

Machine Learning for Biblical Hebrew



Mathias Coeckelbergs

Goals

What level of contextual analysis is useful?

Which features are important?

Data Modelling/pre-processing?

Motivating Example: Lexical Choice

נצעק נזעקת
ויזעקו יזעק
יצעק וזעקו ויזעק
ויצעק נזעקו וזעקתם
תזעק ויצעק מצעק
צעקים צעקה

Motivating Example: Lexical Choice

אירא
יבעתוני יראה
יבעתהו וייראו
נבעתי נבעת ובעתו
ירא תירא
תיראי

ויחרדו
מחריד יפחד ותרגז
הרגזתני תפחד אפחד ורגזו
רגז נפחדת יחרדו תפחדו
ויחרד ויחרד
החריד

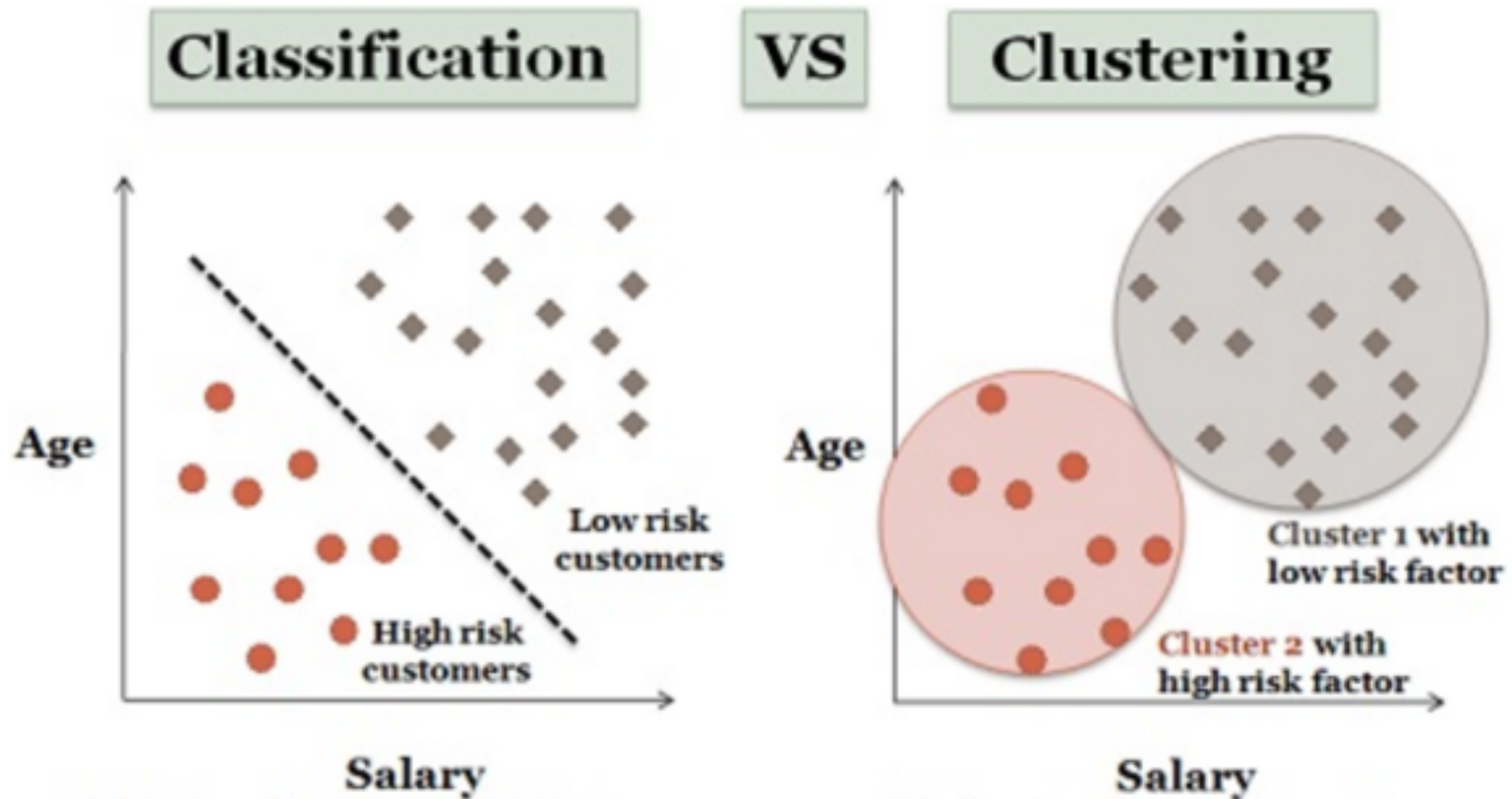
Motivating example: Lexical Choice

קום as 'going to' or 'beginning'.

E.g.: Gen 27:19 : קוּם-נָא שֹׁבָה, וְאָכְלָה מִצִּידֵי

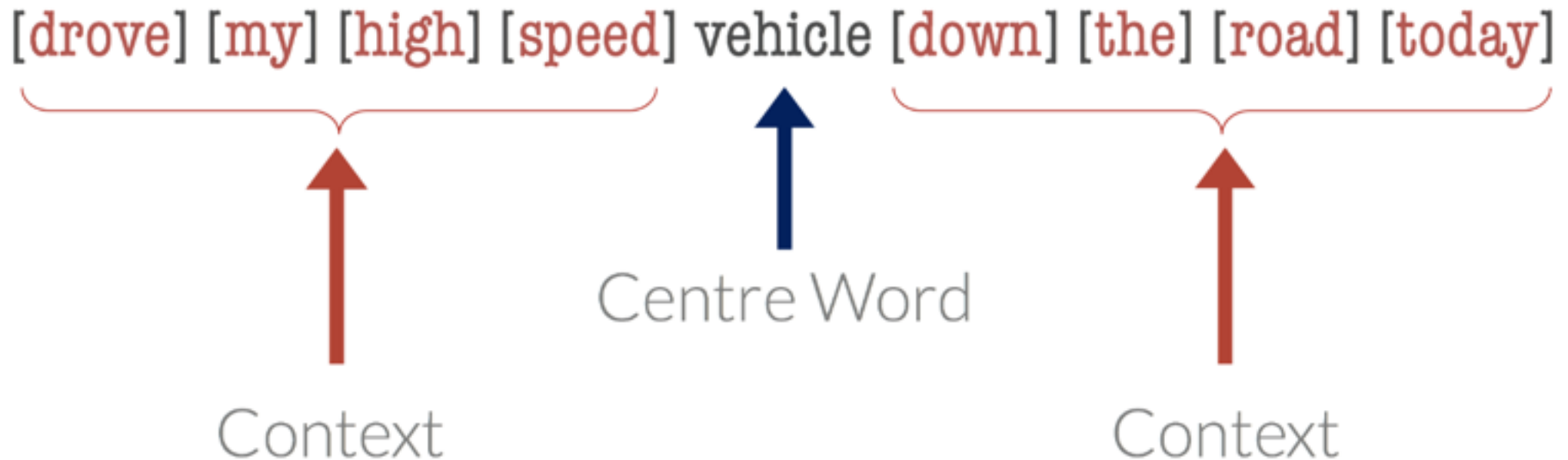
Stand up and sit down, and eat of my venison

Machine Learning



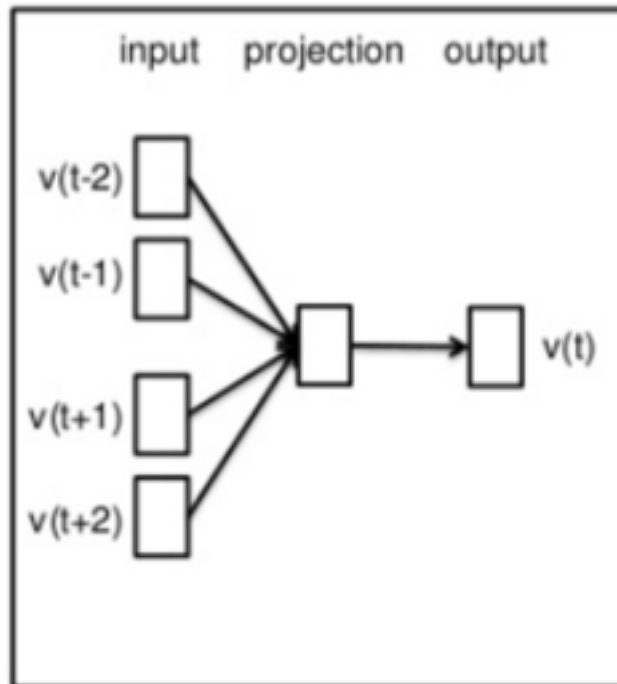
Risk classification for the loan payees on the basis of customer salary

Word Embeddings



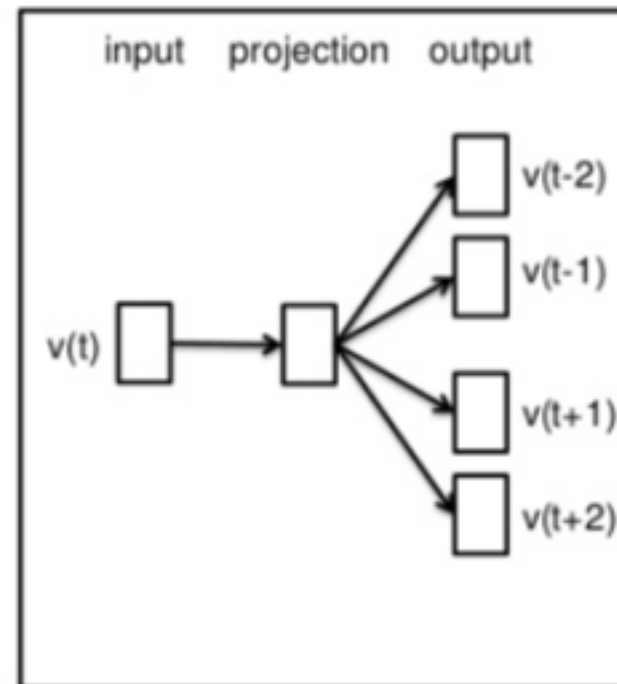
Word Embeddings

CBoW



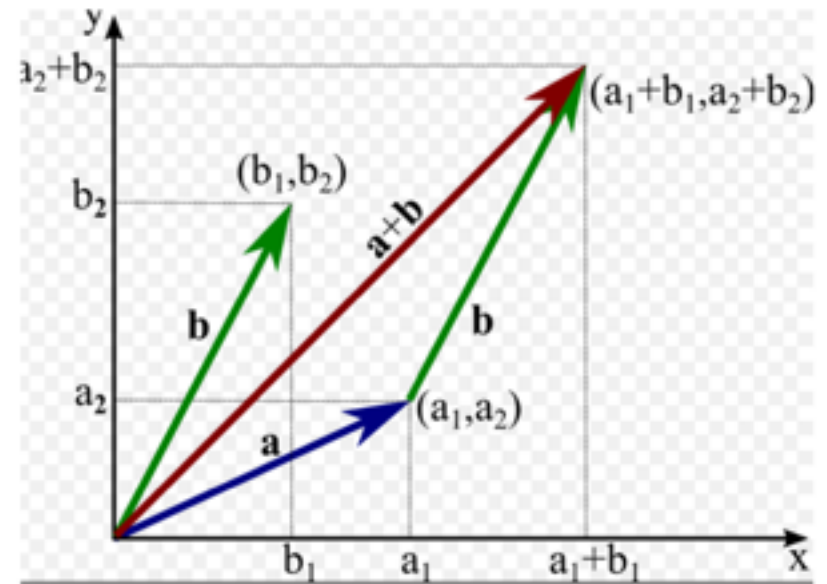
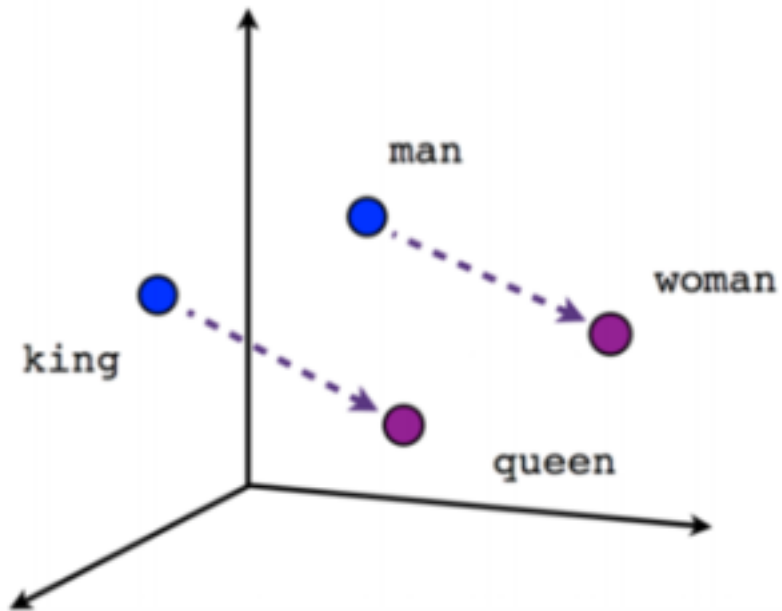
- given context words
- predict a probability of a target word

Skip-gram



- given a target word
- predict a probability of context words

Word Embeddings: Analogies



Topic Modelling

Topics

gene 0.04
dna 0.02
genetic 0.01
...

life 0.02
evolve 0.01
organism 0.01
...

brain 0.04
neuron 0.02
nerve 0.01
...

data 0.02
number 0.02
computer 0.01
...

Documents

Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK—How many **genes** does an **organism** need to **survive**? Last week at the genome meeting here,* two genome researchers with radically different approaches presented complementary views of the basic genes needed for **life**. One research team, using **comparative** analyses to compare known **genomes**, concluded that today's **organisms** can be sustained with just 250 genes, and that the earliest life forms required a mere 125 **genes**. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

Although the numbers don't match precisely, these **predictions**

"are not all that far apart," especially in comparison to the 75,000 **genes** in the human genome, notes Siv Andersson, a geneticist at Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a **numbers** game. "Particularly, more and more **genomes** are being sequenced and sequenced. "It may be a way of organizing any newly **sequenced genomes**," explains Arcady Mushegian, a **computational** molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing

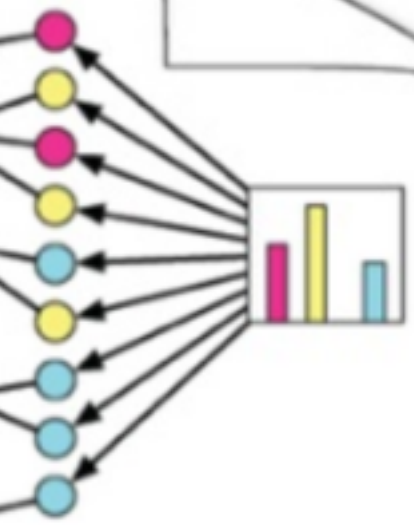


Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

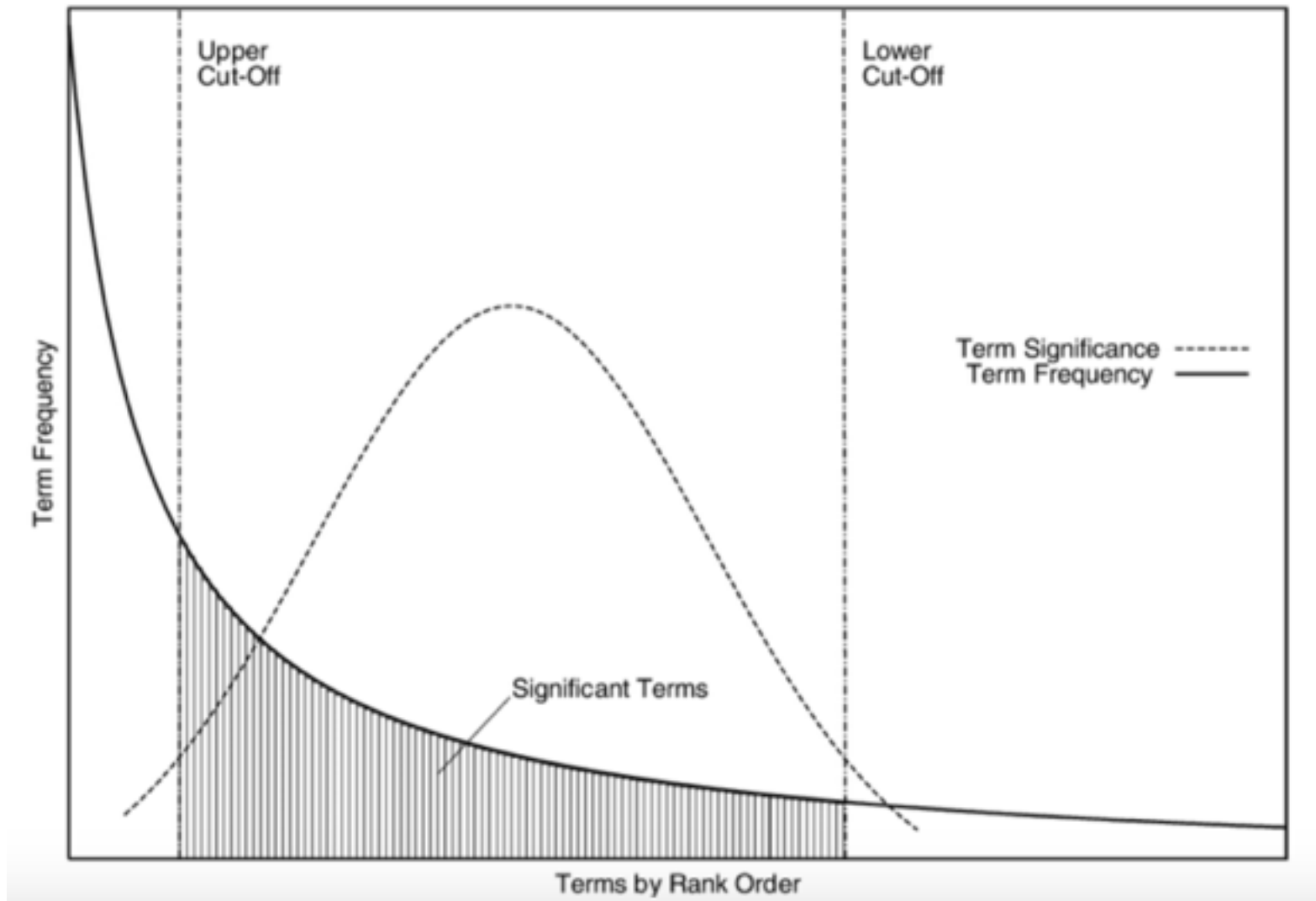
* Genome Mapping and Sequencing. Cold Spring Harbor, New York, May 8 to 12.

SCIENCE • VOL. 272 • 24 MAY 1996

Topic proportions and assignments



Zipf Distribution



Intertextuality

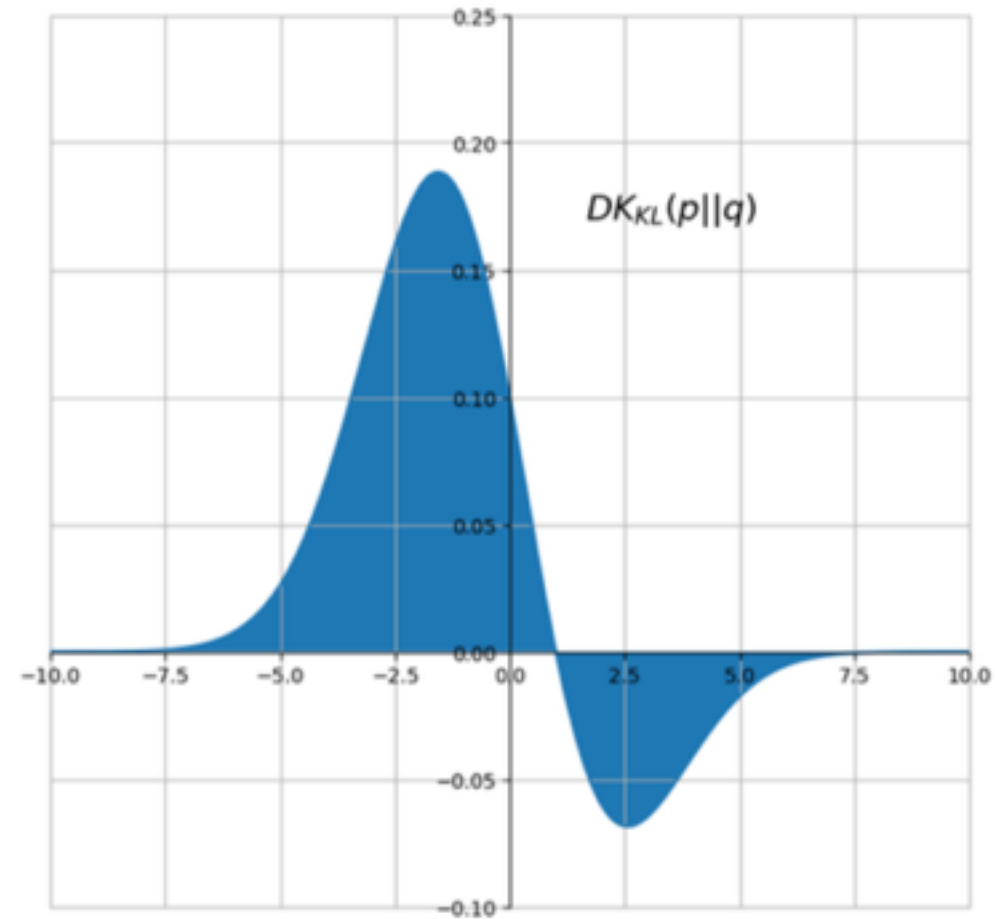
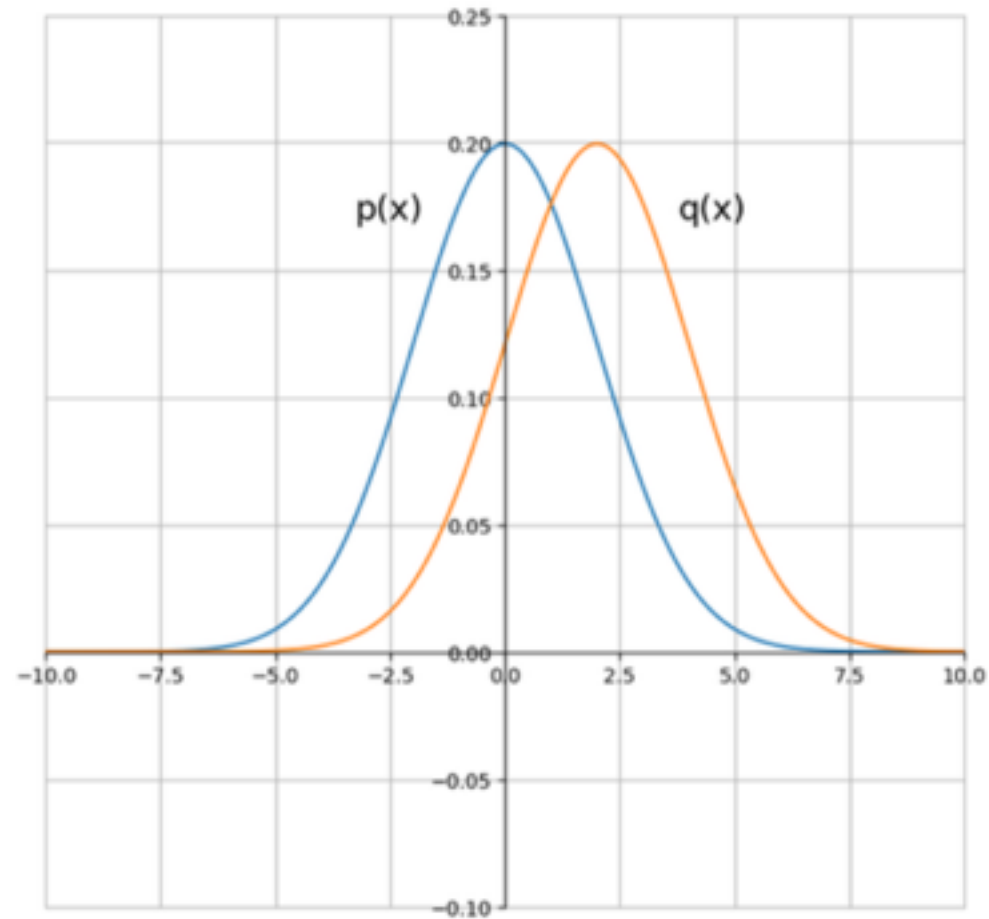
Demarcation of a passage

What is an allusion/marker?

-> can we provide a quantitative assessment?

How remarkable are certain influences/citations given language change?

Local and Global Similarity



Literal citation:

Ps 64:4a:
1QH^a V:13 [XIII:15]:

אֲשֶׁר שָׁנְנוּ כִּחְרַב לְשׁוֹנָם.
וּתְצַל נַפְשׁ עֲנִי בַמַּעוֹן אֲרִיּוֹת אֲשֶׁר שָׁנְנוּ כִּחְרַב לְשׁוֹנָם.

Free citation:

Ps 69:22:
1QH^a IV[XII]:11–12:

וַיִּתְּנוּ בְּבִרְוֹתַי רֹאשׁ וְלִצְמֵאֵי יִשְׁקוּנִי חֶמֶץ.
וַיַּעֲצֹרוּ מִשְׁקָה דַעַת מִצְמָאִים וְלִצְמָאֵם יִשְׁקוּם חוֹמֶץ.

Local and Global Similarity

מִשָּׂא בְּהִמּוֹת נֶגֶב בְּאֶרֶץ צָרָה וְצוּקָה לְבִיא וְלִישׁ מֵהֶם אֲפֶסֶה וְשָׂרָף מְעוֹפֵף (Isa 30:6)

The burden of the beasts of the South. Through the land of trouble and anguish, from whence come the lioness and the lion, the **viper** and flying serpent.

בִּיצֵי צְפֹעוֹנֵי בְּקִעוֹ, וְקוּרֵי עֶכְבִּישׁ יִאָּרְגוּ. הָאֵכֵל מִבֵּיצֵיהֶם יָמוּת וְהַזּוּרָה תִּבְקַע אֲפֶסֶה. (Isa 59:5)

They hatch basilisks' eggs, and weave the spider's web; he that eateth of their eggs dieth, and that which is crushed breaketh out into a **viper**.

Local and Global Similarity

אֶפְעָה (Job 20:16) רֹאשׁ-פְּתָנִים יִינֶק תִּהְרָגְהוּ לְשׁוֹן

He shall suck the poison of asps; the **viper**'s tongue shall slay him.

Local and Global Similarity

- IIB1b For children have come up to the *throes* of death
(10) and the *mother* of a boy is distressed in her pains.
- IIB1c For in the *throes* of death she gives birth to a son,
and in the pains of Sheol bursts forth (11) from the crucible of the
mother a wonderful counsellor with his might,
and a boy is delivered from the *throes*.

הרית גבר

- IIB2a In his *mother* come quickly all (12) the *throes*
and PAINS OF AGONY²⁵ in their birth
and TERROR to their *mothers*;
- IIB2(b) and in his birth all PANGS COME SUDDENLY
- IIB2(b) (13) IN THE CRUCIBLE²⁶ of the *mother*
- IIB2c And the *mother* of VANITY for AGONISING PAIN
and the *throes* of the pit for all WORKS of TERROR.

הרית אפעה