# Name Extraction and Translation for Distillation

**Heng Ji and Ralph Grishman**
New York University

**Dayne Freitag, Matthias Blume and Zhiqiang (John) Wang**
Fair Isaac Corp.

**Shahram Khadivi, Richard Zens and Hermann Ney**
RWTH

## Abstract

Name translation is important well beyond the relative frequency of names in a text: a correctly translated passage, but with the wrong name, may lose most of its value. The Nightingale team has built a name translation component which operates in tandem with a conventional phrase-based statistical MT system, identifying names in the source text and proposing translations to the MT system. Versions have been developed for both Chinese-to-English and Arabic-to-English name translation. The system has four main components, a name tagger, translation lists, a transliteration engine, and a context-based ranker. This chapter presents these components in detail and investigates the impact of name translation on cross-lingual spoken sentence retrieval.

## 1 Introduction

Traditional MT systems focus on the overall fluency and accuracy of the translation but fall short in their ability to translate certain informationally critical words. In particular, the translation of names is fundamentally different from the translation of other lexical items. Table 1 shows the wide range of cases that must be addressed in translating Chinese names into English, according to whether a name is rendered phonetically (P), semantically (S), or a mixture of both (M).

We may be expected to translate source language tokens which do not appear in the training corpus, based on our knowledge of transliteration correspondences (e.g. "*You shen ke*" to "*Yushchenko*") and of contexts in the target language (e.g. to distinguish "*Yasser Arafat*" from "*Yasir Arafat*"). In addition, some source names may appear in abbreviated form and may be mistranslated unless they are recognized as abbreviations. For example, "以" is the abbreviation for 'Israel' but can also be translated into the common word 'as'. Furthermore, errors may be compounded when part of an OOV name is mistakenly segmented into common words. For example, "*瓦斯涅夫斯基(Kwasniewski)*" is mistakenly translated into "*gas(瓦斯) Novsky(涅夫斯基)*" by a phrase-based statistical MT system; "*博贝列夫 (Bobylev)*" receives incorrect translations from different MT systems because it is not recognized as a name: "*German Gref*", "*Bo, yakovlev*", "*Addis Ababa*", "*A. Kozyrev*" and "*1988 lev*".

Name translation is important well beyond the relative frequency of names in a text: a correctly translated passage, but with the wrong name, may lose most of its value. Many GALE distillation templates involve names, so name processing is the key for accurate distillation from foreign languages. We found that distillation performed notably worse on machine translated texts than on texts originally written in English, and our error analysis indicated that a major cause was the low quality of name translation. Thus, it appears that better entity name translation can substantially improve the utility of machine translation and the amount of information that can be gleaned from foreign sources. To meet these challenges, the Nightingale team has built a name translation component which operates in tandem with a conventional phrase-based statistical MT system, identifying names in the source text and proposing translations to the MT system. Versions have been developed for both Chinese-to-English and Arabic-to-English name translation.

## 2 System Overview

The overall system pipeline is summarized in Figure 1. This system runs a source-language name tagger, then uses a variety of strategies to translate the names it identifies. We shall present each of the main components in the following sections.

| Name Translation Types | | Chinese | English |
|---|---|---|---|
| Context/ Ethnic Independent | P→ P | *尤申科 (You shen ke)* | *Yushchenko* |
| | P→ S | *可伶可俐 (Ke Ling Ke Li)* | *Clean Clear* |
| | P→ M | *欧佩尔吧* | *Opal Bar* |
| | S→ P | *旧金山 (Old Golden Mountain)* | *San Francisco* |
| | S→ S | *解放之虎* | *Liberation Tiger* |
| | S→ M | *长江 (Long River)* | *Yangtze River* |
| | M→ P | *清华大学学报 (The Journal of Tsinghua University)* | *Tsinghua Da Xue Xue Bao* |
| | M→ S | *百斯百网站* | *Best Buy Website* |
| | M→ M | *尤干斯克石油天然气公司* | *Yuganskneftegaz Oil and Gas Company* |
| Context/ Ethnic Dependent | P→P | *亚西尔·阿拉法特* | *Yasser Arafat (PLO Chairman)* |
| | | | *Yasir Arafat (Cricketer)* |
| | | *潘基文* | *Pan Jiwen (Chinese)* |
| | | | ***Ban Ki-Moon** (Korean Foreign Minister)* |
| | S→S | *红军* | *Red Army (in China)* |
| | | | *Liverpool Football Club (in England)* |
| | M→M | *圣地亚哥市* | *Santiago City (in Chile)* |
| | | | *San Diego City (in CA)* |

Table 1. Examples for Diverse Name Translation Types

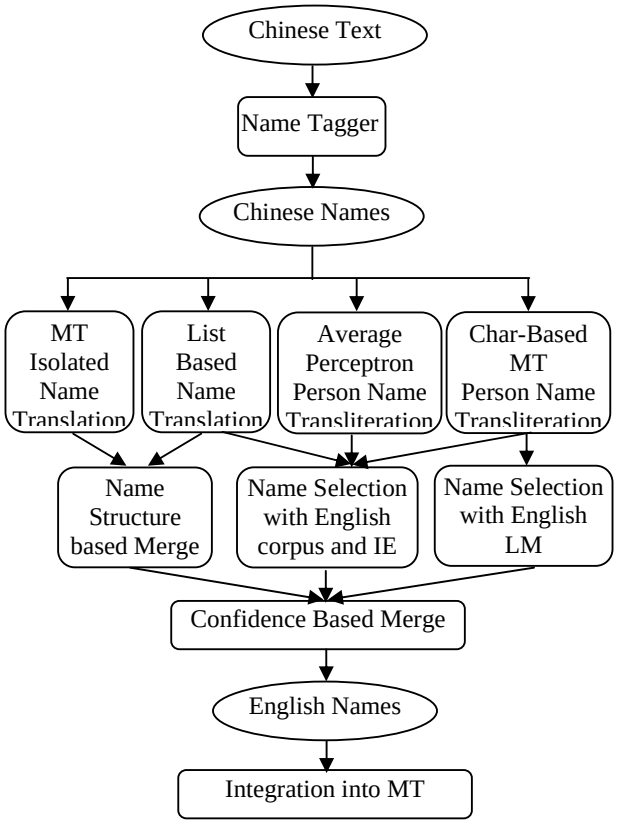


Figure 1. System Overview

## 3 Name Tagging

Both Arabic and Chinese name taggers are trained on several years of ACE (Automatic Content Extraction[1]) corpora, and can identify names and classify them as PER (persons), ORG (organizations), GPE ('geo-political entities' – locations which are also political units, such as countries, counties, and cities) and LOC (other locations). In the following we will describe Arabic and Chinese entity extraction systems respectively.

### 3.1 Arabic Name Tagging

To identify Arabic names we trained a structured perceptron model, as detailed in (Farber et al., 2008). Structured perceptrons are in a class of models, which also includes conditional random fields and hidden Markov SVMs, that can exploit arbitrary features of the input in search of an optimal assignment of labels to a given input. We compensate for Arabic's lack of name-relevant orthographic features (i.e., capitalization) and its relative lack of lexical resources (e.g., gazetteers) by expanding the feature set available to the model. First, we derive term clusters through a statistical analysis of a large volume of unlabeled Arabic newswire text, and treat a given term's cluster membership as a feature (Freitag, 2004). Second, we apply MADA, a tool from Columbia for Arabic morphological analysis and word sense disambiguation. Two Boolean features are derived from the output of MADA, the first reflecting whether the analysis is successful

[1] http://www.nist.gov/speech/tests/ace/

(whether the word was in or out of vocabulary), and the second indicating whether the English gloss returned by MADA is capitalized.

### 3.2 Chinese Name Tagging

The Chinese name tagger consists of a HMM tagger augmented with a set of post-processing rules (Ji and Grishman, 2006). The HMM tagger generally follows the Nymble model (Bikel et al., 1997). Within each of the name class states, a statistical bigram model is employed, with the usual one-word-per-state emission. The various probabilities involve word co-occurrence, word features, and class probabilities. Since these probabilities are estimated based on observations seen in a corpus, several levels of "back-off models" are used to reflect the strength of support for a given statistic, including a back-off from words to word features, as for the Nymble system. To take advantage of Chinese names, we extend the model to include a larger number of states, 14 in total. The expanded HMM can handle name prefixes and suffixes, and has separate states for transliterated foreign names. Finally a set of post-processing heuristic rules are applied to correct some omissions and systematic errors.

## 4 Candidate Name Translations

We first apply the source language name taggers to identify named entities, and then apply the following various techniques to generate a set of candidate translations for each name.

### 4.1 MT Isolated Name Translation

As a baseline we translate each source phrase referring to a named entity (a name 'mention') in isolation into English. The only difference from full text translation is that, as subsentential units are translated, sentence boundaries are not assumed at the beginning and end of each input name mention[2].

The RWTH Aachen Chinese-to-English machine translation system (Zens and Ney, 2004; Zens et al., 2005) is used. It's a statistical, phrase-based machine translation system which memorizes all phrasal translations that have been observed in the training corpus. The posterior probability is modeled directly using a weighted log-linear combination of various models: an n-gram language model, phrase translation models and word-based lexicon models as well as a lexicalized reordering model. The model scaling factors are tuned on a development set to maximize the translation quality (Och, 2003). The bilingual training data consists of about 8 million Chinese-English sentence pairs with more than 200 million words in each language. The language model was trained on the English part of the bilingual training data and additional monolingual data from the GigaWord corpus, about 650 million English words in total. The Chinese text is segmented into words using the tool provided by the Linguistic Data Consortium (LDC).

### 4.2 Name Pair Mining

We exploited a variety of approaches to automatically mine about 80,000 name pairs, as follows.

- *Extracting cross-lingual name titles from Wikipedia pages*. We run a web browser (Artiles et al., 2008) to extract titles from Chinese Wikipedia pages and their corresponding linked English pages (if the link exists). Then we apply heuristic rules based on Chinese name structure to detect name pairs, for example, foreign full names must include a dot separator, Chinese person names must include a last name from a closed set of 437 family names.

- *Tagging names in parallel corpora*. Within each sentence pair in a parallel corpus, we run the Chinese Name tagger as described in section 3.2 and the NYU English name tagger (Grishman et al., 2005). If the types of the name tags on both sides are identical, we extract the name pairs from this sentence. Then at the corpus-wide level, we count the frequency for each name pair, and only keep the name pairs that are frequent enough. Each member of the name pair then becomes the translation of the other member. The corpora used for this approach were all GALE MT training corpora and ACE 07 Entity Translation training corpora.

- *Using patterns for Web mining*: we constructed heuristic patterns such as "Chinese name (English name)" to extract NE pairs

[2] We tried an alternative approach in which mentions are translated in context and the mention translations are then extracted using word alignment information produced by the MT system, but it did not perform as well. The word alignments are indirectly derived from phrase alignment and can be quite noisy. As a result, noise in the form of words from the target language context is introduced into the mention translations. Manual evaluation on a small development set showed that isolated translation obtains (about 14%) better F-measure in translating names.

from web pages with mixed Chinese and English.

In addition, we exploited an LDC bilingual name dictionary and a Japanese-English person name dictionary including 20126 entries (Kurohashi et al., 1994).

Besides full string matching, we also parse name structures (e.g. first name and last name for persons; modifier and suffix word for organizations) and then match each name component separately. We have also developed a name ethnicity identification component based on heuristic rules so that we can match non-Asian person names by pinyin to enhance the matching rate.

### 4.3 Statistical Name Transliteration

If a name can be translated using one of the lists described above, name transliteration is not required. Note that the writing systems of both Arabic and Chinese make transliteration non-trivial. Because non-native names in both languages are typically rendered phonetically, we require methods for "back-transliteration" in order to recover likely English renderings: Chinese renders foreign names by stringing together words that approximate a name's constituent sounds (e.g. "Bu Lai Er" is the pinyin representation for "Blair"), while Arabic omits short vowels. Both languages lack and must approximate some English sounds.

We adopted a data-driven approach to address this problem, and trained two foreign-to-English character transliteration models to generate multiple transliteration hypotheses (multiple plausible English spellings of a name from Chinese), using bilingual name lists assembled from several sources, as described above in section 4.2. We pursued two quite different approaches.

In one, we applied state-of-the-art statistical machine translation techniques to the problem, to translate a sequence of characters which form a foreign name to a sequence of characters which form an English name. We created a 6-gram character-based language model (LM) trained from a large list of English names to rank the candidate transliterations. In this approach, no reordering model is used due to the monotonicity property of the task, and model scaling factors are tuned for maximum transliteration accuracy.

In the other, we trained a structured perceptron to emit character edit operations in response to a foreign string, thereby generating a Romanized version. The two approaches achieve comparable accuracy. A detailed description and empirical comparison of these approaches can be found in (Freitag and Khadivi, 2007). Experiments showed that the combination of the two achieved 3.6% and 6% higher accuracy than each alone.

## 5 Name Translation Selection

In general, the output of the above approaches is a set of candidate English names. In order to ensure that the output is an acceptable English name, we *select* the best translation from a large number of candidates using Information Extraction (IE) results and Language Models (LM) built from large English corpora.

### 5.1 Name Selection with English Corpus and IE

To choose amongst these transliterations we first consult a large English corpus from a similar time period and its corresponding Information Extraction results, similar to the techniques described in (Al-Onaizan and Knight, 2002; Kalmar and Blume, 2007). We prefer those name spellings which appear in the corpus and whose time of appearance and global context overlaps the time and context of the document being translated. Possible contexts include co-occurring names, nominal phrases, and document topic labels.

We first process a large English corpus for names, their corresponding titles, the years in which they appear, and the frequencies under different categories, resulting in a large database of person entities in English. A source name's time of appearance and the translation of any co-occurring titles are then compared to the entries in this database for any candidate translations. For GPE name translation, the associated country information of the target name is used for comparison with the mined lists. Weights are optimized to combine the edit distance, temporal proximity and context similarity metrics. The candidate translation ranked highest according to this combined score is then selected. For example, in the following Chinese text:

*… 据国际文传电讯社和伊塔塔斯社报道，〈PER〉**格里戈里·帕斯科**〈/PER〉的 **律师**〈PER〉**詹利·雷兹尼克**〈/PER〉向俄最高法院提 出上诉。*

The name transliteration components generate the following candidates for the name "***詹利·雷兹尼克*** (zhan li . lei zi ni ke)":

| | |
|---|---|
| 24.11 amri | 28.31 reznik |
| 23.09 obry | 26.40 rezek |
| 22.57 zeri | 25.24 linic |

20.82 henri 23.95 riziq
20.00 henry 23.25 ryshich
19.82 genri 22.66 lysenko
19.67 djari 22.58 ryzhenko
19.57 jafri 22.19 linnik

In a large English corpus we find the following sentence: "**Genri Reznik**, Goldovsky's **lawyer**, asked Russian Supreme Court Chairman **Vyacheslav Lebedev**….". By matching the Chinese entity extraction results ("*律师 (lawyer)*" referring to "*詹利·雷兹尼克*") against English IE results ("*lawyer*" referring to "Genri Reznik"), we can select "Genri Reznik" as the correct translation. Without re-ranking with global context, the transliteration component would have produced "Amri Reznik", an incorrect translation.

### 5.2 Name Selection with English LM

In addition, we have built a unigram word-based LM from a large English name list to penalize those transliteration hypotheses which are unlikely to be English names.

For example, for the name "保尔森" in the sentence "***财政部长保尔森****访问中国 (**Paulson, the Treasury Secretary,** visited China)*", the transliteration component produces the following top hypotheses: "*Bauerson*", "*Paulsen*", "*Paulson*" and "*Baulson*". We assign a low score to "*Bauerson*" and "*Baulson*" because they don't exist in the English unigram LM.

Each of the above translation and re-ranking steps produces a scaled confidence score; at the end we produce the final N-Best name translations with token based weighted voting.

## 6 Integration into MT

The source text, annotated with name translations, is then passed to a statistical, phrase-based MT system. To integrate name translation results into MT, two critical decisions have to be taken: the position of the name in the target sentence and the translation of the name. The position of the name is decided by the MT system, and the translation of the name can be performed with or without its context. Therefore, we have two approaches: a simple transfer method, and an MT-derived method.

### 6.1 Simple Transfer based Integration

The first method simply transfers the best translation of the source name to the target side. This approach ensures all name translations appear in the MT output, so it can principally benefit the distillation task. However, it doesn't take into account word re-ordering or the words found in a name's context. For example, the sentence "*<NAME TYPE="PER" TRANSLATION="Jiang Zemin"> 江 泽 民 </NAME> 和 <NAME TYPE="PER" TRANSLATION="Liu Huaqing"> 刘 华 清 </NAME> 会 见 <NAME TYPE="GPE" TRANSLATION="Thailand"> 泰 国 </NAME> 总 理。(Jiang Zemin and Liu Huaqing met with the premier of Thailand.)* " is translated into "*Jiang Zemin and Liu Huaqing met Thailand premier.*" in which the context words such as 'with', 'the' 'of' are missing, and "premier" and "Thailand" are not in English order.

### 6.2 Phrase Table and LM based Integration

Therefore, we adopt the second method for improving full-text MT. The MT system considers the provided list of name-entity translations as a secondary phrase table, and then uses this phrase table as well as its own phrase table and language model to decide which translation to choose when encountering the names in the text. In order to avoid the problem of word segmentation inconsistency, we add all possible segmentations for each name. In order to obtain the best translation BLEU score, the name phrase table receives a coefficient optimized on a development set.

We then apply a 4-gram word based English LM to select the final translation from the name translation or MT phrase tables. For the example in section 5.2, assuming name translation suggests a translation 'Paulson' while the regular MT phrase table produces 'Paulsen', then we can compare the LM scores based on the following phrases in the LM training corpus:

For "*Paulson*":
***Paulson , Jr. to*** *be the 74th Secretary of the Treasury*
***Paulson , the Treasury*** *secretary , is a good guy*
*Treasury* ***Secretary Henry M. Paulson, the former*** *head of Goldman Sachs*
*Listen to US Secretary of the* ***Treasury Henry M. Paulson discuss economic issues***

For "*Paulsen*":
***Paulsen was worried that*** *her vacation in Los Angels*
***Paulsen , chief investment*** *officer at Norwest Investment Management*
***Paulsen , chief investment*** *strategist at Wells Capital Management*
***Paulsen , professor of*** *agronomy at Kansas State University*

| System Type | Y1 Baseline | Y2 Baseline | Y2 1Best-Phrase | Y2 Simple-Transfer | Y2.5 1Best-Phrase | Y2.5 NBest-Phrase | Y2.5 Simple-Transfer |
|---|---|---|---|---|---|---|---|
| PER | **59.63** | 58.28 | 66.89 | 69.59 | 67.91 | 62.84 | **70.44** |
| GPE | 92.24 | 93.18 | 93.25 | 94.12 | 93.49 | 93.18 | 94.27 |
| ORG | 78.15 | 84.37 | 84.71 | 84.54 | 85.88 | 84.54 | 85.88 |
| ALL | **83.0** | 84.54 | 86.50 | 87.44 | 87.07 | 85.56 | **87.98** |

Table 2. Approximate Accuracy of Name Translation (%)

We then can choose "Paulson" as the best transliteration because its context "***Paulson, the Treasury***" matches those in the English LM.

## 7 Experimental Results and Analysis

### 7.1 Approximate Name Translation Accuracy

In this section we present the overall performance of our Chinese to English name translation system.

#### 7.1.1 Data and Scoring Metric

We conduct experiments on the text set of the NIST 2005 MT evaluation. By annotating names in the four reference translations, we identify on average 592 Person names, 1275 GPE names and 595 organization names - 2762 names in total. We apply the RWTH Aachen statistical phrase-based MT system (Zens and Ney, 2004; Zens et al., 2005) as our baseline, which uses the uniform translation model for names and non-names. The following metric is defined to measure the accuracy of name translation:

*Approximate Accuracy =*
*(# Target reference names found in MT Output)/*
*(# Total reference names)*

Using a manually assembled name variant table, we also support the matching of name variants (e.g., "World Health Organization" and "WHO").

#### 7.1.2 Results

Table 2 summarizes the approximate accuracy results for Y1 and Y2 baselines, and the MT system integrated with name translation by two different approaches: simple transfer and name phrase table.

Table 2 shows that the name translation system provided a 29.29% relative error reduction on overall name translation, and a 26.78% relative error reduction on person names. It's also interesting to see that adding N-Best name translation does not provide improvement over 1-Best, which indicates that the name re-ranking approaches described in section 5 are effective to select the best translations. Experiments also show that the 1-Best name phrase approaches can achieve about 0.2%-0.3% improvement on BLEU score over the MT baselines.

## 8 Impact on Cross-lingual Sentence Retrieval

In this section we describe an experiment to measure the impact of name translation on cross-lingual spoken sentence retrieval: given a person name query in English the system should retrieve the sentences containing this name from Mandarin speech.

We used part of the GALE Y2 audio MT development corpus as our candidate sentence set (in total 668 sentences drawn from 45 shows). 53 queries were constructed by selecting person names from the reference translations of reference transcripts for these shows. For each query, relevant sentences from the entire corpus were then manually labeled as answer keys, using reference translations as the basis for selection.

For each query, we then search the documents as translated by Machine Translation and by Name Translation. We compute the matching confidence between the query and a sentence substring based on the edit distance, with each operation assigned unit cost. We then define the sentence-level match confidence as the maximum confidence between the query and any sentence substring starting and ending on a token boundary.

We found that, over a range of confidence thresholds, searching name translation output yields comparable precision with an absolute improvement in recall of about 23% over searching MT output. More details are described in (Ji et al., 2009).

## 9 Conclusions and Future Work

The name extraction and translation system reduces the number of incorrectly translated names by about 30% (from 17% to 12%). People names remain the most difficult: even with name translation, the error rate remains about 30% (decreased from 40%).

We analyzed the sources of these remaining errors for person names. They reveal the following major shortcomings of both name extraction and name translation.

Chinese name extraction errors contribute 5.5% to the person name translation error rate. As noted in (Ji and Grishman, 2006), boundary errors are dominant in Chinese name identification due to word segmentation errors. In addition, without indicative contexts, the name tagger tends to miss some rare names and confuse foreign Person and GPE names. In order to address these problems we will exploit character based HMM and propagate multiple name tagging hypotheses to name translation to increase recall.

Automatic Speech Recognition (ASR) errors bring more challenges to name tagging; we will attempt to develop a phone-aware name tagger, using coreference resolution to correct name ASR errors.

In (Ji and Grishman, 2009) we demonstrated that feedback from name translation can be used to improve name tagging. We plan to extend the approach by incorporating more feedback from name transliteration confidence and mined name pairs as features.

Limitations in the name transliteration models contributed 8.8% to the translation error rate. It may be very challenging for our edit-distance based models to insert consonants at the end of syllables. For example, "Abdelrahman" is mistakenly transliterated into "Abderaman". In the future we intend to use cross-lingual Wikipedia titles to capture more name pairs, especially more Arabic-origin and Russian-origin names.

A 5.8% error rate was due to the mismatch of some famous person names in uncommon spellings. For example, "鲍尔 (Powell)" does not exist in our mined name lists but its more common spelling "鲍威尔" appears with correct translation. Therefore it will be important to harvest name clusters by using within-document and cross-document coreference, so we that we can provide the clusters instead of each individual name to the name translation pipeline as input.

Our novel approach of using English IE to rerank name translation results produces promising results (Kalmar and Blume, 2007). However, it also relies heavily on the characteristics and size of selected English corpora to provide such 'background knowledge'. We found that 4.6% in the error rate can be traced to a lack of corresponding English contexts for the name candidates. In the future, besides exploiting more unstructured corpora, we will also attempt to incorporate structured knowledge such as the social network databases existing on the web.

In the experiments reported here, the underlying statistical MT system treats names just like any other tokens. Ultimately we will explore a new approach of training a name-aware MT system based on co-ordinated name annotation of source and target bitexts and conducting information-driven decoding.

## References


Javier Artiles, Satoshi Sekine and Julio Gonzalo. 2008. Web People Search - Results of the first evaluation and the plan for the second. *Proc. WWW 2008*. Beijing, China.

Daniel M. Bikel, Scott Miller, Richard Schwartz, and Ralph Weischedel. 1997. Nymble: a high-performance Learning Name-finder. *Proc. ANLP1997*. pp. 194-201., Washington, D.C.

Benjamin Farber, Dayne Freitag, Nizar Habash and Owen Rambow. 2008. Improving NER in Arabic Using a Morphological Tagger. *Proc. LREC 2008*.

Dayne Freitag. 2004. Trained named entity recognition using distributional clusters. *Proc. EMNLP 2004*.

Dayne Freitag and Shahram Khadivi. 2007. A Sequence Alignment Model Based on the Averaged Perceptron. *Proc. EMNLP-CONLL 2007*.

Ralph Grishman, David Westbrook and Adam Meyers. 2005. NYU's English ACE 2005 System Description. *Proc. ACE 2005 Evaluation Workshop*. Washington, US.

Heng Ji and Ralph Grishman. 2006. Analysis and Repair of Name Tagger Errors. In Proceedings of COLING/ACL 06, Sydney, Australia.

Heng Ji, Ralph Grishman and Wen Wang. 2008. Phonetic Name Matching For Cross-lingual Spoken Sentence Retrieval. Proc. IEEE-ACL SLT08. Goa, India.

Heng Ji and Ralph Grishman. 2009. Collaborative Entity Extraction and Translation. *Recent Advances in Natural Language Processing*. John Benjamins Publishers (Amsterdam & Philadelphia).

Paul Kalmar and Matthias Blume. 2007. Web Person Disambiguation Via Weighted Similarity of En-

tity Contexts. *Proc. ACL07 workshop on SemEval*. Prague, Czech.

Sadao Kurohashi, Toshihisa Nakamura, Yuji Matsumoto and Makoto Nagao. 1994. Improvements of Japanese Morphological Analyzer JUMAN. *Proc. The International Workshop on Sharable Natural Language Resources*, pp.22-28.

Franz Josef Och. 2003. Minimum Error Rate Training for Statistical Machine Translation. *Proc. ACL 2003*, Japan, Sapporo.

Richard Zens and Hermann Ney. 2004. Improvements in Phrase-Based Statistical Machine Translation. *Proc. HLT-NAACL 2004*, pp. 257-264, Boston, MA.

Richard Zens, Oliver Bender, Sasa Hasan, Shahram Khadivi, Evgeny Matusov, Jia Xu, Yuqi Zhang and Hermann Ney. 2005. The RWTH Phrase-based Statistical Machine Translation System. *Proc. IWSLT 2005*, pp.155-162, Pittsburgh, PA.