

---

# Assessing the Helpfulness of Learning Materials with Inferred-Based Learner-Like Agent

Yun-Hsuan Jen<sup>1</sup>, Chieh-Yang Huang<sup>2</sup>, Mei-Hua Chen<sup>3</sup>,  
Ting-Hao (Kenneth) Huang<sup>2</sup>, and Lun-Wei Ku<sup>1</sup>

<sup>1</sup>Academia Sinica, Taipei, Taiwan. <sup>2</sup>Pennsylvania State University,  
University Park, PA, USA. <sup>3</sup>Tunghai University, Taichung, Taiwan

—

**How can we use modeling  
to find example sentences  
that clarify English  
near-synonym words?**

---

## Near-Synonym Word Pairs

- refuse vs. reject
- real vs. authentic

---

## Near-Synonym Word Pairs

- refuse vs. reject → grammatical usage difference
- real vs. authentic → sentiment or contextual difference

---




## Near-Synonym Word Pairs

- There is a **real** threat that the school could close.
- This letter is now deemed **authentic** by the experts.




# Automatically Example Sentence Extraction

# Automatically Example Sentence Extraction

- Linguistics search engine: Linggle and Netspeak

linggle<sup>10</sup><sup>12</sup>     [Sign in](#)

N-gram	Percent	Count	Example
refuse to accept	8.7 %	170,000	+
refuse to be	5.6 %	110,000	+
refuse to give	4.4 %	86,000	+

linggle<sup>10</sup><sup>12</sup>     [Sign in](#)

N-gram	Percent	Count	Example
reject to be	8.5 %	310	+
reject to accept	4.5 %	160	+
reject to reject	2.8 %	100	+

# Automatically Example Sentence Extraction

- Kilgarrriff et al's high quality example sentence extraction

simple <b>historically</b> 250,238 (13.61 per million)		Sort GDEX (Default configuration) ✕
1	ⓘ	Two <b>historically</b> original temporal motifs distinguish this story .
2	ⓘ	<b>Historically</b> lobster pots were constructed with wood .
3	ⓘ	Obesity has never been a problem <b>historically</b> until recently .
4	ⓘ	This is surely false , <b>historically</b> speaking .
5	ⓘ	Perhaps restraint has figured <b>historically</b> as a more familiar evil .
6	ⓘ	The position that these symbols are offensive is <b>historically</b> ridiculous .
7	ⓘ	And these companies are <b>historically</b> highly successful at changing behaviors .
8	ⓘ	This year saw division within groups that <b>historically</b> band together .
9	ⓘ	These people have <b>historically</b> made group decisions .
10	ⓘ	<b>Historically</b> recycling was a matter of necessity .

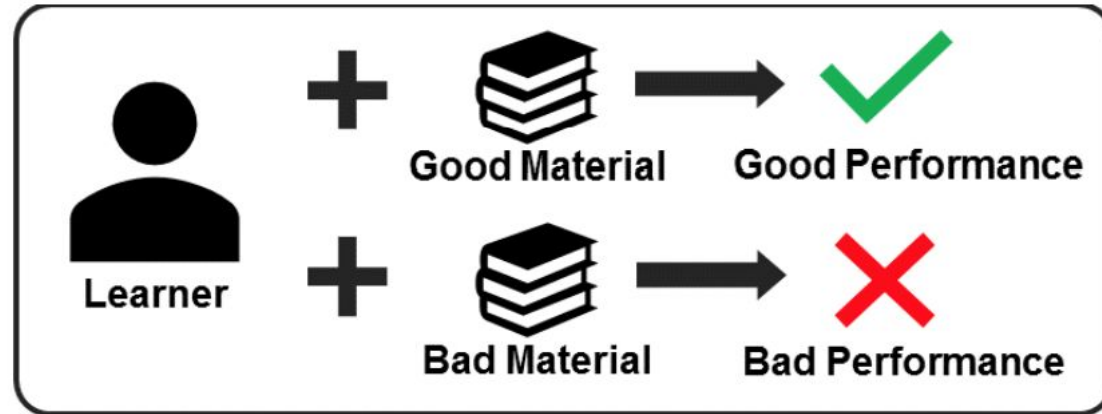


# —

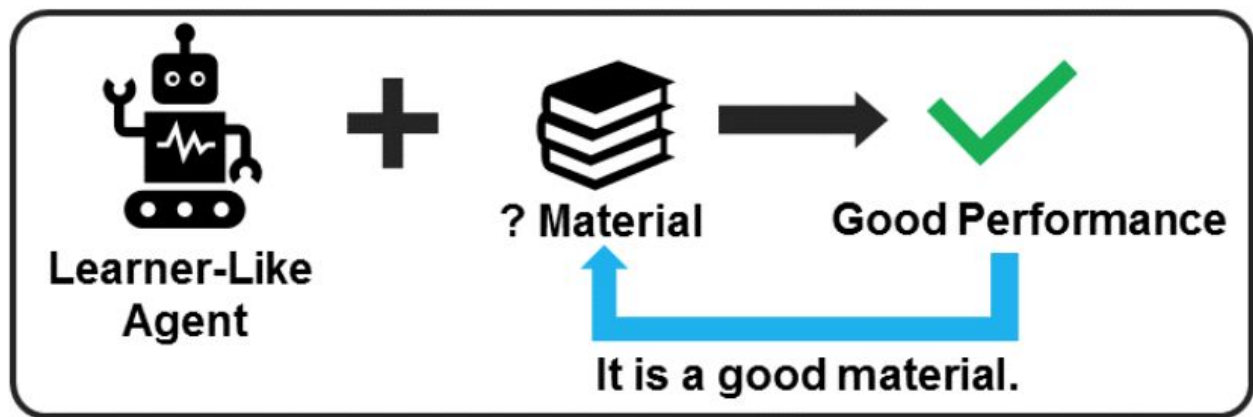
## Automatically Example Sentence Extraction

- Huang et al's handcrafted scoring for near-synonyms.
- refuse vs. reject → grammatical usage difference
- real vs. authentic → sentiment or contextual difference

# Learner's Behavior



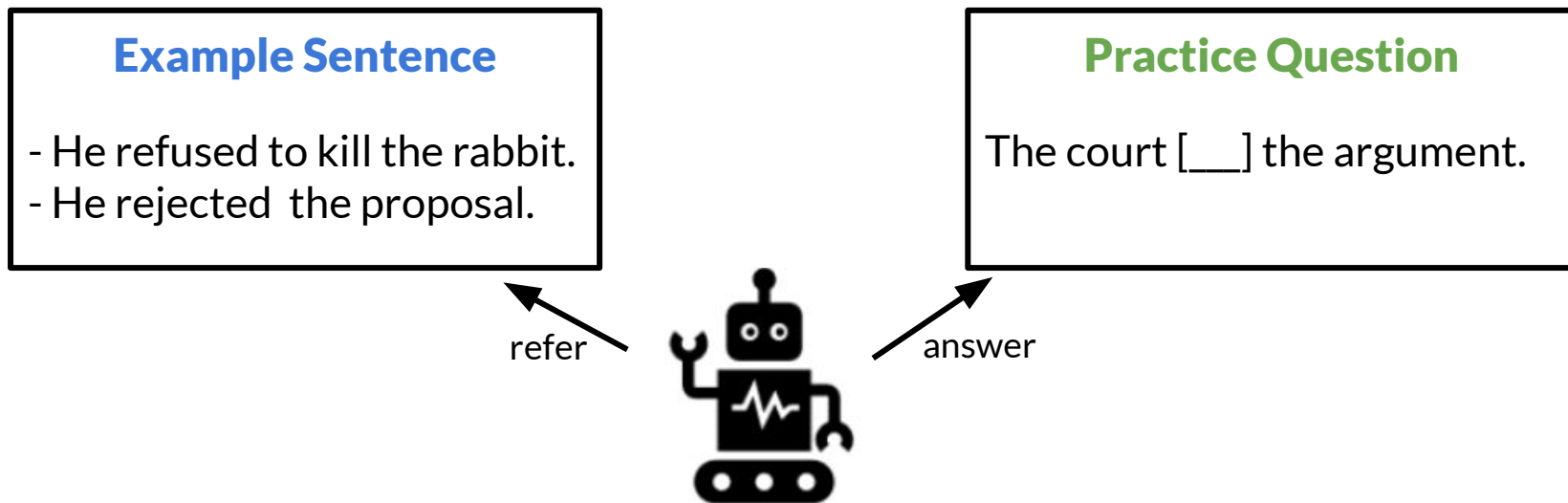
# Learner-Like Agent



# Learner-like Agent

# Learner-like agent

- A model that answers **questions** according to **examples**



# —

## Entailment **M**odeling **L**earner-Like **A**gent (EMLA)

- Similar to **natural language inference (NLI)**
- Premise: example sentences
- Hypothesis: practice question

# Entailment **M**odeling **L**earner-Like **A**gent (EMLA)

## Example Sentence

He rejected the proposal.

refer



## Practice Question

The court rejected the argument.

Ans: **entail**, not entail

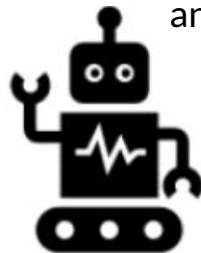
answer

## Perturbed Instances

### Example Sentence

He refused the proposal.  
—(rejected)—

refer



answer

### Practice Question

The court rejected the argument.

Ans: entail, **not entail**



# Models

## Context Modeling Learner-like Agent (CMLA)

### Example Sentence

He refused to kill the rabbit.  
He rejected the proposal.

refer



### Practice Question

The court [\_\_\_] the argument  
Ans: refused, **rejected**

answer

# —

## Data

- 30 near-synonym word-pairs
- Wikipedia
- 5000 sentences for each word; 4000 sentences for each word were used for training, 1000 were used for testing

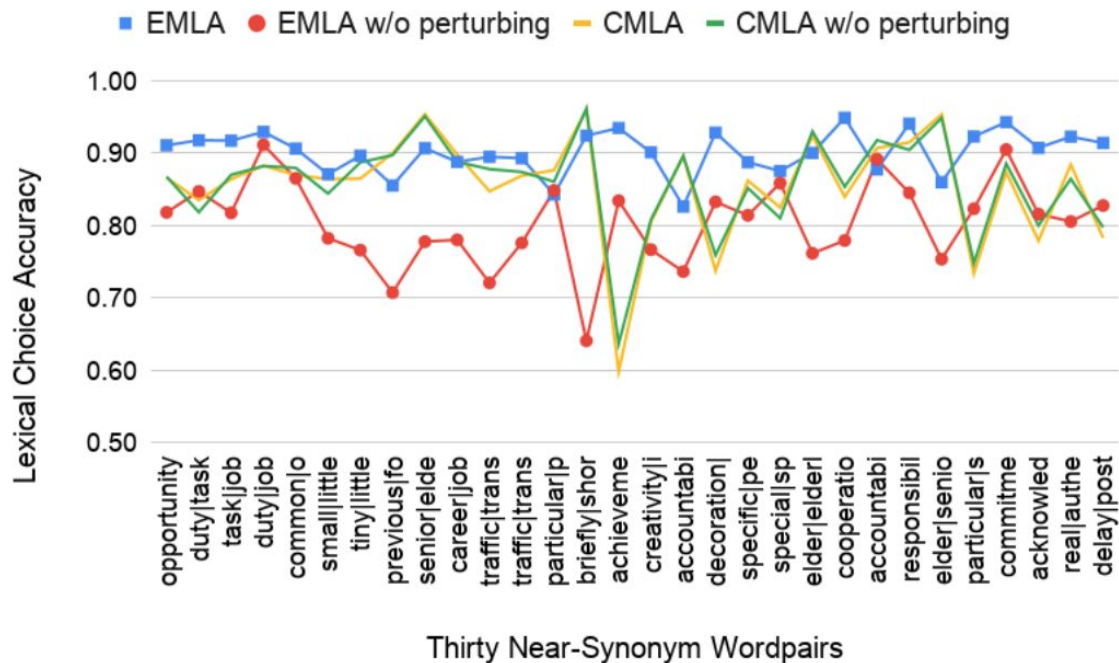
## Do the models know the difference between synonyms?

- Lexical Choice Task
- Fill-in-the-blank quiz

	<b>Acc</b>
accountability, responsibility	0.83
particular, peculiar	0.84
previous, former	0.86
elder, senior	0.86
small, little	0.87
special, specific	0.88
accountability, liability	0.88
specific, peculiar	0.89
career, job	0.89
traffic, transport	0.89

# Result

- Perturbed instances enhanced the inference ability of EMLA but didn't affect CMLA.



# Do these models behave like a learner?

- Behavior Check
- Two fill-in-the-blank quizzes
  - Authentic sentences as appropriate learning material (Models should answer this right!)
  - Automatically generated inappropriate example (Models should answer this wrong!)
  - Same practice question
- t-test

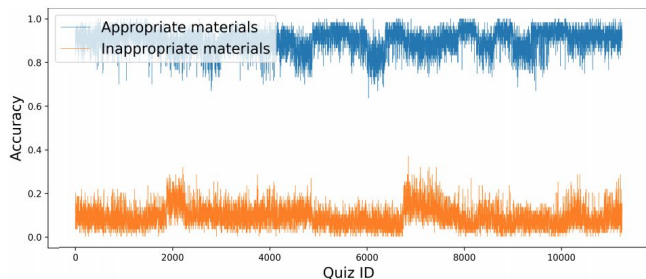
## Inappropriate Example Sentence

He ~~refused~~ **rejected** to kill the rabbit.

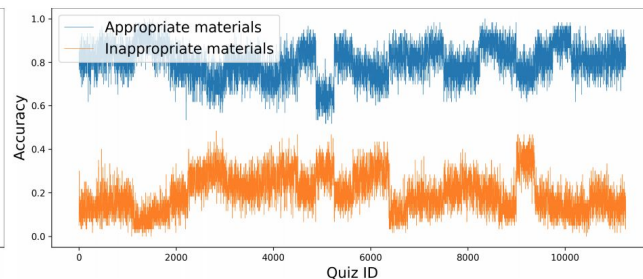
He ~~rejected~~ **refused** the proposal.

# Result

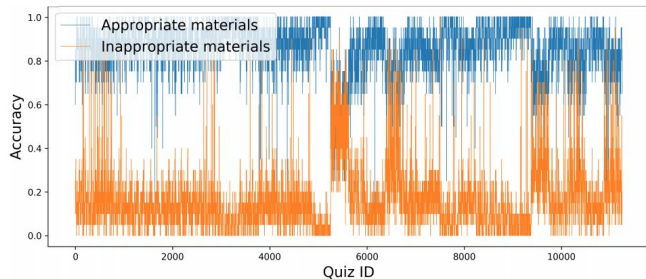
- CMLA relies on perturbed instance to learn the material quality difference while EMLA already can distinguished the difference.



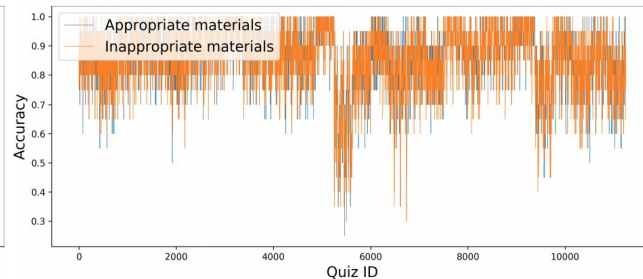
(a) EMLA



(b) EMLA w/o perturbing



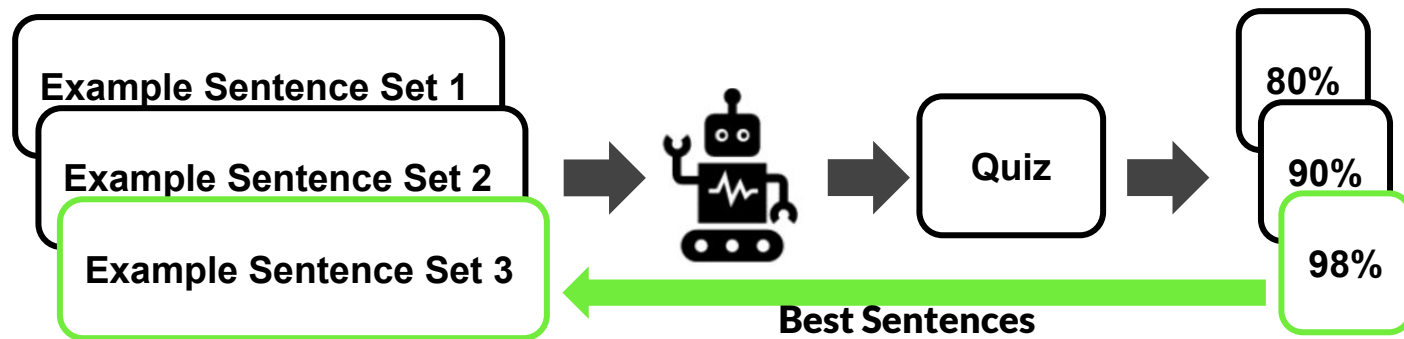
(c) CMLA



(d) CMLA w/o perturbing

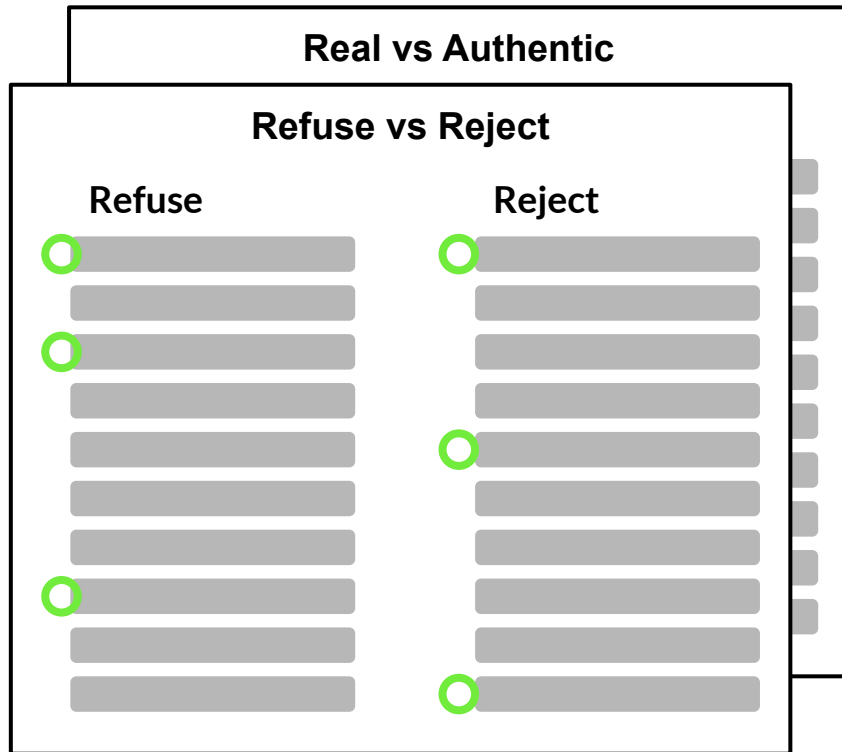
# Can the learner-like agents really select useful example sentences?

- Sentence Selection Task
- The ability of the learner-like agent to select useful example sentences
- We selected an example sentence combination that helps EMLA or CMLA to achieve the highest accuracy in a quiz.
  - A quiz contains 100 fill-in-the-blank questions



## Sentence Selection - Data

- A total of 600 examples sentences (30 word-pairs \* 20 sentences)
- Expert selects 3 best example sentences out of 10 randomly selected, grammatical, and pragmatically correct examples for each word in all near-synonym word pairs.
- Consider: suitability, informatives, diversity, sentence complexity, and lexical complexity





## Sentence Selection - Result

Compare EMLA and CMLA with Huang et al. (2017)'s Gaussian mixture model(GMM).

- When comparing with the human annotations, EMLA achieves the highest F1 of 0.45.

	<b>Precision</b>	<b>Recall</b>	<b>F1</b>
<b>EMLA</b>	0.33	0.71	<b>0.45</b>
<b>CMLA</b>	0.31	0.56	0.38
<b>GMM</b>	<b>0.37</b>	0.34	0.35

# Learner Study

- An English proficiency exam is given for further analysis
- Pre-test
  - Answer the questions (B)
- Post-test
  - The same questions (B) are using
  - The example panel (A) is shown

## (A) the example sentence panel

**A**

Please answer the following questions!

**Job**

1. Ten Cents a Dance is a popular song in which a taxi dancer laments the hardships of her **job**.
2. Many women take on the largest portion of the domestic obligations of the home , even when they are working full-time **jobs**.

readmore+

**Task**

1. Joan is asked by God to perform **tasks** that often appear to be trivial or contrary , but always end up positively improving a larger situation .

readmore+

**B**

1. Three years later, she left her \_\_\_\_\_ to become a full-time writer.  
(A) job  
(B) task
2. The \_\_\_\_\_ came with a salary of £600.  
(A) job  
(B) task
3. It defines three purposes and six \_\_\_\_\_.  
(A) jobs  
(B) tasks

## (B) the test panel.

## Learner Study Result

- Numbers of Students Improved: EMLA 16, CMLA 12, Random 11.
- Separate students into 2 groups using their English proficiency test scores.
  - EMLA helps above-average student to improve 0.75 points, 0.42 for CLMA, 0 for random.

		EMLA	CMLA	Random
Improvement	Above	0.75	0.42	0.00
	Below	0.18	-0.24	0.47

---

# Conclusion

- EMLA can differentiate the helpfulness of learning materials using **inference**.
- **Context Modeling** in the learner-like agent **relies on the perturbed instances** to mimic learner behavior, whereas EMLA already has this ability.
- Learner-like Agent can mimic **learner's behavior** and be used to evaluate the helpfulness of learning materials or to select the best material.