

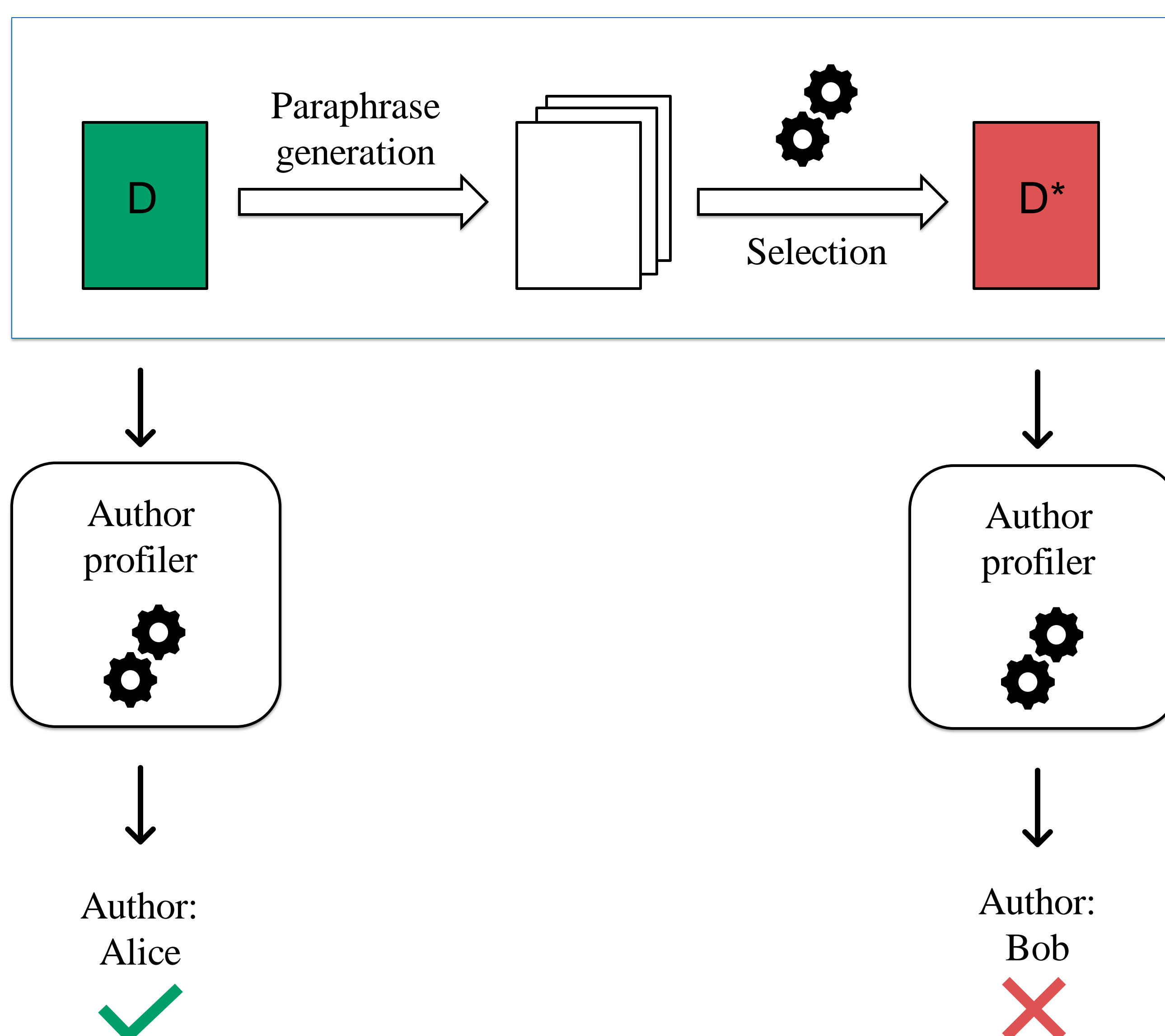
ParChoice: Effective writing style imitation using combinatorial paraphrasing

Author Imitation

- Authors can be **profiled** by **writing style**.
- Author **imitation** tricks a profiler into misclassifying the text.
- Challenge: **semantic retainment**.

ParChoice

- Generate many semantically equivalent candidates using **combinatorial paraphrasing**.
- Use **surrogate model** to **select paraphrase** that achieves best imitation.
- Paraphrasing techniques
 1. **Grammatical transformations (EAT2seq)**
A saw B → B was seen by A
 2. **Simple changes**
do not → don't
 3. **PPDB [2] replacements**
place of work → working place
 4. **WordNet [3] replacements**
car → automobile



Experiments

- **Gender imitation** on **Yelp restaurant reviews**.
- **Baseline techniques**: **CAE** [4], **BT** [5], **A4NT** [6].
- **Semantic retainment**: METEOR score [1], manual evaluation, user study.
- **Imitation success**: accuracy change on two deep learning author profilers: **CNN** [5] and **LSTM** [6].

Technique	Direction	METEOR	Paraphrases (manual)	User study (grade 0-5)	
				Mean	5
CAE	f→m	19.63	2%	0.8	2%
	m→f	19.62	2%		
BT	f→m	21.97	4%	0.9	3%
	m→f	19.80	2%		
A4NT	f→m	43.77	28%	1.7	9%
	m→f	46.10	26%		
ParChoice	f→m	44.80	56%	2.7	24%
	m→f	46.98	42%		

Semantic retainment scores (highest scores in green; lowest scores in red)
Paraphrases measured by manual evaluation on 50 sentences in both directions.
User study on 400 sentences annotated by 20 human evaluators.

Direction	Author profiler	Profiler accuracy				
		Original	CAE	BT	A4NT	ParChoice
f→m	CNN	0.75	0.44	0.42	0.57	0.32
	LSTM	0.83	0.55	0.54	0.64	0.46
m→f	CNN	0.77	0.59	0.53	0.73	0.59
	LSTM	0.82	0.48	0.45	0.70	0.48

Imitation success (best = lowest results in green; worst = highest results in red)

Discussion

- **CAE** and **BT** failed in **semantic retainment**.
- **A4NT**'s seeming success in semantic retainment was due to **sentence reproduction**.
- **ParChoice** uniquely produced **semantically acceptable transformations**.
- **ParChoice** outperformed **state-of-the-art** baseline (**A4NT**) in imitation.