# On the Importance of Pivot Language Selection for Asian Language Translation

Michael Paul*†, Hirofumi Yamamoto†‡, Eiichiro Sumita† and Satoshi Nakamura†
† NICT/ATR, Hikaridai 2-2-2, Keihanna Science City, 619-0288 Kyoto, Japan
‡ Kinki University School of Science and Engineering, Higashi-Osaka City, 577-8502, Japan
Email: Michael.Paul@nict.go.jp

***Abstract*—Recent research on multilingual statistical machine translation focuses on the usage of *pivot languages* in order to overcome resource limitations for certain language pairs. Due to the richness of available language resources, *English* is in general the pivot language of choice. In this paper, we investigate the appropriateness of Asian languages as pivot languages. Experimental results using state-of-the-art statistical machine translation techniques to translate between eight Asian languages revealed that the translation quality of 73% of the language pairs improved when a non-English pivot language was chosen.**

## I. Introduction

The translation quality of state-of-the-art phrased-based statistical machine translation (SMT) approaches heavily depends on the amount of bilingual language resources available to train the statistical models. For frequently used language pairs like *French-English* or *Chinese-English*, large-sized text data sets are readily available. There exist several international data collection initiatives like the *Linguistic Data Consortium* (LDC, http://www.ldc.upenn.edu), the *European Language Resource Association* (ELRA, http://www.elra.info), or *GSK* (http://www.gsk.or.jp/catalog.html) amassing and distributing large amounts of textual data. However, for less frequently used language pairs, e.g., most of the Asian languages, only a limited amount of bilingual resources are available, if at all.

In order to overcome such language resource limitations, recent research on multilingual statistical machine translation focuses on the usage of *pivot languages* [1]. Instead of a direct translation between two languages where only a limited amount of bilingual resources is available, the *pivot translation* approach makes use of a third language that is more appropriate due to the availability of more bilingual corpora and/or its relatedness towards either the source or the target language. In a first step, the source language input is translated into the pivot language using statistical translation models trained on the *source-pivot* language resources. In the second step, the obtained pivot language result is translated into the target language using a second translation engine trained on the *pivot-target* language resources.

In previous research on pivot translation approaches, the pivot language was mainly selected based on (a) the *availability of bilingual language resources* or (b) *language relatedness between source/pivot languages*. For most recent research efforts, *English* is the pivot language of choice due to the richness of available language resources. For example, [1] exploits the Europarl corpus for comparing pivot translation approaches between *French*/*German*/*Spanish* via *English*. Moreover, several research efforts tried to exploit the closeness between specific language pairs to achieve high-quality translation hypotheses in the first step to minimize the detoriation effects of the pivot approach. For example, [2] proposes a method to translate *Catalan*-to-*English* via *Spanish*.

However, both of the above criteria might not be sufficient to choose the best pivot language, especially for Asian languages where, besides for *Chinese*, only few parallel text corpora for Asian languages and *English* are publicly available. Moreover, language families in Asia are quite diverse.

This paper investigates the appropriateness of Asian languages as pivot languages to support future research on machine translation between Asian languages. Pivot translation experiments using state-of-the-art statistical machine translation techniques to translate between eight Asian languages (*Chinese*, *Hindi*, *Indonesian*, *Japanese*, *Korean*, *Malay*, *Thai*, *Vietnamese*) are carried out and the effects of selecting a non-*English* language as the pivot language are compared towards the *English* pivot approach.

Section II gives an overview on recent machine translation research efforts in Asia and summarizes the availability of Asian language resources. Section III introduces the pivot translation approach within the framework of statistical machine translation. Section IV outlines the pivot translation experiments, discusses the effects of using non-*English* pivot languages and identifies the best-suited pivot language for translations between Asian languages.

## II. Machine Translation for Asian Languages

The Asia-Pacific Association for Machine Translation (AAMT, http://www.aamt.info) lists a variety of machine translation products and machine translation services translating from/to Asian languages including *Hindi* (hi), *Indonesian* (id), *Japanese* (ja), *Korean* (ko), *Malay* (ms) *Thai* (th), *Vietnamese* (vi), and *Chinese* (zh). Table I summarizes the amount of publicly available translation systems/services for the respective language pairs. Most of the MT products and services focus on *Japanese* ⇔ *English* translations, followed by *Japanese* ⇔ *Korean* and *Chinese* ⇔ *English*. However, only a few MT tools and MT research prototypes are available for language translations involving *Hindi*, *Indonesian*, *Malay*, *Thai*, or *Vietnamese* [3].

Similarly, the amount of publicly available bilingual text corpora is quite limited. Table II summarizes the amount of bilingual text corpora or dictionaries available from *LDC*, *ELRA*, *GSK* and other publicly available multilingual resources used in recent MT evaluation campaigns like NIST or IWSLT. Most language resources are available for *Chinese* ⇔ *English*

TABLE I
ASIAN LANGUAGE MT PRODUCTS/SERVICES

| | en | hi | id | ja | ko | ms | th | vi | zh |
|---|---|---|---|---|---|---|---|---|---|
| en | – | 4 | 5 | 92 | 24 | 2 | 3 | 3 | 34 |
| hi | 4 | – | 3 | 3 | 3 | 1 | 1 | 3 | 3 |
| id | 5 | 3 | – | 3 | 3 | 1 | 1 | 3 | 3 |
| ja | 89 | 3 | 3 | – | 39 | 1 | 4 | 3 | 32 |
| ko | 20 | 3 | 3 | 36 | – | 1 | 1 | 3 | 9 |
| ms | 2 | 1 | 1 | 1 | 1 | – | 1 | 1 | 1 |
| th | 4 | 1 | 1 | 1 | 1 | 1 | – | 1 | 1 |
| vi | 3 | 3 | 3 | 3 | 3 | 1 | 1 | – | 3 |
| zh | 34 | 3 | 3 | 30 | 9 | 1 | 1 | 3 | – |

TABLE II
ASIAN LANGUAGE RESOURCES

| | en | hi | id | ja | ko | ms | th | vi | zh |
|---|---|---|---|---|---|---|---|---|---|
| en | 132 | 2 | 1 | 11 | 6 | 3 | 1 | - | 72 |
| hi | 3 | - | - | - | - | - | - | - | |
| id | - | - | - | - | - | 1 | 1 | 1 | |
| ja | 13 | - | 1 | 1 | - | - | - | - | |
| ko | 14 | - | - | - | - | - | - | - | |
| ms | 2 | 1 | - | 1 | - | - | - | 1 | |
| th | - | - | - | 1 | - | - | - | 1 | |
| vi | - | - | - | - | - | - | - | - | |
| zh | 64 | | | | | | | | |

and *Japanese* ⇔ *English*. However, almost no bilingual data is available for the other language pairs.

In order to fill the gap, recent research activities on spoken language translation conducted by the *Asian Speech Translation Consortium* (ASTAR, http://www.slc.atr.jp/AStar), which is an international partnership of research laboratories that focuses on the development of large-scaled spoken language corpora in Asia. As a first result of these activities, a Japanese-English speech corpus [4] comprising tourism-related sentences has been translated into the native languages of the ASTAR members resulting in a multilingual sentence-aligned corpus [5]. This corpus is introduced in detail in Section IV and is exploited to investigate the effects of pivot language selection for Asian languages.

## III. PIVOT TRANSLATION

*Pivot translation* is a translation from a source language (SRC) to a target language (TRG) through an intermediate *pivot* (or *bridging*) *language* (PVT). Within the SMT framework, the following coupling strategies have already been investigated:

1) *cascading of two translation systems* where the first MT engine translates the source language input into the pivot language and the second MT engine takes the obtained pivot language output as its input and translates it into the target language.
2) *pseudo corpus* approach that (a) creates a "noisy" SRC-TRG parallel corpus by translating the pivot language parts of the SRC-PVT training resources into the target language using an SMT engine trained on the PVT-TRG language resources and (b) directly translates the source language input into the target language using a single SMT engine that is trained on the obtained SRC-TRG language resources [2].
3) *phrase-table composition* in which the translation models of the SRC-PVT and PVT-TRG translation engines are combined to a new SRC-TRG phrase-table by merging SRC-PVT and PVT-TRG phrase-table entries with identical pivot language phrases and multiplying posterior probabilities [1], [6].
4) *bridging at translation time* where the coupling is integrated into the SMT decoding process by modeling the pivot text as a hidden variable and assuming independence between source and target sentences [7].

In order to investigate the effects of the pivot language selection for statistical machine translation involving Asian languages, the simplest method of *cascading two SMT systems* is exploited in the pivot translation experiments reported in Section IV.

## IV. EXPERIMENTS

The effects of pivot language selection are investigated using the *Basic Travel Expressions Corpus* (BTEC), which is a collection of sentences that bilingual travel experts consider useful for people going to or coming from another country. The corpus currently covers 18 of the major world languages [5]. Besides *English*, we selected the eight Asian languages listed in Section II for the pivot translation experiments.

These languages differ largely in *word order* (SVO, SOV), *segmentation unit* (phrase, word, none), and *morphology* (poor, medium, rich). Concerning word segmentation, the corpora were preprocessed using word-segmentation tools for languages that do not use white-space to separate word tokens like *Chinese*, *Japanese*, *Korean*, and *Thai*. All data sets were case-sensitive with punctuation marks preserved.

However, in a real-world application, identical language resources covering three or more languages are not necessarily to be expected. In order to avoid a trilingual scenario for the pivot translation experiments described in this paper, the full BTEC corpus consisting of 160k sentence-aligned data sets was randomly split into two subsets of 80k sentences each, whereby the first set of sentence pairs was used to train the source-to-pivot translation models ($80k^{sp}$) and the second subset of sentence pairs was used to train the pivot-to-target translation models ($80k^{pt}$).

The characteristics of the utilized BTEC corpus data sets are summarized in Table III. The sentence length is given as the average number of words per sentence. In order to get an idea of how difficult the translation task for the different languages is supposed to be, we calculated the language perplexity of the target language evaluation data sets according to a standard 5-gram language model trained on the respective training data sets. For each language, the *language perplexity* and the *total entropy*, i.e., the entropy multiplied by the number of words of the evaluation data set, are listed. The higher the total entropy, the more difficult the translation task is supposed to be.

For the training of the statistical models, standard word alignment (GIZA++ [8]) and language modeling (SRILM [9]) tools were used. For translation, the *Cleopatra* decoder, an in-house phrase-based SMT decoder [10], was used.

For the evaluation of translation quality, a test data set consisting of 510 sentences of the BTEC corpus reserved for evaluation purposes was translated by the respective translation engines and evaluated using the standard automatic evaluation metrics BLEU [11] and METEOR [12]. For the experimental

TABLE III
PIVOT TRANSLATION LANGUAGE RESOURCES

| BTEC Corpus | | training 80k$^{sp}$ | training 80k$^{pt}$ | eval set | perplexity (total entropy) |
|---|---|---|---|---|---|
| # of sentences | | 80,000 | 80,000 | 510 | – |
| en | vocabulary | 12,264 | 11,129 | 896 | 25.6 |
| | avg. length | 7.8 | 7.1 | 7.5 | (17899.0) |
| hi | vocabulary | 20,725 | 12,320 | 1162 | 104.9 |
| | avg. length | 8.4 | 7.5 | 8.1 | (27644.8) |
| id | vocabulary | 14,585 | 13,343 | 1,000 | 50.2 |
| | avg. length | 7.0 | 6.5 | 7.1 | (20467.0) |
| ja | vocabulary | 13,868 | 12,621 | 959 | 18.1 |
| | avg. length | 8.8 | 8.2 | 8.5 | (18136.4) |
| ko | vocabulary | 13,546 | 12,381 | 946 | 17.8 |
| | avg. length | 8.3 | 7.8 | 8.1 | (17135.1) |
| ms | vocabulary | 15,113 | 13,752 | 995 | 50.9 |
| | avg. length | 7.1 | 6.6 | 7.1 | (20402.1) |
| th | vocabulary | 6,103 | 5,637 | 738 | 36.5 |
| | avg. length | 8.1 | 7.5 | 7.7 | (20516.8) |
| vi | vocabulary | 7,980 | 7,388 | 884 | 26.1 |
| | avg. length | 9.4 | 8.7 | 9.2 | (22043.2) |
| zh | vocabulary | 11,084 | 10,220 | 908 | 28.3 |
| | avg. length | 6.6 | 6.6 | 7.0 | (17136.2) |

TABLE IV
AUTOMATIC EVALUATION METRICS

| | |
|---|---|
| BLEU: | the geometric mean of n-gram precision by the system output with respect to reference translations. Scores range between 0 (worst) and 1 (best) [11] |
| METEOR: | a metric that calculates unigram overlaps between translation and reference texts taking into account various levels of matches (*exact*, *stem*, *synonym*). Scores range between 0 (worst) and 1 (best) [12] |

results in this paper, the given scores are calculated as the average of the respective BLEU and METEOR scores obtained for each system output.

### A. Direct Translation

The automatic evaluation scores for all source and target language pair combinations of the direct translation approach are summarized in Table V. For each target language, the highest evaluation scores are marked in boldface.

Despite slight score differences, quite similar levels of translation quality were obtained by the SRC-PVT and the PVT-TRG models for most of the language pairs. The highest evaluation scores were achieved for closely related language pairs like *Japanese* $\Leftrightarrow$ *Korean* and *Indonesian* $\Leftrightarrow$ *Malay*. In addition, relatively high translation quality was achieved for *Japanese* $\Leftrightarrow$ *Chinese*. As indicated by the total entropy figures in Table III, the hardest translation tasks were those translating to *Hindi* followed by *Malay*, *Thai*, and *Indonesian*.

Interestingly, all language pairs having *English* as the source language achieved not always the highest scores, but performed better than most of the remaining languages which is in contrast to the language pairs translating into *English* where only mid-level scores were achieved. This indicates, that a relatively larger detoriation in translation quality is to be expected for the SRC-PVT translation step when *English* is used as the pivot language compared to other pivot languages where higher evaluation scores for translations into the pivot language were obtained.

### B. Pivot Translation

The automatic evaluation scores for all pivot translation language-pair combinations (*SRC-PVT-TRG*) are summarized in Table VI whereby for each source-target language pair, the results of the pivot translation experiments using (a) *English*(*en*) and (b) the best performing language (*best*) as the pivot language are listed. The experimental results show that *English* is indeed the best pivot language when translating from *Hindi* or *Malay* into most of the other target languages. However, for *Korean* and *Chinese* as the source language, significantly lower evaluation scores were obtained for *English* compared to several other Asian pivot languages for all translation directions. In the case of *Korean*, *Japanese* is the pivot language of choice when translating into other Asian languages, and vice versa. For *Chinese*, *Indonesian*, *Thai*, and *Vietnamese*, the optimal pivot language depends largely on the respective target language.

However, the selection of the optimal pivot language is not symmetric for most of the language pairs. Quite a different picture is obtained when analyzing the results according to the respective target language. For translations into *Hindi*, either *Japanese* or *Korean* should be preferred towards *English* as the pivot language. In the case of *Malay* as the target language, *Indonesian* outperforms all other pivot languages. Nevertheless, the *English* pivot approach seems to be effective for most translations into *Korean* and *Japanese*.

Comparing the results of the pivot translation experiments towards the direct translation results, we can see that for 43% of the language pairs, the pivot approach outperforms the direct translation approach significantly when the optimal pivot language is selected. This phenomena is caused mainly by (a) the *unrelatedness between SRC and TRG* (18$\times$), (b) the *closeness between PVT and TRG* (3$\times$) and (c) the *closeness between SRC and PVT* (1$\times$).

### C. Pivot Language Selection

Besides the automatic evaluation scores, Table VI lists also the optimal pivot language for each source-target language pair in boldface. Moreover, for each language, the amount of language pairs that achieved the highest scores using this language as the pivot is given in Table VII. The experimental results show that the *English* pivot approach still achieves the highest scores for the majority of the examined language pairs. However, in 73% of the cases, an Asian pivot language, mainly *Japanese*, *Malay*, *Indonesian*, and *Korean* is to be preferred.

## V. CONCLUSION

In this paper, the effects of using Asian pivot languages for translations between eight Asian languages were compared to the standard English pivot translation approach. The experimental results revealed that *English* was indeed more frequently (26.8% out of 56 language pairs) selected as the best pivot language over any other Asian language. However, its usage is mainly limited to translations from *Hindi* and *Malay* or translations into *Korean*. Otherwise, the *English* pivot approach is significantly outperformed by using Asian languages as the pivot languages, especially *Japanese*, *Malay*, *Indonesian*, or *Korean*. In contrast to previous research on

TABLE V
TRANSLATION QUALITY OF DIRECT TRANSLATION APPROACHES

| SRC\TRG | en | hi | id | ja | ko | ms | th | vi | zh |
|---|---|---|---|---|---|---|---|---|---|
| en | – | **0.4370** | 0.6471 | 0.8157 | 0.7214 | 0.6215 | **0.6423** | **0.7045** | 0.7101 |
| hi | 0.4906 | – | 0.4695 | 0.5542 | 0.4210 | 0.4719 | 0.4175 | 0.5174 | 0.4714 |
| id | 0.6526 | 0.4135 | – | 0.7233 | 0.6218 | **0.7898** | 0.5956 | 0.6556 | 0.6203 |
| ja | **0.6736** | 0.4224 | 0.6392 | – | **0.8460** | 0.5980 | 0.6147 | 0.6888 | **0.7817** |
| ko | 0.6538 | 0.4277 | 0.6208 | **0.8763** | – | 0.5921 | 0.6107 | 0.6708 | 0.7675 |
| ms | 0.6405 | 0.3900 | **0.7907** | 0.7101 | 0.6151 | – | 0.5940 | 0.6386 | 0.6110 |
| th | 0.6021 | 0.3946 | 0.6024 | 0.6969 | 0.6205 | 0.5647 | – | 0.6768 | 0.6557 |
| vi | 0.6320 | 0.4060 | 0.5934 | 0.7026 | 0.6171 | 0.5695 | 0.5897 | – | 0.6241 |
| zh | 0.6513 | 0.4029 | 0.6238 | 0.7891 | 0.7371 | 0.5797 | 0.6047 | 0.6806 | – |

TABLE VI
TRANSLATION QUALITY OF PIVOT TRANSLATION APPROACHES

| SRC | PVT | hi | id | ja | ko | ms | th | vi | zh |
|---|---|---|---|---|---|---|---|---|---|
| hi | en | – | 0.5105 | 0.6261 | 0.5248 | 0.4973 | 0.4301 | 0.5352 | 0.5240 |
| | *best* | – | (**en**) 0.5105 | (**ms**) 0.6287 | (**en**) 0.5248 | (**en**) 0.4973 | (**en**) 0.4301 | (**en**) 0.5352 | (**en**) 0.5240 |
| id | en | 0.4102 | – | 0.7251 | 0.6451 | 0.6135 | 0.5692 | 0.6256 | 0.6223 |
| | *best* | (**ja**) 0.4221 | – | (**en**) 0.7251 | (**en**) 0.6451 | (**vi**) 0.6227 | (**ms**) 0.5858 | (**ms**) 0.6563 | (**ja**) 0.6447 |
| ja | en | 0.4217 | 0.6043 | – | 0.7189 | 0.5952 | 0.5864 | 0.6685 | 0.6964 |
| | *best* | (**ko**) 0.4250 | (**ms**) 0.6160 | – | (**zh**) 0.7286 | (**id**) 0.6119 | (**ko**) 0.6079 | (**en**) 0.6685 | (**ko**) 0.7729 |
| ko | en | 0.4145 | 0.5871 | 0.7612 | – | 0.5718 | 0.5727 | 0.6536 | 0.6948 |
| | *best* | (**ja**) 0.4266 | (**ja**) 0.6207 | (**zh**) 0.7739 | – | (**id**) 0.5994 | (**ja**) 0.6123 | (**ja**) 0.6802 | (**ja**) 0.7499 |
| ms | en | 0.4234 | 0.6325 | 0.7313 | 0.6301 | – | 0.5687 | 0.6188 | 0.6182 |
| | *best* | (**en**) 0.4234 | (**en**) 0.6325 | (**en**) 0.7313 | (**en**) 0.6301 | – | (**id**) 0.5789 | (**id**) 0.6372 | (**ko**) 0.6254 |
| th | en | 0.4031 | 0.5903 | 0.7103 | 0.6141 | 0.5738 | – | 0.6319 | 0.6234 |
| | *best* | (**vi**) 0.4062 | (**en**) 0.5903 | (**ms**) 0.7236 | (**zh**) 0.6242 | (**id**) 0.5803 | – | (**zh**) 0.6446 | (**vi**) 0.6401 |
| vi | en | 0.3907 | 0.5707 | 0.7262 | 0.6477 | 0.5508 | 0.5504 | – | 0.6266 |
| | *best* | (**ko**) 0.4014 | (**ms**) 0.5956 | (**th**) 0.7295 | (**en**) 0.6477 | (**id**) 0.5614 | (**ms**) 0.5588 | – | (**th**) 0.6454 |
| zh | en | 0.4039 | 0.5935 | 0.7565 | 0.6794 | 0.5618 | 0.5705 | 0.6473 | – |
| | *best* | (**ko**) 0.4191 | (**ms**) 0.6062 | (**ko**) 0.7841 | (**ja**) 0.7140 | (**id**) 0.5866 | (**ja**) 0.6032 | (**th**) 0.6568 | – |

TABLE VII
PIVOT LANGUAGE SELECTION

| PVT | usage (%) | | PVT | usage (%) | |
|---|---|---|---|---|---|
| en | 15 | (26.8) | ko | 7 | (12.5) |
| ja | 9 | (16.1) | zh | 4 | (7.2) |
| ms | 8 | (14.3) | th | 3 | (5.3) |
| id | 7 | (12.5) | vi | 3 | (5.3) |

pivot translation approaches, the most prominent criteria on how to select the best pivot language was not the *language relatedness between source and pivot languages*, but the *unrelatedness between source and target languages*. In such a case, the pivot language approach frequently achieved a better translation quality than the one of the direct translation of the source language input into the target language, because the translation quality of the source-pivot and pivot-target engines were significantly higher than the direct translation engines.

Future research will have to investigate in detail, what kind of features are important to select a pivot language for new Asian source and target language pairs. Besides the translation quality of SMT engines, automatic metrics to measure the closeness of a language pair should also be taken into account to find optimal pivot languages and improve the usability of machine translation between Asian languages further.

## ACKNOWLEDGMENT

This work is partly supported by the Grant-in-Aid for Scientific Research (C) Number 19500137 and "Construction of speech translation foundation aiming to overcome the barrier between Asian languages", the Special Coordination Funds for Promoting Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

## REFERENCES

[1] M. Utiyama and H. Isahara, "A comparison of pivot methods for phrase-based statistical machine translation," in *Proc. of HLT*, New York, USA, 2007, pp. 484–491.

[2] A. Gispert and J. Marino, "Catalan-english statistical machine translation without parallel corpus: bridging through spanish," in *Proc. of 5th LREC*, Genoa, Italy, 2006, pp. 65–68.

[3] H. Isahara et al (Ed.), "Special Issue on Machine Translation Activities in Asia," *AAMT Journal*, pp. 1–40, 2005.

[4] G. Kikui, S. Yamamoto, T. Takezawa, and E¿ Sumita, "Comparative study on corpora for speech translation," *IEEE Transactions on Audio, Speech and Language*, vol. 14(5), pp. 1674–1682, 2006.

[5] M. Paul et al, "Multilingual Mobile-Phone Translation Services," in *Proc. of 22nd COLING*, Manchester, UK, 2006, pp. 165–168.

[6] H. Wu and H. Wang, "Pivot Language Approach for Phrase-Based Statistical Machine Translation," in *Proc. of ACL*, Prague, Czech Republic, 2007, pp. 856–863.

[7] N. Bertoldi, M. Barbaiani, M. Federico, and R¿ Cattoni, "Phrase-Based Statistical Machine Translation with Pivot Languages," in *Proc. of the IWSLT*, Hawaii, USA, 2008, pp. 143–149.

[8] F. Och and H. Ney, "A Systematic Comparison of Statistical Alignment Models," *Computational Linguistics*, vol. 29(1), pp. 19–51, 2003.

[9] A. Stolcke, "SRILM - an extensible language modeling toolkit," in *Proc. of ICSLP*, Denver, 2002, pp. 901–904.

[10] A. Finch et al, "The NICT/ATR Speech Translation System for IWSLT 2007," in *Proc. of the IWSLT*, Trento, Italy, 2007, pp. 103–110.

[11] K. Papineni, S. Roukos, T. Ward, and W. Zhu, "BLEU: a Method for Automatic Evaluation of Machine Translation," in *Proc. of the 40th ACL*, Philadelphia, USA, 2002, pp. 311–318.

[12] S. Banerjee and A. Lavie, "METEOR: An Automatic Metric for MT Evaluation," in *Proc. of the ACL*, Ann Arbor, US, 2005, pp. 65–72.