

# POS Tagging for Code-Switched, Transliterated Texts without Explicit Language Identification

Kelsey Ball

kelseyball@gmail.com

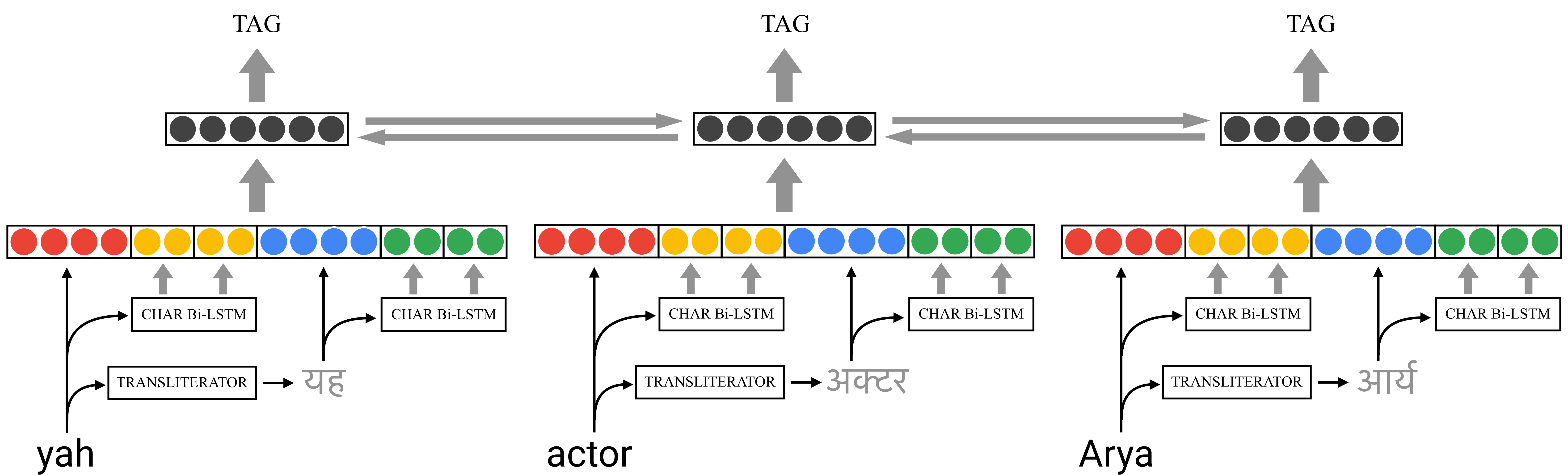
Dan Garrette

dhgarrette@google.com

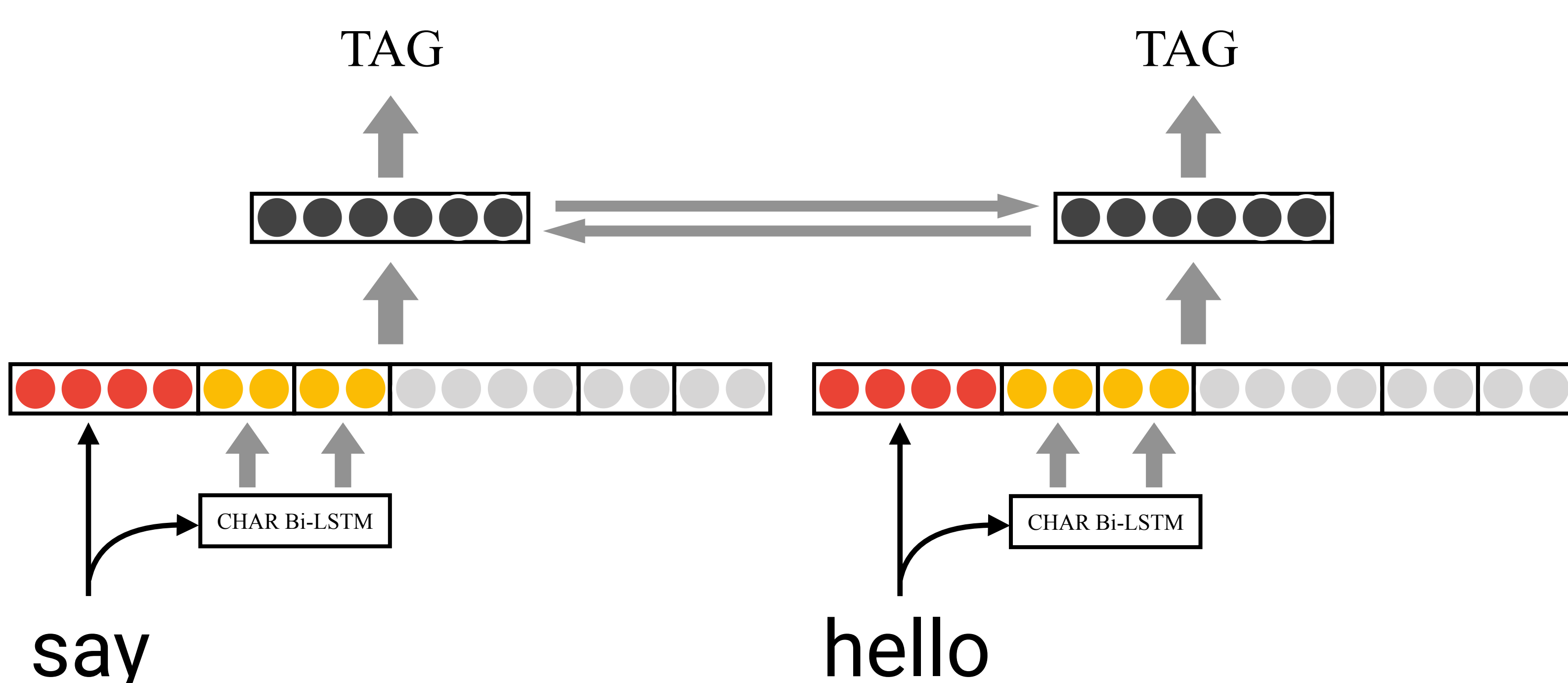
**Task:** Train a model that can be run on *code-mixed, transliterated* text.

But, the only available training data is *monolingual and standard-script*.

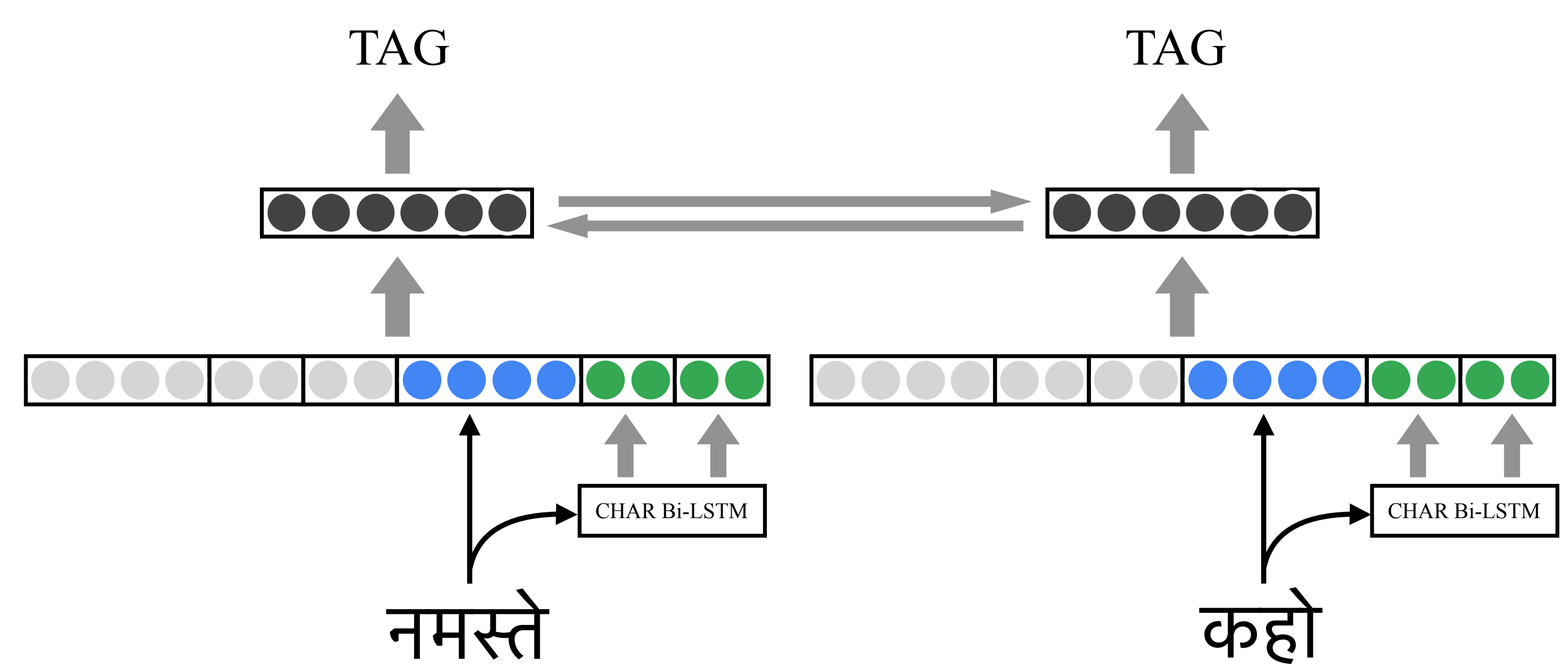
**Our Approach:** Treat every word at inference time as *both* languages simultaneously, by concatenating monolingual representations.



Training from English data



Training from Hindi data



Model	Acc.
Baseline (monolingual representation)	70.92
Ours (multilingual representation)	75.29
Ours, w/ forced language choice	76.04
Ours, w/ languages weighted by unsupervised HMM	<b>77.40</b>
Ours, w/ oracle language choice	80.53
Bhat et al. (2018) (w/ supervision)	90.53