Jin-song SU, Xiao-dong SHI, Yan-zhou HUANG, Yang LIU, Qing-qiang WU, Yi-dong CHEN, Huai-lin DONG, 2014. Topic-aware pivot language approach for statistical machine translation. *Journal of Zhejiang University-SCIENCE C (Computers & Electronics),* **15**(4):241-253. [doi:10.1631/jzus.C1300208]

Topic-aware pivot language approach for statistical machine translation

Key words: Natural language processing, Pivot-based statistical machine translation, Topical context information

Corresponding author: Jin-song Su
E-mail: jssu@xmu.edu.cn

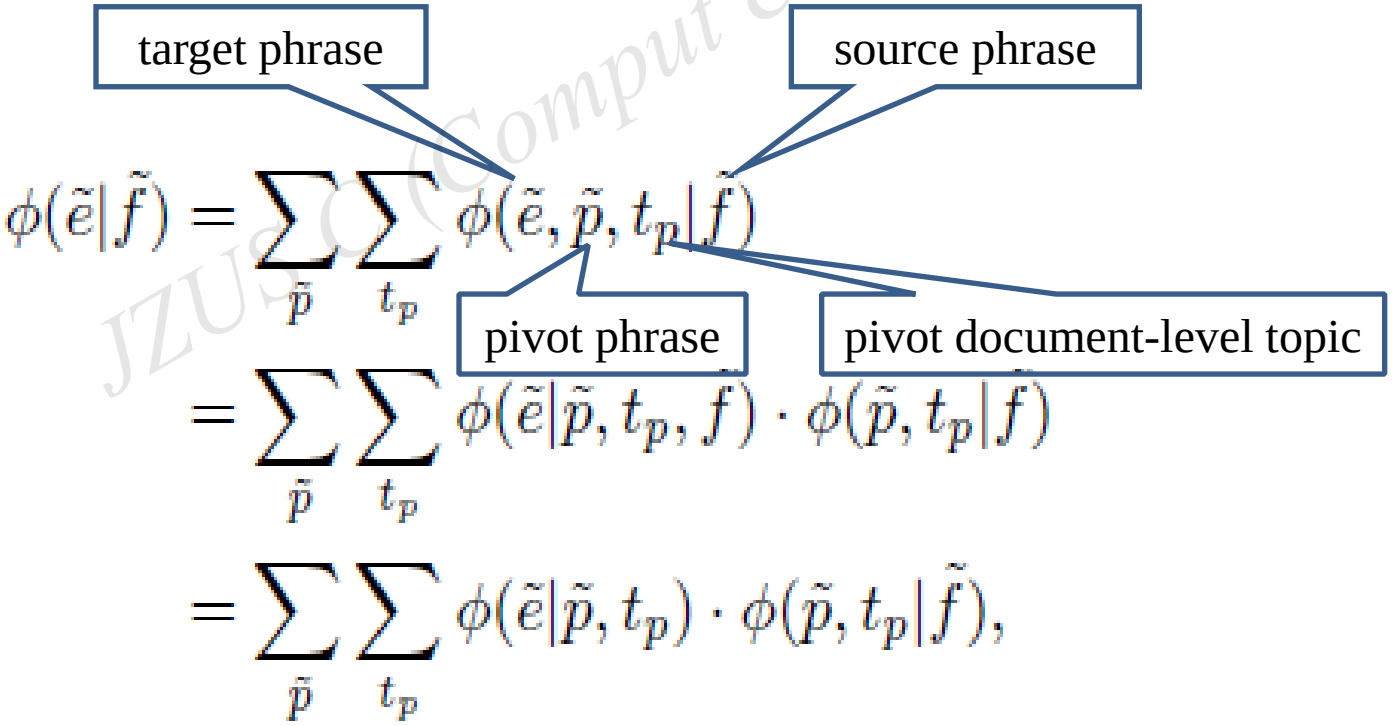
Motivation

- **Disadvantages of existing methods:** The pivot language approach for statistical machine translation (SMT) is a good method to break the resource bottleneck for certain language pairs. However, in the implementation of conventional approaches, pivot-side context information is far from being fully used, resulting in erroneous estimations of translation probabilities.
- **Our method: 1.** Take advantage of document-level context by assuming that the bridged phrase pairs should be similar in the document-level topic distributions. **2.** Focus on the effect of local context. **3.** Build an interpolated model bringing the above methods together to further enhance the system performance.

Framework of our method (I)

Method 1: Phrase table multiplication using pivot document-level topics as hidden variables

The most informative instance:



$$\phi(\tilde{e}|\tilde{f}) = \sum_{\tilde{p}} \sum_{t_p} \phi(\tilde{e}, \tilde{p}, t_p|\tilde{f})$$

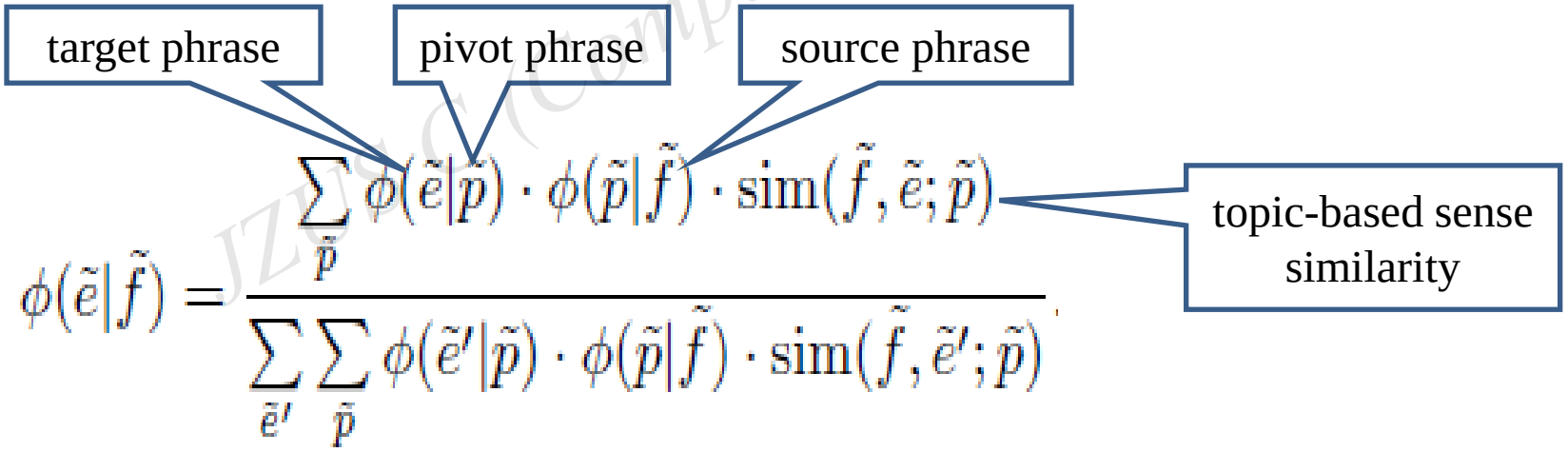
$$= \sum_{\tilde{p}} \sum_{t_p} \phi(\tilde{e}|\tilde{p}, t_p, \tilde{f}) \cdot \phi(\tilde{p}, t_p|\tilde{f})$$

$$= \sum_{\tilde{p}} \sum_{t_p} \phi(\tilde{e}|\tilde{p}, t_p) \cdot \phi(\tilde{p}, t_p|\tilde{f}),$$

Framework of our method (II)

Method 2: Translation probability embedded with the topic-based sense similarity

The most informative instance:



Summary

- **Motivation:** Exploit topic-based context information to improve pivot-based SMT.
- **Methodology:** We propose two topic-aware pivot language approaches to capture different levels of pivot-based context, which are incorporated into translation probability estimation in pivot-based SMT.
- **Performance:** Experimental results on French-Spanish and French-German translations using English as the pivot language demonstrated the effectiveness of topic-based context in pivot-based SMT.