



Controlling NMT Formality with Synthetic Supervision

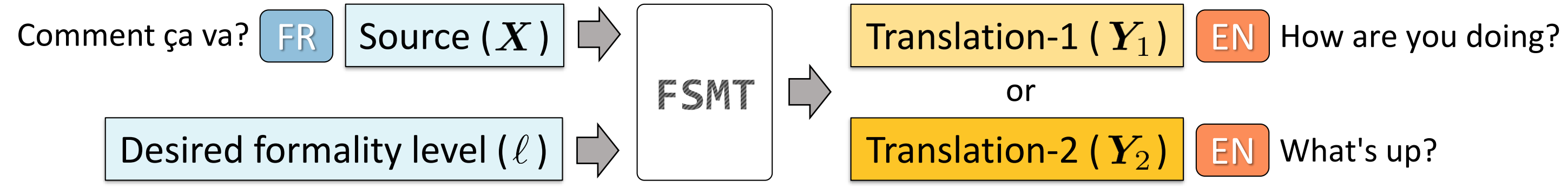


Amazon AI

Xing Niu^{1,2}, Marine Carpuat²
¹Amazon AI ²University of Maryland

PROBLEM

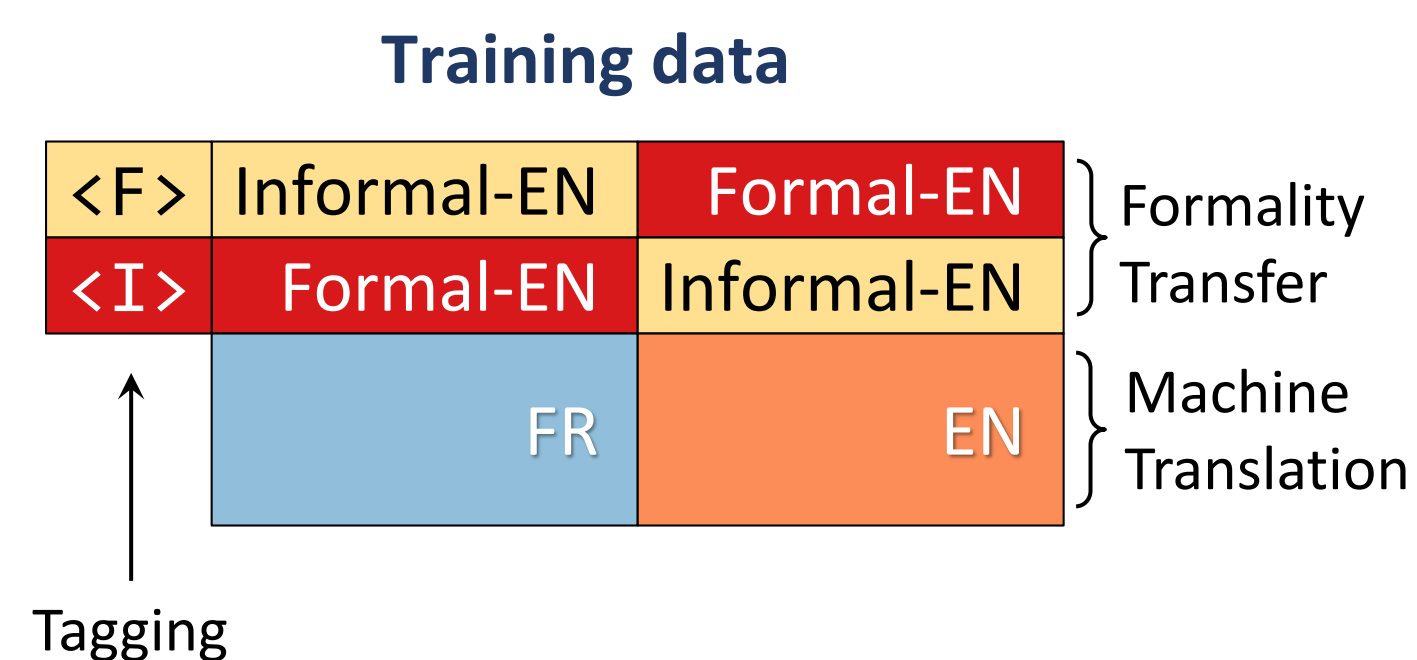
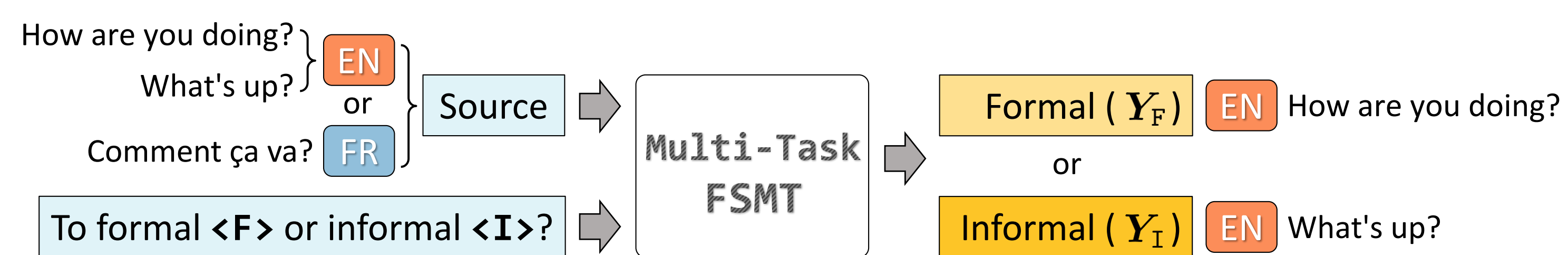
Formality-Sensitive Machine Translation (FSMT)



Ideal training triplets: not available

X	l ₁	Y _F
X	l ₂	Y _I

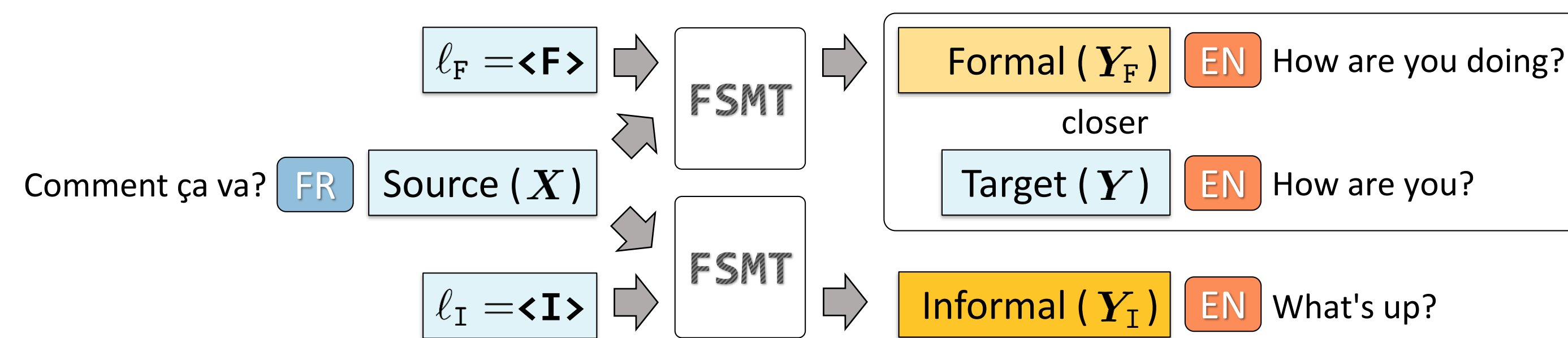
Prior Work – Multi-Task Learning



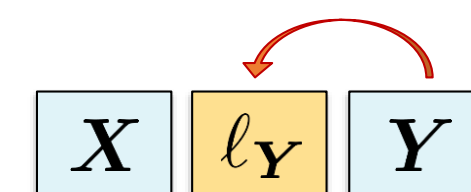
Zero-shot FSMT lags behind when supervision is provided.

APPROACH

Online Style Inference (OSI)



Complete the FSMT triplets on the fly



- Input: (X, Y)
- Pre-trained Multi-Task FSMT produces both informal and formal translations of X.
- Y is labeled as formal because it is closer to the formal translation than the informal translation.
- $(X, Y) \Rightarrow (X, l_Y, Y)$ if $CED(Y_I, Y_F) = H_Y(Y_I) - H_Y(Y_F) > \tau$

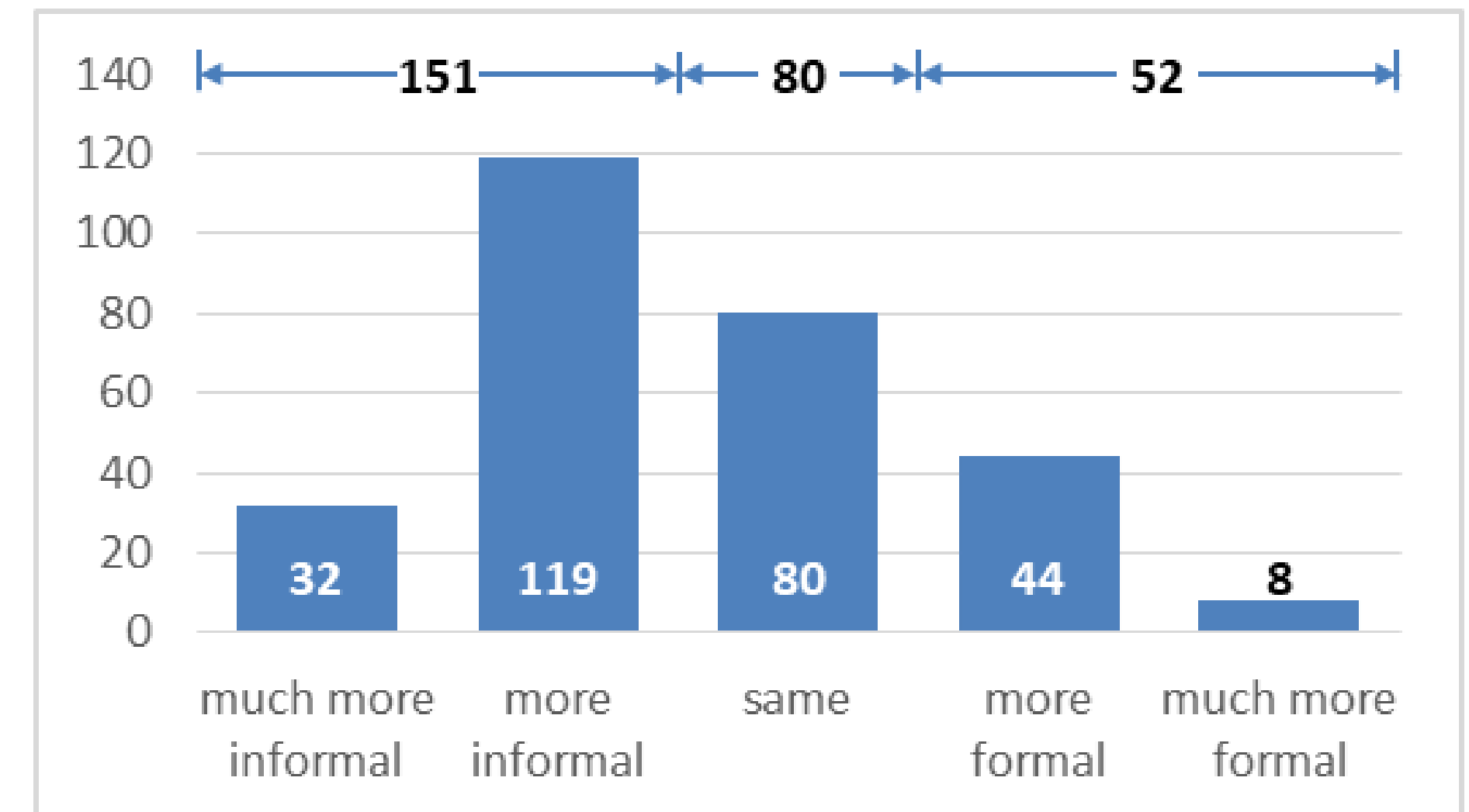
EXAMPLES

Type	Informal translation	Formal translation
Filler	And I think his wife has family there.	I think his wife has family there.
Completeness ▼		
Quotation	The gas tax is simply not sustainable, said Lee.	“The gas tax is simply not sustainable,” said Lee.
Yes-No	You like shopping?	Do you like shopping?
Subject	Sorry it’s my fault.	I’m sorry it’s my fault.
Article	Cookies where I work.	The cookies where I work.
Relativizer	Other stores you can’t buy.	The other stores where you can’t buy.
Paraphrasing ▼		
Contraction	I think he’d like that, but we’ll see.	I think he would like that, but we will see.
Possessive	Fay’s innovation perpetuated over the years.	The innovation of Fay has perpetuated over the years.
Adverb	I told you already .	I already told you .
Idiom	Hi, how’s it going?	Hi, how are you?
Slang	You gotta let him digest.	You have to let him digest.
Word-1	Actually my dad’s some kind of technician so he understands, but my mom’s very old.	In fact, my father is some kind of technician so he understands, but my mother is very old.
Word-2	Maybe a little more in some areas.	Perhaps a little more in certain areas.
Word-3	It’s really necessary for our nation.	This is essential for our nation.
Phrase-1	Yeah, me neither .	Yeah, neither do I .
Phrase-2	I think he’s moving to California now .	I think he is moving to California at the moment .
Phrase-3	It could be a Midwest thing .	This could be one thing from the Midwest .

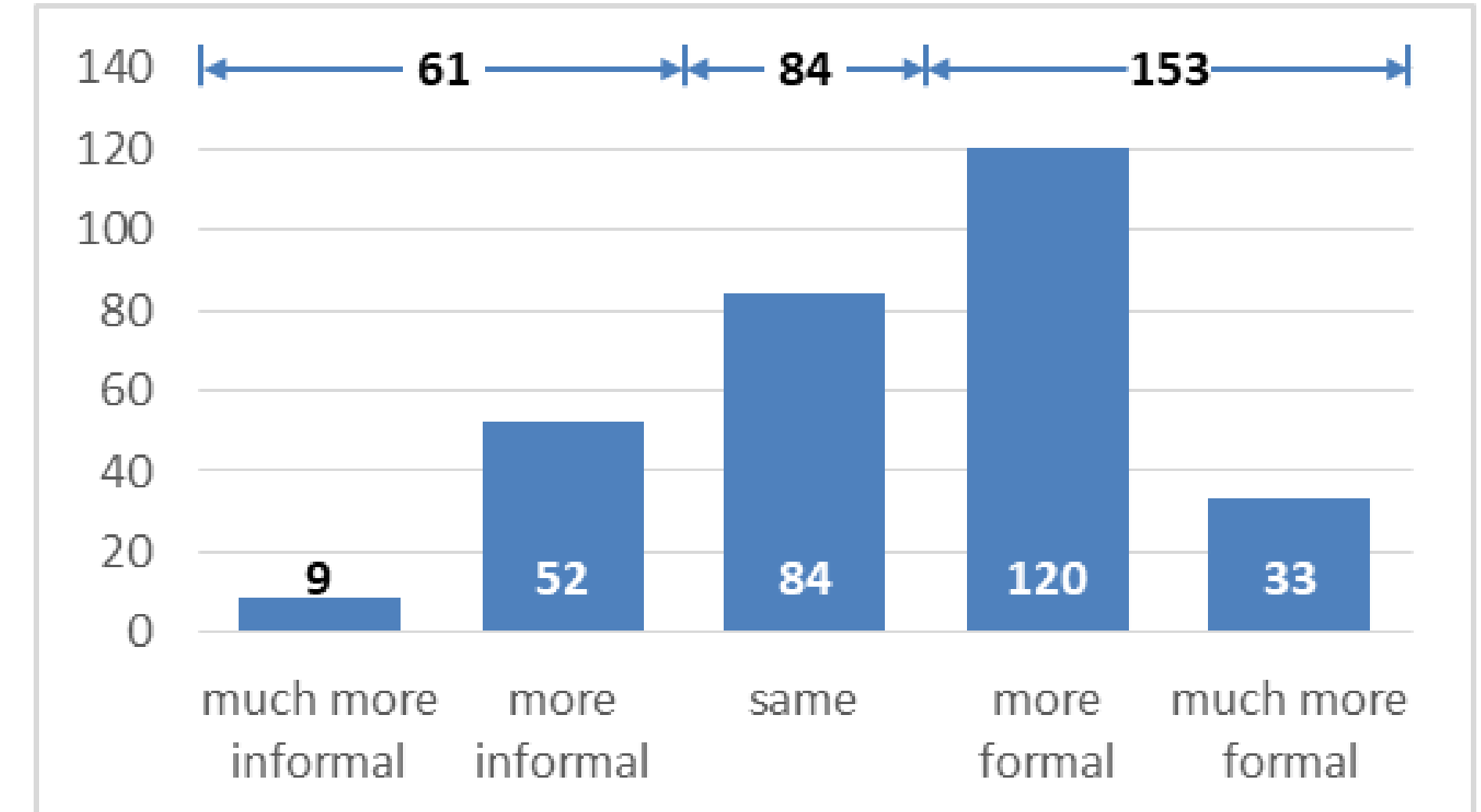
HUMAN EVALUATION

Online Style Inference ↔ Multi-Task

- (In)formal translations generated by OSI are annotated as more (in)formal than Multi-Task.

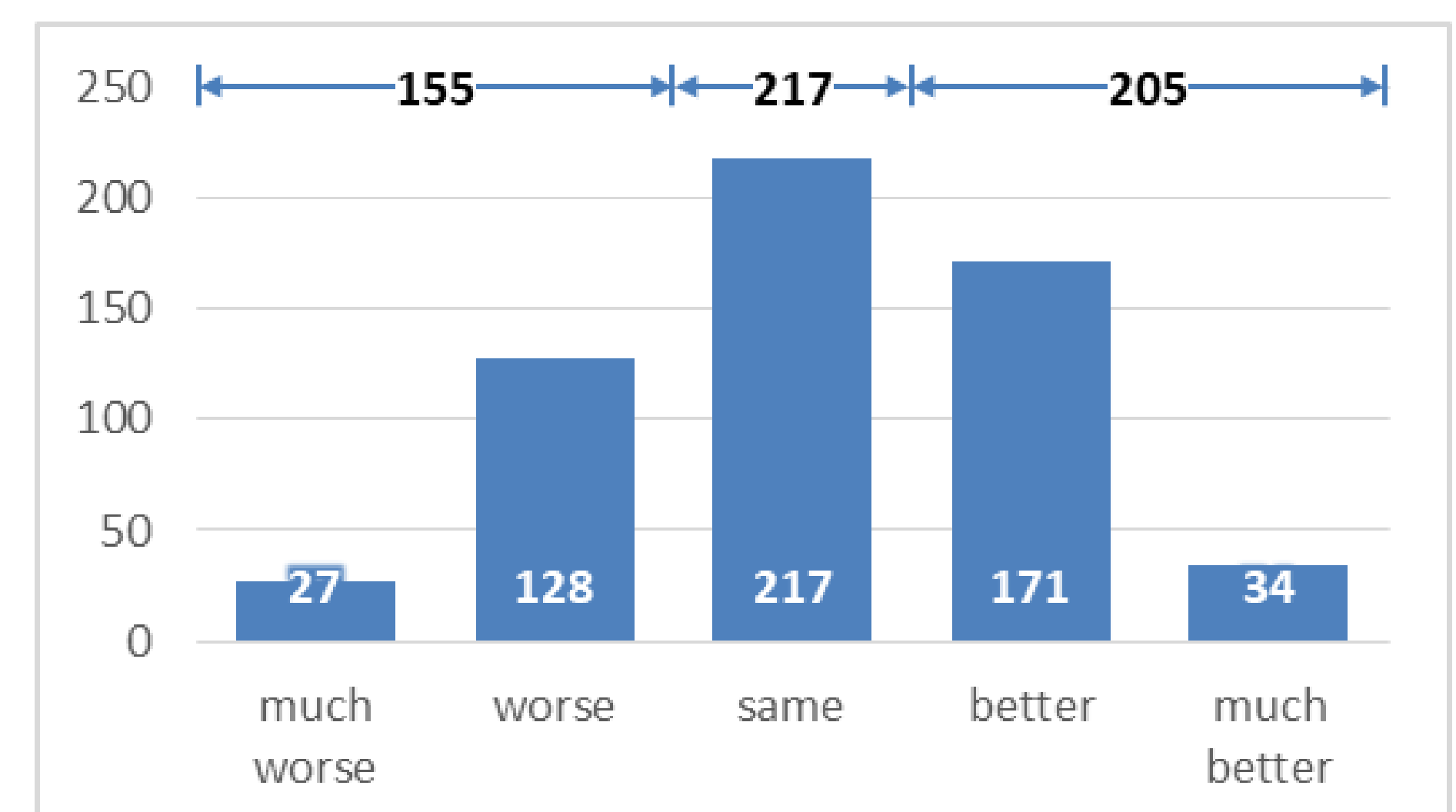


Informal translations



Formal translations

- The OSI model gets more instances that better preserve the meaning.



DATA

Formality Transfer

- Grammarly’s Yahoo Answers Formality Corpus (105K training)

Machine Translation (French→English)

- Europarl + News-Commentary + OpenSubtitles (18M training)

DOWNLOAD



Paper



Code & Data