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
  - and if i go out, i will stop by your place
  - i could not bring it to you
  - i've never seen a pig there
  - you're too delighted to be homesick

- SMT output
  - and if i went.
  - into its enemies.
  - i am seen pig there.
  - 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:

<table>
<thead>
<tr>
<th>source phrase</th>
<th>target phrase</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>路上</td>
<td>on the road</td>
<td>0.4</td>
</tr>
<tr>
<td>路上</td>
<td>on the way</td>
<td>0.3</td>
</tr>
<tr>
<td>路上</td>
<td>on your way</td>
<td>0.2</td>
</tr>
<tr>
<td>点</td>
<td>dot</td>
<td></td>
</tr>
<tr>
<td>点</td>
<td>point</td>
<td></td>
</tr>
</tbody>
</table>

source phrase not in phrase table: **SEEN error**

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

target phrase not in phrase table: **SENSE error**

Word Alignment Driven Evaluation: results

Word-level error statistics for Arabic-English benchmarks

- Correct
- Seen
- Sense
- Score

News 1
News 2
Weblogs
Comments
CTS
Chat
SMS

Relative frequency

News
UG
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