP, en → hi)	hauts (ADJ, fr → hi)	saja (ADV, id → cs)
30	कई (DET, hi → ar)	Byla (AUX, cs → ar)	contraceptives (NOUN, en → hi)	Station (NOUN, en → hi)	lives (NOUN, en → hi)	nasional (ADJ, id → ar)
31	हालांकि (SCONJ, hi → ar)	माइकल (PROPN, hi → en)	steel (NOUN, en → hi)	hearing (NOUN, en → hi)	Brasil (PROPN, id → fr)	ele (PRON, pt → hi)
32	उसने (PRON, hi → ar)	अधिकतर (DET, hi → ar)	vaguely (ADV, en → hi)	internet (NOUN, id → fr)	crecer (VERB, es → hi)	الذين (PRON, ar → es)
33	Sem (SCONJ, is → ar)	apuñalamiento (NOUN, es → is)	gondola (NOUN, en → hi)	crying (VERB, en → hi)	distrito (NOUN, es → hi)	before (ADP, en → ar)
34	((PUNCT, is → hi)	konser (NOUN, id → cs)	Scottish (ADJ, en → hi)	métro (NOUN, fr → hi)	vejiga (NOUN, es → hi)	transition (NOUN, en → ar)
35	Australia (PROPN, cs → ar)	खेलों (NOUN, hi → en)	seeds (NOUN, en → hi)	indiqué (VERB, fr → hi)	الإيديوم (NOUN, ar → hi)	national (ADJ, en → ar)
36	Vladivostok (PROPN, en → cs)	ट्रम्प (PROPN, hi → en)	spokesman (NOUN, en → hi)	prénoms (NOUN, fr → hi)	النووي (ADJ, ar → hi)	लोगों (NOUN, hi → fr)
37	10 (NUM, hi → ar)	11 (NUM, cs → hi)	behaviour (NOUN, en → hi)	vessie (NOUN, fr → hi)	climatique (ADJ, fr → hi)	fiable (ADJ, fr → es)
38	O (PRON, pt → en)	helicopters (NOUN, en → hi)	asphyxiant (ADJ, fr → cs)	judiciaire (ADJ, fr → hi)	Shenzhen (PROPN, fr → hi)	الطيران (NOUN, ar → es)
39	? (PUNCT, is → ar)	لم (PART, ar → en)	way (NOUN, en → hi)	الهرمونية (ADJ, ar → hi)	ejecutivo (ADJ, es → hi)	they (PRON, en → ar)
40	RHS (PROPN, cs → is)	لا (PART, ar → en)	punish (VERB, en → hi)	rhetoric (NOUN, en → hi)	time (NOUN, en → hi)	pero (CCONJ, es → fr)

Table 4: Top dominated tokens at increasing dominance thresholds, excluding duplicates from earlier thresholds. The cells represent Token | POS | Source → Target) taken from BLOOM.