

# Linguistic Resources for Entity Linking Evaluation: from Monolingual to Cross-lingual

Xuansong Li, Stephanie M. Strassel, Heng Ji, Kira Griffitt, Joe Ellis

Linguistic Data Consortium, University of Pennsylvania  
Philadelphia, PA 19104 USA

Computer Science Department and Linguistics Department,  
Queens College and Graduate Center, City University of New York  
New York, NY 10016 USA

Email: {xuansong, strassel, kiragrif, joellis}@ldc.upenn.edu hengjicuny@gmail.com

## Abstract

To advance information extraction and question answering technologies toward a more realistic path, the U.S. NIST (National Institute of Standards and Technology) initiated the KBP (Knowledge Base Population) task as one of the TAC (Text Analysis Conference) evaluation tracks. It aims to encourage research in automatic information extraction of named entities from unstructured texts with the ultimate goal of integrating such information into a structured Knowledge Base. The KBP track consists of two types of evaluation: Named Entity Linking (NEL) and Slot Filling. This paper describes the linguistic resource creation efforts at the Linguistic Data Consortium (LDC) in support of Named Entity Linking evaluation of KBP, focusing on annotation methodologies, process, and features of corpora from 2009 to 2011, with a highlighted analysis of the cross-lingual NEL data. Progressing from monolingual to cross-lingual Entity Linking technologies, the 2011 cross-lingual NEL evaluation targeted multilingual capabilities. Annotation accuracy is presented in comparison with system performance, with promising results from cross-lingual entity linking systems.

**Keywords:** information extraction; named entity linking; Knowledge Base

## 1. Introduction

As one of the three tracks of the TAC (Text Analysis Conference) organized by NIST, the KBP track aims to encourage research in automatic information extraction of named entities from unstructured texts with the ultimate goal of integrating such information into a Knowledge Base (KB). Two tasks are included in this track: Name Entity Linking (NEL) and Slot-Filling (SF). NEL is an independent task or a pre-processing step for the SF task. As a practical extension of named entity recognition (McNamee et al., 2010), it requires systems to effectively retrieve and determine which entity the name refers to in a pre-defined KB. The task started in 2009 using only English texts, and expanded in 2011 to include Chinese documents as well, progressing from monolingual to cross-lingual NEL technologies. Namely, given a set of English and Chinese queries, systems should cluster the queries and link each cluster to the corresponding English KB entry if the linkage exists. Otherwise, a unique NIL-ID should be assigned to the cluster.

For NEL systems, entity disambiguation and clustering are major challenges that are addressed using two different approaches: unsupervised versus supervised learning. In unsupervised or weakly-supervised learning, annotated data is minimally used to tune thresholds and parameters, and the similarity measure is largely based on unlabeled contexts. In supervised learning, pairs of entities and KB nodes are modeled as instances for classification or ranking. Such a classifier or a ranker can be learned from the annotated training data based on various features. Supervised learning usually has better

performance than unsupervised learning (Ji et al, 2010). LDC has developed linguistic resources to facilitate NEL system training and evaluation. This paper describes how such resources are created. The rest of the paper is structured as follows: Section 2 introduces source and training corpora; Section 3 details methodologies; Section 4 elaborates on annotation process; Section 5 presents annotation accuracy; and Section 6 concludes the paper.

## 2. Linguistic Data

### 2.1 Source Data

The source corpora for NEL tasks include source document corpora and a KB source corpus. LDC distributes source document and KB source corpora to NEL teams for system development. The corpora are also used at LDC for training and evaluation data production.

The document corpus for the monolingual English NEL task is a subset selected from LDC's existing English-language collections of newswire articles as well as some web data and audio transcripts, resulting in a total of 1.7 million documents. An extra collection of 1 million Chinese newswire documents was added for the cross-lingual entity linking (CLEL) task. Document selection particularly focused on data concurrent with the epoch of the KB source for an intensive coverage of specific entities (Simpson et al., 2010). The selection focus for the English task is the ACE (Automatic Content Extraction) data epoch plus 2007-2008 data. For the cross-lingual task, we selected newswire stories from Xinhua News Agency, Agence France Presse, and People's Daily Online, also focusing on 2007-2008 Chinese sources.

The KB source data is in English and was used for both English and CLEL tasks. The corpus has about 800,000 KB entries, each with a canonical name and title for a Wikipedia page, and each corresponding to a unique entity of one of the four types: person (PER), organization (ORG), geo-political (GPE), or unknown (UKN). Entries were derived from pages in the October 2008 Wikipedia snapshots that contain semi-structured ‘infoboxes’ or tables of attributes of subjects. Some of the snapshot pages were removed due to ill-formatted infoboxes.

It is a challenge for handling and annotating such corpora due to the large size (1.7 million for the English task, 2.7 million for cross-lingual task, and approximately 800K of KB data). Data pre-processing can also be challenging for systems, such as sentence segmentation, KB indexing, extremely long files, and very noisy web data.

## 2.2 Training and Evaluation Data

LDC creates human annotations for supervised NEL system training and evaluation. The following table displays NEL human annotation corpora up to date. Two types of NEL data are included: training and evaluation. Evaluation data of the previous years is usually used as training data for a new evaluation year. Most corpora include NE linking and NIL co-reference annotation except for the 2010 English training data. The data is available to KBP participants under an agreement license.

and GPE entities with high levels of ambiguity or confusability and diversity, well balanced between training and evaluation sets for each year and across years.

## 3.1 Queries

A query used for the NEL task has three parameters: QueryID, a name string corresponding to a name mention of a PER, ORG, or GPE entity, and a reference document containing that mention. The purpose of the reference document is to provide context for disambiguating a name-string in the Knowledge Base. Reference document is especially important if the name string is ambiguous without context and refers to multiple entities. Queries are formatted with xml tags, as shows in the following examples. Chinese name strings are added for the cross-lingual task.

```
<query id="EL_CLCMN_02111">
  <name>Abbas Moussawi</name>
  <docid>LTW_ENG_19960311.0047</docid>
</query>
<query id="EL_CLCMN_02173">
  <name>丁一汇</name>
  <docid>XIN_CMN_20030323.0033</docid>
</query>
<query id="EL_CLCMN_02174">
  <name>Yihui Ding</name>
  <docid>XIN_ENG_20030327.0034</docid>
</query>
```

Corpus Title (Dataset)	Type	LDC Catalog	Language	Size (Queries)
TAC 2009 KBP Gold Standard Entity Linking Entity Type List	NEL Evaluation	LDC2009E86	English	567 GPE
				627 PER
				2710 ORG
TAC 2010 KBP Evaluation Entity Linking Gold Standard	NEL Evaluation	LDC2010E82	English	749 GPE
				741 PER
				750 ORG
TAC 2010 KBP Training Entity Linking	NEL Training	LDC2010E31	English	500 GPE
				500 PER
				500 ORG
TAC 2011 KBP Cross-lingual Training Entity Linking	NEL Training	LDC2011E55	Chinese English	685 GPE
				817 PER
				660 ORG
TAC 2011 KBP English Evaluation Entity Linking Annotation v1.1	NEL Evaluation	LDC2011R36	English	750 GPE
				750 PER
				750 ORG
TAC 2011 KBP Cross-lingual Evaluation Entity Linking Annotation V1.1	NEL Evaluation	LDC2011R38	Chinese English	642 GPE
				824 PER
				710 ORG

Table 1: TAC KBP Training and Evaluation Data for Entity Linking Task

## 3. Annotation Methodologies

The NEL task requires the computation system to correctly link a named entity to an entry in the KB, or correctly report if it does not have a matching entry. Training data to this end should be sets of queries corresponding to various name mentions of PER, ORG,

## 3.2 Quality Standard for Entity Names

Name strings are ambiguous if they refer to different named entities. Entities are confusable if they are referred to by a diversity of names. One important goal for the information extraction systems is to disambiguate name

strings and resolve mentions to an unambiguous node in a Knowledge Base. Therefore ambiguity and diversity are quality standards for name string selection in the creation of queries.

### 3.2.1 Ambiguity

An entity name can be ambiguous within the same entity type or across types. For instance, given a GPE type, the name “Elizabeth” is ambiguous as it may refer to “Elizabeth, New Jersey” or “Elizabeth, Indiana”. It is also ambiguous across entity types when it refers to “Elizabeth, Queen of England.” A common name is usually more ambiguous than an uncommon one. A single first or last name is more ambiguous than a full name. Names for sports teams of ORGs are highly ambiguous, particularly when they have across-type features (like “Chicago” for “sports team” instead of “the city”).

Given an ambiguous name-string (“seed names” in Table 2), if annotators can find four or five unique entities, they will try to evenly balance them with a total limit of 20 occurrences. This process results in ambiguity clusters.

were typical variance types for GPE names. Abbreviations and acronyms were more common for ORG names while nicknames and honorary names were mostly used as PER variants.

<i>Name Variance</i>	<i>Examples</i>
Alias	“Mark Twain” for “Samuel Clemens”; “鲁迅” for “周树人”
Nicknames	“Alex” or “Alec” for “Alexander”
Abbreviations	“Univ of DE” for “University of Delaware”; “北大” for “北京大学”
Acronyms	“WHO” for “World Health Organization”
Historical	“Beiping” for “Beijing”
Honorary	“Queen of England” for “Elizabeth”
Metaphorical	“Garden State” for “New Jersey”
Word reorder	“Wang Fang” and “Fang Wang”
Misspelling	“Los Angeles” for “Los Angeles”; “复单大学” for “复旦大学”
Deletion	“John Adams” and “John Quincy Adams”

Table 3: Name Variance Types

<b>Corpus Title (Dataset)</b>	<b>Seed Names (Unique Namestrings)</b>	<b>Ambiguity Clusters</b>	<b>Diversity Clusters</b>	<b>Similarity Clusters</b>
TAC 2009 KBP Gold Standard Entity Linking Entity Type List	529	100 (18%)	61 (11%)	341
TAC 2010 KBP Evaluation Entity Linking Gold Standard	756	96 (12%)	6 (1%)	325
TAC 2010 KBP Training Entity Linking	648	43 (6%)	56 (8%)	252
TAC 2011 KBP Cross-lingual Training Entity Linking	562	127(22%)	40 (7%)	245
TAC 2011 KBP English Evaluation Entity Linking Annotation v1.1	1343	175 (13%)	26 (2%)	264
TAC 2011 KBP Cross-lingual Evaluation Entity Linking Annotation V1.1	799	138 (17%)	43 (7%)	282

Table 2: Query Clusters

### 3.2.2 Diversity

Queries are confusable or ambiguous when a diversity of names refers to the same entity. This requires systems to retrieve all possible semantically relevant names for ranking despite variance in name forms. Annotators tried to hunt for name variants to create diversity clusters (Table 2) by searching Wikipedia texts, KB texts and other internet resources. The 2009 evaluation dataset has the most diversity clusters because a GUI tool was particularly developed for searching this kind of cluster.

Table 3 shows some examples of different types of diversity/variance of name strings. During the annotation process, we found that metaphorical and historical names

“Ambiguity clusters” and “diversity clusters” in Table 2 exhibit the total numbers of these two types of query clusters for each dataset. “Seed names” represent unique names. Comparing the seed names in Table 2, we can see the 2011 CLEL evaluation dataset has the most ambiguous clusters (about 22%). Annotators had a hard time catching the rare Chinese GPE clusters because, different from English GPE names, Chinese GPE names are mostly unique and specific.

### 3.3. Similarity Standard for Reference Documents

The “similarity” standard concerns the reference document of a query, which supports supervised system

training in disambiguating entities by providing proper contexts. Systems should adequately recognize a target entity from varying context. Reference documents resemble KB entity texts with various degrees of similarity. We distinguish three degrees of similarity: single-mention, surrounding-context, and whole-text context. With “single-mention” cases, the entity names appear by themselves in the document with no surrounding or remote context as clues for entity disambiguation. “Surrounding-context” clues are close to entity name mentions in documents, usually within the range of the same sentence to two or three sentences. In “whole-text context” documents, entity disambiguation clues appear elsewhere other than in the same sentence or the paragraph where the entity name is first mentioned. A document with multiple mentions of the entity in various locations of the document is also considered as “whole-text context”. The more contexts overlap with the KB text, the higher the similarity. Annotators balance the proportions among “surrounding” and “whole-text” contexts, avoiding “single-mention” documents.

The surrounding-context standard was emphasized in the 2011 cross-lingual task to minimize the impact of current machine translation quality. Rather than translating the whole text, systems learn essential features by catching significant disambiguating information around an entity mention. About 70%-80% of the reference documents in 2011 Chinese queries provided this type of contextual information. The “similarity clusters” in Table 2 reflect clusters of queries of the same entity in the same name, with reference documents of varying contexts.

### 3.4 Quantity Standard

Consistent profiles of training queries and test queries with well-balanced proportion distributions facilitate systems in estimating some parameters in advance. Several quantity standards for query selection are therefore imposed, including entity type, KB link, genre, and language type distributions.

#### 3.4.1 Entity Type Distribution

We basically follow the 1/3 quantity standard for selecting the three types of entity (GPE, PER, ORG). Table 4 indicates type distributions in each dataset, where we see a good representation of the 1/3 quantity standard for entity types in all datasets except for the 2009 evaluation data where ORG type was highly skewed. No hard limit was imposed for the first year NEL task, which helps to explain the disproportion. This disproportion also reveals that it is much easier to select ambiguous ORG queries than selecting other types. We spent more selection time later in 2010 and 2011 to strictly execute this 1/3 quantity standard.

#### 3.4.2 NILs versus Non-NILs

Queries are non-NILs when they are linked to KB entries and NILs when they do not match any KB entry. As a pre-requisite step for the KBP Slot Filling task, systems

are required to properly cluster queries, co-referring not only KB-entry entities but NIL queries as well. To this end, the 1/2 (slightly skewed towards NILs) quantity standard was proposed for non-NIL versus NIL query selection. Numbers in Table 4 indicates a uniform distribution following this standard among all datasets and across all entity types except for a more obvious discrepancy in the 2009 evaluation and 2010 English training datasets.

Dataset	KB link	GPE	PER	ORG	Total
2009 Eng Eval	non-NIL	407	255	1013	1675
	NIL	160	372	1697	2229
2010 Eng Eval	non-NIL	503	213	304	1020
	NIL	246	538	446	1230
2010 Eng Training	non-NIL	404	335	335	1074
	NIL	96	165	165	426
2011 Eng Eval	non-NIL	521	265	338	1124
	NIL	229	485	412	1126
2011 CLEL Training	non-NIL	419	331	251	1001
	NIL	266	486	409	1161
2011 CLEL Eval	non-NIL	416	379	290	1085
	NIL	226	445	420	1091

Table 4: Query Proportion Distribution

#### 3.4.3 Genre Source Distribution

The NEL task targets both newswire domain as well as other genres, such as weblog or broadcast. Newswire article are structured, newsworthy, and full of real-life entities with supporting context while other genres are less structured with sparse supporting context about entities. This requires systems to have the capability to handle both structured and unstructured information. The genre source distribution required for the 2011 monolingual English task is that 2/3 query names should be from newswire source and 1/3 names from other genres. The genres used for the 2011 cross-lingual task were Chinese newswires, English newswire and web/broadcast, with no specific genre proportion requirement for query selection.

#### 3.4.4 Language Source Distribution

The language source distribution standard was uniquely designed for the 2011 CLEL task. The plethora of multilingual information on the internet attracts the attention of the information retrieval (IR) community, encouraging research on methodologies for cross-lingual IR technology. The CLEL task targeted this capability for the KBP systems and has designed the language source distribution standard. Table 5 presents proportions of training and evaluation datasets (“cmn” stands for entities represented only by Chinese queries; “eng” for entities

only by English queries; “both” for entities by both Chinese and English queries). We found that most well-known GPE names usually had KB entries. The majority of GPE names without KB matches were either from Chinese or English documents, but rarely both.

Datasets	cmn	eng	both	total
2011 cross-lingual train	1399	461	302	2162
2011 cross-lingual eval	1408	548	220	2176

Table 5: Language Source Distribution

#### 4. Process for Creating NEL Corpora

Following the aforementioned query selection standards, annotators select a pool of confusable queries/entities for further entity linking or co-reference annotation to create training data. The entire process involves three steps: name seeding, name expansion, and NEL/co-reference annotation.

##### 4.1 Entity Name Seeding

Query creation begins with the entity name selection by targeting name seeds. Seed names are potentially ambiguous names or names with potential variants. Seeds for the 2009 NEL task are from the ACE name pool (Doddington et al., 2004). For 2010 and 2011 tasks, we utilized the output of a bilingual (English and Chinese) named entity tagger (Ji & Grishman, 2006). The output profiles include 309,094 Chinese and 3,296,265 English entity names. Some other information also comes with the tagger output, such as document IDs, frequencies in English source corpus as well as in English KB source. The Chinese name profile has four more fields: Chinese document hit, English name translations, and Chinese KB hits. Such information, coupled with internet search, greatly helped annotators in properly judging the confusability of a name.

From the tagger name pool, annotators identify confusable names as seeds based on the quality and quantity standards (section 3). LDC developed a tool to support this seeding process. Names with the KB hit between 1-7 are reserved for selecting names with more specific features for the slot-filling task while entity linking name seeds fall into zero or >7 KB hits. Annotators were instructed to select name strings with higher document and English KB hits and to avoid fictional as well as non-individual PER entities. “Seed names” in Table 2 show seeding results for each dataset.

The seeding process is more complicated with the cross-lingual task due to the extra language distribution requirement (Section 3.4.4). Before seeding, we re-pooled the tagger name pool into several sub-pools. The Chinese names were first partitioned into NIL and non-NIL pools according to English KB hits, both of which were further divided into pools of English-only, Chinese-only and English-Chinese based on Chinese/English document hits. The sub-pooling process yielded a large pool of

Chinese-only and English-only names. However, the Chinese-English name pool was very small, indicating a sparsity of overlapping entities both in Chinese and English source documents, especially of the NIL type. To assure selection quality, native English annotators are assigned with English-only names, native Chinese annotators with Chinese-only names, and fluent bi-lingual speakers with Chinese-English overlapping names.

##### 4.2 Query Creation and Expansion

The potentially ambiguous seed names from the seeding stage turn into concrete ambiguous queries in the query expansion stage. Given a seed name, annotators look for documents referring to different entities. When several entity clusters can be found for the same name, annotators balance number of queries among these clusters within the limit of 20 queries in total. In the case of only one entity for a given name, annotators can still select up to 20 queries by adding reference documents with different degrees of context support. Redirection and disambiguation links of Wikipedia pages are good resources for catching ambiguity and diversity features. As a result, with this expansion process, seed names grow into a query set consisting of entity clusters. The expansion was performed via GUI interfaces developed at LDC, targeting the three types of query clusters: ambiguity, diversity and similarity clusters (Table 2).

The challenge for the CLEL task at this stage is to select overlapping entities appearing in both Chinese and English documents. We implement an ad hoc process to reach such entities. Bilingual annotators first located ambiguous names from the Chinese-English tagger output pool to formulate Chinese queries. The Chinese source names were subsequently translated into English names that further served as seed names for counterpart English queries of the same entities.

##### 4.3 Entity Linking and NIL-coreference Annotation

If the seeding stage produces potential ambiguous queries, and the query expansion stage creates potential clusters of queries, then the last stage is to realize the true clusters via NEL and NIL-coreference annotation. The annotation first starts with KB linking, where confusable queries from the expansion stage are fed into an entity linking tool (see Figure 1 on the next page) developed at LDC. Annotators searched the KB corpus via the GUI interface, linking a query to a wiki entry if a match is returned. The annotation yielded two types of queries: those with KB matches (non-NIL) and those without KB match (NIL).

With KB linking annotation, the non-NIL queries were clustered/co-referred by KB IDs, whereas for NIL queries, we grouped them with another round of co-reference annotation. The NIL-coreference annotation was performed separately on files of three types: GPE, PER and ORG. Incorporation of both annotations constitutes the final training data.

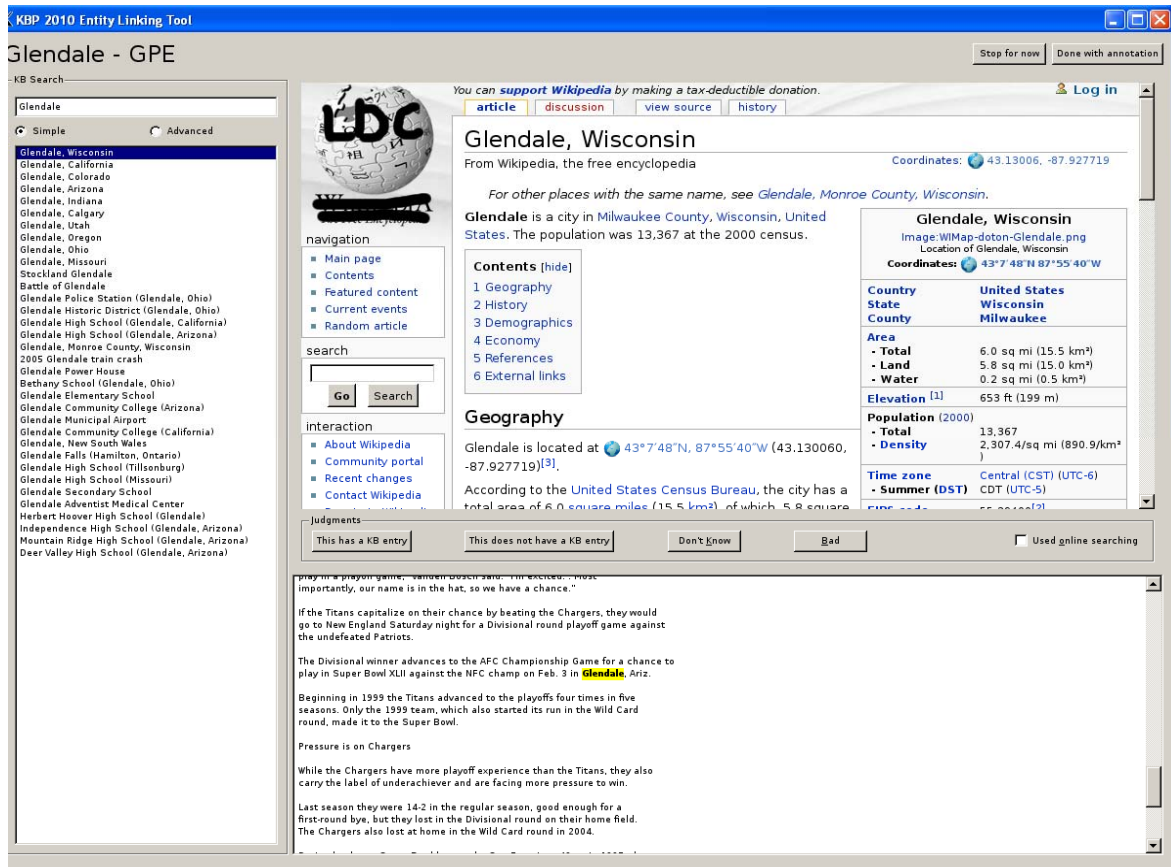


Figure 1: Entity Linking Tool

For the NIL-coreference annotation in the CLEL task, we added two more sub-steps: English annotators co-refed the English query set, and Chinese annotators co-refed the Chinese set separately; afterwards, bilingual annotators co-refed English queries to Chinese queries.

### 5. Annotation Quality

We measured annotation accuracy by computing the micro-averaged accuracy across all queries. Figure 2 delineates the comparison between the averaged human annotators and the top 5 monolingual systems in 2010 on a subset of 200 queries used for an inter-annotator agreement study.

We have seen encouraging advances in the 2011 NEL monolingual task (Ji et al., 2011). Instead of rule-based methods, most systems exploited statistical name variant expansion techniques, such as mining structures or coreference from background documents. Topic modeling was designed for ranking candidates by capturing topic and context features of each query (Kozareva & Ravi, 2011). New supervised learning and ranking algorithms have been introduced. Query classifiers were tuned and trained for each entity type with better performance. Instead of the single query and single KB extraction and disambiguation, systems also enriched the Knowledge

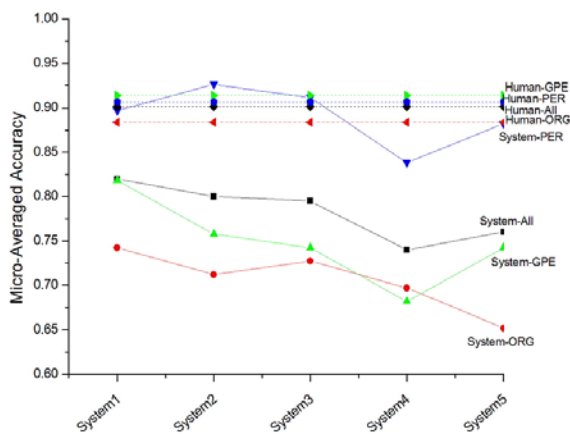


Figure 2: Comparison between System and Human NEL Performance

Base by extracting and disambiguating all entities in the context of a given query.

For the cross-lingual task, it is worth investigating what kinds of challenges have been brought to entity linking because of language barriers. Although this is the first year for cross-lingual track at KBP, the systems achieved

very good performance due to the rich and high-quality training resources provided by LDC. The top cross-lingual entity linking systems in KBP2011 can be ranked at top 4 and 5 in the mono-lingual track, and are better than most mono-lingual entity linking systems.

Systems for the 2011 CLEL task generally used two types of architecture: Name Translation and MT + English Entity Linking; and Chinese Entity Linking + Cross-lingual KB linkages. The former translates a Chinese query and its associated document into English, and then runs English mono-lingual entity linking to link the translated query and document to English KB (McNamee et al., 2011). The latter links a Chinese query to Chinese KB, and then uses cross-lingual KB linkages to map the Chinese KB node to English KB node (Monahan et al., 2011).

For fair comparison, we summarize the performance of Chinese queries and English queries separately in Figure 3. In fact the mean score of cross-lingual queries is only 2.65% lower than that of mono-lingual queries.

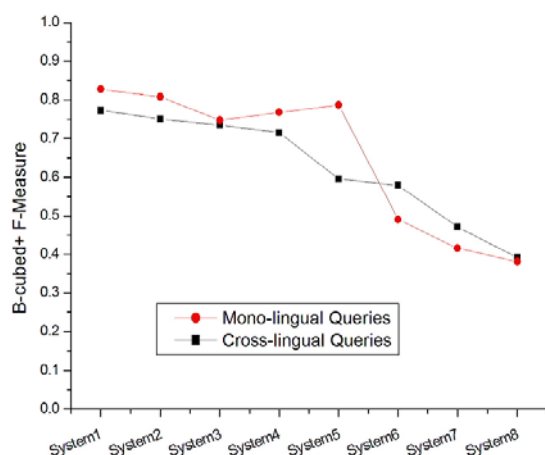


Figure 3: System Performance Comparison of Monolingual Queries and Cross-lingual Queries

## 6. Conclusion

This paper describes the linguistic resource creation efforts in support of Named Entity Linking evaluation of the KBP track, focusing on annotation methodologies, process, and features of corpora from 2009 to 2011, with a detailed discussion on the cross-lingual NEL data. The 2011 KBP NEL evaluation matured with the monolingual task and witnessed a cross-lingual transition. The cross-lingual data enriched the existing NEL corpora, targeting the training and evaluation of the systems' multilingual capabilities. Details discussed herein will facilitate a transparent use of training data for future KBP participants. The linguistic resources described in this paper will be made available to the broader research community via publication in LDC's catalog.

## 7. Acknowledgments

Heng Ji's work in this paper was supported by the U.S. Army Research Laboratory under Cooperative Agreement No. W911NF-09-2-0053 (NS-CTA), the U.S. NSF CAREER Award under Grant IIS-0953149, the U.S. NSF EAGER Award under Grant No. IIS-1144111, the U.S. DARPA Broad Operational Language Translations program and CUNY Junior Faculty Award. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

## 8. References

- Dang, H.T., Lin J., Kelly, D. (2006). Overview of the TREC 2006 Question Answering Track. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*. Genoa, Italy.
- Doddington, G., Mitchell, A., Przybocki, M., Ramshaw, L., Strassel, S., Weischedel, R. (2004). Automatic Content Extraction (ACE) program – task definitions and performance measures. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation*. Lisbon, Portugal.
- Ji, H., Grishman, R., Dang, H.T., Griffitt, K., Ellis, J. (2010). An Overview of the TAC2010 Knowledge Base Population Track. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation*. Valletta, Malta.
- Ji, H., Grishman, R., (2006). Data Selection in Semi-supervised Learning for Name Tagging. In *Proceedings of COLING/ACL 06 Workshop on Information Extraction Beyond Document*.
- Ji, H., Grishman, R., Dang, H. T. (2011). Overview of the TAC2011 Knowledge Base Population Track. In *Proceedings of Text Analysis Conference 2011*. Gaithersburg, Maryland, USA.
- Kozareva, Z., Ravi, S. (2011). Unsupervised Name Ambiguity Resolution Using A Generative Model. In *Proceeding of EMNLP2011 Workshop on unsupervised Learning in NLP*.
- Li, X., Ellis, J., Griffitt, K., Strassel, S., Parker, R. (2011) Linguistic Resources for 2011 Knowledge Base Population Evaluation. In *Proceedings of 2011 Text Analysis Conference*. Gaithersburg, Maryland, USA.
- McNamee, P., Dang, H.T., Simpson, H., Schone, P., Strassel, S. (2010). An Evaluation of Technologies for Knowledge Base Population. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation*. Valletta, Malta.
- McNamee, P., Mayfield, J., Oard, D., Xu, T., Wu, K., Stoyanov, V, Doerman, D. (2011) Cross-Language Entity Linking in Maryland during a Hurricane. In *Proceeding of Text Analysis Conference 2011*. Gaithersburg, Maryland, USA.
- Monahan, S., Lehmann, J., Nyberg, T., Plymale, J., Jung, A. (2011). Cross-Lingual Cross-Document Coreference with Entity Linking. In *Proceedings of Text Analysis Conference 2011*. Gaithersburg, Maryland, USA.
- Simpson, H., Strassel, S., Parker, R., McNamee, P. (2010)

Wikipedia and the Web of Confusable Entities:  
Experience from Entity Linking Query Creation for  
TAC 2009 Knowledge Base Population. In  
*Proceedings of the Seventh International Conference  
on Language Resources and Evaluation*. Valletta,  
Malta.