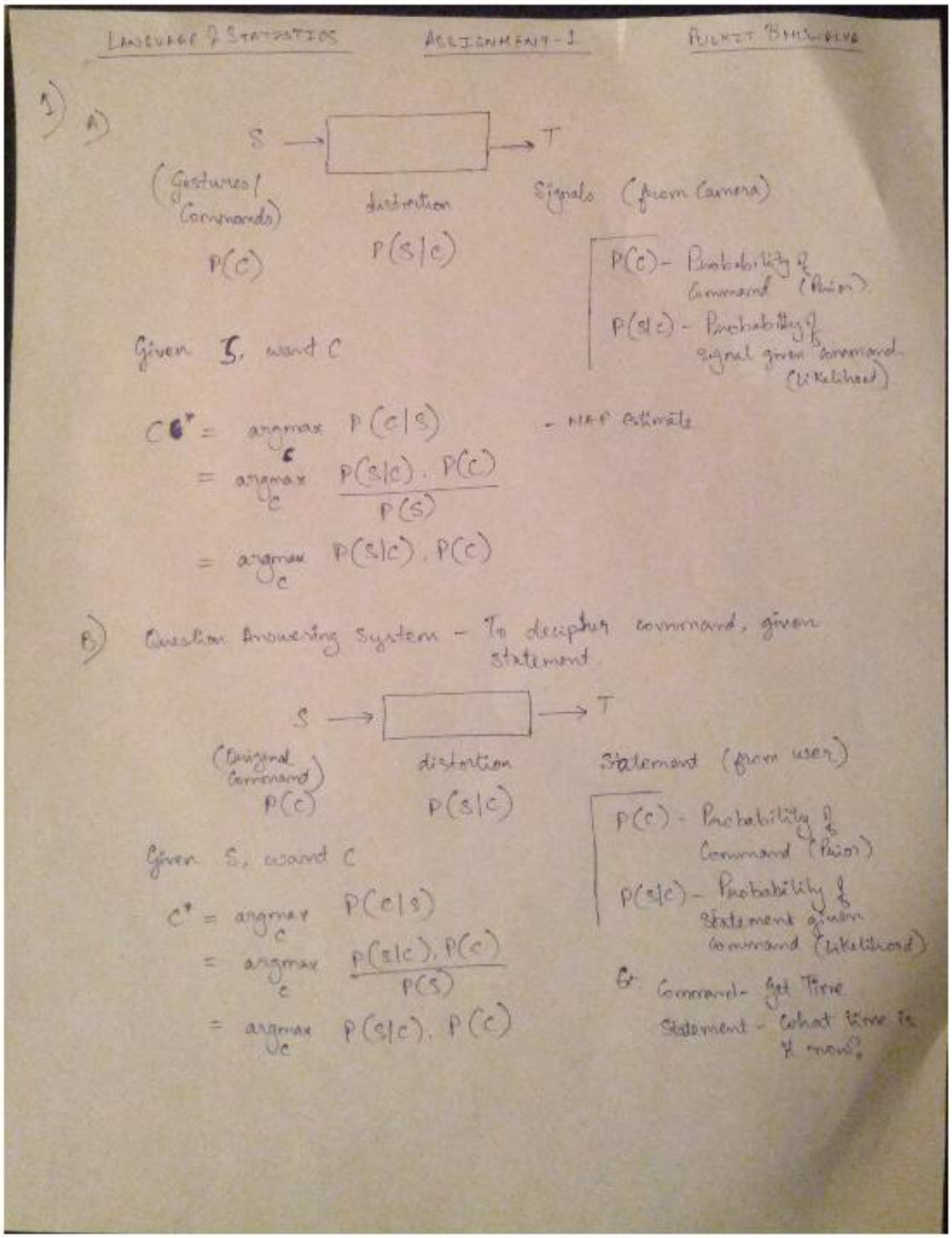
Question 1 : Source Channel Paradigm (image attached)



Question 2 - Sub-Language (Type-token curve estimation)

Sub-Language Description

The sub-languages chosen were the lyrics of the music artists Bob Dylan and the Beatles. I was interested in noticing the kind of words that were most popular in the songs and how that would be different from regular

english. They were also interesting topics of choice given the following diversities -

1. Dylan was an american artist, while the Beatles (despite being very popular in America) were based out of Europe. There are quite a few subtle differences in the english in geography.
2. Both were extremely popular, lived in the same era and are considered legends in great poetry. Dylan in fact is considered by many to be worthy of a nobel prize in literature.
3. Dylan was often very politically motivated. The Beatles as an ensemble were less so, but some members individually could often be.

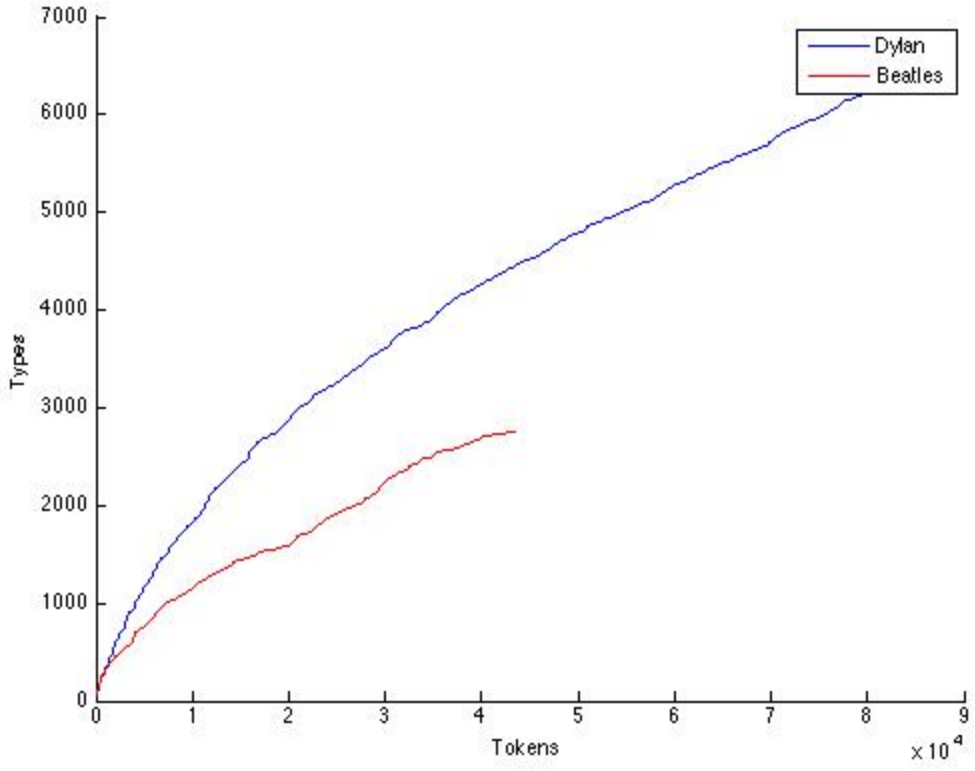
Bob Dylan lyrics - 87964 tokens, 6559 types
The Beatles lyrics - 43381 tokens, 2758 types

Pre-processing

All words with non-word characters such as exclamations, numbers etc. were removed. They were also stripped of leading or training whitespace characters.

Head and Tail lists of both the corpora have been attached.

Type Token Curve



Hypotheses

1. I suspected that the Beatles would have a richer vocabulary given that they are an ensemble of 4 members,

while Dylan is a single artist who wrote the songs himself.

Verdict - I was quite wrong. Dylan had a much richer vocabulary and a higher type-token curve.

2. I expected most of the wordings to very simple since these are simple songs. And given that they had a very large outreach and were extremely popular I expected them to have a very simple vocabulary.

Verdict - I was right. The most popular words - in both the corpora are extremely simple words that would be used in common spoken language and would be extremely understandable.

Other interesting points.
The most popular Beatles words are "you" and "i". "the" is third. also love is as high as the 10th word. This seems consistent that most beatles songs were about love and relationships.

I also looked at the frequency of male/female words in the both the corpora. They are reasonably matched - in first person, male words tend to be more prominent, whereas in third person female words. That makes sense since both the artists were male.

Dylan - man(227), him(226), his(412), he(711), boy(40), men(64), boys(17), himself(10)
woman(61), her(359), she(452), women(21), girl(73), girls(13), herself(4)

Beatles - man(91), him(29), his(84), he(120), boy(33), men(2), boys(18), himself(2)
woman(24), her(269), she(326), women(1), girl(148), girls(12), herself(2)

Dylan Top 50

the 4919
i 2780
you 2599
and 2556
to 2471
a 2076
in 1582
of 1495
my 1224
me 969
that 924
it 891
on 886
your 845
for 759
he 711
is 675
all 654
be 649
was 640

but 636
with 591
they 553
like 481
down 453
she 452
just 452
when 445
got 439
no 433
can 414
his 412
up 411
know 397
so 396
her 359
what 357
are 348
if 347
from 345
out 343
one 331
have 330
not 293
come 291
gonna 290
as 287
love 285
there 279
see 278

Dylan Bottom 50

frontline 1
tossed 1
ringed 1
dumb 1
temporary 1
continual 1
clung 1
slash 1
rice 1
beans 1
worrying 1
permanent 1
frontier 1
pus 1

wichita 1
shaven 1
slap 1
smells 1
railings 1
tent 1
shoots 1
tailgates 1
substitutes 1
strap 1
roots 1
khan 1
supplied 1
coach 1
learning 1
extreme 1
midstream 1
corkscrew 1
cancel 1
careless 1
drummin 1
correct 1
timbukto 1
target 1
direct 1
anne 1
relationship 1
rimbaud 1
compare 1
affair 1
honolulu 1
ashtabula 1
cherish 1
plaything 1
league 1
disembark 1

Beatles Top 50

you 2195
i 1705
the 1420
to 1125
and 982
a 952
me 923
my 582

in 536
love 527
be 488
that 476
it 474
all 448
know 440
your 414
of 388
is 382
do 360
on 343
she 326
so 314
for 296
oh 285
her 269
what 267
but 264
with 256
when 249
if 233
got 227
can 222
yeah 219
like 206
now 206
baby 196
get 194
say 193
see 191
want 187
just 180
come 180
will 178
go 177
no 167
well 167
good 159
little 154
we 154
was 153

Beatles Bottom 50

river 1
marmalade 1

eagle 1
licks 1
bone 1
cellophane 1
towering 1
photograph 1
mist 1
taxis 1
fountain 1
scarlet 1
planned 1
shadow 1
welcome 1
slaggers 1
featuring 1
marshmellow 1
pum 1
grade 1
prrr 1
watched 1
nerve 1
burnt 1
toast 1
health 1
eyeball 1
negotiations 1
investigation 1
jobber 1
lorry 1
limousine 1
lp 1
badly 1
madly 1
split 1
hearing 1
recorded 1
gather 1
clowns 1
engaged 1
age 1
drift 1
ribbon 1
fractured 1
burnette 1
dorsey 1
lame 1
incredibly 1
daughter 1