# SHIFTING SANDS

## TEMPORAL CONTEXT SHIFT IN WORDS

Sep 24, 2024

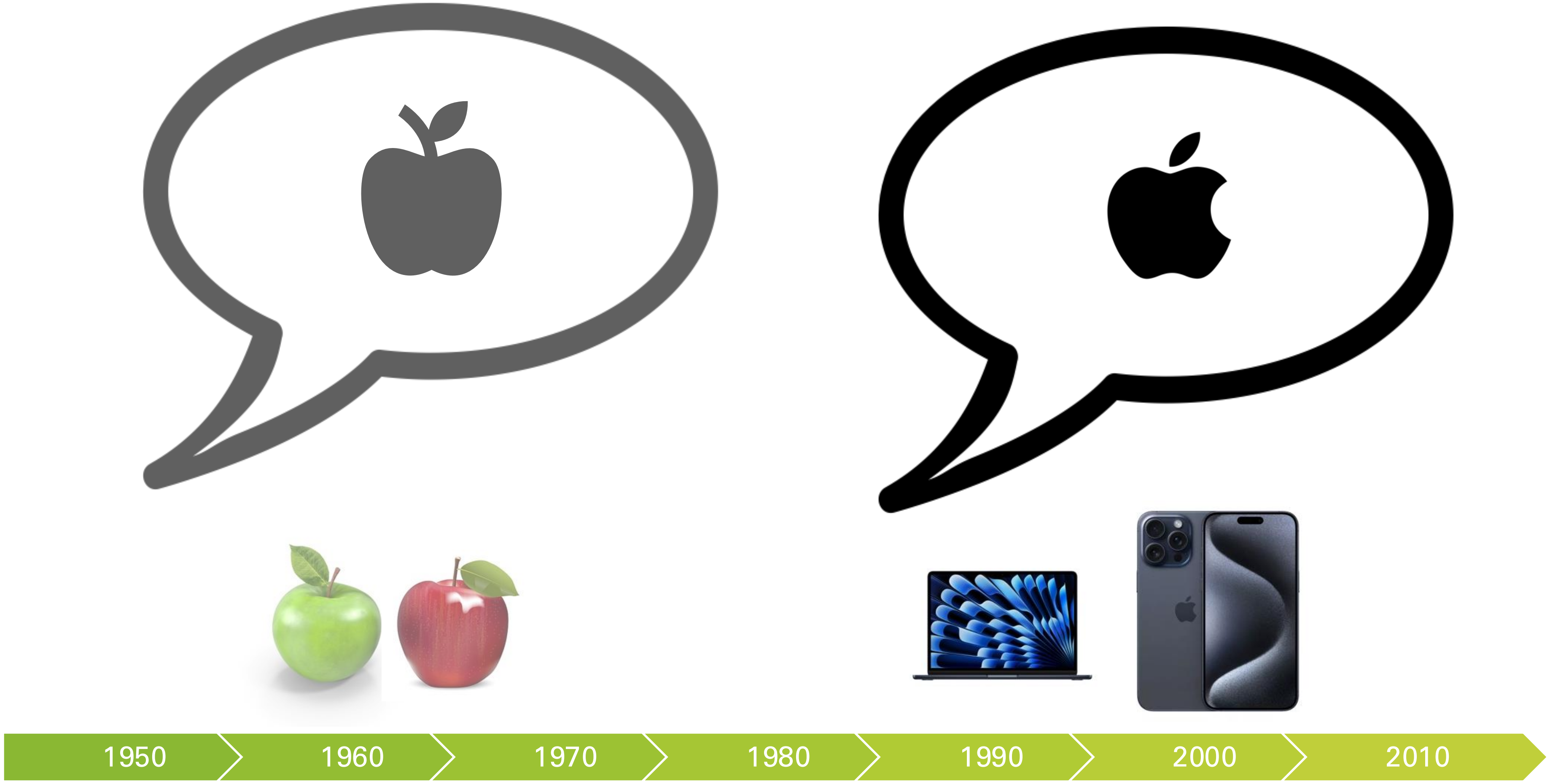
Manisha Keim
Ph.D. Computer Science
*Advisor: Prof. Rishab Nithyanand*

# CONTEXT AROUND A WORD EVOLVES WITH TIME.

"You shall know a word by the company it keeps."

(J. R. Firth)

# Context around a word evolves over time.



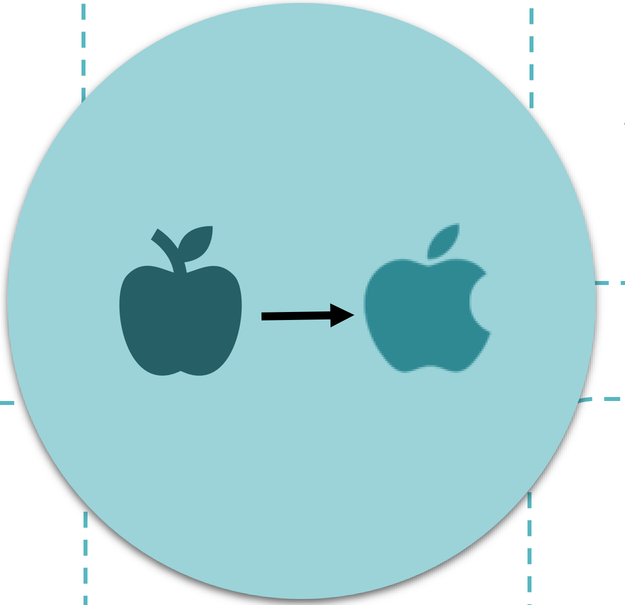
# Words can change meaning in many ways.

**Semantic**

From any kind of fruit to a specific fruit, and then to a tech company.

**Syntactic**

Evolved from a common noun to proper noun (e.g., “Apple Inc.”).



**Broadening**

Expanded to include the meaning of a technology company.

**Narrowing**

Shifted from referring to any fruit to specifically the fruit of the apple tree.

# Words can change meaning in many ways.

## Amelioration

- More positive meaning over time.
- Example: *Knight (cniht)*

Servant

Chivalry & honor

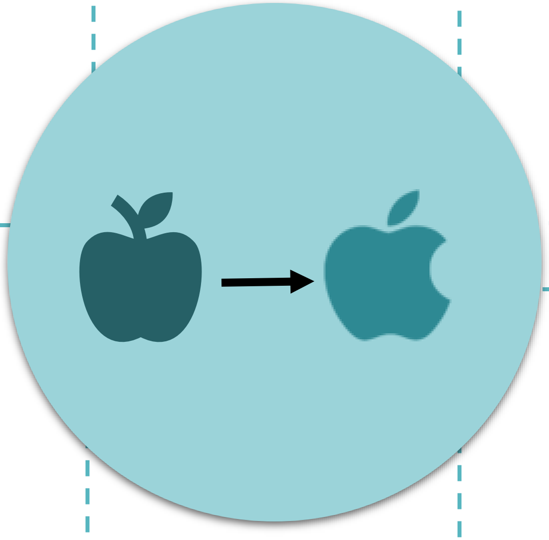
## Pejoration

- More negative meaning over time.
- Example: *Villain (villein)*

Farm laborer

Evil person

# Words can change meaning in many ways.



### Semantic

From any kind of fruit to a specific fruit, and then to a tech company.

### Syntactic

Evolved from a common noun to proper noun (e.g., “Apple Inc.”).

### Broadening

Expanded to include the meaning of a technology company.

### Narrowing

Shifted from referring to any fruit to specifically the fruit of the apple tree.

### Amelioration

*Knight* : From servant or attendant to a noble warrior or honorary title.

### Pejoration

*Villain* : From a farm laborer to an evil character.

Quantifying **linguistic evolution** and identifying mechanisms for detection**.**

# 1.

Statistically Significant Detection of Linguistic Change

Vivek Kulkarni
Rami Al-Rfou
Bryan Perozzi
Steven Skiena

***WWW 2015***

# 2.

Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change

William L. Hamilton
Jure Leskovec
Dan Jurafsky

***ACL 2016***

*Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. (2015) Statistically Significant Detection of Linguistic Change.
**William L. Hamilton, Jure Leskovec, and Dan Jurafsky (2016) Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change

Analyzing **linguistic change** through **statistical methods** and **time series** in Kulkarni et al. (2015).

# Statistically Significant Detection of Linguistic Change


Vivek Kulkarni
Stony Brook University, USA
vvkulkarni@cs.stonybrook.edu

Rami Al-Rfou
Stony Brook University, USA
ralrfou@cs.stonybrook.edu

Bryan Perozzi
Stony Brook University, USA
bperozzi@cs.stonybrook.edu

Steven Skiena
Stony Brook University, USA
skiena@cs.stonybrook.edu



## ABSTRACT

We propose a new computational approach for tracking and detecting statistically significant linguistic shifts in the meaning and usage of words. Such linguistic shifts are especially prevalent on the Internet, where the rapid exchange of ideas can quickly change a word's meaning. Our meta-analysis approach constructs property time series of word usage, and then uses statistically sound change point detection algorithms to identify significant linguistic shifts.

We consider and analyze three approaches of increasing complexity to generate such linguistic property time series, the culmination of which uses distributional characteristics inferred from word co-occurrences. Using recently proposed deep neural language models, we first train vector representations of words for each time period. Second, we warp the vector spaces into one unified coordinate system. Finally, we construct a distance-based distributional time series for each word to track it's linguistic displacement over time.

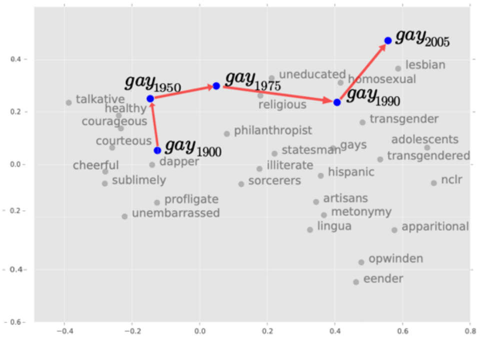




Figure 1: A 2-dimensional projection of the latent semantic space captured by our algorithm. Notice the semantic trajectory of the word **gay** transitioning meaning in the space.

# Key approaches for tracking language evolution.

## Quantify the changes in a word's usage over time.

Creating time series for each word.

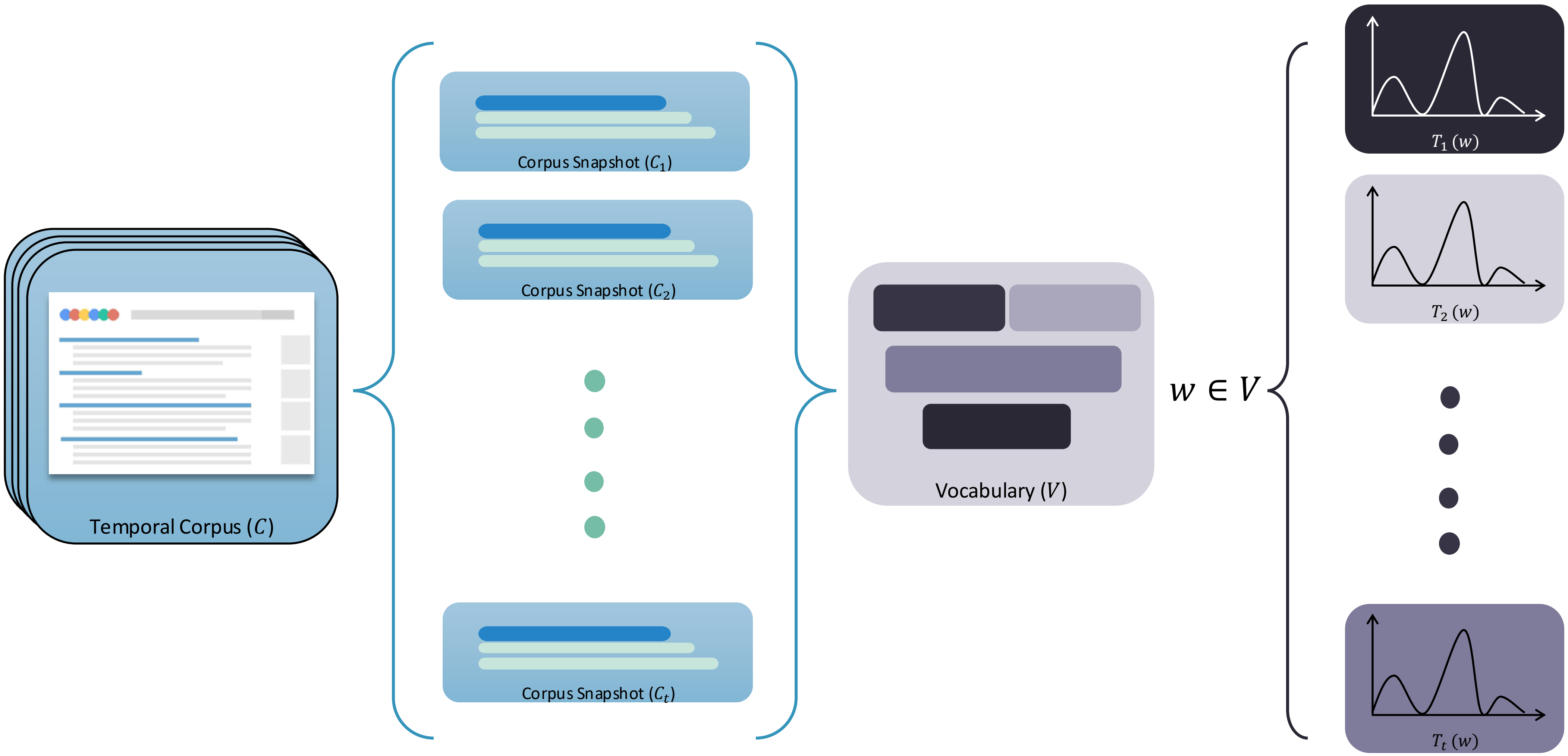
## Detection of precise moment of semantic shift.

Creating algorithm for finding significant shifts.



Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. (2015) Statistically Significant Detection of Linguistic Change.

# Divide and conquer approach for large-scale time series analysis.






Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. (2015) Statistically Significant Detection of Linguistic Change.

# There are three fundamental ways to use time series for measuring shifts.

**Distributional**

**Syntactic**

**Frequency**



Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. (2015) Statistically Significant Detection of Linguistic Change.

# Analyzing variations in a word’s surrounding context to observe shifts in its semantic meaning.

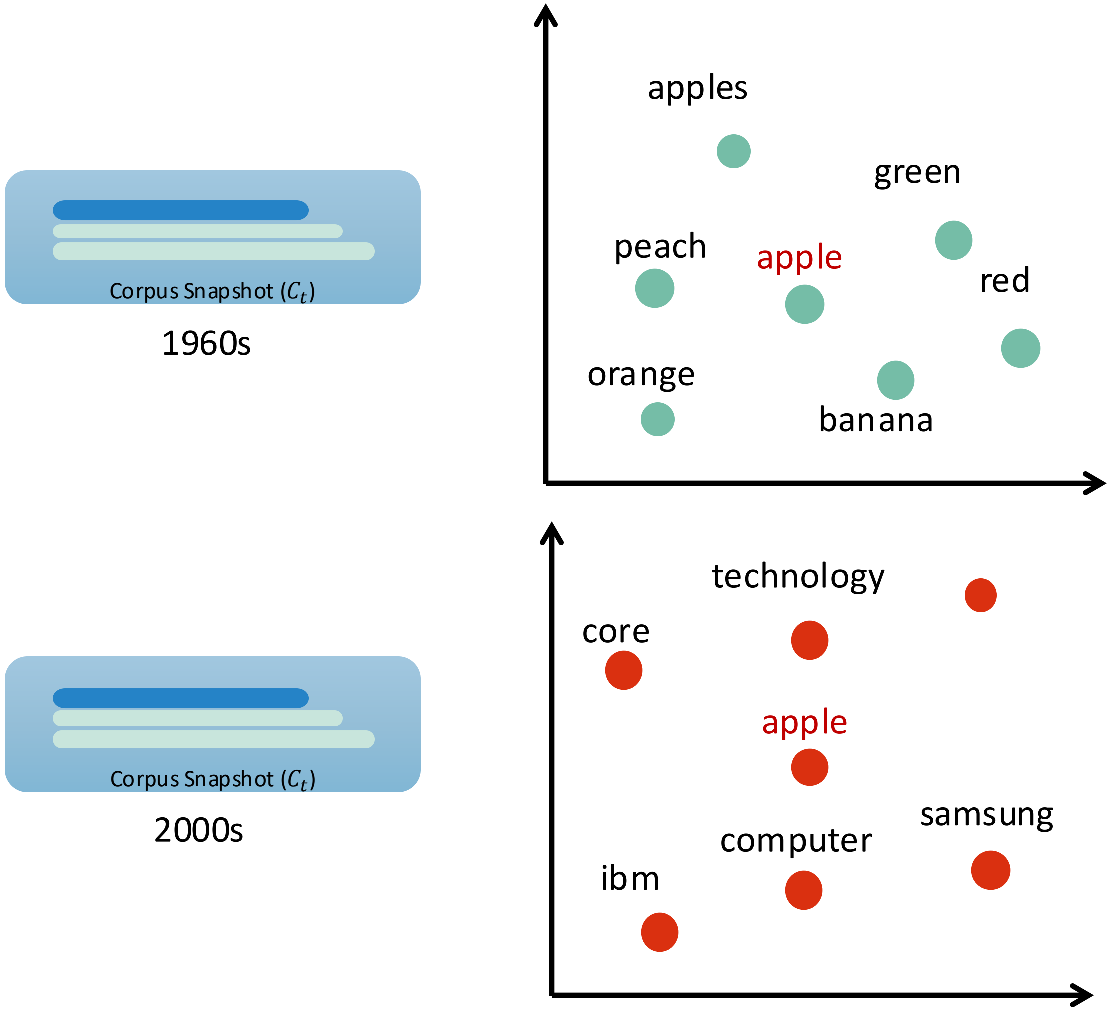
- Captures changes in word meaning based on surrounding words.
- Utilizes word embeddings to model context.
- E.g.: tape, gay, diet, sandy

"You shall know a word by the company it keeps." (J. R. Firth)



Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. (2015) Statistically Significant Detection of Linguistic Change.

# Analyzing variations in a word's surrounding context to observe shifts in its semantic meaning.

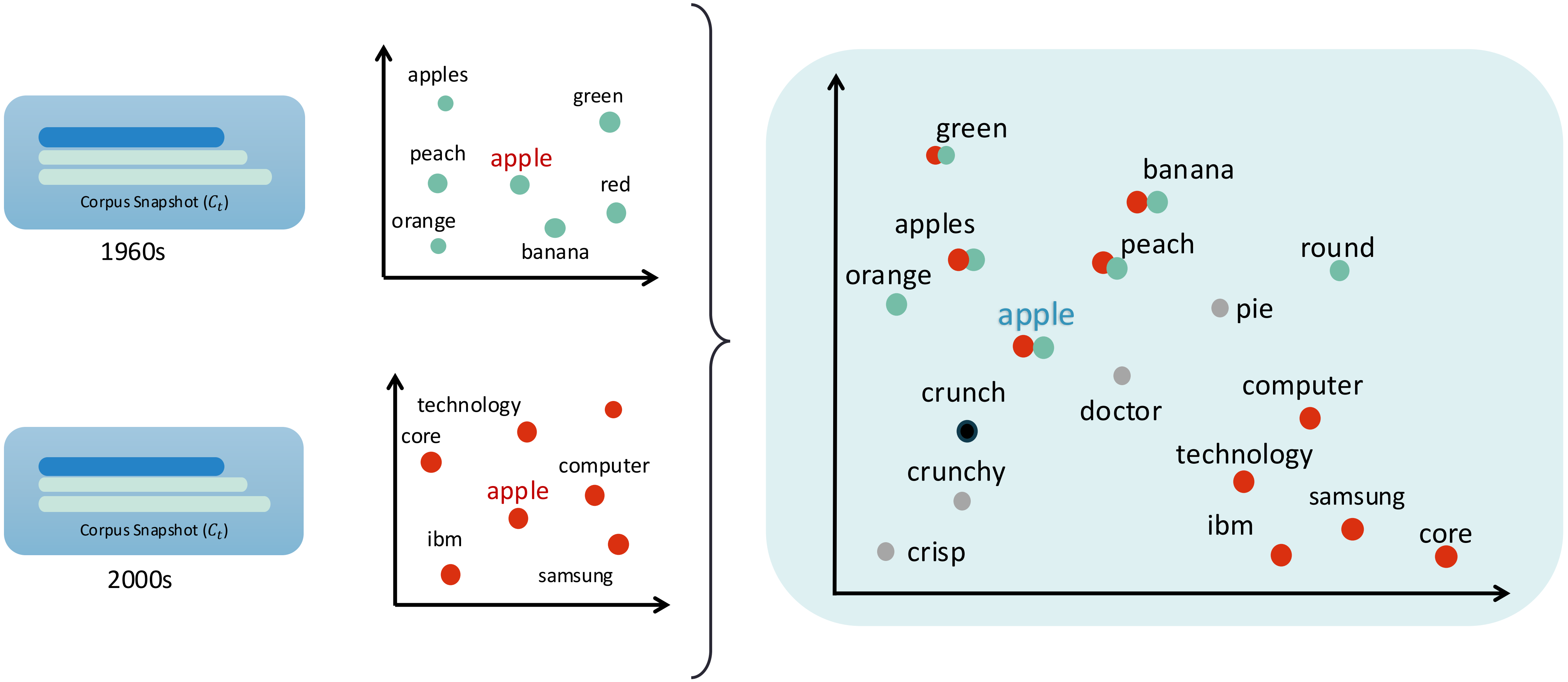



- Captures changes in word meaning based on surrounding words.
- Utilizes word embeddings to model context.
- E.g.: tape, gay, diet, sandy



Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. (2015) Statistically Significant Detection of Linguistic Change.

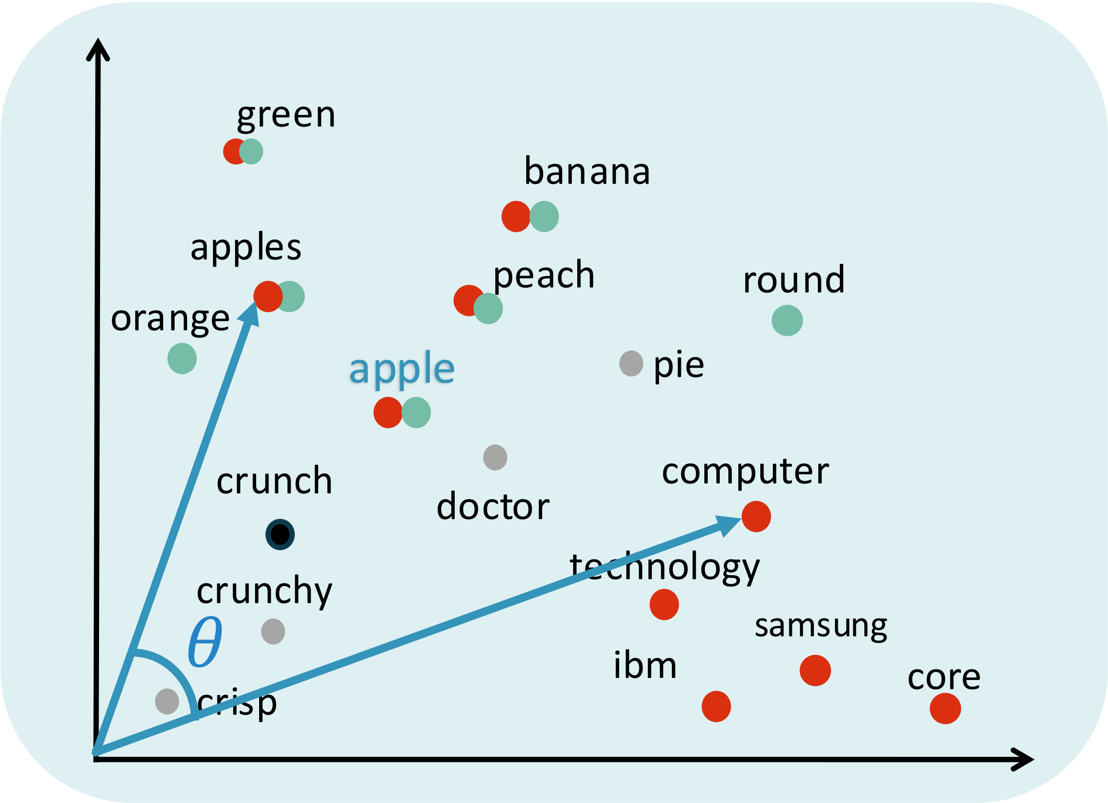
# Aligning word embeddings enables comparison of a word's meaning over time.






Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. (2015) Statistically Significant Detection of Linguistic Change.

# Aligning word embeddings enables comparison of a word’s meaning over time.

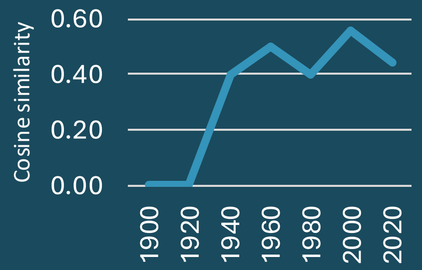



Transforming embeddings from different time snapshots into a common coordinate system, enabling comparison of a word’s meaning over time.

# There are three fundamental ways to use time series for measuring shifts.

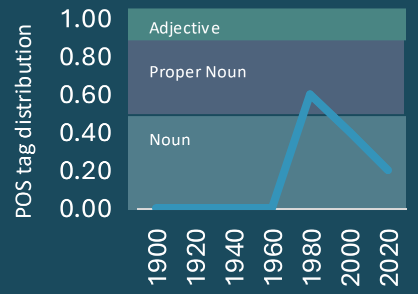
## Distributional

Cosine Similarity



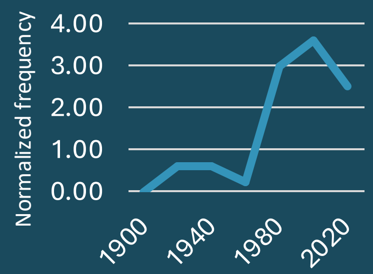

## Syntactic

Part of Speech




## Frequency

Raw count






Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. (2015) Statistically Significant Detection of Linguistic Change.

# Distributional method is the most effective method for capturing semantic shifts.

**Frequency**

Raw count

Sampling errors due to domain or genre bias in corpus.

**Syntactic**

Part of Speech

Semantic shifts are not restricted to POS.

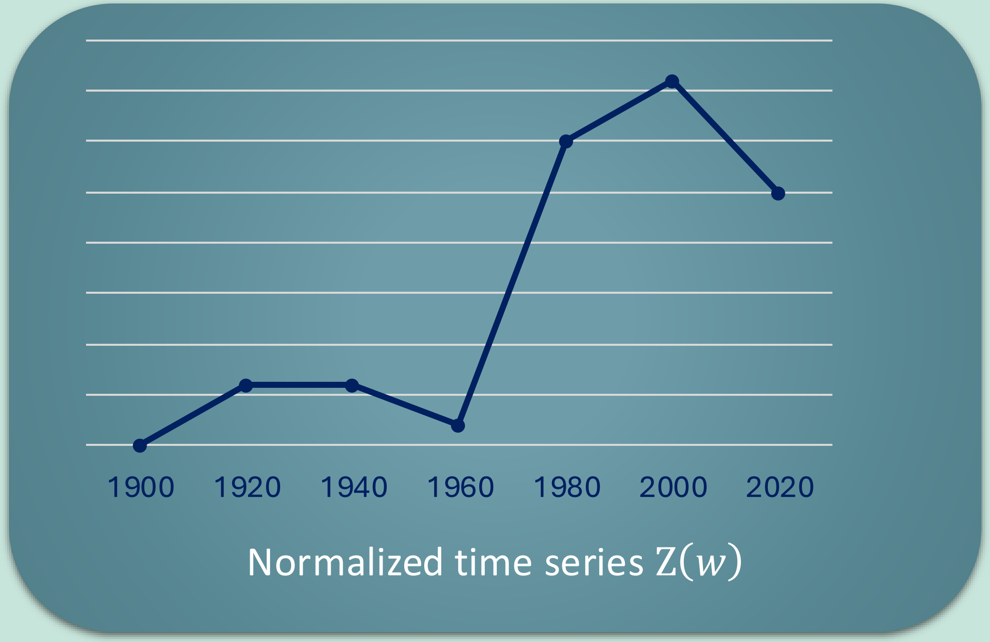
**Distributional**

Cosine Similarity

Sensitive to the quality and size of the corpus.

Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. (2015) Statistically Significant Detection of Linguistic Change.

# Identifying significant shifts through change point algorithm.

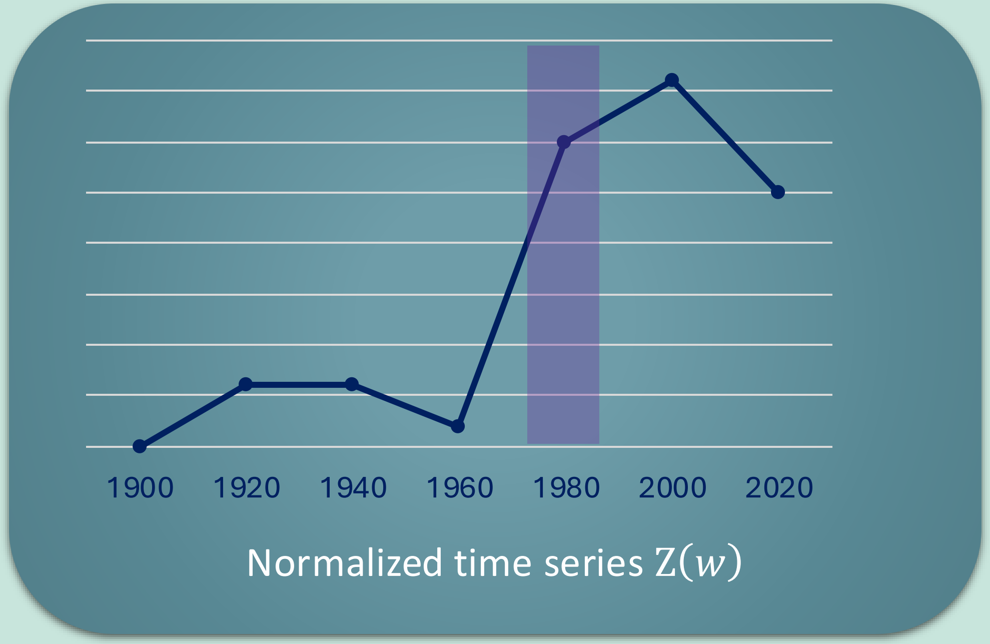



Normalization

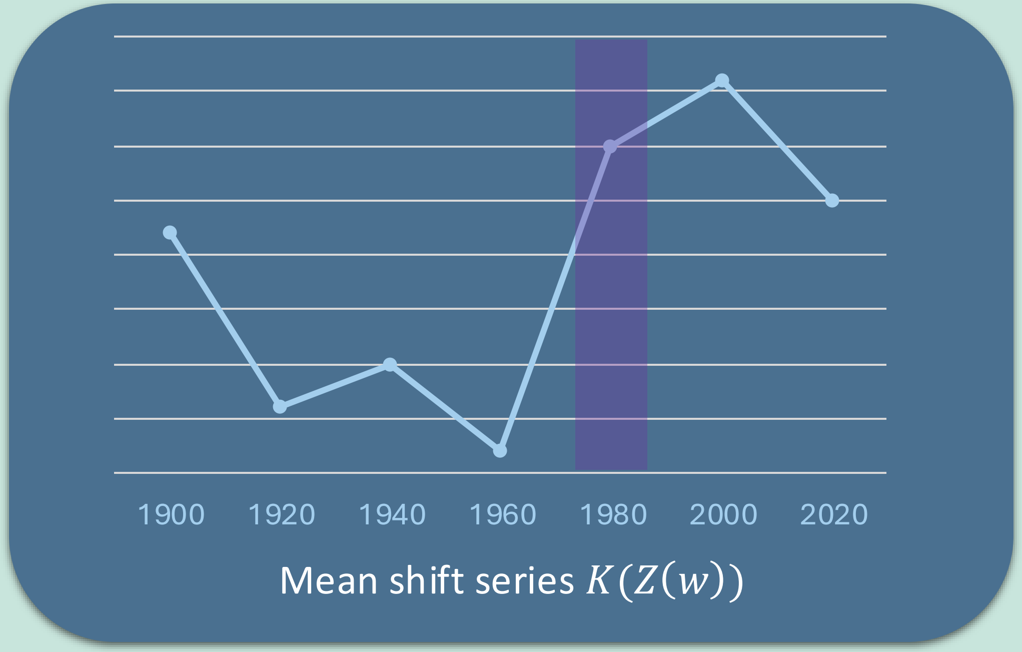


Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. (2015) Statistically Significant Detection of Linguistic Change.

# Identifying significant shifts through change point algorithm.




Normalization

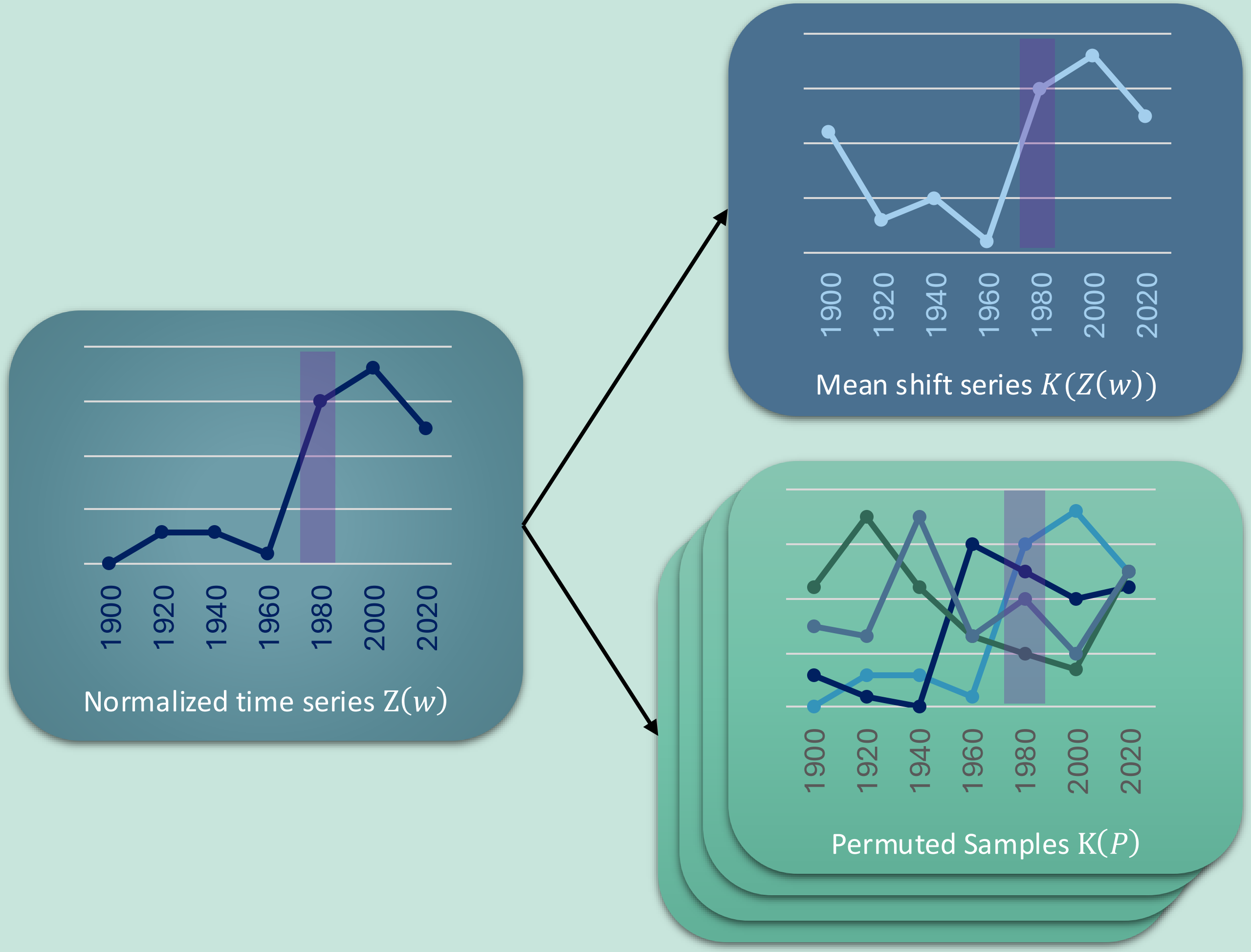



Mean Shift



Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. (2015) Statistically Significant Detection of Linguistic Change.

# Identifying significant shifts through change point algorithm.



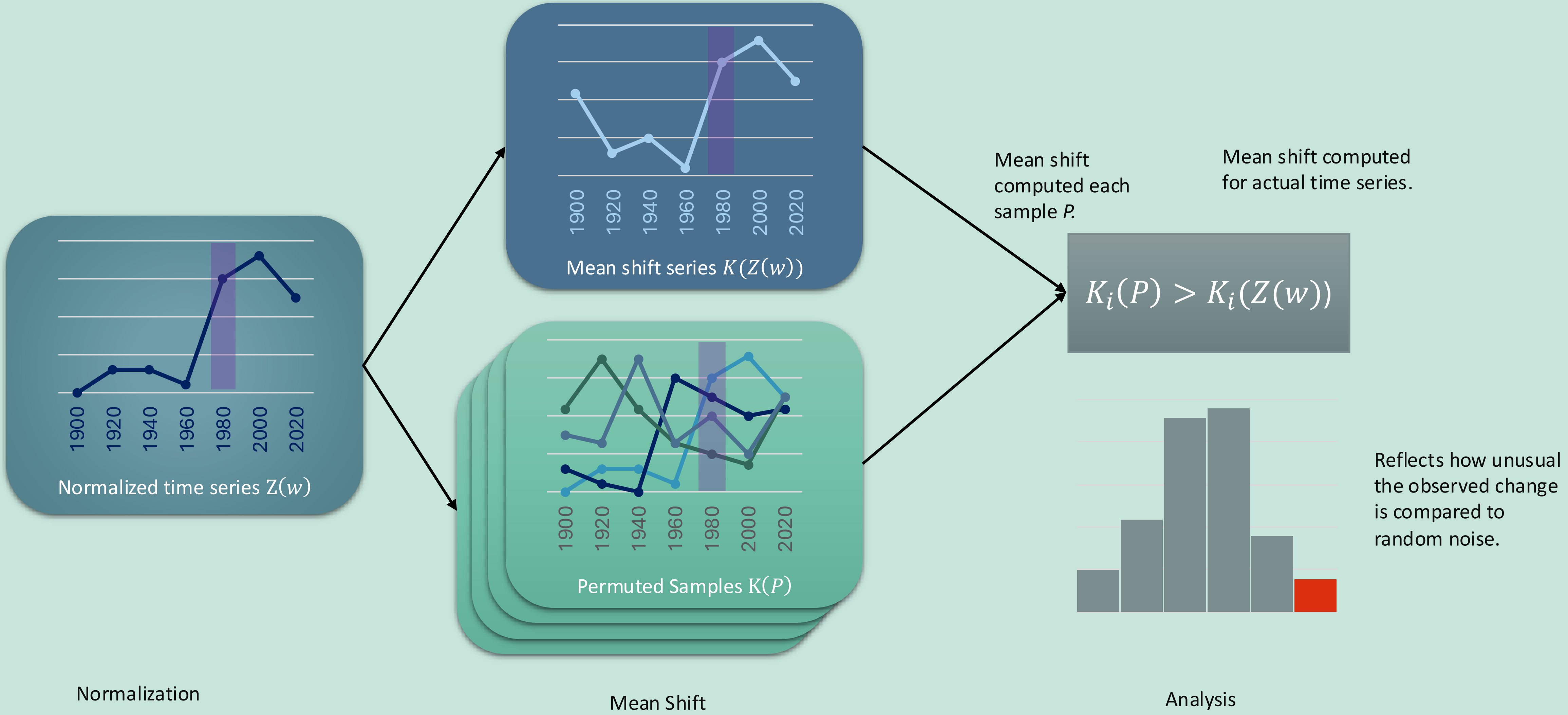

Normalization

Mean Shift



Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. (2015) Statistically Significant Detection of Linguistic Change.

# Identifying significant shifts through change point algorithm.






Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. (2015) Statistically Significant Detection of Linguistic Change.

# Takeaways.

- Use of word embeddings to measure **context** of a word.
- Applying change point algorithm for identifying **substantial shifts** in word meanings.



Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. (2015) Statistically Significant Detection of Linguistic Change.

# Quantifying **semantic change** through **Diachronic** word embeddings in Hamilton et al. (2016).

## Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change

**William L. Hamilton, Jure Leskovec, Dan Jurafsky**
Department of Computer Science, Stanford University, Stanford CA, 94305
wleif,jure,jurafsky@stanford.edu

**Abstract**

Understanding how words change their meanings over time is key to models of language and cultural evolution, but historical data on meaning is scarce, making theories hard to develop and test. Word embeddings show promise as a diachronic tool, but have not been carefully evaluated. We develop a robust methodology for quantifying semantic change by evaluating word embeddings (PPMI, SVD, word2vec) against known historical changes. We then use this methodology to reveal statistical laws of semantic evolution. Using six historical corpora spanning four languages and two centuries, we propose two quantitative laws of seman-

But many core questions about semantic change remain unanswered. One is the role of *frequency*. Frequency plays a key role in other linguistic changes, associated sometimes with faster change—sound changes like lenition occur in more frequent words—and sometimes with slower change—high frequency words are more resistant to morphological regularization (Bybee, 2007; Pagel et al., 2007; Lieberman et al., 2007). What is the role of word frequency in meaning change?

Another unanswered question is the relationship between semantic change and *polysemy*. Words gain senses over time as they semantically drift (Bréal, 1897; Wilkins, 1993; Hopper and Traugott, 2003), and polysemous words[1] occur in more diverse contexts, affecting lexical access speed (Adelman et al., 2006) and rates of L2 learning (Crossley et al., 2010). But we don't

# **Frequency** and **polysemy** drive the evolution of word meanings.

Role of word frequency in stability of word meanings.

*Law of Conformity*

Relation of polysemy to semantic change.

*Law of Innovation*



William L. Hamilton, Jure Leskovec, and Dan Jurafsky (2016) Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change

# There are three ways for capturing **word co-occurrence**.

**Skip-gram with negative sampling (SGNS)**

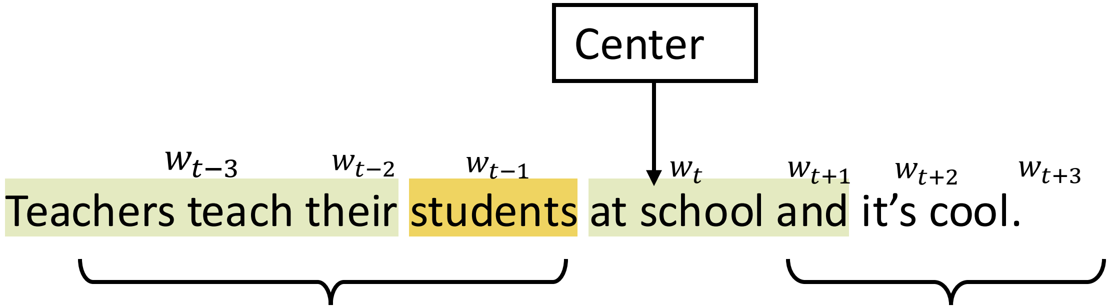
**Positive point-wise mutual information (PPMI)**

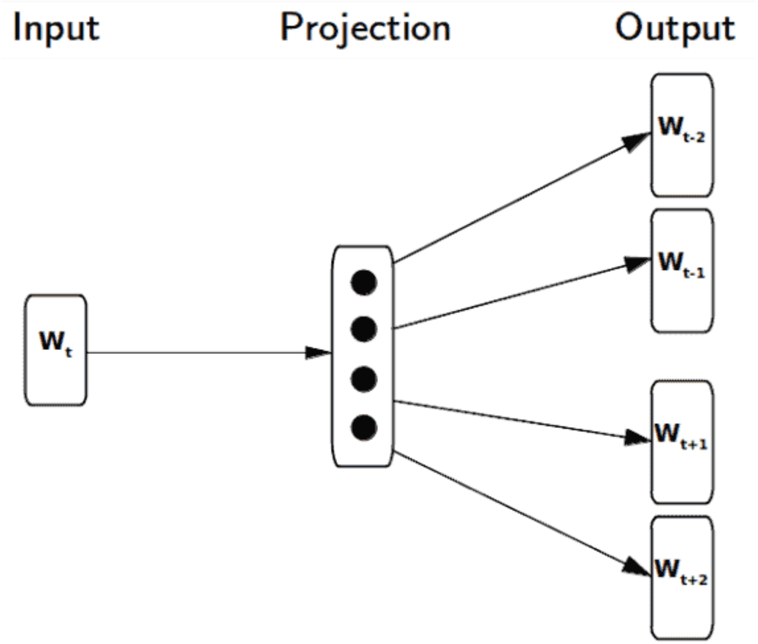
**Singular Value Decomposition (SVD)**

William L. Hamilton, Jure Leskovec, and Dan Jurafsky (2016) Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change

# Unlocking the hidden language of words with skip-gram.

- Skip-gram model is designed to predict the context given a word.




William L. Hamilton, Jure Leskovec, and Dan Jurafsky (2016) Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change

# Unlocking the hidden language of words with skip-gram.

- Skip-gram model is designed to predict the context given a word.
- The model tries to maximize the probability of these context words given the target word, while also minimizing the probability of randomly sampled negative words that do not appear in the context.

**Positive samples**

(students, teacher)

(students, school)

**Negative samples**

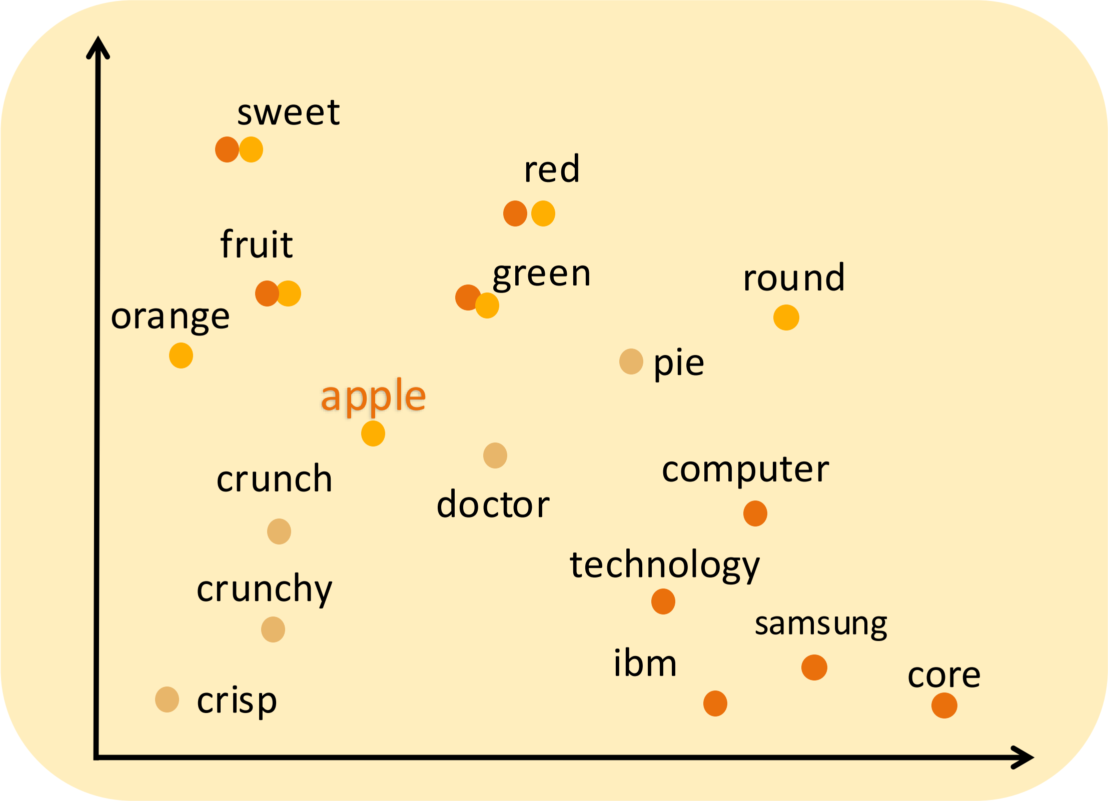
(students, car)

(students, banana)



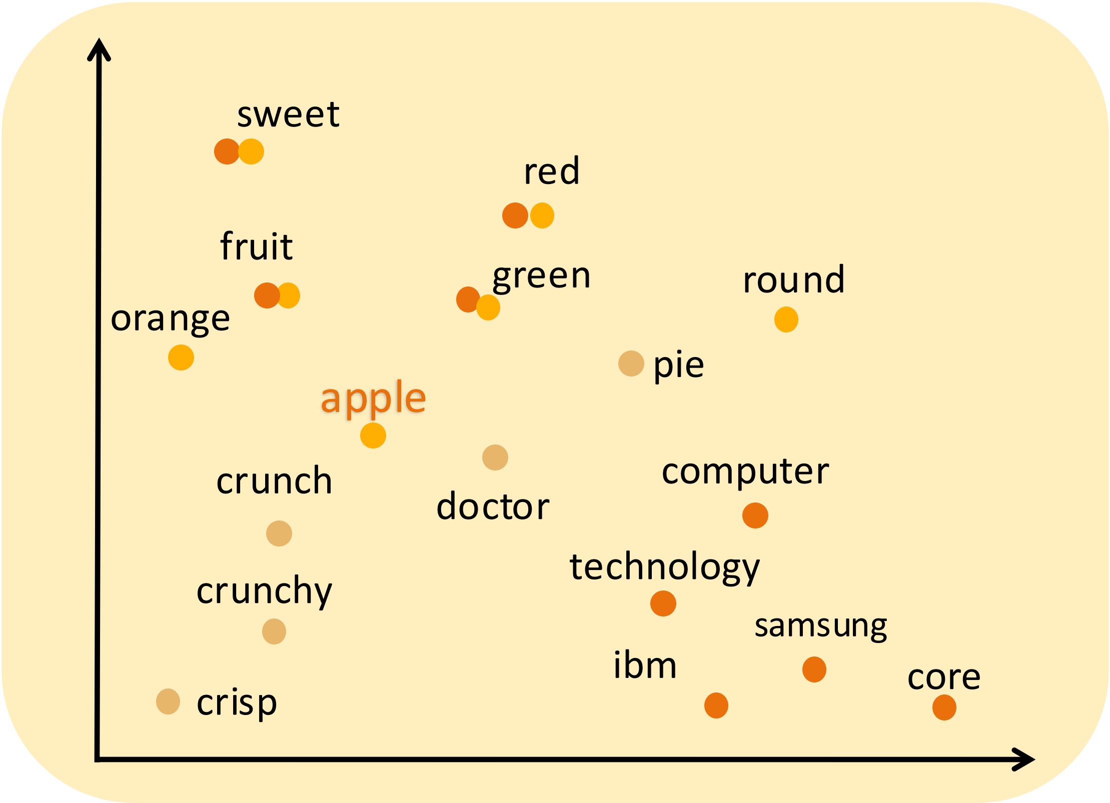
William L. Hamilton, Jure Leskovec, and Dan Jurafsky (2016) Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change

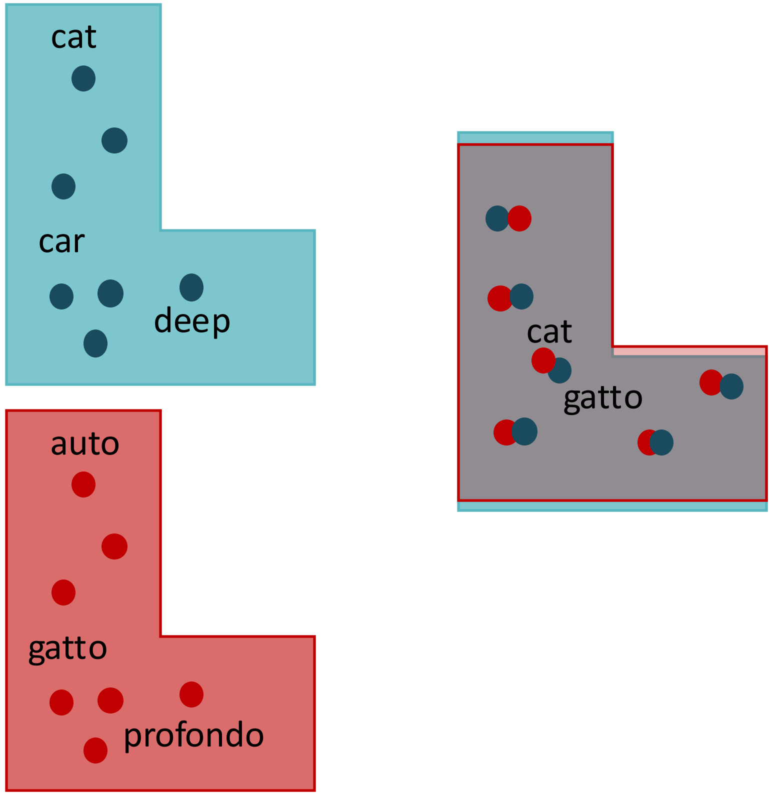
# Aligning word embeddings enables comparison of a word's meaning over time.




Transforming embeddings from different time snapshots into a common coordinate system, enabling comparison of a word's meaning over time.

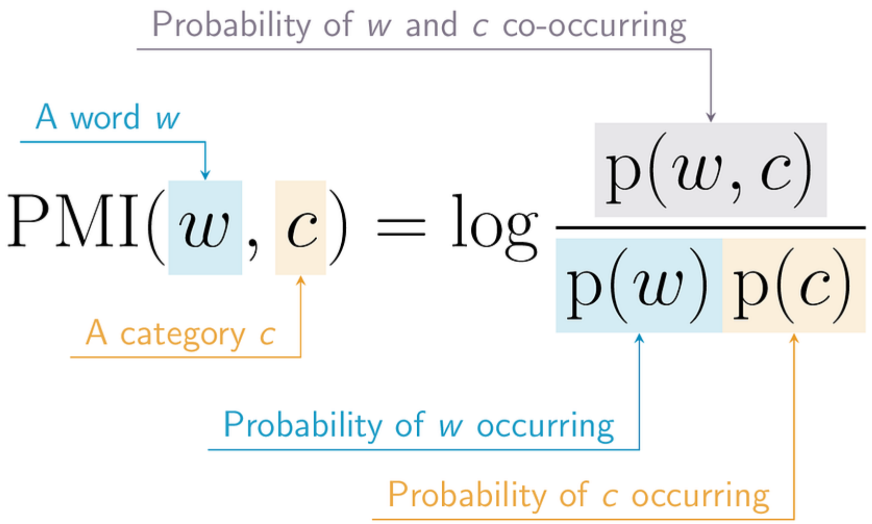
# Aligning word embeddings enables comparison of a word's meaning over time.




Transforming embeddings from different time snapshots into a common coordinate system, enabling comparison of a word's meaning over time.

# Measuring strength of word associations through PPMI.




$$\mathrm{PMI}(w, c) = \log \frac{\mathrm{p}(w, c)}{\mathrm{p}(w)\,\mathrm{p}(c)}$$

- PPMI helps to find words that frequently appear together in the same context.
- A higher PPMI value between two words suggests a stronger semantic relationship.
- The PPMI values for a target word and all context words form a high-dimensional vector.



William L. Hamilton, Jure Leskovec, and Dan Jurafsky (2016) Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change

# Measuring strength of word associations through PPMI.

| | *computer* | *data* | *pinch* | *result* | *sugar* |
|---|---|---|---|---|---|
| *appricot* | 0.00 | 0.00 | 0.05 | 0.00 | 0.05 |
| *pineapple* | 0.00 | 0.00 | 0.05 | 0.00 | 0.05 |
| *digital* | 0.11 | 0.05 | 0.00 | 0.05 | 0.00 |
| *information* | 0.05 | 0.32 | 0.00 | 0.21 | 0.01 |

$$\mathrm{PMI}(w, c) = \log \frac{\mathrm{p}(w, c)}{\mathrm{p}(w)\,\mathrm{p}(c)}$$

- A word $w$
- A category $c$
- Probability of $w$ and $c$ co-occurring: $\mathrm{p}(w, c)$
- Probability of $w$ occurring: $\mathrm{p}(w)$
- Probability of $c$ occurring: $\mathrm{p}(c)$

- PPMI helps to find words that frequently appear together in the same context.
- A higher PPMI value between two words suggests a stronger semantic relationship.
- The PPMI values for a target word and all context words form a high-dimensional vector.

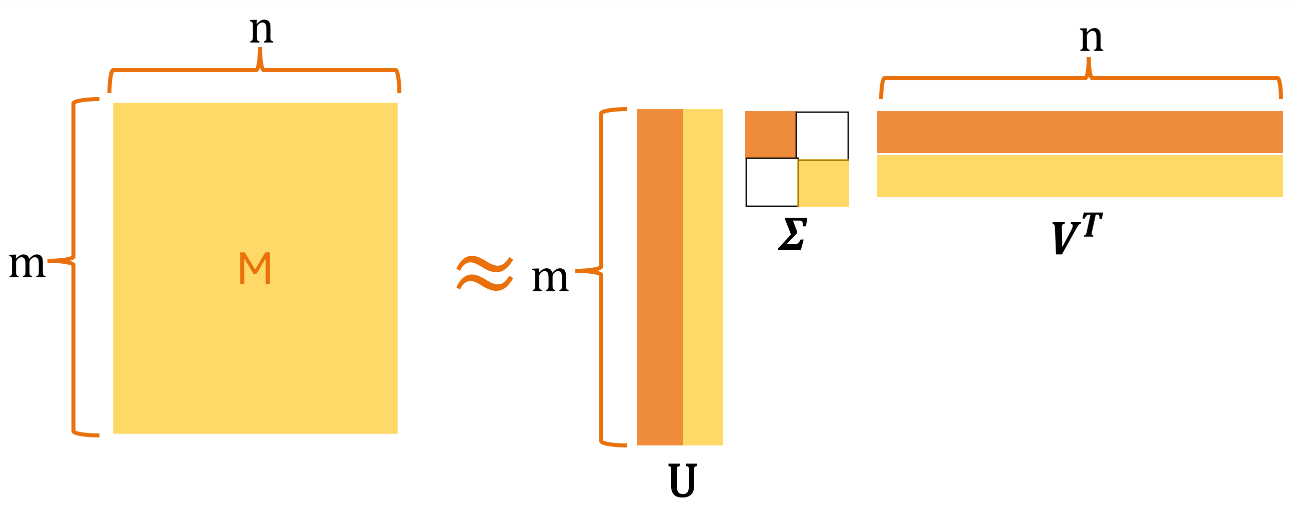


William L. Hamilton, Jure Leskovec, and Dan Jurafsky (2016) Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change

# Extracting important features from PPMI matrix using SVD.

- SVD helps to reduce the dimensionality of the data while preserving the essential structure and patterns in word co-occurrences.
- SVD decomposes the co-occurrence matrix $\mathrm{M}$ into three matrices.

# Extracting important features from PPMI matrix using SVD.




- SVD helps to reduce the dimensionality of the data while preserving the essential structure and patterns in word co-occurrences.
- SVD decomposes the co-occurrence matrix M into three matrices.

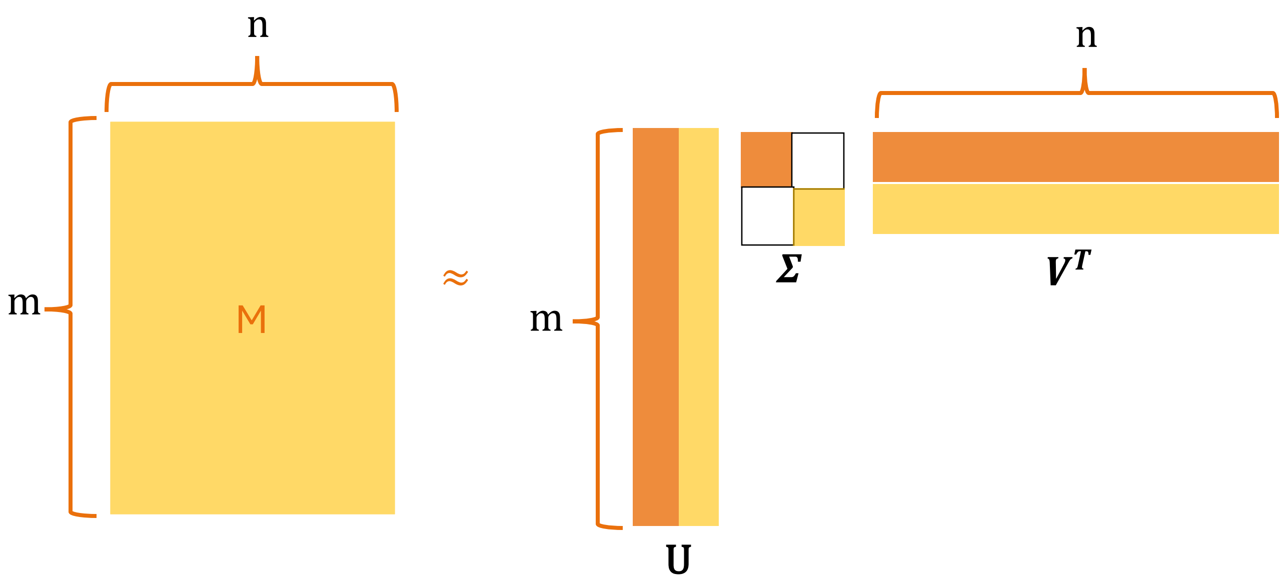


William L. Hamilton, Jure Leskovec, and Dan Jurafsky (2016) Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change

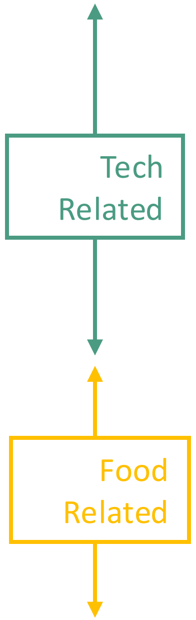
# SVD - Example

| | computer | screen | system | fruit | juice |
|---|---|---|---|---|---|
| laptop | 1 | 1 | 1 | 0 | 0 |
| information | 2 | 2 | 2 | 0 | 0 |
| software | 1 | 1 | 1 | 0 | 0 |
| data | 5 | 5 | 5 | 0 | 0 |
| apple | 0 | 0 | 0 | 2 | 2 |
| banana | 0 | 0 | 0 | 3 | 3 |
| sweet | 0 | 0 | 0 | 1 | 1 |

=

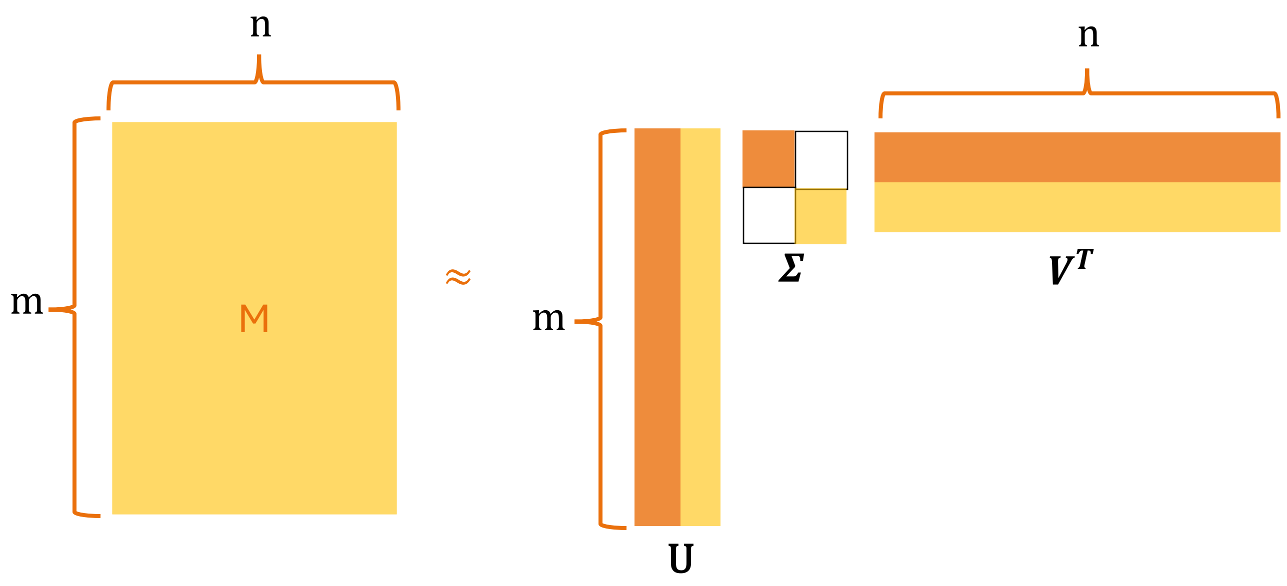


# SVD - Example

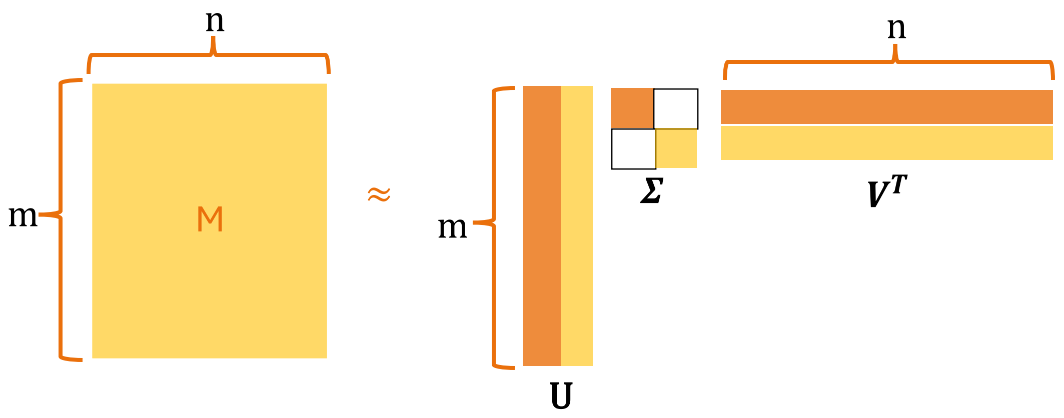



| computer | screen | system | fruit | juice |
|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 0 |
| 2 | 2 | 2 | 0 | 0 |
| 1 | 1 | 1 | 0 | 0 |
| 5 | 5 | 5 | 0 | 0 |
| 0 | 0 | 0 | 2 | 2 |
| 0 | 0 | 0 | 3 | 3 |
| 0 | 0 | 0 | 1 | 1 |

=



# SVD - Example




| | computer | screen | system | fruit | juice |
|---|---|---|---|---|---|
| Tech Related | 1 | 1 | 1 | 0 | 0 |
| | 2 | 2 | 2 | 0 | 0 |
| | 1 | 1 | 1 | 0 | 0 |
| | 5 | 5 | 5 | 0 | 0 |
| Food Related | 0 | 0 | 0 | 2 | 2 |
| | 0 | 0 | 0 | 3 | 3 |
| | 0 | 0 | 0 | 1 | 1 |

=

| Tech Related | Food Related |
|---|---|
| 0.18 | 0.02 |
| 0.36 | 0.07 |
| 0.18 | 0.09 |
| 0.90 | 0.11 |
| 0.15 | 0.53 |
| 0.07 | 0.80 |
| 0.07 | 0.27 |

**U**

Word-concept similarity matrix

×

**Strength** of Tech Related

| | |
|---|---|
| 9.64 | 0 |
| 0 | 5.29 |

$\Sigma$

×

| | computer | screen | system | fruit | juice |
|---|---|---|---|---|---|
| Tech Related | 0.53 | 0.53 | 0.53 | 0.09 | 0.09 |
| Food Related | 0.12 | 0.02 | 0.02 | 0.71 | 0.71 |

$V^T$

context-concept similarity matrix

# Skip-gram models continues to be best suited for measuring shifts.

**Positive point-wise mutual information (PPMI)**

Sparse vector
High-dimensional

**Singular Value Decomposition (SVD)**

Loss of detail

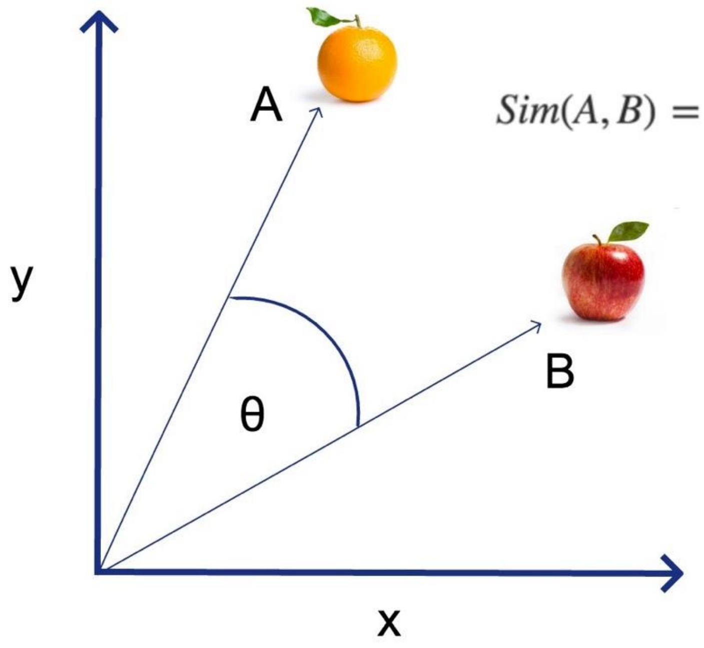
**Skip-gram with negative sampling (SGNS)**

Dense vector
Low-dimensional



William L. Hamilton, Jure Leskovec, and Dan Jurafsky (2016) Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change

# Computing **pair-wise similarity** time-series to track how the similarity between word pairs changes over time.



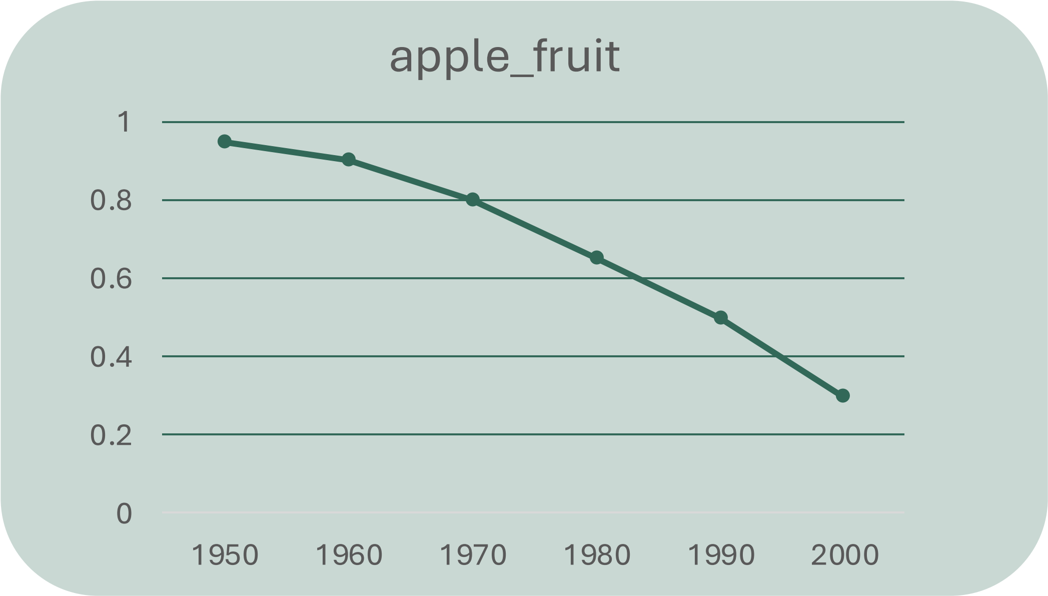

$$Sim(A,B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

- For a given pair of words, the similarity is calculated using <u>cosine similarity</u> between their word embeddings at different time points (e.g., different decades).

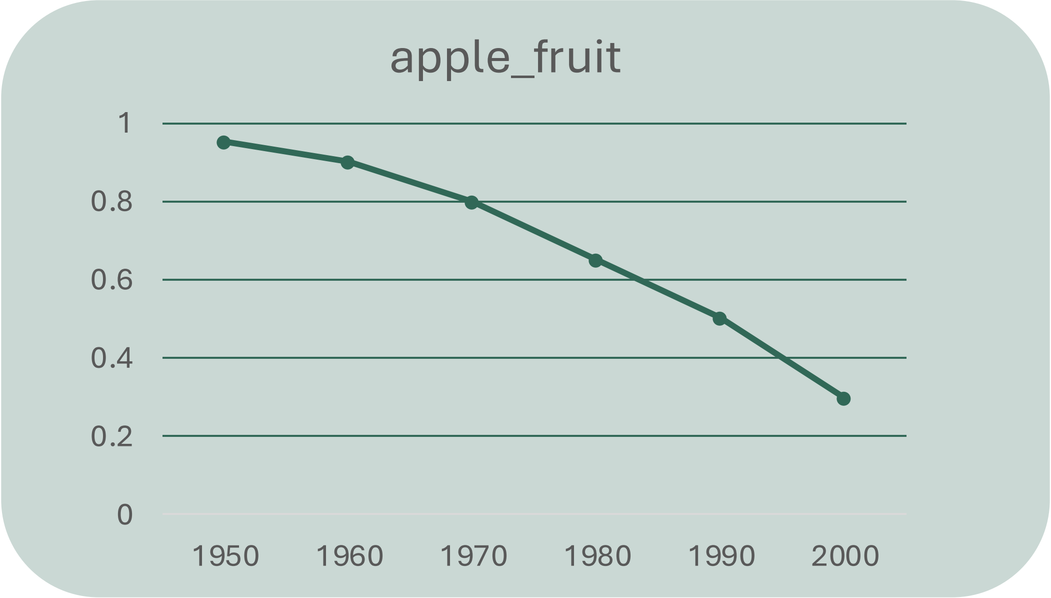


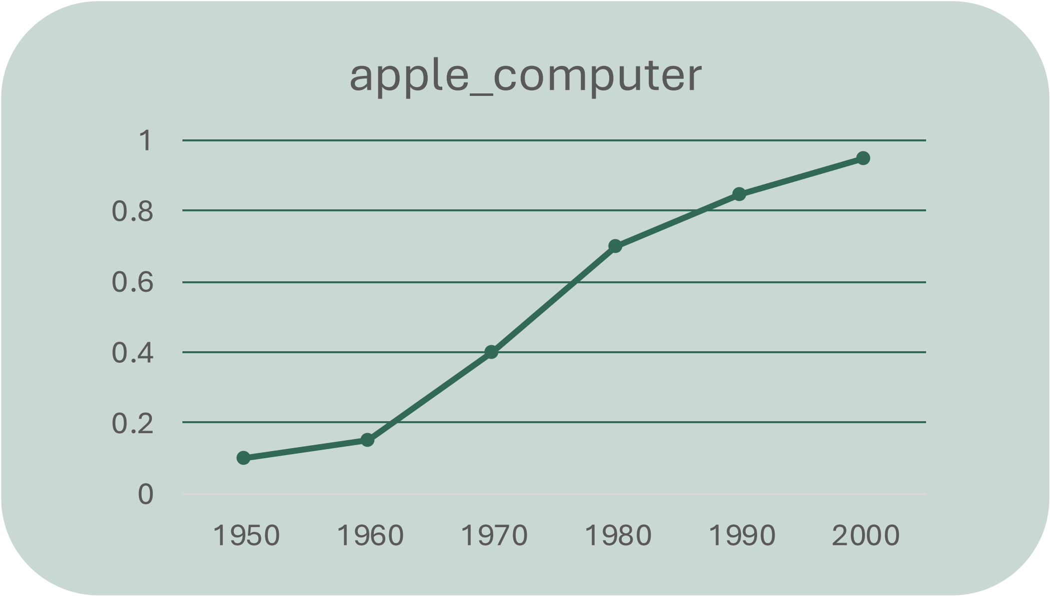
William L. Hamilton, Jure Leskovec, and Dan Jurafsky (2016) Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change

Computing **pair-wise similarity** time-series to track how the similarity between word pairs changes over time.



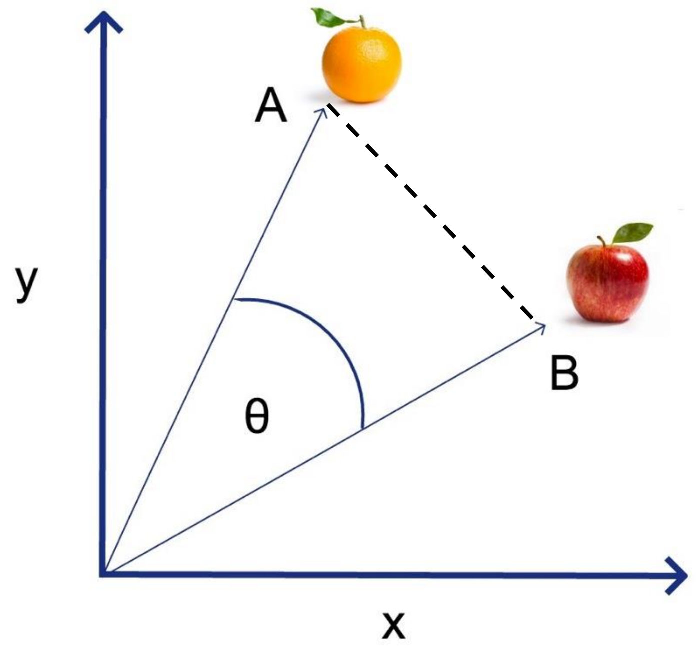
Computing **pair-wise similarity** time-series to track how the similarity between word pairs changes over time.



# Computing **semantic displacement** of a word to measure how much a word's meaning changes over time.

- Semantic displacement refers to the amount of movement or "displacement" of a word's meaning in the semantic space over time.
- After aligning the embeddings, <u>cosine distance</u> is used between the word embeddings of the same word across time periods.





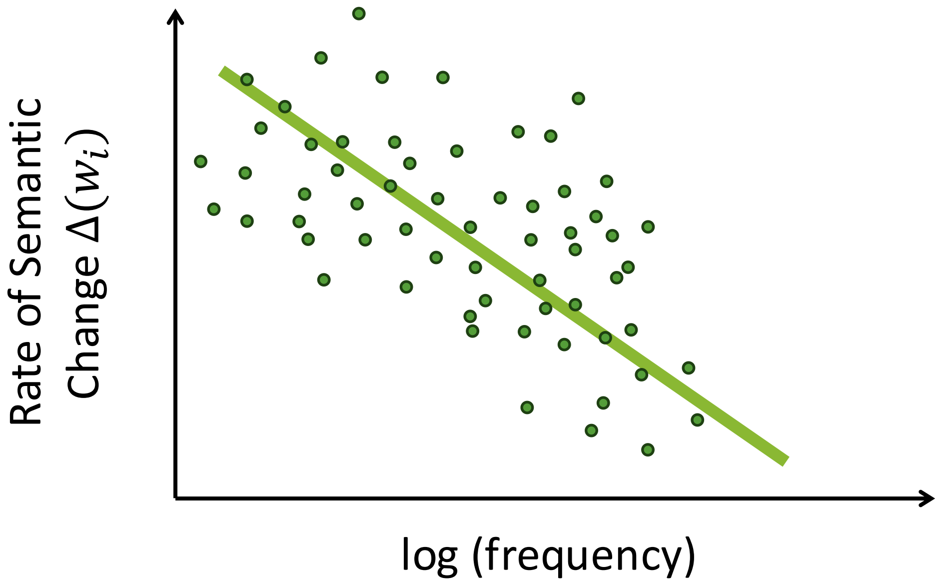

William L. Hamilton, Jure Leskovec, and Dan Jurafsky (2016) Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change

# Frequently used words change at slower rates as stated by Law of Conformity.

Regression analysis was used to explore the relationship between word frequency and the rate of semantic change.

By using linear mixed models, the authors accounted for both fixed effects (like frequency) and random effects (such as variations across different words or contexts).

# Frequently used words change at slower rates as stated by Law of Conformity.




$$\Delta(w_i) \propto f(w_i)\beta^f$$

# Law of Innovation posits that polysemous words change at faster rates.

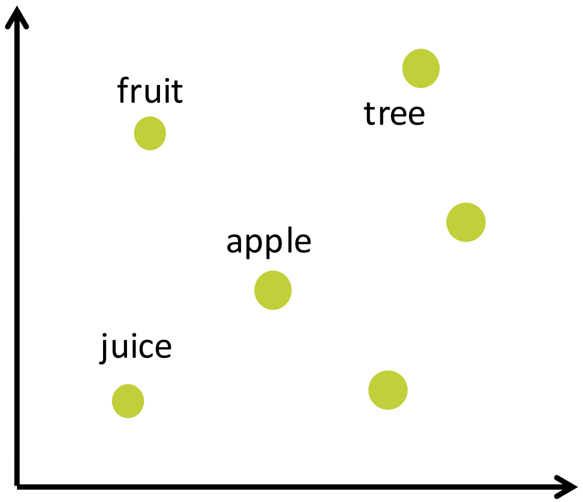
Polysemy is measured by the words that occur in many distinct contexts.

Regression analysis to understand how word frequency and polysemy relate to the rate of semantic change.

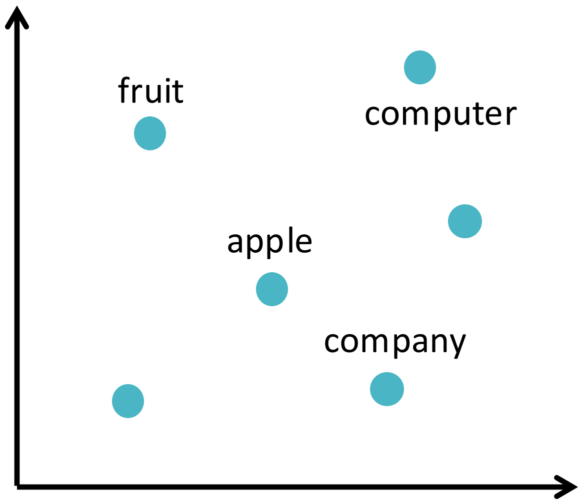
Measured a word's contextual diversity by looking at its co-occurrence patterns with other words in a large corpus.

# Law of Innovation posits that polysemous words change at faster rates.

- Polysemy is measured by the words that occur in many distinct contexts.
- Regression analysis to understand how word frequency and polysemy relate to the rate of semantic change.
- Measured a word's contextual diversity by looking at its co-occurrence patterns with other words in a large corpus.

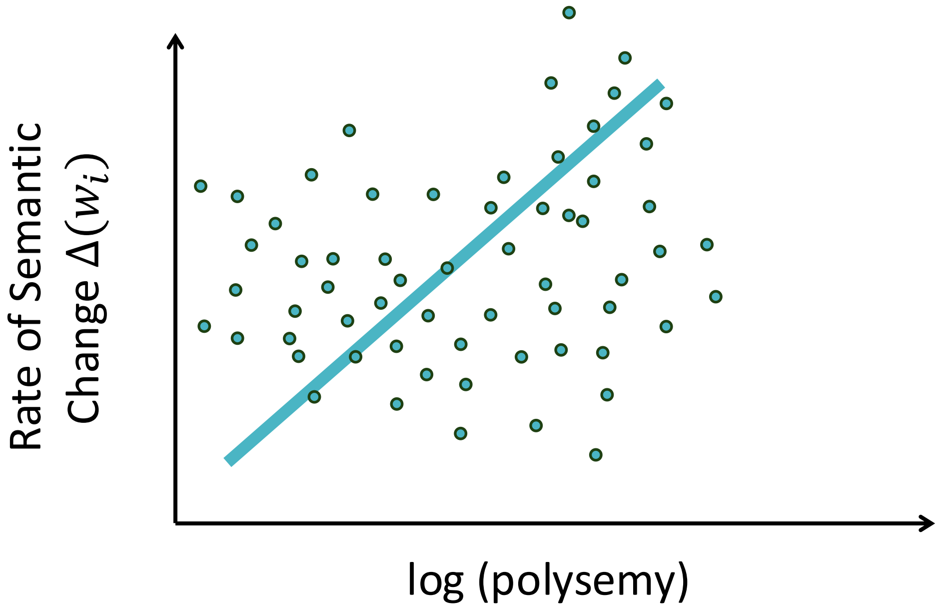



Case 1: High Clustering Coefficient (Low Polysemy)




Case 2: Low Clustering Coefficient (High Polysemy)

# Law of Innovation posits that polysemous words change at faster rates.




$$\Delta(w_i) \propto d(w_i)\beta^d$$

# Takeaways.

- Creation of **diachronic word embeddings** by using PPMI, SVD and SGNS.
- Identification of **statistical laws** of semantic change.