Inferring Minimalist Grammars with an SMT-Solver

Overview. The Minimalist Grammar (MG) formalism (Stabler, 1996) is a well established formal model of syntax inspired by the Minimalist Program (Chomsky, 1995). We introduce (1) a novel parser for MGs, encoded as a system of first-order logic formulae that may be evaluated using a solver for Satisfiability Modulo Theories (SMT) (De Moura and Bjørner, 2008), and (2) a novel procedure for inferring MGs using this parser. The input to this procedure is a sequence of sentences that have been annotated with syntactic relations such as semantic role labels (connecting arguments to predicates) and subject-verb agreement. The output of this procedure is a set of MGs, each of which is able to parse the sentences in the input sequence such that the parse for a sentence has the same syntactic relations as those specified in the annotation for that sentence. We applied this procedure to a set of sentences annotated with syntactic relations and evaluated the inferred grammars using cost functions inspired by the Minimum Description Length principle and the Subset principle. Inferred grammars that were optimal with respect to certain combinations of these cost functions were found to align closely with contemporary theories of syntax, producing the prescribed syntactic structures for a range of constructions that include ditransitive predicates, passivization and Wh-fronting for question formation.

Inference Procedure. Our inference procedure takes the form of a computational model of language acquisition (Chomsky, 1965) consisting of: (1) an initial state, $S_0$, consisting of a system of first-order logical formulae that serve as axioms for deducing the class of minimalist lexicons; (2) the input, consisting of a sequence of $n$ sentences, denoted $I_1, I_2, \ldots, I_n$, each of which is annotated with syntactic relations between pairs of words in the sentence; (3) a function, $Q$, that takes as input a state, $S_i$, and an annotated sentence, $I_i$, and outputs the successor state, $S_{i+1}$; (4) a function, $R$, that maps a state $S_i$ to a set of MG lexicons, $G_i$, with the property that for each sentence $I_j$ in the input sequence, each lexicon $L \in G_i$ can produce a parse $p^L_j$ such that the syntactic relations in $p^L_j$ parse match those specified in the annotation of $s_j$. In the case of the initial state, $S_0$, since there are no constraints yet imposed by the input, $R(S_0)$ will map to the set of all minimalist lexicons. The procedure consumes the input sequence one annotated sentence at a time, using $Q$ to drive the initial state, $S_0$, to the final state, $S_n$; the function $R$ is then applied to $S_n$ to produce a set of MG lexicons, $G_n$, that constitutes the output of the inference procedure.

We implemented this inference procedure by encoding an MG parser as a system of first-order, quantifier-free logical formulas that could be solved with the Z3 SMT-solver (De Moura and Bjørner, 2008). This system of formulas is composed of formulas for MG parse trees that are connected (by way of shared symbols) to a formula for an MG lexicon (i.e. $S_0$); by imposing constraints on the formulas for parse trees (via $Q$), the set of solutions to the lexicon formula is restricted (i.e. $R$ is constrained).

Data. The input to the inference procedure is a sequence of fourteen sentences, $I_1 - I_{14}$ in Table-1, each annotated with predicate-argument relations as well as morphological agreement; the sentences listed include passive constructions ($I_2, I_4, I_{10}$), ditransitive constructions ($I_{11} - I_{14}$), yes/no-questions ($I_4, I_6, I_7, I_{10}$) and wh-questions ($I_1, I_3, I_5, I_6, I_9, I_{11}, I_{13}, I_{14}$).

Analysis. We used our procedure to infer a set of minimalist lexicons, denoted here as $G^*$, from the input sequence described in Table-1. Lexicons sampled from $G^*$ produced parses that do not align with those prescribed by contemporary theories of minimalist syntax. (See Lexicon-A in Table-2 for an example of such a lexicon.)

We filtered out such lexicons by using Z3 to identify lexicons in $G^*$ that were optimal with respect to three cost functions that (respectively): (i) minimized the number of lexical entries in the lexicon; (ii) minimized the total number of selectional and licensing features in the lexicon and the parses (this reduces the size of both the lexicon and the parses); (iii) maximized the number of distinct selectional features in the lexicon (this reduces lexicons that are more exclusive in which structures they generate). We encoded these cost function as first order logical formulae, adding them to the SMT-solver after running the inference procedure, and then re-solving; the resulting set of (inferred) MGs are optimal with respect to the specified cost functions.

This produced a subset of $G^*$, denoted $F^*$, in which each lexicon had exactly: 24 lexical items;
who has eaten
what has mary
who was money

---

Table 1: Model Input — A sequence of sentences annotated with syntactic relations. Some phonetic forms have their category pre-specified, indicated by a suffix of a slash followed by the category. Locality constraints include agreement (Agr) and predicate-argument structure (i.e. a θ grid), with the predicate indicated in the suffix and the subject, object and indirect object components marked by “s”, “o” and “i” respectively. The type of the sentence, declarative or interrogative, is indicated by the end-of-sentence punctuation.

48 features in the lexicon (not including the special feature C); 202 features in the parses; at least five distinct selective features. Lexicons sampled from F* produced parses that respect the syntactic relations prescribed in Table-1 and align with structures prescribed by contemporary theories of minimalist syntax. See Lexicon-B in Table-2 for a representative member of F* – the syntactic phenomenon that Lexicon-B correctly models includes: A’ movement (Wh-fronting for question formation); a (double) VP shell structure that employs V-to-v head-movement (as part of the predicate-argument structure within the parse tree; see (Hale and Keyser, 2002)); T-to-C head-movement (i.e. subj- auxiliary verb inversion) and A- movement (subject raising for morphological agreement).

Conclusion. Our results demonstrate that our procedure for inferring MGs is able to acquire knowledge of syntax from psychologically plausible input and employ movement (i.e. displacement) to establish multiple (crossing and nested) discontinuous relations within a syntactic structure.

References
Noam Chomsky. 1965. Aspects of the theory of syntax. MIT Press.
Noam Chomsky. 1995. The Minimalist Program. MIT Press.
Leonardo De Moura and Nikolaj Bjørner. 2008. Z3: An efficient smt solver. TACAS’08/ETAPS’08, pages 337–340. Springer-Verlag.
Kenneth L. Hale and Samuel J. Keyser. 2002. Prolegomenon to a theory of argument structure, volume 39. MIT press.

| $I_i$ | Sentence | Locality Constraints |
|------|----------|----------------------|
| $I_1$ | who has eaten/V icecream/N? | $\theta_{\text{eating}}$: who, o: icecream, Agr$_{\text{hav}}$: s: who |
| $I_2$ | icecream/N was eaten/V. | $\theta_{\text{eating}}$: icecream, Agr$_{\text{had}}$: icecream |
| $I_3$ | who was eating/V icecream/N? | $\theta_{\text{eating}}$: who, o: icecream, Agr$_{\text{had}}$: s: who |
| $I_4$ | was pizza/N eaten/V? | $\theta_{\text{eating}}$: o: pizza, Agr$_{\text{was}}$: s: pizza |
| $I_5$ | what has john/N eaten/V? | $\theta_{\text{eating}}$: what, Agr$_{\text{was}}$: s: john |
| $I_6$ | has mary/N eaten/V pizza/N? | $\theta_{\text{eating}}$: mary, o: pizza, Agr$_{\text{had}}$: s: mary |
| $I_7$ | was john/N eating/V pizza/N? | $\theta_{\text{eating}}$: o: john, Agr$_{\text{was}}$: s: john |
| $I_8$ | what was mary/N eating/V? | $\theta_{\text{eating}}$: who, o: what, Agr$_{\text{was}}$: s: mary |

| Lexicon-A | Lexicon-B |
|-----------|-----------|
| eaten/V ::= $x_4$, $\sim x_4$ | eaten/V ::= $x_5$, $\sim x_1$ |
| eating/V ::= $x_4$, $\sim x_4$ | eating/V ::= $x_5$, $\sim x_1$ |
| given/V ::= $x_4$, $x_4$, $\sim x_4$ | given/V ::= $x_5$, $x_5$, $\sim x_1$ |
| has/T ::= $x_4$, $\sim x_0$ | icecream/N ::= $x_5$ |
| has/T ::= $x_4$, $\sim x_1$ | icecream/N ::= $x_5$, $\sim x_1$ |
| icecream/N ::= $x_4$, $x_4$, $\sim x_4$ | john/N ::= $x_5$ |
| john/N ::= $x_4$, $\sim x_4$ | money/N ::= $x_5$, $\sim x_1$ |
| john/N ::= $x_4$, $\sim x_1$ | money/N ::= $x_5$, $\sim x_1$ |
| john/N ::= $x_4$, $\sim x_1$ | money/N ::= $x_5$, $\sim x_1$ |
| john/N ::= $x_4$, $\sim x_1$ | money/N ::= $x_5$, $\sim x_1$ |
| john/N ::= $x_4$, $\sim x_1$ | money/N ::= $x_5$, $\sim x_1$ |

Table 2: Examples of inferred lexicons that satisfy the conditions imposed by the input sequence in Table-1. Each lexical item has the form, (PF/CAT::SFS), consisting of a phonetic form (PF), a category (CAT) and a sequence of syntactic features (SFS). The phonetic forms ε is covert (unpronounced). The selectional features are $\{x_0, x_1, \ldots, x_4\}$ and the licensing features are $\{l, r\}$.

Edward Stabler. 1996. Derivational minimalism. In Intl. Conf. on Logical Aspects of Comp. Ling., pages 68–95. Springer.