NLP assignment #2

Andrius Buinovskij - 18-940-270

Q1 a)

$$S \rightarrow NP, VP \tag{1}$$
$$NP \rightarrow Det, N \mid NP, PP \mid "I" \mid "glasses" \tag{2}$$
$$VP \rightarrow VP, PP \mid V, NP \tag{3}$$
$$Det \rightarrow "a" \mid "an" \tag{4}$$
$$N \rightarrow "man" \mid "pencil" \mid "ball" \mid "umbrella" \tag{5}$$
$$PP \rightarrow P, NP \tag{6}$$
$$V \rightarrow "draw" \mid "hit" \tag{7}$$
$$P \rightarrow "with" \tag{8}$$

Q1 b)
Warning: advanced calculations ahead (instead of probabilities We just do counts first):

$$S \rightarrow NP, VP(4) \tag{9}$$
$$NP(14) \rightarrow Det, N(7) \mid NP, PP(2) \mid "I"(4) \mid "glasses"(1) \tag{10}$$
$$VP(6) \rightarrow VP, PP(2) \mid V, NP(4) \tag{11}$$
$$Det(7) \rightarrow "a"(6) \mid "an"(1) \tag{12}$$
$$N(7) \rightarrow "man"(4) \mid "pencil"(1) \mid "ball"(1) \mid "umbrella"(1) \tag{13}$$
$$PP(4) \rightarrow P, NP(4) \tag{14}$$
$$V(4) \rightarrow "draw"(2) \mid "hit"(2) \tag{15}$$
$$P(4) \rightarrow "with"(4) \tag{16}$$

Convert to probabilities (rounded, for precise figures just divide counts on the right by total counts on the left):

$$S \rightarrow NP, VP(1.0) \tag{17}$$

$$NP \rightarrow Det, N(0.5) \mid NP, PP(0.14) \mid "I"(0.29) \mid "glasses"(0.7) \tag{18}$$

$$VP(\rightarrow VP, PP(2 = 0.33) \mid V, NP(0.66) \tag{19}$$

$$Det \rightarrow "a"(0.86(\mid "an"(0.14) \tag{20}$$

$$N \rightarrow "man"(0.57) \mid "pencil"(0.14) \mid "ball"(0.14) \mid "umbrella"(0.14) \tag{21}$$

$$PP \rightarrow P, NP(1) \tag{22}$$

$$V \rightarrow "draw"(0.5) \mid "hit"(0.5) \tag{23}$$

$$P \rightarrow "with"(1) \tag{24}$$

Q1 c)

Our goal is to have different likelihood of expansions on noun phrase depending on whether the expansion is going to be an object or a subject.

I suppose a logical way of going about it would be to replace the noun phrase non-terminal with two other non terminals - object phrase (OP) and subject phrase (SP).

$$S \rightarrow NP, VP(1.0) \tag{25}$$

$$NP \rightarrow SP \mid OP \tag{26}$$

$$SP \rightarrow Det, N(0.5) \mid NP, PP(0.14) \mid "I"(0.29) \mid "glasses"(0.7) \tag{27}$$

$$OP \rightarrow Det, N(0.5) \mid NP, PP(0.14) \mid "I"(0.29) \mid "glasses"(0.7) \tag{28}$$

$$VP(\rightarrow VP, PP(2 = 0.33) \mid V, NP(0.66) \tag{29}$$

$$Det \rightarrow "a"(0.86(\mid "an"(0.14) \tag{30}$$

$$N \rightarrow "man"(0.57) \mid "pencil"(0.14) \mid "ball"(0.14) \mid "umbrella"(0.14) \tag{31}$$

$$PP \rightarrow P, NP(1) \tag{32}$$

$$V \rightarrow "draw"(0.5) \mid "hit"(0.5) \tag{33}$$

$$P \rightarrow "with"(1) \tag{34}$$

Here are the counts as before:

$$S \rightarrow NP, VP (1.0) \tag{35}$$
$$NP(14) \rightarrow SP(4) \mid OP(10) \tag{36}$$
$$SP(4) \rightarrow Det, N(0) \mid NP, PP(0) \mid "I"(4) \mid "glasses"(0) \tag{37}$$
$$OP(10) \rightarrow Det, N(7) \mid NP, PP(2) \mid "I"(0) \mid "glasses"(1) \tag{38}$$
$$VP(\rightarrow VP, PP(2 = 0.33) \mid V, NP(0.66) \tag{39}$$
$$Det \rightarrow "a"(0.86(\mid "an"(0.14) \tag{40}$$
$$N \rightarrow "man"(0.57) \mid "pencil"(0.14) \mid "ball"(0.14) \mid "umbrella"(0.14) \tag{41}$$
$$PP \rightarrow P, NP(1) \tag{42}$$
$$V \rightarrow "draw"(0.5) \mid "hit"(0.5) \tag{43}$$
$$P \rightarrow "with"(1) \tag{44}$$

Finally converting to probabilities:

$$S \rightarrow NP, VP (1.0) \tag{45}$$
$$NP \rightarrow SP(0.29) \mid OP(0.71) \tag{46}$$
$$SP(4) \rightarrow Det, N(0) \mid NP, PP(0) \mid "I"(1) \mid "glasses"(0) \tag{47}$$
$$OP(10) \rightarrow Det, N(0.7) \mid NP, PP(0.2) \mid "I"(0) \mid "glasses"(0.1) \tag{48}$$
$$VP(\rightarrow VP, PP(2 = 0.33) \mid V, NP(0.66) \tag{49}$$
$$Det \rightarrow "a"(0.86(\mid "an"(0.14) \tag{50}$$
$$N \rightarrow "man"(0.57) \mid "pencil"(0.14) \mid "ball"(0.14) \mid "umbrella"(0.14) \tag{51}$$
$$PP \rightarrow P, NP(1) \tag{52}$$
$$V \rightarrow "draw"(0.5) \mid "hit"(0.5) \tag{53}$$
$$P \rightarrow "with"(1) \tag{54}$$

Well this at least shows that subject phrases are likely to be "I", and object phrases are unlikely to be "I", so it's capturing some of what We wanted to.

Q1 d)

$$S \to NP, VP^+ \tag{55}$$
$$NP \to Det, N^+ \mid NP^+, PP \mid "I" \mid "glasses" \tag{56}$$
$$VP \to VP^+, PP \mid V^+, NP \tag{57}$$
$$Det \to "a" \mid "an" \tag{58}$$
$$N \to "man" \mid "pencil" \mid "ball" \mid "umbrella" \tag{59}$$
$$PP \to P, NP^+ \tag{60}$$
$$V \to "draw" \mid "hit" \tag{61}$$
$$P \to "with" \tag{62}$$

Q2 a)

So, it is my understanding that for the projective dependency tree, to obtain the score We would simply take all dependency pairs and multiply the scores corresponding to those dependencies (taking care to pay attention to the direction of the dependency etc.)

To obtain the score of a lexicalized probabilistic context free grammar tree, one also simply multiplies the score for each "rule" applied, except this rule is also lexicalized so You have very many parameters indeed (page 15 from _ click me_.)

With all that out of the way, I would construct a weighted lexicalized CFG as follows:

1. Add root productions $R \to X_j$, $j \in 1 \ldots M$.

2. Add self-productions $X_j \to j$.

3. For each production $\phi(i \to j)$, add bidirectional productions to the language as follows:

$$X_i \to X_i^+, X_j \tag{63}$$
$$X_i \to X_j, X_i^+ \tag{64}$$

And of course they weight for that transformation corresponds to the value of $\phi(i \to j)$.

And that's all She wrote.

Q1 b)

So the dependency tree would simply be $ROOT \to fish$, $fish \to they$, and the score would be $\psi(ROOT, fish) \cdot \psi(fish, they)$.

In our CFG We would end up with $ROOT \to S$, then $S \to NP$,