IHLT Exam – 2022

DO NOT USE THIS PAPER TO PROVIDE YOUR ANSWERS
(answers without justification or with a wrong one will be considered wrong answers).

1. (4 points) Given the following PCFG:

[1] S → NP VP	(1.0)	[8] NN → time	(0.4)
[2] NP → NN NN	(0.25)	[9] NN → flies	(0.2)
[3] NP → DT NN	(0.4)	[10] NN → arrow	(0.4)
[4] NP → NN	(0.35)	[11] DT → an	(1.0)
[5] VP → V NP	(0.6)	[12] ADV → like	(1.0)
[6] VP → V ADVP	(0.4)	[13] V→ flies	(0.5)
[7] ADVP → ADV NP	(1.0)	[14] V → like	(0.5)

and the following input sentence:

"Time flies like an arrow"

a) Write two parse trees that the grammar generates for this sentence.
b) Compute the probability of each tree. Justify the results providing your calculations.
c) Convert the grammar into CNF and provide the resulting grammar. Justify briefly your decisions.
d) Apply probabilistic CKY algorithm and provide the complete, resulting dynamic table. For each component of the table, provide all the information involved. Write the resulting best parse tree and its probability.

2. (3 points) We are performing PoS tagging with a trigram-based CRF model, using tagset T={DT, V, NN, ADV, PREP}, and we defined a history as $h=<t_{i-2}, t_{i-1}, t_i, w_{[1:n]}, i>$, where t_i is the current state, t_{i-2} and t_{i-1} are their two previous states, and $w_{[1:n]}$ is the whole sequence of observations.

a) How many possible histories are there for a given input sequence $w_{[1:n]}$ and a fixed value of i ? Justify briefly your answer.

b) Which of the following are correct features? Justify briefly your answers.

$f_1(h)$: 1 if t_i = V and t_{i-1} = PREP ; 0 otherwise

$f_2(h)$: 1 if t_i = V and w_{i-2} = dog ; 0 otherwise

$f_3(h)$: 1 if t_i = V and t_{i-3} = NN ; 0 otherwise

$f_4(h)$: 1 if t_i = V and t_{i-1} = PREP and w_2 = dog ; 0 otherwise

c) Given history $h = <t_{i-2}, t_{i-1}, t_i, w_{[1:n]}, i> = <$V, DT, NN, the man saw the dog in the park, 5>, which of the following are correct features and yield equal to 1? Justify briefly your answers.

$f_1(h)$: 1 if t_i = NN and w_i = dog ; 0 otherwise

$f_2(h)$: 1 if t_{i-1} = DT and w_i = dog ; 0 otherwise

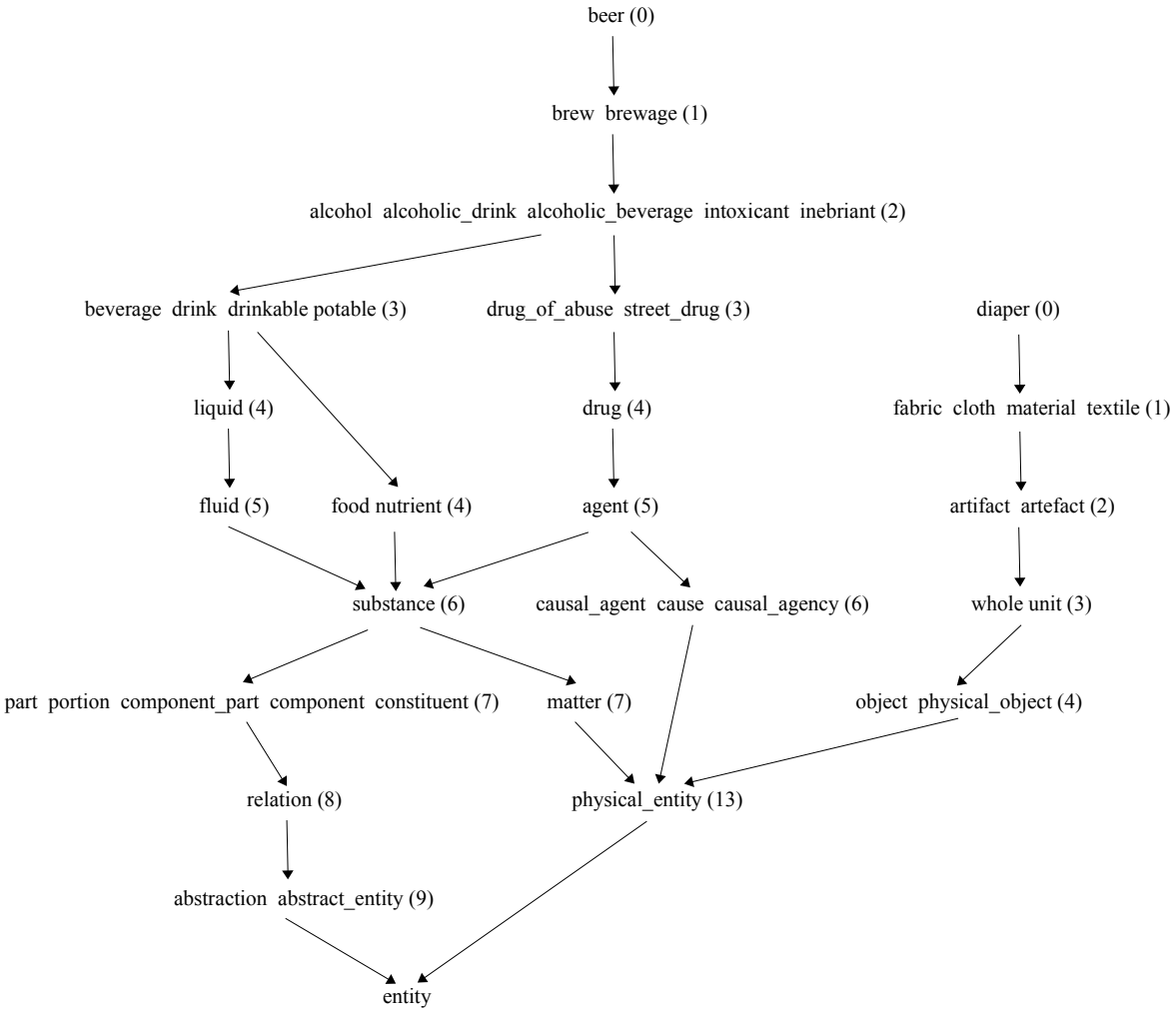
$f_3(h)$: 1 if t_i = DT and w_{i+1} = dog ; 0 otherwise

$f_4(h)$: 1 if t_i = NN and t_{i-1} = DT ; 0 otherwise

3. (3 points) Annotate as *True/False* the following sentences. Justify briefly all your decisions.

a) We want to learn automatically a mention-pair model for identity noun-phrase coreference resolution...
   1. … and, therefor, we can use Support Vector Machines.
   2. … because it is the most informed type of model for identity noun-phrase coreference resolution.
   3. … and, therefor, we can use Hidden Markov Models.

b) Given the following subgraph of WordNet, …



1. *beer (0)* is so similar to *liquid (4)* than to *drug (4)* assuming Shortest Path Length similarity.
2. *beer (0)* is so similar to *liquid (4)* than to *drug (4)* assuming Wu & Palmer similarity.
3. *drug (4)* is less similar to *beer (0)* than to *beverage (3)* assuming Wu & Palmer similarity.