Constraining MGbank: Agreement, L-Selection and Supertagging in Minimalist Grammars

John Torr
School of Informatics, University of Edinburgh
z134432@sms.ed.ac.uk

Minimlist Grammars (MGs)

- MGs (Stabler 1997) are a computationally-oriented (MCFG equivalent) formalisation of Chomsky’s (1995) Minimalist Program.
- Strongly lexicalised formalism with syntactic movement operations, in which lexical categories consist of an ordered list of features which must be checked and deleted as the derivation is constructed. E.g.:

\[
\text{him} :: \text{d -case}
\]

\[
\text{helps} :: d = +\text{CASE} v
\]

\[
\varepsilon, \text{helps}, \varepsilon :: +\text{case} v, \text{him} :: +\text{case}
\]

- In the above grammar, \(he\) and \(him\) have exactly the same categories, so the grammar overgenerates \(him\) helps \(he\).
- We could split -case into \(-\text{nom/-acc}\), and \(+\text{CASE into } +\text{NOM/+ACC}\).
- But this is not very elegant, as it increases \(k\) and in cases of syncretism \((it, this, that, you, [def] etc)\) blows up the lexicon.

Selectional Restrictions 1: Case Agreement

- Instead, we subcategorise the structure building features with selectional properties (NOM, ACC, FIN, DECL, etc) and requirements (+NOM, +ACC, -FIN, -DECL etc).
- \(X\) indicates that the feature selected or licensed must bear property \(X\), while \(-X\) indicates that it must not bear this property.
- The following updated lexical entries will successfully block \(him\) helps \(he\).

\[
he :: d -\text{case} \text{(NOM)}
\]

\[
him :: d -\text{case} \text{(ACC)}
\]

\[
\text{helps} :: d = +\text{CASE} \{ +\text{ACC} \} v
\]

\[
\text{[pres]} :: lv = +\text{CASE} \{ +\text{NOM} \} t
\]

Selectional Restrictions 2: Subcategorization and Subject-Verb Agreement

- Selectional requirement features can also be used to constrain local subcategorization, for example to block Jack wants that she \(\text{help}\)(s).
- However, cases of long-distance subcategorization also exist: Jack demanded that she be there on time.
- Subject-Verb agreement is also non-local in Minimalist analyses, being mediated by the null head.
- To capture such phenomena, we introduce selectional variables, \(x, y, z\) etc, which allow selectional properties and requirements to be percolated up the tree. E.g. helps will bear a \(+3SG\) feature on its \(v\) selectee which will percolate (via \(vP\)) up to the \(+\text{CASE} licensor of } T\) where it will enforce the non-local agreement.

Supertagging with MGs

- Supertagging first introduced by Bangalore and Joshi (1999) for LTAG, but has since proven most effective at making CCG parsing highly efficient.
- Applies Markovian part-of-speech tagging techniques to strongly lexicalised formalisms, thereby reducing the parser’s search space.
- Although MGs are strongly lexicalised, they contain null heads, which is problematic because existing supertaggers can only tag what they can see.
- We present an algorithm for extracting a set of complex LTAG-like categories from a corpus of derivation trees, which anchors null heads to overt ones.

\[
\text{for each derivation tree } t: \text{for each null head } \eta \text{ in } t:
\]

\[
\text{if } \eta \text{ is a pronom: linkWithGovernor}(\eta);
\]

\[
\text{else: linkWithHeadOfComplement}(\eta);
\]

\[
groupLinksIntoSuperTags()
\]

- The resulting supertags are composed of one overt category and zero or more null categories. The following is the supertag for helps from the derivation tree opposite (subscripts indicate obligatory checking relations):

\[
\text{[decl]} :: l = c
\]

\[
\text{[pres]} :: lv^2 = +\text{CASE } t^1
\]

\[
\text{[trans]} :: v^2 = d h^2
\]

\[
\text{helps} :: d = +\text{CASE } v^3
\]

Selectional Restrictions Continued

- Strongly lexicalised formalism with syntactic movement operations, in which lexical categories consist of an ordered list of features which must be checked and deleted as the derivation is constructed. E.g.:

\[
\text{him} :: d -\text{case}
\]

\[
\text{helps} :: d = +\text{CASE} v
\]

\[
\varepsilon, \text{helps}, \varepsilon :: +\text{case} v, \text{him} :: +\text{case}
\]

- To derive the sentence he helps him we expand the lexicon:

\[
he :: d -\text{case}
\]

\[
\text{[trans]} :: >v= =d lv
\]

\[
\text{[pres]} :: lv= +\text{CASE} t
\]

\[
\text{[decl]} :: t = c
\]

- MGs include the movement operations and phonetically null heads of mainstream linguistic Minimalism, both of which can be computationally costly unless carefully constrained (the worst case complexity of MG chart parsing is \(n^{2.3}\) (Fowlie and Koller 2017)).
- A non-statistical parser was used to generate MGbank, the first wide-coverage corpus of MG trees, and we present two ways in which its search space was constrained.

Table: Preliminary results of an LSTM MG supertagger, trained on 13,000 trees.

<table>
<thead>
<tr>
<th>metric</th>
<th>tagset 30</th>
<th>tagset 25</th>
<th>tagset 10</th>
<th>tagset 4</th>
<th>tagset 2</th>
<th>tagset 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>precision</td>
<td>0.87</td>
<td>0.86</td>
<td>0.85</td>
<td>0.84</td>
<td>0.83</td>
<td>0.82</td>
</tr>
<tr>
<td>recall</td>
<td>0.84</td>
<td>0.82</td>
<td>0.81</td>
<td>0.80</td>
<td>0.79</td>
<td>0.78</td>
</tr>
<tr>
<td>f1</td>
<td>0.85</td>
<td>0.81</td>
<td>0.80</td>
<td>0.79</td>
<td>0.78</td>
<td>0.77</td>
</tr>
</tbody>
</table>

University of Edinburgh
http://www.ilcc.inf.ed.ac.uk/