	i	want	to	eat	chinese	food	lunch	spend
i	6	828	1	10	1	1	1	3
want	3	1	609	2	7	7	1	2
To	3	1	5	687	3	1	7	212
eat	1	1	3	19	17	3	43	1
chinese	2	1	1	1	1	83	2	1
food	16	1	16	1	2	5	1	1
lunch	3	1	1	1	1	2	1	1
spend	2	1	2	1	1	1	1	1

$V = 1446$

$$P^*(i|i) = \frac{6}{2533+1446} = 0.0015$$

$$P^*(want|i) = \frac{828}{2533+1446} = 0.208$$

$$P^*(To|i) = \frac{1}{2533+1446} = 0.00025$$

$$P^*(eat|i) = \frac{10}{2533+1446} = 0.0025$$

$$P^*(chinese|i) = \frac{1}{2533+1446} = 0.00025 = P(food|i) = P(lunch|i)$$

$$P^*(spend|i) = \frac{3}{2533+1446} = 0.00075$$

$$c^*(i|i) = \frac{6 \times 2533}{2533+1446} = 3.7995$$

$$c^*(want|i) = \frac{828 \times 2533}{2533+1446} = 526.864$$

$$c^*(to|i) = \frac{1 \times 2533}{2533+1446} = 0.6333 = c^*(chinese|i) = c^*(food|i) = c^*(lunch|i)$$

$$c^*(eat|i) = \frac{10 \times 2533}{2533+1446} = 6.3325$$

$$c^*(spend|i) = \frac{3 \times 2533}{2533+1446} = 1.8998$$

Laplace Smoothing:

- Sharp change in counts & probabilities occur - b

Smoothing:

1) Unsmoothed
2) Laplace - Add - on
3) Add - k
4) Good Turing
5) Combining
   estimation — Backoff

Good Turing Smoothing:

- N_c - Count of things we've seen c times

Sam Sam I am Sam I do not eat

I	3
Sam	2
am	2
do	1
not	1
eat	1

$N_1 = 3$
$N_2 = 2$
$N_3 = 1$

N_c - frequency of frequency c

N - no. of bigrams seen - V^2 = given
N_0 - no. of bigrams with count 0 = ?
N_c - no. of bigrams with count c

$$N_0 = V^2 - N$$

$$c^* = \frac{(c+1) \cdot N_{c+1}}{N_c}$$

c - original (real) word count

$(c+1).N_{c+1}$ - The probability mass for words with frequency c+1

c^* - new (adjusted) word count.

MLE count for N_c is c.

Eg:

10 carp	N1 = 3	N = 18
3 perch	N2 = 1	N_0
2 whitefish	N3 = 1	
1 trout	N10 = 1	
1 salmon	N_0 = 2	
1 eel		

MLE count of hitherto-unseen species (catfish or bass) is 0

P^*_{GT} (things with frequency 0 in training) $= \frac{N_1}{N} = \frac{3}{18}$

P^*_{GT} (bass) = P^*_{GT} (catfish) $= \frac{3}{18} * \frac{1}{N_0}$

	unseen (bass or catfish)	trout
c	0	1
MLE p	$p = \frac{0}{18} = 0$	$\frac{1}{18}$
c^*		$c^*(\text{trout}) = 2 \times \frac{N_2}{N_1} = 2 \times \frac{1}{3} = 0.67$
GT p^*_{GT}	$P^*_{GT}(\text{unseen}) = \frac{N_1}{N} = \frac{3}{18} = 0.17$	$P^*_{GT}(\text{trout}) = \frac{0.67}{18} = \frac{1}{27} = 0.037$

c^* computation isn't there for unseen i.e., for count 0

Revised count.

Consider vocabulary as {a, b, c}
Consider Corpus : b a b a a c b c a c a c
Find list of possible bigrams
Find out count of seen (observed) & unseen (unobserved) bigram
& Their frequencies using Good Turing.

	a	b	c
a	1	1	3
b	2	0	1
c	2	1	0

(ba, ab, ba, aa, ac, cb, bc, ca, ca, ac)

aa = 1 ba = 2 ca = 2
ab = 1 bb = 0 cb = 1
ac = 3 bc = 1 cc = 0

N = 11

3 b
5 a
4 c
N = 12

$N_0 = 2$
$N_1 = 4$
$N_2 = 2$
$N_3 = 1$

c	P_{MLE}	C^*_{GT}	P^*_{GT}	P^*_{GT} (based on no. of words)
c = 0	0	N_1	$\frac{4}{11}$ $\frac{N_1}{N}$	$P^*_{GT}(bb) = P^*_{GT}(cc) = \frac{2}{11}\frac{N_1}{N_0}$
c = 1	$\frac{1}{11}$	1.1	$2 \times \frac{2}{4} * \frac{1}{11} = \frac{1}{11}$	$P^*_{GT}(aa) = P^*_{GT}(ab) = P^*_{GT}(bc) = P^*_{GT}(cb) = \frac{1}{11}$
c = 2	$\frac{2}{11}$	$3 \times \frac{1}{2}$	$\frac{3}{2} \times \frac{1}{11} = \frac{3}{22}$	$P^*_{GT}(ba) = P^*_{GT}(ca) = \frac{3}{22}$
	$\frac{4}{11} + \frac{4}{11} + \frac{6}{22}$		$= \frac{8+8+6}{22}$	

Laplace:

	a	b	c
a	1	1	3
b	2	0	1
c	2	1	0

unsmoothed.

$$c_i^* = (c_{i+1}) * c(w_{i-1})$$

$$P_{Laplace}(w_i/w_{i-1}) = \frac{C(w_{i-1}\cdot w_i)+1}{C(w_{i-1})+V}$$

	a	b	c
a	2	2	4
b	3	1	2
c	3	2	1

Add one smoothed.

$V^2 = 9$

$$P^*(a/a) = \frac{2}{5+9} = \frac{2}{14} = \frac{1}{7}$$

$$P^*(b/a) = \frac{2}{5+9} = \frac{1}{7}$$

$$P^*(c/a) = \frac{4}{5+9} = \frac{4}{14} = \frac{2}{7}$$

$$P^*(a/b) = \frac{3}{3+9} = \frac{3}{12} = \frac{1}{4}$$

$$P^*(b/b) = 0 = P^*(c/c)$$

$$P^*(c/b) = \frac{1}{1+9} = \frac{1}{12}$$

$$P^*(a/c) = \frac{2}{4+9} = \frac{2}{13}$$

$$P^*(b/c) = \frac{1}{4+9} = \frac{1}{13}$$

$$\frac{(C+1)N}{N+V}$$

$$\frac{C(w_{n-1}\cdot w_n)+K}{C(w_{n-1})+KV}$$

a	b	c
6	3	4

	a	b	c
a	2	2	4
b	3	1	2
c	3	2	1

	a	b	c
a	1.25	1.25	2.5
b	1.5	0.5	1
c	1.71	1.14	0.54

Add one reconstructed bigram count c^* $= \frac{(c+1)N}{N+V} = \frac{2 \times 5}{5+3}$

	a	b	c
a	0.25	0.25	0.5
b	0.5	0.0	0.33
c			

$$P(w_i | w_{i-1}) = \frac{C((w_{i-1}.w_i)+1}{C(w_{i-1})+V}$$

$$P(a|a) = \frac{C(a.a)+1}{C(a)+V}$$

$\frac{2}{5} \times 5$

$$= \frac{2}{5+3}$$

$$= 2/8$$

$$= 0.25$$

Interpolation:

- Simple linear interpolation:
- We combine different order N-grams by linearly interpolating all models.
- To combine estimate Trigram probability $P(w_n / w_{n-1} w_{n-2})$
-

In $P(w_n | w_{n-1} w_{n-2}) = \lambda_1 P(w_n | w_{n-1} w_{n-2}) + \lambda_2 P(w_n | w_{n-1}) + \lambda_3 P(w_n)$

BackOff N-gram modelling : Katz backoff

$$P_{katz}(z|x,y) = \begin{cases} P^*(z|x,y) & \text{if } C(x,y,z) > 0 \\ \alpha(x,y) P_{katz}(z|y) & \text{else if } C(y,z) > 0 \\ P^*(z) & \text{otherwise} \end{cases}$$

$$P_{katz}(z|y) = \begin{cases} P^*(z|y) & \text{if } C(y,z) > 0 \\ \alpha(y) P^*(z) & \text{otherwise} \end{cases}$$

Different

Verbs : VB, VBP, VBZ, VBD, VBG, VBN
- base, present-non-3rd, present-3rd, past-ing, -en

Nouns : NNP, NNPS, NN, NNS
- proper/common, singular/plural (singular includes mass+generic)

Adjectives : JJ, JJR, JJS (base, comparative, superlative)

Adverbs : RB, RBR, RBS, RP (base, comparative, superlative, particle)

Pronouns : PRP, PP$ (personal, possessive)

Interrogatives : WP, WP$, WDT, WRB (compare to : PRP, PP$, DT, RB)

Other Closed Class : CC, CD, DT, PDT, IN, MD

Punctuation : # $.,

Examples:

The grand jury commented on a no. of other topics

- Eg : Determiners : the and a
- The adjectives grand & other
- The common nouns jury, number & topics,
- the past tense verb commented
- .
- .

1) $P(t_n / t_{n-1})$ → tag transition probability

2) $P(w_i / t_i)$ → observation likelihood probability / estimate

HMM part of speech tagging

- We want, out of all sequences of n tags $t_1 \ldots t_n$ the single sequence such that $P(t_1 \ldots t_n / w_1 \ldots w_n)$ is highest

$$\hat{t}_1^n = \arg\max_{t_1^n} P(t_1^n / w_1^n)$$

- Hat ^ means 'our estimate is best one'

$$P(x/y) = \frac{P(y/x) \cdot P(x)}{P(y)}$$

$$P(t_1^n / w_1^n) = \frac{P(w_1^n / t_1^n) \cdot P(t_1^n)}{P(w_1^n)}$$

$$=$$

HMM simplifying assumptions

- It is too hard to compute → $\hat{t}_1^n = \arg\max_{t_1^n} P(w_1^n / t_1^n)$ [likelihood]

- Assumption 1 : The probability of a word is dependent only on its part-of-speech, not on the surrounding words or other tags around it.

$$P(w_1^n / t_1^n) \approx \prod_{i=1}^{n} P(w_i / t_i)$$

- Assumption 2 : The probability of a tag appearing is dependent only on the previous tag (bigram assumption)

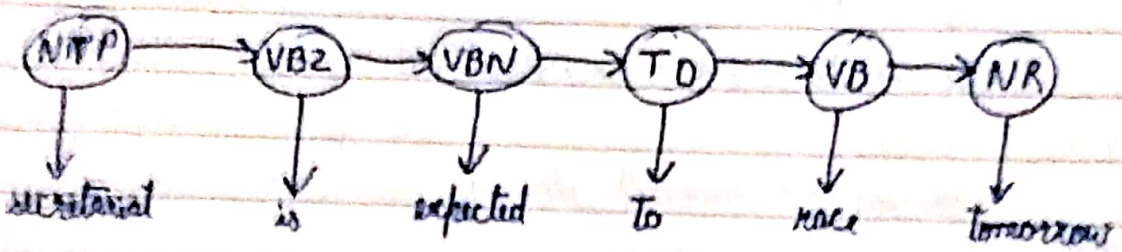
$$P(t_1^n) \approx \prod_{i=1}^{n} P(t_i \mid t_{i-1})$$

word likelihoods, tag transition probabilities

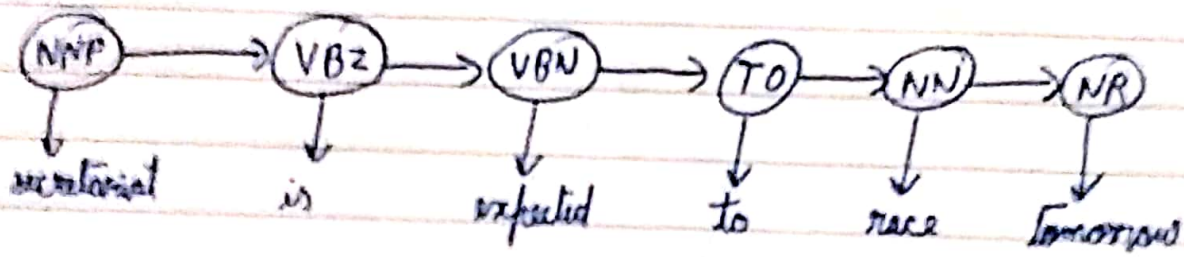
$$\hat{t}_1^n = \arg\max P(t_1^n \mid w_1^n) \approx \arg\max \prod_{i=1}^{n} P(w_i \mid t_i)\, P(t_i \mid t_{i-1})$$

Disambiguating "race"

a)



b)



a) $P(VB \mid TO) \times P(NR \mid VB) \times P(race \mid VB)$

$= 0.83 \times 0.0027 \times 0.00012$

$= \underline{2.69 \times 10^{-7}}$

b) $P(NN \mid TO) \times P(NR \mid NN) \times P(race \mid NN)$

$= 0.00047 \times 0.0012 \times 0.00057$

$= 3.21 \times 10^{-10}$

Start probability, $P(\pi) = 1$

Draw parse

S → NP VP

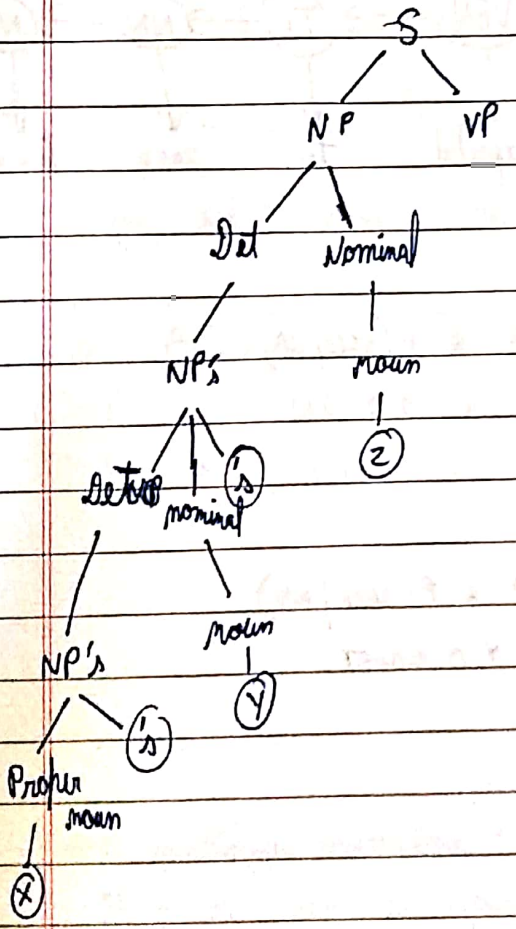
NP → pronoun |
Proper-noun |
Det nominal

Det → NP's

Det → the | a | an | This | these | that

Nominal → Nominal Noun |
Noun

~~Draw~~



S → NP VP

NP → Pronoun | Proper noun | Det Nominal

Nominal → Nominal noun | Noun | Nominal PP (PP) (PP)

VP → Verb | Verb NP | Verb NP PP | Verb PP

PP → Preposition NP

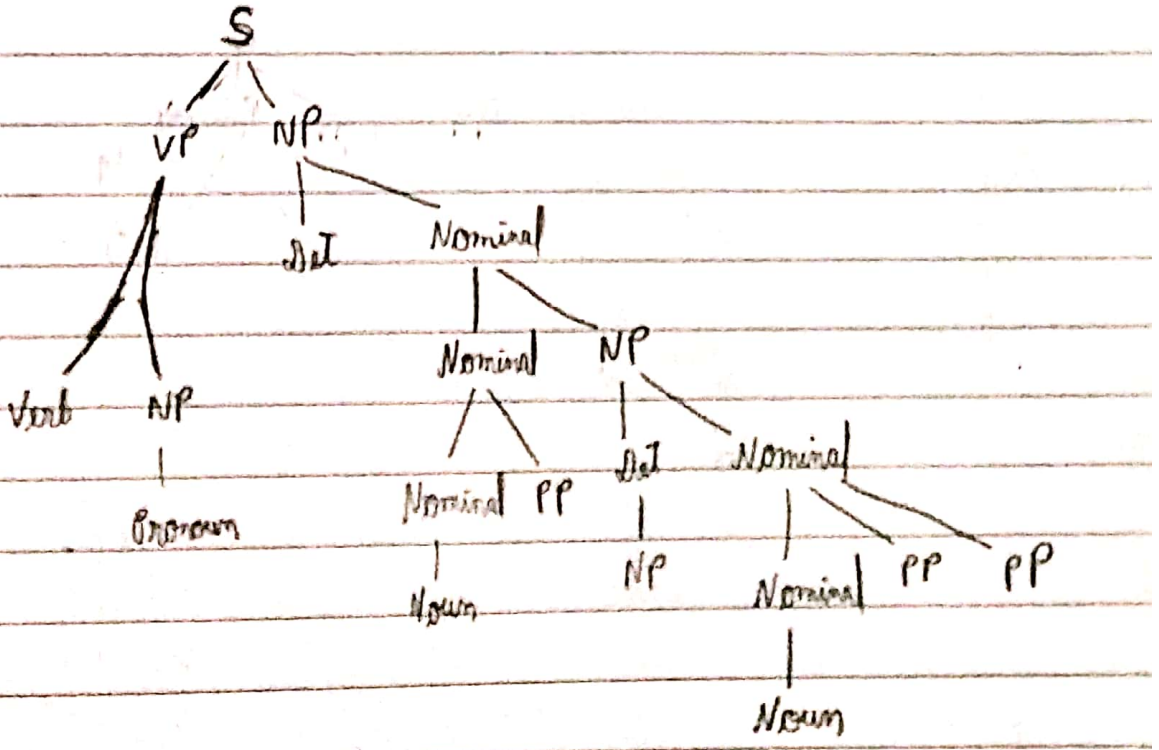
PP postmodifiers
- any

Draw Top-down parse Tree for following

1) [All] [the] [morning flights] [from P1] [To P2] [leaving [before [10]]]
Predet Det Nominal PP PP gerundV PP N gerundVP

Nominal → Nominal gerundVP
gerundVP → gerundV NP | gerundV PP | gerundV | gerundV NPP
gerundV → being | preferring | arriving | leaving | ...

[Show] [me] [the] [meal] [on] [flight Boeing -780] [from P1] [To P2]
Verb pronoun Det Noun Prep NP PP PP



S → NP VP

NP → Pre-Head Modifiers Nominal Post-head-Modifiers

Pre-Head-Modifiers → PreDet Det Post-Determiners

Post-Determiners → (Card) (Ord) (Quant) (AP)

NP → Pronoun / Proper-Noun / Det Nominal

Det → A's

Nominal → Nominal Noun / Noun / Nominal NP

Pre-head-Modifiers → (Card) (Ord) (Quant) (AP)

Nominal → Nominal PP (PP) (PP)

Nominal → Nominal GerundVP

Nominal → Nominal RelClause

RelClause → (who/that) VP

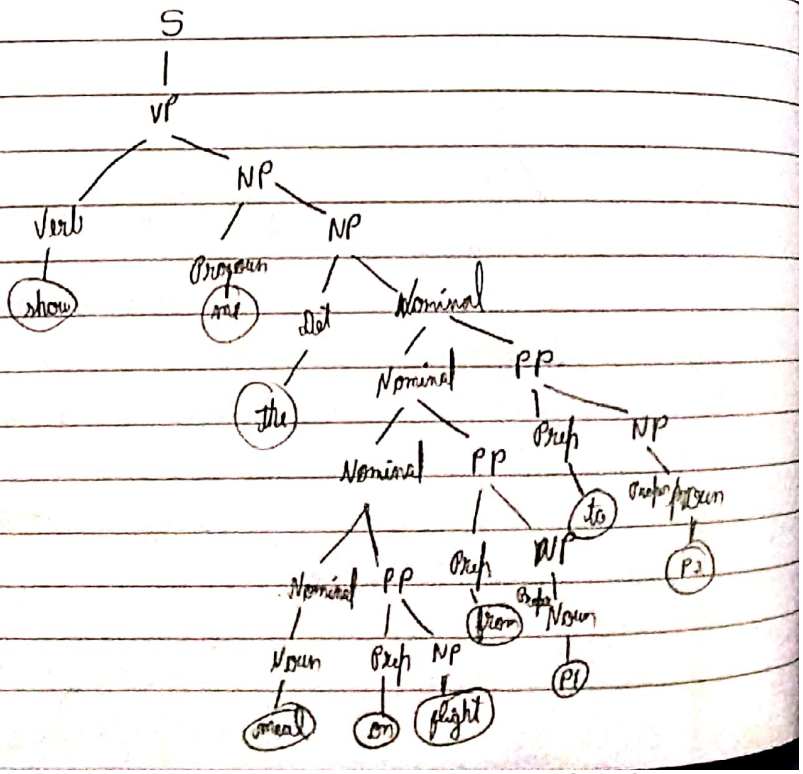
GerundVP → GerundV / GerondV NP / GerundV PP / GerundV NP PP

VP → Verb / Verb NP / Verb PP / Verb NP PP

PP → Preposition NP

Note: Post-head-Modifiers → PP-Modifier / Non-finite-Modifier / Rel-Cl-Modifier

PP-Modifier →



CC	co-ordin conjunction	and, but, or
CD	cardinal no.	one, two, three
DT	determiner	a, an, the
EX	existential 'there'	there
FW	Foreign word	mea culpa
IN	Preposition / sub-conj	of, in, by
JJ	adjective	yellow
JJR	adj. comparative	bigger
JJS	adj superlative	wildest
LS	list item marker	1, 2, one
MD	Modal	can, should
NN	Noun, singular or mass	llama
NNS	Noun plural	llamas
NNP	Proper noun singular	IBM
NNPS	Proper noun, plural	Carolinas
PDT	pre-determiner	all, both
POS	possessive ending	's
PP	personal pronoun	I, you, he
PPS	possessive pronoun	your, one's
RB	adverb	quickly, never
RBR	adverb, comparative	faster
RBS	adverb, superlative	fastest
RP	particle	up, off
SYM	symbol	+, %, &
TO	"to"	to
UH	interjection	oops, oh
VB	Verb, base form	eat
VBD	Verb, past tense	ate
VBG	Verb, gerund	eating
VBN	Verb, past participle	eaten
VBP	Verb, non-3sg pres	eat
VBZ	Verb, 3sg pres	eats
WDT	wh-determiner	which, that
WP	wh-pronoun	what, who

WP$	possessive wh-	whose
WRB	wh- adverb	how, where

Auxillaries
- Modal Verbs
- do
- have
- progressive be
- passive be

modal < perfect < progressive < passive

modal perfect	could have been a contender
modal passive	will be married
perfect progressive	have been feasting
modal perfect passive	might have been prevented

Co-ordination