Exploring Alignment in Shared Cross-Lingual Spaces

Basel Mousi, Nadir Durrani, Fahim Dalvi, Majd Hawasly, Ahmed Abdelali
{bmousi,ndurrani,faimaduddin}@hbku.edu.qa,aabdelali@sdaia.gov.sa

QCRI معهد قطر لبحوث الحوسبة Qatar Computing Research Institute جامعة حمد بن خليفة HAMAD BIN KHALIFA UNIVERSITY

SDAIA الهيئة السعودية للبيانات والذكاء الاصطناعي Saudi Data & AI Authority

Motivation

Multilingual models and embeddings achieve high performance, but how do these represent different languages, their similarities and differences?

To what extent do latent spaces across languages exhibit **alignment** and **overlap** in multilingual models?

What makes phenomena like zero-shot **cross-lingual** transfer, possible?

How does the **alignment** change after finetuning models?

Experimental Setup

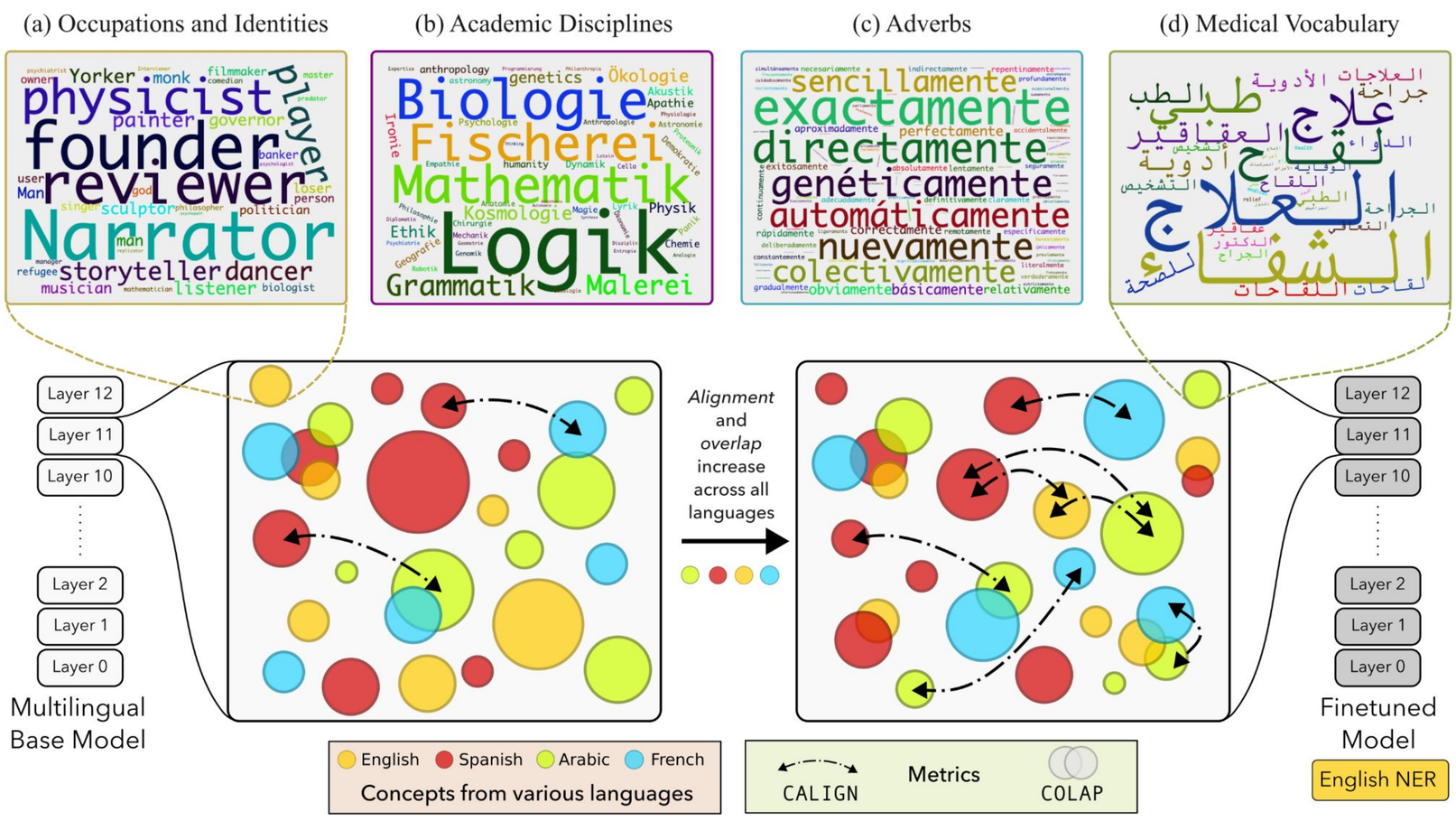
Models: mT5, mBERT, and XLM-R
Languages: English, Spanish, Arabic, French, German
Tasks: Machine translation, Named entity recognition and Sentiment Analysis

Methodology

Discovery: Cluster contextualized representations across layers to discover concepts

Alignment: Measure the extent of semantic alignment between concepts

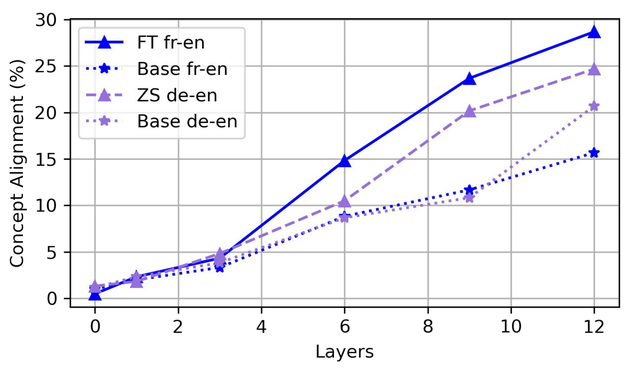
Overlap: Measure how many concepts constitute words from more than one language



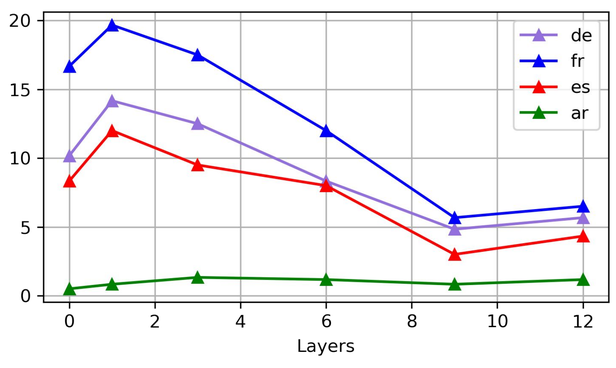
Results

mT5 finetuned on FR-EN translation

Alignment across French/German and English concepts

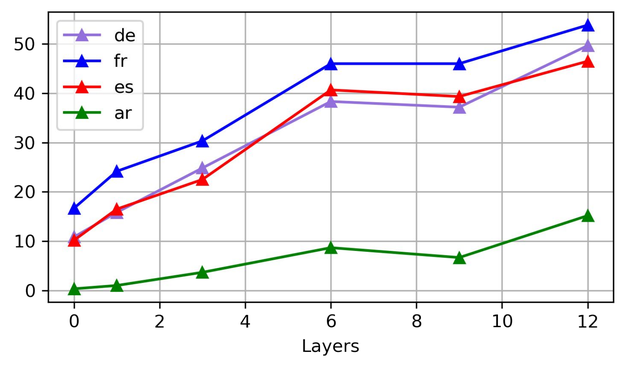


- **Deeper Layers** show **Increased Alignment** due to Semantic Concepts
- **Fine-tuning** calibrates the space towards **higher-alignment**
- **Task-Specific** calibration of latent space facilitates **zero-shot capabilities**



mT5 finetuned decoder
Overlap of various languages with English

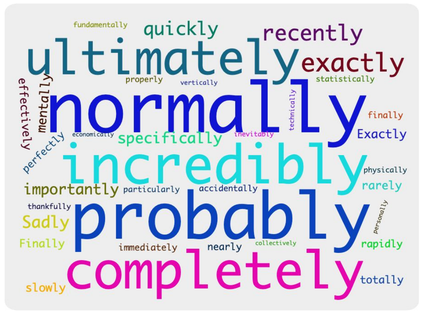
Deeper layers **preserve** language specific concepts after finetuning



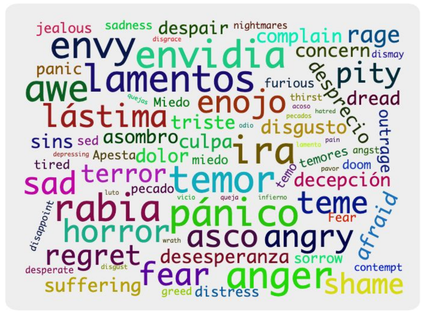
mT5 finetuned encoder
Overlap of various languages with English

Closely related languages **demonstrate higher overlap**

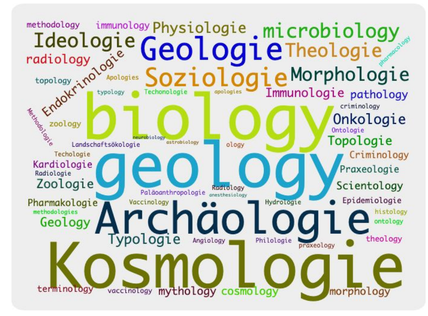
Aligned clusters from mT5's encoder across English, Spanish, Arabic and German representing colors



Aligned clusters from French-English tuned mT5 model representing Adverbs



Emotions in Spanish and English (tuned mT5 encoder)



Shared infix "olog" in German and English (mT5 encoder)