Redundancy Detection in ESL Writings

Huichao Xue and Rebecca Hwa
Department of Computer Science
University of Pittsburgh

Redundancies in ESL essays

Vigorous writing is concise… This requires … that he make every word tell.
—Elementary Principles of Composition, *The Elements of Style (Strunk, 1918)*

- Writing concisely is challenging
  - Especially for Non-native speakers
- Redundancy – extra words/phrases:
  - Do not add to the meaning
  - Make the sentence more awkward to read

This study asks ~~the question of~~ whether …

- Redundancies are prevalent
  - In NUCLE (Dahlmeier and Ng, 2011), 13.71% of the marked problems are redundancy (2nd most frequent)

Examples of Redundancies in NUCLE

- There should be a careful consideration about what ~~are the things that~~ governments should pay for.
- The sodium-cooled technique was started to use since ~~the year~~ 1951.
- Non-renewable energy sources such as fossil fuels will ~~soon~~ be depleted within decades.
- ~~Nowadays,~~ as the population of the world is increasing rapidly , humans are facing severe food crisis .

Goal: Automatically detect redundancy

- Previous work did not directly address redundancy

Related work (XX)	XX but not redundant	Redundant but not XX
Grammar Error Correction (Leacock et al. 2010)	He ~~like~~ likes dogs.	… illustrate the ~~methodological~~ challenge …
Sentence compression – keep words that are specific to the sentence (Jing 2000; McDonald 2006; Clarke and Lapata 2007)	Kurtz completed ~~in high platform~~ diving.	These findings are often unpredictable ~~and uncertain~~.
Sentence simplification (Coster and Kauchak, 2011)	… positive critical reception … → … good reviews …	… not only just …

- To remove redundancy, we need an automatic measure for redundant phrases

Contributions

- We conducted the first study on automatic redundancy detection
- We propose a measure of redundancy
  - A probability value
  - The calculation boils down to looking at the input sentence's alignment with its translation
    - If one word is aligned to nothing → redundant
    - If two words are aligned to the same word → redundant
    - If deleting one word/phrase hurts fluency → non-redundant
- The proposed measure out-performs several baselines by a large margin

Redundancy – words that do not tell

- We consider a word/phrase redundant if …
  - Deleting it results in a fluent English sentence that conveys the same meaning as before
- Our definition suggests two factors for redundancy:
  - Contribution to fluency — We can capture with language models
  - Contribution to meaning — How do we capture this?

Adds to fluency

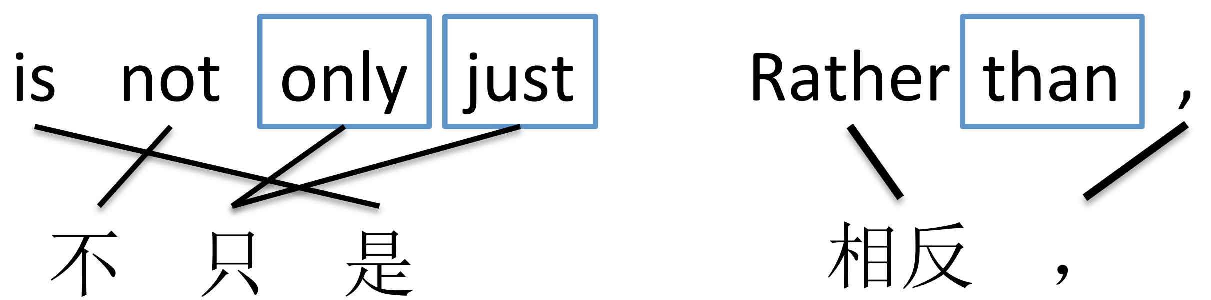
redundant

Conveys meaning

This is just an example of the received achievements .

Approximating Meaning with Translation

- Sentence's meaning can be represented by its translation in another language. (Hermet et al. 2009, Madnani et al. 2012)
- A word's alignment suggests how much meaning it conveys



Carrying same meaning as other words'

Not semantically meaningful

Modeling Redundancy with Translation

A phrase $e_s \ldots e_t$ in e is deemed redundant if we translate sentence e into foreign language f and then back into English, we are likely to obtain the rest of the sentence $e_{-}^{s,t}$

$$R(s,t;e) = \log \sum_{F=f} \left(\Pr(f|e)\Pr(e_{-}^{s,t}|f)\right)$$

$$\approx \log\left(\Pr(f^*|e)\Pr(e_{-}^{s,t}|f^*)\right)$$

$$= \boxed{\log \Pr(f^*|e)} + \boxed{\log \Pr(e_{-}^{s,t}|f^*)}$$

common over all sub-phrases

We want to calculate this number

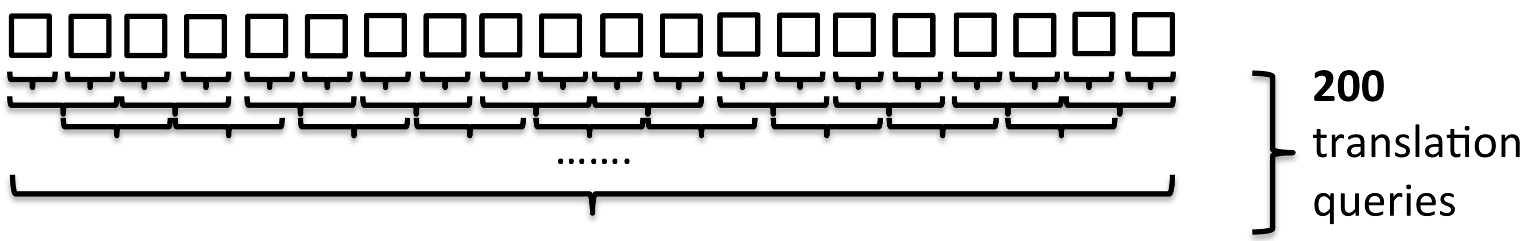
- We consider the one best translation f^* of e
- E.g. e = "I really like it", f^* = "我真的喜欢它"
  R(really) = $\boxed{\text{log Pr("I like it" | "我真的喜欢它")}}$ + $\boxed{\text{C}(e)}$

We want to calculate this number

common over all sub-phrases

We don't directly query SMT systems for $\Pr(e_{-}^{s,t} | f^{*})$

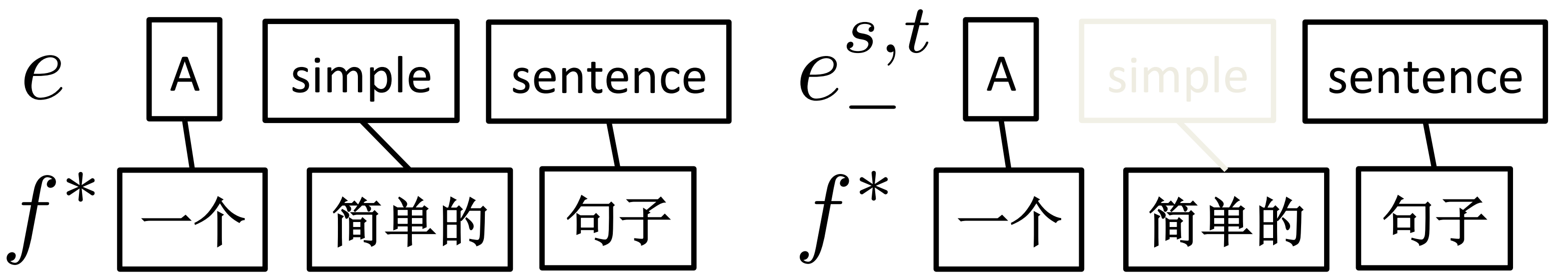
- It is expensive: for every sub-phrase, we need one translation query
  - Consider enumerating all sub-phrases in a 20-word English sentence



- It is inconvenient: many translation systems do not have APIs for it
- We propose an approximation
  - Less expensive: 1 translation query per sentence
  - Convenient: uses normal MT system output, translation and alignment

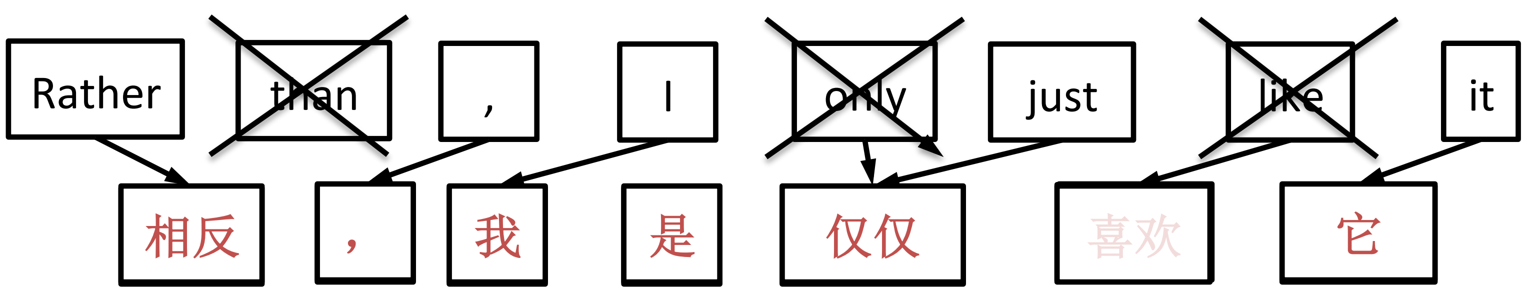
Approximating $\Pr(e_{-}^{s,t} | f^*)$

- SMT systems roughly compute it in two steps
  1. Align the two sentences
  2. Calculate the probability given the alignment
- Our approximations
  1. We reuse the alignments between e and f^*



Approximating $\Pr(e_{-}^{s,t}|f^*)$

- SMT systems roughly compute it in two steps
  1. Align the two sentences
  2. Calculate the probability given the alignment
- Our approximations
  1. We reuse the alignments between e and f^*
  2. IBM model 1 (Brown et al, 1993)– each word contributes to its aligned slot
     - Deleting a word risks losing the its aligned word

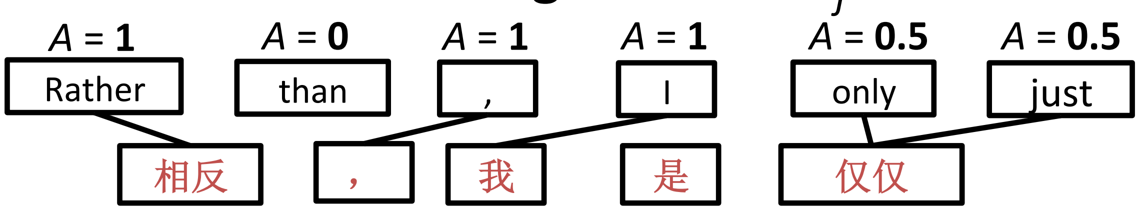


Proposed Redundancy Measure

Fluency w/o $e_s \ldots e_t$ | Per word meaning redundancy | const

$$R(s,t;e) \approx \boxed{\mathrm{LM}(e_{-}^{s,t})} + \boxed{\sum_{s \le j \le t} A(j) \log \Pr(e_j)} + \boxed{C(e)}$$

- $\mathrm{LM}(e_{-}^{s,t})$: log likelihood of sentence without $e_s \ldots e_t$
  - A phrase is redundant, if deleting it does not hurt fluency
- Meaning Redundancy
  - $A(j)$: number of words aligned with e_j



  - $\Pr(e_j)$: unigram probability of e_j
    - Rare words are often less redundant

Experimental Setup

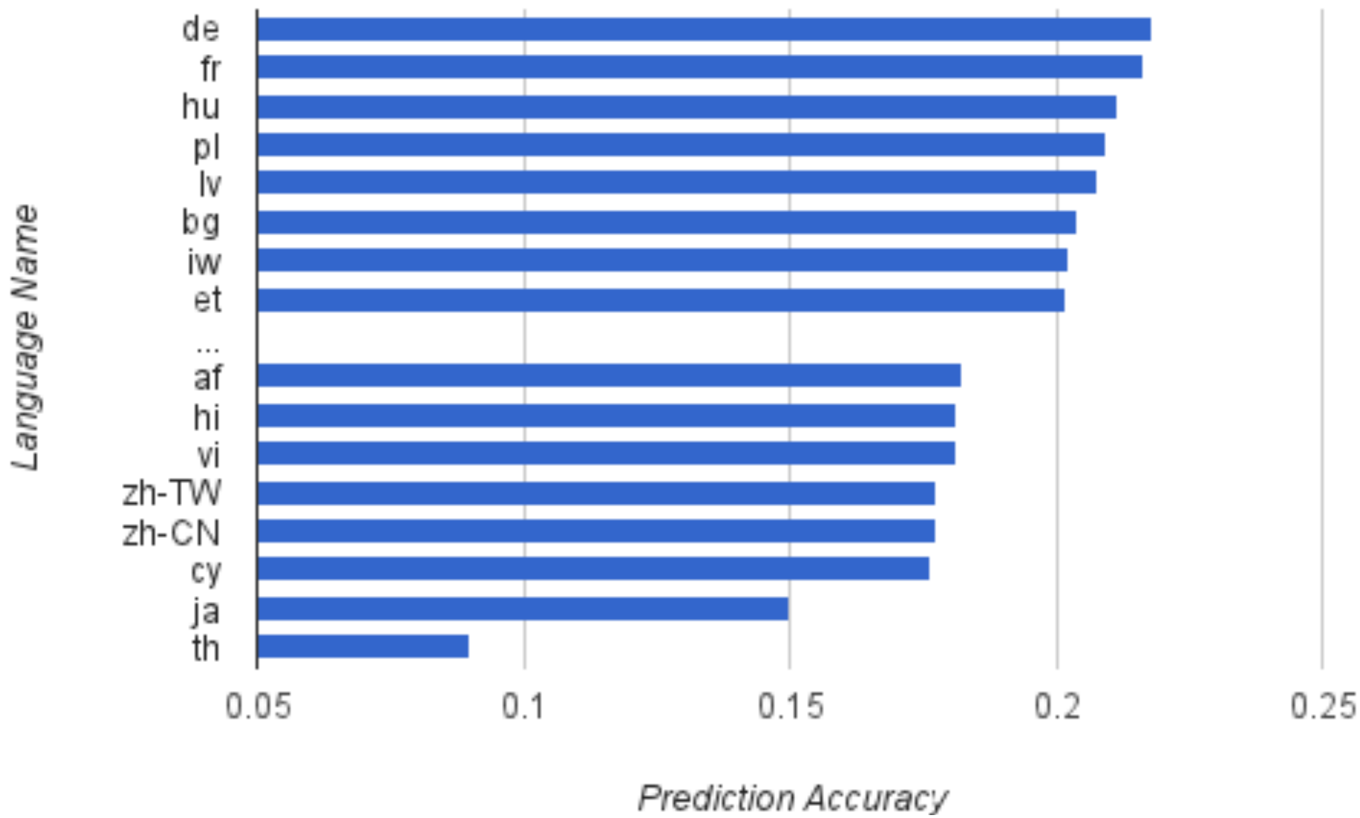
- Evaluation Data: NUCLE (Dahlmeier and Ng, 2011)
  - Redundancies are explicitly marked
  - Evaluation set:
    - 527 sentences (from 200 essays)
    - Each sentence has exactly one redundant phrase
- Task:
  - Pick the most redundant phrase *for a given length*
    - pick one from ≈20
  - Evaluation Metric: accuracy
- Tools:
  - Fluency: *trigram* language model (trained on English Gigaword)
  - Google translate (French as pivot)

Different Redundancy Measures

Metric	Explanation	accuracy
random	the random baseline	4.44%
R(s, t; e)	proposed method	**21.63%**
LM($e_{-}^{s,t}$)	Fluency, by *trigram* language model	8.06%
meaning-red	Per-word meaning redundancy	8.59%
sig-score	sentence compression (Clarke et al. 2007)	10.71%
round-trip	number of words disappeared after a round-trip translation	10.69%
trigram + α round-trip		14.80%
trigram + α sig-score		11.01%

Using translation as an approximation for sentence meaning is plausible

Using different pivot languages for redundancy measurement



European languages generally work better.

Influence from meaning components

- Google translate organizes output into characters for Asian languages
  - Characters are not the minimum meaning component

explain

解 to solve

释 to release

- We merged characters/alignments using tokenization result for zh-CN

language	accuracy
De	21.82%
Zh-CN	17.74%
Zh-CN (char-merged)	20.11%

Getting close

improved

What types of redundancies do LM($e_{\_}^{s,t}$)/meaning-red capture?

- We measure **recalls**: percentage of redundant function/content words correctly detected
  - Function words: determiners and prepositions
  - Content words: others

measure	accuracy	Recall (function)	Recall (content)
LM($e_{\_}^{s,t}$)	8.06%	3.95%	**9.73%**
meaning-red	8.59%	**20.23%**	3.87%
R(s, t; e)	**21.63%**	38.16%	14.93%

- Fluency (**trigram**) helps detect redundant content words
- Meaning redundancy (**meaning-red**) helps detect redundant function words
- The accuracies of these two components add up

Conclusions

- We conducted the first study in redundancy detection
- We proposed to account for redundancies by comparing one sentence with its translation
  - The measure accounts for one phrase's contribution to meaning and fluency
- The proposed measure shows promise for redundancy detection
  - Outperforms other metrics by a large margin
  - Five-times more accurate than random baseline

Thank you!

- This work is supported by U.S. National Science Foundation Grant IIS-0745914
- We thank the anonymous reviewers, Joel Tetreault, Janyce Wiebe, Wencan Luo, Fan Zhang, Lingjia Deng, Jiahe Qian, Nitin Madnani and Yafei Wei for helpful discussions