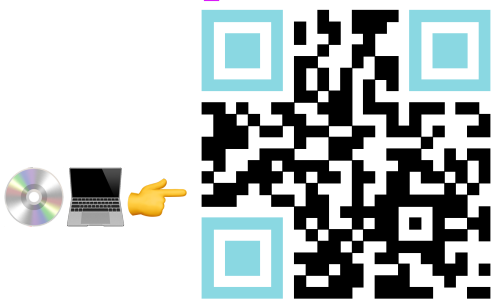# The ELCo Dataset:
Bridging Emoji and Lexical Composition
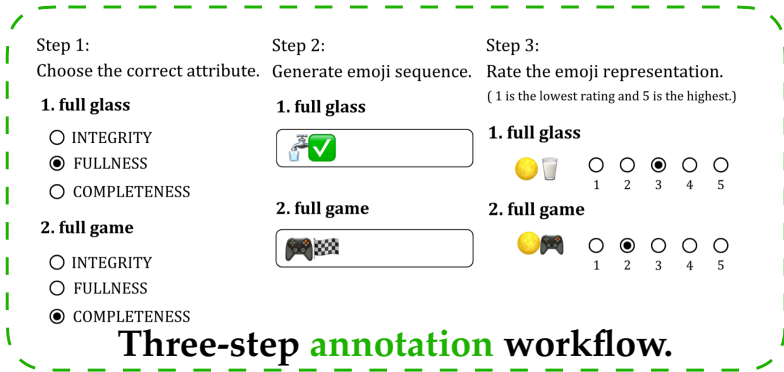
👩‍🎓👩‍🎓👨‍🎓 Zi Yun Yang, Ziqing Zhang, Yisong Miao

📥📥 yisong@comp.nus.edu.sg

👨‍🏫 Advisor: Prof Min-Yen Kan

Dataset: How to annotate emoji compositions?

Structure: How are emojis composed?

Meaning: Do LMs understand emoji compositions?

Three-step annotation workflow.

| English Phrase	Attribute	ELCo's Annotations (length = 2)	ELCo's Annotations (length > 2)	Average length of ELCo's Annotation	Emojinating's Output	Emojinating Rating
full attention	INTEGRITY	💯❗ and 🧐✍️	🧑‍💼💡🏛️	2.43	🪙🔔	3.1
full glass	FULLNESS	🥛満 and 🚰✅	🥛➕🚰🚰満	2.71	🪙🥤	3.0
full game	COMPLETENESS	🎮満 and 🎮💯	💯🆓 FREE	2.29	🪙🎮	2.7
full auditorium	FULLNESS	満🏛️	👩‍👩‍👧🏛️ and 👩‍👩‍👧満	4.14	🪙🤸	2.0
full life	FULLNESS	😊❤️	❤️👨‍👩‍👧💰💰💯 and 😆👨‍👩‍👧👧	4.00	🪙😌	1.1

Dataset: ELCo's annotations have variable lengths, avoid literal emoji choice, and offer metaphorical representations.

	Compositional Strategy	EN Phrase	Emoji Sequence
Ex. 1	Direct	right man	✔️👨
Ex. 2	Metaphorical	right man	💑👨
Ex. 3	Metaphorical	clear explanation	🧑‍🏫💯🗣️
Ex. 4	Metaphorical	fresh bread	🍞🕐🍞♨️
Ex. 5	Semantic list	bright future	😎🏙️👨
Ex. 6	Reduplication	big group	👨‍👨‍👦👨‍👨‍👦
Ex. 7	Single	right thing	✔️

Five prevalent structures in ELCo; Taxonomy adopted from Cohn et al 2019

Distribution of the compositional structures.

Metaphorical structure ∝ Diversity

Emoji-based Textual Entailment (EmoTE) Task

Input	Golden Label
Premise: This is 🧐✍️. Hypothesis: This is full attention.	Entailment
Premise: This is 🚰✅. Hypothesis: This is full attention.	Non-Entailment

EmoTE Task: Formalization.

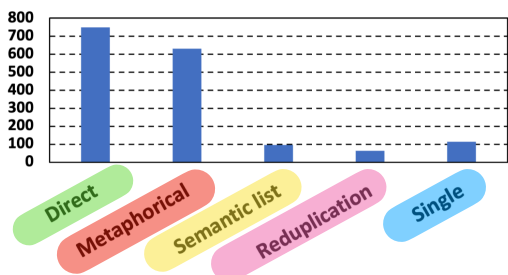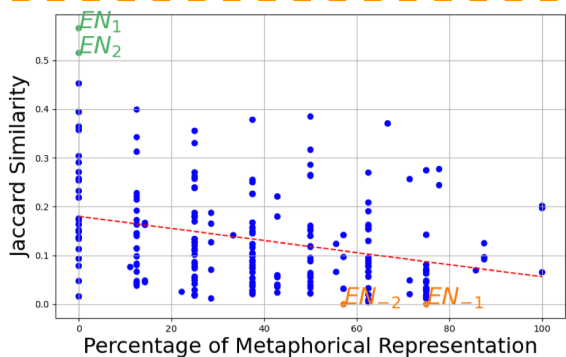|	w/o fine-tuning on ELCo					Fine-tuned on ELCo				
	$BERT_{base}$	$RoBERTa_{base}$	$RoBERTa_{large}$	$BART_{large}$	Avg	$BERT_{base}$	$RoBERTa_{base}$	$RoBERTa_{large}$	$BART_{large}$	Avg
Direct	34.7	35.6	41.5	51.7	40.9	$85.1_{2.3}$	$89_{3.3}$	$91.9_{2.2}$	$88.5_{2.8}$	88.6
Metaphorical	19.4	24.7	34.4	36.6	28.8	$68.4_{3.9}$	$73.1_{2.6}$	$80.4_{1.9}$	$82.8_{4.3}$	76.2
Semantic list	33.3	41.7	50.0	58.3	45.8	$86.7_{7.5}$	$78.3_{4.6}$	$85.0_{3.8}$	91.7_{0}	85.4
Reduplication	13.3	0	6.7	0	5.0	$65.4_{3.0}$	$52.0_{7.3}$	$62.7_{6.0}$	$88.0_{8.7}$	67.0
Single	19.0	19.0	19.0	52.4	27.4	$66.7_{3.4}$	$88.6_{2.6}$	$85.7_{4.8}$	$83.8_{2.6}$	81.2
Negative	83.4	83.0	90.3	82.2	84.7	$84.3_{4.1}$	$87.2_{1.9}$	$85.2_{0.7}$	$84.8_{1.2}$	85.4
Overall	55.0	55.8	62.9	62.9	59.2	$80.4_{1.5}$	$84.0_{0.8}$	$85.2_{0.9}$	$85.5_{0.9}$	83.8

RQ2: Fine-grained analysis reveals challenging classes.

RQ1: Overall performance on EmoTE.
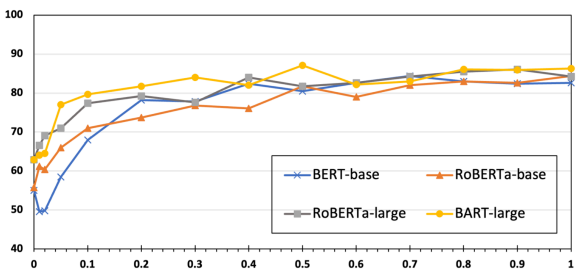
RQ3: ~50% of the ELCo training data is sufficient for model convergence.
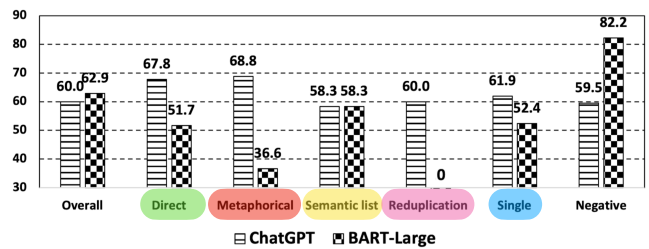
RQ4: ChatGPT is similar to BART-Large in overall performance, but it's more balanced.