Robust Parsing in HPSG: Bridging the Coverage Chasm

Dan Flickinger (Stanford Univ.) and Woodley Packard (Univ. of Washington)

Grammar implementations which are guided by linguistic theory will normally lack coverage of even some well-formed utterances, since no current theory exhaustively characterizes all of the phenomena in any language. For many uses of a grammar, approximate or robust analyses of the out-of-grammar utterances would be better than nothing, and a variety of approaches have been developed for such robust parsing. In this work, we present an implemented method which adds two simple "bridging" rules to an existing broad-coverage HPSG grammar, the English Resource Grammar (ERG: Flickinger (2011)), allowing any two constituents to combine. This method relies on a parser which can efficiently pack the full parse forest for an utterance, and then selectively unpack the most likely N analyses guided by a statistical model trained on a manually constructed treebank.

Motivations are numerous for extending a manually constructed grammar to include approximate analyses of out-of-grammar sentences, including the use of such grammars for analysis of dialogue-based phenomena such as coreference resolution or VP ellipsis, where even an approximate parse may often include successful analyses of the relevant phrases, sufficient to sustain coreference chains across multiple sentences. For the grammarian interested in identifying occurrences of a phenomenon under study in a large corpus, these approximate parses will again often exhibit such occurrences if that phenomenon itself is within the coverage of the non-robust grammar. For computational linguists or NLP application developers, comparison of accuracy of different grammar/parser engines at the level of bi-lexical semantic dependencies can be more informative when (nearly) every sentence in an evaluation corpus receives some kind of analysis.

Previous work on adding a robust "safety net" to manually constructed grammars has explored both relatively simple strategies and more ambitious hybrid approaches. Kasper et al (1999) tried concatenation of the longest candidate phrases found by a parser for inputs that received no full covering analysis. Fouvry (2003) proposed a much more computationally expensive method, allowing relaxation of all unification constraints so that any two signs could combine. Closer in spirit to the present proposal, Schneider & McCoy (1998) manually extended a grammar with *mal*-rules that explicitly admit particular sequences of constituents not otherwise licensed by the grammar. None of these approaches achieved the desired balance in scalability of human effort, accuracy, processing efficiency, and interaction with the core linguistically principled grammar.

We propose a radical variant of the *mal*-rule approach, adding just two *bridging* rules to a standard HPSG grammar, designed to (recursively) construct a robust "phrase" from any two constituents, and relying on a more sophisticated statistical model to select the most useful analysis from the very large space of candidates that results from using such permissive rules. One of the two bridge rules is unary, converting any well-formed constituent into a "bridge-head", and the other rule is a binary one which admits any sequence of two bridge signs, either unary or binary. Below are these two bridging rules as implemented in the ERG, with the unary rule on the left and the binary rule on the right.

While the unary rule adds no semantics beyond that provided by its daughter, the binary rule adds to the resulting semantics a binary bridge_x predication that takes as its two arguments the indices of each of its two daughters, so that the resulting semantics for a bridged analysis is still fully connected, here expressed in terms of Minimal Recursion Semantics (MRS: Copestake et al. (2005)). One important benefit of this approach is that every sentence analysis has an associated semantic representation in the same formalism, whether it employed bridging rules or not, so that applications making use of the grammar need not make any special accommodation for the robustness, beyond some attention to the particular binary bridging predication.

$$\left[\begin{array}{l} \textit{unary-bridge} \\ \text{INFLECTED na} \\ \text{SYNSEM} \left[\begin{array}{l} \text{LOC} \left[\begin{array}{l} \text{CAT.HEAD bridge-head} \left[\text{MOD } \langle\rangle\right] \\ \text{VAL saturated-valence} \end{array}\right] \\ \text{NONLOC non-local-none} \end{array}\right] \\ \text{C-CONT.RELS } \langle\rangle \\ \text{ARGS} \left\langle \left[\begin{array}{l} \text{INFLECTED +} \\ \text{SYNSEM.LOCAL.CAT.HEAD non-frag} \end{array}\right] \right\rangle \end{array}\right]$$

$$\left[\begin{array}{l} \textit{binary-bridge} \\ \text{INFLECTED -} \\ \text{SYNSEM} \left[\begin{array}{l} \text{LOC.CAT} \left[\begin{array}{l} \text{HEAD bridge-head} \left[\text{MOD } \langle\rangle\right] \\ \text{VAL saturated-valence} \end{array}\right] \\ \text{NONLOC non-local-none} \end{array}\right] \\ \text{C-CONT} \left[\begin{array}{l} \text{HOOK} \left[\text{LTOP } \boxed{\textit{lbl}} \text{ INDEX } \boxed{\textit{arg0}}\right] \\ \text{RELS} \left\langle \left[\begin{array}{l} \text{PRED bridge\_x\_rel, LBL } \boxed{\textit{lbl}}, \\ \text{ARG0 } \boxed{\textit{arg0}}, \text{ ARG1 } \boxed{\textit{arg1}}, \text{ ARG2 } \boxed{\textit{arg2}} \end{array}\right] \right\rangle \end{array}\right] \\ \text{ARGS} \left\langle \begin{array}{l} \left[\begin{array}{l} \text{INFLECTED na} \\ \text{SYNSEM.LOCAL} \left[\begin{array}{l} \text{CAT.HEAD bridge-head} \\ \text{CONT.HOOK.INDEX } \boxed{\textit{arg1}} \end{array}\right] \end{array}\right], \\ \left[\begin{array}{l} \text{SYNSEM.LOCAL} \left[\begin{array}{l} \text{CAT.HEAD bridge-head} \\ \text{CONT.HOOK.INDEX } \boxed{\textit{arg2}} \end{array}\right] \end{array}\right] \end{array} \right\rangle \end{array}\right]$$

To date, we have used the enhanced grammar to exhaustively parse the first 1000 sentences of the usual Wall Street Journal corpus, using the ACE parser `moin.delph-in.net/AceTop`, and then to manually treebank each sentence, using the ACE Treebanker to select the best analysis from the parse forest produced by the parser. We thus created training data for computing a statistical model that enables parse selection which considers robust phrases as viable candidates in competition with well-formed constituents. In a first proof-of-concept experiment, we trained a model using 500 of the sentences, and evaluated the accuracy of that model when parsing the second 500, and found that with few exceptions, the model trained on this small amount of data could already correctly prefer a non-bridged analysis over a bridged one when a good analysis was available.

In further experiments, we enlarged the treebank to some 2500 sentences, retrained the parse selection model, and evaluated the accuracy and efficiency of the parser on existing annotated data from the English Wikipedia. We also manually produced 'gold' target semantic dependencies for 45 sentences that the ERG did not parse, and used these to measure both exact tree match and the more fine-grained matching of elementary dependencies. We also compared these measures to those of a robust PCFG, a re-implementation of Zhang & Kordoni (2008), which currently outperforms our bridging approach, but has other drawbacks in resource consumption and brittleness to continued improvements in the ERG.

More work will be needed to confirm the scalability and efficiency of the bridging approach, but our first experiments with the ERG leave us optimistic that this approach to robust parsing can preserve the benefits of using a linguistically precise HPSG grammar implementation while enabling useful if partial analyses of those sentences in any corpus which remain outside the reach of that grammar.

References

Copestake, Ann, Dan Flickinger, Carl Pollard & Ivan A. Sag. 2005. Minimal recursion semantics: An introduction. *Journal of Research on Language and Computation* 3(4):281-332.

Flickinger, Dan. 2011. Accuracy vs. Robustness in Grammar Engineering. In E.M. Bender and J.E. Arnold, eds. *Language from a Cognitive Perspective: Grammar, Usage, and Processing*, pp. 31–50. CSLI Publications, Stanford.

Kasper, Walter, Bernd Kiefer, Hans-Ulrich Krieger, C.J. Rupp, and Karsten L. Worm. 1999. Charting the depths of robust speech parsing. In *Proceedings of the ACL-99 Thematic Session on Robust Sentence-Level Interpretation*

Schneider, David and Kathleen McCoy. 1998 Recognizing Syntactic Errors in the Writing of Second Language Learners. *Proceedings of Coling-ACL 1998*, pp.1198-1204. Montreal.

Zhang, Yi and Valia Kordoni. 2008 Robust Parsing with an HPSG Grammar. *Proceedings of LREC 2008*. Marakesh.