Ordered Context-Free Grammars

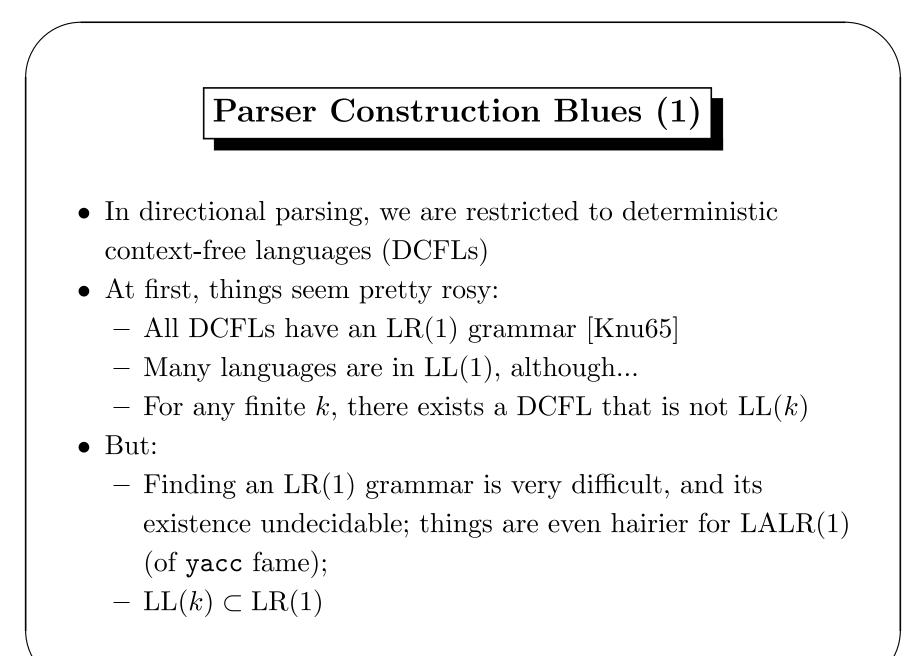
Ziemowit Laski

Department of Information and Computer Science University of California, Irvine Irvine, CA 92697-3425, USA laski@ics.uci.edu

April 30, 1999 (Updated: June 16, 1999)

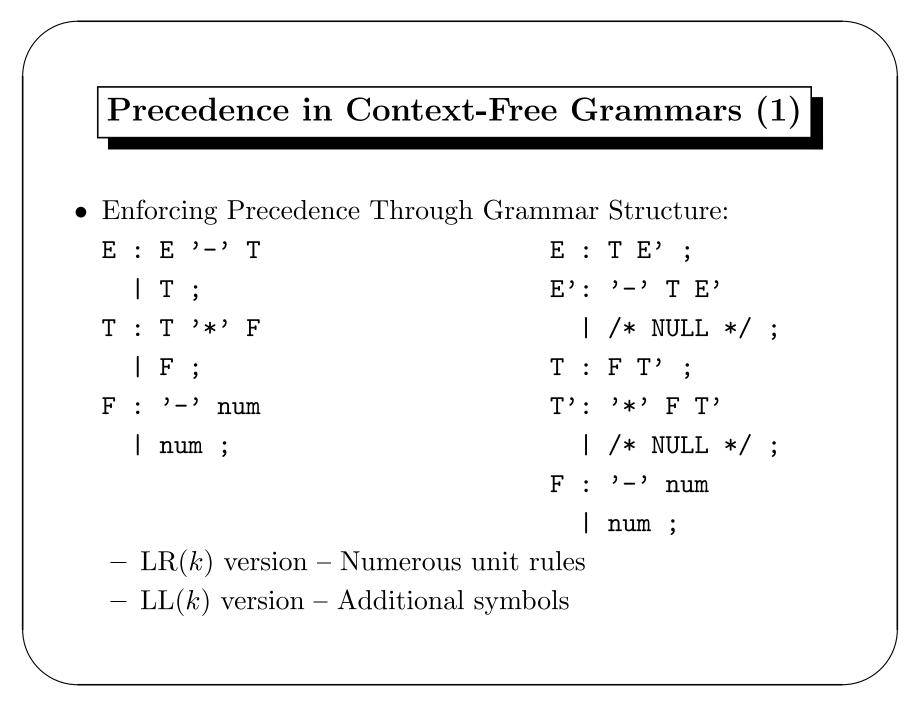
Overview

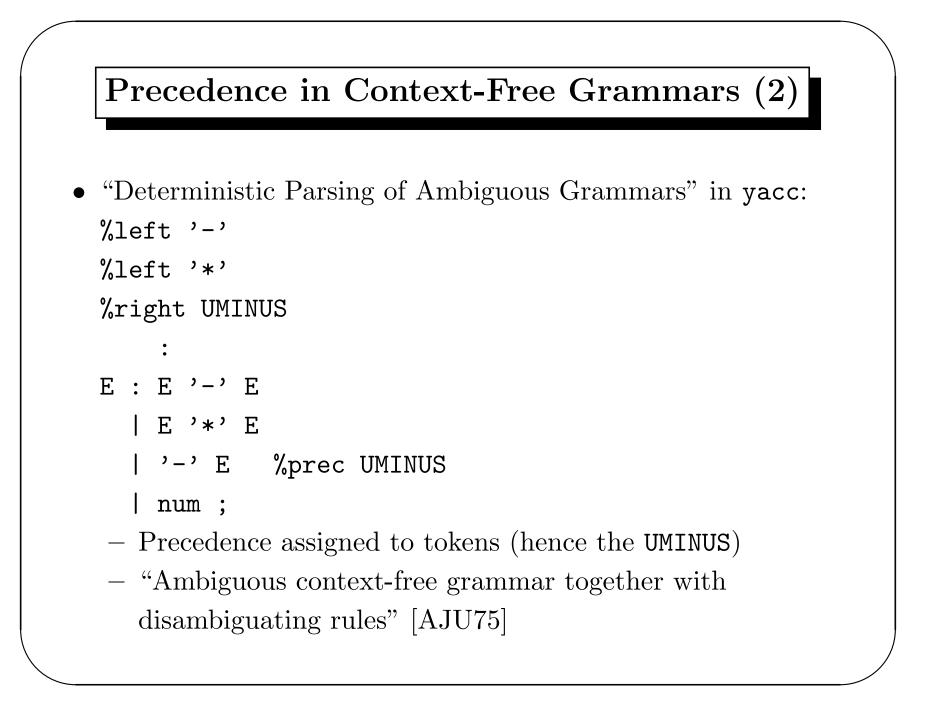
- Parser Construction Blues
- Precedence in Context-Free Grammars
- Ordered Context-Free Grammars
- Derivation Properties of OCFGs
- Parsing: $\mathfrak{Bertha}^{\mathsf{TM}}$ to the Rescue
- Possible Future Work

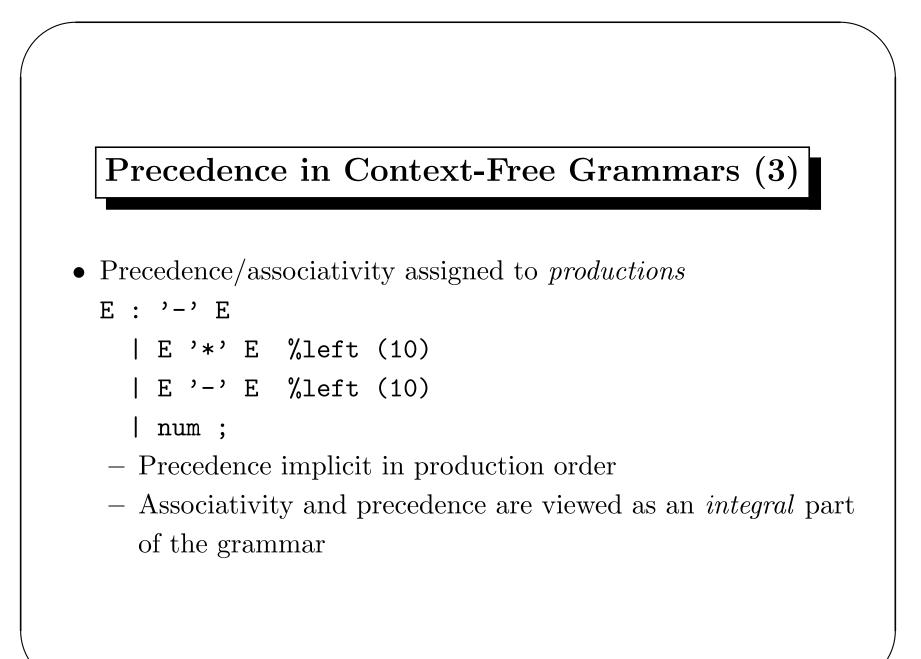


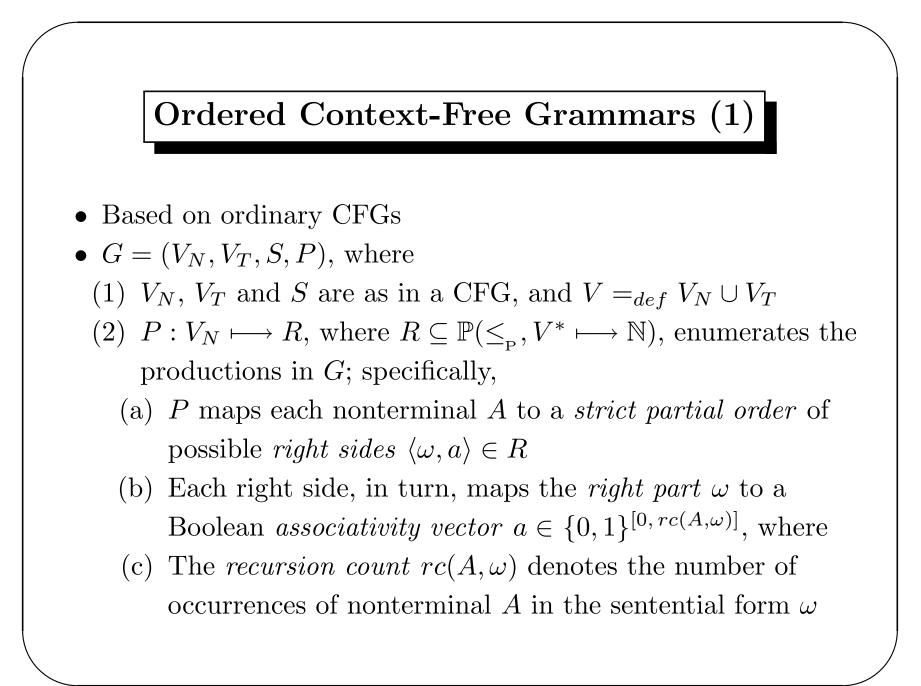
Parser Construction Blues (2)

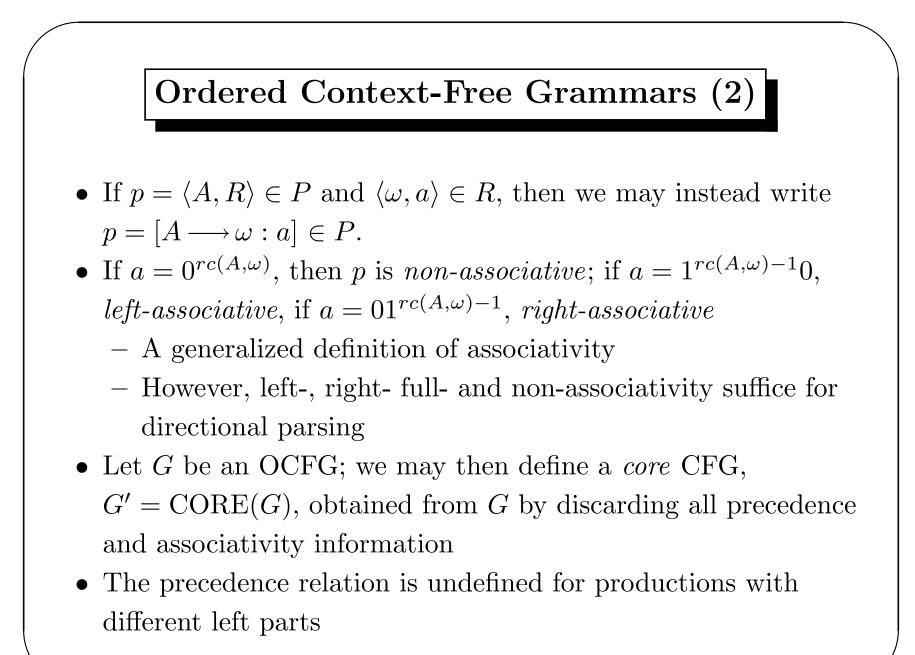
- Even if is possible to find a suitable LL(k) or LALR(1) grammar for our DCFL,
 - "Language translation is a harder and more important problem than language recognition." [PQ96]
 - Semantic bindings (i.e., associating grammar rules with actions) become very tricky
 - Beneficial to leave the structure of the grammar intact, even at the expense of larger lookahead (k > 1)
- Our claim: Precedence is an alternative/complementary disambiguation tool to k > 1





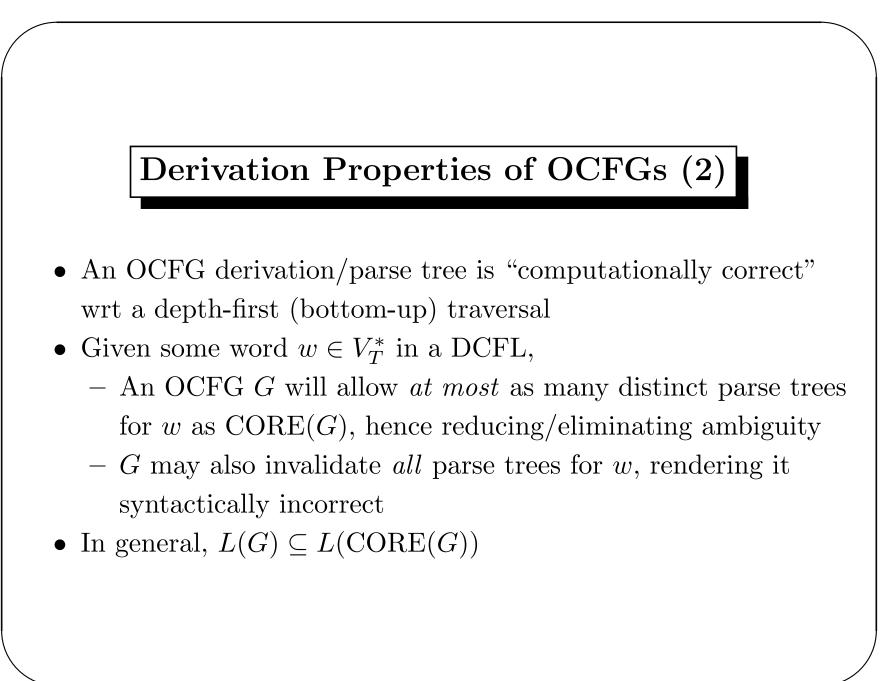






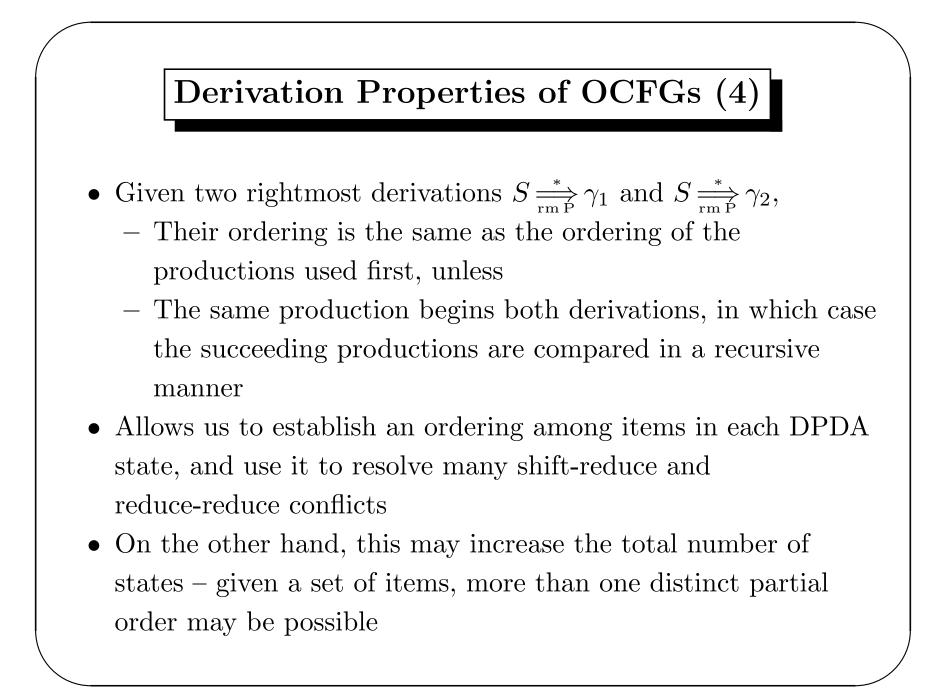
Derivation Properties of OCFGs (1)

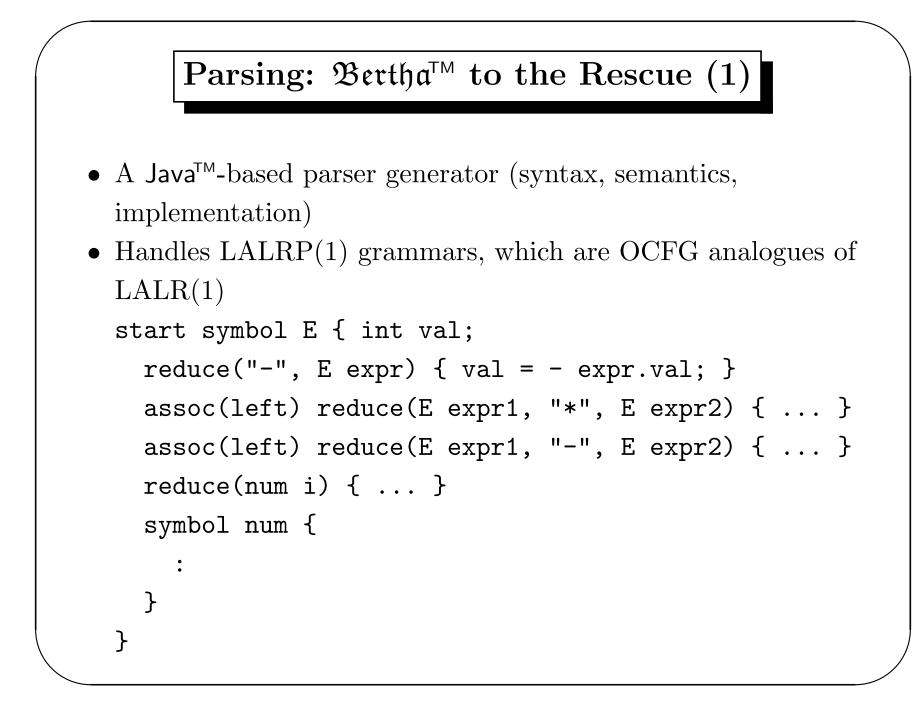
- An OCFG derivation $S \stackrel{*}{\Longrightarrow} \gamma$ is a CFG derivation $S \Longrightarrow \gamma$ that additionally respects the following invariant:
 - If two productions with the same left part $A \in V_N$ are used in *immediate* succession to construct some path in the derivation tree, then either:
 - (1) The production being used first has lower precedence, or
 - (2) The two productions have incomparable precedence, or
 - (3) They have equal precedence, in which case the associativity of the parent production $[A \longrightarrow \omega : a]$ must be obeyed:
 - (a) The *i*-th occurrence of A in ω may be expanded via the child production only if a[i] = 1.
 - (b) Productions with higher precedence may expand any A in ω .

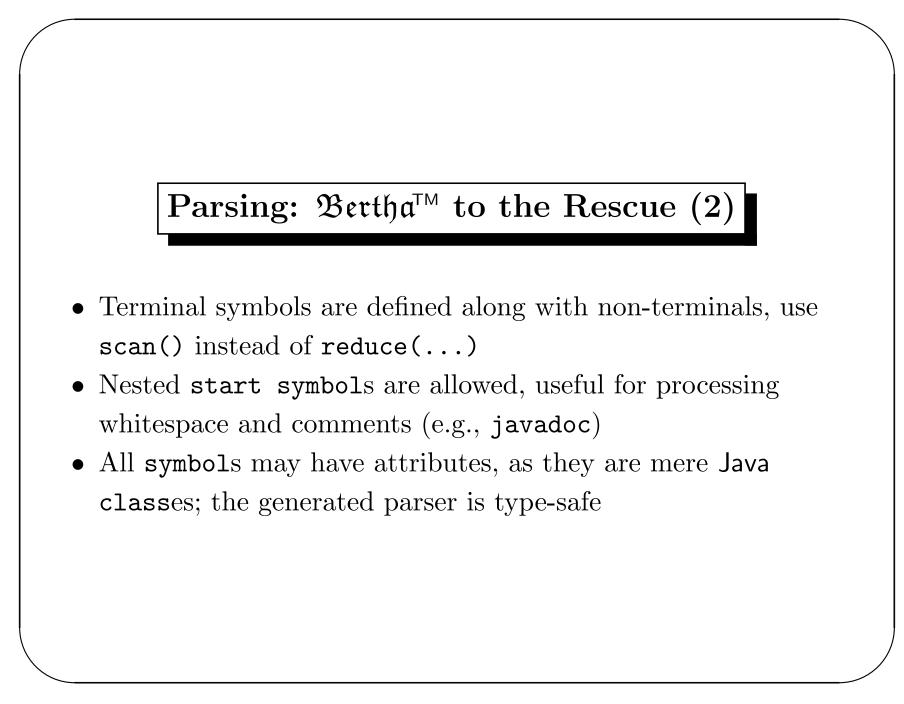


Derivation Properties of OCFGs (3)

- Given a partial order \leq_{P} among productions in G with the same left part, we can induce a partial order $\leq_{rm P}$ among entire OCFG derivations $S \xrightarrow{*}_{rm P} \gamma$ in G
 - Must choose consistent method of expanding nonterminals in intermediate sentential forms
 - Righttmost derivation seems the logical choice
 - * Corresponds to a bottom-up, left-to-right scan of the input
 - * Can be nicely mapped to closure computations within parser states, and to state transitions in a deterministic pushdown automaton (DPDA)







Possible Future Work

- Investigate the formal conditions under which the LRP(k) and LALRP(k) parser construction algorithms create additional states, beyond those created by the LR(k)/LALR(k) algorithms
 Empirical observation: they practically never occur
- Apply Pager's [Pag77] or Spector's [Spe88] compaction algorithms to extend **Bertha** to handle LRP(1) grammars
- It is possible to nesting symbols within one another, so...
 - Within the scope of the outer symbol, it is possible to use inherited attributes
 - May lead to some precedence-enhanced variant of left-corner (LC(k)) parsing, a method which combines LL(k) and LR(k) techniques