

A Dataset for Metaphor Detection in Early Medieval Hebrew Poetry

The corpus of late antique and medieval Hebrew is vast and represents a crucial linguistic and cultural bridge between Biblical and modern Hebrew. Poetry is prominent in texts from these time periods and one of its characteristics is the use of metaphor. Distinguishing figurative and literal words or phrases is a major task for scholars in the Humanities, especially in the fields of literature, linguistics, and hermeneutics. We present a new, challenging dataset of early medieval Hebrew poetry with expert annotations of metaphors with baseline results.

The Dataset

- Early medieval Hebrew poems.
- 180 labeled texts, 7,000 sentences, 45,000 words.
- 15% of the words are metaphors.

Hebrew	English	Meaning
טבענו בגזרות, כשטופי ים	We drowned in decrees, like sea washed	There are too many decrees
הצית נשמה ויבער נר	Ignite a soul, fire a candle	Create a soul
עד דור עשירי לא עשו פרי	Until the tenth generation they did not bear fruit	For ten generations, they did no good deeds

Inter Annotator Agreement

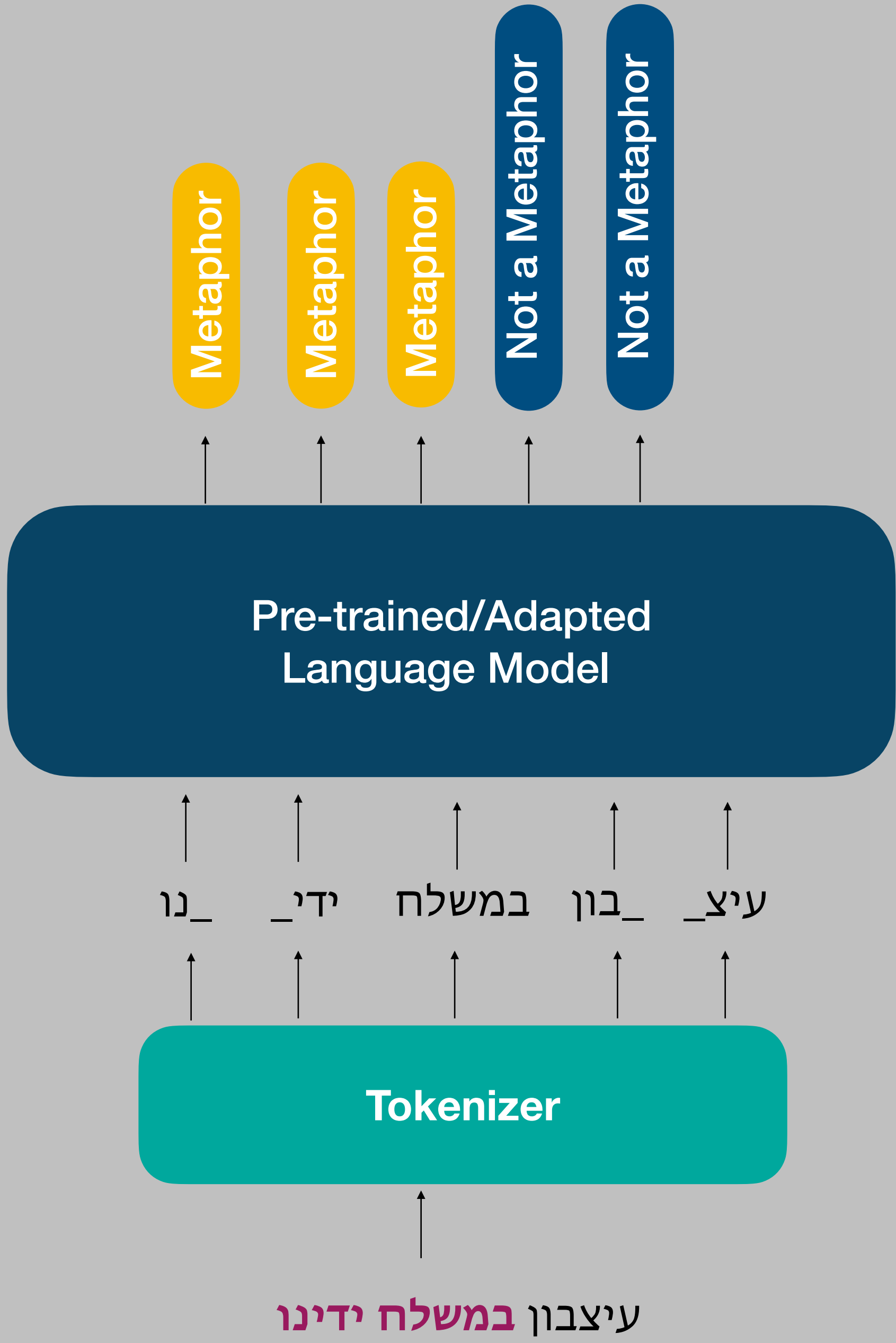
Inter-annotator agreement evaluates how consistently different annotators label data for NLP research (measured in Cohen's kappa). The score for our dataset is 0.618 (considered substantial).



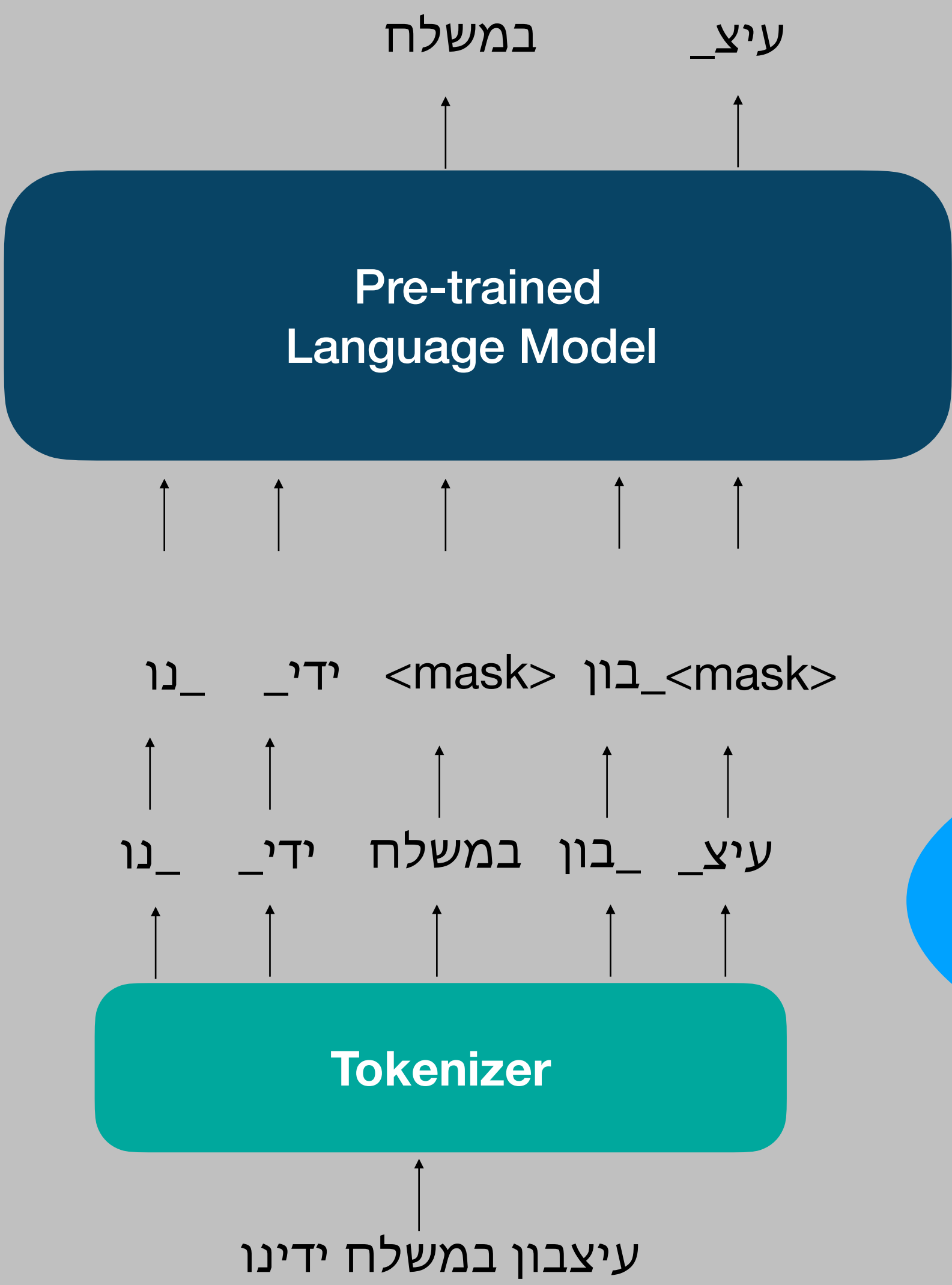
A medieval Hebrew poem from the Worms Prayerbook, Germany, 13th Century, The National Library of Israel, "Ktiv" Project

Methods

Fine-tuning for metaphor detection



LM adaptation



Language models (pre-trained & adapted)

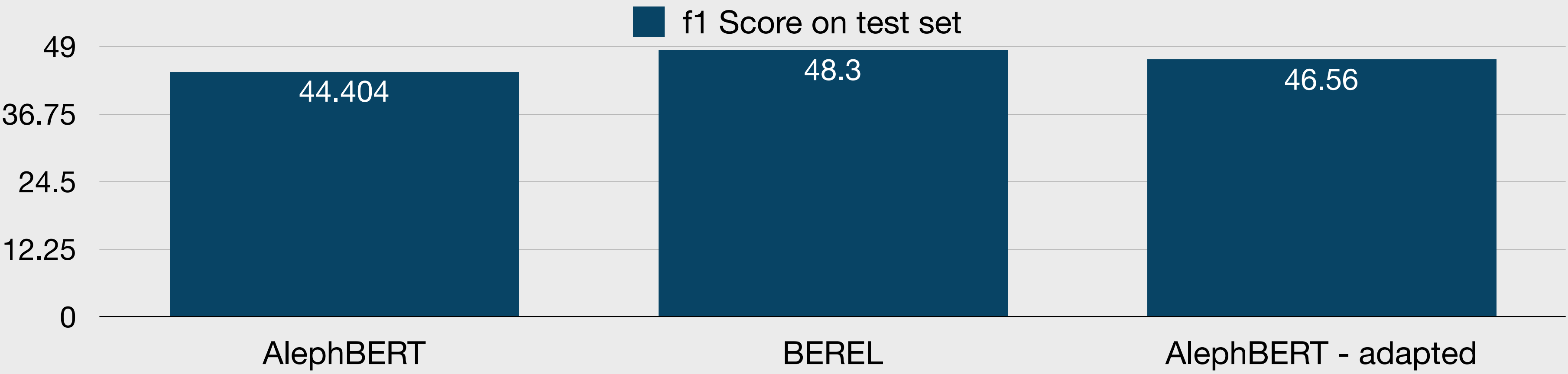
	Data Size (words)	Data Type
BEREL	220 Million	Medieval
AlephBERT	2 Bilion	Modern
AlephBERT Adapted	2 Bilion + 10 Milion	Modern + Medieval

Trained on additional prose and poetic medieval texts (Project Ben-Yehuda*) + Piyyut dataset

*Project Ben-Yehuda is a volunteer-based online library of free Hebrew literature.

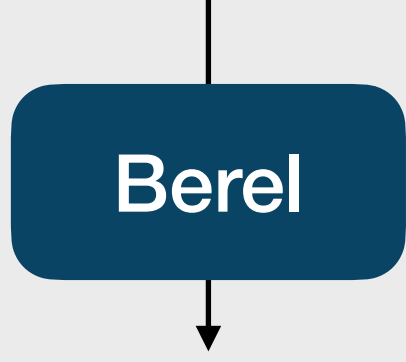
Results

- AlephBERT is outperformed by AlephBERT Adapted.
- **BEREL**, pre-trained on medieval texts, shows the best results.



Error Analysis

ילד לניחוח חשקתיו (A child for a scent - I have desired)



ילד לניחוח חשקתיו

Expert explanation: The child is a non-metaphorical epithet for Isaac. “Scent” is a metonymic term for “sacrifice”.