Striking a Balance: Human and Computer Contributions to Learning through Semantic Analysis

Jesse English and Sergei Nirenburg
Department of Computer Science and Electrical Engineering
University of Maryland, Baltimore County
Baltimore, Maryland 21250

***Abstract*—Manual acquisition of high-quality, broad-coverage knowledge needed by knowledge-based NLP systems is commonly considered too expensive a procedure, and has been known to cause ”the knowledge acquisition bottleneck”. The use of the web as a corpus to support automating knowledge acquisition has been gaining in popularity in the recent years. This approach tends to introduce noise at the early stages of learning which can have a compounding impact on the quality of the final results. If the goal is to alleviate the expense of manual knowledge acquisition in the short term, a combination of automatic knowledge learning and human validation / correction must be considered. People can either post-edit automatically produced candidate knowledge elements or intervene at various stages in the acquisition process to facilitate high-quality automatic output. In this paper, we report on a sequence of experiments analyzing the utility of the latter methodology in the framework of a mutual-bootstrapping environment in which new knowledge resources are acquired both for and through the operation of an automatic meaning extraction system.**

I. Introduction

Automating acquisition of ontologies and lexicons is an important task aimed at overcoming the notorious knowledge acquisition bottleneck problem of AI. The traditional approaches to manual knowledge acquisition have been amply criticized and even dismissed out of hand as infeasible (see, e.g., [1]). Many experiments have been carried out with the aim of alleviating this bottleneck. Existing NLP tools have been used to assist in the acquisition of ontological and lexical knowledge (e.g., by [2] and [3] among many others). Typically, language learning uses fully automatic but rather limited-depth language processing approaches, concentrating on corpus-based heuristics (e.g., [4] or [5], again, among many others). Our approach to this task follows a mutual bootstrapping methodology in which an existing semantic text analyzer is used to support continuous acquisition of ontological and lexical knowledge while at the same time benefiting from the continuous enhancements to its ontology and lexicon as a result of this learning process. Figure 1 illustrates the above process.

We have been experimenting with both fully automatic and human-validated learning environments. To use terminology from machine translation, we used both the postediting approach, validating automatically generated knowledge elements (a methodology also used by, among others, [6] and [7]) and the pre-editing approach in which the quality of inputs to the learning process were validated by people. However, the latter method differeed from the almost universally accepted practice of hand-annotating corpora in that the annotations themselves were originally generated automatically with subsequent validation by people. In this paper we concentrate on learning ontological concepts and present preliminary results on a small test set. We must also note that the use of the web as a corpus is has become quite unexceptional in the last five or so years (see, e.g., http://webascorpus.sourceforge.net/ for more information). This paper is organized as follows. In Section 2, we present OntoSem, our semantic analysis engine and its static knowledge resources; in the Section 3 we introduce our learning environment; and in Section 4 we describe the sequence of experiments on isolating each phase of the learner and compare the results of the individual experiments.

II. OntoSem

Basic semantic analysis in the Ontosem environment uses the information (mutual constraints) in lexicon entries, the ontology and the results of pre-semantic processing to carry out word sense disambiguation and establish basic semantic dependencies in the input text (e.g., [8], [9]). Extended semantic analysis includes such advanced-reasoning tasks as reference resolution (e.g., [10], [11]), processing inputs containing out-of-vocabulary lexical items, sentence fragments (e.g., [12]), non-literal language (e.g., metaphors [13]) and indirect speech act analysis (e.g., *It is cold in here* can be a request to close the window). Extended semantic analysis is carried out by procedural semantic routines that we call *meaning procedures* (e.g., [11]), the algorithmic substrate for which is recorded in evolving microtheories of each of the phenomena in question. The result of the semantic analysis of an input is a *text meaning representation (TMR)* that is written in the same metalanguage used for all knowledge representation and reasoning in our environment. As an example, consider the TMR for the sentence *Do you have chest pain?*, shown below in an abbreviated, pretty-printed format. For reasons of space, we only briefly point out some highlights.

```
REQUEST-INFO-1
   from-sense        *question-mark*-punct1
```

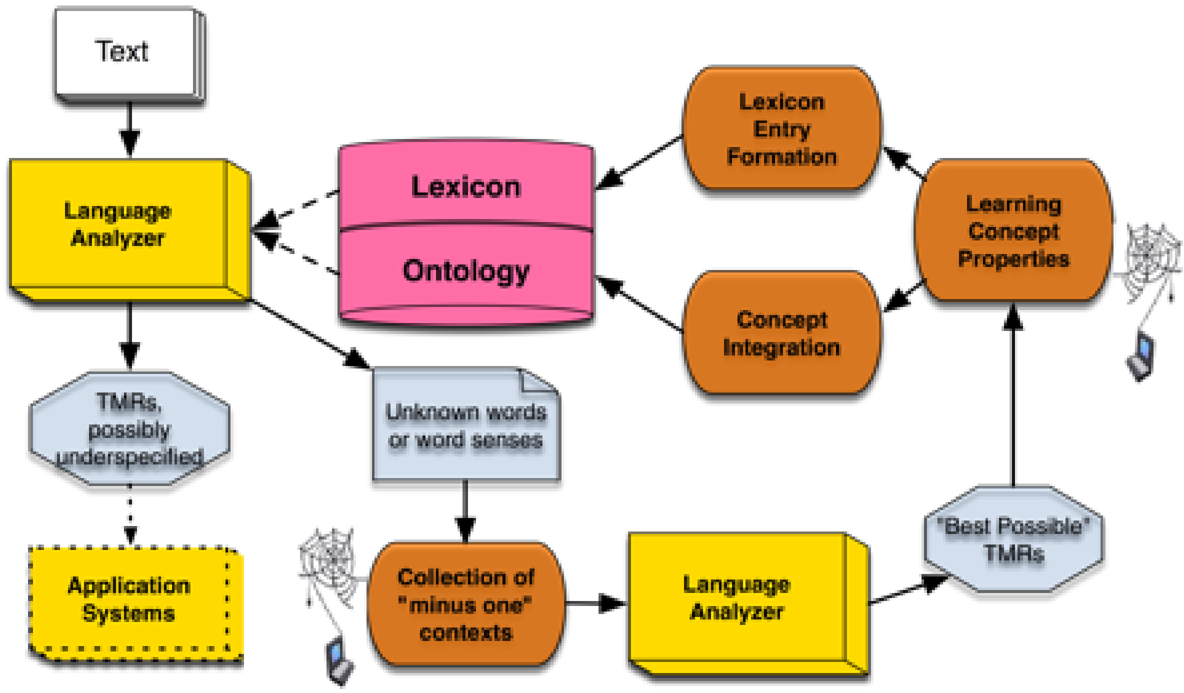

Fig. 1. The mutual bootstrapping ontology learning environment. The ”minus-one” contexts refer to sentences in which all the lexical units are attested in the system lexicon other than the one that we intend to learn.

THEME	MODALITY-1.VALUE
MODALITY-1	
TYPE	EPISTEMIC
SCOPE	PAIN-1
EXPERIENCER-1	
from-sense	have-v7
DOMAIN	PAIN-1
RANGE	HUMAN-1
PAIN-1	
from-sense	chest-pain-n1
EXPERIENCER	HUMAN-1
LOCATION	CHEST-1

The question mark matches lexical sense **question-mark*-punct1*, whose syntactic structure expects the input to be a yes/no question, and whose semantic structure indicates that a REQUEST-INFO event must be instantiated whose THEME is an instance of epistemic modality scoping over the proposition asked about. (The value of the epistemic modality will be 1 if the response is ”yes” and 0 if it is ”no”.) The scope of the modality is an instance of PAIN, whose LOCATION is CHEST and whose EXPERIENCER is the person being asked the question, HUMAN-1. This sentence includes two collocations that have been treated specially in the lexicon: *have (+ symptom/disease)* and *chest pain*. *Have* is a so-called ”light verb” that has many different senses depending on the arguments it takes. The 7th verbal sense of *have* in the OntoSem lexicon (*have-v7*) expects the subject to refer to an ANIMAL and the direct object to refer to a SYMPTOM or DISEASE. In such semantic contexts, *have* indicates that the ANIMAL is the EXPERIENCER-OF the SYMPTOM/DISEASE, which is what the TMR shows. Other senses of *have* have different syntactic and/or semantic constraints, permitting automatic disambiguation. The collocation *chest pain* happened to be recorded as a phrasal in the lexicon with the meaning (PAIN (LOCATION CHEST)), but it could also have been analyzed productively, since the first nominal sense of *pain* permits an optional preposed nominal which, when referring to a BODY-PART, means that the LOCATION of the PAIN is that BODY-PART. As TMRs are generated, the Fact Extractor extracts relevant (as determined by the agent's active plans and goals) information to be stored in the agent's Fact Repository for later use as input to decision functions.

The main points to take away from this brief example are: (1) text meaning representations contain the results of often complex language- and world-oriented reasoning, (2) they are unambiguous, (3) they contain *knowledge* that can be stored in an intelligent agent's memory and directly utilized in decision-making and (4) although **text** meaning representations derive from language, *meaning* representations that are formally identical could ultimately be derived from any type of sensory input: vision, haptics, graphics, etc. While pursuing long-term basic research, our team has also concentrated on building practical systems for applications ranging from machine translation (e.g., [14]) to question answering (e.g., [15]) to generating Semantic Web content (e.g., [16]) and information extraction (e.g., [17]).

III. The Learning Experiment

Experimentation on which we report in this paper is the continuation of work reported in [18] and [19]. Using the web as an open corpus of raw text, and an unknown target word (or phrase) as input, the learner extracts semantic knowledge relating to the unknown target, clusters the information learned into a set of word senses, treats each of the word senses as a candidate ontological concept and proposes a location for each candidate in the ontology. The experiment can be broken down into three main phases, each of which is a prerequisite for the subsequent one. In this section we describe each of these phases and the output they produce.

A. Raw Text Corpus Building

The first phase of the learner involves constructing a corpus of raw texts containing the input target word from the web. If the learning process is triggered as a side effect of the operation of the analyzer, the target word is identified automatically. In a standalone mode, the target word is supplied manually. A search engine is then queried for at least 500 results matching the target word. Each of these pages is then archived, the HTML markup and other extraneous metadata is removed, and the resultant text is run through a sentence breaker. Any sentences from the pages that contain the target word are retained, while all others are discarded, resulting in a corpus of sentences involving some sense of the target word. At this point we make no distinction regarding the polysemy of the target word: it is assumed that most words have multiple meanings, and for the purposes of extracting a corpus or raw sentences, we do not attempt to subdivide the texts by word sense (this is done after extraction of the property-filler pairs from the TMRs: see section C below).

To ensure a base level of quality, the texts are then filtered through multiple checks, including an English-sentence probability filter (a custom implementation of a Hidden Markov Model) as well as a parse complexity filter (a system designed to estimate the complexity of the syntactic parse tree that OntoSem will generate and use). Each of these filters is designed to filter out input sentences that are garbled or in general not sufficiently well formed to provide useful semantic knowledge about the target word. The original corpus consisted of over 15,000 sentences; after filtering, 4,768 sentences were retained for further processing.

For the experiment described in this paper we hand-selected ten target words. Five of these words (*dragster*, *truck*, *freeway*, *skid*, and *curb*) were selected as "independent," meaning the query issued to the search engine consisted only of the target word itself. The second set of five words (*chicken*, *jackknife*, *bottleneck*, *rubberneck*, and *hitchhike*) were "dependent," meaning that the query issued contained, in addition to the target word itself, at least one of the above independent target words. These words were originally selected for a separate set of experiments (not reported here) on learning domain interdependencies from a non-domain-specific corpus. This choice did not impact our current experimentation. A total of 4,768 sentences was acquired for the ten target words.

B. TMR Corpus Building

The above corpus was processed through the semantic analyzer (OntoSem), which produced a set of automatically generated TMRs. These TMRs consist of a collections of frames in the OntoSem metalanguage that express the meaning of input elements and relations among them; the target words remain unknown at this point.

C. Candidate Generation and Placement

OntoSem incorporates means for alleviating well-known brittleness of knowledge-based NLP systems associated with the need to process out-of-vocabulary lexical material. In OntoSem such unexpected input is treated by relaxing selectional restrictions on lexical units standing in specific relations with the unknown lexical unit in the input. In other words, when the standard selectional restriction matching process, the core method of knowledge-based disambiguation, cannot be carried out because constraints on one of the two elements being compared are unknown, selectional restrictions are applied unidirectionally, that is the restrictions of the known lexical unit are assigned to the unknown one. This operation is similar to unification of feature structures.

By aggregating the results of this unidirectional selectional restriction application, we produce a statistically probable cumulative representation of the set of selectional restrictions on the target word (at this stage still covering all its senses in one place). This set of selectional restrictions provides values for an important subset of properties (mostly, case roles) of a candidate ontological concept expressing the meaning of the target word. The instance frames generated in each TMR from the target word are crawled for property-filler pairs, which are collated and then passed to a custom word-sense clusterer based on the WEKA [20] implementation of the EM algorithm [21]. A matrix of property-filler occurrences is constructed and used as input to the clustering algorithm to separate the monolithic candidate concept into a set of concepts expressing the meaning of separate senses of the target word.

Each cluster is viewed as a candidate concept for which an appropriate place must be found in the onological subsumption structure. For this purpose we have developed a tree-crawling algorithm (emulating a decision tree), and a custom metric that determines similarity using an aggressive pairwise property-filler comparison tool. The algorithms navigates the subsumption links of the ontology until it settles on the existing concept that appears to be most similar to the candidate: the existing concept is considered the "proposed parent" of the candidate, and the learner terminates. If a candidate exactly matches an existing ontological concept, it is listed as a synonym in the lexicon entry for each lexical unit that references this concept. If no appropriate position can be found for a candidate, it is sent to an acquirer for manual correction.

D. Evaluation

Evaluation was performed using a custom similarity metric (a system designed to process a pair-wise property/filler comparison on two ontological concepts, accounting for subsumption in the ontology, resulting in a 0 - 1 similarity score), comparing each candidate concept to a gold-standard concept that was manually acquired for this purpose. These gold standards are represented by a fully manually acquired concept for the target word, created to the same level of detail as other entries in the ontology. The similarity metric reports precision, and recall, which also produces an f-measure.

IV. Isolating The Pieces

In order to evaluate the impact of each phase of the learner on the final results, we isolated the components one at a time, and replaced them with increasing levels of manually-corrected outputs. Intuitively, we intend to show that each component is bound to introduce cascading error, and by removing that error (by replacing the automatic outputs with manually-corrected ones) we can improve the final results and also make it easier - and, therefore, less expensive - to validate and correct. In order to test this hypothesis, we ran four independent experiments, each with an increasing level of human intervention: the first experiment provides a fully-automatic baseline, as described in the previous section; the second experiment replaces raw input with a manually trimmed version, and continues automatically as before; the third experiment replaces the TMR corpus that was produced in the second experiment with a manually corrected set of TMRs, and the fourth experiment enhances the TMRs with manually acquired lexical and ontological knowledge so that the only unknown element in the input is the target word. Table I shows the frequency of occurrence of each words in the Baseline section, as well as the final three sections. The final results of each experiment were compared, and are shown at the end of this section.

Our f-measure scores are determined by comparing precision and recall of all individual property / filler pairs in the candidate concept with those found in the gold-standard (manually annotated) concept. The semantics of precision and recall we use differs from that used in standard information retrieval: we define precision as the number of property / filler pairs found in the candidate that were also found in the gold-standard (with penalties applied to account for the hierarchical nature of the ontology) out of the total property / filler pairs in the candidate. Recall is calculated as the number of property / filler pairs in the gold-standard that were found in the candidate (with penalties applied both for not finding a property as well as failing to match a property / filler pair) out of the total property / filler pairs in the gold-standard. It is important to note that there are in excess of 200 properties defined in the ontology, many of which have thousands of possible fillers. Average existing concepts in the ontology contain hundreds of property / fillers of varying degrees of specificity. In order to even hope to obtain a 100% precision and recall, the learner must at the very least have access to text containing expressions denoting every property / filler, with no extraneous information. Our corpus work, though extensive, did not afford such an opportunity.

	Baseline	**AS-IS, Edited, Lex.**
VOLUNTARY-VISUAL-EVENT	50	9
TRAVEL-EVENT	53	7
SPACE	76	9
KNIFE	103	11
ACCIDENT	103	11
AUTOMOBILE	140	10
SPORTS-ROLE	140	10
SLIDE	232	10
WHEELED-VEHICLE	232	10
ROAD-SYSTEM-ARTIFACT	397	9
RESTRAIN	397	9
CHICKEN	421	8
SPORTS-COMPETITION	421	8
HIGHWAY	448	15
AUTOMOBILE	1778	18
WHEELBARROW	1778	18
TRANSFER-OBJECT	1778	18

TABLE I
FREQUENCY OF OCCURRENCE OF TARGET WORDS.

A. Baseline

The baseline experiment was a fully automatic learning experiment as described in Section 3. The raw input corpus of 4768 sentences, each containing one of the ten target words, was semantically analyzed by OntoSem, and the TMRs were used to construct a set of word senses that were mapped to the ontology with no manual intervention. Figure 2 shows the results of this experiment, which had an average f-measure score of 11.84%. The x-axis (here and in all the rest of the experiments) is labeled with the (gold-standard) ontological concepts to which the our results were compared. The concepts are arranged in the order of the frequency of target word occurrence in the corpus, and are plotted against their precision, recall and f-measure scores; additionally linear trend-lines are included to illustrate the effect an increased corpus size has on the results. As can be seen, the quality of results begins to drop as input frequency increases, due to the cumulative effects of the noise inherent in the raw corpus. To give but one example of the "noisy" sentences, consider the following: *These statistics prompted the community to mobilize and develop Project CURB* returned when searched for the target word *curb*.

B. AS-IS

The second experiment, AS-IS, replaced the automatically collected raw text corpus of 4768 sentences with 92 hand-selected sentences from the original corpus. The intention was to show the impact of a smaller, but higher-quality (less noisy) corpus of raw input texts. The texts were selected for being semantically well-formed (sufficient to describe the meaning of the target word) while being relatively straight-forward syntactically. The number of sentences selected for each target word was mostly proportional to the number of sentences found in the full corpus, to maintain consistency (although minor variance was introduced). The remainder of the experiment was carried out fully automatically as before. The manual effort involved in this phase has been minimal: it took on the order of just one hour to compose the small

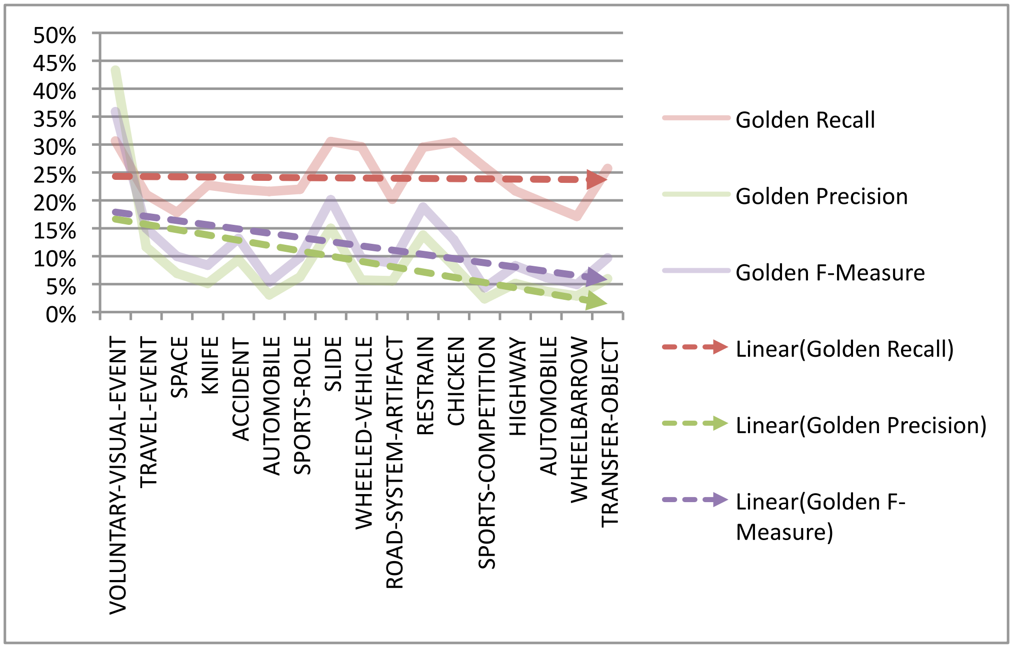

Fig. 2. Baseline precision and recall

corpus of 92 sentences. The results of the AS-IS experiment are shown in Figure 3, with an average f-measure score of 5.96%.

C. Edited

For the third phase of the experiment, we maintained the reduced corpus of 92 texts, but rather than using the fully automatically created TMRs from OntoSem, we manually corrected each to the highest standard that could be produced given the current state of the static knowledge resources in OntoSem. These knowledge resources were not in any way modified for this experiment, so any TMR elements that required knowledge unavailable in the system were not manually corrected. As an example, consider the text: *The truck was involved in an accident*. In this case, the target word is *truck*, however the existing static knowledge resources to not have the correct sense of the word *accident*, having only a general (non-automotive) meaning. The edited TMR for this text would verify that the case-roles for the *truck* were properly filled, however the word-sense for *accident* would have to remain untouched. These TMRs represented the best we could have hoped for OntoSem to produce at its current state of development. The manual effort involved in preparing this experiment took approximately twenty hours. Figure 4 shows the results of the Edited experiment, which averaged 11.74% (f-measure).

D. Lexicalized

In the final phase of the experiment, we further corrected the TMRs by manually acquiring all word senses for the entire input corpus that were not originally covered by the OntoSem knowledge resources (excepting the target words). The TMRs were then updated to reflect the new knowledge. To extend the previous example, another word-sense for *accident* would be acquired (the automotive meaning), and the TMR would be updated to reflect this. This presented us with the absolute best possible TMRs for the input sentence. The manual effort involved in this phase lasted approximately five hours, but required the prior manual effort from the previous experiment. Figure 5 presents the results of the Lexicalized experiment, producing an average f-measure of 13.82%.

E. Evaluation

The purpose of these experiments was to evaluate the change in results given higher quality inputs to each main phase of the learner's pipeline. Table II shows a comparison of the average f-measure scores during each phase of the experiment; these scores reflect our initial hypothesis that increased quality of inputs has a higher benefit than increased quantity of inputs. Although the AS-IS scores are lower than the baseline, (covering approximately half of the knowledge of the baseline phase), the AS-IS phase was working with ~2% of the input. The edited phase shows the most promise, as the average score is nearly as good as the baseline, but again with a substantially reduced corpus of higher quality TMRs. Finally, the edited phase surpasses the score of the baseline, showing that a higher quality of inputs has produced better results than a higher quatity.

V. Future Work and Conclusion

The inclusion of manual corrections throughout the learning pipeline is not the only method of hybrid human-computer acquisition: another possibility includes a fully automatic acquisition followed by human correction of the output. In future research, we intend to test this process by identifying the volume of corrections required (as well as man-hours required)

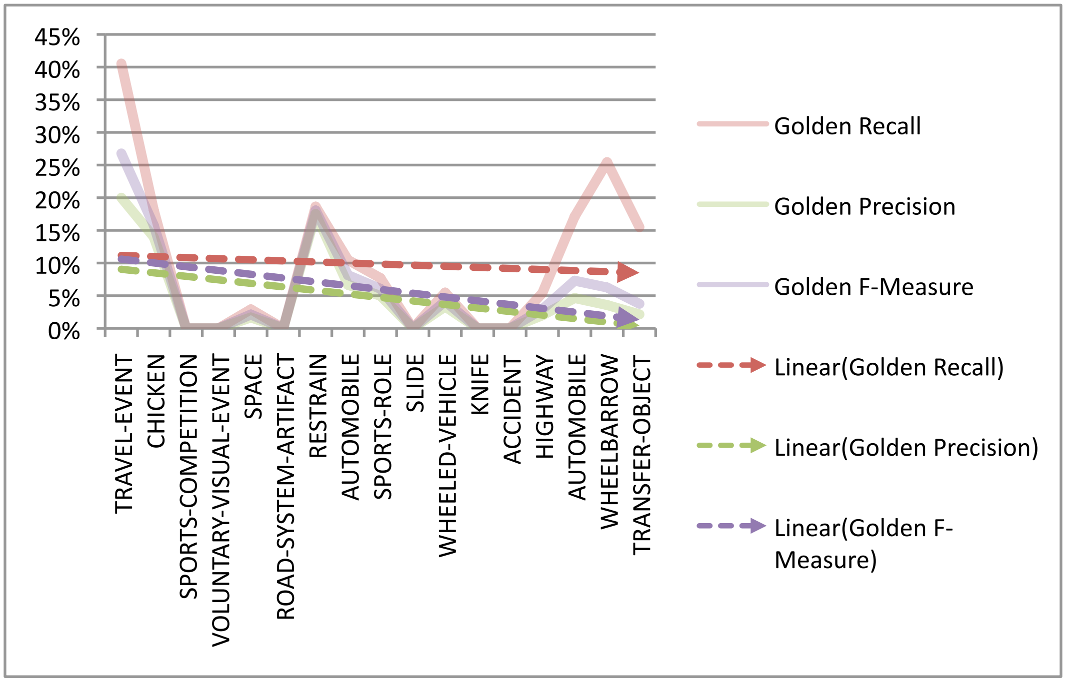

Fig. 3. AS-IS precision and recall

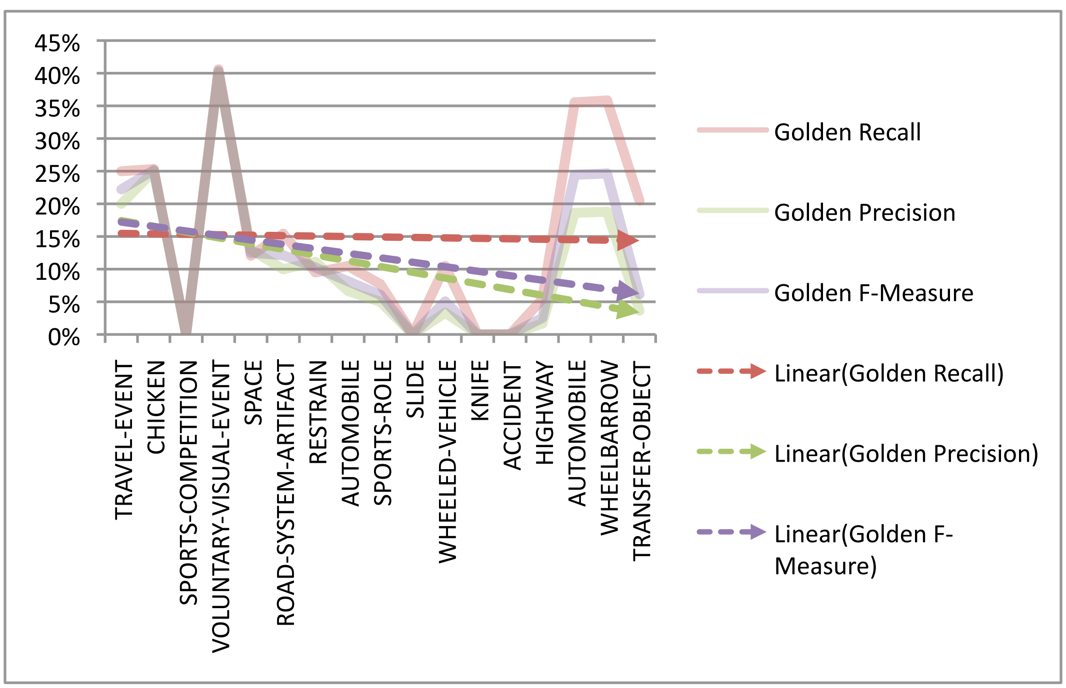

Fig. 4. Edited precision and recall

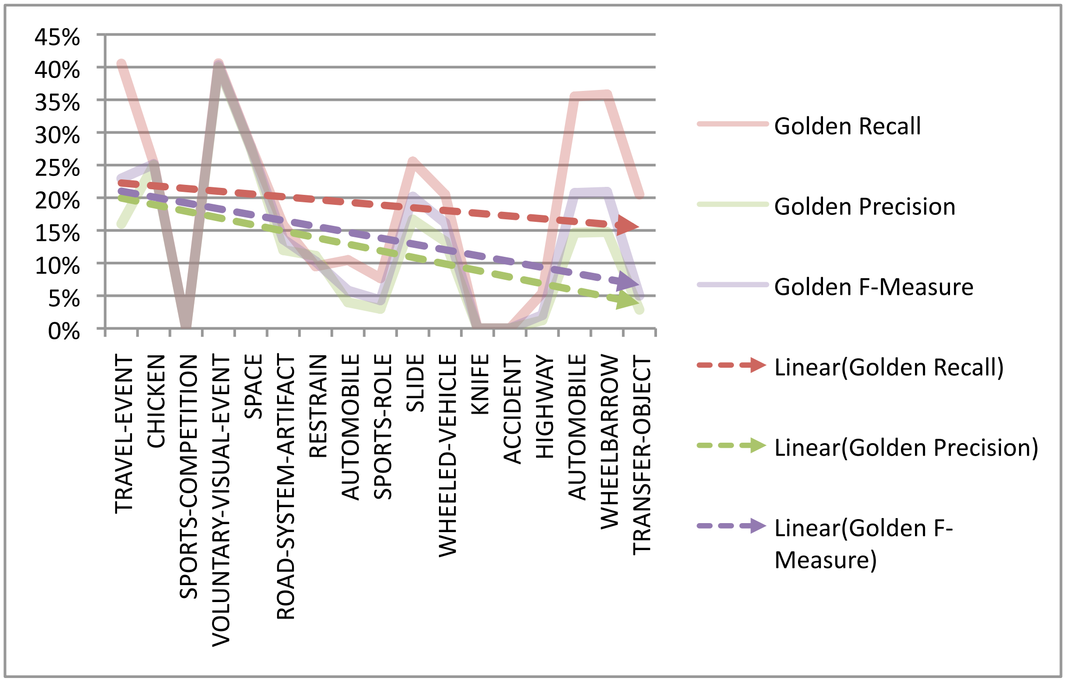

Fig. 5. Lexicalized precision and recall

	Avg. f-measure	# Sentences
Baseline	11.84	4768
AS-IS	5.96	92
Edited	11.74	92
Lexicalized	13.82	92

TABLE II
AVERAGE F-MEASURES ACROSS PHASES.

to align the candidate concepts that were manually acquired with gold-standards. Further, we can modify our evaluation metric to include an evaluation of the TMRs produced by OntoSem (e.g., [9]) in order to more accurately assess the impact of low-quality analyses on the learner.

Additionally, we will need to investigate the learning (either automatic, or semi-automatic) of lexical entries realizing the learned ontological concepts in specific languages. The sem-struc zones of OntoSem lexicon entries for specific word senses encode the meaning of these word senses, typically by referring to ontological concepts in a metalanguage of the ontology. Ontology-lexicon mappings may be univocal, with the content of the sem-struc zone simply pointing to an ontological conept, or constrained, when ontological concepts are imported into the lexicon and the modified *in situ*. For example, there is a choice of expressing the meaning of the English verb *taxi* either as a reference to an ontological concept, say TAXI-EVENT, or as an instance of the concept MOVE-ON-SURFACE in which the theme of movement is an airplane. Even if we choose the simplifying option of always using univocal mappings, we will still have to address the other zones of the sense, including the syn-struc (which defines the lexical entry's syntactic structure, see [22]) as well as the issues connected with linking syntactic modifiers into semantic case roles (and other properties).

Further, we intend to extend the contextual base for learned words by incorporating entire text documents in lieu of single sentences. The nature of our work using full semantic realizations of text requires the use of a comprehensive reference resolution engine designed for use with OntoSem; work in this area is ongoing (see, e.g., [23]), and will be incorporated into our automatic knowledge research in the future.

At the time of writing we haven't had an opportunity for testing the practical utility of human intervention during the process of automatic acquisition and comparing this methodology to pure post-editing of automatically produced results. However, we have developed a toolkit that supports such intervention and informally compared the judgments of knoweldge engineers about the relative ease of interactive editing versus post-editing. After the initial learning curve of using the toolkit, knowledge engineers uniformly preferred the interactive option. We have also observed a seemingly surprising phenomenon: drastically reducing the size but enhancing the quality of the input corpus (from 4863 texts down to only 92) by manually selecting the best quality inputs had a telling eventual impact on results with minimal manual interference, implying that using human help at the earliest phase of raw text corpus construction may contain the best cost-benefit ratio. Among many potential improvements of the approach we plan to experiment with using even more points at which human-computer interaction is licensed to occur, for example, allowing the user to correct results of automatic syntax-to-semantics linking (for which we already have a dedicated interface). On a larger scale, we are also planning an experiment in "lifelong" learning, whereby the results of knowledge acquisition according to the above methodology are continuously fed back into the knowledge resources of OntoSem with the expectation that eventually the amount of manual work made necessary by knowledge lacunae will taper off. It is also quite possible that uses will be found even for the less-than-optimum results of automatic learning at the present stage in the development of OntoSem knowledge resources, though at the moment we do not plan to actively investigate this option.

REFERENCES

[1] C. Brewster, J. Iria, F. Ciravegna, and Y. Wilks, "The Ontology: Chimaera or Pegasus," in *In Proc. Dagstuhl Seminar on Machine Learning for the Semantic Web*, 2005, pp. 13–18.

[2] U. Hahn and K. Schnattinger, "Knowledge mining from textual sources," in *CIKM '97: Proceedings of the sixth international conference on Information and knowledge management*. New York, NY, USA: ACM, 1997, pp. 83–90.

[3] C. Cardie, "Embedded Machine Learning Systems for Natural Language Processing: A General Framework," in *Lecture Notes in Artificial Intelligence Series*. Springer, 1996, pp. 315–328.

[4] R. Basili, M. Pazienza, and M. Vindigni, "Lexical Learning for Improving Syntactic Analysis," in *ACAI99 Workshop on Machine Learning in Human Language Technology*, 1999.

[5] M. laure Reinberger, "Discovering Knowledge in Texts for the Learning of DOGMA-Inspired Ontologies," in *ECAI 2004 Workshop on Ontology Learning and Population*, 2004.

[6] D. Faure and C. Nedellec, "Knowledge Acquisition of Predicate Argument Structures from Technical Texts Using Machine Learning: The System ASIUM," in *EKAW '99: Proceedings of the 11th European Workshop on Knowledge Acquisition, Modeling and Management*. London, UK: Springer-Verlag, 1999, pp. 329–334.

[7] N. Ogata and N. Collier, "Ontology Express: Statistical and Non-Monotonic Learning of Domain Ontologies from Text," in *ECAI 2004 Workshop on Ontology Learning and Population*, 2004.

[8] S. Nirenburg, M. McShane, and S. Beale, "Operative strategies in Ontological Semantics," in *Proceedings of HLT-NAACL-03 Workshop on Text Meaning*, 2003.

[9] S. Nirenburg, S. Beale, and M. McShane, "Evaluating the Performance of the OntoSem Semantic Analyzer," in *Proceedings of the ACL-04 Workshop on Text Meaning*, 2004.

[10] M. McShane, *A Theory of Ellipsis*. Oxford University Press, 2005.

[11] M. McShane, S. Beale, and S. Nirenburg, "Some Meaning Procedures of Ontological Semantics," in *Proceedings of LREC-2004*, 2004.

[12] M. McShane, S. Nirenburg, and S. Beale, "Semantics-Based Resolution of Fragments and Underspecified Structures," *Traitement Automatique des Langues 46*, vol. 1, pp. 163–184, 2005.

[13] J. Barnden, "Metaphor, Semantic Preferences and Context-Sensitivity," in *Ahmad, Brewster and Stevenson (eds.) Words and Intelligence II: Essays in Honor of Yorick Wilks*, 2007.

[14] S. Beale, S. Nirenburg, and K. Mahesh, "Semantic analysis in the Mikrokosmos machine translation project," in *Proceedings of the Symposium on Natural Language Processing*, 1995.

[15] S. Beale, B. Lavoie, M. McShane, S. Nirenburg, and T. Korelsky, "Question answering using Ontological Semantics," in *Proceedings of ACL-2004 Workshop on Text Meaning and Interpretation*, 2004.

[16] A. Java, S. Nirenburg, M. McShane, T. Finin, J. English, and A. Joshi, "Using a natural language understanding system to generate Semantic Web content," *International Journal on Semantic Web and Information Systems*, vol. 3, pp. 50–74, 2007.

[17] S. Nirenburg, M. McShane, and S. Beale, "Enhancing recall in information extraction through ontological semantics," in *Proceedings of the Workshop on Ontologies and Information Extraction*, 2003.

[18] J. English and S. Nirenburg, "Ontology Learning from Text Using Automatic Ontolgical-Semantic Text Annotation and the Web as the Corpus," in *Proceedings of the AAAI 2007 Spring Symposium Series on Machine Reading*, 2007.

[19] S. Nirenburg, T. Oates, and J. English, "Learning by Reading by Learning to Read," in *Proceedings of ICSC-07*, 2007.

[20] S. R. Garner, "WEKA: The Waikato Environment for Knowledge Analysis," in *In Proc. of the New Zealand Computer Science Research Students Conference*, 1995, pp. 57–64.

[21] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[22] M. McShane, S. Nirenburg, S. Beale, and T. O'Hara, "Semantically Rich Human-aided Machine Annotation," in *Proceedings the Workshop on Frontiers in Corpus Annotation II: Pie in the Sky, ACL-05*, 2005.

[23] M. McShane, "Advances in difficult aspects of reference resolution: Working Notes," Institute for Language and Information Technologies, Tech. Rep., 2009.