Proceedings of the Australasian Language Technology Workshop 2007, pages 83-91

Two-Step Comprehensive Open Domain Text Annotation with Frame Semantics

Bahadorreza Ofoghi
Centre for Informatics and Applied Optimization, University of Ballarat
bofoghi@students.ballarat.edu.au

John Yearwood
Centre for Informatics and Applied Optimization, University of Ballarat
j.yearwood@ballarat.edu.au

Liping Ma
Centre for Informatics and Applied Optimization, University of Ballarat
l.ma@ballarat.edu.au

Abstract

With shallow semantic parsing tasks receiving more attention in many natural language applications, there is a need for labeled corpora for learning the specific tags under consideration. In this paper, we discuss a two-step semantic class and semantic role assignment based on the FrameNet elements over a subset of the AQUAINT collection with a reasonable coverage over the semantic frames in FrameNet. The quality of the annotation task is examined through inter-annotator agreement. The methodology described in this work for measuring inter-annotator agreement can be adapted for similar tasks. Some central aspects of the task are also detailed in this paper.

1 Introduction

Open domain text labeling with the tags of lexical semantic structures is a time consuming intensive task that is necessary to initiate the semi-(or fully-) automated process of text annotation with lexical semantic information. The extra knowledge that such labeling can contribute to texts develops a higher level of text understanding for high level applications in the domain of natural language processing e.g. question answering (QA), information extraction (IE), machine translation, and etc.

The learning parsers which can afford the automated semantic labeling of texts are highly dependent on the training sets with example annotated texts. The existing semantic parsers, better known in this context as *shallow semantic parsers*, rely on the syntactic and/or semantic similarities of the training sets and the test sentences (syntactical and semantic features) (Erk and Pado 2005) (Pradhan, Ward et al. 2004). The SHALMANESER parser (Erk and Pado 2006), as one of such parsers, benefits from two classifiers to add the FrameNet (Baker, Fillmore et al. 1998) Frames and Frame Elements (FEs) to open domain texts. It is mostly dependent on the syntactic features in the texts. The ASSERT shallow semantic parser (Pradhan, Ward et al. 2004) exploits the SVM classifiers trained on both syntactical and semantic features of texts to assign semantic roles to sentence elements (i.e. arguments).

One of the first issues that highly affect the accuracy of shallow semantic parsers is the quality of the training set. To avoid any unnecessary behavioral bias towards the annotation applied by a coder in the training set, it is prudent that the training annotation passes a reasonable inter-annotator agreement measure with respect to the different aspects of the labeling task.

In annotating a text collection with Frame Semantics (Fillmore 1976), the two main aspects are considered to be the frame and FE assignment to the text. While the first aspect is more a matter of word sense disambiguation (Erk 2004), the latter is a challenge with syntactic and semantic analysis of the constituents of a sentence and boundary detection for each argument of a predicate to be assigned to a semantic role (i.e. an FE).

In this paper, we describe the activity of annotating a subset of the AQUAINT collection with the

FrameNet frames and their corresponding FEs. The three major motivations of this task are as follows:

1- The labeled corpus can be exploited for the purpose of training shallow semantic parsers as it contains appropriate semantic class and role assignment to the text,
2- The methodology of measuring the inter-annotator agreement with regard to the tasks of frame assignment and FE labeling has not been tackled and formulated comprehensively before and our methodology can be adapted for similar tasks,
3- As the corpus contains the answer passages for a large amount of the TREC 2004 factoid question set (Voorhees 2004), it can be used for answer extraction technique learning and articulating the FrameNet-based answer candidate identification.

The task of adding frames and FEs to the texts has been performed in the two steps of *automated annotation* and *human expert augmentation*.

This paper is organized as follows. A brief definition of frame semantics in section 2 is followed by the description of the FrameNet-based text labeling in section 3. Section 4 analyzes the different aspects of our annotation task and section 5 concludes the paper.

2 Frame Semantics

Frame Semantics, basically developed from Charles Fillmore's Case Structure Grammar (Fillmore 1968) (Cook 1989), emphasizes the continuities between language and human experience (Fillmore 1976) (Lowe, Baker et al. 1997) (Petruck 1996). The main idea behind frame semantics is that the meaning of a single word is dependent to the essential knowledge related to that word. With such an understanding of frame semantics, the required knowledge about each single word is stored in a *frame*. In order to encapsulate frame semantics in such frames, the FrameNet project (Baker, Fillmore et al. 1998) has been developing a network of inter-related frames which is a lexical resource for English now used in many natural language applications.

The main entity in FrameNet is the frame which develops a kind of semantic normalization over concepts semantically related to each other. The semantic relation between concepts in a frame is realized with regard to the scenario of a real situation which may happen and cover the participant concepts rather than synonymy or other peer-to-peer relations. In this regard, the frames encode the base definitions necessary to understand the semantics and the scene of each contained term. In other words, real-world knowledge about real scenarios and their related properties are encoded in the frames (Lowe, Baker et al. 1997). Each frame contains some *frame elements* (FEs) as representatives of different semantic and syntactic roles regarding a target word inside the frame. The semantic roles are common properties among all of the terms that are inherited from a frame. This ensures a suitable inclusion over the English terms which either have similar meanings or share the context and/or the scenario in which they could occur in the sentences of the language.

The current version of the FrameNet data (release 1.3) contains about 795 semantic frames and more than 135,000 annotated sentences from the British National Corpus.

3 FrameNet-based Text Annotation

The task of shallow semantic parsing of sentences with respect to the predicates (e.g. verbs, nouns, etc.) mainly consists of two phases: i) sense disambiguation of the predicate, and ii) role assignment to the arguments of the predicate with regard to the specific sense (i.e. class) of it (Erk and Pado 2006). In the context of FrameNet, the class is realized as the specific frame which is evoked in the true sense of the context of the sentence (Erk 2004), while the roles are the different FEs in that frame.

Frame: Manufacturing		
Definition	A **Manufacturer** *produces* a **Product** from **Resource** for commercial purposes.	
FEs	FACTORY	Thos machines were *manufactured* **in the Miami plant**.
	MANUFACTURER	**General Electric** *produces* electric appliances.
	PRODUCT	The company *manufactured* **many T-shirts**.
Lexical units	fabricate.v, fabrication.n, industrial.a, make.v, maker.n, …, production.n	

Table 1. An example frame in FrameNet

Table 1 shows an example frame "Manufacturing" with its definition, core FEs, and lexical units. The main semantic roles that are necessary for the scenario to be complete are those known as the core FEs Factory, Manufacturer, and Product. Different predicates with different parts-of-speech (e.g.

noun, verb, and adjective) are inherited form this frame.

With respect to this frame, the annotation of an example sentence "*the company makes different types of doors in the Australia plant*" with respect to the predicate "make.v" consists of two stages: i) to identify the right semantic class (i.e. frame), and ii) to assign the different parts of the sentence to the semantic roles (i.e. FEs of the frame). Figure 1 visualizes this annotation task. There are different semantic classes that can be realized with the predicate "make.v" like arriving, building, cooking-creation, causation, manufacturing, etc. The task of finding the right frame from this set of related semantic classes is a problem formulated as word sense disambiguation in (Erk 2004). Having the right frame identified, the semantic role assignment to connect the FEs and the sentence constituents is the next challenge. As already mentioned, there are different studies which attack the problem using syntactic and/or semantic features in the text.

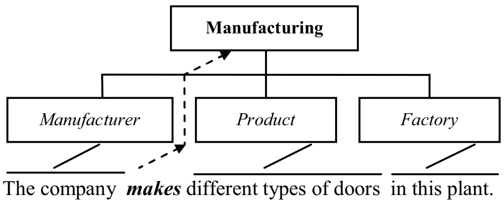

Figure 1. Shallow semantic analysis of an example sentence

The semantic annotation of the sentence in Figure 1, however, is not complete with regard to all of the existing predicates in the sentence. There are other frame-evoking elements: "company", "different", "types", "doors", and "plant" each of which can evoke at least a single frame in FrameNet. Figure 2 shows a complete annotation of frame evocation.

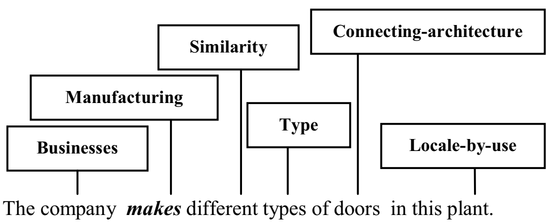

Figure 2. Complete frame annotation of an example sentence

The semantic annotations with respect to the FEs are eliminated in the figure above for readability, though in a complete annotation they need to be properly assigned to the corresponding sentence segments.

There are more than 135,000 annotated sentences in the FrameNet database. However, the standard of the annotations is different from that in Figure 2. The annotations in this FrameNet database are frame-oriented which result in annotating sentences per predicate inherited from the current frame. Consequently, there is not more than one frame evoked per example sentence.

In the task of automated shallow semantic parsing the training phase is vital. Most existing shallow semantic parsers are trained with the data provided in the FrameNet database. We believe that there are two main drawbacks on this training paradigm:

1- The concurrence of the frames is ignored as a meta semantic feature that could potentially be informative in correct semantic class identification,
2- The example sentences annotated per frame are semantically focused on the class covered by the frame.

The former could easily affect the context analysis of the sentences as a potentially useful feature that has to be eliminated in the training phase with the current FrameNet annotations. The latter produces bias in the classifiers which are supposed to be effective on the open domain text with lots of concurrent semantic scenarios.

We have been providing a comprehensive annotation with all of the frames and FEs evoked in an exhaustive manner to overcome the two abovementioned shortcomings.

4 Comprehensive Text Annotation

With the first aim of training a frame-semantics-based answer extraction module for QA, we have comprehensively annotated a text collection. The output, which will be publicly available for research purposes, can be exploited for other natural language learner systems. The different aspects of the annotation task are explained in the following sub-sections.

4.1 Data

A subset of the AQUAINT text collection[1] which contains the news articles from the New York Times News Service (1998-2000), Xinhua News Service (1996-2000), and Associated Press Worldstream News Service (1998-2000) has been annotated using the method explained in section 4.2. The statistical information about the annotated corpus is summarized in Table 2.

Element	Measure
# Passage	1379
# Sentence	3451
Ave. # sentences per passage	2.502
# Single word	89434
Ave. # single words per sentence	25.915
# Single word (unique)	9291
# Predicate	53215
# Predicate (unique)	8121

Table 2. Annotated corpus statistics

The 1379 passages have been extracted in response to the information request of a subset of the TREC 2004 factoid questions including 143 questions (out of the entire set of 230 factoid questions) for which the retrieval systems retrieves passages actually containing the correct answers. The limitation for the task of passage retrieval was set to retrieve the top 10 passages per question. For a few questions, the retrieval system could not retrieve exactly 10 passages (in some occasions less number of passages) as there was not enough information text in the collection specifically related to the question. The modified version of the MultiText passage retrieval algorithm (Ofoghi, Yearwood et al. 2006) has been used for this purpose which interprets the passages as variable-sized strings starting and ending with pairs of query keywords at any position in the corpus documents.

4.2 Method

The annotation task of the corpus has been performed in a two-step process including *automated annotation* and *manual augmentation*.

In the automated annotation, the SHALMANESER shallow semantic parser (Erk and Pado 2006) version 1.0 has been used. The instance of SHALMANESER used in the task, benefits from the two learner classifiers FRED and ROSY trained with the FrameNet data release 1.2. The FRED classifier identifies the semantic class (i.e. the frame) of the frame-evoking elements while the ROSY classifier assigns segments of the sentences to the semantic roles (i.e. the FEs). According to the evaluations in (Erk and Pado 2006), FRED performs relatively better than the ROSY system in terms of precision and recall.

Step	System	Version
POS-tagging	TNT	2
Lemmatization	TreeTagger	-
Syntactic Parsing	Collins' Parser	1.0
Machine learning	Mallet	mallet 0.4

Table 3. SHALMANESER settings at each step

As SHALMANESER is a loosely coupled tool chain for automated annotation, there are different learner systems that are supported by the tool and can be exploited at its different modules. Table 3 shows the different systems that we have used for each task in the SHALMANESER settings.

To enhance the semantic class and role labeling accuracy on the output SALSA/TIGER xml files (Erk and Pado 2004), an intensive manual augmentation over the automated outputs has been conducted using the SALSA annotation tool known as SALTO (Burchardt, Erk et al. 2006).

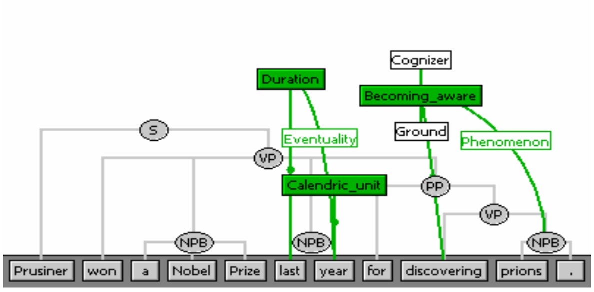

Figure 3. Incomplete automated annotation of an example sentence

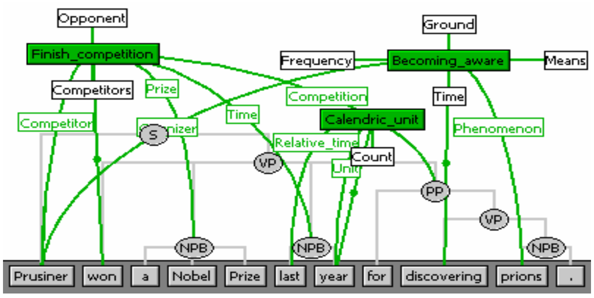

Figure 4. Comprehensive annotation of an example sentence after manual augmentation

The task of manual annotation (i.e. augmentation of the automated annotation) is an exhaustive process which thoroughly examines each sentence word-by-word. It includes:

[1] http://www.ldc.upenn.edu/Catalog/docs/LDC2002T31/

- *Frame evocation*: if a predicate could have evoked a correct frame in FrameNet with respect to the sense of the predicate,
- *Frame change*: in case the frame already evoked (by SHALMANESER/FRED) is not of the correct semantic class of the predicate,
- *FE assignment*: when parts of a sentence could have been assigned to FEs,
- *FE assignment correction*: where there is a need for changing the connectivity of the sentence segments to the FEs of a frame (as indicated by SHALMANESER/ROSY).

Figure 3 shows an example sentence annotated by SHALMANESER visualized in SALTO. Figure 4 shows that after the task of manual augmentation.

In the manual annotation process of the example sentence, the frame "Finish-competition" has been added, the wrongly assigned frame "Duration" has been eliminated, and the FEs of the two frames "Calendric-unit" and "Becoming-aware" have been corrected in their corresponding sentence segments.

The manual annotation, in order to develop the most comprehensive and up-to-date annotation, uses the FrameNet data version 1.3 with 795 semantic frames.

4.3 Statistics of Annotation

The manual annotation process includes changing many of the frames and FEs assignments. To have a better picture of the task, the two subtasks of the manual augmentation (i.e. the frame changing and the FEs corrections) are separately analyzed in a statistical approach as shown in Table 4.

Frame level	Frames changed	FEs changed
SHALMANESER evoked frames	N/A	35.741
Manually augmented frames – FN1.2	59.926	158.356
Manually augmented frames – FN1.3	74.160	181.860

Table 4. Average number of frames and FEs changed in manual augmentation per 10 passages (per question) – raw measures

As shown in Table 4, the FE-oriented augmentation of the frames evoked by SHALMANESER includes changing of the 35.741 FEs of those frames on average, where there are no frames added. In the two next levels, there are frames added with respect to the two versions of the FrameNet data. It is obvious that with respect to the FrameNet dataset 1.3, there are more frames and FEs that can be added to the text.

Table 5 translates the measures in Table 4 when normalized against the sentences.

Frame level	Frames changed	FEs changed
SHALMANESER evoked frames	N/A	1.560
Manually augmented frames – FN1.2	2.547	6.844
Manually augmented frames – FN1.3	3.182	7.848

Table 5. Average number of frames and FEs changed in manual augmentation per 10 passages (per question) – sentence number normalized

To have more FrameNet-oriented statistical sense of the annotation, Table 6 shows the overall measures with respect to the frames and FEs assigned to the corpus sentences.

Element	Measure
# Frames evoked	21741
# Frames evoked (unique)	592
# FEs assigned	40589
# FEs assigned (unique)	2586

Table 6. FrameNet-oriented statistics of the annotated corpus

The total number of unique frames evoked in the corpus, i.e. 592, covers 74.465% of the total frames in the FrameNet data release 1.3 containing 795 semantic frames. On the other hand, the overall frame count on the corpus (i.e. 21741) represents the concurrency rate of the frames over the sentences as 6.299 frames per sentence on average.

Having all this statistical information about the manual annotation of the SHALMANESER outputs, we have analyzed the accuracy of SHALMANESER and the other levels of annotation against the human level annotation. Table 7 considers the baseline human level augmentation with respect to the FrameNet data 1.2 and Table 8 shows the average accuracy of the tasks with taking the FrameNet data 1.3 frames into account in the baseline human level annotation.

$$Accuracy = \frac{\#correct\_items}{\#items} \quad (1)$$

In order to calculate the accuracy of the labeling task at each level, the percentage of correct frames and/or FEs assigned at each level is measured over the total number of items (i.e. frames or FEs) assigned at the highest level of assignment (i.e. the human level). Equation 1 shows the formulations to measure the accuracies over the item (i.e. frame and FE) assignment at each level where *correct_items* refers to the items correctly assigned at the level under consideration and *items* is the whole set of assignments at that level. This formula is used for both frames and FEs.

As shown in Table 7 and Table 8, the overall accuracy of the fully automated shallow semantic parsing on the open domain texts of the AQUAINT collection is not promisingly high. The task of FE assignment, especially, seems to be a challenging process where the overall accuracy is not reaching more than 17.000%.

The low performance of the automated shallow semantic parsing is an evidence for the need for more comprehensively labeled training sets of text annotations. We believe that SHALMANESER could have achieved much higher accuracies if it had been trained against a more extensive training set, though the classification features are of importance as well. Our work is in line with this requirement (i.e. a comprehensive labeled training set with concurrent frames evoked) for current and future semantic parsers.

Frame level	Frame evocation accuracy	FE assignation accuracy
SHALMANESER evoked frames	41.765%	17.000%
SHALMANESER evoked frames – FEs augmented	41.765%	43.539%
Manually augmented frames – FN1.2	100%	100%

Table 7. Average accuracy of annotation at each frame level – FN 1.2 in baseline human level annotation

Frame level	Frame evocation accuracy	FE assignation accuracy
SHALMANESER evoked frames	38.003%	15.665%
SHALMANESER evoked frames – FEs augmented	38.003%%	40.042%
Manually augmented frames – FN1.3	100%	100%

Table 8. Average accuracy of annotation at each frame level – FN 1.3 in baseline human level annotation

4.4 Quality

An important aspect of the manual augmentation is the quality of the output annotation with respect to the two main subtasks namely frame evocation and FE assignment to the sentence segments. The manual augmentation process, in our work, has been conducted by one coder; however, there has been a method for validating the output annotation with respect to the inter-annotator agreement rates.

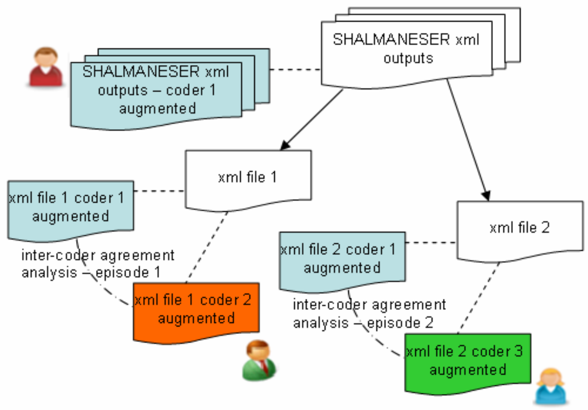

Figure 5. Inter-annotator agreement analysis scenario

After finishing the manual augmentation task by the sole coder, two separate portions (10 passages each) of the same SHALMANESER outputs (not the whole set) have been annotated by two other coders (three coders in total). Each portion is then augmented by a coder i.e. portion 1 by coder 2 and portion 2 by coder 3. With this setting, there are two portions annotated by two coders where the pairs are coder1-coder2 and coder1-coder3. In two separate episodes, the inter-annotator agreement has been measured in the sense of frame evocation and FE assignation. Figure 5 depicts the scenario.

The overall agreement has been then calculated as the average values on the two measure sets.

The alpha statistics has been used in other similar tasks for frame agreement calculation between annotators (Erk, Kowalski et al. 2003). In this task, we use the Kappa statistics (Cohen 1960) as shown in Equation 2 where *P(A)* indicates the observed agreement among the coders (i.e. the probability of the agreed items over the total number of items coded) and *P(E)* is the expected agreement.

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)} \quad (2)$$

The computation of *P(E)* as the probability of agreement among coders by chance is the challenging part in the Kappa statistics which can be approached in different ways. We benefit from the Siegel and Castellan's agreement table (Eugenio and Glass 2004) to compute the expected agreement *P(E)*.

$$P(E) = \sum_j \left(\frac{\sum_i n_{ij}}{Nk}\right)^2 \quad (3)$$

Equation 3 shows how they calculate the *P(E)* measure for any number of possible labels, where *N* is the total number of observations, *k* is the total number of labels that coders can assign to each item, and n_{ij} is the number of codings of label *j* to item *i*.

For each predicate in the corpus (i.e. items), we consider 4 labels *no-frame*, $frame_{coder1}$, $frame_{coder2}$, and *frame*, where *no-frame* is used for the predicates that are not assigned to any frame by the coders, $frame_{coder1}$ indicates that a frame has been chosen by the coder1 which is not the same as the frame selected by the coder2 as $frame_{coder2}$ indicates. In cases that the coders agree on the same frames, the tag *frame* is chosen for both coders. Table 9 depicts an example agreement table according to 10 predicates {P1, P2, …, P10}.

Predicate	No-frame	$Frame_{coder1}$	$Frame_{coder2}$	Frame
P1	1	1	0	0
P2	1	1	0	0
P3	1	0	1	0
P4	2	0	0	0
P5	2	0	0	0
P6	2	0	0	0
P7	0	1	1	0
P8	0	0	0	2
P9	0	0	0	2
P10	0	0	0	2

Table 9. An example frame agreement table for 10 predicates and 2 coders

With this example agreement table, $K_{S\&C}$ is calculated as follows. First, *P(A)* is calculated as 6/10=0.600 (at 6 rows there are agreements indicated by the number 2). Second, for each label *j*, p_j (i.e. the proportion of predicates assigned to label *j*) is calculated using the formula in Equation 4.

$$p_j = \frac{1}{Nk} \sum_i n_{ij} \quad (4)$$

With k=4 and N=10, we have $p_{no\text{-}frame}$=0.225, $p_{frame\text{-}coder1}$=0.075, $p_{frame\text{-}coder2}$=0.050, and p_{frame}=0.150. Having these values per label, the overall *P(E)* is equal to 0.079. Finally, the $K_{S\&C}$ measure is (0.600-0.079)/(1-0.079)=0.565.

There are different possibilities for measuring the frame evocation agreement with regard to the total number of predicates (i.e. *N*). We calculate the agreement with respect to three predicate counts: i) all predicates in the corpus, ii) all FEEs (frame evoking elements) from the coders' point of view, and iii) all FEEs from the FrameNet point of view. The second set is the maximum number of FEEs identified by either coder, where the third set contains all of the predicates which inherit from an existing frame in FrameNet.

The inter-annotator agreement on the FE assignment task is, however, more problematic due to a few challenges:

- The different coders may assign slightly different string segments to the same FEs as there is no boundary detection performed to identify and unify the set of arguments in the sentences prior to the manual annotation,
- The task of comparison between the FEs assigned by the two coders is not very well addressed as it is not obvious which FEs need to be aligned,
- The total number of FEs over which the agreement is calculated is not constant. That is, the identification of a baseline set of the FEs to calculate the agreement on is a challenge.

Analysis episode	Frame agreement		
	All predicates	FEEs – coders' view	FEEs – FN view
coder1-coder2	0.804	0.387	0.661
coder1-coder3	0.789	0.378	0.708
Average agreement	*0.796*	*0.382*	*0.684*

Table 10. Inter-annotator frame agreement rates

With respect to the above-mentioned challenges, we have set different measurement strategies for FEs agreement calculation. We consider both exact and partial matches between the instances (arguments) assigned to the FEs. On the other hand, we consider two overall sets of FEs to calculate the agreement over: i) the union set of the FEs assigned by the two coders, and ii) the maximum set (i.e. number) of the FEs assigned by either coder. The method of calculation of the FE agreement is based on the percentage of the agreed FEs over the total number of FEs according to one of the two overall sets mentioned above.

Frame level	FE agreement (%)			
	Exact match		Partial match	
	Max	Union	Max	Union
coder1-coder2	17.100	14.420	25.278	21.316
coder1-coder3	29.032	31.629	36.363	39.616
Average agreement	*23.066*	*23.024*	*30.820*	*30.466*

Table 11. Inter-annotator FE agreement rates

Table 10 and Table 11 summarize the two episodes of the agreement analysis for the frame and FE agreement in the annotations. We expect that the calculated agreements over the sub-corpora can be generalized to the whole set of annotation.

The overall agreement on frame evocation for the predicates is much higher than that of the FE assignments to the text of the corpus. This was expected though not such a drastic difference as in the two tables. We believe that the low FE agreement is due to three main reasons: i) different coders' skills on the annotation task results in different standards of annotation which damage the FE assignment task more than the frame evocation process. This happens as the total number of FE assignations is much more than that in terms of frames, ii) different coders' knowledge in frame semantics and more specifically in FrameNet initiates different understandings of the annotation task. Once again, this is more strongly affecting the FE assignment task as there are lots of FEs with different definitions in FrameNet, and iii) dissimilar interpretations of the sentences and clauses by the coders yield an undesired difference in annotations.

5 Conclusion

In line with the current trend for natural language applications based on a high level of semantic information, the necessity of the labeled training sets which include reasonable amounts of annotation has emerged. In this paper, the different aspects of a comprehensive open domain text annotation task are described. We have performed an exhaustive annotation on a subset of the AQUAINT collection which can be used for both shallow semantic parser training and answer extraction module tuning in QA systems that work on the basis of semantic class identification and role labeling approaches. The annotated corpus contains the tags of frames and FEs of FrameNet with a reasonable coverage over the total number of the semantic frames in the FrameNet dataset 1.3 (see section 4.3).

We have approached the comprehensive annotation in a two-step process of automated shallow semantic parsing and manual augmentation of the outputs of the first sub-process. We have shown the different statistical information of the corpus and its annotated version which will be publicly available soon. In addition, the complete inter-annotator agreement calculation methodology has been detailed with respect to the two main subtasks of FrameNet-based annotation (i.e. frame agreement and FE agreement). One of the next goals in this direction is to study and formulate methods to reach higher levels of inter-annotator agreement, especially with respect to the task of FE assignments. We believe that the methodological attributes of our task can shed more light on the similar activities and their challenges. We may re-conduct the inter-annotator agreement calculation process with the annotators with more specific knowledge in the FrameNet concepts.

Acknowledgment

We would like to thank Katrin Erk from the SALSA project, Saarland University, for providing us with SHALMANESER and her assistance with SHALMANESER and SALTO.

References

Baker, C. F., C. J. Fillmore and J. B. Lowe. 1998. *The Berkeley FrameNet project*. In *Proceedings of the International Conference on Computational Linguistics*, pp. 86 - 90.

Burchardt, A., K. Erk, A. Frank, A. Kowalski and S. Pado. 2006. *SALTO -- a versatile multi-level annotation tool*. In *Proceedings of the LREC-06*, 2006.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **20**: 37-46.

Cook, W. A. (1989), *Case Grammar Theory*, Georgetown University Press.

Erk, K. 2004. *Frame assignment as word sense disambiguation*. In *Proceedings of the 2006 IAENG International Workshop on Computer Science, IWCS'06*, Tilburg University, Tilburg, The Netherlands, 2004.

Erk, K., A. Kowalski, S. Pado and M. Pinkal. 2003. *Towards a Resource for Lexical Semantics: A Large German Corpus with Extensive Semantic Annotation*. In *Proceedings of the ACL-03*, Sapporo, Japan, 2003.

Erk, K. and S. Pado. 2004. *A powerful and versatile XML format for representing role-semantic annotation*. In *Proceedings of the LREC-04*, Lisbon, Portugal, 2004.

Erk, K. and S. Pado. 2005. *Analysing models for semantic role assignment using confusability*. In *Proceed-*

ings of the HLT/EMNLP-05, Vancouver, B.C., Canada, 2005.

Erk, K. and S. Pado. 2006. *SHALMANESER - a toolchain for shallow semantic parsing*. In *Proceedings of the LREC-06*, 2006.

Eugenio, B. D. and M. Glass. 2004. The Kappa Statistics: A Second Look. *Association for Computational Linguistics* **30**(1): 95-101.

Fillmore, C. J. 1968. The case for case. *Universals in Linguistic Theory*: 1-88.

Fillmore, C. J. 1976. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech* **280**: 20-32.

Lowe, J. B., C. F. Baker and C. J. Fillmore. 1997. *A frame-semantic approach to semantic annotation*. In *Proceedings of the SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How?*, April 4-5, 1997.

Ofoghi, B., J. Yearwood and R. Ghosh. 2006. *A semantic approach to boost passage retrieval effectiveness for question answering*. In *Proceedings of the 29th Australian Computer Science Conference*, January 16-19, 2006. pp. 95-101.

Petruck, M. R. L. (1996), *Frame semantics*. Philadelphia, John Benjamins.

Pradhan, S., W. Ward, K. Hacioglu, J. Martin and D. Jurafsky. 2004. *Shallow semantic parsing using support vector machines*. In *Proceedings of the Human Language Technology Conference/North American chapter of the Association for Computational Linguistics annual meeting (HLT/NAACL-2004)*, Boston, May 2-7, 2004.

Voorhees, E. M. 2004. *Overview of the TREC 2004 question answering track*. In *Proceedings of the Thirteenth Text Retrieval Conference (TREC-2004)*, 2004.