Chương 6

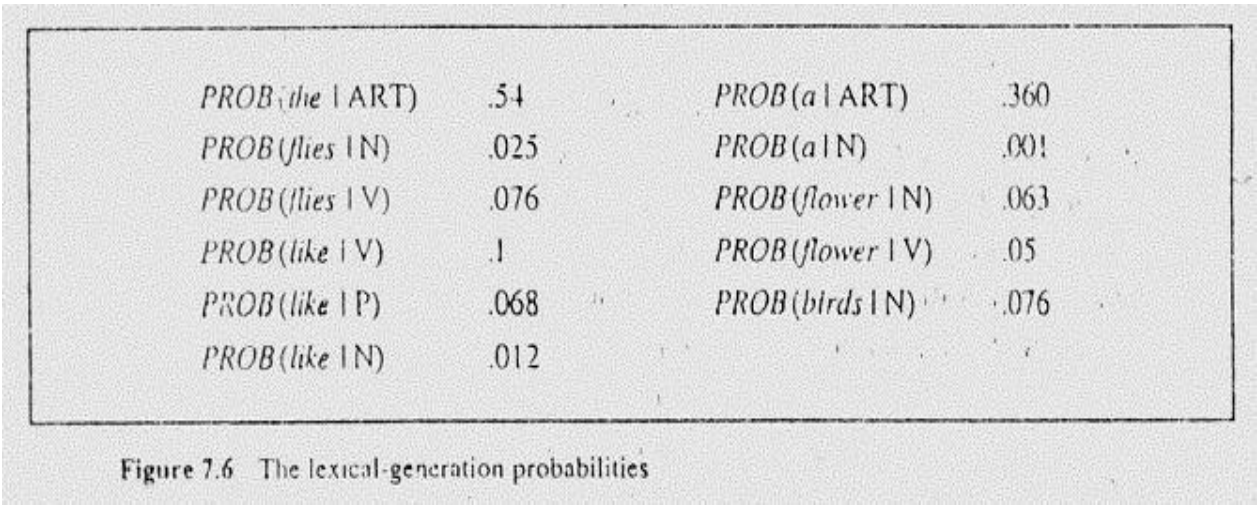
1. Hand simulate the Viterbi algorithm using the data and probability estimates in Figures 7.4- 7.6 on the sentence *Flower flowers like flowers*. Draw transition network as in Figure 7.10-7.12 for the problem, and identify what part of speech the algorithm identifies for each word.

Giải

Flower flowers like flowers

L = V, N, ART, P

category	Count at i	pair	Count at i,i+1	bigram	estimate
O	300	O, ART	213	PROB(ART/O)	0.71
O	300	O, N	87	PORB(N/O)	0.29
ART	558	ART, N	558	PROB(N/ART)	1.0
N	833	N,V	258	PROB(V/N)	0.43
N	833	N, N	108	PROB(N/N)	0.13
N	833	N, P	366	PROB(P/N)	0.44
V	300	V, N	75	PROB(N/V)	0.35
V	300	V, ART	194	PROB(ART/V)	0.65
P	307	P, ART	226	PROB(ART/P)	0.74
P	307	P, N	81	PROB(N/P)	0.26



PROB(the | ART)	.54	*PROB(a | ART)*	.360
PROB(flies | N)	.025	*PROB(a | N)*	.001
PROB(flies | V)	.076	*PROB(flower | N)*	.063
PROB(like | V)	.1	*PROB(flower | V)*	.05
PROB(like | P)	.068	*PROB(birds | N)*	.076
PROB(like | N)	.012		

Figure 7.6 The lexical-generation probabilities

i=1 to N	**SEQ(1,1)** = PROB(Flower/V) * PROB (V/O)	= 0.05 * 10^-4 = 5 * 10^-6
	SEQ(2,1) = PROB(Flower/N) * PROB (N/O)	**= 0.063 * 0.29 = 0.01827**
	SEQ(3,1) = PROB(Flower/ART) * PROB (ART/O)	= 0
	SEQ(4,1) = PROB(Flower/P) * PROB (P/O)	= 0
t=2 to 4 i=1 to 4	**SEQ(1,2)** = max $_{j=1,4}$ (SEQ(1,1) * PROB(V/V), SEQ (2,1) * PROB(V/N)) * PROB(flowers/V)	= max(5 * 10^-6 * 10^-4, 0.01827 * 0.43) * 0.05 = **0.000392805**
	SEQ(2,2) = max $_{j=1,4}$ (SEQ(1,1) * PROB(N/V), SEQ (2,1) * PROB(N/N)) * PROB(flowers/N)	= max(5 * 10^-6 * 0.35, 0.01827 * 0.13) * 0.063 = 0.0001496313
	SEQ(3,2) = max $_{j=1,4}$ (SEQ(1,1) * PROB(ART/V), SEQ (2,1) * PROB(ART/N)) * PROB(flowers/ART)	= 0
	SEQ(4,2) = max $_{j=1,4}$ (SEQ(1,1) * PROB(P/V), SEQ (2,1) * PROB(P/N)) * PROB(flowers/P)	= 0
t=3 to 4 i=1 to 4	**SEQ(1,3)** = max $_{j=1,4}$ (SEQ(1,2) * PROB(V/V), SEQ (2,2) * PROB(V/N)) * PROB(like/V)	= max(0.000392805 * 10^-4, 0.0001496313 * 0.43) * 0.1 **= 0.0000064341459 = 6.43 * 10^-6**
	SEQ(2,3) = max $_{j=1,4}$ (SEQ(1,2) * PROB(N/V), SEQ (2,2) * PROB(N/N)) * PROB(like/N)	= max(0.000392805 * 0.35, 0.0001496313 * 0.13) * 0.012 = 0.000001649781 = 1.65 * 10^-6
	SEQ(3,3) = max $_{j=1,4}$ (SEQ(1,2) * PROB(ART/V), SEQ (2,2) * PROB(ART/N)) * PROB(like/ART)	= 0
	SEQ(4,3) = max $_{j=1,4}$ (SEQ(1,2) * PROB(P/V), SEQ (2,2) * PROB(P/N)) * PROB(like/P)	= max(0.000392805 * 10^-4, 0.0001496313 * 0.44) * 0.068 = 0.000004476968496 = 4.48 * 10^-6
t=4 i=1 to 4	**SEQ(1,4)** = max $_{j=1,4}$ (SEQ(1,3) * PROB(V/V), SEQ(2,3) * PROB(V/N), SEQ (4,3) * PROB(V/P)) * PROB(flowers/V)	= max(6.43 * 10^-6 * 10^-4, 1.65 * 10^-6 * 0.43, 4.48 * 10^-6 * 10^-4) * 0.05 = 3.5475 * 10^-8
	SEQ(2,4) = max $_{j=1,4}$ (SEQ(1,3) * PROB(N/V), SEQ(2,3) * PROB(N/N), SEQ (4,3) * PROB(N/P)) * PROB(flowers/N)	= max(6.43 * 10^-6 * 0.35, 1.65 * 10^-6 * 0.13, 4.48 * 10^-6 * 0.26) * 0.063 **= 1.417815 * 10^-7**
	SEQ(3,4) = max $_{j=1,4}$ (SEQ(1,3) * PROB(ART/V), SEQ(2,3) * PROB(ART/N),	= 0

	SEQ (4,3) * PROB(ART/P)) * PROB(flowers/ART) **SEQ(4,4)** = max $_{j=1,4}$ (SEQ(1,3) * PROB(P/V), SEQ(2,3) * PROB(P/N), SEQ (4,3) * PROB(P/P)) * PROB(flowers/P)	= 0

Xác định chuỗi từ loại:

C(4) = 2, C(3) = 1, C(2) = 1, C(1) = 2

Vậy chuỗi từ loại là **N V V N**

2. Using the bigram and lexical generation probabilities given in this chapter, calculate the word probabilities using the forward algorithm for the sentence *The a flies like flower* (involving a very rare use of the word a as a noun, as in the a flies, the b flies, and so on). Remember to use 0.0001 as a probability for any bigram not in the table. Are the results you get reasonable? If not, what is the problem and how might it be fixed ?

Giải

The a flies like flower

L = ART, N, V, P

	N	V	ART	P	TOTAL
flies	21	23	0	0	44
fruit	49	5	1	0	55
lik	10	30	0	21	61
a	1	0	201	0	202
the	1	0	300	2	303
flower	53	15	0	0	68
flowers	42	16	0	0	58
birds	64	1	0	0	65
others	592	210	56	284	1142
TOTAL	833	300	558	307	1998

t=1

SEQSUM(1,1) = PROB(The/ART) * PROB(ART/O) = 0.54 * 0.71 = 0.3834

SEQSUM(2,1) = PROB(The/N) * PROB(N/O) = 1/833 * 0.29 = 3.4814 * 10^-4

SEQSUM(3,1) = PROB(The/V) * PROB(V/O) = 0 * 10^-4 = 0

SEQSUM(4,1) = PROB(The/P) * PROB(P/O) = 2/307 * 10^-4 = 6.5147 * 10^-7

t=2

SEQSUM(1,2) =

(PROB(ART/ART) * SEQSUM(1,1) + PROB(ART/N) * SEQSUM(2,1) +

PROB(ART/V) * SEQSUM(3,1) + PROB(ART/P) * SEQSUM(4,1))

* PROB(a/ART)

= (10^-4 * 0.3834 + 10^-4 * 3.4814 * 10^-4 + 0 + 0.74 * 6.5147 * 10^-7) * 0.36

= 1.3988 * 10^-5

SEQSUM(2,2) =

(PROB(N/ART) * SEQSUM(1,1) + PROB(N/N) * SEQSUM(2,1) +

PROB(N/V) * SEQSUM(3,1) + PROB(N/P) * SEQSUM(4,1)) * PROB(a/N) =

= (1 * 0.3834 + 0.13 * 3.4814 * 10^-4 + 0 + 0.26 * 6.5147 * 10^-7) * 0.001

= 3.8346 * 10^-4

SEQSUM(3,2) =

(PROB(V/ART) * SEQSUM(1,1) + PROB(V/N) * SEQSUM(2,1) +

PROB(V/V) * SEQSUM(3,1) + PROB(V/P) * SEQSUM(4,1)) * PROB(a/V) = 0

SEQSUM(4,2) =

(PROB(P/ART) * SEQSUM(1,1) + PROB(P/N) * SEQSUM(2,1) +

PROB(P/V) * SEQSUM(3,1) + PROB(P/P) * SEQSUM(4,1)) * PROB(a/P) = 0

t=3

SEQSUM(1,3) =

(PROB(ART/ART) * SEQSUM(1,2) + PROB(ART/N) * SEQSUM(2,2) +

PROB(ART/V) * SEQSUM(3,2) + PROB(ART/P) * SEQSUM(4,2))

* PROB(flies/ART) = 0

SEQSUM(2,3) =

(PROB(N/ART) * SEQSUM(1,2) + PROB(N/N) * SEQSUM(2,2) +

PROB(N/V) * SEQSUM(3,2) + PROB(N/P) * SEQSUM(4,2)) * PROB(flies/N)

= (1 * 1.3988 * 10^-5 + 0.13 * 3.8346 * 10^-4 + 0 + 0) * 0.025 = 1.5959 * 10^-6

SEQSUM(3,3) =

(PROB(V/ART) * SEQSUM(1,2) + PROB(V/N) * SEQSUM(2,2) +

PROB(V/V) * SEQSUM(3,2) + PROB(V/P) * SEQSUM(4,2)) * PROB(flies/V)

= (10^-4 * 1.3988 * 10^-5 + 0.43 * 3.8346 * 10^-4 + 0 + 0) * 0.076 = 1.2532 * 10^-5

SEQSUM(4,3) =

(PROB(P/ART) * SEQSUM(1,2) + PROB(P/N) * SEQSUM(2,2) +

PROB(P/V) * SEQSUM(3,2) + PROB(P/P) * SEQSUM(4,2)) * PROB(flies/P) = 0

t=4

SEQSUM(1,4) =

(PROB(ART/ART) * SEQSUM(1,3) + PROB(ART/N) * SEQSUM(2,3) +

PROB(ART/V) * SEQSUM(3,3) + PROB(ART/P) * SEQSUM(4,3))

* PROB(like/ART) = 0

SEQSUM(2,4) =

(PROB(N/ART) * SEQSUM(1,3) + PROB(N/N) * SEQSUM(2,3) +

PROB(N/V) * SEQSUM(3,3) + PROB(N/P) * SEQSUM(4,3)) * PROB(like/N)

= (0 + 0.13 * 1.5959 * 10^-6 + 0.35 * 1.2532 * 10^-5 + 0) * 0.012

= 5.5124 * 10^-8

SEQSUM(3,4) =

(PROB(V/ART) * SEQSUM(1,3) + PROB(V/N) * SEQSUM(2,3) +

PROB(V/V) * SEQSUM(3,3) + PROB(V/P) * SEQSUM(4,3)) * PROB(like/V)

= (0 + 0.43 * 1.5959 * 10^-6 + 10^-4 * 1.2532 * 10^-5 + 0) * 0.1

= 6.8749 * 10^-8

SEQSUM(4,4) =

(PROB(P/ART) * SEQSUM(1,3) + PROB(P/N) * SEQSUM(2,3) +

PROB(P/V) * SEQSUM(3,3) + PROB(P/P) * SEQSUM(4,3)) * PROB(like/P)

= (0 + 0.44 * 1.5959 * 10^-6 + 10^-4 * 1.2532 * 10^-5 + 0) * 0.068

= 4.7835 * 10^-8

t=5

SEQSUM(1,5) =

(PROB(ART/ART) * SEQSUM(1,4) + PROB(ART/N) * SEQSUM(2,4) +

PROB(ART/V) * SEQSUM(3,4) + PROB(ART/P) * SEQSUM(4,4))

* PROB(flower/ART) = 0

SEQSUM(2,5) =

(PROB(N/ART) * SEQSUM(1,4) + PROB(N/N) * SEQSUM(2,4) +

PROB(N/V) * SEQSUM(3,4) + PROB(N/P) * SEQSUM(4,4)) * PROB(flower/N)

= (0 + 0.13 * 5.5124 * 10^-8 + 0.35 * 6.8749 * 10^-8 + 0.26 * 4.7835 * 10^-8) * 0.063

= 2.7509 * 10^-9

SEQSUM(3,5) =

(PROB(V/ART) * SEQSUM(1,4) + PROB(V/N) * SEQSUM(2,4) +

PROB(V/V) * SEQSUM(3,4) + PROB(V/P) * SEQSUM(4,4)) * PROB(flower/V)

= (0 + 0.43 * 5.5124 * 10^-8 + 10^-4 * 6.8749 * 10^-8 + 10^-4 * 4.7835 * 10^-8) * 0.05

= 1.1857 * 10^-9

SEQSUM(4,5) =

(PROB(P/ART) * SEQSUM(1,4) + PROB(P/N) * SEQSUM(2,4) +

PROB(P/V) * SEQSUM(3,4) + PROB(P/P) * SEQSUM(4,4)) * PROB(flower/P) = 0

Tính xác suất từ vựng

1. SEQSUM(1,1) + SEQSUM(2,1) + SEQSUM(3,1) + SEQSUM(4,1) = 0.3837

PROB(the|ART|the) = SEQSUM(1,1) / 0.3837 = 0.9992

PROB(the|N|the) = SEQSUM(2,1) / 0.3837 = 0.00091

PROB(the|V|the) = SEQSUM(3,1) / 0.3837 = 0

PROB(the|P|the) = SEQSUM(4,1) / 0.3837 = 0.0000017

2. SEQSUM(1,2) + SEQSUM(2,2) + SEQSUM(3,2) + SEQSUM(4,2) = 3.9745 * 10^-4

PROB(a|ART|the a) = SEQSUM(1,2) / (3.9745 * 10^-4) = 0.0352

PROB(a|N|the a) = SEQSUM(2,2) / (3.9745 * 10^-4) = 0.9648

PROB(a|V|the a) = SEQSUM(3,2) / (3.9745 * 10^-4) = 0

PROB(a|P|the a) = SEQSUM(4,2) / (3.9745 * 10^-4) = 0

3. SEQSUM(1,3) + SEQSUM(2,3) + SEQSUM(3,3) + SEQSUM(4,3) = 1.4128 * 10^-5

PROB(flies|ART|the a flies) = SEQSUM(1,3) / (1.4128 * 10^-5) = 0

PROB(flies|N|the a flies) = SEQSUM(2,3) / (1.4128 * 10^-5) = 0.1130

PROB(flies|V|the a flies) = SEQSUM(3,3) / (1.4128 * 10^-5) = 0.8870

PROB(flies|P|the a flies) = SEQSUM(4,3) / (1.4128 * 10^-5) = 0

4. SEQSUM(1,4) + SEQSUM(2,4) + SEQSUM(3,4) + SEQSUM(4,4) = 1.7171 * 10^-7

PROB(like|ART|the a flies like) = SEQSUM(1,4) / (1.7171 * 10^-7) = 0

PROB(like|N|the a flies like) = SEQSUM(2,4) / (1.7171 * 10^-7) = 0.3210

PROB(like|V|the a flies like) = SEQSUM(3,4) / (1.7171 * 10^-7) = 0.4004

PROB(like|P|the a flies like) = SEQSUM(4,4) / (1.7171 * 10^-7) = 0.2786

5. SEQSUM(1,5) + SEQSUM(2,5) + SEQSUM(3,5) + SEQSUM(4,5) = 3.9366 * 10^-9

PROB(flowers|ART|the a flies like flowers) = SEQSUM(1,5) / (3.9366 * 10^-9) = 0

PROB(flowers|N|the a flies like flowers) = SEQSUM(2,5) / (3.9366 * 10^-9) = 0.6988

PROB(flowers|V|the a flies like flowers) = SEQSUM(3,5) / (3.9366 * 10^-9) = 0.3012

PROB(flowers|P|the a flies like flowers) = SEQSUM(4,5) / (3.9366 * 10^-9) = 0

Như vậy **the a flies like flowers** là **ART N V V N**.

Kết quả trên không hợp lý, khi từ loại của flies như ta mong muốn là N, như kết quả phân tích lại thu được flies có từ loại là V.

Lý do có thể vì kích thước tập mẫu dữ liệu đầu vào chưa đủ lớn.

Cách giải quyết dễ thấy nhất là tăng kích thước tập mẫu lên.

3. 3. Consider an extended version of Grammar 7.17 with the additional rule:

10. VP → V PP

The revised rule probabilities are shown here (Any not mentioned are the same as in Grammar 7.17):

VP → V 0.32 VP → V NP PP 0.20

VP → V NP 0.33 VP → V PP 0.15

In addition, the following bigram probabilities differ from those in Figure 7.4:

PROB(N/V) = 0.53 PROB(ART/V) = 0.32 PROB(P/V) = 0.15

a) Hand simulate (or implement) the forward algorithm on *Fruit flies like birds* to produce the lexical probabilities.

b) Draw out the full chart for *Fruit flies like birds*, showing the probabilities of each constituent.

Giải

a) L = ART, N, V, P

	N	V	ART	P	TOTAL
flies	21	23	0	0	44
fruit	49	5	1	0	55
lik	10	30	0	21	61
a	1	0	201	0	202
the	1	0	300	2	303
flower	53	15	0	0	68
flowers	42	16	0	0	58
birds	64	1	0	0	65
others	592	210	56	284	1142
TOTAL	833	300	558	307	1998

t=1

SEQSUM(1,1) = PROB(fruit/ART) * PROB(ART/O) = 1/558 * 0.71 = 1.2724 * 10^-3

SEQSUM(2,1) = PROB(fruit/N) * PROB(N/O) = 49/833 * 0.29 = 0.0171

SEQSUM(3,1) = PROB(fruit/V) * PROB(V/O) = 5/300 * 10^-4 = 1.6667 * 10^-6

SEQSUM(4,1) = PROB(fruit/P) * PROB(P/O) = 0

t=2

SEQSUM(1,2) =

(PROB(ART/ART) * SEQSUM(1,1) + PROB(ART/N) * SEQSUM(2,1) +

PROB(ART/V) * SEQSUM(3,1) + PROB(ART/P) * SEQSUM(4,1))

* PROB(flies/ART)

= 0

SEQSUM(2,2) =

(PROB(N/ART) * SEQSUM(1,1) + PROB(N/N) * SEQSUM(2,1) +

PROB(N/V) * SEQSUM(3,1) + PROB(N/P) * SEQSUM(4,1))

* PROB(flies/N)

= (1 * 1.2724 * 10^-3 + 0.13 * 0.0171 + 0.53 * 1.6667 * 10^-6 + 0) * 0.025

= 8.7407 * 10^-5

SEQSUM(3,2) =

(PROB(V/ART) * SEQSUM(1,1) + PROB(V/N) * SEQSUM(2,1) +

PROB(V/V) * SEQSUM(3,1) + PROB(V/P) * SEQSUM(4,1)) * PROB(flies/V)

= (10^-4 * 1.2724 * 10^-3 + 0.43 * 0.0171 + 10^-4 * 1.6667 * 10^-6 + 0) * 0.076

= 5.5884 * 10^-4

SEQSUM(4,2) =

(PROB(P/ART) * SEQSUM(1,1) + PROB(P/N) * SEQSUM(2,1) +

PROB(P/V) * SEQSUM(3,1) + PROB(P/P) * SEQSUM(4,1)) * PROB(flies/P)

= 0

t=3

SEQSUM(1,3) =

(PROB(ART/ART) * SEQSUM(1,2) + PROB(ART/N) * SEQSUM(2,2) +

PROB(ART/V) * SEQSUM(3,2) + PROB(ART/P) * SEQSUM(4,2))

* PROB(like/ART)

= 0

SEQSUM(2,3) =

(PROB(N/ART) * SEQSUM(1,2) + PROB(N/N) * SEQSUM(2,2) +
PROB(N/V) * SEQSUM(3,2) + PROB(N/P) * SEQSUM(4,2))
* PROB(like/N)
= (0 + 0.13 * 8.7407 * 10^-5 + 0.53 * 5.5884 * 10^-4 + 0) * 0.012
= 3.6906 * 10^-6

SEQSUM(3,3) =
(PROB(V/ART) * SEQSUM(1,2) + PROB(V/N) * SEQSUM(2,2) +
PROB(V/V) * SEQSUM(3,2) + PROB(V/P) * SEQSUM(4,2))
* PROB(like/V)
= (0 + 0.43 * 8.7407 * 10^-5 + 10^-4 * 5.5884 * 10^-4 + 0) * 0.1
= 3.7641 * 10^-6

SEQSUM(4,3) =
(PROB(P/ART) * SEQSUM(1,2) + PROB(P/N) * SEQSUM(2,2) +
PROB(P/V) * SEQSUM(3,2) + PROB(P/P) * SEQSUM(4,2))
* PROB(like/P)
= (0 + 0.44 * 8.7407 * 10^-5 + 0.15 * 5.5884 * 10^-4 + 0) * 0.068
= 8.3154 * 10^-6

t=4

SEQSUM(1,4) =
(PROB(ART/ART) * SEQSUM(1,3) + PROB(ART/N) * SEQSUM(2,3) +
PROB(ART/V) * SEQSUM(3,3) + PROB(ART/P) * SEQSUM(4,3))
* PROB(birds/ART)
= 0

SEQSUM(2,4) =
(PROB(N/ART) * SEQSUM(1,3) + PROB(N/N) * SEQSUM(2,3) +
PROB(N/V) * SEQSUM(3,3) + PROB(N/P) * SEQSUM(4,3))
* PROB(birds/N)

= (0 + 0.13 * 3.6906 * 10^-6 + 0.53 * 3.7641 * 10^-6 + 0.26 * 8.3154 * 10^-6) * 0.076

= 3.5239 * 10^-7

SEQSUM(3,4) =

(PROB(V/ART) * SEQSUM(1,3) + PROB(V/N) * SEQSUM(2,3) +

PROB(V/V) * SEQSUM(3,3) + PROB(V/P) * SEQSUM(4,3))

* PROB(birds/V)

= (0 + 0.43 * 3.6906 * 10^-6 + 10^-4 * 3.7641 * 10^-6 + 10^-4 * 8.3154 * 10^-6) * 1/300

= 5.2939 * 10^-9

SEQSUM(4,4) =

(PROB(P/ART) * SEQSUM(1,3) + PROB(P/N) * SEQSUM(2,3) +

PROB(P/V) * SEQSUM(3,3) + PROB(P/P) * SEQSUM(4,3))

* PROB(birds/P)

= 0

Tính xác suất từ vựng

1. SEQSUM(1,1) + SEQSUM(2,1) + SEQSUM(3,1) + SEQSUM(4,1) = 0.0184

PROB(fruit|ART|fruit) = SEQSUM(1,1) / 0.0184 = 0.069152173913

PROB(fruit|N|fruit) = SEQSUM(2,1) / 0.0184 = 0.929347826087

PROB(fruit|V|fruit) = SEQSUM(3,1) / 0.0184 = 0.0000905815217391

PROB(fruit|P|fruit) = SEQSUM(4,1) / 0.0184 = 0

2. SEQSUM(1,2) + SEQSUM(2,2) + SEQSUM(3,2) + SEQSUM(4,2) = 0.000647

PROB(flies|ART|fruit flies) = 0

PROB(flies|N|fruit flies) = SEQSUM(2,2) / 0.000647 = 0.1363

PROB(flies|V|fruit flies) = SEQSUM(3,2) / 0.000647 = 0.8637

PROB(flies|P|fruit flies) = 0

3. SEQSUM(1,3) + SEQSUM(2,3) + SEQSUM(3,3) + SEQSUM(4,3) = 0.00001577

PROB(like|ART|fruit flies like) = 0

PROB(like|N|fruit flies like) = SEQSUM(2,3) / 0.00001577 = 0.2340

PROB(like|V|fruit flies like) = SEQSUM(3,3) / 0.00001577 = 0.2387

PROB(like|P|fruit flies like) = SEQSUM(4,3) / 0.00001577 = 0.5273

4. SEQSUM(1,4) + SEQSUM(2,4) + SEQSUM(3,4) + SEQSUM(4,4) = 3.5768 * 10^-7

PROB(birds|ART|fruit flies like birds) = 0

PROB(birds|N|fruit flies like birds) = SEQSUM(2,4) / (3.5768 * 10^-7) = 0.9852

PROB(birds|V|fruit flies like birds) = SEQSUM(3,4) / (3.5768 * 10^-7) = 0.0148

PROB(birds|P|fruit flies like birds) = 0

Như vậy **fruit flies like birds** là **N V P N**.

b)

<table>
<tr><td colspan="2">S4
1 NP6, 2 VP9, 1 * 0.0383 * 0.33 * 30/61 * 0.1378 = 8.56 * 10^-4</td></tr>
<tr><td colspan="2">S3
1 NP5, 2 VP9, 1 * 4.77 * 10^-3 * 0.33 * 30/61 * 0.1378 = 1.07 * 10^-4</td></tr>
<tr><td colspan="2">S2
1 NP1, 2 VP11, 1 * 0.14 * 49/55 * 0.15 * 23/44 * 0.0685 = 6.69 * 10^-4</td></tr>
<tr><td colspan="2">S1
1 NP1, 2 VP10, 1 * 0.14 * 49/55 * 0.33 * 23/44 * 0.0145 = 3.12 * 10^-4</td></tr>
<tr><td></td><td>VP11
1 V2, 2 PP1, 0.15 * 23/44 * 0.0685</td></tr>
<tr><td></td><td>VP10</td></tr>
</table>

	1 V2, 2 NP8, 0.33 * 23/44 * 0.0145		
		VP9 1 V3, 2 NP4, 0.33 * 30/61 * 0.1378	
	VP8 1 V2, 2 NP8, 0.33 * 23/44 * 0.0145		
	VP7 1 V2, 2 NP3, 0.33 * 23/44 * 0.0656		
VP6 1 V1, 2 NP7, 0.33 * 5/55 * 7.04 * 10^-3			
VP5 1 V1, 2 NP2, 0.33 * 5/55 * 0.0668			
		PP1 1 P1, 2 NP4, 1 * 21/61 * 0.14 * 64/65 = 0.0685	
		NP8 1 N3, 2 N4, 0.09 * 10/61 * 64/65 = 0.0145	
	NP7 1 N2, 2 N3, 0.09 * 21/44 * 10/61 = 7.04 * 10^-3		
NP6 1 N1, 2 N2, 0.09 * 49/55 * 21/44 = 0.0383			
NP5 1 ART1, 2 N2, 0.55 * 1/55 * 21/44 = 4.77 * 10^-3			
VP1 1 V1, 0.32 * 5/55	VP2 1 V2, 0.32 * 23/44	VP3 1 V3, 0.32 * 30/61	VP4 1 V4, 0.32 * 1/65
NP1 1 N1	NP2 1 N2	NP3 1 N3	NP4 1 N4

0.14 * 49/55	0.14 * 21/44 = 0.0668	0.14 * 10/61 = 0.0656	0.14 * 64/65 = 0.1378
ART1 1/55		P1 21/61	
V1 5/55	V2 23/44	V3 30/61	V4 1/65
N1 49/55	N2 21/44	N3 10/61	N4 64/65
Fruit	flies	like	birds

Như vậy **fruit flies like birds** là **N N V N**.

4. Specify PMI between two words, Positive PMI between two words in the below table

	aadvark	computer	data	pinch	result	sugar
Apricot	0	0	0	0	1	0
Pineapple	0	0	0	0	1	0
Digital	0	2	1	0	1	0
Information	0	1	6	0	4	0

Giải

p(w=Digital | c=computer) = 2/17

p(w=Digital) = 4/17

p(c=computer) = 3/17

Tương tự, ta có bảng sau

	p(w, c)						p(w)
	aadvark	computer	data	pinch	result	sugar	
Apricot	0	0	0	0	1/17	0	1/17
Pineapple	0	0	0	0	1/17	0	1/17
Digital	0	2/17	1/17	0	1/17	0	4/17
Information	0	1/17	6/17	0	4/17	0	11/17
p(c)	0	3/17	7/17	0	7/17	0	

Suy ra:

pmi(Apricot, result) = $\log_2(1/17 / (7/17 * 1/17)) = 1.2801$

Tương tự ta có bảng sau:

	pmi(w, c)

	aadvark	computer	data	pinch	result	sugar
Apricot					1.2801	
Pineapple					1.2801	
Digital		1.5025	-0.7199		-0.7199	
Information		-0.9569	0.4056		-0.1793	

	ppmi(w, c)					
	aadvark	computer	data	pinch	result	sugar
Apricot					1.2801	
Pineapple					1.2801	
Digital		1.5025				
Information			0.4056			