

Context-Free Languages

- There are languages CFGs can't generate (non-context-free languages), notably those that incorporate *cross-serial dependencies*, such as Swiss German.
- Perhaps more importantly, CFGs are cumbersome and inefficient for representing natural language syntax.
- Most (but not all) modern theories of syntax include a notion of phrase structure (CFG), and then extend it.

Swiss German example (Shieber 1985) (1/2)

...mer d'chind em Hans es huus lönd hälfe aastriiche

...we the children-ACC Hans-DAT the house-ACC let help paint

‘...we let the children help Hans paint the house’

- Cross-serial dependency:
 - *let* governs the case on *children*
 - *help* governs the case on *Hans*
 - *paint* governs the case on *house*

Swiss German example (Shieber 1985) (2/2)

- Define a new language $f(\text{Swiss German}) =$

$$f(\text{d'chind}) = a \quad f(\text{Jan säit das mer}) = w$$

$$f(\text{em Hans}) = b \quad f(\text{es huus}) = x$$

$$f(\text{lönde}) = c \quad f(\text{aastriiche}) = y$$

$$f(\text{hälfe}) = d \quad f([\text{other}]) = z$$

- Let r be the regular language $wa^*b^*xc^*d^*y$.
- $f(\text{SwissGerman}) \cap r = wa^mb^nc^md^ny$
- $wa^mb^nc^md^ny$ is not context-free
- Context free languages are closed under intersection
- \therefore Swiss German is not context-free.

Strongly v. weakly context-free

- A language is *weakly* context-free if the set of strings in the language can be generated by a CFG.
- A language is *strongly* context-free if it is weakly context free and the set of structures assigned to the strings by the CFG are the right ones.
- Shieber's proof shows that Swiss German is *weakly* not context-free and therefore *a fortiori strongly* not context-free.
- A prior paper by Bresnan et al had argued that Dutch was *strongly* not context-free, but the argument was dependent on linguistic analyses.