Five Shades of Noise: Analyzing Machine Translation Errors in User-Generated Text

Marlies van der Wees, Arianna Bisazza, Christof Monz
Statistical Machine Translation

News sentence: 印度金融中心孟买亦受到波及。
(mumbai, india's financial center, was also affected.)

india's financial center mumbai also affected.
Statistical Machine Translation

SMS sentence: 你路上慢点
(be careful on your way / take your time)

you are on the road to slow points
# SMT for user-generated text is often bad

| Reference | SMT output |
|-----------|------------|
| and if i go out, i will stop by your place | and if i went. |
| i could not bring it to you | into its enemies. |
| i've never seen a pig there | i am seen pig there. |
| you're too delighted to be homesick | anytime you |
Towards improving SMT quality for UG

- To target specific error types, we need to know why mistakes are made:
  - in UG versus formal text
    - contrast UG with newswire
  - in different types of UG
    - five shades of noise: weblogs, comments, speech (CTS), SMS, and chat messages
  - in different language pairs
    - Arabic-English & Chinese-English
Analyzing SMT errors in UG text

- What translation choices were made by the SMT system?

- What translation choices could have been made by the SMT system?

- Why did the SMT system make the choices that it made?
Word Alignment Driven Evaluation: approach*

* For each word alignment link in the test (e.g. 你 — your) that is translated wrongly, determine:

| source phrase | target phrase       | probability |
|---------------|---------------------|-------------|
| 路上          | on the road         | 0.4         |
| 路上          | on the way          | 0.3         |
| 路上          | on your way         | 0.2         |
| 点            | dot                 |             |
| 点            | point               |             |

source phrase not in phrase table: SENSE error

source and target phrases both in table, but other translation preferred: SCORE error

target phrase not in phrase table: SEEN error

* Approach adopted from Irvine et al., *Measuring Machine Translation Errors in New Domains*, 2013
Word Alignment Driven Evaluation: results

Word-level error statistics for Arabic-English benchmarks

- Correct
- Seen
- Sense
- Score
Word Alignment Driven Evaluation: findings

- SMT errors for UG text differ
  - from SMT errors for news
    - many SEEN and SENSE errors for UG
  - between different types of UG
    - SMS and chat messages are most affected
  - between different language pairs
    - differences in Chinese-English are more subtle than in Arabic-English
Analyzing SMT errors in UG: what we learned

- Common errors in UG are due to:
  - misspellings or Arabic dialectal forms
  - formal lexical choices
  - idioms translated word by word
  - dropped pronouns in Chinese
- UG suffers from low model coverage
  - generate new translation candidates
  - normalize existing translation candidates
More Error Analysis?

- Visit the **poster** for:
  - Model coverage analysis
  - Arabic-English versus Chinese-English results
  - Qualitative Examples

- Read the **paper** for:
  - Phrase-length analysis
  - Detailed explanation and discussions
Thank you!

- Marlies van der Wees
- m.e.vanderwees@uva.nl