|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | ID | Section Contributed | Section Edited | Other Contributions |
| Thompson, Josh | 3018 | 2/3 Corpus Collection, lexical diversity, longest sentence.  Write up | Code | Citations.  Excel sheet to keep track of corpus. |
| Ty, David | 301334955 | 1/3 Corpus collection,  Top ten vowels  Stemming  Collocation  Length | Code | Created the github. |

1. Corpus of Eurogamer reviews – Genre: Video Game Articles

1. The length (in words).

5858

2. The lexical diversity.

0.29085392284351974

3. The longest sentence (type the sentence and also give the number of words). Hint: look at the Gutenberg part of Section 2.1 in NLTK.

['As', 'a', 'surprisingly', 'linear', 'adventure', 'it', 's', 'difficult', 'to', 'talk', 'much', 'about', 'Estelle', 's', 'journey', 'without', 'giving', 'something', 'important', 'away', 'and', 'as', 'her', 'story', 'is', 'essentially', 'as', 'long', 'or', 'as', 'brief', 'as', 'you', 'want', 'it', 'to', 'be', 'you', 'can', 'belt', 'through', 'it', 'and', 'be', 'done', 'within', 'a', 'few', 'hours', 'or', 'savour', 'each', 'new', 'environment', 'and', 'scour', 'every', 'inch', 'of', 'it', 'for', 'days', 'I', 'imagine', 'there', 'may', 'be', 'some', 'people', 'or', 'places', 'you', 'could', 'miss', 'entirely', 'and', 'complete', 'the', 'game', 'never', 'knowing', 'that', 'they', 'existed']

84 Words long

As a surprisingly linear adventure it s difficult to talk much about Estelle s journey without giving something important away and as her story is essentially as long or as brief as you want it to be you can belt through it and be done within a few hours or savour each new environment and scour every inch of it for days I imagine there may be some people or places you could miss entirely and complete the game never knowing that they existed

4. The top collocations.

[('A', 'Space'), ('Fire', 'Emblem'), ('Space', 'for'), ('the', 'Unbound'), ('This', 'is'), ('even', 'if'), ('Three', 'Houses'), ('for', 'the'), ('Emblem', 'Engage'), ('you', 'can')]

5. The top ten words that start with each of the vowels (involves using FreqDist).

[('a', 183), ('and', 175), ('an', 37), ('as', 33), ('are', 30), ('at', 20), ('all', 15), ('army', 10), ('about', 9), ('also', 8)]

[('engage', 18), ('even', 15), ('each', 14), ('every', 13), ('estelle', 10), ('emblem', 10), ('enemy', 6), ('end', 5), ('entirely', 4), ('enemies', 4)]

[('in', 78), ('it', 70), ('i', 69), ('is', 56), ('its', 21), ('into', 18), ('if', 18), ('inkulinati', 13), ('instead', 7), ('isn', 7)]

[('of', 125), ('or', 24), ('on', 19), ('one', 18), ('off', 16), ('out', 15), ('over', 13), ('other', 10), ('own', 8), ('only', 7)]

[('up', 16), ('unbound', 11), ('units', 9), ('unique', 5), ('until', 4), ('unit', 3), ('unless', 2), ('underneath', 2), ('used', 2), ('upon', 2)]

6. A stemmed version of the longest sentence (extracted above in 3).

as

a

surpris

linear

adventur

it

s

difficult

to

talk

much

about

estell

s

journey

without

give

someth

import

away

and

as

her

stori

is

essenti

as

long

or

as

brief

as

you

want

it

to

be

you

can

belt

through

it

and

be

done

within

a

few

hour

or

savour

each

new

environ

and

scour

everi

inch

of

it

for

day

i

imagin

there

may

be

some

peopl

or

place

you

could

miss

entir

and

complet

the

game

never

know

that

they

exist

1. Noam Chomsky Interview – Genre: Interview

1. The length (in words).

11165

2. The lexical diversity.

0.1660031548110869

3. The longest sentence (type the sentence and also give the number of words). Hint: look at the Gutenberg part of Section 2.1 in NLTK.

['If', 'I', 'go', 'back', 'to', 'early', 'childhood', 'some', 'of', 'the', 'reading', 'that', 'had', 'a', 'lasting', 'impact', 'which', 'I', 'don', 't', 'even', 'think', 'it', 's', 'in', 'English', 'was', 'the', 'Hebrew', 'essays', 'of', 'a', 'turn', 'of', 'the', 'century', 'essayist', 'that', 'went', 'by', 'the', 'name', 'of', 'Ahad', 'Ha', 'am', 'writing', 'about', 'partly', 'intellectual', 'contributions', 'which', 'were', 'significant', 'partly', 'talking', 'about', 'the', 'developing', 'situation', 'in', 'what', 'was', 'then', 'Palestine', 'which', 'had', 'a', 'large', 'effect', 'on', 'my', 'thinking', 'ever', 'since']

75 words long

If I go back to early childhood some of the reading that had a lasting impact which I don t even think it s in English was the Hebrew essays of a turn of the century essayist that went by the name of Ahad Ha am writing about partly intellectual contributions which were significant partly talking about the developing situation in what was then Palestine which had a large effect on my thinking ever since

4. The top collocations.

[('EZRA', 'KLEIN'), ('NOAM', 'CHOMSKY'), ('going', 'to'), ('I', 'think'), ('in', 'the'), ('of', 'the'), ('a', 'lot'), ('I', 'mean'), ('United', 'States'), ('I', 'don')]

5. The top ten words that start with each of the vowels (involves using FreqDist).

[('a', 267), ('and', 217), ('are', 72), ('about', 68), ('as', 46), ('at', 29), ('all', 28), ('an', 27), ('any', 19), ('american', 15)]

[('ezra', 35), ('even', 15), ('example', 7), ('every', 7), ('end', 6), ('early', 6), ('economic', 6), ('everything', 6), ('ever', 5), ('essentially', 5)]

[('it', 261), ('in', 188), ('is', 158), ('i', 153), ('if', 54), ('into', 19), ('industry', 14), ('idea', 8), ('international', 7), ('its', 6)]

[('of', 356), ('on', 55), ('or', 31), ('one', 31), ('out', 31), ('our', 24), ('other', 18), ('over', 16), ('own', 12), ('others', 7)]

[('up', 18), ('us', 16), ('united', 8), ('used', 7), ('under', 7), ('understand', 6), ('use', 5), ('u', 5), ('using', 4), ('unfortunately', 4)]

6. A stemmed version of the longest sentence (extracted above in 3).

if

i

go

back

to

earli

childhood

some

of

the

read

that

had

a

last

impact

which

i

don

t

even

think

it

s

in

english

was

the

hebrew

essay

of

a

turn

of

the

centuri

essayist

that

went

by

the

name

of

ahad

ha

am

write

about

part

intellectu

contribut

which

were

signific

part

talk

about

the

develop

situat

in

what

was

then

palestin

which

had

a

larg

effect

on

my

think

ever

sinc

1. Wizard of Oz Script – Genre: Movie Script

1. The length (in words).

29996

2. The lexical diversity.

0.07287356905348123

3. The longest sentence (type the sentence and also give the number of words). Hint: look at the Gutenberg part of Section 2.1 in NLTK.

['Gale', 'Sitting', 'room', 'Aunt', 'Em', 'and', 'Miss', 'Gulch', 'seated', 'Dorothy', 'enters', 'carrying', 'Toto', 'in', 'her', 'arms', 'CAMERA', 'TRUCKS', 'forward', 'on', 'them', 'PANS', 'to', 'right', 'with', 'Dorothy', 'to', 'Uncle', 'Henry', 'then', 'Pans', 'her', 'left', 'to', 'Aunt', 'Em', 'and', 'Miss', 'Gulch', 'Miss', 'Gulch', 'shows', 'order', 'to', 'Aunt', 'Em', 'Uncle', 'Henry', 'enters', 'looks', 'at', 'the', 'order', 'Miss', 'Gulch', 'picks', 'up', 'basket', 'rises', 'Dorothy', 'screams', 'at', 'Miss', 'Gulch', 'Miss', 'Gulch', 'tries', 'to', 'take', 'Toto', 'away', 'from', 'Dorothy', 'Uncle', 'Henry', 'takes', 'Toto', 'puts', 'him', 'into', 'basket']

81 words long

Gale Sitting room Aunt Em and Miss Gulch seated Dorothy enters carrying Toto in her arms CAMERA TRUCKS forward on them PANS to right with Dorothy to Uncle Henry then Pans her left to Aunt Em and Miss Gulch Miss Gulch shows order to Aunt Em Uncle Henry enters looks at the order Miss Gulch picks up basket rises Dorothy screams at Miss Gulch Miss Gulch tries to take Toto away from Dorothy Uncle Henry takes Toto puts him into basket

- Longest sentence originally had a lot of dashes in it, we used regex in our code to filter to alphabetical words.

4. The top collocations.

[('Tin', 'Man'), ('TIN', 'MAN'), ('CAMERA', 'PANS'), ('CAMERA', 'TRUCKS'), ('DOROTHY', 'Oh'), ('Aunt', 'Em'), ('LAP', 'DISSOLVE'), ('DISSOLVE', 'TO'), ('shooting', 'past'), ('Dorothy', 'and')]

1. The top ten words that start with each of the vowels (involves using FreqDist).

[('and', 768), ('as', 344), ('a', 331), ('at', 167), ('all', 115), ('aunt', 69), ('about', 66), ('are', 64), ('away', 52), ('around', 40)]

[('em', 101), ('els', 43), ('exits', 32), ('enters', 26), ('ever', 23), ('enter', 23), ('eyes', 22), ('exit', 19), ('emerald', 14), ('even', 12)]

[('i', 474), ('in', 360), ('it', 261), ('is', 102), ('into', 72), ('if', 58), ('int', 30), ('isn', 9), ('images', 6), ('incubator', 5)]

[('of', 449), ('oh', 278), ('on', 233), ('o', 225), ('out', 158), ('oz', 128), ('over', 59), ('one', 40), ('others', 37), ('old', 32)]

[('up', 181), ('uncle', 45), ('us', 27), ('under', 17), ('use', 6), ('uh', 5), ('upon', 3), ('until', 3), ('used', 2), ('ugly', 2)]

6. A stemmed version of the longest sentence (extracted above in 3).

gale

sit

room

aunt

em

and

miss

gulch

seat

dorothi

enter

carri

toto

in

her

arm

camera

truck

forward

on

them

pan

to

right

with

dorothi

to

uncl

henri

then

pan

her

left

to

aunt

em

and

miss

gulch

miss

gulch

show

order

to

aunt

em

uncl

henri

enter

look

at

the

order

miss

gulch

pick

up

basket

rise

dorothi

scream

at

miss

gulch

miss

gulch

tri

to

take

toto

away

from

dorothi

uncl

henri

take

toto

put

him

into

basket

References

Blake, V. (2023, January 26). *Season: A letter to the future review - a thoughtful premise with some minor snags*. Eurogamer. Retrieved February 2, 2023, from https://www.eurogamer.net/season-a-letter-to-the-future-review-a-thoughtful-premise-with-some-minor-snags

Cassidy, R. (2023, January 30). *Inkulinati Review – deep and daft tactical play*. Eurogamer. Retrieved February 2, 2023, from https://www.eurogamer.net/inkulinati-review-deep-and-daft-tactical-play

Klein, E., & Chomsky, N. (2021, April 23). Transcript: Ezra Klein Interviews Noam Chomsky. *The New York Times*. other. Retrieved February 2, 2023, from https://www.nytimes.com/2021/04/23/podcasts/ezra-klein-podcast-noam-chomsky-transcript.html.

Langley, N., Ryerson, F., & Woolf, E. A. (1939). *Wizard of Oz, The Script*. *The Internet Movie Script Database (IMSDb)*. Retrieved February 2, 2023, from https://imsdb.com/Movie%20Scripts/Wizard%20of%20Oz,%20The%20Script.html.

Stockdale, H. (2023, January 17). *Fire emblem engage review - a sideways step for the series that celebrates the classics*. Eurogamer. Retrieved February 2, 2023, from https://www.eurogamer.net/fire-emblem-engage-review-a-sideways-step-for-the-strategy-series-that-celebrates-the-classics

Stockdale, H. (2023, January 23). *Forspoken Review - a slow burner that's not without its charms*. Eurogamer. Retrieved February 2, 2023, from https://www.eurogamer.net/forspoken-review-a-slow-burner-thats-not-without-its-charms

Tapsell, C. (2023, February 1). *A space for the unbound review - a slice of life, and all its pain*. Eurogamer. Retrieved February 2, 2023, from https://www.eurogamer.net/a-space-for-the-unbound-review-a-slice-of-life-and-all-its-pain