Hyperedge Replacement Grammar (HRG) Learning

Xiaochang Peng

University of Rochester

Motivation

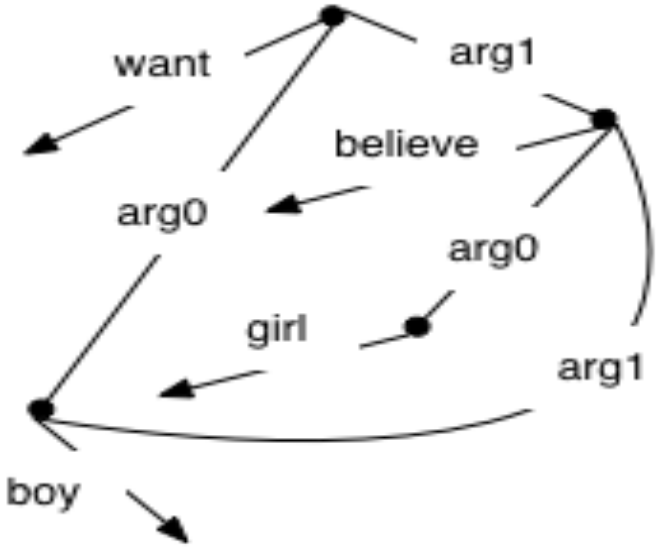
- Learning Hyperedge Replacement Grammar (HRG) and Synchronous HRG (SHRG) from AMR bank
  - Input: AMR graphs and sentences
  - Output: HRG and SHRG rules
- HRG: test the quality of AMR graph
- SHRG: strings to AMRs, AMRs to strings

String to AMR parsing

- Input: English string
- Output: English AMR graph

Input: The boy wants the girl to believe him

Output:



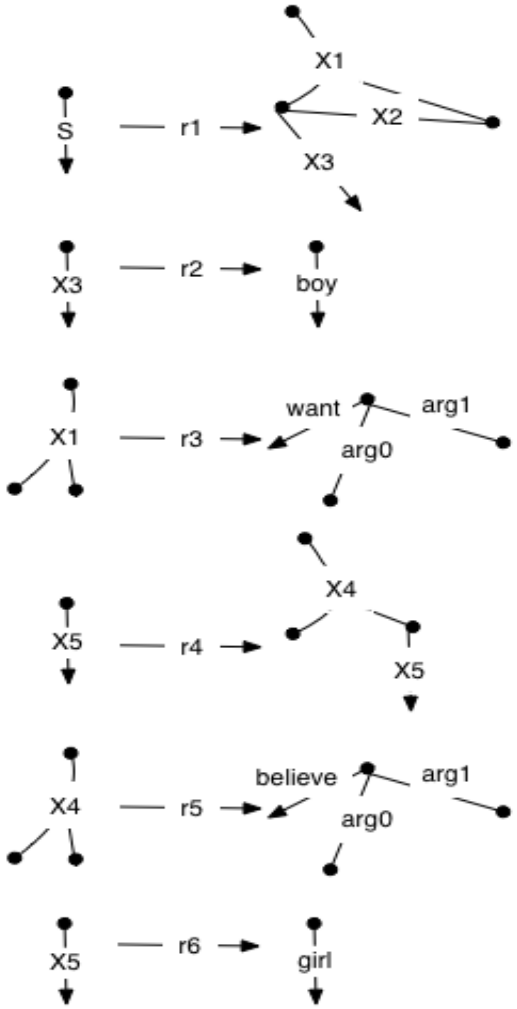
Comparison with other methods

	JAMR	Transition based	SHRG
Input	string	Dependency tree	string
Output	AMR	AMR	AMR
Formalism	grammarless	grammarless	SHRG
Model	Edge-factored	Transition based	generative
Training	perceptron	perceptron	MCMC
Inference	MST	Greedy search	CYK

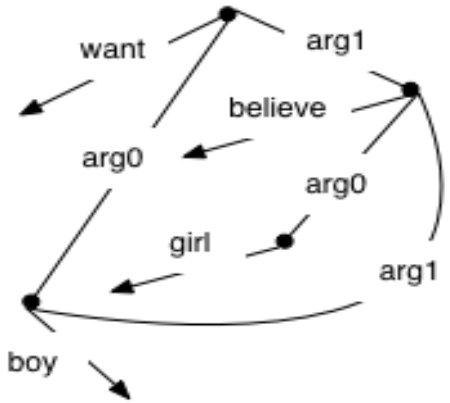
AMR graph parsing

- HRG-based parsing
- Given: HRG rules
- Input: English AMR graph
- Output: the series of rules applied to derive this graph

AMR graph parsing

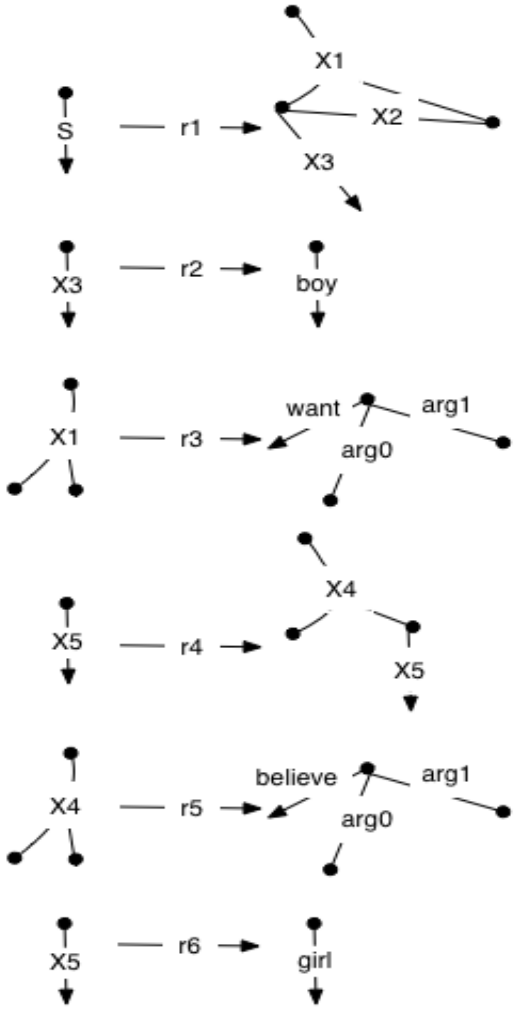


Input:

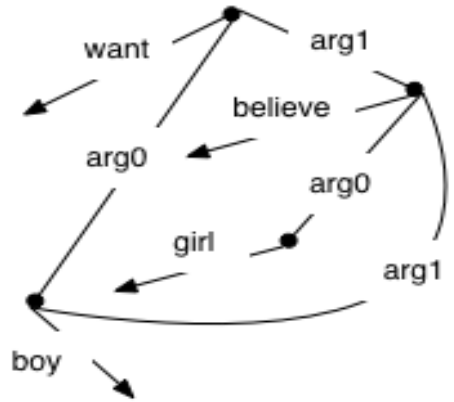


The boy wants the girl to believe him

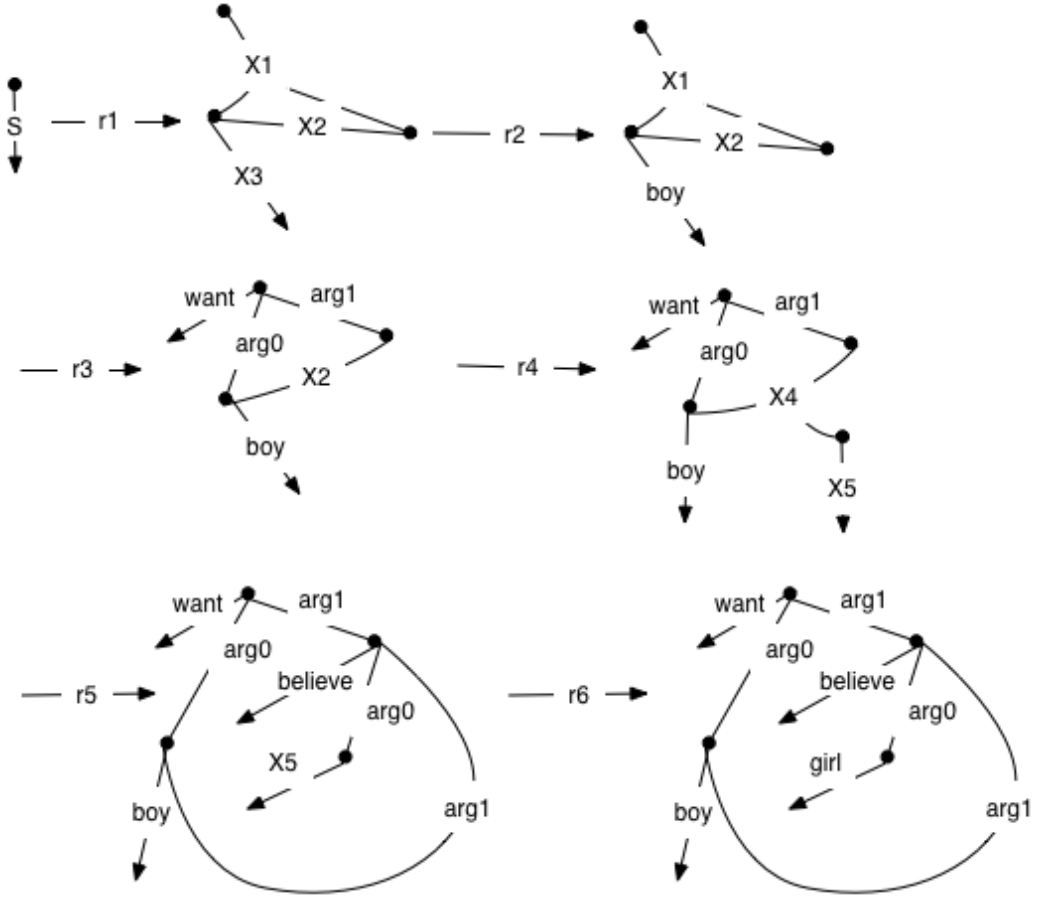
AMR parsing

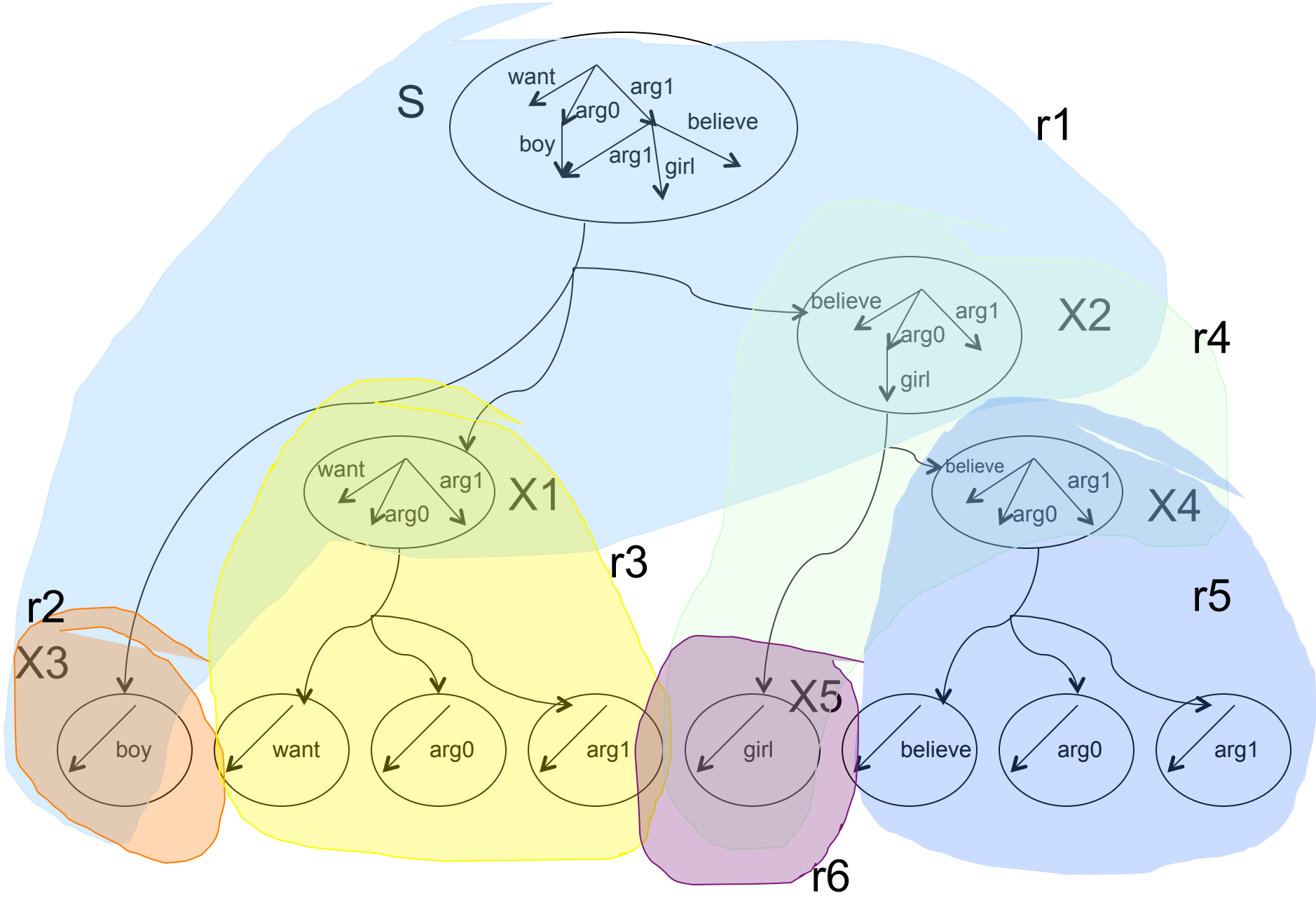


Input:

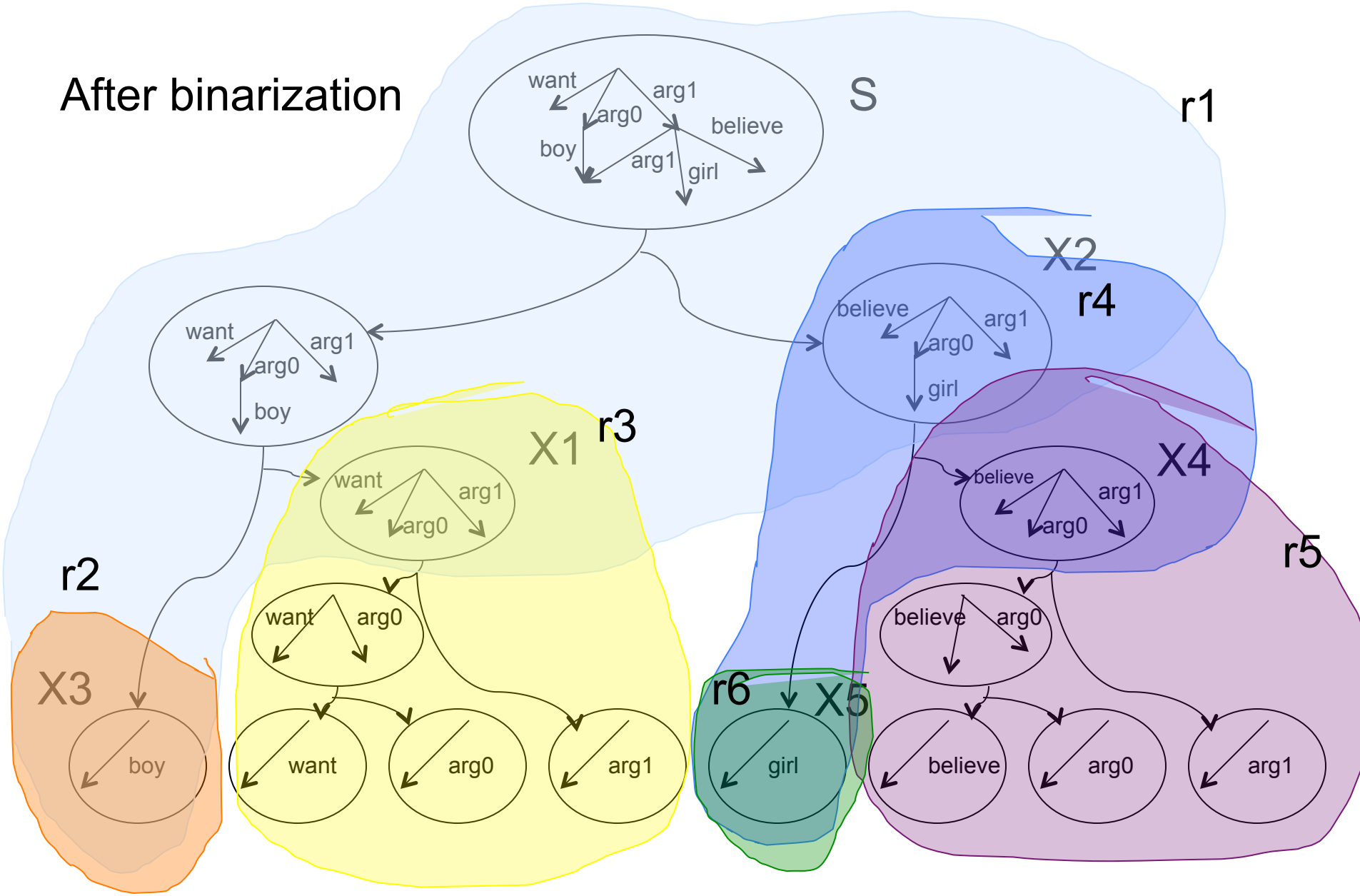


Output:




S
want
arg1
arg0
believe
boy
arg1
girl
r1
believe
arg1
arg0
girl
X2
r4
want
arg1
arg0
X1
believe
arg1
arg0
X4
r3
r2
r5
X3
X5
boy
want
arg0
arg1
girl
believe
arg0
arg1
r6

After binarization



Intuitions for HRG learning

- Context free derivation: graph fragment decomposition tree
- Derivation rule: tree fragment
- HRG learning strategy: build a forest, then sample tree fragment from forest

Problem

- Graph v.s. strings: exponential numbers of fragments v.s. n^2 number of phrases (substrings)
- Our strategy: fragments v.s. substrings, fragment forest v.s. forest of substrings
- A big assumption: projective

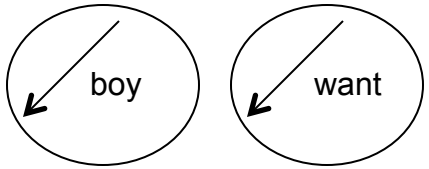
Constructing fragment decomposition forest

- Extracting all concept fragments using JAMR, order them by the position in the strings
- Insert unaligned graph fragments in the middle of aligned concepts fragments
- Keep building larger fragments bottom up

The boy wants the girl to believe him

Constructing fragment decomposition forest

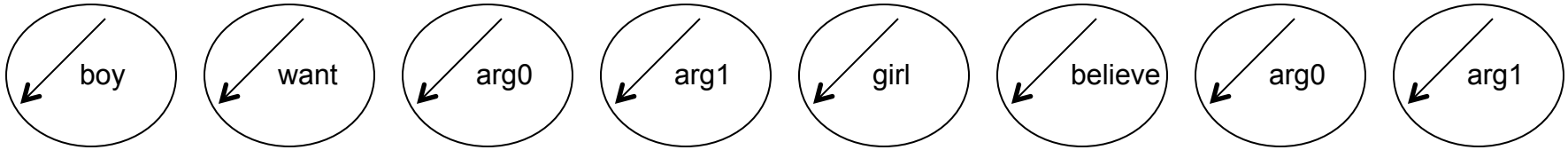
- Extracting all concept fragments using JAMR, order them by the position in the strings
- Insert unaligned graph fragments in the middle of aligned concepts fragments
- Keep building larger fragments bottom up



The boy wants the girl to believe him

Constructing fragment decomposition forest

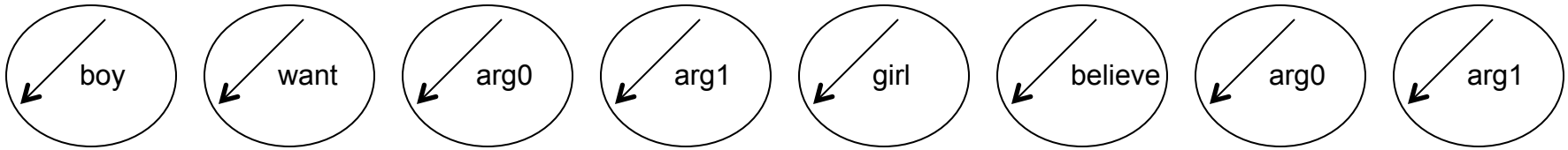
- Extracting all concept fragments using JAMR, order them by the position in the strings
- Insert unaligned graph fragments in the middle of aligned concepts fragments
- Keep building larger fragments bottom up



The boy wants the girl to believe him

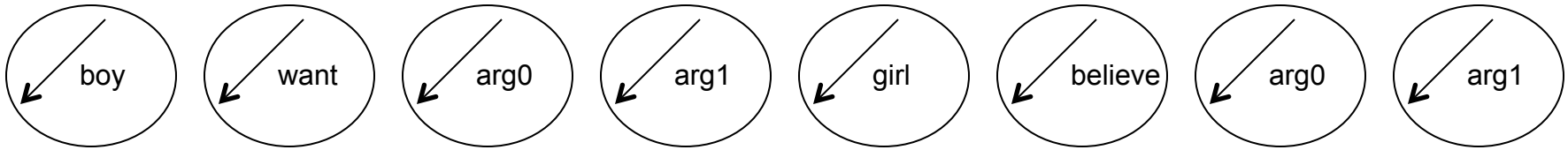
Constructing fragment decomposition forest

- Extracting all concept fragments using JAMR, order them by the position in the strings
- Insert unaligned graph fragments in the middle of aligned concepts fragments
- Keep building larger fragments bottom up



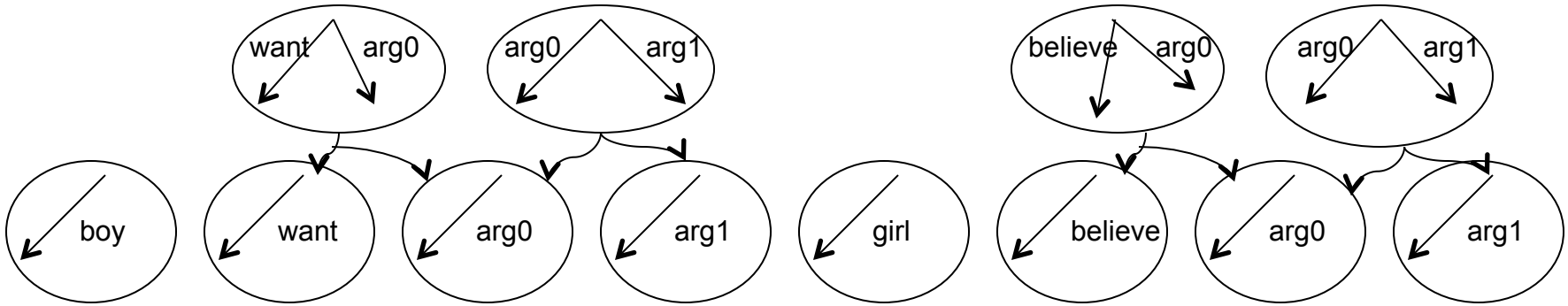
The boy wants the girl to believe him

Building fragment forest

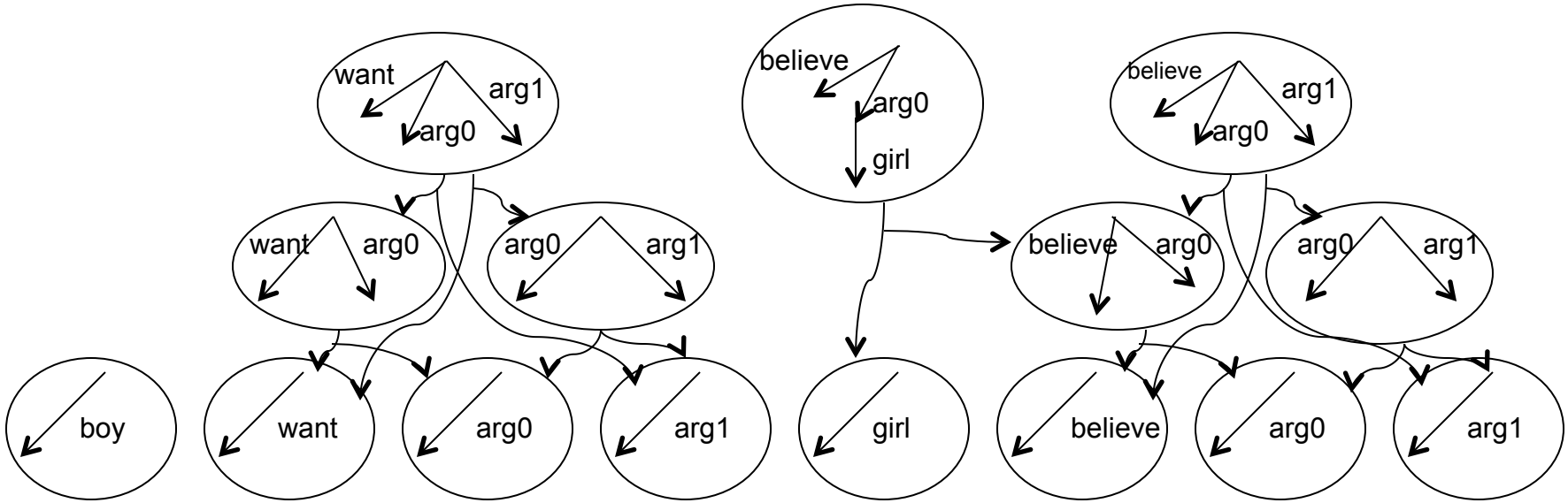


The boy wants the girl to believe him

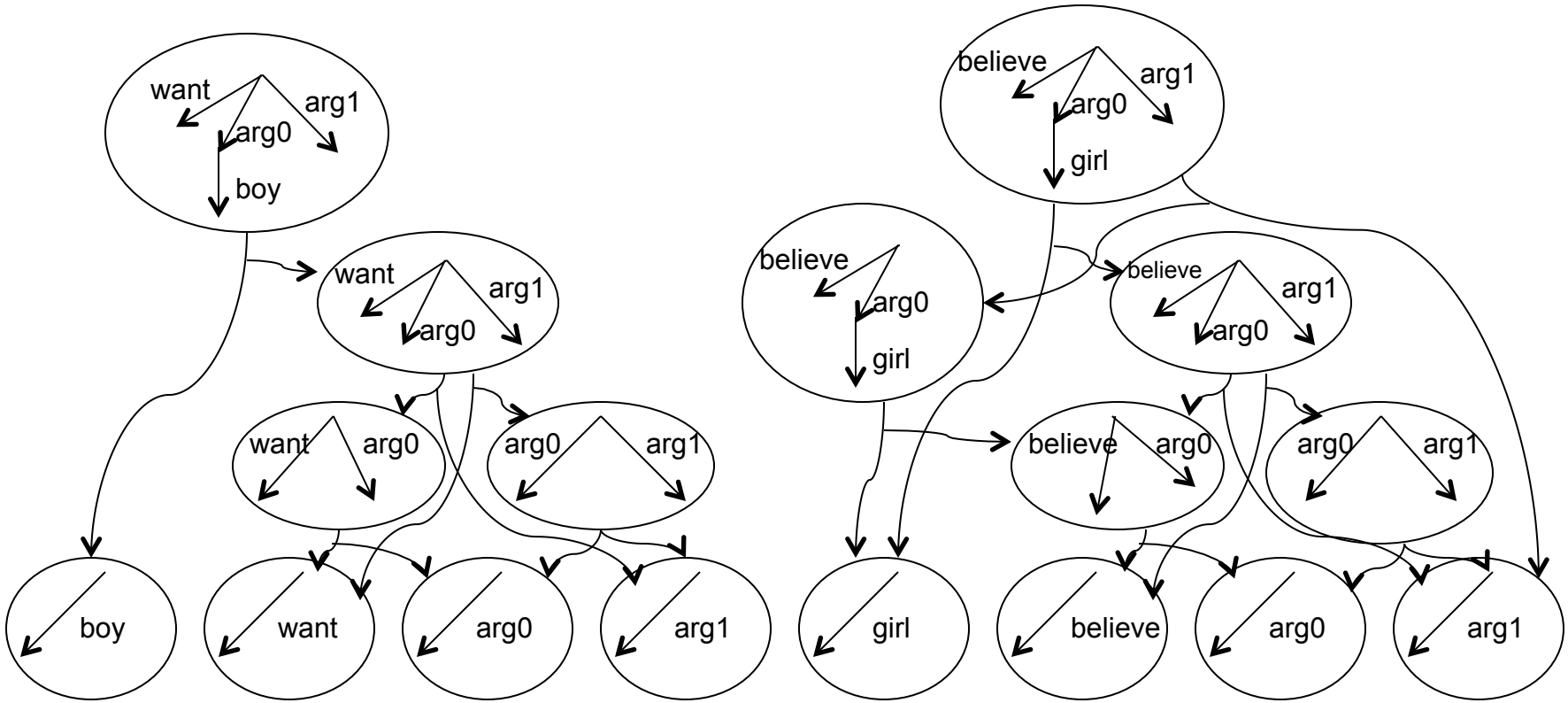
Building fragment forest



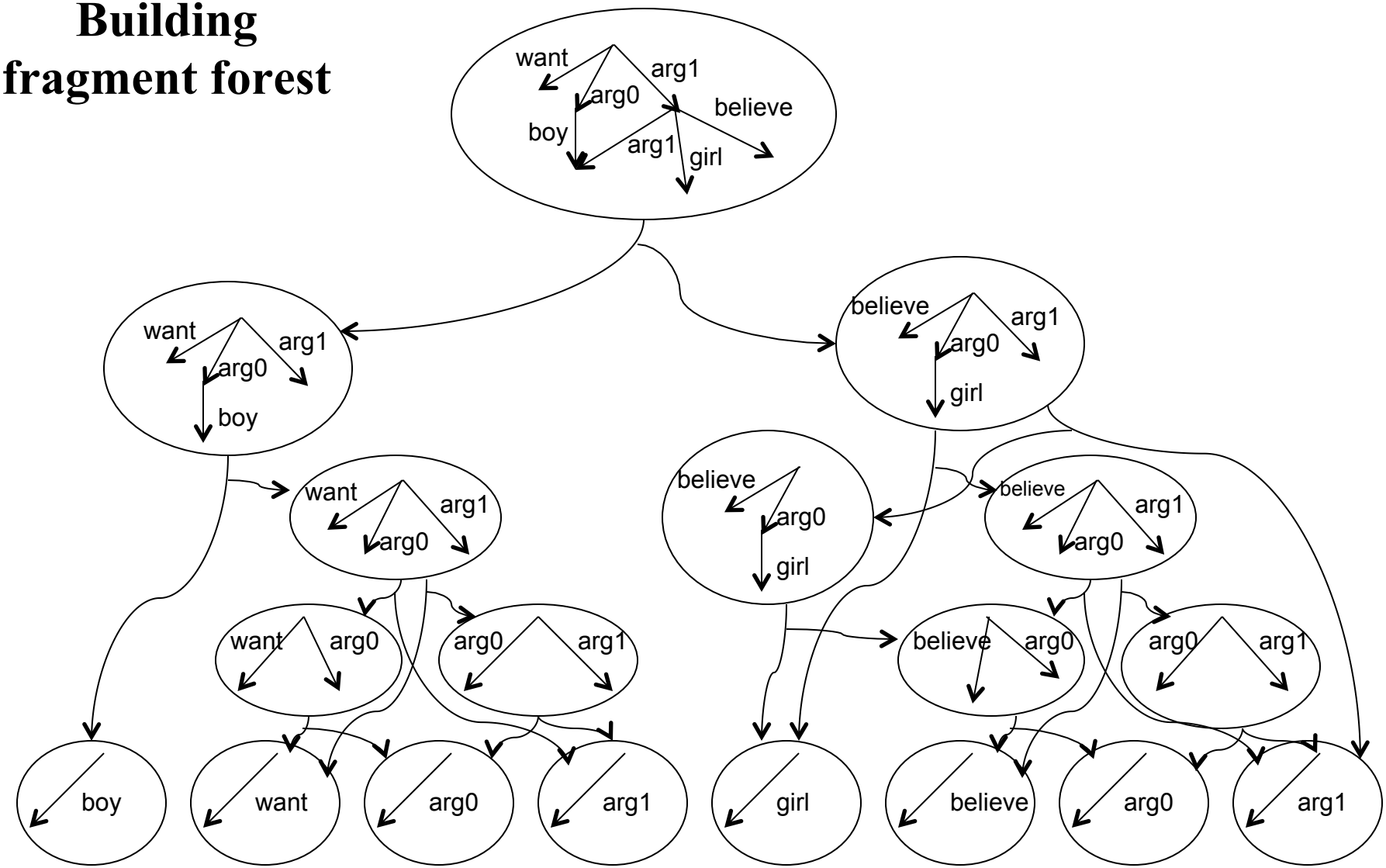
Building fragment forest



Building fragment forest

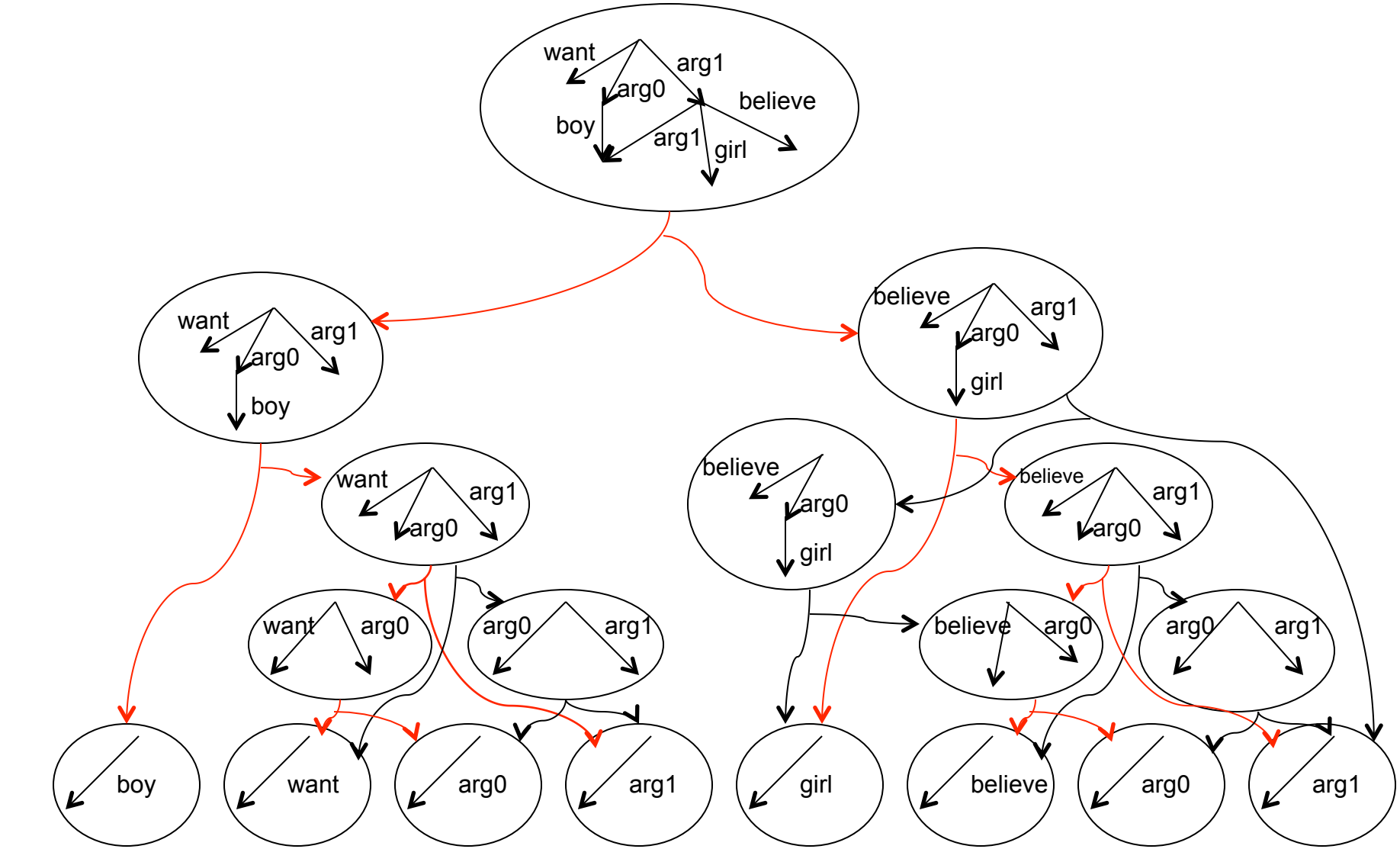


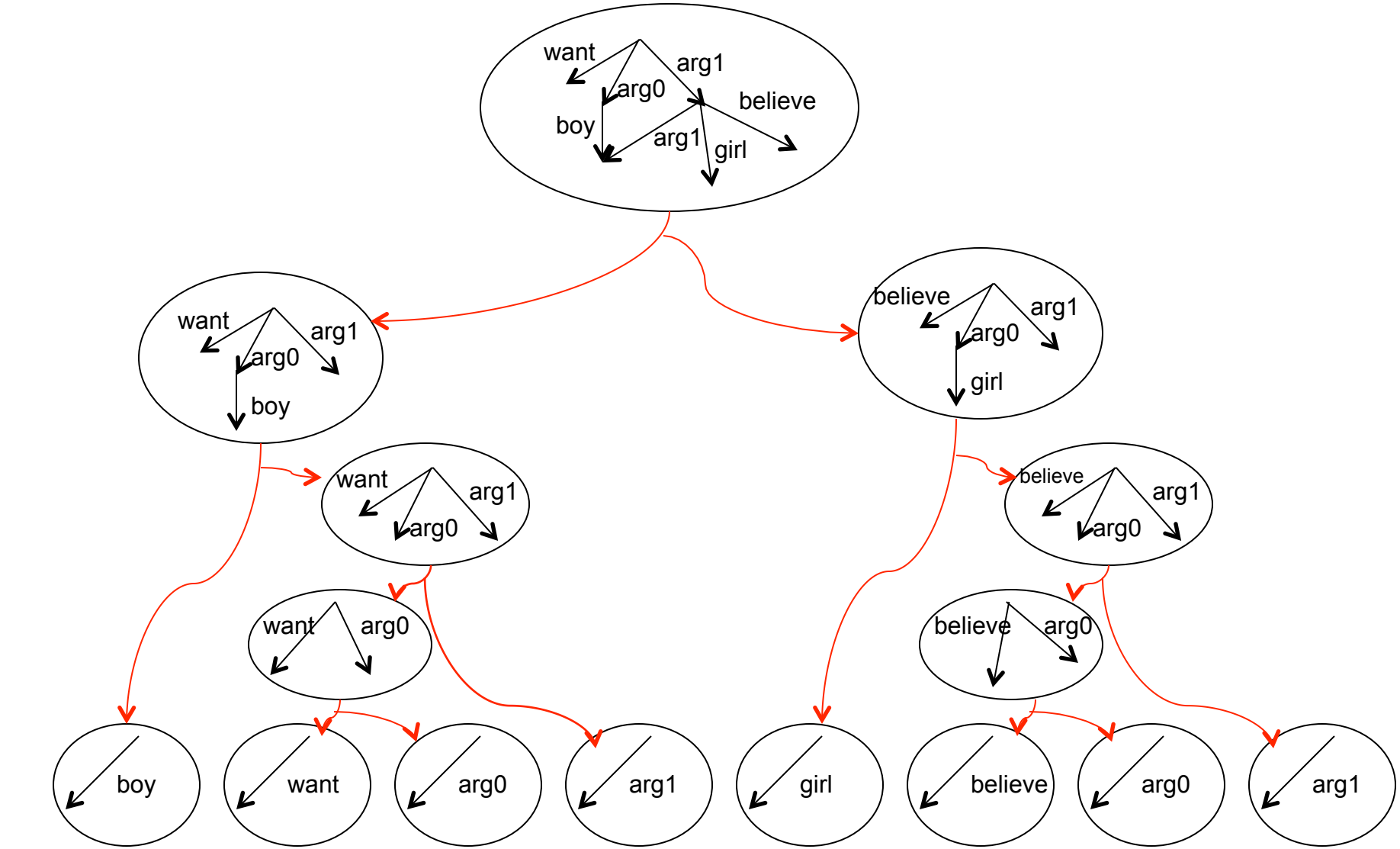
Building fragment forest

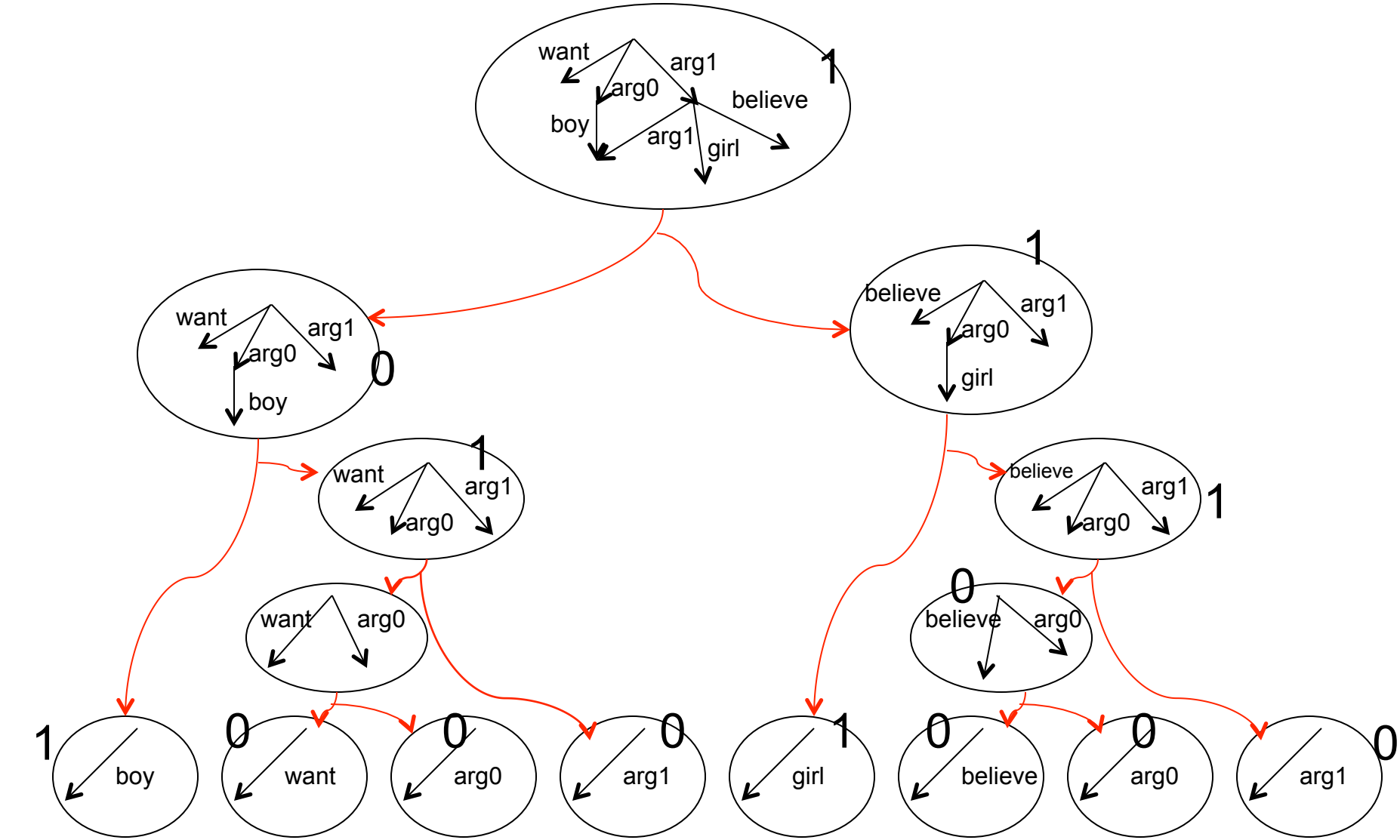


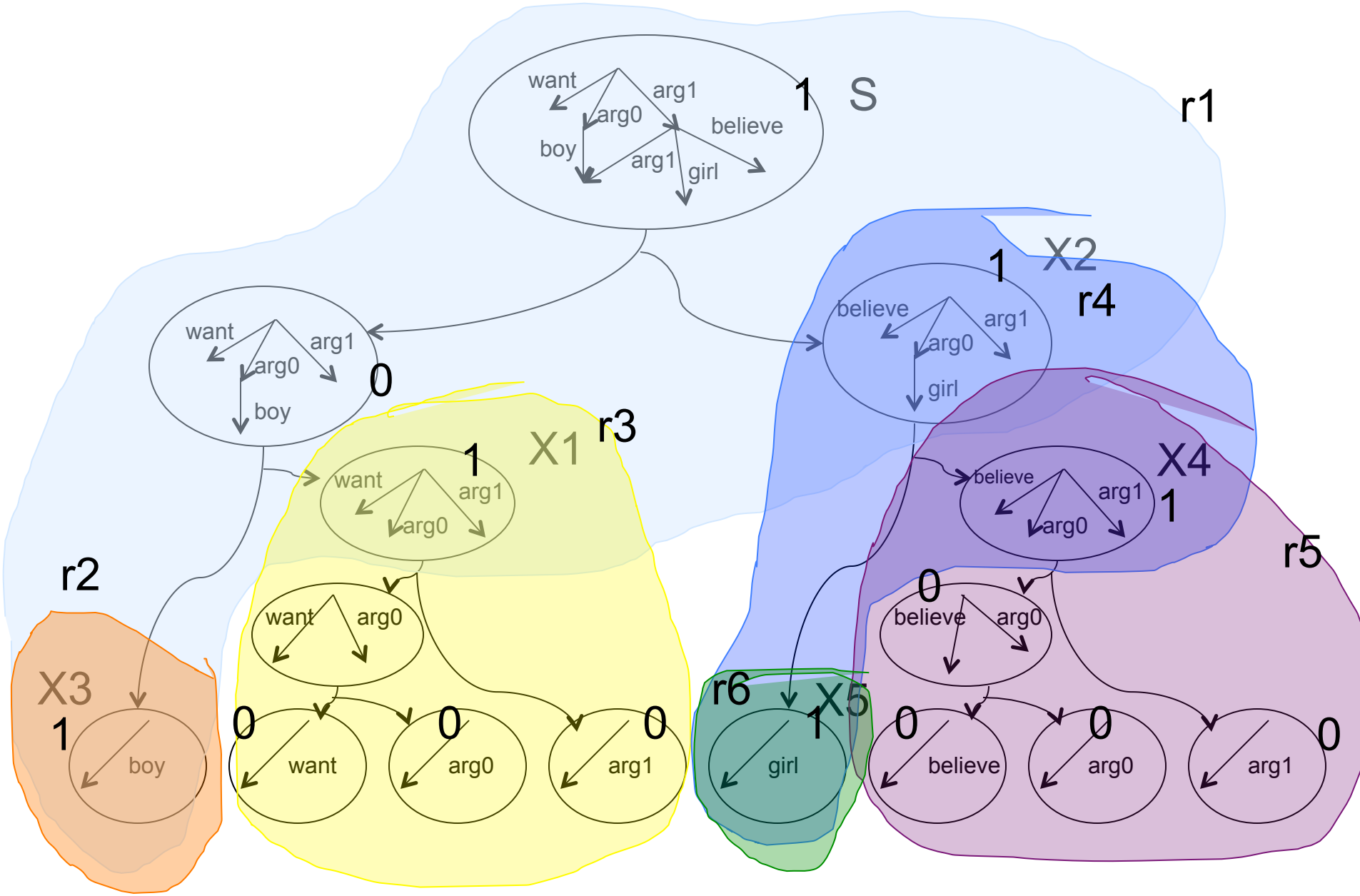
HRG extraction

- Strategy: sampling tree fragments from forest
- Similar to TSG learning (Chung, Gildea 2014)
- Gibbs sampling, top down, sample edge, sample cut


want
arg1
arg0
believe
boy
arg1
girl
want
arg1
arg0
boy
believe
arg1
arg0
girl
want
arg1
arg0
believe
arg0
girl
believe
arg1
arg0
want
arg0
arg0
arg1
believe
arg0
arg0
arg1
boy
want
arg0
arg1
girl
believe
arg0
arg1
The boy wants the girl to believe him


want
arg1
arg0
believe
boy
arg1
girl
want
arg1
arg0
boy
believe
arg1
arg0
girl
want
arg1
arg0
believe
arg1
arg0
want
arg0
believe
arg0
boy
want
arg0
arg1
girl
believe
arg0
arg1
The boy
wants
the girl
to believe
him


want
arg1
arg0
boy
believe
girl
1
0
The boy
wants
the girl
to believe
him


want
arg1
arg0
believe
boy
arg1
girl
1
S
r1
want
arg1
arg0
boy
0
1
X2
r4
believe
arg1
arg0
girl
r3
X1
1
want
arg1
arg0
X4
1
believe
arg1
arg0
r2
r5
want
arg0
0
believe
arg0
X3
1
boy
0
want
0
arg0
0
arg1
r6
X5
1
girl
0
believe
0
arg0
0
arg1

Synchronous HRG learning

- Just keep the strings!
- When we keep building larger fragments, we also build larger strings
- Same top-down sampling schedule

Results

- Forest construction: about 3700 English AMRs from AMR bank, almost all of them have at least one binarized fragment derivation
- Markov Chain Monte Carlo (MCMC) sampling