Generic parsing for multi-domain semantic interpretation

Myroslava Dzikovska∗, **Mary Swift**†, **James Allen**†, **William de Beaumont**†
∗ Human Communication Research Centre
University of Edinburgh, 2 Buccleuch Place, Edinburgh EH8 9LW, United Kingdom
m.dzikovska@ed.ac.uk
† Department of Computer Science University of Rochester, Rochester, NY 14627-0226
{swift, james, wdebeaum}@cs.rochester.edu

1 Introduction

Producing detailed syntactic and semantic representations of natural language is essential for practical dialog systems such as plan-based assistants and tutorial systems. Development of such systems is time-consuming and costly as they are typically hand-crafted for each application, and dialog corpus data is more difficult to obtain than text. The TRIPS parser and grammar addresses these issues by providing broad coverage of common constructions in practical dialog and producing semantic representations suitable for dialog processing across domains. Our system bootstraps dialog system development in new domains and helps build parsed corpora.[1]

Evaluating deep parsers is a challenge (e.g., (Kaplan et al., 2004)). Although common bracketing accuracy metrics may provide a baseline, they are insufficient for applications such as ours that require complete and correct semantic representations produced by the parser. We evaluate our parser on bracketing accuracy against a statistical parser as a baseline, then on a word sense disambiguation task, and finally on full sentence syntactic and semantic accuracy in multiple domains as a realistic measure of system performance and portability.

2 The TRIPS Parser and Logical Form

The TRIPS grammar is a linguistically motivated unification formalism using attribute-value struc-

```
(SPEECHACT sa1 SA_REQUEST :content e123)
(F e123 (:* LF::Fill-Container Load)
        :Agent pro1 :Theme v1 :Goal v2)
(IMPRO pro1 LF::Person :context-rel *YOU*)
(THE v1 (SET-OF (:* LF::Fruit Orange)))
(THE v2 (:* LF::Vehicle Truck))
```

Figure 1: LF for *Load the oranges into the truck.*

tures. An unscoped neo-Davidsonian semantic representation is built in parallel with the syntactic representation. A sample logical form (LF) representation for *Load the oranges into the truck* is shown above. The TRIPS LF provides the necessary information for reference resolution, surface speech act analysis, and interpretations for a wide variety of fragmentary utterances and conventional phrases typical in dialog. The LF content comes from a domain-independent ontology adapted from FrameNet (Johnson and Fillmore, 2000; Dzikovska et al., 2004) and linked to a domain-independent lexicon (Dzikovska, 2004).

The parser uses a bottom-up chart algorithm with beam search. Alternative parses are scored with factors assigned to grammar rules and lexical entries by hand, because due to the limited amount of corpus data we have not yet been able to train a statistical model that outperforms our hand-tuned factors.

3 Evaluation

As a rough baseline, we compared the bracketing accuracy of our parser to that of a statistical parser (Bikel, 2002), Bikel-M, trained on 4294 TRIPS

parse trees from the Monroe corpus (Stent, 2001), task-oriented human dialogs in an emergency rescue domain. 100 randomly selected utterances were held out for testing. The gold standard for evaluation is created with the help of the parser (Swift et al., 2004). Corpus utterances are parsed, and the parsed output is checked by trained annotators for full-sentence syntactic and semantic accuracy, reliable with a kappa score 0.79. For test utterances for which TRIPS failed to produce a correct parse, gold standard trees were manually constructed independently by two linguists and reconciled. Table 1 shows results for the 100 test utterances and for the subset for which TRIPS finds a spanning parse (74). Bikel-M performs somewhat better on the bracketing task for the entire test set, which includes utterances for which TRIPS failed to find a parse, but it is lower on complete matches, which are crucial for semantic interpretation.

	All test utts (100)			Spanning parse utts (74)		
	R	P	CM	R	P	CM
BIKEL-M	79	79	42	89	88	54
TRIPS	77	79	65	95	95	86

Table 1: Bracketing results for Monroe test sets (R: recall, P: precision, CM: complete match).

Word senses are an important part of the LF representation, so we also evaluated TRIPS on word sense tagging against a baseline of the most common word senses in Monroe. There were 546 instances of ambiguous words in the 100 test utterances. TRIPS tagged 90.3% (493) of these correctly, compared to the baseline model of 75.3% (411) correct.

To evaluate portability to new domains, we compared TRIPS full sentence accuracy on a subset of Monroe that underwent a fair amount of development (Tetreault et al., 2004) to corpora of keyboard tutorial session transcripts from new domains in basic electronics (BEETLE) and differentiation (LAM) (Table 2). The only development for these domains was addition of missing lexical items and two grammar rules. TRIPS full accuracy requires correct speech act, word sense and thematic role assignment as well as complete constituent match.

Error analysis shows that certain senses and subcategorization frames for existing words are still

Domain	Utts	Acc.	Cov.	Prec.
Monroe	1576	70%	1301	84.1%
BEETLE	192	50%	129	75%
LAM	934	42%	579	68%

Table 2: TRIPS full sentence syntactic and semantic accuracy in 3 domains (Acc: full accuracy; Cov.: # spanning parses; Prec: full acc. on spanning parses).

needed in the new domains, which can be rectified fairly quickly. Finding and addressing such gaps is part of bootstrapping a system in a new domain.

4 Conclusion

Our wide-coverage grammar, together with a domain-independent ontology and lexicon, produces semantic representations applicable across domains that are detailed enough for practical dialog applications. Our generic components reduce development effort when porting to new dialog domains where corpus data is difficult to obtain.

References

D. Bikel. 2002. Design of a multi-lingual, parallel-processing statistical parsing engine. In *HLT-2002*.

M. O. Dzikovska, M. D. Swift, and J. F. Allen. 2004. Building a computational lexicon and ontology with framenet. In *LREC workshop on Building Lexical Resources from Semantically Annotated Corpora*.

M. O. Dzikovska. 2004. *A Practical Semantic Representation For Natural Language Parsing*. Ph.D. thesis, University of Rochester.

C. Johnson and C. J. Fillmore. 2000. The FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure. In *ANLP-NAACL 2000*.

R. M. Kaplan, S. Riezler, T. H. King, J. T. Maxwell III, A. Vasserman, and R. S. Crouch. 2004. Speed and accuracy in shallow and deep stochastic parsing. In *HLT-NAACL 2004*.

A. J. Stent. 2001. *Dialogue Systems as Conversational Partners*. Ph.D. thesis, University of Rochester.

M. D. Swift, M. O. Dzikovska, J. R. Tetreault, and J. F. Allen. 2004. Semi-automatic syntactic and semantic corpus annotation with a deep parser. In *LREC-2004*.

J. Tetreault, M. Swift, P. Prithviraj, M. Dzikovska, and J. Allen. 2004. Discourse annotation in the Monroe corpus. In *ACL workshop on Discourse Annotation*.