YOUR NAME: REGISTRATION #

# (F) Grammar Rules! (1/3) [10 points]

One way for computers to understand language is by parsing sentences to figure out the role of each word. A context free grammar (CFG) (also called phrase structure grammar) is a set of rules for forming sentences. Only sentences that can be generated using such a set of rules are then deemed grammatically correct and 'well-formed'. Computer scientists and linguists use CFGs to define and parse languages, where a "language" is defined as any and all sentences that a given CFG can generate. S is the starting symbol for each sentence.

The following rules make up a simple CFG:

S → N V  N → children  N → squirrels  V → sing  V → eat

Each rule says that the element to the left of the arrow can be expanded into the elements to the right of the arrow. By repeatedly replacing symbols, this CFG can expand the symbol S into "squirrels sing", "children sing", "squirrels eat", and "children eat". It cannot, however, generate "children eat squirrels" or "squirrels eat children" or just "children" – you can see that there is no possible sequence of replacements that turns S into any of these.

The following is another simple CFG. The rules have been numbered for your convenience, but the numbers are not part of the rules.

| | | | |
|---|---|---|---|
| 1. S → NP VP | 2. VP → VP PP | 3. PP → P | 4. IV → runs |
| 5. NP → N | 6. VP → VP CONJ VP | 7. PP → P NP | 8. C → that |
| 9. NP → D N | 10. N → squirrel | 11. TV → chases | 12. P → in |
| 13. NP → NP CONJ NP | 14. N → he | 15. TV → eats | 16. P → away |
| 17. VP → IV | 18. N → John | 19. TV → catches | 20. CONJ → and |
| 21. VP → IV PP | 22. N → Mary | 23. TV → tells | 24. D → the |
| 25. VP → TV NP | 26. N → dog | 27. TV → sees | |
| 28. VP → TV C S | 29. N → tree | 30. IV → sits | |

YOUR NAME: REGISTRATION #

# (F) Grammar Rules! (2/3)

**F1**. Here is a simple story. Several of the following sentences are, according to the above CFG, not well formed, meaning they cannot be derived from S by repeated substitution of symbols. List the sentences that the CFG above *can* generate in the box below; ignore the periods.

A. John sees the dog and Mary sees the dog.

B. The dog sees John and Mary.

C. The dog sees a squirrel.

D. The squirrel sits in the tree.

E. That squirrel sees the dog.

F. The squirrel was seen by the dog.

G. The dog runs.

H. The squirrel in the tree runs.

I. The dog chases the squirrel and eats the squirrel.

J. The dog eats.

K. John sees that the dog eats the squirrel.

L. John tells Mary that the dog eats the squirrel.

M. The dog sees that John sees that he eats the squirrel.

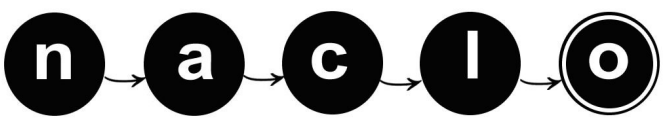
N. And the dog runs away.

O. Mary and John chase the dog.

P. John chases and catches the dog.

Q. John eats dog.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | | |



YOUR NAME: REGISTRATION #

# (F) Grammar Rules! (3/3)

**F2**. Not all of the sentences that this CFG can generate are actually sentences of English. For example, "The dog and the squirrel sits" can be generated but this isn't a correct sentence of English.

Give three more examples of sentences that can be generated by this CFG but are not correct English sentences; enter one word per box. Skip the periods at the end of the sentences.

| | | | | | | |
|---|---|---|---|---|---|---|
| A. | | | | | | |
| | | | | | | |
| | | | | | | |
| B. | | | | | | |
| | | | | | | |
| | | | | | | |
| C. | | | | | | |
| | | | | | | |
| | | | | | | |

F3. One of the rules in the CFG above is redundant: any sentence that can be generated using this rule can already be generated by a combination of other rules. Write the number of the redundant rule below.