Semi-Supervised Word Sense Disambiguation for Mixed-Initiative Conversational Spoken Language Translation

Sankaranarayanan Ananthakrishnan, Sanjika Hewavitharana, Rohit Kumar, Enoch Kan, Rohit Prasad, and Prem Natarajan
Speech, Language, and Multimedia Business Unit
Raytheon BBN Technologies
10 Moulton Street, Cambridge, MA 02138, U.S.A.
{sanantha,shewavit,rkumar,ekan,rprasad,pnataraj}@bbn.com

Abstract

Lexical ambiguity can cause critical failure in conversational spoken language translation (CSLT) systems due to the wrong sense being presented in the target language. In this paper, we present a framework for improving translation of ambiguous source words that (a) constrains statistical machine translation (SMT) decoding with phrase pair clusters to select a desired sense for translation; (b) automatically predicts the intended sense of an ambiguous source word given its context; and (c) combines the above to define a set of interactive strategies to confirm the intended sense of an ambiguous word and guide the system to the correct translation. The novel use of this framework in a real-world CSLT system distinguishes our approach from the existing work focusing on word sense disambiguation (WSD) for non-interactive, batch-mode SMT. In addition to reporting metrics that evaluate this approach in an interactive spoken language translation system, we also present offline assessments of the component technologies, viz. constrained SMT decoding with sense-specific phrase pair clusters, and automated word sense prediction.

1 Introduction

Lexical ambiguity arises when a single word can refer to different concepts depending on the context. The underlying *senses* of these ambiguous words may be related as in *polysemous words* (e.g. *wood* can refer to *a piece of a tree* or *forested area*), or completely unrelated as in general *homographs* (e.g. *fair* can mean *pale complexion*, *equitable*, or *carnival*). Statistical machine translation (SMT) of spontaneous conversational speech is particularly susceptible to word sense errors arising from these ambiguous words due to two primary factors.

First, phrase-based SMT constructs hypotheses based on phrase pairs with limited context. Thus, it is liable to mis-translate less frequent senses of an ambiguous source word if the latter is not disambiguated by the surrounding words. Further, translation errors can occur even if the local context fully disambiguates the word, if that context has not been observed in SMT training data.

Second, spontaneous conversational speech often depends on dialog context to fully understand and interpret. For instance, the word *fair* in the utterance "*the village head is a fair man*" could refer to *pale complexion* or *equitable*, depending on the dialog context. In such cases, phrase-based SMT will usually translate the ambiguous source word in the sense that occurs most frequently in the training data, even if a different sense was intended.

Table 1 illustrates this problem in a few conversational utterances for English-to-Iraqi Arabic phrase-based SMT using the baseline system of Section 2. In all of these cases, the translated sense of the ambiguous English word is incorrect, even though the context clearly disambiguates the intended sense. These examples further illustrate that *word sense translation errors affect not just polysemous words, but also homographs referring to entirely unconnected concepts.*

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

Sima'an, K., Forcada, M.L., Grasmick, D., Depraetere, H., Way, A. (eds.)
Proceedings of the XIV Machine Translation Summit (Nice, September 2–6, 2013), p. 43–50.

English Input	Arabic Translation
we need to **fast** twenty days next month	MHnA lAzm **bsrEp** {*quickly*} ERryn ywm AlRhr AljAy
we are going to visit your grandfather's **grave** tomorrow	MHnA rH nzwr jdk **kbyrp** {*big/large*} bAkr
after our **late** leader died our town mourned for several weeks	bEd mAltnA **mtJxr** {*delayed*} AlqAQd mAt bldtnA km JsbwE
this fifty pound **note** will cover the cost of dinner	hCA xmsyn **mlAHZp** {*remark*} rH ygTy tklfp AlERAG

Table 1: Baseline phrase-based SMT of English utterances to Iraqi Arabic (Buckwalter transliteration). The ambiguous English word and its Arabic translation are shown in boldface. The translated sense (in curly braces) is incorrect in all cases.

1.1 Novel Contributions

We present novel techniques and strategies that alleviate the word sense problem in real-world, interactive CSLT systems. First, we propose semi-supervised constrained k-means for partitioning bilingual phrase pairs, and describe how these partitions can be used in conjunction with automatically- or user-derived sense labels to constrain the SMT hypothesis space and obtain better translations for ambiguous source words.

Second, we develop a supervised word sense disambiguation (WSD) system that uses maximum-entropy (maxent) classifiers to predict the sense of an ambiguous word using contextual and dependency features, and show how it is combined with constrained decoding in our interactive CSLT system to identify and implement appropriate word sense translation error-resolution strategies. Our interactive system supports *mixed-initiative* error resolution, meaning that both the system and the user can take action to discover and correct word sense translation errors.

A key contribution of this effort is the development of a *high-precision* framework for detection and correction of word sense problems in interactive CSLT systems. The system relies on a set of pre-defined ambiguity classes and corresponding senses to give users precise feedback and receive targeted guidance with the objective of correctly translating ambiguous source words.

1.2 Previous Work

There is a large body of work devoted to automatic discovery of word senses using monolingual or bilingual corpora (Schutze, 1998; Diab and Resnik, 2002; Ng et al., 2003). The general approach is to induce word senses by clustering ambiguous words based on the distributional similarity of their context. Our work differs from these approaches in that the objective is to disambiguate the sense of a given word from among a set of pre-defined word senses, specifically to improve the precision of an SMT system.

Yarowsky's (1995) well-known unsupervised word sense disambiguation approach starts with a small number of training examples representative of each sense of an ambiguous word, followed by iterative self-training to label training examples. Our semi-supervised approach deviates from his work in the following important ways. First, unlike Yarowsky's monolingual word sense labeling technique, our approach partitions the bilingual translation phrase pairs into different senses; our primary goal is the integration of WSD within SMT via constrained decoding as a building block towards interactive resolution of word sense translation errors in a real-world CSLT system. Second, our constrained K-means clustering approach produces mutually exclusive sense partitions, whereas Yarowsky's decision-list algorithm produces a set of collocations that can overlap. Finally, Yarowsky makes the assumption that there is one sense per discourse, whereas our approach processes one utterance at a time independently, and makes no such assumptions about the dialogue.

Integration of WSD within SMT is an actively explored area. Carpuat and Wu (2007) and Chan et al. (2007) integrated WSD system into a phrase-based SMT system, and a Hiero system, respectively. They reported significant performance improvement on a Chinese-English translation task. Bansal et al. (2012) employed unsupervised clustering to build sense-based clusters of translations and annotate them with usage examples. All this work focuses exclusively on *offline, batch-mode, non-interactive* translation. To our knowledge, the framework proposed in this paper represents the first attempt at integrating clarification-based WSD within an interactive CSLT system.

2 Baseline SMT System

The data corpus for building the phrase-based SMT system is derived from the DARPA TransTac English-Iraqi parallel two-way spoken dialogue collection. These data span a variety of domains including force protection (e.g. checkpoint, reconnaissance, patrol), medical diagnosis and aid, maintenance and infrastructure, etc., and are conversational in genre. The parallel corpus consists of approximately 773K sentence pairs (7.3M English words). Phrase pairs were extracted from bidirectional IBM Model 4 word alignment. A 4-gram target LM was trained on Iraqi Arabic transcriptions.

Our phrase-based decoder, similar to Moses (Koehn et al., 2007), uses the phrase pairs and target LM to perform beam search stack decoding based on a standard log-linear model, whose parameters were tuned with MERT on a held-out development set (3,534 sentence pairs, 45K words). The BLEU and METEOR scores of this system on a separate test set (3,138 sentence pairs, 38K words) were 16.1 and 42.5, respectively. This SMT system is part of a large-vocabulary, real-world CSLT system capable of facilitating communication between English and Iraqi speakers in the specified domains.

3 Constrained SMT Decoding

We introduce the idea of constrained SMT decoding for accurate translation of ambiguous source words, based on a set of pre-defined ambiguous words and corresponding senses.

3.1 Defining Ambiguity Classes and Senses

We define an *ambiguity class* as a set of minor morphological variants of a base-form ambiguous word that are used in similar contexts. For instance, the ambiguity class **FAIR** = {*fair*} contains only one word, while **BEAR** = {*bear, bears*} contains two. A variety of heuristics (e.g. WordNet, public domain homograph lists, words with multiple part-of-speech tags, etc.) was used to identify a set of 240 pre-defined ambiguity classes in source sentences of the parallel corpus. Based on a dictionary/thesaurus, a native English-speaking annotator with basic linguistic training provided a list of pre-defined senses for each ambiguity class. Not all pre-defined senses of an ambiguity class actually occur in the parallel corpus.

Ambiguity classes are essential because many words exhibit lexical ambiguity only in certain morphological forms. For instance, *fair* is ambiguous but *fairs* is completely unambiguous since it can only have one meaning in any context (viz., “carnivals”). Thus, it is not possible to use stemming or lemmatization to reduce ambiguous words to a base-form.

3.2 Decoding with Phrase Pair Partitions

As described in Section 4, we establish, for each ambiguity class, mutually exclusive sense-specific partitions over phrase pairs whose source phrases contain any of the words belonging to that class. Constrained SMT decoding reduces to (a) obtaining intended sense labels for ambiguous source words; and (b) restricting phrase pair choices for source phrases spanning the ambiguous word(s) to those belonging to the partition corresponding to the intended sense.

To enable constrained decoding, each ambiguous source word in the test input is tagged with its corresponding ambiguity class and intended sense. The decoder then chooses a translation from the corresponding phrase pair partition, generating a target hypothesis that reflects the intended sense of the ambiguous source word(s). Thus, constrained SMT decoding is a form of dynamic pruning of the hypothesis search space where the source phrase spans an ambiguous word. Because the search space is constrained only in the regions of source ambiguity, we are able to preserve the intended sense in the target language, while generating fluent translations.

4 Semi-Supervised Phrase-Pair Clustering

Constrained SMT decoding hinges on the ability to construct phrase pair partitions representing the different senses of a contained ambiguous source word. Standard clustering algorithms (e.g. k-means) provide a natural way to automatically partition phrase pairs belonging to an ambiguity class into multiple constituent senses. To facilitate this, phrase pairs can be represented in a bilingual term-frequency vector space indexed by a unified bilingual (e.g. English + Iraqi) vocabulary. While automated, this approach will produce noisy clusters due to (a) inability to exploit constraints intrinsic to the data; and (b) lack of a principled method to initialize the cluster centers.

We propose a novel, semi-supervised technique that uses constrained k-means clustering to partition phrase pairs based on a sparse annotation of sense-specific key-phrases. Key-phrases are n-gram subsets of source phrases of phrase pairs that are definitive indicators of a specific sense. These key-phrases are used to (a) establish a set of *must-link* and *cannot-link* constraints over the set of phrase pairs corresponding to a given ambiguity

Class	Variants	Senses	Sample Key-phrases
ABSENT	*absent*	*missing* *preoccupied*	*absent*people, employees*absent* *absent*minded, absent*mindedly*
LATE	*late*	*delayed* *deceased*	*came*late, working*late* *late*family, late*mother*
NOTE	*note, notes*	*brief record* *give attention* *currency bill*	*note*signed, note*down* *take*note* *dollar*note*
IMPACT	*impact, impacts*	*collision* *effect*	*impact*area* *negative*impact, impacts*us*

Table 2: Excerpts of ambiguity classes in English-Iraqi parallel data with sample sense-specific key-phrases. The asterisk represents whitespace between words in the key-phrases.

class; and (b) anchor initial sense-specific cluster centers as well as the mapping between k-means clusters and sense labels. Thus overcoming both limitations of simple k-means outlined above.

4.1 Key-phrase Generation

We annotate a small set of sense-specific key-phrases for each ambiguity class, enabling us to:

1. Identify an initial subset of sense-specific phrase pairs for each ambiguity class, establishing a robust set of starting centroids for k-means clustering.
2. Establish an initial set of link constraints between phrase pairs, to be expanded via *transitive closure*.
3. Conveniently map constrained k-means clusters back to high-level sense labels.
4. Perform "quick-and-dirty" sense prediction in the interactive CSLT system (Section 7).

Key-phrase generation is a very light annotation process that takes about 5-15 minutes per ambiguity class in our real-world English-Iraqi parallel corpus. A native English speaker without any special linguistic training took only about 24 working hours to generate key-phrases for pre-defined senses of the 240 ambiguity classes. We again emphasize that key-phrases simply capture the "flavor" of word senses as a starting point for constrained clustering and are not intended to exhaustively cover all possible sense-specific contexts in our phrase table. Table 2 shows key-phrases for different senses of a few ambiguity classes observed in our data. Of 240 pre-defined ambiguity classes, the annotator generated key-phrases for multiple senses for only 73 (the remaining 167 occurred in only one sense in our parallel corpus).

4.2 Key-phrase Constraints

Key-phrase constraints establish an initial set of relationships between phrase pairs of a given ambiguity class. Two phrase pairs (A, B) are related by a *must-link* constraint if their source phrases both contain key-phrases associated with the same sense label. They are related by a *cannot-link* constraint if their source phrases contain key-phrases corresponding to different sense labels.

Because key-phrases are not exhaustive, this initial set of constraints is relatively small. We further expand the set of constraints through instance-based relationships and transitive closure.

4.3 Instance-based Constraints

Phrase pairs are extracted from a many-to-many bidirectional word alignment over the parent sentence pairs. Depending on the word alignment, the same ambiguous word may be contained in multiple phrase pairs extracted from a given sentence pair. Figure 1 illustrates a case where the word alignment permits extraction of two phrase pairs spanning the ambiguous word *fair*.

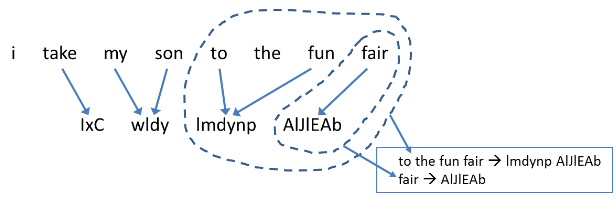

Figure 1: Example shows word alignment and phrase pairs spanning the ambiguous word *fair*. The intended sense (*carnival*) is not evident from the shorter phrase pair, but must be the same as that of the longer one.

Both phrase pairs must be placed within the same cluster, because they are both derived from the same instance and obviously refer to the same sense. Thus, even if shorter phrase pairs lack context, they can "ride the coattails" of longer phrase pairs and ultimately be assigned to the desired partition. We enforce this by establishing *instance-based must-link* constraints between phrase pairs with a common parent sentence pair.

Algorithm 1 Transitive Closure with the Modified Floyd-Warshall Algorithm

$\mathbf{L} \equiv l_{ij} \leftarrow \{true \mid false \mid undef\} \quad \forall (i,j) \in (1 \ldots N, 1 \ldots N)$
for $k \leftarrow 1$ to N **do**
  for $i \leftarrow 1$ to N **do**
    for $j \leftarrow 1$ to N **do**
      if $l_{ik} = undef \vee l_{kj} = undef \vee (l_{ik} = false \wedge l_{kj} = false)$ **then**
        skip to next loop iteration
      end if
      if $l_{ij} = undef$ **then**
        $l_{ij} \leftarrow l_{ik} \wedge l_{kj}$
      end if
    end for
  end for
end for
return L

4.4 Transitive Closure

The constraints established by key-phrases and instance-based co-occurrences can be expanded significantly by exploiting the transitive nature of these relationships. Consider two phrase pairs (A, B); if A must link to B, and B must link to C, then A must link to C. Conversely, if A must link to B, and B cannot link to C, then A cannot link to C. *Transitive closure* propagates the initial set of constraints across all two-tuples of phrase pairs, usually resulting in a far larger set of constraints that lead to well-formed, noise-free clusters.

We implement transitive closure as a modified version of the Floyd-Warshall all-pairs shortest-path algorithm. The entire set of phrase pairs (numbering N) corresponding to a given ambiguity class is treated as a set of nodes, with constraints represented by edges connecting them. The "weight" assigned to each edge is a ternary boolean value with *true* representing a must-link constraint, *false* representing a cannot-link constraint, and *undefined* representing no constraint. The entire graph is represented by a $N \times N$ ternary boolean matrix. Transitive closure is summarized in Algorithm 1. Its complexity is $O\left(N^3\right)$, making it feasible for all ambiguity classes in our corpus.

4.5 Constrained K-means

Armed with the full set of constraints, we apply the constrained k-means clustering algorithm (Wagstaff et al., 2001) to establish a mutually exclusive partition independently over the samples corresponding to each ambiguity class. Two samples that are related by a must-link constraint will always be assigned to the same cluster. Conversely, two samples related by a cannot-link constraint will always be placed in different clusters.

We obtain a reliable, deterministic set of initial centroids by averaging all bilingual term-frequency vectors corresponding to phrase pairs that contain the key-phrases identifying each sense. When the k-means clusters have been mapped back on to high-level sense labels using the key-phrases, we integrate partition information within the SMT phrase table by augmenting each phrase pair with two additional fields, viz. ambiguity class and sense identity.

5 Offline Translation Evaluation

Standard offline evaluation sets for MT usually do not provide a balanced representation of senses for most ambiguity classes. This precludes evaluation of translation effectiveness over the less frequent senses. The use of low-precision, automated corpus-level metrics such as BLEU for measuring translation success on specific ambiguous source words further compounds this problem. We address both issues in our evaluation framework.

First, we created an offline test set consisting of 164 English sentences covering all senses of the 73 ambiguity classes that appeared in multiple senses in our training data. This test set is perfectly balanced, containing exactly one representative of each sense for every ambiguity class. Each test set sentence contains exactly one ambiguous word tagged with its corresponding ambiguity class and intended sense identity. The uniform distribution of senses, while not reflective of real-world data, allows us to evaluate how the systems perform when confronted with less frequent senses of an ambiguous word. This is an important consideration for interactive CSLT systems and is not captured by typical SMT evaluations with natural sense distributions.

Second, we used precise human judgments to evaluate translations of ambiguous source words. We presented each input sentence and its translation to the bilingual judge, with the ambiguous source word and the corresponding target word(s) both highlighted. The evaluator passes a binary judgment; *correct*, implying correct translation of the intended sense, or *incorrect*, indicating an incorrect sense substitution. This evaluation, summarized in Table 3, was performed on the baseline system as well as constrained SMT.

Method	correct	incorrect	unk
Unconstrained	**95**	**68**	**1**
Constrained	**108**	**22**	**34**
Improvement	**13.7%**	**67.6%**	n/a

Table 3: Concept transfer for ambiguous words in offline decoding evaluation.

The third column in Table 3, titled **unk**, merits further explanation. Due to the uneven distribution of senses of several ambiguity classes in our training data, clusters corresponding to low frequency senses may be very sparse; the decoder may not be able to find a contextually suitable phrase pair to translate the corresponding ambiguous word, causing it to be untranslatable. Analysis showed that the baseline system substituted the wrong translation sense in most cases where the constrained decoder was unable to translate the word. This feature is particularly valuable for interactive CSLT systems, serving to trigger appropriate user feedback as described in Section 7. Table 4 illustrates how constrained decoding is able to resolve the correct translation for ambiguous source words where the baseline system failed (compare to Table 1 in Section 1).

Pairwise bootstrap resampling (Koehn, 2004) was used to ascertain statistical significance of the improvements in Table 3. The non-parametric Wilcoxon signed-rank test returned a p-value of 3.67×10^{-10}, allowing us to strongly reject the null hypothesis that the two systems do not differ in performance.

6 Supervised Word Sense Prediction

In this section, we describe traditional, maxent-based supervised word sense classifiers that predict sense labels for each ambiguity class. These classifiers are not directly integrated within the SMT system, but are used to define and select clarification strategies in the mixed-initiative CSLT system. For instance, we engage the user in a clarification dialog to obtain the intended sense of an ambiguous word if its predicted sense does not match the translated sense. We provide further details on these strategies in Section 7.

The supervised sense predictors were trained using sense labels annotated at the sentence level. We selected up to 250 representative source sentences independently for each ambiguity class from the parallel training corpus, using a diversity selection mechanism to obtain a balanced representation of different senses. Using the annotated sentences, we trained a separate maxent classifier for each ambiguity class, with sense identities as target labels. A total of 110 classifiers were trained with the a set of contextual lexical, dependency, and part of speech features. Figure 2 illustrates the features used in our system with an example.

English Input	Arabic Translation
we need to **fast** twenty days next month	MHnA lAzm **#unk#** {*untranslatable*} ERryn ywm AlRhr AljAy
we are going to visit your grandfather's **grave** tomorrow	MHnA rH nzwr jdk **qbr** {*tomb*} bAkr
after our **late** leader died our town mourned for several weeks	bEd mAltnA **AlmrHwm** {*deceased*} AlqAQd mAt bldtnA km JsbwE
this fifty pound **note** will cover the cost of dinner	hCA xmsyn **Alwrqp** {*bill*} rH ygTy tklfp AlERAG

Table 4: Constrained SMT decoding of English utterances. The ambiguous English word and its Arabic translation are in boldface. The first case is untranslatable. The rest are correctly translated. Compare to Table 1.

Contextual lexical features: Indicators representing the local context of the ambiguous word, viz. the previous and next words.

Dependency features: Indicators extracted from a dependency parse of the sentence, specifying parent and child of the ambiguous word.

Part-of-speech features: For better generalization, POS-tag features for both contextual and dependency features are also included.

We conducted an off-line evaluation of the classifiers by using them to predict the sense of ambiguity classes in held out test sentences. The most frequent sense of an ambiguity class in the training data served as a baseline for that class. The baseline word sense predication accuracy rate over 110 ambiguity classes covering 2,324 held-out sentences containing ambiguous words (with a "natural distribution" of senses observed in the training data) was **73.7%**. The accuracy of the supervised classifiers for these sentences was **88.1%**.

7 Mixed-Initiative Interactive CSLT

A set of four *disambiguation strategies* (Figure 3) built around word sense classifier predictions, clarification input from the user, and constrained decoding with phrase pair partitions pre-empts and

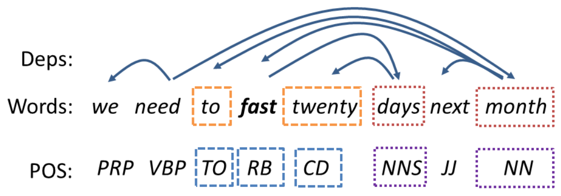

Figure 2: Contextual lexical, dependency, and POS features for ambiguous word *fast*

recovers from errors in translation of ambiguous source words in our interactive CSLT system.

We begin by performing unconstrained SMT decoding and determine the translated sense of each pre-defined ambiguous source word by looking up the sense partition to which the corresponding Viterbi phrase pair belongs. We then attempt to identify the intended sense by checking if the attendant source phrase contains any of the key-phrases associated with the translated sense. If so, the translated sense is correct and no further action is needed. This is the WSD_FILTERED strategy.

If the WSD_FILTERED strategy falls through, we proceed to the WSD_NO_MISMATCH strategy. Here, we check if a word sense classifier is available for the given ambiguity class, and invoke it to predict word sense. If the predicted and translated senses match, no further action is required.

If the translated and classifier-predicted senses do not match, we invoke the WSD_MISMATCH strategy to resolve the sense ambiguity through clarification. The system presents a description of the translated sense, asking the user to confirm if that is the intended sense. If the user confirms, we retain the unconstrained translation. If the user rejects the translated sense, a choice of alternate sense candidates, translatable in the current context, is offered. Constrained SMT decoding is performed with the chosen sense to obtain the final translation. If no such candidates exist, the user is asked to rephrase the utterance, replacing the ambiguous word with an unambiguous equivalent.

If none of the above strategies is applicable, we employ the WSD_BACKOFF strategy, where the user is immediately asked to choose the intended sense, followed by constrained translation. At all times, the user can rephrase the initial utterance or force the system to proceed with the current translation. This allows the user to override system false alarms as needed. Thus, the system uses *mixed-initiative* strategies to recover from potential errors. When it is confident of the translation, it does not perform clarification. When in doubt, it seeks user intervention to obtain the intended sense.

The end-to-end CSLT pipeline incorporating constrained decoding, word sense prediction, and clarification strategies was recently evaluated in a competitive setting. The results showed that, on a set of 18 spoken utterances that triggered one of the above interactive WSD strategies (out of a larger test set that invoked other error types, e.g. OOV words, etc.), the high-level concept transfer rate of the translations improved from **22.2%** without clarification to **55.6%** with clarification and constrained SMT decoding where applicable. This corresponds to a **33.4%** improvement in concept transfer rate over the baseline CSLT system when confronted with word sense ambiguities. The average clarification load of our system for these test utterances was 0.89, i.e. less than one clarification turn per utterance.

8 Discussion and Future Directions

Phrase-based SMT systems are susceptible to word sense translation errors, causing critical communication failures in CSLT systems. Our proposed end-to-end framework for detecting and correcting these errors is comprised of constrained SMT decoding backed by semi-automatically derived sense-specific phrase pair partitions, with further support from supervised word sense classifiers and interactive error resolution strategies.

We used constrained k-means clustering to obtain sense-specific phrase pair partitions. The clustering process was guided by an initial set of constraints obtained from light key-phrase annotations, and further expanded by instance-based constraints followed by transitive closure. Constrained SMT decoding of a balanced, sense-tagged evaluation set with these partitions resulted in a 13.7% increase in correct concept transfer rate and a 67.6% reduction in incorrect concept transfer rate in the English to Iraqi direction.

The above framework was integrated, along with supervised maxent-based word sense classifiers, within a mixed-initiative English-Iraqi CSLT system. Interactive strategies designed around these components detected and corrected word sense translation errors over pre-defined ambiguity classes, while minimizing user clarification load. The end-to-end system improved high-level concept transfer rate by 33.4% on an evaluation set containing ambiguous words that triggered one of

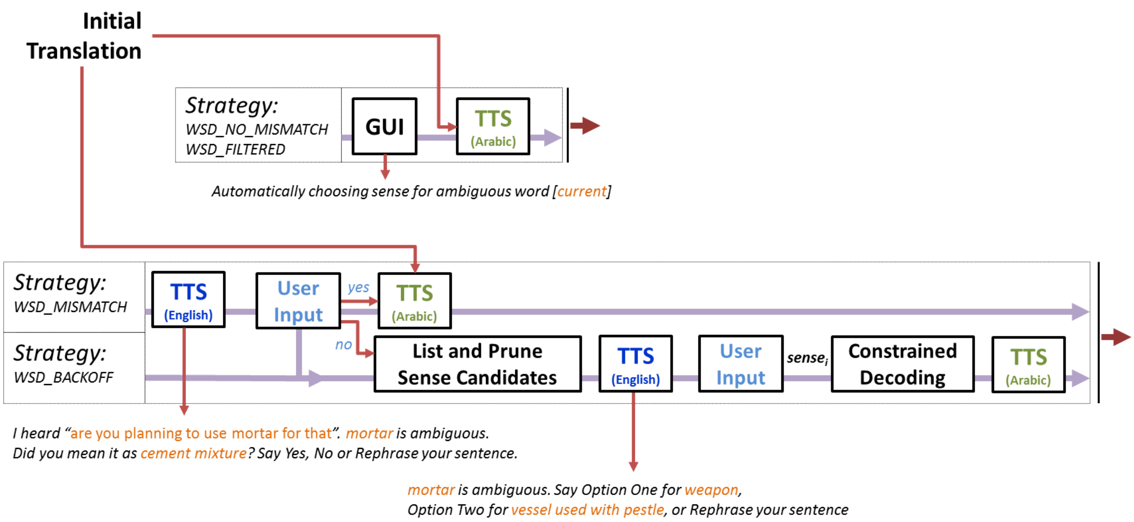

Figure 3: End-to-end word-sense disambiguation and error-recovery strategies for CSLT.

the resolution strategies.

Our constrained clustering approach could also be used to project sense partition information on to individual instances of source words in the parallel corpus. Data annotated in this semi-supervised fashion can be used for a variety of tasks including vocabulary expansion as well as training classifiers for supervised WSD.

References

Bansal, Mohit, John DeNero, and Dekang Lin. 2012. Unsupervised translation sense clustering. In *Processings of the Conference of the American Association of Computational Linguistics: Human Language Technologies*, pages 773–782, Montreal, Canada, June.

Carpuat, Marine and Dekai Wu. 2007. Improving statistical machine translation using word sense disambiguation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 61–72, Prague, Czech Republic, June.

Chan, Yee Seng, Hwee Tou Ng, and David Chiang. 2007. Word sense disambiguation improves statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 33–40, Prague, Czech Republic, June.

Diab, Mona and Philip Resnik. 2002. An unsupervised method for word sense tagging using parallel corpora. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 255–262, July.

Koehn, Philipp, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, ACL '07, pages 177–180, Stroudsburg, PA, USA. Association for Computational Linguistics.

Koehn, Philipp. 2004. Statistical significance tests for machine translation evaluation. In *EMNLP*, pages 388–395, Barcelona, Spain, July.

Ng, Hwee Tou, Bin Wang, and Yee Seng Chan. 2003. Exploiting parallel texts for word sense disambiguation: An empirical study. In *Proceedings of 41st Annual Meeting on Association for Computational Linguistics*, pages 455–462, July.

Schutze, Hinrich. 1998. Automatic word sense discrimination. *Journal of Computational Linguistics*, 24:97–123.

Wagstaff, Kiri, Claire Cardie, Seth Rogers, and Stefan Schrödl. 2001. Constrained k-means clustering with background knowledge. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, pages 577–584, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Yarowsky, David. 1995. Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*, ACL '95, pages 189–196, Stroudsburg, PA, USA. Association for Computational Linguistics.

This paper is based upon work supported by the DARPA BOLT program. The views expressed here are those of the author(s) and do not reflect the official policy or position of the Department of Defense or the U.S. Government.