Factored grammar and performance models

Edward Stabler, UCLA

ENS Paris 2012

- Factors in grammars and performance
  - MG (merge, move) vs MCFG (→): strongly \(\equiv\) but \(\neq\)
  - Factors in incremental parsing
    - Another factor: MG+\(\phi\)Agree vs MG
    - Nonissue: Traces vs none
- Certain varieties of structure dependence matter:
  how to defend these claims
Computational models beyond level 1: Basics first

(1) All relevant responses in range. “descriptive adequacy, level 1”

We need not have the correct model; but a class containing it.

(2) Among adequate models, how to choose?

- $\mathcal{O}(n)$ differences in space/time are insignificant.

Equivalents with a few symbols more or less mainly uninteresting.

- $\mathcal{O}(2^n)$ or $\mathcal{O}(2 \cdot 2^n)$ differences significant.

In/significant comparisons confused in literature; contrasted here.
2 arguments + a model

Conclusions

MG vs. MCFG
A performance model
MG+\phi Agree vs. MG
Traces or not?

Edward Stabler, UCLA
Factored grammar and performance models
Minimalist grammars (MGs)

derived tree

\[
\begin{array}{c}
\text{CP} \\
\text{DP}(0) \quad \text{C'} \\
\quad \text{D'} \quad \text{C} \\
\quad \quad \text{DP} \quad \text{VP} \\
\quad \quad \quad \text{D} \quad \text{NP} \\
\quad \quad \quad \text{N} \\
\end{array}
\]

\[
\begin{array}{c}
\text{which} \quad \text{N'} \\
\quad \text{N} \\
\end{array}
\]

\[
\begin{array}{c}
\text{race} \quad \text{t(0)} \\
\quad \text{D'} \quad \text{V} \\
\quad \text{DP} \\
\end{array}
\]

\[
\begin{array}{c}
\text{horse} \\
\quad \text{they} \\
\end{array}
\]

derivation tree

\[
\begin{array}{c}
\epsilon := V + \text{wh} \ C \\
\quad \text{they} := D \\
\quad \text{race} := D = D \ V \\
\quad \text{which} := N \text{D} \ - \text{wh} \ 	ext{horse} := N \\
\end{array}
\]

isomorphic to MCFG derivation
MGs $\equiv$ MCFGs

- MG features treated as MCFG categories: relation is transparent!
- This translation always works – every MG strongly equiv to MCFG

(Michaelis’98,’01; Harkema’01)
Movement in MG

\[ s : +f \gamma, \mu \uplus \{ t : -f \} \]

\[ ts : \gamma, \mu \]

\((\circ_1)\)

Edward Stabler, UCLA

Factored grammar and performance models
Movement in MCFG

\[
\begin{align*}
\epsilon ::= & V +wh C \\
\text{they} ::= & D \\
\text{race} ::= & D = D V \\
\text{which} ::= & N D -wh \quad \text{horse} ::= N
\end{align*}
\]

\[
\begin{align*}
\langle 0, C \rangle (\text{which horse they race}) \\
\langle 0, +wh C, -wh \rangle (\text{they race, which horse}) \\
\langle 1, = V +wh C \rangle (\epsilon) \\
\langle 0, V, -wh \rangle (\text{they race, which horse}) \\
\langle 0, = D V, -wh \rangle (\text{race, which horse}) \\
\langle 1, = D = D V \rangle (\text{race}) \\
\langle 0, D -wh \rangle (\text{which horse}) \\
\langle 1, = N D -wh \rangle (\text{which}) \\
\langle 1, N \rangle (\text{horse})
\end{align*}
\]

\[
\begin{align*}
\langle 0, = D V, -wh \rangle (x, y) & \rightarrow \langle 1, = D = D V \rangle (x) \quad \langle 0, D -wh \rangle (y) \\
\langle 0, C \rangle (yx) & \rightarrow \langle 0, +wh C, -wh \rangle (x, y)
\end{align*}
\]
MG vs MCFG movement: significantly different

- MG treats movement configurations \([+f_\alpha] \ldots [-f_\beta] \ldots\) alike, but MCFG needs a separate rule for every instance.

- This allows us to prove: MGs can be exponentially smaller than strongly equivalent MCFGs.

For any \(k\) we show how to define MG with \(k\) movers that can be introduced to a XP in any order; any equivalent MCFG needs at least \(2^k\) rules. \(\square\)
I do not know whether English is... literally beyond the bounds of phrase structure description... When we turn to the question of the complexity of description... however, we find that there are ample grounds for the conclusion that this theory of linguistic structure is fundamentally inadequate. (Chomsky’56, p.119)

- Pullum: “[A weak] non-CF-ness result itself, Chomsky has repeatedly told us, is of little importance.”

- Does the idea that Gs should distinguish movements across categorial differences have a bearing on performance models? Is it supported by evidence from performance?
In most models, grammar size $\propto$ parser size.

Often both grammar + ops explicit: 1 step/node in derivation...
Model: TD beam parser for MGs

\[
\begin{align*}
\epsilon::=& V \ C \quad \text{knows::=} C = D \ V \\
\epsilon::=& V + wh \ C \quad \text{says::=} C = D \ V \\
\epsilon::=& V \quad \text{prefers::=} D = D \ V \\
\epsilon::=& V \quad \text{drinks::=} D = D \ V \\
\end{align*}
\]

\[
\begin{align*}
\epsilon::=& V \\
\epsilon::=& V + wh \\
\epsilon::=& = C \\
\epsilon::=& knows \\
\epsilon::=& says \\
\epsilon::=& prefers \\
\epsilon::=& drinks \\
\epsilon::=& D \\
\epsilon::=& = C \\
\epsilon::=& = D \\
\epsilon::=& king \\
\epsilon::=& queen \\
\epsilon::=& wine \\
\epsilon::=& beer \\
\epsilon::=& the \\
\epsilon::=& which \\
\epsilon::=& the
\end{align*}
\]

Edward Stabler, UCLA  
Factored grammar and performance models
<table>
<thead>
<tr>
<th>step</th>
<th>remaining input</th>
<th>rule</th>
<th>queue</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>which wine the queen prefers start</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

```
C
=V +wh
ε =V
ε knows

V

=D king

ε says

N
=queen wine beer

ε prefers

D

=beer

ε drinks

=the

-D

-wh

N
=which
```
2 arguments + a model
Conclusions
MG vs. MCFG
A performance model
MG+\phi Agree vs. MG
Traces or not?

1. which wine the queen prefers

\[
\begin{align*}
C &= V \\
V &= +wh \\
\epsilon &= V \\
\epsilon &= knows \\
\epsilon &= says \\
\epsilon &= prefers \\
\epsilon &= drinks \\
N &= V \\
N &= D \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &= N \\
D &
Factored grammar and performance models

2 arguments + a model
Conclusions

MG vs. MCFG
A performance model
MG+φAgree vs. MG
Traces or not?

2. which wine the queen prefers

\[
\begin{array}{c}
\text{step} \quad \text{remaining input} \quad \text{rule queue} \\
2. \quad \text{which wine the queen prefers} \quad \bullet_1 \quad 1 \quad 2
\end{array}
\]

\[
C = V + wh = D \quad \text{king} \quad \text{queen} \quad \text{wine} \quad \text{beer} = N \quad D
\]

\[
\epsilon = V \quad \text{knows} \quad \text{says} \quad \text{prefers} \quad \text{drinks} \quad \epsilon = D \quad \text{the} \quad \text{which} = N
\]

Edward Stabler, UCLA
3. which wine the queen prefers

C

=V +wh

=V

ε

ε knows says prefers drinks

V

=V

ε

=V

κing queen wine beer =N

D

prefers drinks which

N

N

D

the =N

D

wh

Factored grammar and performance models
2 arguments + a model

Conclusions

MG vs. MCFG

A performance model

MG+φAgree vs. MG

Traces or not?

---

Factored grammar and performance models

---

Edward Stabler, UCLA
2 arguments + a model
Conclusions
MG vs. MCFG
A performance model
MG+φAgree vs. MG
Traces or not?

step | remaining input | rule queue
---|-----------------|----------
5.  | which wine the queen prefers | ●₁ 1 2 3 4 5

C
\[ \epsilon \]
\[ =V \]

+wh

\[ =V \]

\[ \epsilon \]

\[ \epsilon \]

V
\[ \epsilon \]

\[ =C \]

knows

says

N
\[ \epsilon \]

\[ =D \]

prefers
drinks

D
\[ =N \]

the

beer

king

queen

wine

which

Edward Stabler, UCLA
Factored grammar and performance models
Conclusions

MG vs. MCFG

Factored grammar and performance models

MG+φ Agree vs. MG

Traces or not?

Edward Stabler, UCLA

2 arguments + a model

A performance model

MG vs. MCFG

MG+φ Agree vs. MG

Traces or not?
2 arguments + a model

Conclusions

MG vs. MCFG

A performance model

MG+ φ Agree vs. MG

Traces or not?

---

Factored grammar and performance models
2 arguments + a model
Conclusions
MG vs. MCFG
A performance model
MG+ϕAgree vs. MG
Traces or not?

Edward Stabler, UCLA
Factored grammar and performance models
Conclusions

MG vs. MCFG

A performance model

MG+φ Agree vs. MG

Traces or not?

Factored grammar and performance models

Edward Stabler, UCLA
which wine the queen prefers scan 1 2

C
\[=V \quad +wh\]
\[\varepsilon \quad =V =C =D\]
\[\varepsilon \quad knows \quad says \quad prefers \quad drinks\]

V
\[=D \quad king \quad queen \quad wine \quad beer =N =D\]

N
\[-wh\]

D
\[the =N wh which\]
which wine the queen prefers scan 1

C

V

N

D

-wh

=V +wh

=D king queen wine beer =N D

ε =V

=ε knows says prefers drinks

ε

ε

Edward Stabler, UCLA

Factored grammar and performance models
step remaining input rule queue

12. which wine the queen prefers scan $\epsilon$

- this method works for any MG (sound, complete)
- MG representation is transparent; succinctness evident
Independent evidence: movement distinguished

- **locality effects in movement**:  
  - self-paced reading effect $\propto$ distance to antecedent (Gibson’98; Hale’03; Bartek&al’11, ...)
  - island effects (Aoshima&al’09, Sag&al’07, Yoshida’06, ...)

- **patterns of acquisition**: acquisition of wh-movement, etc.  
  (Friedmann&Lavi’06, vanKampen97’)

- **neural correlates?**  
  Brodman 45 for movement (Santi&Grodzinsky’10)
\[ \phi \text{ agreement} \]

Agree domain: Bejar & Rezac 2009

Kobele’11: Many proposals of this kind are regular constraints: Graf’11: enforceable in the MG category system
ϕ agreement in MGs

How to allow a derivation like this for each ϕ specification \( a, b \) that the language allows?

\[
\begin{align*}
\epsilon &::= V + \text{wh} \ C \\
\text{they}_b &::= D_b \\
\text{race}_{ab} &::= D_a = D_b \ V \\
\text{which}_a &::= N D_a - \text{wh} \ 	ext{horse} ::= N
\end{align*}
\]
φ agreement in MGs

If DP has \( k \) features with \( j \) values, then \( j^k \) possibilities. If \( n \) args agree, ‘categorial infrastructure’ multiplied by \( 2^{kn} \), missing generalizations about match of verb+arguments.
suppose each arg has its own probe

then instead of $j^{nk}$ grammar, we have $nj^k$

$(U1)$ perhaps $n \leq 2^{1/2}$?

$(U2)$ perhaps $j, k$ bounded too? still $j^{nk}$ can be large
MG+$\phi$Agr

Each $\phi$Agr probe realized by regular tree automaton

States of each $\phi$Agr probe not multiplied through Lex, parser smaller, generalization captured.
2 arguments + a model
Conclusions

MG vs. MCFG
A performance model
MG+φAgree vs. MG
Traces or not?

(D) derivation tree

ε::=V +wh C

race::=D =D V

which::=N D -wh horse::N

(T) derived tree (traces)

CP

DP(0)

D'

C'

VP

which N' D' V' D P

horse N they race t(0)

Edward Stabler, UCLA
Factored grammar and performance models
Insignificantly different alternatives

(D) derivation tree

\[ \epsilon ::= V + \text{wh} C \]
\[ \text{race} ::= D = D V \]
\[ \text{which} ::= N D - \text{wh} \text{ horse} ::= N \]

(M) derived tree (multidominance)

\[ \epsilon ::= V + \text{wh} C \]
\[ \text{race} ::= D = D V \]
\[ \text{which} ::= N D - \text{wh} \text{ horse} ::= N \]

D, T, M are easily, $O(n)$ computable from each other.

(Graf’11; Kobele’11; Kobele&al’07; Mönich’07)
Varieties of structure dependence

[subj-aux inv] refers to the abstract label “noun phrase,” a grouping of words into constituents, and consequently is called *structure dependent* -Berwick, Pietroski, Yankama, Chomsky ’11

- Rules of Gs here define constituents, ’groupings of words’ like NP
- MG rules ‘groups constituents’ that trigger movement, and MG+ϕAgree ‘groups constituents’ relevant to agreement.

⇒ non-construction-specific, more succinct, captures generalizations.

Stipulation of grammatical constructions (interrogative, passive, etc.), with their independent properties, was overcome . . . by analyzing them into components that function generally, also eliminating redundancies – Chomsky ’12
Questions about models and methods

Q1  How should models be developed and empirically assessed?
   * Desc adequacy sets stage for next questions
     Math. characterization of classes of possibilities!

Q2  Most significant gaps in our understanding?
    q1  What are the basic mechanisms of grammar?
        Beyond level 1: what is basic, how to defend?
        What performance measures bear most directly on grammar?
    q2  How does the language learner work?

Q3  Experimental approaches vs computational modeling??
   * q1 assessment difficult. cf. Newton/Hooke, field/math bio
**APPENDICES**
A. Relative succinctness of MGs vs. MCFGs

For each $i \in \mathbb{N}$, the lexicon of $MG_i$ contains the lexical items specified in (i-iv) and nothing else, with $A$ the ‘start’ category:

i. the following lexical item, with $i + 2$ syntactic features:

$$a :: =B +1+2 \ldots +i A$$

ii. the following lexical item, with 3 syntactic features:

$$b :: =B=C B$$

iii. the following lexical item, with 1 syntactic feature:

$$d :: B$$

iv. And for each $1 \leq j \leq i$, this lexical item with 2 syntactic features,

$$c :: =C -j.$$
\[ MG_2 \text{ has } (3 \times 2) + 6 = 12 \text{ features; } 2! \text{ derivations of } c^2ab^2d: \]
B. Relative succinctness of $\text{MG} + \phi \text{Agr}$ vs. $\text{MG}$

We define a series of MGs, $\text{MG}_i$ for $i = 1, 2, \ldots$ For simplicity, assume each argument is a pronoun $k = 1$ feature with $j = 2$ values, 0 or 1; each verb in $\text{MG}_n$ selects and agrees with $n$ arguments. So $\text{MG}_n$ needs, for $j, j_i \in \{0, 1\}$,

$$\text{pronoun-}j::D_j$$
$$\text{verb-}j_n\ldots- j_1::=D_{j_n}\ldots=D_{j_1} \text{ V}$$

That is $2 + 2^n$ lexical items in each $\text{MG}_n$.

For comparison, we can use probes to define the same languages. 2 lexical items define what Baker (2001) calls ‘categorial infrastructure’:

$$\text{pronoun::D}$$
$$\text{verb::=}D\ldots=D \text{ V}$$

For each of the $n$ arguments of $\text{V}$, an automaton (probe) matches an affix 0 or 1 with the corresponding affix of $\text{V}$. (Assume verb,pronoun dominate their affixes, so the automation can find them.) That’s 2 lexical items and a collection of $n$ automata each of which has a size $\mathcal{O}(n)$.
C. Structure dependence vs. linear order

Chomsky (1968, 1971, 1975, 1980a,b) discusses ‘structure dependence’ in many places, but leaves it not quite clear. G is structure dependent iff it has rules that generate expressions from other expressions, so that it can, for example, have a rule that applies to ‘groupings of words’ like the ‘noun phrases’ to form other expressions like ‘determiner phrases’. In this sense...

- ‘Orthogonal to’ nested, hierarchical structure?
  No. Grammars for nested structure are all structure dependent.
  e.g. \([a] [b] [c]\) => \([[b] [[a] [c]]]\) refers to nested structures

- Distinct from ‘linear structure’? What is that?
  ‘the nth verb’ is definable even in a regular grammar.

- Hierarchical structure unable to count?
  No. Regular grammars can count to \(k\); others arbitrarily high.
D. The TD beam parser for MG

Like CF beam parser except that stack is replaced by priority queue, sorted by linear precedence as explained in Stabler’11,’12. Basically just inverting the standard bottom-up MG definitions, the parsing rules for MGs in Stabler’12, used in the example above:

\[
\text{input}, \ (C(x), \emptyset)_\varepsilon \quad \text{(start) } \ell[C(x)], \text{ for start category } C
\]

\[
w \ast \text{input}, \ (t[w], \emptyset)_i \ast q \quad \text{(scan)} \Rightarrow \text{input}, \ q
\]

\[
\text{input}, \ (t=[f(x)], \mu)_i \ast q \quad \text{(\(\bullet_1\)) } \ell[f(y)] \land \Sigma x \neq \varepsilon
\]

\[
\text{input}, \ (=f(\Sigma x), \emptyset)_i \ast (f(y), \mu)_i \ast q
\]
\[\begin{align*}
\text{input, } (t=[f(x)], \mu)_i * q & \quad (\bullet_2) \quad \ell[f(y)] \land \Sigma x \neq \epsilon \\
\text{input, } (=f(\Sigma x), \mu)_i_1 * (f(y), \emptyset)_i_0 * q & \\
\text{input, } (t=[f(x)], u[f(y)]_j \cup \mu)_i * q & \quad (\bullet_3) \quad \Sigma x \neq \epsilon \\
\text{input, } (=f(\Sigma x), \emptyset)_i * (f(y), \mu)_j * q & \\
\text{input, } (t=[f(x)], u[f(y)]_j \cup \mu)_i * q & \quad (\bullet_4) \quad \Sigma x \neq \epsilon \\
\text{input, } (=f(\Sigma x), \mu)_i * (f(y), \emptyset)_j * q & \\
\text{input, } (t+[f(x)], \mu)_i * q & \quad (\circ_1) \quad \ell[-f(y)] \\
\text{input, } (+f(x), -f(y))_i_0 \cup \mu)_i_1 * q & \\
\text{input, } (t+[f(x)], u[-f(y)]_j * \mu)_i * q & \quad (\circ_2) \\
\text{input, } (+f(x), -f(y))_j \cup \mu)_i * q &
\end{align*}\]
Bartek, B., Lewis, R. L., Vasishth, S., and Smith, M. R.
In search of on-line locality effects in sentence comprehension.

Barton, G. E., Berwick, R. C., and Ristad, E. S.
Computational Complexity and Natural Language.

Béjar, S., and Rezac, M.
Cyclic agree.
Linguistic Inquiry 40, 1 (2009), 35–73.

Chomsky, N.
Three models for the description of language.

Chomsky, N.
Language and Mind.

Chomsky, N.
Problems of Knowledge and Freedom: The Russell Lectures.

Chomsky, N.
On cognitive structures and their development.

Chomsky, N.
Problems of projection.
Lingua forthcoming (2012).

Culy, C.
The complexity of the vocabulary of Bambara.
Linguistics and Philosophy 8, 3 (1985), 345–352.

Friedmann, N., and Lavi, H.
On the order of acquisition of A-movement, wh-movement and V-C movement.

Gibson, E.
Linguistic complexity: Locality of syntactic dependencies.

Graf, T.
Closure properties of minimalist derivation tree languages.

Graf, T.
Locality and the complexity of minimalist derivation tree languages.
to appear.

Hale, J.
Grammar, Uncertainty, and Sentence Processing.

Harkema, H.
A characterization of minimalist languages.

Harkema, H.

*Parsing Minimalist Languages.*

Kobele, G. M.

Minimalist tree languages are closed under intersection with recognizable tree languages.

Michaelis, J.

Derivational minimalism is mildly context-sensitive.

Michaelis, J.

*On Formal Properties of Minimalist Grammars.*
*Linguistics in Potsdam* 13, Universitätsbibliothek, Potsdam, Germany.

Michaelis, J.

Transforming linear context free rewriting systems into minimalist grammars.

Preminger, O.

*Agreement as a Fallible Operation.*

Pullum, G. K.
Footloose and context-free.

Sag, I., Hofmeister, P., and Snider, N.
Processing complexity in subjacency violations: the complex noun phrase constraint.

Santi, A., and Grodzinsky, Y.
fMRI adaptation dissociates syntactic complexity dimensions.

Shieber, S. M.
Evidence against the context-freeness of natural language.

Stabler, E. P.
Computational perspectives on minimalism.

Stabler, E. P.
After GB theory.

Stabler, E. P.
After GB theory.
Bayesian, minimalist, incremental syntactic analysis.

van Kampen, J.
*First steps in wh-movement.*

Vijay-Shanker, K., Weir, D., and Joshi, A. K.
Characterizing structural descriptions produced by various grammatical formalisms.

Yoshida, M.
*Constraints and Mechanisms in Long-Distance Dependency Formation.*