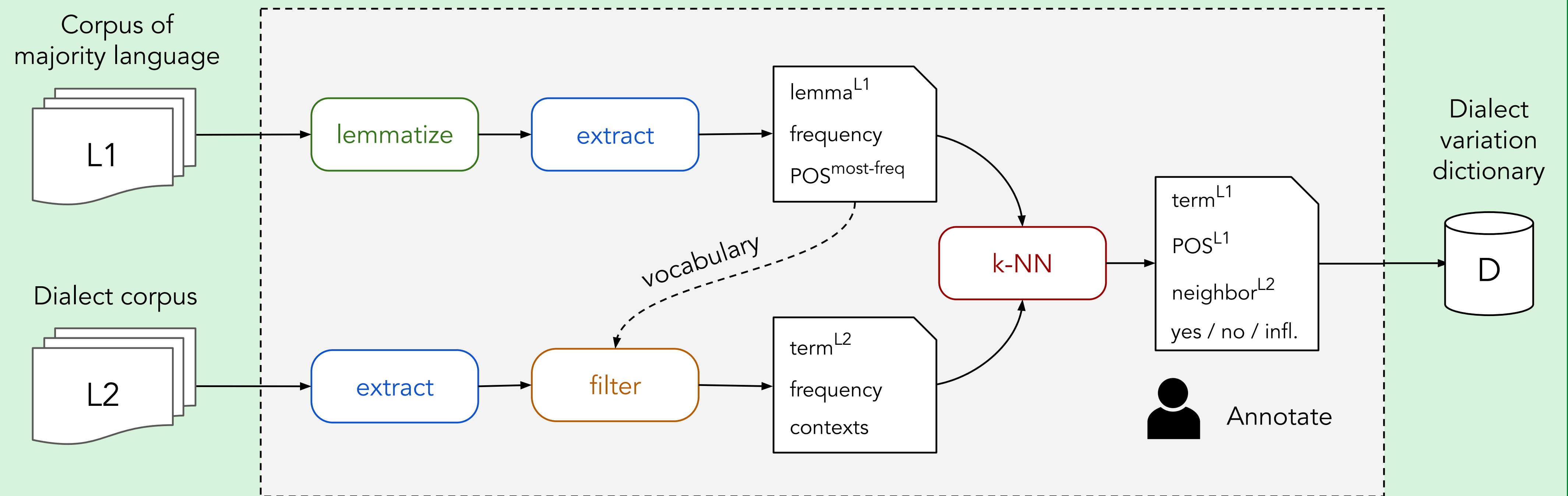


Annotation Framework

- **DiaLemma**: A novel framework to build dialect variation dictionaries.
- We **annotated 100K Bavarian words** with high string similarity to German lemmas: translation, inflection, unrelated?
- **Variation Dictionary**: From the annotated dataset, we extract dialect variants for 5,124 German lemmas.

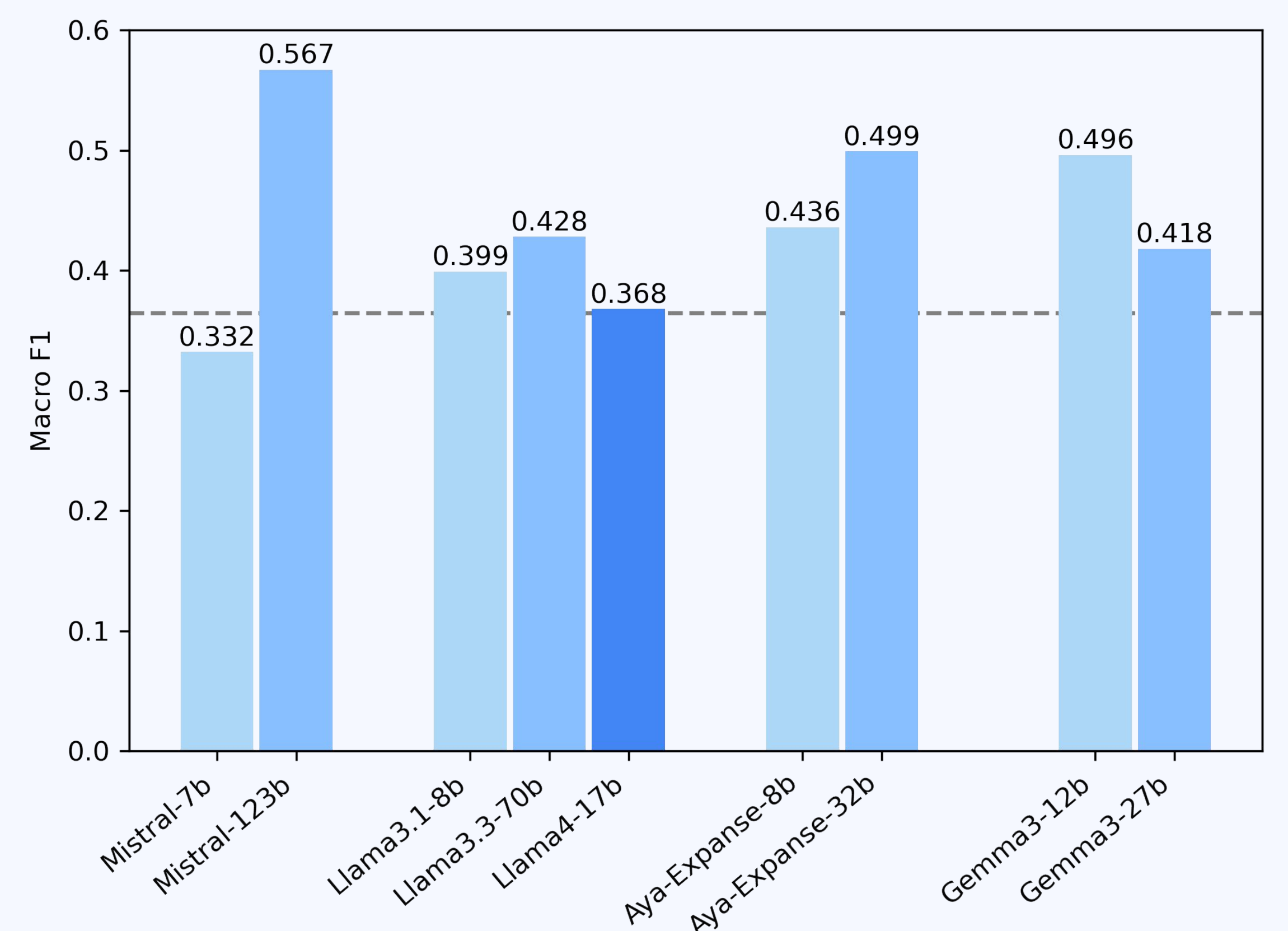


Judging Translation Candidates

German Lemma	Bavarian Candidate	Translation?
zweisprachig ("bilingual")	zwaasprochig	yes
	zwasprächig	yes
	zwoasprachign	inflected
	measprochig ("multilingual")	no
	...	...

--- Logistic regression

- **LLMs perform poorly as dialect judges** (best F1: 0.567).
- Smaller LLMs are **biased towards predicting "no"**, larger models identify dialect variants ("yes") more accurately (see below).

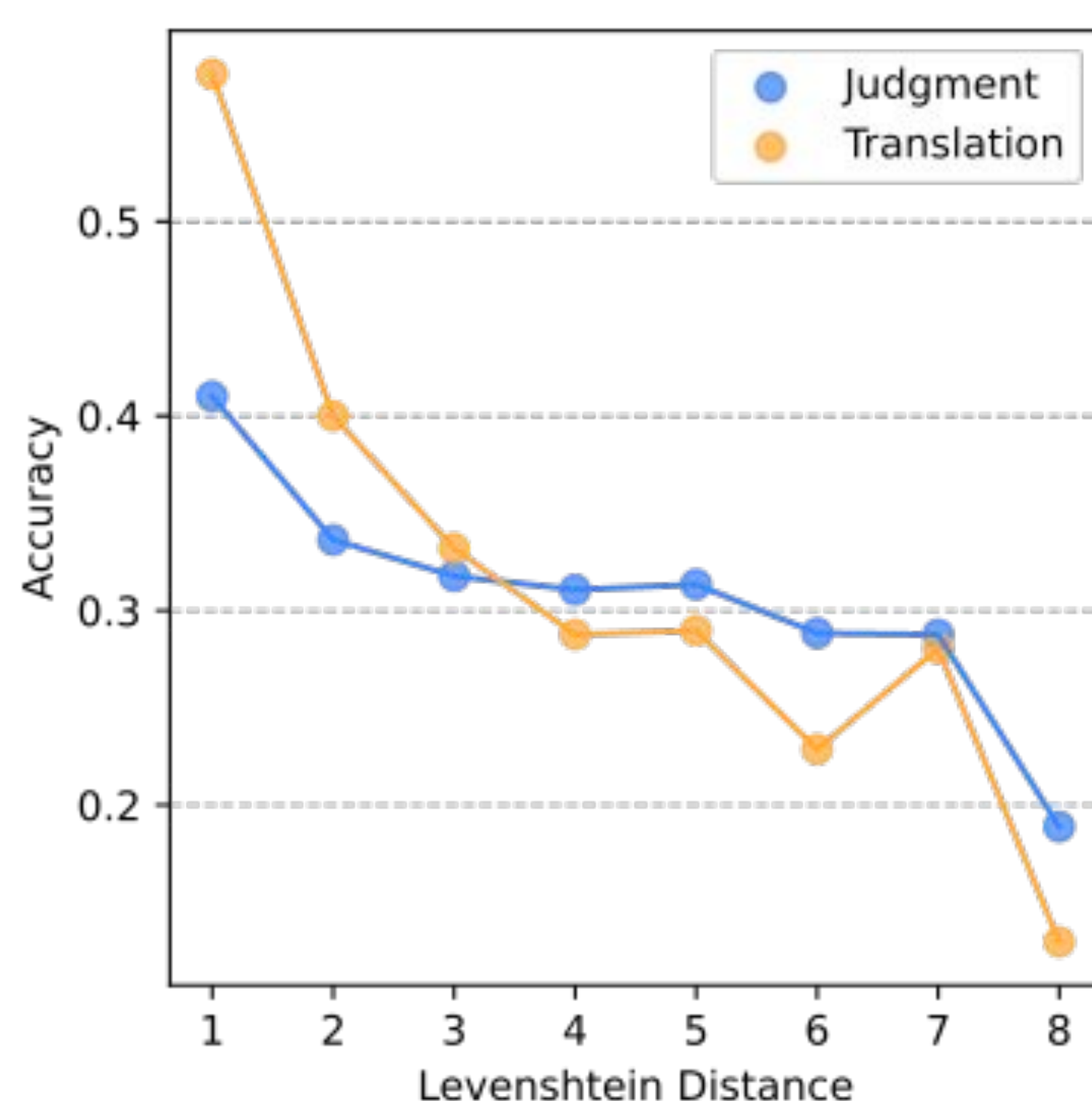
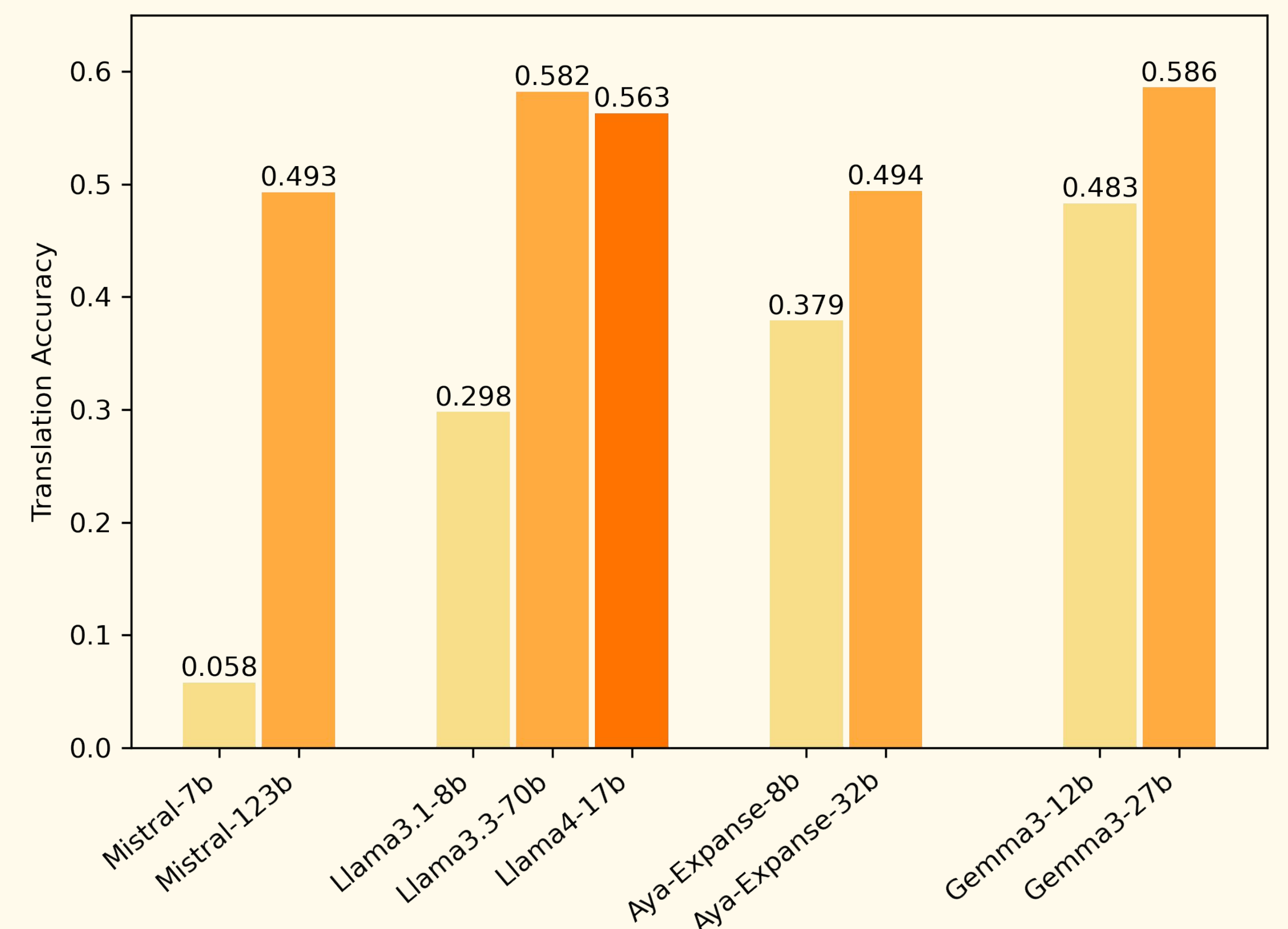


Dialect-to-Standard Translation

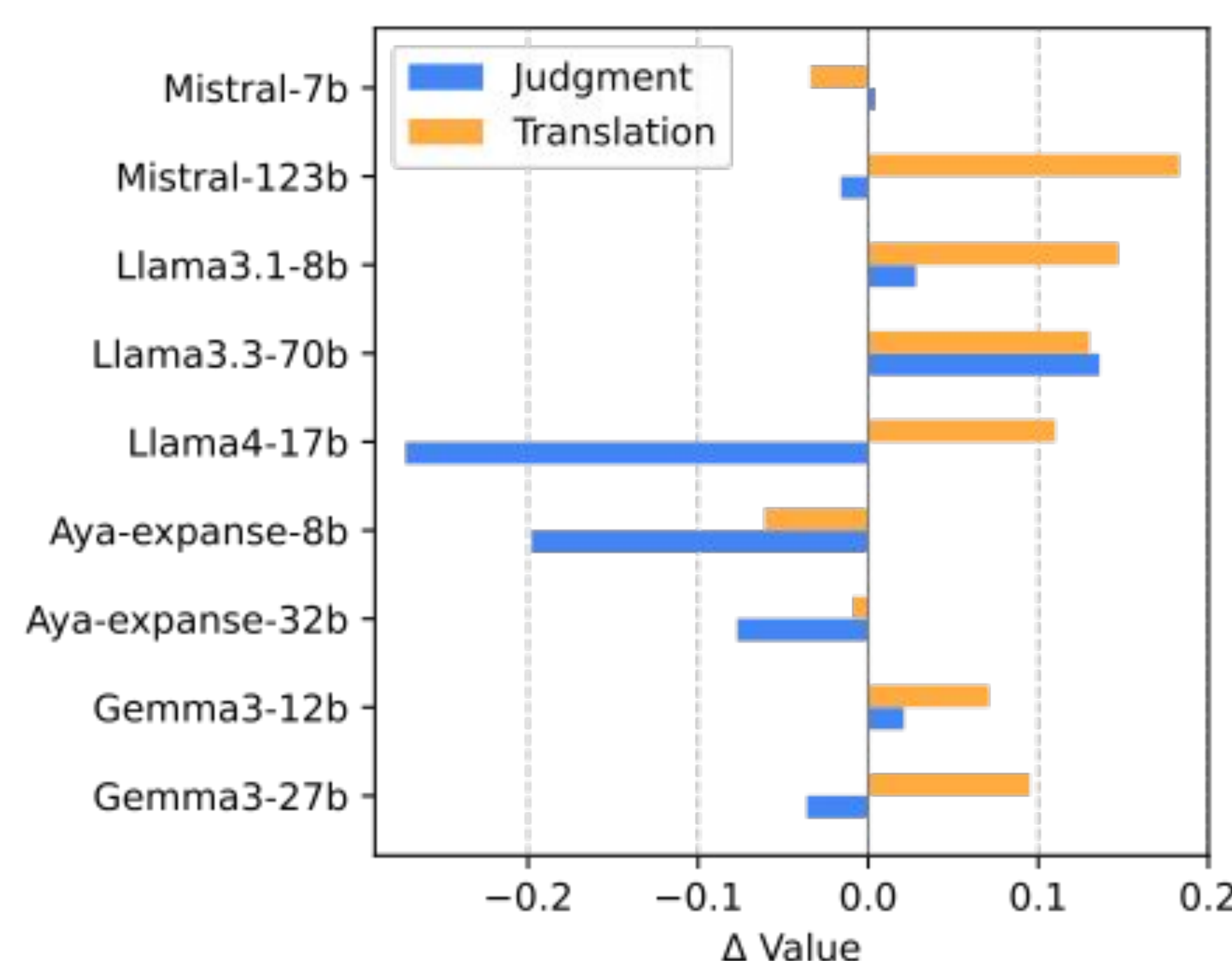
Bavarian (source)	German Lemma (target)
zwaasprochig	zweisprachig
zwasprächig	zweisprachig
dozwischn	dazwischen
...	...

Types of errors made by
Gemma3.27b

- Wrong derivational morphology:
 - **Literaturwissnschoftla** ("literary scholar"; German: Literaturwissenschaftler) gets translated to **Literaturwissenschaft** ("literary studies").
- Only part of the word is translated correctly:
 - **labaprfung** ("audit", lit. "over+test"; German: Überprüfung) becomes **Jahresprüfung** ("yearly test", lit. "year+test").



Accuracies for **judging translation candidates** (excl. majority class 'no') and for **translation** with respect to Levenshtein distances.



Context effects - Overall changes (Δ) in macro F1 for **judging translation candidates** and **translation accuracy**.

Predicted	Actual		
	yes	inflected	no
yes	0.76%	0.41%	0.52%
inflected	5.50%	7.09%	3.16%
no	93.46%	92.29%	96.27%
IF Error	0.28%	0.21%	0.05%

Confusion matrix for Mistral-7b: F1 scores per class: 0.015 (yes), 0.093 (inflected), 0.888 (no).

Predicted	Actual		
	yes	inflected	no
yes	72.05%	47.93%	8.01%
inflected	1.02%	15.97%	1.62%
no	25.86%	34.89%	87.45%
IF Error	1.07%	1.20%	2.91%

Confusion matrix for Mistral-123b: F1 scores per class: 0.550 (yes), 0.234 (inflected), 0.902 (no).