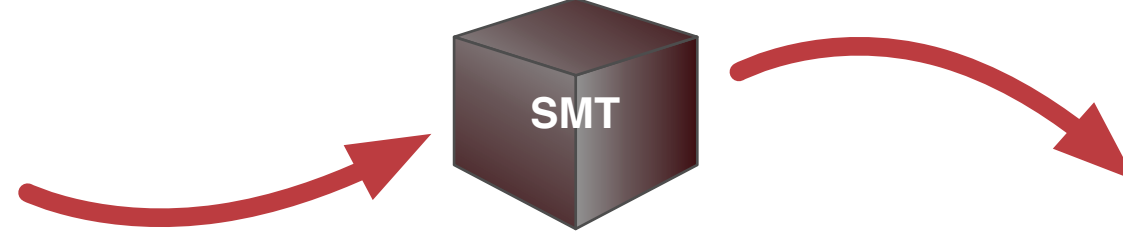


Five Shades of Noise: Analyzing Machine Translation Errors in User-Generated Text

Marlies van der Wees Arianna Bisazza Christof Monz
Informatics Institute, University of Amsterdam

Motivation

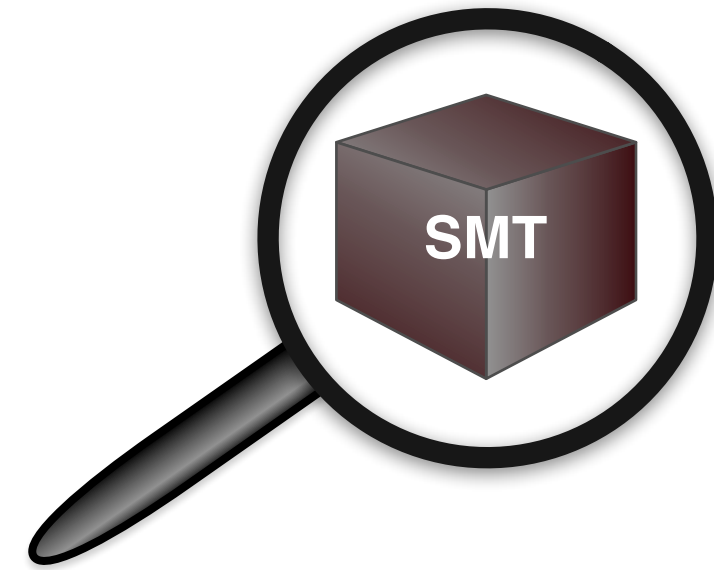
Statistical machine translation (SMT) of user-generated (UG) text
input SMS message: **你路上慢点**
(= be careful on your way / take your time)



output translation: **you are on the road to slow points**

Understanding SMT errors in UG text
why does SMT make the errors that it makes on UG?

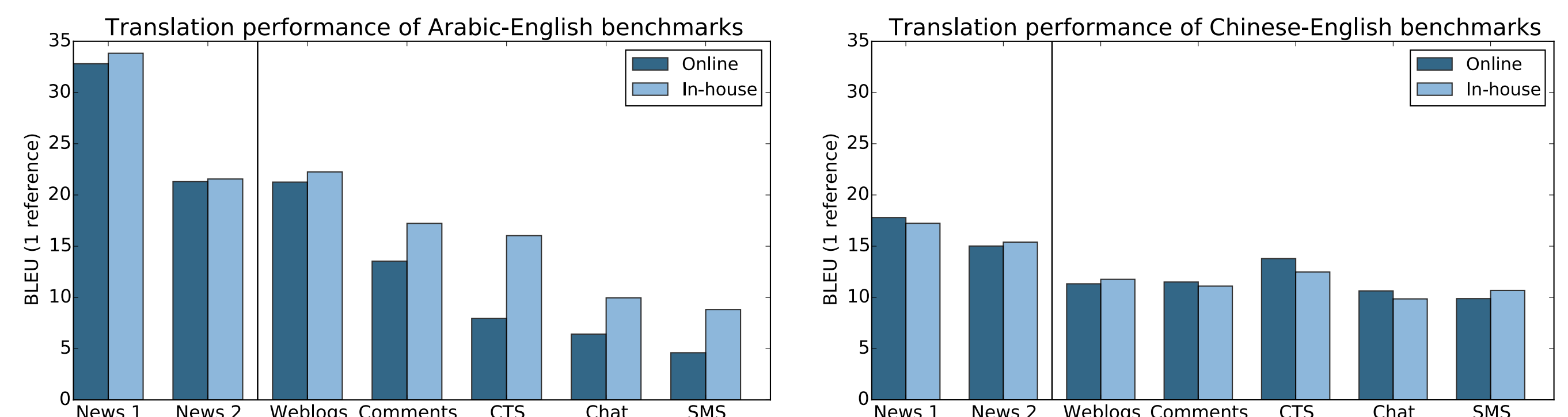
- low model coverage?
 - poor scoring of translation options?
- what errors are observed for various types of UG?



Five Shades of Noise

Two language pairs Arabic-English & Chinese-English
Five UG sets weblogs, comments, speech, SMS, chat
Two news sets different sources, to contrast with UG

Lower translation quality for UG than for news



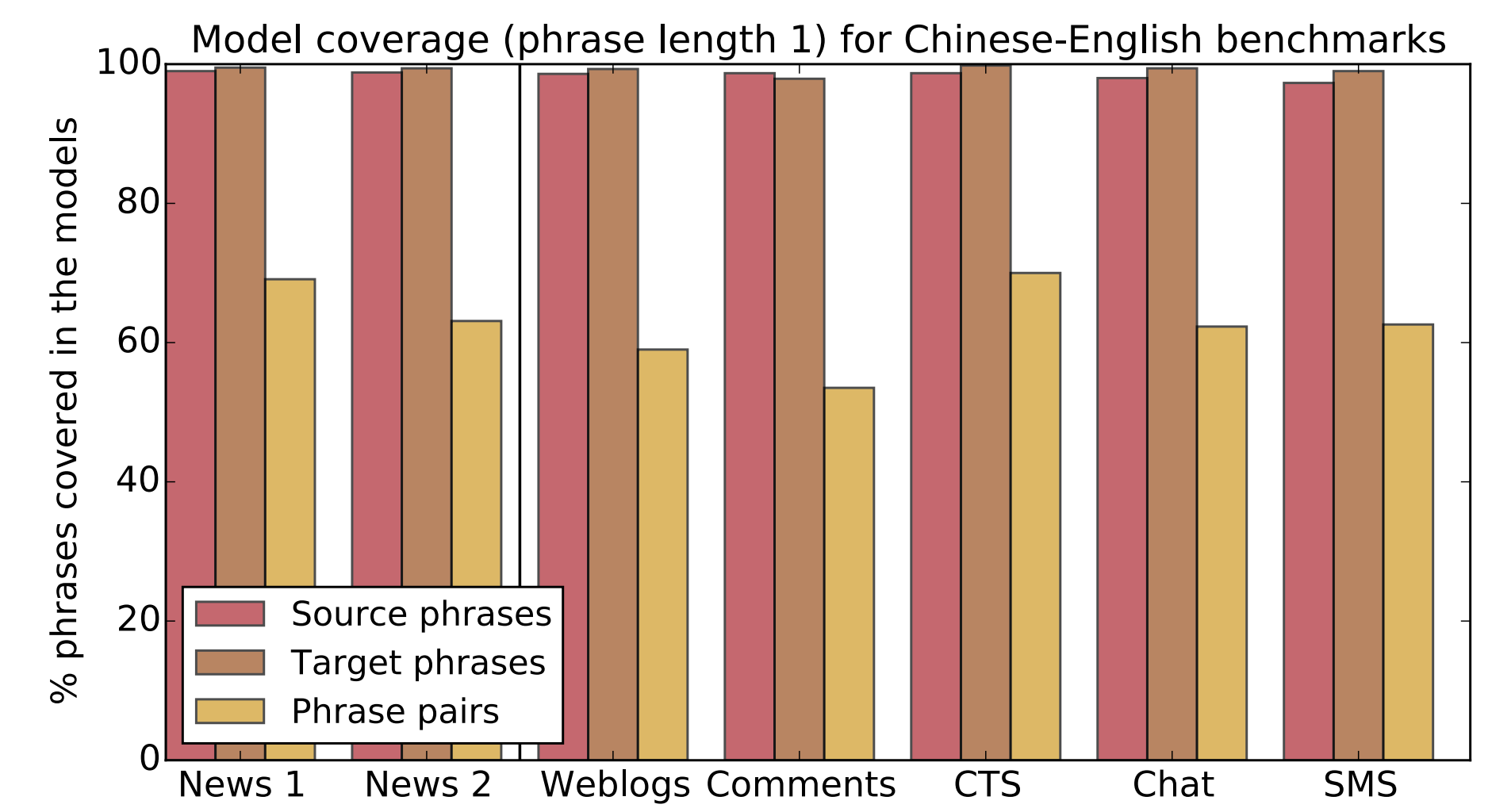
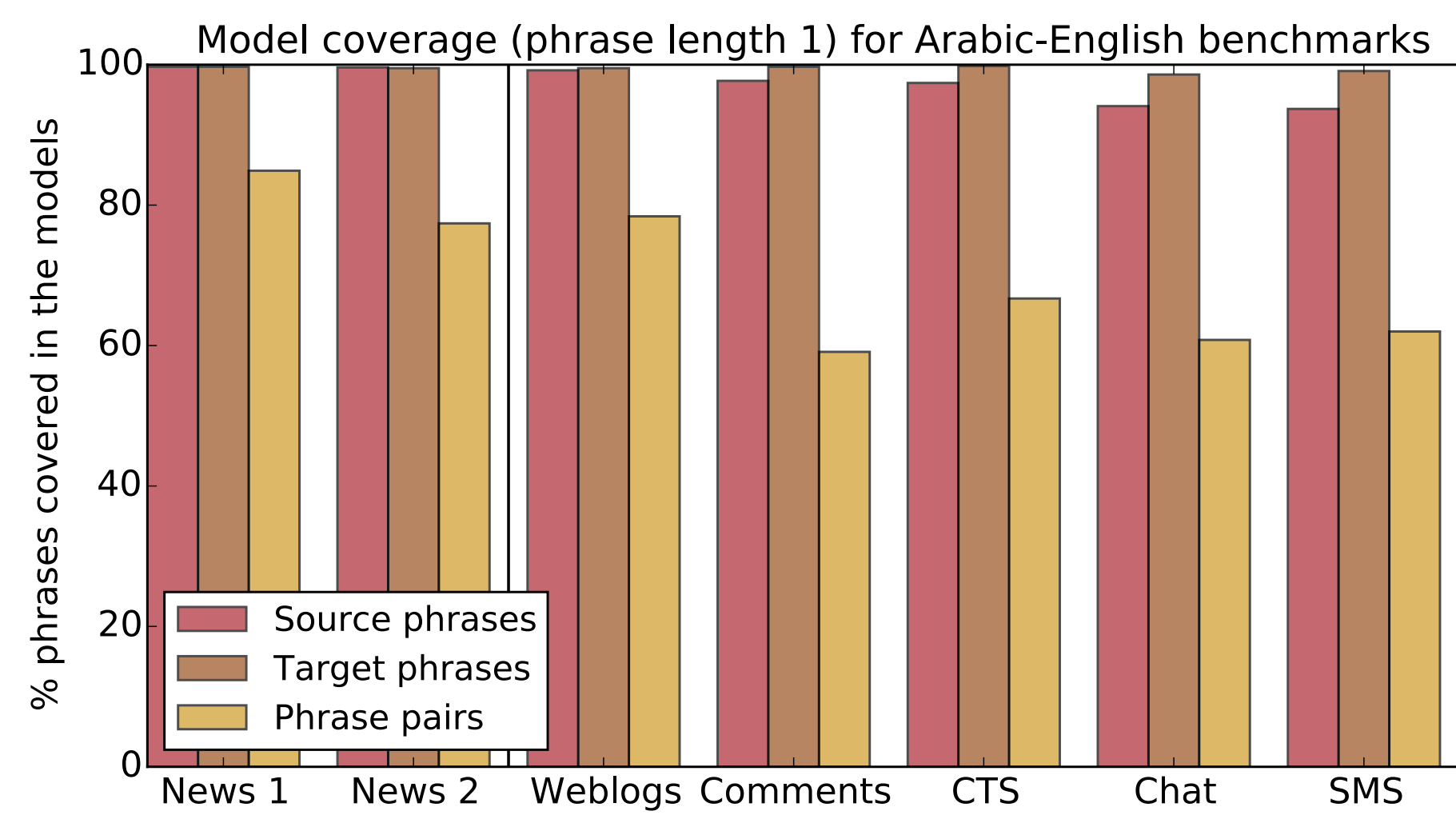
Quantitative Analysis: SMT Model Coverage

Approach

for each phrase pair in the test set (e.g. 你路上慢点 / take your time), determine:

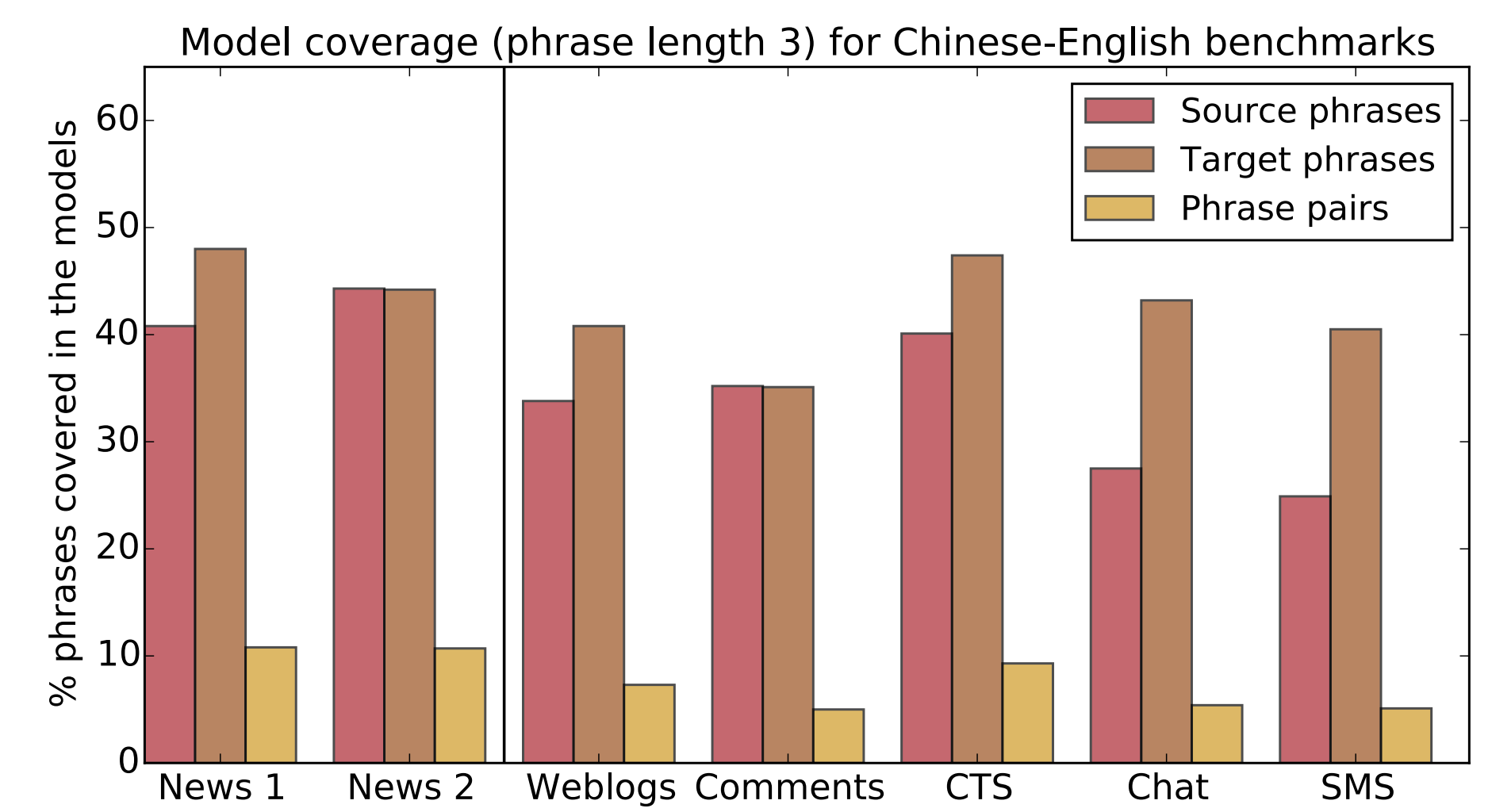
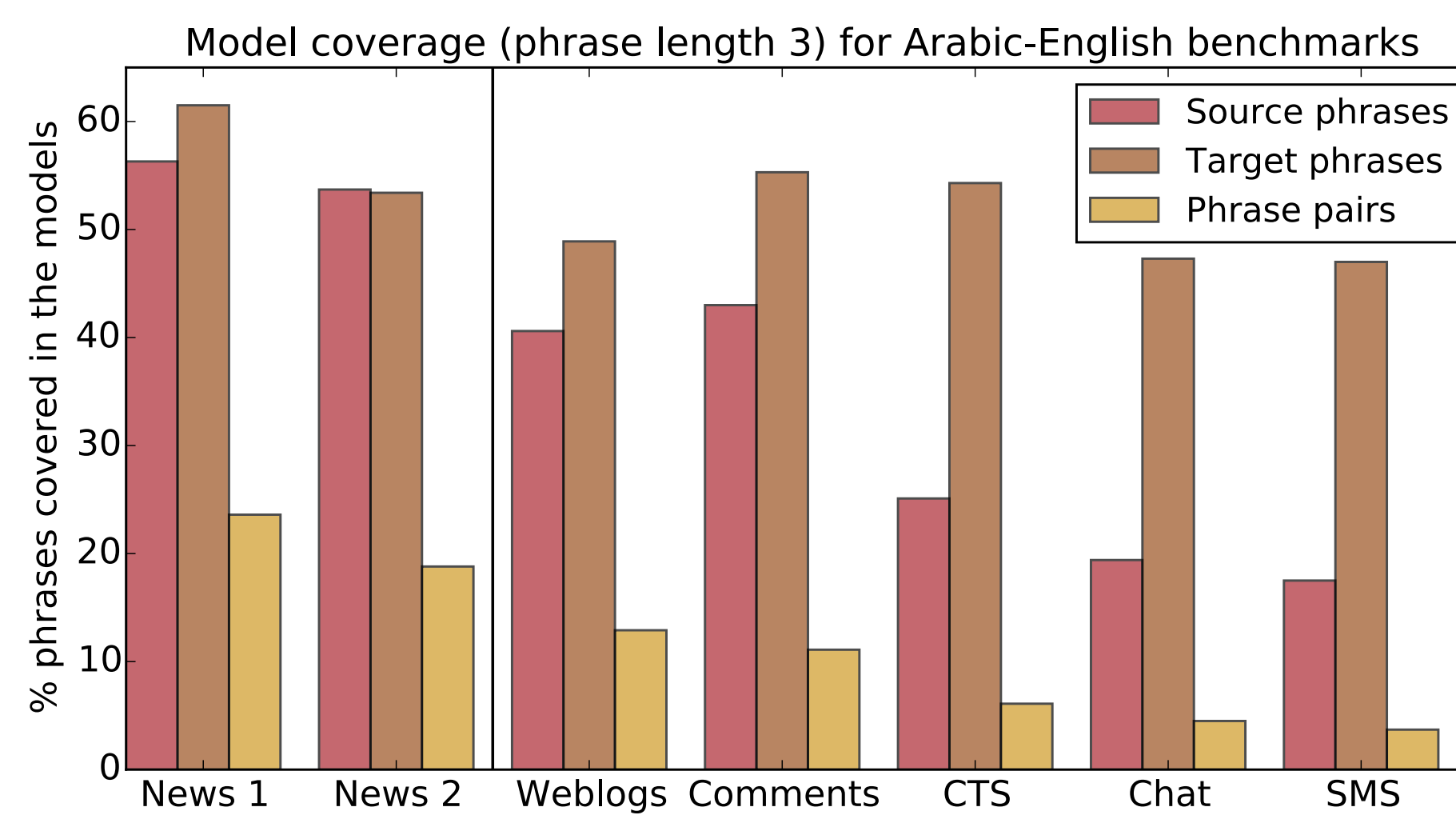
- **source phrase covered** in the SMT models
- **target phrase covered** in the SMT models
- **phrase pair covered** in the SMT models

all computed for various phrase lengths

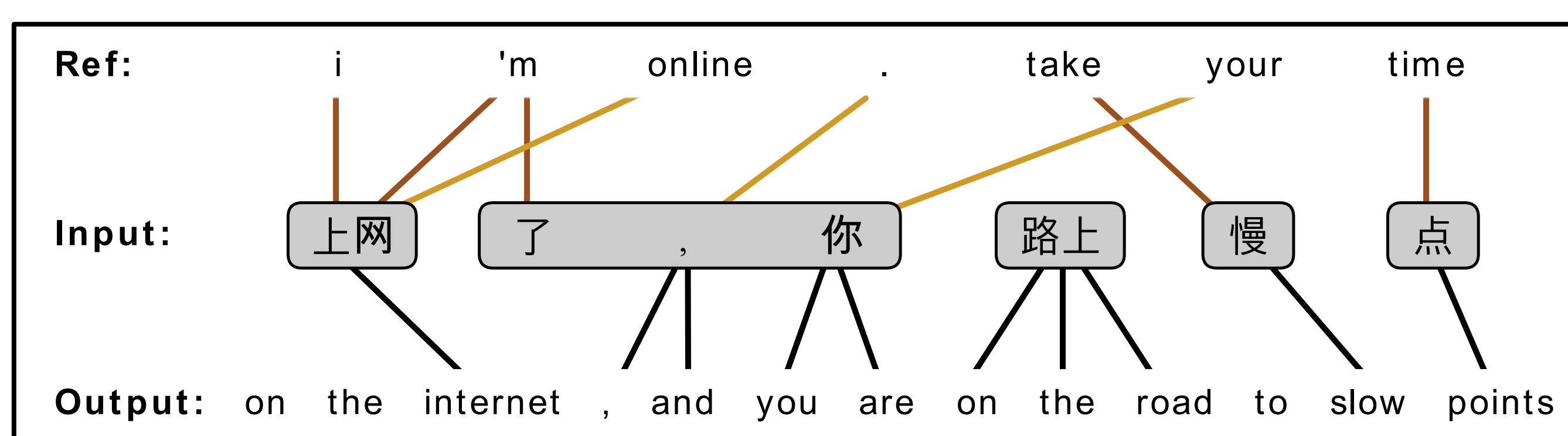


Findings

- coverage of **source phrases** and **phrase pairs** is lower for UG than for news
- coverage of **target phrases** is more balanced among test sets
- coverage dramatically decreases for longer phrases
- SMS and chat suffer most from low coverage

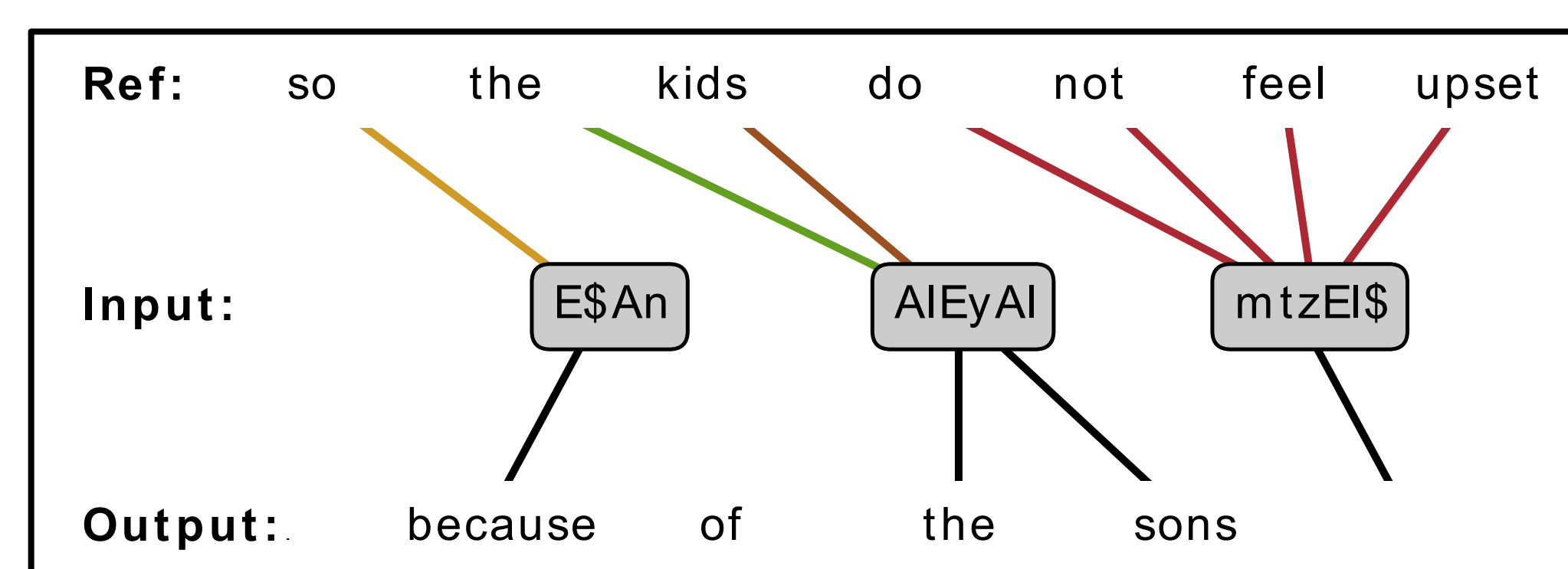


Qualitative Analysis: Word Alignment Driven Evaluation*



missing pronoun
not inferred by SMT system

idiom translated in small chunks
losing its meaning as a phrase



lexical choices that are too formal
not reflecting colloquial language

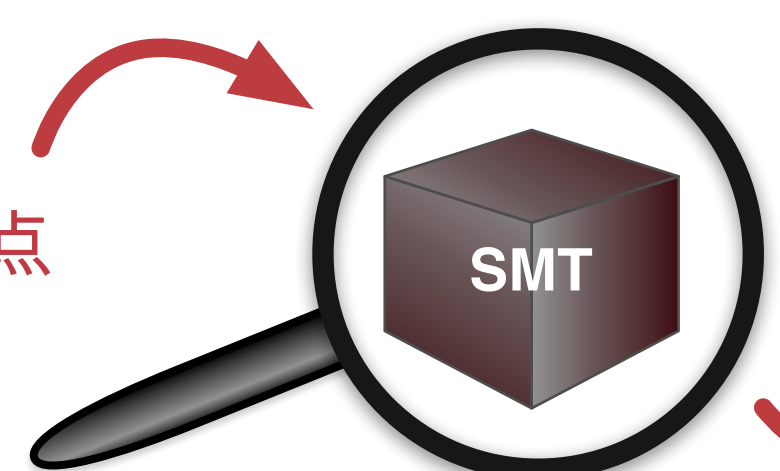
out-of-vocabulary (OOV)
due to dialect or misspellings

* Irvine et al., *Measuring Machine Translation Errors in New Domains*, 2013

Conclusions

UG text

你路上慢点



SMT errors for UG text differ

- from SMT errors for news
- between different types of UG
- between different language pairs

promising solutions include

- improving scoring for news
- increasing phrase pair coverage for UG
- increasing source phrase coverage for SMS & chat

UNIVERSITY OF AMSTERDAM

This research was funded in part by the Netherlands Organization for Scientific Research (NWO) under project number 639.022.213

