Towards Named Entity Extraction and Translation in Spoken Language Translation

Fei Huang, Stephan Vogel, Alex Waibel

Language Technologies Institute,
School of Computer Sciences
Carnegie Mellon University
Pittsburgh, PA USA
{fhuang, vogel, ahw}@cs.cmu.edu

Abstract

In this paper we propose a new method of detecting and translating named entities (NE) from spoken language, e.g., Chinese broadcast news. This approach detects possible NE regions from less reliably recognized hypotheses using confidence measures. Each possible NE boundary within the region is compared with candidate NEs from retrieved documents based on their acoustic similarities and semantic correlations. These candidate NEs are re-ranked by additionally incorporating general and topic-specific language models to measure the NE context consistency. This approach, combined with the HMM-based NE extraction on confidently recognized words, improves NE extraction F-score from 66% to 71% and NE translation quality from 69% to 73% over the baseline method. Systematic comparisons on NE translation quality with different speech input quality are also presented.

1. Introduction

Named entity (NE) refers to the structured information designating particular objects, such as persons, locations, and organizations. Since NEs capture key information from human communication, detecting and translating them benefits many natural language processing tasks. The efficiency of a human analyst can be greatly improved when reliable NE identification extracts factual information from unstructured documents, such as text and speech. Correct NE translations also broaden the scope of information access by incorporating facts presented in a foreign language. For example, translated NEs are either the key queries in cross-lingual information retrieval or the right answers in multilingual question/answer pairs. In machine translation, meaningful information from original documents is often lost due to incorrect NE translations. Incorrect NE translations also introduce distorted semantic context through an unwanted translation, thereby degrading the overall translation quality.

Although NE extraction from well-formatted text input has been intensively investigated and achieved satisfactory performance, NE extraction from speech remains under-explored. (Kubala et. al. 1998) and (Miller et.al. 2000) applied a text-based NE extraction model on the first best recognition hypothesis of English broadcast news. (Palmer et. al. 2000) extract NEs directly from recognition lattices, and (Zhai et. al. 2004) use N-best lists for Chinese NE extraction.

On the other hand, translating NEs is also a challenging problem (Knight and Graehl 1997). Part of the reason is that NEs are sometimes phonetically transliterated (translation by pronunciation, for example, "阿/a 巴/ba 拉/la 契/qi 亚/ya" is transliterated as "Appalachian"), sometimes semantically translated (translation by meaning, for example, "山脉/shanmai" is translated into "mountains"), and sometimes both strategies are used together. It is difficult to decide which strategy should be used in different contexts. Additionally, neither a transliteration model nor a translation model has a one-to-one mapping between a source unit (e.g., a word, a syllable or a letter) and its target translation; thus one source NE may be translated into several correct target translations. Furthermore, since new names occur infinitely, pre-constructed NE translation dictionaries cannot cover all NEs, and the Out-of-Vocabulary (OOV) problem is always a main issue to be dealt with in both speech recognition and translation. Finally, if the NEs are automatically extracted from text and speech input, NE tagging errors and speech recognition errors further complicate the translation problem (Meng et. al. 2001).

2. Overall Technical Approach

We propose a new approach to NE detection and translation from recognized speech input. We combine confidence measures from speech recognition, a statistical framework for NE extraction, and multiple features for NE translation to tackle this problem.

Less frequently occurring NE words are often not included in a speech recognizer's vocabulary list, thus prevents proper recognition, detection and translation for these NEs. What is even worse is that the recognizer

will misrecognize them as a word from the vocabulary with a similar pronunciation, and thus may change the local context of the NE, leading to more recognition errors around the NE and more difficulties in NE detection and translation.

With speech recognition confidence measures, we can distinguish reliably and less reliably recognized word hypotheses. We use a standard word-based statistical model to detect NEs from reliably recognized words. For recognition hypotheses with low confidence, we search for the most likely NE boundaries (if there is an NE) within the region, then find the best-matched candidate NEs from topic relevant documents based on their acoustic similarity and semantic correlations. We re-evaluate these candidate NEs by additionally incorporating general and topic-specific language models. The topic-specific language model is trained on relevant source documents, which are retrieved using the recognized hypothesis as the query. Although in this paper we mainly discuss our approach applied in the retrieved source documents, we also give an example of the feasibility of directly detecting and translating NEs from target documents.

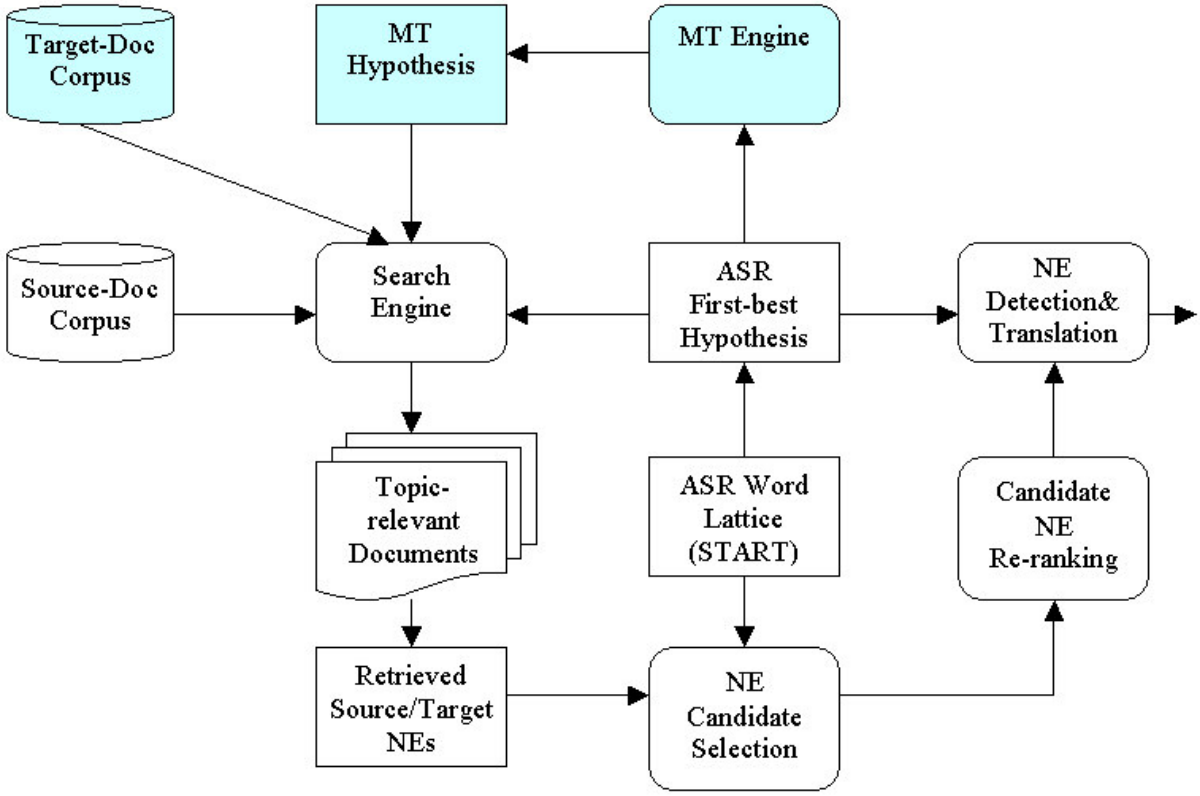

Figure 1. Overall Architecture for NE Detection and Translation

The overall architecture of the proposed method is illustrated in Figure 1. Given a recognition word lattice, the first-best recognition hypothesis is determined based on its acoustic and language model scores. Moreover, the confidence score of each recognized word is computed, and adjacent words with low confidence scores are grouped as "candidate NE regions". This recognized hypothesis is formulated as a query for a search engine. After searching in a pre-indexed corpus, the search engine returns a set of documents which are topically relevant to the recognized speech. NE words in the retrieved documents are extracted and compared with all possible NE boundaries within the detected "NE region" based on their phonetic similarity and semantic correlation, and candidate NEs are returned. In the next step, candidate NEs are re-evaluated based on their acoustic similarity and language model scores, and the one with the minimum matching cost is selected to replace the hypothesized NE words in the first run. Note that, if we apply a machine translation process on the first-best word hypothesis, we are able to access crosslingual information relevant to the source NEs, for example, the translation of a person's name. This additional information is helpful if the true source NE is not included in the retrieved source documents. In

this case, a transliteration from the target NE to the source NE may act as an alternative. It is also helpful for the following NE translation process.

In the following, we will describe candidate NE extraction in section 3, candidate NE re-ranking in section 4, and NE translation in section 5. Finally, we will present some experiment results.

3. Candidate NE Extraction from ASR Hypothesis

3.1. NE Extraction from Reliably Recognized Hypothesis

NE extraction from text input has been intensively investigated, and several successful frameworks have been proposed. We adopt the HMM-based NE extraction framework, where an NE generative model is used to capture the sentence generation process. Given a sequence of words $\vec{W} = (w_1, w_2, ..., w_n)$ and its corresponding NE type sequence $\vec{N} = (N_1, N_2, ..., N_n)$, the probability of generating the words from the NE type sequence is defined as

$$P(\vec{W} \mid \vec{N}) = \prod_{i=1}^{n} p(w_i, N_i \mid w_{i-1}, N_{i-1}) \qquad (1)$$

where $p(w_i, N_i \mid w_{i-1}, N_{i-1})$ denotes the transitional probability from w_{i-1} to w_i, given that the corresponding NE types are N_i and N_{i-1}. This probability can be decomposed as the product of the following three probabilities (for more details, please refer to (Bikel et. al. 1998)):

1. Generating next NE class given its previous NE class and previous word, $P_c(N \mid w_{-1}, N_{-1})$;
2. Generating the first word in a new NE class given the current and previous NE classes, $P_f(w_1 \mid N, N_{-1})$;
3. Generating the subsequent word within the current NE class, given its previous word, $P_b(w \mid w_{-1}, N)$. This is an NE class-dependent bigram model.

All these probabilities can be calculated from annotated corpora by means of supervised learning. A 93-94% F-score (a combined measure of precision and recall) can be achieved on English newswire text.

Directly applying the above word-based NE extraction model to a ASR hypothesis is not optimal, because some words in the hypothesis may include recognition errors, and the NE generative model is applied to distorted NE and context words. In fact, (Kubala et. al. 1998) and (Miller et.al. 2000) reported F-score degradation on speech recognition hypotheses with increasing WER: on average, 1% increase of WER corresponds to 0.7 drop in F-score. However, with speech recognition confidence measures, we can identify reliably recognized words in the hypothesis, and only apply the word-based NE extraction model to those.

3.2. NE Extraction from Unreliably Recognized Hypothesis

Because of speech recognition errors, some NEs, especially less frequently occurring NEs, are likely to be misrecognized. As a result, the NE and its context words are decoded as words with similar pronunciation. This might lead to different NE extraction errors: deletion, insertion and substitution. For example, in the following broadcast news segment, the person name "谷/gu 源/yuan 洋/yang" was misrecognized as an organization name "国/guo 务/wu 院/yuan 向/xiang", which is a substitution error:

REF: *ORG*{中国社会科学院} *ORG*{世界 经济 与 政治 研究所} *PER*{谷 源 洋}

ASR: *ORG*{中国社会科学院} *ORG*{世界 经济 与 政治 研究所} *ORG*{国务院} 向

To deal with such problems, it is necessary to reliably detect the boundaries of actual NEs. In other words, we need to detect the boundary of an actual NE within a given speech segment, even though it may be misrecognized as other words. Text-based NE extraction will fail because of wrong word identities (both NE words and context words). We use their confidence measures from speech recognition to identify low confidence recognition regions, then select possible NE boundaries within the region based on its match to candidate NEs.

Confidence measures have been widely used to estimate the reliability of the recognition hypothesis (Huang, et. al. 2001). The posterior probability $P(W \mid X)$ is a good measure for recognizing a word hypothesis W given its acoustic input X. With Bayes rule,

$$P(W \mid X) = \frac{P(W)P(X \mid W)}{\sum_{W'} P(W')P(X \mid W')}, \qquad (2)$$

using an N-best list of W' generated from word lattices, the ratio can be efficiently computed. For each recognized word, we compute its confidence score, and adjacent words with lower confidence scores are further grouped into an "unsure word region". Notice that as misrecognized Chinese words may have segmentation errors (e.g., a three-character word may be recognized as three single-character words), we convert unsure words into their character sequences, thus obtaining an "unsure character region" (UCR).

To detect candidate NE boundaries from UCR, we use a variable-length window sliding within the region.

For each window position, we treat it as a hypothesized NE, and find the best matched NEs from retrieved topic-relevant documents. The matching is based on their acoustic similarity modeled by a transliteration model, and semantic correlation modeled by a context vector model. NEs with high matching probabilities are registered with their corresponding window positions, which is further evaluated in the next re-ranking step.

3.2.1. Acoustic Similarity: Transliteration model

The transliteration model captures the pronunciation similarities between two NE representation forms, which could be a source NE string, its romanization script (pinyin for Chinese NEs), the recognized phoneme sequence, or its translation in a target language.

To compare the acoustic similarity between a pseudo-NE from the window and a Chinese NE from retrieved documents, we convert both into their pinyin scripts, and calculate their transliteration probability based on (Huang et. al. 2004). With the independence assumption about letter transliteration we obtain,

$$P_{trl}(e \mid f) \approx \prod_{i=1}^{m} p(e_i \mid y_i) = \prod_{i=1}^{m} \prod_{j=1}^{m_i} p(e_{i,k} \mid y_{i,j}). \quad (3)$$

That is, the transliteration probability between two pinyin scripts is approximated by the product of their letter transliteration probabilities over aligned letter pairs, where the alignment path can be searched through dynamic programming based on letter alignment probabilities.

3.2.2. Semantic Correlation: Context Vector Model

This model infers the semantic similarity of an NE pair based on their context words' semantic correlations, with the assumption that reliably recognized context words might represent the NE's role (e.g., the title of the person) in that sentence. Context vectors are composed of certain context words within a predefined window, and each word has a different weight determined by its POS tag and the distance to the NE (Huang et. al. 2004). A context vector can be represented as a set of (w, g) pairs, where w is the context word, and g is its weight. For speech-based NE detection and translation, constructing context vectors requires confidence information about the recognized context words because we do not want to include misrecognized words into context vectors. One option is to only consider reliably recognized context words, even if they are beyond the predefined context window, as in our current implementation. While the semantic correlations between cross-lingual context vectors can be effectively computed via IBM Model-1, similarities between two source context vectors can be calculated in terms of WordNet, latent semantic analysis, or syntactic structures. In our current implementation, we only consider the number of words occurring in both vectors. For a context vector $C1$ with l (w, g) pairs and a context vector $C2$ with m (w, g) pairs, their similarity measure is defined as

$$P_{cv}(C1, C2) = \frac{\sum_{w_t \in (C1 \cap C2)} g_{1t} g_{2t}}{\sum_{i=1}^{l} g_{1i} \sum_{j=1}^{m} g_{2j}} \quad (4)$$

Overall, the similarity between a hypothesized NE N_h and a retrieved NE N_r is defined as:

$$Sim(N_h, N_r) = \lambda \log P_{trl}(N_h \mid N_r) + (1-\lambda) P_{cv}(C(N_h), C(N_r)) \quad (5)$$

where $C(N)$ is the context vector of the NE N. This similarity is the linear interpolation of their transliteration cost and context vector similarity. The interpolation weight λ is chosen empirically.

3.2.3. Search for Candidate NEs

Given a hypothesized NE (or its pronunciation approximation, i.e., detected "NE" words which may contain recognition and NE extraction errors) together with its context, we want to find documents containing the true NE. After automatically tagging all NEs in the retrieved text, we can compare the hypothesized NE with each retrieved NE based on their acoustic and semantic similarities. Finally we choose the best-matched NEs as candidate NEs, which are to be re-ranked in the following step. Assuming that documents containing the same NE share common topics, our task is to search for topic relevant documents using the recognition hypothesis as the query.

We use the whole discourse recognition hypothesis as the query, searching a pre-indexed Chinese document corpus to find topic-relevant Chinese documents. The indexed corpus is composed of 63,092 Chinese documents from the Xinhua News Agency, which corresponds to over 444K sentences and 22M words. The retrieved top 100 Chinese documents are selected for the NE extraction pool. NEs in these documents are automatically detected and their context vectors are identified. Each NE in retrieved documents is compared with the recognized hypothesized NE words, and the best-matched retrieved NEs will be selected as the candidate NEs to be re-ranked.

4. Candidate NE Reranking

To re-rank candidate NEs, in addition to the acoustic and context semantic similarities, we introduce language models to evaluate their linguistic fitness to the given context.

Mathematically, for each hypothesized NE N_h, we want to find the candidate NE N_r^*, so that

$$\begin{aligned} N_r^* &= \arg\max_{N_r} P(N_r \mid N_h) \\ &= \arg\max_{N_r} \frac{\exp(\sum_{m=1}^{M} \lambda_m h_m(N_r, N_h))}{\sum_{N_r'} \exp(\sum_{m=1}^{M} \lambda_m h_m(N_r', N_h))} \\ &= \arg\max_{N_r} \sum_{m=1}^{M} \lambda_m h_m(N_r, N_h) \end{aligned} \quad (6)$$

Here N_r' is the set of candidate NEs, h_m refers to different similarity feature functions, including phonetic similarity based on the transliteration model, context semantic similarity based on the context vector model and context consistency based on N-gram language models. The phonetic and semantic similarities are computed as described in formulae (3) and (4). The context consistency requires the appropriate encapsulation of the candidate NE N_r within the original context $C(N_h)$. We first extract the hypothesized NE together with its several left and right context characters, then replace the hypothesized NE with the candidate NE to construct a new context string. We estimate the context string's N-gram generative probability by incorporating two language models (LMs): a general LM P_g trained from broad topic newswire text and a topic-dependent LM P_d trained from retrieved documents.

$$\begin{aligned} &h_{lm}(N_h, N_r) = \\ &\lambda_g P_g(N_r \mid C(N_h)) + (1-\lambda_g) P_d(N_r \mid C(N_h)) \end{aligned} \quad (7)$$

The latter model captures its topical relevancy, while the former model allows for a smooth back-off in case of data sparseness.

The best matched candidate NE under the re-ranking function is selected if its re-ranking cost is above a certain threshold. This recovered NE replaces the hypothesized NE in the corresponding matching window, and is marked with the appropriate NE tag from the relevant documents.

5. NE Translation

Once NEs are identified from the recognition hypothesis, which may be either extracted from reliably recognized words or searched from relevant documents for unsure character regions, translating them is straightforward. We can directly apply the text-based NE translation techniques to the speech recognition hypothesis: either the NE translation lookup from a pre-constructed bilingual NE dictionary or the target NE retrieval from topic-relevant target documents.

Additionally, if the context words of an NE can be reliably recognized and the boundary of the NE can be reliably detected from its context (which is sometimes true for PERSON NEs), we can find the translations of these (possibly misrecognized) NEs directly from the target topic-relevant documents, without the need of recovering its original source NE. For example, the original transcript "美国 副 助理 国务卿 卡塔曼" was misrecognized as "美国 副 助理 国务卿 他 爆满", where only the person's name "卡塔曼/kataman" was misrecognized as "他 爆满/ta baoman". If we can detect the boundary of the source NE from the context (in this example, the context of "US Deputy Assistant Secretary"), we can directly search the retrieved relevant English documents, finding the NE "Kartman", which is phonetically similar to "他 爆满/ta baoman" and semantically relevant to "US Deputy Assistant Secretary". Research in this direction is still in progress.

6. Preliminary Experiments

We did initial experiments on Chinese-English NE detection and translation, both on manually transcribed speech and on ASR output. 898 NEs (1550 words) corresponding to 284 unique NEs were extracted from a one-hour Chinese broadcast news speech segment, which contained 114 sentence-level utterances and 9176 Chinese words after automatic word segmentation.

We manually annotated NEs from the manual speech transcription. These NEs (as well as the automatically extracted NEs) were translated by combining a bilingual NE translation dictionary and a crosslingual NE translation retrieval technique. The NE dictionary was constructed by aligning automatically tagged Chinese and English NEs from a sentence alignment parallel corpus, and included 12500 NE translation pairs. We evaluated the NE translation quality by means of precision, recall and F-score. Precision was calculated as the correctly translated NEs divided by the total number of translated NEs, while recall was calculated as the number of correctly translated NEs divided by the total number of correct NEs in the manual annotation. F-score is defined as 2PR/(P+R). Due to errors from NE extraction and translation, we classified NE translation results into three categories:

- *Correct*, where neither NE extraction nor translation had any errors;
- *Acceptable*, where there were minor errors in either NE extraction or translation, but the result was acceptable, e.g., two NEs "邓亚萍" and "杨影" were detected as one NE "邓 亚萍 杨 影" and translated as "deng yaping / yang ying";
- *Wrong,* where there were significant errors either in NE detection or translation.

In the first scenario, we only translated NEs from manually transcribed and annotated speech, thus there were no errors from speech recognition and NE extraction. This illustrated the performance of the NE translation module, as shown in Table 1. We evaluated the NE translation quality in both the *Correct* and the *Acceptable* cases. NE type referred to the total number of unique NEs, while NE token referred to the total number of NEs. Since the NEs to be translated were the same as the correct NEs in the manual annotation, precision equaled recall and F-score.

Table 1. NE translation performance on manually transcribed and manually annotated NEs

	Correct		Acceptable	
	Type	Token	Type	Token
Precision	72.18	83.07	75.70	85.74
Recall	72.18	83.07	75.70	85.74
F-score	**72.18**	**83.07**	**75.70**	**85.74**

In the second scenario, we applied our HMM-based NE extraction module to the manual transcriptions, and then translated those automatically detected NEs. We got the NE extraction F-score of 85.59. The NE translation results are presented in Table 2. Notice that 15% NE extraction errors led to an additional 7-10% NE translation errors.

Table 2. NE translation performance on manually transcribed and automatically extracted NEs

	Correct		Acceptable	
	Type	Token	Type	Token
Precision	57.19	75.66	63.01	79.93
Recall	58.80	73.05	64.78	77.17
F-score	**57.98**	**74.33**	**63.87**	**78.52**

In the third scenario, we applied the same techniques to extract NEs from ASR output, which had a character error rate of 18.2%. We got a much lower NE extraction quality; an F-score of 66.25. Detailed analysis in Figure 2 showed that with misrecognized NE words, both deletion and false insertion NE errors increased by 15%.

Translating these detected NEs, we got even lower NE translation quality. As shown in Table 3, 18% character error rate led to 19.3% increase on NE extraction errors, which jointly led to an additional 9-12% NE translation errors.

Table 3. NE translation performance on ASR hypothesis and automatically extracted NEs

	Correct		Acceptable	
	Type	Token	Type	Token
Precision	44.09	70.28	54.00	77.63
Recall	48.59	56.34	59.51	62.25
F-score	**45.96**	**62.54**	**56.62**	**69.10**

Finally, we tested the proposed NE detection and repairing method on the ASR hypothesis. As the candidate NE replacement was on the character level, we also converted the recognized word hypothesis into a character sequence. Extracting NEs on the corrected hypothesis improved the NE extraction F-score from 66.25 to 71.73. Further analysis showed that improvement came from both the character-based NE detection model (which avoided the word segmentation problem) and the corrected NEs from retrieved documents, as shown in the following example:

Man.Transcr: 任命 @PER{列显伦} @PER{沈澄} @PER{包致金} 为 终审 法院 常设 法官

Man.Transl: appoint @PER{Henry Litton} @PER{Charles Ching} @PER{Kemal Bokhary} as Court of Final Appeal permanent judges

ASR Ouput: 任命 电线 论 县城 包 之 心 为 终审 法院 常设 法官

ASRTransl: appointed wire by the county package of mind for final trials court permanent judges

CHP: 任 命 @{列 显 伦} @PER{沈 澄} @PER{包 致 金 生} 终 审 法 院 常 设 法 官

CHPTrans: appointed @PER{patrick chan} @PER {charles ching} @PER{bokhary} for final trials court permanent judges

The first two sentences are manually transcribed speech and manually annotated and translated NEs. The middle two sentences are ASR output and automatic NE extraction and translation results. Due to recognition errors for person names, all words were translated according to their individual semantic meanings, which were neither coherent nor related to the source NEs. From the last two sentences, one may notice the three person names recovered from the initial recognition errors (although an additional character is added into one person name) in the corrected hypothesis (**CHP**). **CHPTrans** shows the translations of these newly detected NEs, where one NE was correctly translated and another was partially correctly translated. Table 4 shows the overall translation quality on the corrected hypothesis, with an absolute 4-6% improvement over the direct ASR NE translation. Detailed analysis showed that most of the corrected NEs are PERSON NEs for which the context words had been correctly recognized, which makes it possible to accurately model context correlations and consistencies.

Table 4. NE translation performance on improved ASR hypothesis and automatically extracted NEs

	Correct		Acceptable	
	Type	Token	Type	Token
Precision	53.42	76.91	59.93	80.47
Recall	57.75	64.92	64.79	67.93
F-score	**55.50**	**70.41**	**62.26**	**73.67**

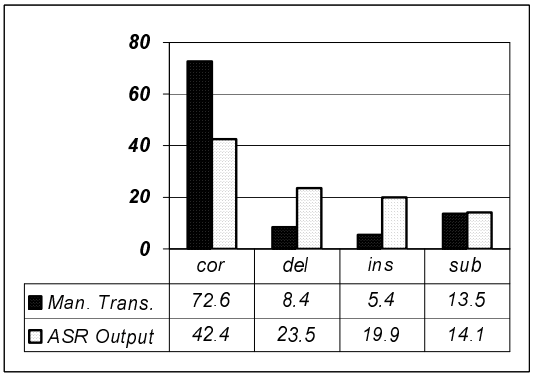

Figure 2. Percent of NE extraction errors in reference and ASR hypothesis

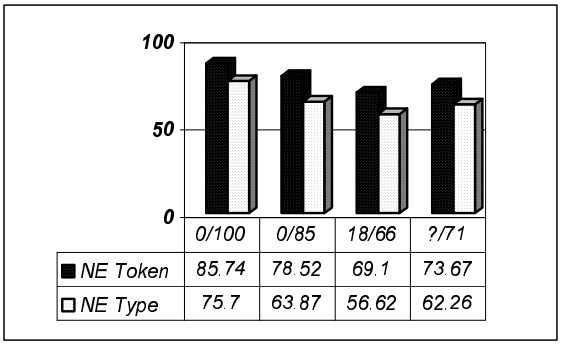

Figure 3. NE translation quality with degraded speech input

Figure 3 shows the overall NE type and token translation degradation on speech input with different qualities and NEs extracted with different accuracies (from left to right): manually transcribed and annotated NEs, manually transcribed but automatically extracted NEs, automatically recognized and extracted NEs, and the partial correction of the third NE input. *n*/*m* at the X-axis means the input has a character error rate of *n* and an NE extraction F-score of *m*. Obviously, with higher character recognition error rate and lower NE extraction F-score, NE translation quality decreases. However compared with the degradation on speech recognition and NE extraction, NE translation quality decreases much slower.

7. Conclusions

We proposed a method of detecting and translating named entities from Chinese broadcast news. This approach detects possible NE regions on unreliably recognized words. Candidate NEs from retrieved documents are compared with each possible NE boundary within the region, and re-ranked based on their acoustic similarity, semantic correlation, and context consistency. This approach, combined with the HMM-based NE extraction of confidently recognized words, improves NE extraction F-score by an absolute 5% and NE translation quality by an absolute 4.5% over the baseline method.

8. References

(Bikel et.al. 1997) D. Bikel, S. Miller, R. Schwarz and R. Weischedel. 1997. Nymble: A High-performance Learning Name-finder. *In the Proceedings of Applied Natural Language Processing,* pp.194-201, Washington DC, 1997.

(Huang et. al. 2001) X. Huang, A. Acero and H. Hon. *Spoken Language Processing, A Guide to Theory, Algorithm and System Development* Prentice Hall Inc. Upper Saddle River, NJ, 2001.

(Huang et. al. 2004) F. Huang, S. Vogel and A. Waibel. Improving Named Entity Translation Combining Phonetic and Semantic Similarities, *in the Proceedings of the HLT/NAACL2004,* Boston, MA, May, 2004.

(Knight and Graehl, 1997) K. Knight and J. Graehl. Machine Transliteration. *In the Proceedings of the ACL '97.* pp.128-135, Somerset, NJ, 1997.

(Kubala et.al. 1998) F. Kubala, R. Schwartz, R. Stone, and R. Weischedel, Named Entity Extraction from Speech, in the *Proceedings of DARPA Broadcast News Transcription and Understanding Workshop*, Lansdowne, VA, February 1998.

(Meng et. al. 2001) H. Meng, W. K. Lo, B. Chen and K. Tang. 2001. Generating Phonetic Cognates to Handle Named Entities in English-Chinese Cross-Language Spoken Document Retrieval. *In the Proceedings of the Automatic Speech Recognition and Understanding Workshop*, Trento, Italy, December 2001.

(Miller et.al. 2000) D. Miller, S. Boisen, R. Schwartz, R. Stone and R. Weischedel, Named Entity Extraction from Broadcast News, *in the Proceedings of the sixth conference on Applied Natural Language Processing,* Seattle, WA. pp316– 324, 2000.

(Palmer et. al. 2000) D. Palmer, M. Ostendorf, and J. Burger Robust Information Extraction from Automatically Generated Speech Transcriptions, *Speech Communication*, vol. 32, pp. 95--109, 2000.

(Zhai et. al. 2004) L. Zhai, P. Fung, R. Schwartz, M. Carpuat and D. Wu, Using N-best Lists for Named Entity Recognition from Chinese Speech, *in the Proceedings of the HLT/NAACL2004,* Boston, MA, May, 2004.