Collaborative Entity Extraction and Translation

Heng Ji Ralph Grishman
Department of Computer Science
New York University
New York, NY, 10003, USA
hengji@cs.nyu.edu grishman@cs.nyu.edu

Abstract

Entity extraction is the task of identifying names and nominal phrases ('mentions') in a text and linking coreferring mentions. We propose the use of a new source of data for improving entity extraction: the information gleaned from large bitexts and captured by a statistical, phrase-based machine translation system. We translate the individual mentions and test properties of the translated mentions, as well as comparing the translations of coreferring mentions. The results provide feedback to improve source language entity extraction. Experiments on Chinese and English show that this approach can significantly improve Chinese entity extraction (2.2%-relative improvement in name tagging F-measure, representing a 15.0% error reduction), as well as Chinese to English entity translation (9.1% relative improvement in F-measure), over state-of-the-art entity extraction and machine translation systems.

Keywords
Entity extraction, Bitext, Entity translation, Joint Inference

1. Introduction

Named entity tagging has become an essential component of many NLP systems, such as question answering and information extraction. Building a high-performance name tagger, however, remains a significant challenge. The challenge is greater for languages such as Chinese and Japanese with neither capitalization nor overt tokenization to aid name detection, or Semitic languages such as Arabic that do not exhibit differences in orthographic case.

This challenge is now generally addressed by constructing, by hand, a large name-annotated corpus. Because of the cost of such annotation, several recent studies have sought to augment this approach through the use of un-annotated data, for example by constructing word classes (Miller et al., 2004) or by annotating additional data automatically and selecting the most confident annotations as further training (Ji and Grishman, 2006).

One further source of information for improving name taggers are bitexts – corpora pairing the text to be tagged with its translation into one or more other languages. Such bitexts are becoming increasingly available for many language pairs, and now play a central role in the creation of machine translation and name translation systems. By aligning the texts at the word level, we are able to infer properties of a sequence *s* in language *S* from the properties of the sequence of tokens *d* with which it is aligned in language *D*. For example, knowing that *d* is a name, or merely that it is capitalized (for *D* = English) makes it more likely that *s* is a name. So if we have multiple, closely competing name hypotheses in the source language *S*, we can use the bitext to select the correct analysis.

Huang and Vogel (2002) used these observations to improve the name tagging of a bitext, and the NE (named entity) dictionary learned from the bitext. We wish to take this one step further by using information which can be gleaned from bitexts to improve the tagging of data for which we do not have pre-existing parallel text. We will use a phrase-based statistical machine translation system trained from these bitexts; we will translate the source-language entities using the MT and name translation systems; and then we will use this translation to improve the tagging of the original text.

This approach is an example of joint inference across quite disparate knowledge sources: in this case, combining the knowledge from named entity tagging and translation to produce better results for each. Such symbiosis of analysis components will be essential for the creation of high-performance NLP systems.

The translation knowledge source has an additional benefit: because name variants in *S* may translate into the same form in *D*, translation can also aid in identifying name coreference in *S*.

The rest of this paper is structured as follows. Section 2 briefly introduces the terminology used throughout the paper. Section 3 presents some detailed linguistic intuitions behind the use of bitexts for name tagging. Section 4 introduces our method for using translations, including the use of confidence estimation via cross-lingual cache models. Section 5 presents the detailed inference rules and Section 6 summarizes the system architecture. Section 7 then presents an application of our method to the Chinese to English language pair and its experimental results. Section 8 compares our method to some relevant previous work. Section 9 concludes the paper, pointing to directions for future work.

2. Terminology

We shall use the terminology of ACE[1] to explain our central ideas.

entity: an object or a set of objects in one of the semantic categories of interest, referred to by a set of mentions

mention: a reference to an entity (typically, a noun phrase)

name mention: a reference by name to an entity

nominal mention: a reference by a common noun or noun phrase to an entity

In this paper we consider five types of entities in ACE evaluation: PER (persons), ORG (organizations), GPE ('geo-political entities' – locations which are also political units, such as countries, counties, and cities), LOC (other locations), FAC (facility). *Entity extraction* can then be viewed as a combination of mention detection and classification with coreference analysis, which links coreferring mentions.

3. Motivation for Using Bitexts

We present first our motivation for using word-aligned bitexts to improve source language (*S*) entity extraction. Many languages have special features that can be employed for entity extraction. By using the alignment between the entity extraction results in language *S* and their translations in target language *D*, the language-specific information in *D* will enable the system to perform more accurate extraction than a model built from the monolingual corpus in *S* alone. In the following we present some examples for Chinese-English pair.

- **Chinese → English**

Chinese does not have white space for tokenization or capitalization, features which, for English, can help identify name boundaries and distinguish names from nominals. Using Chinese-English bitexts allows us to capture such indicative information to improve Chinese name tagging. For example,

(a) Results from Chinese name tagger

美德联盟立刻委任了一名执行人员出任 <ENAMEX TYPE="ORG">三菱新 </ENAMEX>总裁。

(b) Bitext

Chinese: *三菱* | *新*

English: ***Mitsubishi*** | ***new***

(c) Name tagging after using bitext

美德联盟立刻委任了一名执行人员出任 <ENAMEX TYPE="ORG">三菱</ENAMEX>新总裁。

Based on the title context word "*总裁(president)*" the Chinese name tagger mistakenly identifies "*Mitsubish new*" as an organization name. But the un-capitalized English translation of "*new*" can provide a useful clue to fix this boundary error.

- **English → Chinese**

On the other hand, Chinese has some useful language-specific properties for entity extraction. For example, standard Chinese family names are generally single characters drawn from a fixed set of 437 family names, and almost all first names include one or two characters. The suffix words (if there are any) of ORG and GPE names belong to relatively distinguishable fixed lists. This feature – particular character or word vocabulary for names – can be exploited as useful 'feedback' for fixing name tagging errors.

(a) Results from English name tagger

*The flashpoint in a week of bitter <ENAMEX TYPE="ORG">**West Bank** </ENAMEX> clashes.*

(b) Bitext

English: ***West Bank***

Chinese: *西岸*

(c) Name tagging after using translation

*The flashpoint in a week of bitter <ENAMEX TYPE="LOC"> **West Bank**</ENAMEX> clashes…*

"Bank" in English can be the suffix word of either a ORG or LOC name, while its Chinese translation "*岸(shore, side)*" indicates that "*West Bank*" is more likely to be a LOC name.

These examples indicate how aligned bitexts can aid entity extraction. However, in most cases the texts from which we wish to extract entities will not be part of such bitexts. We shall instead use a statistical MT system which in effect distills the knowledge in its training bitexts. We will use this MT system to generate entity translations, and then use these translations as we did the bitexts in the examples above.

[1] The Automatic Content Extraction evaluation program of the U. S. Government. The ACE guidelines are at http://www.ldc.upenn.edu/Projects/ACE/

4. General Approach

4.1 Combining Entity Extraction and Translation

We propose a new framework to improve source language *S* entity extraction through the indirect use of bitexts as follows.

We first apply a source language 'baseline' entity extraction system trained from a monolingual corpus to produce entities (*SEntities*), and then translate these entities into target language *D (DEntities)*. The *DEntities* carry information from a machine translation system trained from large bitexts, information which may not have been captured in the monolingual entity extraction. The *DEntities* can be used to provide *cross-lingual feedback* to confirm the results or repair the errors in *SEntities*. This feedback is provided by a set of rules which are applied iteratively.

However, in such a framework we face the problem that the translations produced by the MT system will not always be correct. In this paper we address this problem by using confidence estimation based on voting among translations of coreferring mentions, which we shall refer to as a *mention cache*. In section 4.2 and 4.3 we shall verify the two hypotheses which are required to apply the cache scheme, and in section 4.4 we shall explain the details of these caches.

4.2 One Translation per Named Entity

Named entities may have many variants, for example, "IOC" and "International Olympic Committee" refer to the same entity; and "New York City" alternates with "New York"; but all these different variants tend to preserve 'name heads' – a brief "key" alternation that represent the *naming function* (Carroll, 1985). Unlike common words for which *fluency* and *vitality* are most required during translation, translating a named entity requires preserving its *functional* property – the real-world object that the name is referring to. Inspired by this linguistic property we propose a hypothesis:

- **Hypothesis (1).** *One Translation per Named Entity*: The translation of different name mentions is highly consistent within an entity.

This hypothesis may seem intuitive, but it is important to verify its accuracy. On 50 English documents from ACE 2007 Chinese to English Entity Translation training data with human tagged entities, we measure this hypothesis' *accuracy* by:

$$\frac{|\text{ Coreferred mention pairs with } \textit{consistent} \text{ translations }|}{|\text{Coreferred mention pairs}|}$$

We consider two translations *consistent* if one is a name component or acronym of the other

The *accuracy* of this hypothesis for different name types are: 99.6% for PER, 99.5% for GPE, 99.0% for ORG and 100% for DOC. This clearly indicates that Hypothesis (1) holds with high reliability.

4.3 One Name per Translation

Based on Hypothesis (1), we can select a single 'best (maximal) name translation' for each entity with a name; and this best translation can be used as '*feedback*' to determine whether the extracted name mentions in source language are correct or not. If they are incorrect (if their translations are not consistent with the best translation), they can be replaced by a 'best source language name'. This is justified by:

- **Hypothesis (2).** *One Name per Translation*: Names that have the same translation tend to exhibit *consistent* spellings in the source language.

In reviewing 101 Chinese documents with human translations from ACE07 entity translation training data, the accuracy of this hypothesis for all entity types was close to 100%; the exceptions appeared to be clear translation errors.

Therefore, if we require the name mentions in one entity to achieve consistent translation as well as extraction (name boundary and type), then we can fix within-doc or cross-doc entity-level errors, with small sacrifice of (<1%) exceptional instances.

4.4 Cross-lingual Voted Caches

Given an entity in source language *SEntity* and its translation *DEntity*, let *SName(i)* be a name mention in *SEntity* and have translation *DName(i)*. Then the above two properties indicate that if string *DName(i)* appears frequently in *DEntity*, then *SName(i)* is likely to be correct. On the other hand, if *DName(i)* is infrequent in *DEntity* and conflicts with the most frequent translation in boundary or word morphology, then *SName(i)* is likely to be a wrong extraction.

For a pair of languages *S* (source language) → *D* (target language), we build the following voted cache models in order to get the best *assignment* (extraction or translation candidate) for each entity:

- ***Inside-S-D-Cache***

For each name mention in one entity (inside a single document), record its unique translations and frequencies;

- ***Cross-S-D cache***

Corpus-wide (across documents), for each name and its consistent variants, record its unique translations and their frequencies;

- ***Cross-D-S cache***

Corpus-wide, for each consistent translation in *D*, record its corresponding names in *S* and their frequencies.

The caches incorporate simple filters based on properties of language *D* to exclude translations which are not likely to be names. For *D*=English, we exclude empty translations, translations which are single un-capitalized tokens, and, for person names, translations with any un-capitalized tokens. In addition, in counting translations in the cache, we group together consistent translations. For English, this includes combining person name translations if one is a subsequence of the tokens in the other. The goal of these simple heuristics is to take advantage of the general properties of language *D* in order to increase the likelihood that the most frequent entry in the cache is indeed the best translation.

For each entry in these caches, we get the frequency of each unique *assignment*, and then use the following *margin* measurement to compute the confidence of the best assignment:

Margin = Frequency (Best Assignment) – Frequency (Second Best Assignment)

A large margin indicates greater confidence in the assignment.

5. Inference Rules

We can combine the language-specific information in *SEntity*, and its entry in the cross-lingual caches to detect potential extraction errors and take corresponding corrective measures. We construct the following inference rules and an example for some particular rules below.

Based on hypothesis (1) and (2), for a test corpus we aim to achieve a group of entities in both source and target languages, with high consistency on the following levels:

Rule(1): Adjust Source Language Annotations to Achieve Mention-level Consistency:

Rule (1-1): Adjust Mention Identification

If a mention receives translation that has small margin as defined in section 4.4 and violates the linguistic constraints in target language, then do not classify the mention as a name.

Rule (1-2): Adjust Isolated Mention Boundary

Adjust the boundary of each mention in *SEntity* to be consistent with the mention receiving the best translation.

Rule (1-3): Adjust Adjacent Mention Boundary

If two adjacent mentions receive the same translation with high confidence, merge them into one single mention.

Rule (2): Adjust Source Language Annotations to Achieve Entity-level Consistency:

If one entity is translated into two groups of different mentions, split it into two entities.

Rule (3): Adjust Target Language Annotations to Achieve Mention-level Consistency:

Enforce entity-level translation consistency by propagating the high-confidence best translation through coreferred mentions.

For example, for the following Chinese document,

<TEXT>

<sent 1>加拿大第３７届联邦议会２９日举行会议，选举自由党议员***米利肯***为众议院新议长。</sent>
The 37th Canadian Federal Parliament held a meeting on the 29th and elected Liberal MP ***Miliken*** as House of Commons speaker.

<sent2>今年５４岁的***彼得.米利肯***是来自加拿大安大略省金斯敦地区的议员。</sent>
The 54-year-old ***Peter Miliken*** is a MP from Kingston, Ontario, Canada.

<sent3>***米利肯***是在５轮投票后当选的。</sent>
Miliken was elected after five rounds of voting.

</TEXT>

The baseline system extracts and translates the following entity:

{米利/Mili, 彼得.米利肯/ Peter Miliken, 米利肯/Miliken}

By applying rule (1), we can fix the boundary of the first name mention "*米利*" into "*米利肯*" because "*米利肯*" has the (maximal) best translation "*Miliken*":

{米利肯/Mili,彼得.米利肯/Peter Miliken, 米利肯/Miliken}

then by applying rule (5) we can change the translation "*Mili*" into the more frequent translation "*Miliken*":

{米利肯/Miliken,彼得.米利肯/Peter Miliken, 米利肯/Miliken}

These inference are formalized in Table 1. They are applied repeatedly until there are no further changes; improved translation in one iteration can lead to improved *S* entity extraction in a subsequent iteration.

6. System Pipeline

The overall system pipeline for language pair (*S, D*) is summarized in Figure 1.

Terms	
DConstraint	Some constraint that name entities must satisfy in language *D*. For example, in the setting of *S*=Chinese and *D*=English, it includes the capitalization constraint.
CorefMentionNum(i)	the number of name mentions coreferring to *SName(i)* in *SEntity*
BestDName(Cache)	the best (most frequent) translation in *Cache*
FreBestDName(Cache)	the frequency of the best (most frequent) translation in *Cache*
FreSeBestDName(Cache)	the frequency of the second best (most frequent) translation in *Cache*
Margin(i, Cache)	the *margin* (defined in section 4.4) of name *SName(i)* in *Cache*
Predicates	
ViolateDConstraint(i)	*DName(i)* does not satisfy *DConstraint*
HasBestTran(j, Cache)	*SName(j)* has translation *BestDName(Cache)* in *Cache*
ConflictBoundary(i, j)	*SName(i)* is consistent with *SName(j)* at one boundary but not the other
HasFewCorefMentions(i)	*CorefMentionNum(i)* $< \delta_1$
HasLowConf(i, Cache)	*Margin(i, Cache)* $< \delta_2$
ShareTranslation(i, j)	*DName(i) = DName(j)*
Adjacent(i, j)	*SName(i)* and *SName(j)* are adjacent to each other
EqualConf(SEntity)	*FreBestDName(Inside-S-D-Cache)* $> \delta_3 \wedge$ *FreSeBestDName(Inside-S-D-Cache)* $> \delta_4$
Overlap(i, j)	*SName(i) and SName(j)* overlap in spelling
Rule (1-1): Adjust Mention Identification	
if *(ViolateDConstraint(i)* $\wedge$ *HasFewCorefMentions(i)* $\vee$ *HasLowConf(i, Cross-D-S-Cache)* then Change *SName(i)* into nominal or delete it	
Rule (1-2): Adjust Isolated Mention Boundary	
for all j ≠ i do if *(ViolateDConstraint(i)* $\wedge$ *HasBestTran(j, Inside-S-D-Cache)* $\wedge$ *ConflictBoundary(i, j))* $\vee$ *(HasBestTran(j, Cross-D-S-Cache)* $\wedge$ *ConflictBoundary(i, j))* then Replace *SName(i)* with *SName(j)* or split it into *SName(j)* and another mention	
Rule (1-3): Adjust Adjacent Mention Boundary	
for all j ≠ i do if *ShareTranslation(i, j)* $\wedge$ *Adjacent(i, j)* then Merge *SName(i)* and *SName(j)* into a single mention	
Rule (2): Adjust Entity-level Consistent Source Language Annotation (Coreference Resolution)	
if *EqualConf(SEntity)* $\wedge$ $\neg$*Overlap(i, j)* then Split *SEntity* into two entities	
Rule (3): Adjust Mention-level Consistent Target Language Annotation (Mention Translation)	
if $\neg$*HasLowConf(i, Inside-S-D-Cache)* then Replace *DName(i)* with *BestDName(Inside-S-D-Cache)* if $\neg$*HasLowConf(i, Cross-S-D-Cache)* then Replace *DName(i)* with *BestDName(Cross-S-D-Cache)*	

Table 1. Inference Rules of Using Translation to Improve *SEntity* Extraction

Type	Baseline	After Using Inference Rules
PER	89.9%	91.2%
GPE	87.0%	86.9%
ORG	85.7%	88.5%
LOC	89.7%	90.6%
FAC	80.9%	85.3%
ALL	87.3%	89.2%

Table 2. F-Measure (%) of Name Tagging

Type	Baseline	After Using Inference Rules
PER	34.8%	36.7%
GPE	44.7%	49.8%
ORG	37.0%	39.9%
LOC	18.3%	18.1%
FAC	23.1%	23.3%
ALL	35.1%	38.3%

Table 3. F-Measure (%) of Entity Translation

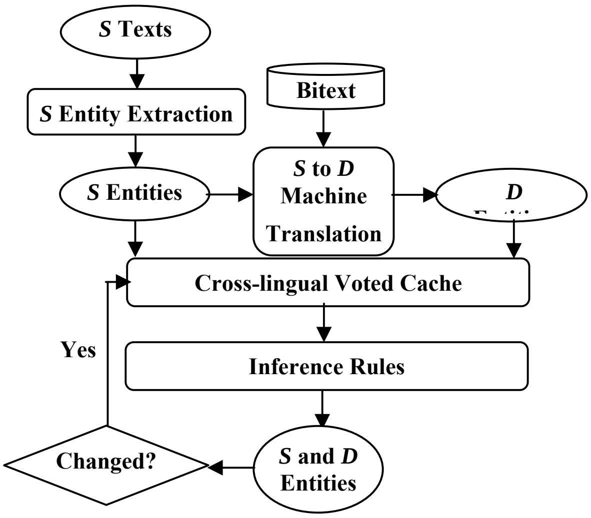

Figure 1. A Symbiotic Framework of Entity Extraction and Translation

7. Experiments on Chinese to English

In this section we shall present an example of applying this method using Chinese-to-English translation to improve Chinese entity extraction.

7.1 Baseline Systems

We used a Chinese entity extraction system described in (Ji et al., 2005) and a statistical, phrase-based machine translation system (Zens and Ney, 2004) for our experiments.

7.2 Rule Restriction

We tested the rules on a development set, and added a few source-language-specific restrictions on their applicability to improve performance. Also, where the rules allowed for two alternative corrections, we added a language-specific criterion for choosing the correction.[2]

7.3 Data

We took the Chinese newswire data from the ACE 2007 Entity Translation training and evaluation corpus as our blind test set, and evaluated our system. The test set includes 67 news texts, with 2077 name mentions and 1907 entities.

7.4 Improvement in Entity Extraction

Tables 2 shows the name tagging performance on different entity types. Except for the small loss for GPE names, our method achieved positive corrections on most entity types. Significant improvements were achieved on ORG and FAC names for all three language sources, mainly because organization and facility names in English texts have less boundary ambiguity than in Chinese texts. So they are better aligned in bitexts and easier to translate. The small loss in GPE names for the Chinese source is due to the poor quality of the translation of country name abbreviations.

The rules can also improve nominal tagging by disambiguating mention types (name vs. nominal), and improve coreference by merging or splitting incorrect entity structures. All of these improvements benefit entity extraction.

7.5 Improvement in Entity Translation

A further benefit of our system is a boost in the translation quality of Chinese entities. We used the official ACE 2007-ET scorer[3] to measure the F-scores. The performance for translating different entity types is presented in Table 3. The inference based on voting over mentions of an entity particularly improved GPE name abbreviation translation and fixed translated person foreign name boundaries. Thus we have succeeded in using the interaction of entity extraction and translation to improve the performance of both.

7.6 Error Analysis

The errors reveal both the shortcomings of the MT system and consistent difficulties across languages.

For a name not seen in training bitexts the MT system tends to mistakenly align part of the name with an un-capitalized token. Also, there are words where the ambiguity between name and nominal exists in both Chinese and English, such as "国会–parliament". Rule (2) fails in these cases by mistakenly changing correct names into nominal mentions. In these and other cases, we could apply a separate name transliteration system developed from larger name-specific bitexts to re-translate these difficult names. Or we could incorporate the confidence values such as (Ueffing and Ney, 2005) generated from the MT system into our cross-lingual cache model. Nevertheless, as Table 2 and 3 indicate, the rewards of using the bitext/translation information outweigh the risks.

8. Related Work

The work described here complements the research described by (Huang and Vogel, 2002). They presented an

[2] Specifically: for Rule (1) we added a check that *SName(i)* and *SName(j)* are not a name and its acronym. Also for Rule (1), if *SName(i)* includes a conjunction the rule splits the name into two names, otherwise replacing it by *SName(j)*. For Rule (2), since in Chinese most ambiguities between name and nominal arise in GPE or ORG names, GPE or ORG names are corrected into nominals, while deleted PER names are deleted. Rule (3) was limited to merging mentions of selected entity type pairs, such as "PER-GPE" and "ORG-LOC" because they are unlikely appear adjacent in Chinese.

[3] The description of the ACE entity translation metric can be found at http://www.nist.gov/speech/tests/ace/ace07/doc/ ET07-evalplan-v1.6.pdf

effective integrated approach that can improve the extracted named entity translation dictionary and the entity annotation in bilingual training corpus. We expand their idea of alignment consistency to the task of entity extraction in a *monolingual test* corpus. Unlike their approach requiring reference translations in order to achieve highest alignment probability, we only need the source language unlabeled document.. So our approach is more broadly applicable and also can be extended to additional information extraction tasks (nominal tagging and coreference).

Aligned bitexts have also been used to project name tags from French to English by Riloff et al. (2002) and from Japanese to English by Sudo et al.(2004), but their approaches only use the entity information from the source language.

Bitexts have been used for several other NLP tasks besides MT. Bannard and Callison-Burch (2005) have applied bilingual corpus to obtain paraphrases. Ng et al. (2003) have sought to acquire sense-tagged training data from English-Chinese bitexts, and used them for word sense disambiguation (WSD). These methods are based entirely on the bitext correspondences, while we combined evidence from the source language with bitexts.

In addition, our approach represents a cross-lingual joint inference example, which complements the joint inference in the monolingual analysis pipeline as described in (Ji and Grishman, 2005) and (Roth and Yi, 2004).

9. Conclusion and Future Work

Bitexts can provide a valuable additional source of information for improving named entity tagging. We have demonstrated how the information from bitexts, as captured by a phrase-based statistical machine translation system, and then used to generate translations, can be used to correct errors made by a source-language named-entity tagger. While our approach has only been tested on Chinese and English so far, we can expect that it is applicable to other language pairs. The approach is independent of the baseline tagging/extraction system, and so can be used to improve systems with varied learning schemes or rules.

There are a number of natural extensions and generalizations of the current approach. In place of correction rules, we could adopt a joint inference approach based on generating alternative source language name tags (with probabilities), estimating the probabilities of the corresponding target language features, and seeking an optimal tag assignment. Although the current approach only relies on limited target language features, we could use a full target-language entity extractor (as Huang and Vogel (2002) did), providing more information as feedback (for example, name type information). Furthermore, we intend to pass the name tagging hypotheses to a name transliteration system and use the transliteration results as additional feedback in assessing name hypotheses.

10. Acknowledgements

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR0011-06-C-0023. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Defense Advanced Research Projects Agency (DARPA).

11. References

[1] Colin Bannard and Chris Callison-Burch. 2005. Paraphrasing with Bilingual Parallel Corpora. In ACL 2005:597-604. Ann Arbor, US.

[2] John M. Carroll. 1985. *What's in a Name?: An Essay in the Psychology of Reference.* New York, US.

[3] Fei Huang and Stephan Vogel. 2002. Improved Named Entity Translation and Bilingual Named Entity Extraction. *In ICMI 2002:253-258*. Pittsburgh, PA, US.

[4] Heng Ji and Ralph Grishman. 2005. Improving Name Tagging by Reference Resolution and Relation Detection. *Proc. ACL2005.* pp. 411-418. Ann Arbor, USA.

[5] Heng Ji and Ralph Grishman. 2006. Data Selection in Semi-supervised Learning for Name Tagging. *In ACL 2006 Workshop on Information Extraction Beyond the Document:48-55*. Sydney, Australia.

[6] Heng Ji, Adam Meyers and Ralph Grishman. 2005. NYU's Chinese ACE 2005 EDR System Description. ACE 2005 PI Workshop. Washington, US.

[7] Scott Miller, Jethran Guinness and Alex Zamanian. 2004. Name Tagging with Word Clusters and Discriminative Training. *In HLT/NAACL 2004: 337-342.* Boston, Massachusetts, US.

[8] Hwee Tou Ng, Bin Wang, and Yee Seng Chan. Exploiting Parallel Texts for Word Sense Disambiguation: An Empirical Study. *In ACL 2003: 455-462.* Sapporo, Japan.

[9] Dan Roth and Wen-tau Yih. 2004. A Linear Programming Formulation for Global Inference in Natural Language Tasks. *Proc. CONLL2004.* pp. 1-8

[10] Ellen Riloff, Charles Schafer, and David Yarowsky. 2002. Inducing Information Extraction Systems for New Languages via Cross-Language Projection. *In COLING 2002:828-834.* Taipei, Taiwan.

[11] Kiyoshi Sudo, Satoshi Sekine and Ralph Grishman. 2004. Cross-lingual Information Extraction System Evaluation. *In COLING 2004:882-888.* Geneva, Switzerland.

[12] Nicola Ueffing and Hermann Ney. 2005. Word-Level Confidence Estimation for Machine Translation using Phrase-Based Translation Models. *In HLT/EMNLP 2005:763-770.* Vancouver, Canada.

[13] Richard Zens and Hermann Ney. 2004. Improvements in Phrase-Based Statistical Machine Translation. In HLT/NAACL 2004. New York City, NY, US