

## Problem: Existing BLI datasets are machine generated

- Using machine translation as “ground truth” for another machine translation
  - ▶ not always bad, but never ideal
- This results in weird artifacts:

Thai “Word”	English “Translation”
แคลอรี	calories
โคมไฟ	lanterns
annie	annie
bdfutbol	bdfutbol
getparent	getparent
roca	roca

- Quality varies tremendously between languages  
(EN-ES pretty good... EN-TH pretty bad... others???)
- ? suggest that future research  
*“avoids drawing conclusions from quantitative results on [the MUSE]  
BLI dataset”*