A Radically Simple, Effective Annotation and Alignment Methodology for Semantic Frame Based SMT and MT Evaluation

Chi-kiu LO and **Dekai WU**
HKUST
Human Language Technology Center
Department of Computer Science and Engineering
Hong Kong University of Science and Technology
jackielo,dekai @cs.ust.hk

Abstract

We introduce a radically simple yet effective methodology for annotating and aligning semantic frames inexpensively using untrained lay annotators that is ideally suited for practical semantic SMT and evaluation applications. For example, recent work by Lo and Wu (2011) introduced MEANT and HMEANT, which are state-of-the-art metrics that evaluates translation meaning preservation via Propbank style of semantic frames. For such applications, however, we argue that the Propbank annotation are too complex and detailed, since they are aimed at training linguists to annotate semantic frames with gold standard accuracy. Instead, we believe that annotating semantic frames for such purposes should be as intuitive as understanding the basic event structure of a sentence, which any untrained human does effortlessly. We propose a simplified set of annotation guidelines consisting of half a page plus three annotated examples. Together with a graphical user interface designed to facilitate the annotation and comparison process by guiding untrained humans step by step, only 5 to 15 minutes are needed to train lay annotators. This allows the lay annotators to focus on understanding the translation to provide consistent and efficient annotation and comparison. The methodology is 'cloud' based to be truly platform independent, installation-free and portable.

1 Introduction

We present a practical alternative to linguistically sophisticated but expensive methodologies semantic frame annotation and alignment, designed in particular with an eye to semantic statistical machine translation (SMT) and MT evaluation. Our approach contrasts with, for example, the complex guidelines for Propbank annotation (Bonial *et al.*, 2010) used to train linguists to annotate semantic frames with gold standard accuracy. Though excellent for their intended purpose, Propbank style guidelines are long and full of linguistic terminology, making them highly unsuitable for training lay persons.

Our efforts are motivated by the increasing needs of recent work on semantic SMT and semantic MT evaluation. In semantic SMT for example, the SRL-for-SMT work of Wu and Fung (2009a) and Wu and Fung (2009b) relies on cross-lingual matching of semantic role labels. In semantic MT evaluation, the metrics MEANT and HMEANT from Lo and Wu (2011a,b,c) are also based on SRL matching.

New research directions of this kind demand quick, inexpensive, relative accurate semantic frame annotation and alignment. We argue that the methodology for annotating semantic frames for such purposes should be as easily intuitive as comprehending the basic event structure of a sentence — which any untrained native speaker does naturally and effortlessly.

Our alternative methodogy achieves this by combining (1) a streamlined, highly simplified and intuitive set of annotation guidelines with (2) an easy-to-use graphical user interface that guides untrained lay annotators step-by-step through the annotation process within (3) a convenient 'cloud' based platform that flexibly supports distributed workflows involving physically separated annotators working on any standard browser.

The streamlined annotation guidelines consist of a mere half-page of instructions — mostly whitespace — supplemented with three annotated examples for reference. The simplicity of the guidelines allows lay annotators to focus on understanding the translation to provide consistent and efficient annotation and comparison. Training an annotator typically takes on the order of five minutes. Despite (or perhaps because) of the simplicity, interannotator agreement is nevertheless quite high.

A graphical user interface is specifically designed to address the risk of annotation inconsistency that arises from using unskilled humans rather than linguistic experts to annotate semantic frames. The guidelines are incorporated into a GUI that guides annotators to label semantic predicate argument structure. The system guides

the annotators to first identify the predicate of a frame, and then specify the span and the role of its associated arguments one by one. Every time when the annotators label a predicate, they start the process of annotating a new semantic frame in the sentence. Each annotated frame is marked up with a different color, so that annotators can clearly distinguish the multiple semantic frames within a single sentence.

Convenient annotation workflows across distributed locations are facilitated by the 'cloud' based approach. The cross-platform web interface is accessible from any mordern Javascript-enabled browser. Annotation of translation is currently supported in any language encoded in UTF8 with left to right orthography. Text is expected to be segmented into sentences, reflecting the common assumption of nearly all present day MT systems.

In the following sections, we first contrast our approach with related work on the process methodology for Propbank annotation. We then propose a concrete set of annotation guidelines. Next, we describe the design of a graphical user interface specifically tailored to guide lay annotators step by step through the process of annotating semantic frames with our simplified set of role labels. Following this, we propose a set of guidelines for aligning and comparing semantic frames for translations, again designed to be easy for lay annotators and yet sufficiently accurate. We also describe the design of the graphical user interface for alignment of semantic frames. Finally, we present experimental results on timing lay annotators, demonstrating the efficiency and low cost of this methodology (which has been shown elsewhere to produce state-of-the-art results for semantic MT evaluation).

2 Related Work

The Propbank annotation guidelines (Bonial *et al.*, 2010) are aimed at training linguists to annotate semantic frames to gold standard accuracy, and are unnecessarily long and technical for lay persons. Propbank requires annotators to determine the word sense for each predicate and is built on top of the syntactic structure in the sentences, using the software tool Jubilee (Choi *et al.*, 2010) to support the complex Propbank annotation and viewing process. Thus, the annotator training cost of Propbank annotation is disproportionately high for applications such as semantic MT evaluation.

Recent works in semantic SMT and MT evaluation show an increasing demand for low-cost semantic frame annotation and comparison. In semantic SMT, for example, Wu and Fung (2009a) and Wu and Fung (2009b) apply SRL to SMT decoding, using an SRL based reordering model that returns improved translations containing fewer semantic role confusion errors. The SRL based reordering model relies on cross-lingual SRL matching. In semantic MT evaluation, a new generation of automatic and semi-automatic MT evaluation metrics proposed by Lo and Wu (2011a,b,c) captures similarities and differences between the reference translation and MT output semantic structures. This approach also relies on SRL matching between reference translation and MT output.

The Propbank annotation guidelines consist of 70 pages, of which 59 pages are annotation instructions and 11 pages cover the menu for the annotation tools. The annotation instructions detail the annotation process, the definition of the argument labels, exception handling for tagging, the handling of null elements in syntax trees and the handling of special cases and spoken data. Since Probank is built on top of the syntactic structure of the sentences, Propbank annotators, i.e. readers of the guidelines, are expected to have prior knowledge of word senses, syntactic structure annotations, and other linguistic information (e.g. null elements). However, all the details in the Propbank annotation guidelines are only necessary when the goal of annotation is to provide consistent and high quality gold standard semantic frame annotation. In contrast, extracting semantic information for practical applications such as semantic SMT and evaluation should be as intuitive as understanding the basic event structure of a sentence which any untrained human does effortlessly.

Since Propbank aims to provide gold standard semantic frame annotation, the annotations are subsequently adjudicated. Therefore, Jubilee, the Propbank instance annotation editor has complex use cases and consists of two modes: the 'normal' mode and the 'gold mode'.

The normal mode is used by annotators to determine the word sense for each predicate in the sentence and annotate the arguments with semantic role labels. Since the Propbank annotation is built on top of the syntactic structure of the sentences and requires annotators to first determine the word sense of the predicate, the normal mode consists of three panels — the treebank view, the frameset view and the argument view. Annotators must navigate around these panels in the different steps of annotation.

One the other hand, the gold mode is used by the adjudicators who select the most appropriate annotation of the instance as the gold standard or correct the annotations if necessary. To determine which annotation of the instance is the most appropriate as the gold standard, in addition to the three panels in the normal mode, the gold mode includes one more panel showing all the annotations for the instance. Similarly, adjudicators must navigate between all the four panels in the different step of adjudication.

The complexity of Jubilee is only necessary when the goal of annotation is providing consistent and high quality gold standard semantic frame annotation. In contrast, for practical applications such as semantic SMT and evaluation, the semantic frame annotation tool should be

Introduction to Semantic Roles

The idea of semantic roles is to summarize the meaning of a sentence (or phrase) using a simple semantic structure that captures essential aspects of the meaning like "*Who* did *What* to *Whom*, *When*, *Where*, *Why* and *How*". Phrases or clauses that express meanings can be identified as playing a particular semantic role in the sentence. In other words, semantic roles are the systemic abstraction of meanings in a sentence.

The following is the list of the semantic roles to be used in this assignment:

Agent (Who)	Action (did)	Patient (What)
Benefactive (Whom)	Temporal (When)	Locative (Where)
Purpose (Why)	Manner (How)	Degree or Extent (How)
Modal (How) [e.g. may, should, etc.]	Negation (How) [e.g. not]	
Other adverbial argument (How)		

** Please note that the "**Action**" role should involve **exactly ONE word** only. **

Figure 1: Instruction of semantic frame annotation for MT evaluation

straightforward and require minimal training instructions in using the tool itself. A software tool supporting these kinds of annotation should be easy to use so that lay annotators can concentrate on evaluating the meaning of the translation and provide consistent annotations for evaluation.

3 Annotating Semantic Frames

To minimize the labor cost of running the semantic MT evaluation metric so that it can be driven by untrained monolingual human, the instructions for annotating semantic frames have to be clear, simple and intuitive. MEANT (Lo and Wu, 2011a) adopted Propbank SRL style predicate-argument framework, which captures the basic event structure in a sentence. The original Propbank annotation specification is designed for readers with strong linguistic background who can distinguish different word senses of predicates. We present the intuitive guidelines and step-by-step guided interface that make semantic role labeling, i.e. identifying the basic event structure—"who did what to whom, when, where and why" (Pradhan *et al.*, 2004) — a task that even untrained monolingual readers can do.

3.1 Simplified set of labels and minimal guidelines

In contrary to the 89 pages of Propbank annotation guidelines, we simplified the instructions of annotation into half of a page intuitively. We first clearly state the objective of semantic role labeling using lay person terminologies. Then, according to the basic event structures—"who did what to whom, when, where and why", we simplified the set of Propbank style semantic role labels into a set of 10 to 12 role labels. Figure 1 shows the half-page instructions with the simplified set of roles.

The "did" event which corresponds to the predicate in the semantic frame is defined as "Action".

The "who" event which corresponds to the subject of the predicate (i.e. ARG0) in the semantic frame is defined as"Agent".

The "what" event which corresponds to the object of the predicate (i.e. ARG1) in the semantic frame, (in other words, "the argument which undergoes the change of state or is being affected by the action" (Bonial *et al.*, 2010)), is defined as "Patient".

The "whom" event which corresponds to the benefactive argument of the predicate (i.e. ARG2) in the semantic frame is defined as "Benefactive".

The "when" event which corresponds to the temporal argument of the predicate (i.e. ARGM-TMP) in the semantic frame is defined as "Temporal".

The "where" event which corresponds to the locative argument of the predicate (i.e. ARGM-LOC, ARGM-DIR) in the semantic frame is defined as "Locative".

The "why" event which corresponds to the cause or purpose argument of the predicate (i.e. ARGM-CAU, ARGM-PRP) in the semantic frame is defined as "Purpose".

Since the "how" event which corresponds to the more detailed modifiers of the predicate, Lo and Wu (2011b) presented experiments on different variants of sub-categorizing the "how" event.

To concretize the lay human annotators' understanding of the role labels, three annotated examples were provided. The examples were shown in the order of advance-

Example 1:
I send a message to you on Saturday at home for making appointment .

Agent:	I
Action:	send
Patient:	a message
Benefactive:	to you
Temporal:	on Saturday
Locative:	at home
Purpose:	for making appointment

Example 2:
A few days ago , the National Development Bank successfully issued 30 billion yen of samurai bonds to Japan 's capital market .

Agent:	the National Development Bank
Action:	issued
Patient:	30 billion yen of samurai bonds
Temporal:	A few days ago
Locative:	Japan 's capital market
Manner:	successfully

Example 3:
South Korea 's Ministry of Agriculture and Forestry said this evening that an Asan City duck farm reported to the relevant department on the 11th that since the 5th of this month , the number of egg production of over 9,000 ducks in the duck farm had fallen sharply .

Agent 1:	South Korea 's Ministry of Agriculture and Forestry
Action 1:	said
Patient 1:	an Asan City duck farm reported to the relevant department on the 11th that since the 5th of this month , the number of egg production of over 9,000 ducks in the duck farm had fallen sharply
Temporal 1:	this evening
Agent 2:	an Asan City duck farm
Action 2:	reported
Patient 2:	since the 5th of this month , the number of egg production of over 9,000 ducks in the duck farm had fallen sharply
Benefactive 2:	the relevant department
Temporal 2:	on 11th
Agent 3:	the number of egg production of over 9,000 ducks in the duck farm
Action 3:	fallen
Temporal 3:	since the 5th of this month
Manner 3:	sharply

Figure 2: Annotated examples

ment of semantic structures. The first two examples contained one predicate only and the last example contained three predicates. Figure 2 shows the three annotated examples provided to annotators to concretize their understanding of the simplified set of semantic role labels.

3.2 Semantic frame annotation web interface

Annotators are allowed to view and annotate only translations that are assigned to them. Therefore, users have to login to the system. A login page is shared by the annotation web interface and the comparison web interface that is introduced in later section. After logging in, annotators are sent to the annotation dataset claiming page, where they can see the list of datasets that is assigned to themselves and the list of datasets that they have already annotated. Figure 3 shows the task claiming page.

Figure 4 shows the annotation page with an annotation in progress. The page can be divided into three panels.

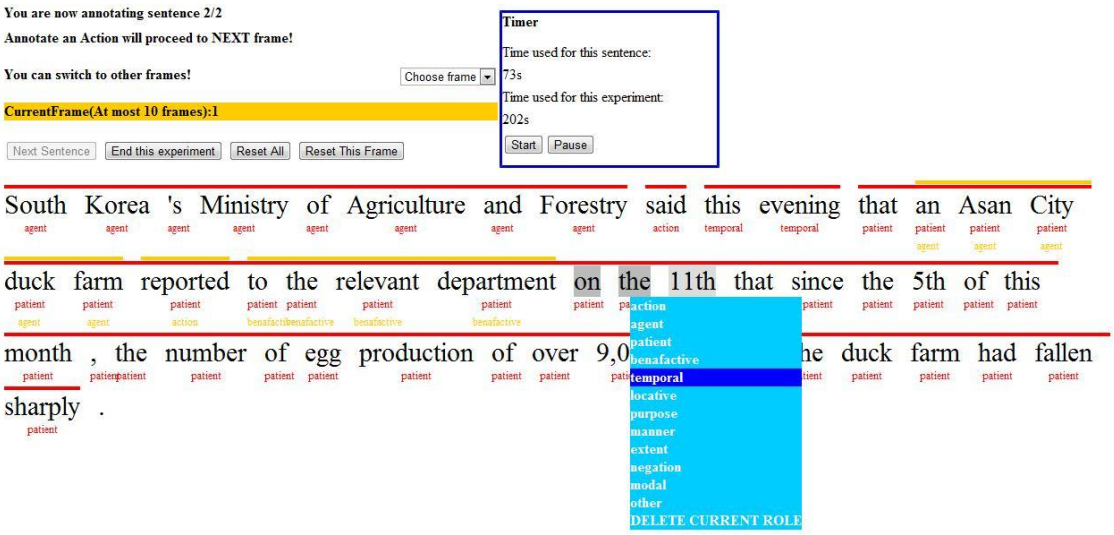

Figure 4: Semantic frame annotation web interface

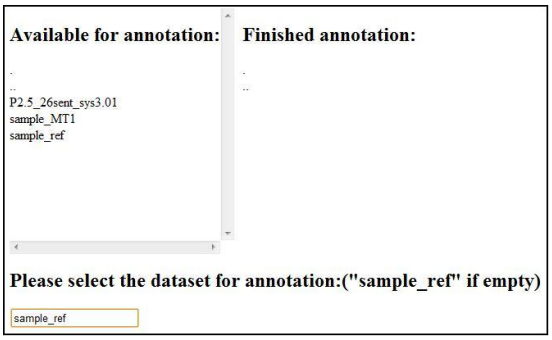

Figure 3: Annotation task claiming page

The top left corner is the information and control panel where annotators receive information about the progress and control the annotation process. The top right corner is the timer panel. The lower panel is the annotation panel.

The first line on the top left corner shows the progress of the annotation task. In this screen shot, the annotator is annotating the second sentence in a dataset of 2 sentences. The next line reminds the annotators to annotate the action first to start annotating a new frame. This is designed according the linguistic formation of predicate-argument structure of semantic frame. The third line on the top left corner reminds the annotators if they find any error in previous annotated frames, they can choose the corresponding frame from the associated combo box at the end of the line. The fourth line is colored to show the annotator clearly which frame they are currently editing. Following the first four lines, there are four buttons. When the annotator finishes annotating the current sentence, he/she should either click "Next Sentence" if there are more sentences in the data set for annotation, or "End this experiment" if there is no more sentence in the data set. "Reset All" allows the annotators to remove all annotations in all frames of the current sentence. "Reset this Frame" allows the annotators to remove all annotations in the current frame.

On the top right corner, there is the timer showing the time used for the current sentence and the current task. There are two buttons in the timer, "Start" and "Pause". The sentence will be covered up if the timer is not started to ensure accurate timing.

The lower half of the page is the annotation panel. The current sentence is shown in the annotation panel. The colored lines above the sentence indicate the span of the semantic role. The colored labels below the sentence indicate the label of the semantic role. One frame is represented by one color. the annotations in all frames are shown to the annotators at the same time in the same panel so that the annotators can see the whole event structure they annotated and verify the annotations easily.

The annotators click on the word token at the begining of a role span and click on the work token at the end of the same role span to specify the span of the semantic role. After that, a pop up menu will be shown to let annotators to determine the role label. After selecting the role label, the pop up menu will be hidden again and the annotators can continue annotating other roles or frames.

4 Aligning/comparing Semantic Frames

After annotating the semantic frames, we must then determine the translation accuracy of the role fillers. To overcome the disadvantages of resorting to excessively permissive bag-of-words matching or excessively restrictive exact string matching, human judges were employed

Evaluating translation accuracy

The goal of this task is the evaluate translation accuracy of the machine translation output action-by-action, role-by-role systematically.

Given a pair of machine translation and human translation that are labeled with semantic roles, please follow the steps below to evaluate the translation accuracy of the machine translation output.

1. Match the labeled "Action" in the machine translation with the labeled "Action" that expresses the same meaning in the human translation.

2. For each matched pair of "Action", align the labeled role fillers in the machine translation with the labeled role fillers that expresses the closest meaning in the human translation. ** Note that each labeled role fillers can ONLY be aligned **ONCE**. **

3. For each pair of aligned role fillers, mark the translation accuracy of the role fillers in the machine translation. The following are the choices of translation accuracy defined in this assignment.

   Correct - Role fillers in MT express the same meaning as the aligned role fillers in the human translation.

   Partial - Role fillers in MT express part of the meaning of the aligned role fillers in human translation. Extra meaning is **NOT** penalized unless it belongs in other role fillers in the human translation.

Figure 5: Semantic frame comparison guidelines for MT evaluation

to evaluate the correctness of each role filler translation between the reference and machine translations. However, with the ultimate goal of automating this step, the definition of translation correctness in meaning must be well-defined. Moreover, to facilitate a finer-grained measurement of translation utility, the definition of translation correctness must also be finer-grained. We present the fine-grained but well-defined choices of translation correctness and minimal guidelines for semantic MT evaluation.

4.1 Fine-grained but well-defined choices of correctness and minimal guidelines

To avoid the inconsistency among human judges , instead of adopting 5-point or 7-point scales used in translation adequacy judgment, we define the translation correctness of role fillers as three cardinal marks, i.e. "correct", "partial" and "incorrect". Since predicate verb is exactly one word, either the machine translation express the same action or not the same action, we only define "correct" and "incorrect' for predicate. Figure 5 shows the fine-grained but well-defined choices of translation correctness.

> Role fillers in MT, that express the same meaning as that in the reference translation, is considered as a "correct" translation.
>
> Role fillers in MT, that express a part of the meanings of that in the reference translation, is considered as a "partial" correct translation. Extra meaning is not penalized unless it belongs in another role.

We also assume that a wrongly translated predicate means that the entire semantic frame is incorrect; therefore, the "correct" and "partial" argument counts are collected only if their associated predicate is correctly translated in the first place.

4.2 Semantic frame comparison web interface

Similar to the annotation web interface, human judges are allowed to view and judge only translations that are assigned to them. After logging in, human judges are sent to the comparison task claiming page, where they can see the list of datasets that is assigned to themselves and the list of datasets that they have already compared.

We assume that a wrongly translated predicate means that the entire semantic frame is incorrect; therefore, human judges are required to pick a pair of correctly translated predicate in the reference translation and the machine translation before judging the translation accuracy of the arguments associated with it. After picking a pair of matched predicates, the annotated machine translation and reference translation are shown to the human judges simultaneously. The reference translation is shown on the left and the machine translation is shown on the right. Human judges will then align each argument in the machine translation with one argument in the reference translation which expresses meaning that is closest to each other and mark the translation correctness of that argument. Figure 6 shows the comparison page when a comparison is in progress.

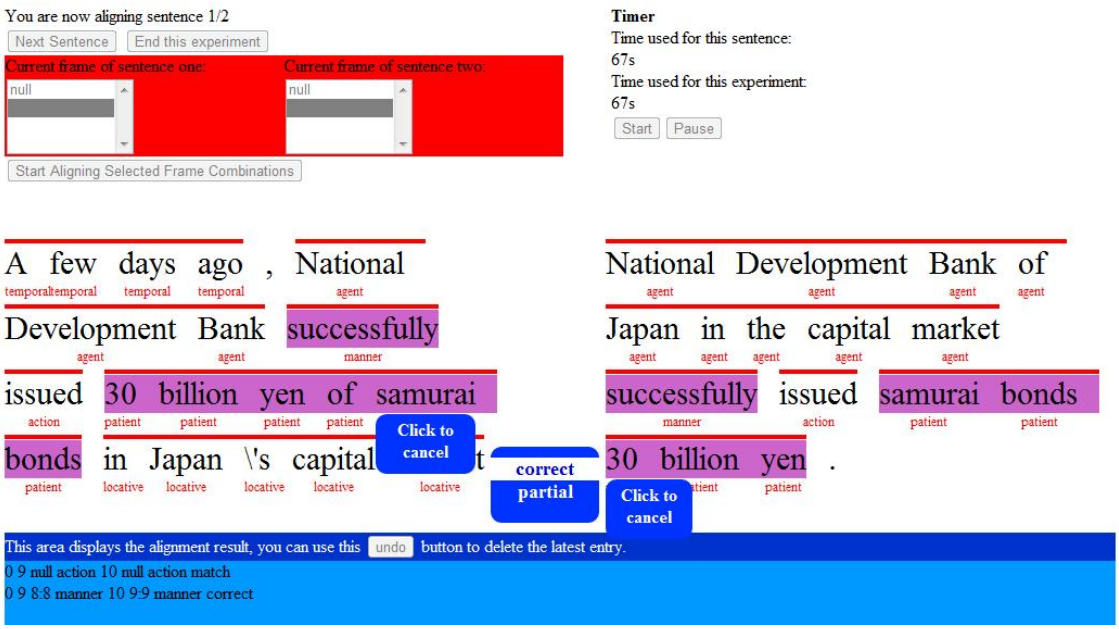

Figure 6: Semantic frame comparison web interface

5 Experiments

To assess the efficiency of the guidelines and the interface, we measured the time required by human judges to perform either the semantic frame annotation and comparison task, on two different data sets.

We also analyzed the inter-annotator agreement to show that despite of the simplicity of the annotation guidelines, the annotators are nevertheless quite consistent to each other.

Lo and Wu (2011a, b, c) have already presented state-of-the-art results in semantic MT evaluation using the proposed methodology. That is, semantic MT evaluation metrics using low-cost lay annotators for semantic frame annotation correlates with human adequacy judgement higher than automatic fluency-oriented metric, BLEU, and non-automatic expensive metric, HTER.

5.1 Setup

We had two set of data samples annotated and compared. Each sample was randomly drawn from a translation evaluation corpus containing Chinese input sentences, English reference translations, and the machine translation outputs from three different state-of-the-art systems. A set of 35 sentences drawn from the subset of the DARPA GALE program Phase 2.5 newswire evaluation dataset in which both the Chinese and English sentences have been annotated with PropBank semantic role labels. Another set of samples was drawn from the NIST MetricsMaTr meta-evaluation dataset (Callison-Burch *et al.*, 2010), with 39 sentences of the broadcast news genre. We employed Chinese-English bilinguals to annotate the semantic roles using the proposed annotation guidelines. Each translation is annotated by at least two annotators to support the consistency analysis.

5.2 Results on efficiency

The collected timing data is detailed in Table 1 in terms of sentences, frames, roles and words. The training on the annotation guidelines and briefing on the graphical user interface require typically 5 to 10 minutes of preparation, at most 15 minutes, including any necessary time for annotators or judges asking questions.

The results bear out the efficiency of our methodology, in spite of the fact that annotation was performed solely by inexpensive computer science undergraduate students with no linguistic background training. The time used for annotating semantic frames averaged about 1-1.5 minutes per sentence, depending on the complexity of the sentences—much less time than required for gold standard Propbank annotation. The time used for comparing the role fillers between the semantic frames in the reference and machine translations, similarly, averaged under 2 minutes per sentence.

Furthermore, note that these timing figures are for completely unskilled non-experts. In fact, the time required tends to decrease even further as annotators gain experience.

5.3 Results on consistency

With the easy-to-use graphical user interface, the annotations from different annotators are even more consistent

Table 1: Timing statistics for human semantic role annotation and role filler comparison tasks, for both the MetricsMaTr and GALE samples. t/s, t/f, t/r, and t/w indicate time per sentence, frame, role, word, respectively.

	#frames	#roles	#words	min t/s	max t/s	avg t/s	avg t/f	avg t/r	avg t/w
MetricsMaTr REF annotation	1.85	6.86	12.69	15.00	485.00	127.12	68.59	18.53	5.01
MetricsMaTr MT annotation	1.39	5.19	10.59	2.00	428	75.94	54.40	14.54	3.49
MetricsMaTr MT comparison	——	——	——	5.00	183	26.75	5.05	1.35	0.33
GALE REF annotation	2.79	11.07	21.44	18.00	416.00	131.30	47.13	11.71	3.06
GALE MT annotation	2.49	7.46	15.53	4.00	376	96.22	38.99	11.03	2.68
GALE MT comparison	——	——	——	9.00	401	141.33	41.61	13.10	4.89

than that reported in Lo and Wu (2011a). The IAA on role identification is 78% for reference translation and 75% for MT output. The IAA on role classification is 70% and 69% for reference translation and MT output respectively. By guiding the annotators step by step through the process of annotation, the IAA on both tasks show a 1-4% improvement from that reported in Lo and Wu (2011a). The high IAA suggests that the simple and intuitive annotation guidelines are in general sufficient for practical application such as semantic SMT and MT evaluation.

6 Web Access to the System

For research uses, please register for the full cloud based interface at `http://www.cs.ust.hk/~dekai/meant`.

7 Conclusion

We have presented a new, radically simple yet effective methodology for inexpensively annotating semantic frames using minimally trained lay annotators, that we believe to be ideal for practical semantic SMT and evaluation applications. Instead of using skilled linguists to annotate gold standard Propbank semantic frame annotation, we showed that annotating semantic frames for MT evaluation can be as intuitive as understanding the basic event structure of a sentence, which any untrained human does naturally and effortlessly. We simplified the annotation guidelines into half a page plus three annotated examples. We described a graphical user interface for both semantic frame annotation and semantic frame alignment/comparison, that guides untrained humans step by step. Restricted guidelines with this easy-to-use GUI allow untrained humans to focus on understanding the translation to provide consistent and efficient annotation and comparison. Our convenient 'cloud' based implementation of this is platform independent, installation-free, and portable as it is developed using technologies supported by any mordern web browser. Thus, we have presented in detail a semantic frame annotation and alignment methodology with minimal labor cost.

References

Claire Bonial, Olga Babko-Malaya, Jinho D. Choi, Jena Hwang, and Martha Palmer. PropBank Annotation Guidelines. In *http://www.ldc.upenn.edu/Catalog/docs/LDC2011T03/propbank/english-propbank.pdf*, 2010.

Chris Callison-Burch, Philipp Koehn, Christof Monz, Kay Peterson, Mark Pryzbocki, and Omar Zaidan. Findings of the 2010 Joint Workshop on Statistical Machine Translation and Metrics for Machine Translation. In *Proceedings of the Joint 5th Workshop on Statistical Machine Translation and MetricsMATR*, pages 17–53, Uppsala, Sweden, 15-16 July 2010.

Jinho D. Choi, Claire Bonial, and Martha Palmer. Propbank Instance Annotation Guidelines Using a Dedicated Editor, Jubilee. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC2010)*, 2010.

Chi-Kiu Lo and Dekai Wu. MEANT: An Inexpensive, High-Accuracy, Semi-Automatic Metric for Evaluating Translation Utility based on Semantic Roles. In *Proceedings of the Joint conference of the 49th Annual Meeting of the Association for Computational Linguistics : Human Language Technologies (ACL-HLT-11)*, 2011.

Chi-Kiu Lo and Dekai Wu. SMT vs. AI redux: How semantic frames evaluate MT more accurately. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)*, 2011.

Chi-Kiu Lo and Dekai Wu. Structured vs. Flat Semantic Role Representations for Machine Translation Evaluation. In *Proceedings of the 5th Workshop on Syntax and Structure in Statistical Translation (SSST-5)*, 2011.

Sameer Pradhan, Wayne Ward, Kadri Hacioglu, James H. Martin, and Dan Jurafsky. Shallow Semantic Parsing Using Support Vector Machines. In *Proceedings of the 2004 Conference on Human Language Technology and the North American Chapter of the Association for Computational Linguistics (HLT-NAACL-04)*, 2004.

Dekai Wu and Pascale Fung. Can Semantic Role Labeling Improve SMT? In *Proceedings of the 13th Annual Conference of the European Association for Machine Translation (EAMT-2009)*, pages 218–225, Barcelona, Spain, May 2009.

Dekai Wu and Pascale Fung. Semantic Roles for SMT: A Hybrid Two-Pass Model. In *Proceedings of the 2009 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT-09)*, pages 13–16, 2009.