Computational Linguistics Assignment #7

Stephanie Lund (2555914)
Laura Frädrich (2556722)
Francesco Tombini (2554549)

Exercise 1.

P(Start DT DT DT End | man the boat) =
P(DT | Start) * P(man | DT) * P(DT | DT) * P(the | DT) * P(DT | DT) * P(boat | DT) * P(End | DT)
= 0.4 * 0.05 * 0.05 * 0.4 * 0.05 * 0.05 * 0.1 = $1.0 * 10^{-7}$
P(Start DT DT NN End | man the boat) =
P(DT | Start) * P(man | DT) * P(DT | DT) * P(the | DT) * P(NN | DT) * P(boat | NN) * P(End | NN)
= 0.4 * 0.05 * 0.05 * 0.4 * 0.7 * 0.8 * 0.2 = $4.48 * 10^{-5}$
P(Start DT DT VB End | man the boat) =
P(DT | Start) * P(man | DT) * P(DT | DT) * P(the | DT) * P(VB | DT) * P(boat | VB) * P(End | VB)
= 0.4 * 0.05 * 0.05 * 0.4 * 0.1 * 0.05 * 0.2 = $4.0 * 10^{-7}$

P(Start DT NN DT End | man the boat) =
P(DT | Start) * P(man | DT) * P(NN | DT) * P(the | NN) * P(DT | NN) * P(boat | DT) * P(End | DT)
= 0.4 * 0.05 * 0.7 * 0.05 * 0.2 * 0.05 * 0.1 = $7.0 * 10^{-7}$
P(Start DT NN NN End | man the boat) =
P(DT | Start) * P(man | DT) * P(NN | DT) * P(the | NN) * P(NN | NN) * P(boat | NN) * P(End | NN)
= 0.4 * 0.05 * 0.7 * 0.05 * 0.15 * 0.8 * 0.2 = $1.68 * 10^{-5}$
P(Start DT NN VB End | man the boat) =
P(DT | Start) * P(man | DT) * P(NN | DT) * P(the | NN) * P(VB | NN) * P(boat | VB) * P(End | VB)
= 0.4 * 0.05 * 0.7 * 0.05 * 0.3 * 0.05 * 0.2 = $2.1 * 10^{-6}$

P(Start DT VB DT End | man the boat) =
P(DT | Start) * P(man | DT) * P(VB | DT) * P(the | VB) * P(DT | VB) * P(boat | DT) * P(End | DT)
= 0.4 * 0.05 * 0.1 * 0.05 * 0.3 * 0.05 * 0.1 = $1.5 * 10^{-7}$
P(Start DT VB NN End | man the boat) =
P(DT | Start) * P(man | DT) * P(VB | DT) * P(the | VB) * P(NN | VB) * P(boat | NN) * P(End | NN)
= 0.4 * 0.05 * 0.1 * 0.05 * 0.3 * 0.8 * 0.2 = $4.8 * 10^{-6}$
P(Start DT VB VB End | man the boat) =
P(DT | Start) * P(man | DT) * P(VB | DT) * P(the | VB) * P(VB | VB) * P(boat | VB) * P(End | VB)
= 0.4 * 0.05 * 0.1 * 0.05 * 0.1 * 0.05 * 0.2 = $1.0 * 10^{-7}$

P(Start NN DT DT End | man the boat) =
P(NN | Start) * P(man | NN) * P(DT | NN) * P(the | DT) * P(DT | DT) * P(boat | DT) * P(End | DT)
= 0.1 * 0.7 * 0.2 * 0.4 * 0.05 * 0.05 * 0.1 = $1.4 * 10^{-6}$
P(Start NN DT NN End | man the boat) =
P(NN | Start) * P(man | NN) * P(DT | NN) * P(the | DT) * P(NN | DT) * P(boat | NN) * P(End | NN)
= 0.1 * 0.7 * 0.2 * 0.4 * 0.7 * 0.8 * 0.2 = $6.272 * 10^{-4}$
P(Start NN DT VB End | man the boat) =
P(NN | Start) * P(man | NN) * P(DT | NN) * P(the | DT) * P(VB | DT) * P(boat | VB) * P(End | VB)
= 0.1 * 0.7 * 0.2 * 0.4 * 0.1 * 0.05 * 0.2 = $5.6 * 10^{-6}$

P(Start NN NN DT End | man the boat) =

P(NN | Start) * P(man | NN) * P(NN | NN) * P(the | NN) * P(DT | NN) * P(boat | DT) * P(End | DT)
= 0.1 * 0.7 * 0.15 * 0.05 * 0.2 * 0.05 * 0.1 = $5.25 * 10^{-7}$
P(Start NN NN NN End | man the boat) =
P(NN | Start) * P(man | NN) * P(NN | NN) * P(the | NN) * P(NN | NN) * P(boat | NN) * P(End | NN)
= 0.1 * 0.7 * 0.15 * 0.05 * 0.15 * 0.8 * 0.2 = $1.26 * 10^{-5}$
P(Start NN NN VB End | man the boat) =
P(NN | Start) * P(man | NN) * P(NN | NN) * P(the | NN) * P(VB | NN) * P(boat | VB) * P(End | VB)
= 0.1 * 0.7 * 0.15 * 0.05 * 0.3 * 0.05 * 0.2 = $1.575 * 10^{-6}$

P(Start NN VB DT End | man the boat) =
P(NN | Start) * P(man | NN) * P(VB | NN) * P(the | VB) * P(DT | VB) * P(boat | DT) * P(End | DT)
= 0.1 * 0.7 * 0.3 * 0.05 * 0.3 * 0.05 * 0.1 = $1.575 * 10^{-6}$
P(Start NN VB NN End | man the boat) =
P(NN | Start) * P(man | NN) * P(VB | NN) * P(the | VB) * P(NN | VB) * P(boat | NN) * P(End | NN)
= 0.1 * 0.7 * 0.3 * 0.05 * 0.2 * 0.8 * 0.2 = $3.36 * 10^{-5}$
P(Start NN VB VB End | man the boat) =
P(NN | Start) * P(man | NN) * P(VB | NN) * P(the | VB) * P(VB | VB) * P(boat | VB) * P(End | VB)
= 0.1 * 0.7 * 0.3 * 0.05 * 0.1 * 0.05 * 0.2 = $1.05 * 10^{-6}$

P(Start VB DT DT End | man the boat) =
P(VB | Start) * P(man | VB) * P(DT | VB) * P(the | DT) * P(DT | DT) * P(boat | DT) * P(End | DT)
= 0.2 * 0.1 * 0.3 * 0.4 * 0.05 * 0.05 * 0.1 = $6.0 * 10^{-7}$
P(Start VB DT NN End | man the boat) =
P(VB | Start) * P(man | VB) * P(DT | VB) * P(the | DT) * P(NN | DT) * P(boat | NN) * P(End | NN)
= 0.2 * 0.1 * 0.3 * 0.4 * 0.7 * 0.8 * 0.2 = $2.688 * 10^{-4}$
P(Start VB DT VB End | man the boat) =
P(VB | Start) * P(man | VB) * P(DT | VB) * P(the | DT) * P(VB | DT) * P(boat | VB) * P(End | VB)
= 0.2 * 0.1 * 0.3 * 0.4 * 0.1 * 0.05 * 0.2 = $2.4 * 10^{-6}$

P(Start VB NN DT End | man the boat) =
P(VB | Start) * P(man | VB) * P(NN | VB) * P(the | NN) * P(DT | NN) * P(boat | DT) * P(End | DT)
= 0.2 * 0.1 * 0.2 * 0.05 * 0.2 * 0.05 * 0.1 = $2.0 * 10^{-7}$
P(Start VB NN NN End | man the boat) =
P(VB | Start) * P(man | VB) * P(NN | VB) * P(the | NN) * P(NN | NN) * P(boat | NN) * P(End | NN)
= 0.2 * 0.1 * 0.2 * 0.05 * 0.15 * 0.8 * 0.2 = $4.8 * 10^{-6}$
P(Start VB NN VB End | man the boat) =
P(VB | Start) * P(man | VB) * P(NN | VB) * P(the | NN) * P(VB | NN) * P(boat | VB) * P(End | VB)
= 0.2 * 0.1 * 0.2 * 0.05 * 0.3 * 0.05 * 0.2 = $6.0 * 10^{-7}$

P(Start VB VB DT End | man the boat) =
P(VB | Start) * P(man | VB) * P(VB | VB) * P(the | VB) * P(DT | VB) * P(boat | DT) * P(End | DT)
= 0.2 * 0.1 * 0.1 * 0.05 * 0.3 * 0.05 * 0.1 = $1.5 * 10^{-7}$
P(Start VB VB NN End | man the boat) =
P(VB | Start) * P(man | VB) * P(VB | VB) * P(the | VB) * P(NN | VB) * P(boat | NN) * P(End | NN)
= 0.2 * 0.1 * 0.1 * 0.05 * 0.2 * 0.8 * 0.2 = $3.2 * 10^{-6}$
P(Start VB VB VB End | man the boat) =
P(VB | Start) * P(man | VB) * P(VB | VB) * P(the | VB) * P(VB | VB) * P(boat | VB) * P(End | VB)
= 0.2 * 0.1 * 0.1 * 0.05 * 0.1 * 0.05 * 0.2 = $1.0 * 10^{-7}$

The two paths P(Start NN DT NN End | man the boat) and P(Start VB DT NN End | man the boat) produce the highest probabilities, namely $6.272 * 10^{-4}$ and $2.688 * 10^{-4}$.

Exercise 3.

a. 0.55568*
b. 0.529
c. 0.96521

*This result isn't deterministic, since the tag it chooses to assign to words with zero probabilities depends on the ordering of the Python dict.

Exercise 5.

a. 0.953038
b. 1.0
c. 0.953038