

# Computer-Linguistische Anwendungen

CLA | B.Sc. | LMU



# Vorlesung: Korpus-basierte semantische Ähnlichkeit; WordSpace

Philipp Wicke, PhD  
Centrum für Sprach- und Informationsverarbeitung  
Ludwig-Maximilians-Universität München  
[pwicke@cis.lmu.de](mailto:pwicke@cis.lmu.de)

20''



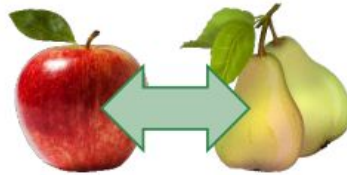
# Übersicht

- PPMI - Semantische Wortähnlichkeit
- PPMI - Modellierung von Wortähnlichkeit
- PPMI - Wordspace
- PPMI - Sprachwandel
- SVD - Wordspace
- SVD - Linear Algebra
- SVD - Input Matrix
- SVD - Matrix Factorization

# Was ist semantische Wortähnlichkeit?

# Was ist semantische Wortähnlichkeit?

Wann sind die Bedeutungen zweier Wörter ähnlich?



Apfel	← →	Birne
Tag	← →	Nacht
Schicht	← →	schlicht
sehen	← →	blicken
und	← →	Flugzeug
...		

# Was ist semantische Wortähnlichkeit?

Sichtweise 1:

- Die Wörter teilen bestimmte Eigenschaften

Tag	–	Nacht
Nomen		Nomen
Singular		Singular
Zeiteinheit		Zeiteinheit
Tag+Nacht		Nacht

- Die Eigenschaften bestimmen die Wortverwendung
  - *Intension* der Wortbedeutung: im Wort angelegte Eigenschaften
  - Sprachsystem

# Was ist semantische Wortähnlichkeit?

Sichtweise 2:

- Die Wörter können **durch einander** ersetzt werden, und eine **Aussage** bleibt sinnvoll:

“Ich musste einen **Tag** warten, bevor ich aufbrechen konnte”

“Ich musste eine **Nacht** warten, bevor ich aufbrechen konnte”

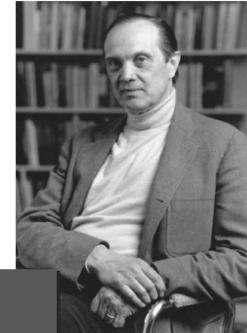
→ je mehr Kontexte dies erlauben, umso ähnlicher die Worte

- In welchen (sinnvollen Kontexten ein Wort vorkommt, kann beobachtet werden
  - Extension der Wortbedeutung: Umfang der möglichen Verwendungen
  - Sprach**gebrauch**

# Die lexikalische Hypothese



Gottfried Wilhelm Leibniz (17. Jahrhundert):  
"Die Dinge sind gleich, von denen eines durch  
das andere ersetzt werden kann, ohne den  
Wahrheitsgehalt zu beeinträchtigen."  
*Lat. "Eadem sunt quorum unum potest substitui  
alteri salva veritate."*

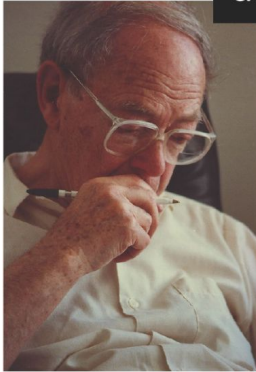


George A Miller (1991):  
"Those things are similar of  
which one can be  
**substituted** for the other  
without loss of plausibility"



# Die lexikalische Hypothese

Zellig Harris (1954): "difference in meaning correlates with difference of distribution."



John Rupert Firth (1957):  
"You shall know a word by  
the company it keeps."



# Verallgemeinerte Ähnlichkeit

- Die Wörter haben "etwas miteinander zu tun"

LKW            <->            Autobahn

Publikum      <->            Konzert

- Es besteht eine Relation zwischen den Wörtern

(X-Verwendet-Y, X-ist-Teil-von-Y, ...)

- Die Wörter kommen oft zusammen vor

# Wortähnlichkeit - wozu?

Anwendung: Für die automatische Erkennung von Wortähnlichkeit nützlich ist

# Wortähnlichkeit - wozu?

- **Suchmaschinen**

- Anfrage “E-Auto mit großer Reichweite”
- Treffer: “ ... mit diesen 5 kommen Sie weit, mit wenigen Ladungen ... ”

- **Koreferenz:** Welche Nomen in einem Text beziehen sich auf dieselbe Entität (Person, Gegenstand, Sachverhalt ...)

*“Beeindruckend von außen und wunderschön, innen ist der **Dom** sicher einer der Schönsten. Als historisches **Denkmal** der Stadt bietet diese wunderschöne **Kathedrale** einen beeindruckenden Blick auf die Altstadt ...”*

- ... viele weitere Anwendungen

# Modellierung von Wortähnlichkeit

# Modellierung mit dem Computer

**Modellierung:** **Formalisierung**, die von einem Computerprogramm verarbeitet werden kann.

**Korpus:** Textsammlung, z.B. Nachrichtenartikel, Wikipedia, ...

## Wissensbasiert

- Die Wörter teilen bestimmte Eigenschaften
- Intension, Sprachsystem
- Worтеigenschaften werden von Experten katalogisiert (elektronische Lexika, z.B. WordNet)

## Korpus-basiert

- Die Wörter werden ähnlich verwendet
- Extension, Sprachgebrauch
- Ähnlichkeitsmodell basiert auf Wortkontexten im Korpus

# Korpus-basierte Modellierung: Vorteile/Nachteile

## **Vorteile**

- Wortverwendung kann direkt beobachtet werden
- Es muss kein elektronisches Lexikon von Experten erstellt werden
- Alle Wörter in einer Textsammlung sind abgedeckt

## **Nachteile**

- Ungenau (vor allem für seltene Wörter)
- Oft wird generalisierte Ähnlichkeit modelliert (Zusammenvorkommen vs. Austauschbarkeit)

# Demo/Übung

- mit englischem Nachrichtenkorpus (GoogleNews) modellierte Wortähnlichkeit
- Überlegen Sie sich ein Wort (auf Englisch), und Wörter, die dazu ähnlich sind  
z.B.: apple – ....
- Weichen die gefundenen Wörter von der Erwartung ab? Inwiefern?

10''

<https://www.cis.uni-muenchen.de/schuetze/intro/tmp/index.html>



# Korpus-basiertes Ähnlichkeitsmodell

# Korpus-basiertes Ähnlichkeitsmodell


"Wordspace mit PPMI-Gewichtung und Kosinus-Ähnlichkeit"

- **Wordspace:**

- Wörter werden durch die Verteilung ihrer Nachbarwörter repräsentiert.
- Wortstatistik, als Vektor dargestellt
- "You shall know a word by the company it keeps."

- **PPMI-Gewichtung:**

- Positive Pointwise Mutual Information
- Nur besonders stark korrelierte Nachbarwörter werden berücksichtigt
- Effekt: Funktionswörter (die, der, und, in, zu,...) werden herausgefiltert


$$\text{ppmi}(x; y) \equiv \max \left( \log_2 \frac{p(x, y)}{p(x)p(y)}, 0 \right)$$

- **Kosinus-Ähnlichkeit:**

- Misst Überlappung der gewichteten Wortstatistiken

# Korpus-basiertes Ähnlichkeitsmodell: Wordspace

- Wörter werden durch ihre Nachbarwörter repräsentiert
- Wörter haben 2 Rollen, als
  1. zu modellierendes Ziel-Wort
  2. Wort im Kontext
- Statistik für Zielwort "Tofu", Kontextwörter:
  - "Temperatur" : 2
  - "Soja" : 5
  - "Marken" : 2
  - "und" : 7
  - ...

Verfahren zur Herstellung von Sojamilch und Tofu, das die Anwendung einer Behandlung mit hoher Temperatur, vermindertem [...]  
↳ [v3.espacenet.com](#)

[...] während die Temperatur der Milch 60°C oder höher ist, wodurch Tofu hergestellt wird; und anschliessendes Einfrieren; worin im [...]  
↳ [v3.espacenet.com](#)

[...] Graubünden, Spätzli im Mangoldmantel an Sauce Mornay serviert mit Bohnen, Gemüse und Tofu" als Hauptspeise auf dem Menü.  
↳ [swiss.com](#)

[...] Label Délicorn eine neue Linie von vegetarischen Produkten auf der Basis von Soja/Weizen, Tofu/Weizen und Bio-Getreide vor.  
↳ [coop.ch](#)

[...] klanglichen und visuellen Unterschied der ersten Bestandteile "Tofu" und "Curry" der in Rede stehenden Marken und durch den Zwischenraum [...]  
↳ [oami.europa.eu](#)

[...] Marken angeht, ist darauf hinzuweisen, dass der Bestandteil "Tofu" ein deutsches Wort ist und der Bestandteil "King" ein englischer [...]  
↳ [oami.europa.eu](#)

[...] Produktion lediglich auf die Verarbeitung von Sojabohnen, um Tofu bester Qualität zu gewinnen, erstreckte sich dann aber auf [...]  
↳ [biolab-eu.com](#)

Fleisch, Fisch, Ei, Käse, Tofu und Quorn sind reich an Eiweissen, die im Körper vor allem als Baustoff dienen.  
↳ [hug-luzern.ch](#)

[...] besten ein Produkt, das mit Calcium angereichert ist, da Sojamilch, im Gegensatz zu Tofu, nur sehr wenig Calcium enthält.  
↳ [maggi.ch](#)

[...] es noch Alternativen aus anderen Kulturkreisen: Aus Japan Tofu (aus Sojabohnen) und Seitan\* (Weizeneiweiss), aus Indien Paneer\* (aus frischer Milch).  
↳ [good-will.ch](#)

# Korpus-basiertes Ähnlichkeitsmodell: WordSpace

- Wörter werden durch ihre Nachbarwörter repräsentiert
- Wörter haben 2 Rollen, als
  1. zu modellierendes Ziel-Wort
  2. Wort im Kontext
- Statistik für Zielwort "Tomate", Kontextwörter:
  - "Temperatur" : 0
  - "Soja" : 1
  - "Marken" : 0
  - "und" : 6
  - ...

Die gegenwärtige Regelung für Verarbeitungserzeugnisse aus **Tomaten** soll insofern geändert werden, als die besten Bestandteile [...] [eur-lex.europa.eu](#)

Vorrichtung zur Abwicklung von Schnur zum Anbau von **Tomaten** im Treibhaus mittels des sogenannten Wachstums durch seitliche [...] [v3.espacenet.com](#)

Hergestellt mit Vollkornmehl, gewürzt mit den sogenannten Sponsilli-**Tomaten**, die unter Bogengängen aufbewahrt werden, mit Olivenöl, [...] [costiera.it](#)

[...] Sojasauce, mit Zusatz von Essig, Gewürzen, Salz, Zucker, Zuckercouleur und Senf), **Tomaten-Ketchup** (auf der Grundlage von Tomatenmark, Zucker, Essig, Salz und Gewürzen) [...] [ezv.admin.ch](#)

[...] Druckerkartuschen und den Ersatz von Scheiben in einer Erntemaschine für **Tomaten** sämtlich als zulässige Reparatur angesehen. [aiippi.org](#)

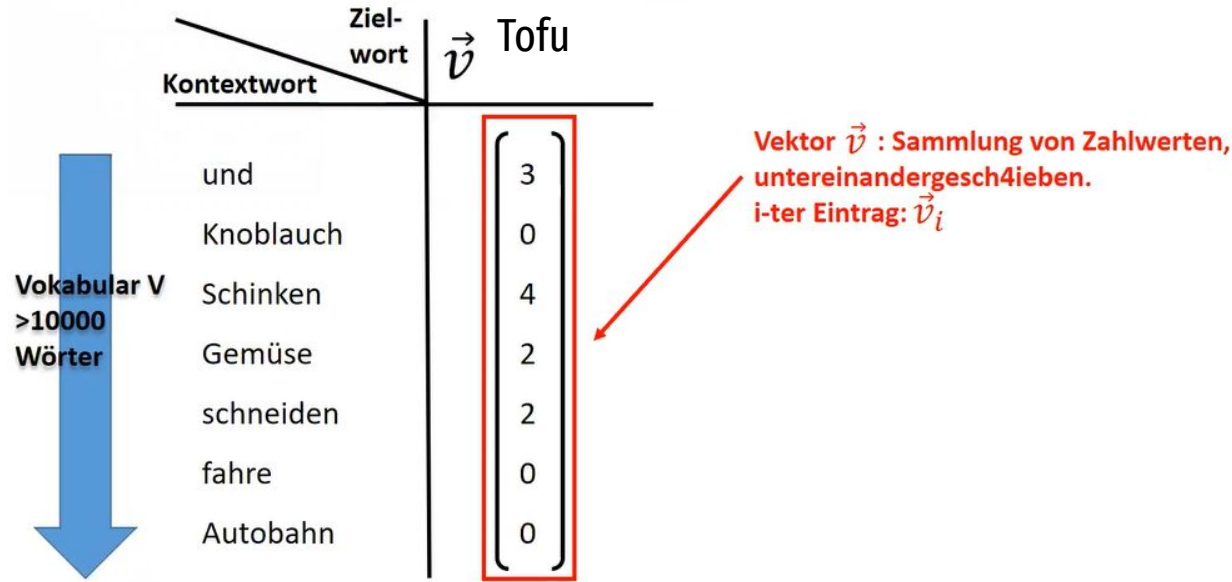
Einer von ihnen sagte mir: ?Wenn ich einen Container **Tomaten** beispielsweise von Ramallah nach Gaza bringe, kostet mich das [...] [europarl.europa.eu](#)

[...] Notwendigkeit zusammen, in den Wintermonaten über frische **Tomaten** zu verfügen, um die seit jeher in der Gegend von Neapel [...] [eur-lex.europa.eu](#)

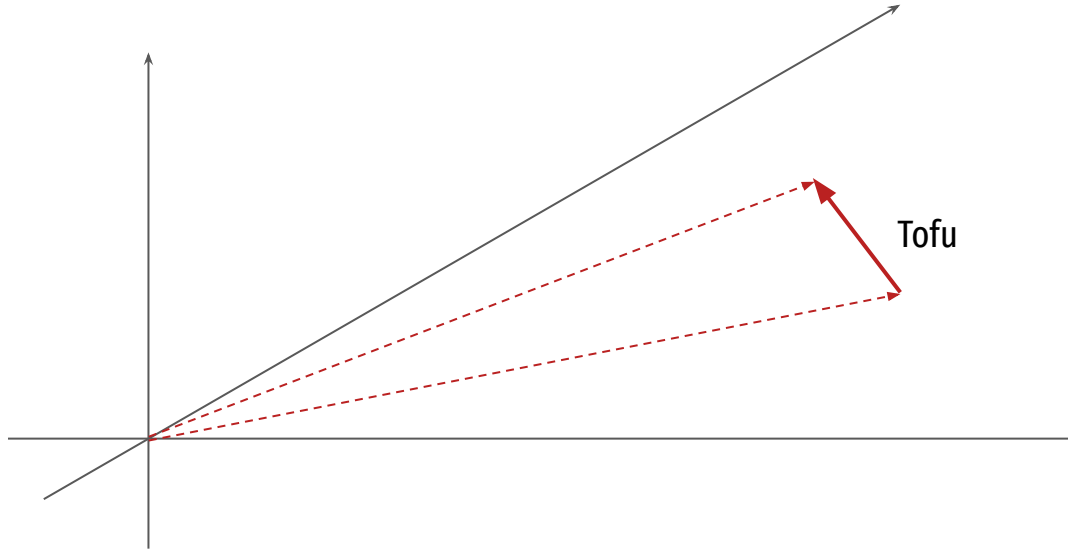
[...] vorzubereiten, die die Zwiebeln schmückt, fein im Öl, dann die geschälten **Tomaten** hinzuzufügen, springt und ohne Samen. [guguigueb.com.ar](#)

Andere Gerichte sind malloreddus, Grieß Knödel gewürzt mit **Tomaten**-, Wurst- und Safran, die culurgiones, große Ravioli gefüllt mit Kartoffelpüree, Eier, Minze, Zwiebeln und Käse, gewürzt mit **Tomaten**-Soße und Fleisch, Brot frattau, Blatt Brot weich und in Brühe serviert mit Pecorino Käse, Tomatensauce und pochiertem Ei; fregula der Paste, Mehl verarbeitet, serviert trocken oder mit Muscheln in Brühe als Suppe, die Suppe Gallura, auf der Grundlage von Brot, **Tomaten**, Brühe und Pecorino. [sardegna.italiaguida.it](#)

# Darstellung als Vektoren



# Vektorraum

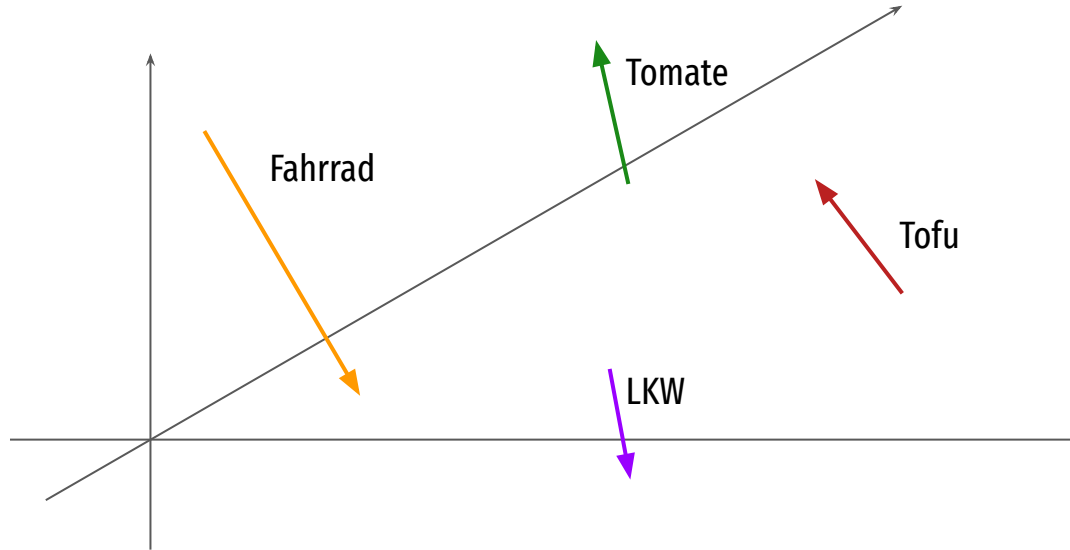


Jeder erzeugte Vektor liegt im N-dimensionalen Vektorraum, hier  $N = >10.000$  Vokabular-Wörter. Wir stellen diesen Vektorraum hier 3 Dimensional dar.

# Darstellung als Vektoren

<div> <div>Ziel- wort</div> <div>Kontextwort</div> </div>	$\vec{v}$ Tofu	$\vec{v}$ Tomate	$\vec{v}^{LKW}$	$\vec{v}^{Fahrrad}$
und	$\begin{bmatrix} 3 \end{bmatrix}$	$\begin{bmatrix} 3 \end{bmatrix}$	$\begin{bmatrix} 4 \end{bmatrix}$	$\begin{bmatrix} 3 \end{bmatrix}$
Knoblauch	$\begin{bmatrix} 0 \end{bmatrix}$	$\begin{bmatrix} 2 \end{bmatrix}$	$\begin{bmatrix} 0 \end{bmatrix}$	$\begin{bmatrix} 0 \end{bmatrix}$
Schinken	$\begin{bmatrix} 4 \end{bmatrix}$	$\begin{bmatrix} 1 \end{bmatrix}$	$\begin{bmatrix} 0 \end{bmatrix}$	$\begin{bmatrix} 0 \end{bmatrix}$
Gemüse	$\begin{bmatrix} 2 \end{bmatrix}$	$\begin{bmatrix} 1 \end{bmatrix}$	$\begin{bmatrix} 0 \end{bmatrix}$	$\begin{bmatrix} 0 \end{bmatrix}$
schneiden	$\begin{bmatrix} 2 \end{bmatrix}$	$\begin{bmatrix} 0 \end{bmatrix}$	$\begin{bmatrix} 0 \end{bmatrix}$	$\begin{bmatrix} 0 \end{bmatrix}$
fahre	$\begin{bmatrix} 0 \end{bmatrix}$	$\begin{bmatrix} 0 \end{bmatrix}$	$\begin{bmatrix} 2 \end{bmatrix}$	$\begin{bmatrix} 3 \end{bmatrix}$
Autobahn	$\begin{bmatrix} 0 \end{bmatrix}$	$\begin{bmatrix} 0 \end{bmatrix}$	$\begin{bmatrix} 1 \end{bmatrix}$	$\begin{bmatrix} 0 \end{bmatrix}$

# Vektorraum



Wie berechnen Sie die Ähnlichkeit zweier Vektoren?



# Berechnung der Kosinus-Ähnlichkeit

Der Kosinus erlaubt uns die Ähnlichkeit zweier Wortvektoren zu vergleichen:

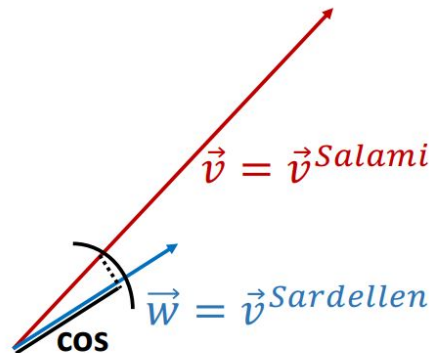
$$\cos(\vec{v}, \vec{w}) = \frac{\vec{v} \cdot \vec{w}}{|\vec{v}| |\vec{w}|}$$

1. Skalar-Produkt: ~ Überlappung der Wortvektoren

$$\vec{v} \cdot \vec{w} = \sum_i \vec{v}_i \vec{w}_i$$

2. Normierung durch Länge der Vektoren (L2-Norm)

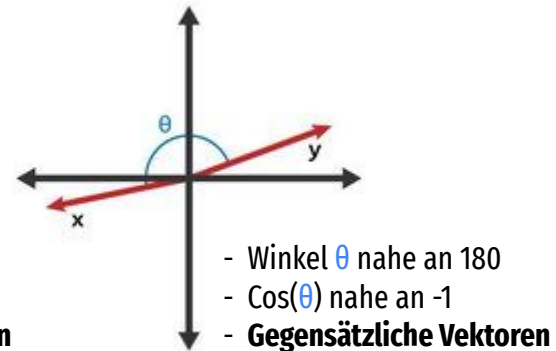
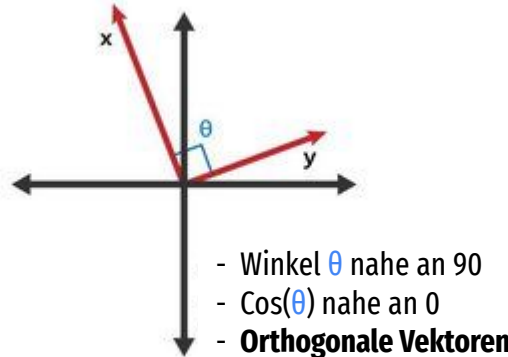
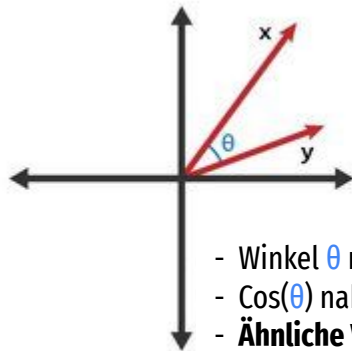
$$|\vec{v}| = \sqrt{\sum_i \vec{v}_i^2}$$



# Berechnung der Kosinus-Ähnlichkeit

Der Kosinus erlaubt uns die Ähnlichkeit zweier Wortvektoren zu vergleichen:

$$\cos(\vec{v}, \vec{w}) = \frac{\vec{v} \cdot \vec{w}}{|\vec{v}| |\vec{w}|}$$



# Berechnung der Kosinus-Ähnlichkeit: Beispiel

$$\cos(\vec{v}^{\text{Tofu}}, \vec{v}^{\text{Tomate}}) = \frac{\vec{v}^{\text{Tofu}} \cdot \vec{v}^{\text{Tomate}}}{|\vec{v}^{\text{Tofu}}| |\vec{v}^{\text{Tomate}}|}$$
$$= \frac{3 \cdot 3 + 4 \cdot 1 + 2 \cdot 1}{\sqrt{3^2 + 4^2 + 2^2 + 2^2} \sqrt{3^2 + 2^2 + 1^2 + 1^2}}$$
$$\approx 0,67$$

$\vec{v}^{\text{Tofu}}$

$$\begin{bmatrix} 3 \\ 0 \\ 4 \\ 2 \\ 2 \\ 0 \\ 0 \end{bmatrix}$$

$\vec{v}^{\text{Tomate}}$

$$\begin{bmatrix} 3 \\ 2 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

# Berechnung der Kosinus-Ähnlichkeit: Beispiel

<div><div>Ziel- wort</div><div>Kontextwort</div></div>	$\vec{v}^{\text{Tofu}}$	$\vec{v}^{\text{Tomate}}$	$\vec{v}^{\text{LKW}}$	$\vec{v}^{\text{Fahrrad}}$
und	$\begin{bmatrix} 3 \\ 0 \\ 4 \\ 2 \\ 2 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 2 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \\ 0 \\ 2 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \\ 0 \\ 3 \\ 0 \end{bmatrix}$
Knoblauch				
Schinken				
Gemüse				
schneiden				
fahre				
Autobahn				

$$\cos(\vec{v}^{\text{LKW}}, \vec{v}^{\text{Fahrrad}}) = 0.92$$

$$\cos(\vec{v}^{\text{Tofu}}, \vec{v}^{\text{Tomate}}) = 0.67$$

$$\cos(\vec{v}^{\text{Tofu}}, \vec{v}^{\text{Fahrrad}}) = 0.36$$

$$\cos(\vec{v}^{\text{LKW}}, \vec{v}^{\text{Tomate}}) = 0.68$$

...

# Gewichtung der Vektoren

- **Problem:** Sehr häufige, nicht-informative Wörter (*die, der, und, in, zu,...*) beeinflussen die Ähnlichkeitsberechnung zu stark
- **Lösung:** Gewichtung nach Informativität

→ Dazu verwenden wir statistische Information über die Verteilung von Wörtern und Wortpaaren

# Gewichtung der Vektoren

- Die Wahrscheinlichkeiten werden aus den relativen Frequenzen berechnet

- Wahrscheinlichkeit eines Wortpaares:

$$P(w_{Ziel}, w_{Kontext}) = \frac{\text{Wie oft kommen } w_{Ziel} \text{ und } w_{Kontext} \text{ zusammen vor}}{\text{Gesamtvorkommen aller Wortpaare}}$$

- Wahrscheinlichkeit, dass Wort als Ziel- bzw. Kontextwort vorkommt:

$$P(w_{Ziel}) = \frac{\text{Wie oft kommt } w_{Ziel} \text{ in Wortpaaren als Ziel vor}}{\text{Gesamtvorkommen aller Wortpaare}}$$

$$P(w_{Kontext}) = \frac{\text{Wie oft kommt } w_{Kontext} \text{ in Wortpaaren als Kontext vor}}{\text{Gesamtvorkommen aller Wortpaare}}$$

# Gewichtung der Vektoren

Was wird gezählt?	Notation
<b>Zusammenvorkommen</b> von Ziel- und Kontextwort (Wortpaar $w_{\text{Ziel}}, w_{\text{Kontext}}$ )	$n(w_{\text{Ziel}}, w_{\text{Kontext}})$
<b>Vorkommen</b> als Ziel-  bzw. Kontextwort	$n_Z(w_{\text{Ziel}}) = \sum_{w' \in V} n(w_{\text{Ziel}}, w')$ $n_K(w_{\text{Kontext}}) = \sum_{w' \in V} n(w', w_{\text{Kontext}})$
<b>Vorkommen</b> aller Wortpaare (Summe)	$N$

# Wahrscheinlichkeiten: Aus relativen Frequenzen

Was wird berechnet?	Notation
<b>Wahrscheinlichkeit</b> eines Wortpaares	$P(w_{Ziel}, w_{Kontext}) = \frac{n(w_{Ziel}, w_{Kontext})}{N}$
<b>Wahrscheinlichkeit</b> , dass Wort als Ziel- bzw. Kontextwort vorkommt	$P(w_{Kontext}) = \frac{n_K(w_{Kontext})}{N}$ $P(w_{Ziel}, w_{Kontext}) = \frac{n_Z(w_{Ziel})}{N}$



# Gewichtung der Vektoreinträge

Gewichtung durch Positive Pointwise Mutual Information:

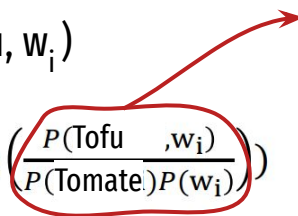
$$\vec{v}_i^{w_{\text{Tofu}}} = \text{PPMI}(\text{Tofu}, w_i)$$

$$\text{PPMI}(\text{Tofu}, w_i) = \max(0, \log \left( \frac{P(\text{Tofu}, w_i)}{P(\text{Tomate})P(w_i)} \right))$$

# Gewichtung der Vektoreinträge

Gewichtung durch Positive Pointwise Mutual Information:

$$\vec{v}_i^{w_{\text{Tofu}}} = \text{PPMI}(\text{Tofu}, w_i)$$

$$\text{PPMI}(\text{Tofu}, w_i) = \max(0, \log \left( \frac{P(\text{Tofu}, w_i)}{P(\text{Tomate})P(w_i)} \right))$$


Wörter kommen **überdurchschnittlich oft zusammen** vor, wenn die gemeinsame Wahrscheinlichkeit im Verhältnis zu einzelnen Wahrscheinlichkeiten größer als 1 ist.

→ denn: Wenn zwei  $w_k$  unabhängig voneinander sind dann ist die gemeinsame  $w_k$  gleich deren Produkt.

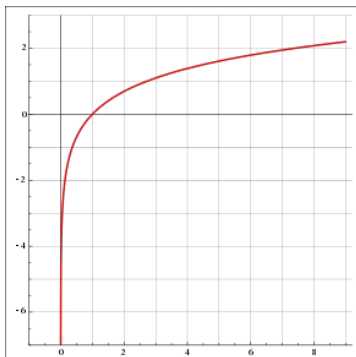
$$P(A \text{ und } B) = P(A) * P(B)$$
$$P(A) / P(B) > 1$$

# Gewichtung der Vektoreinträge

Gewichtung durch Positive Pointwise Mutual Information:

$$\vec{v}_i^{w_{\text{Tofu}}} = \text{PPMI}(\text{Tofu}, w_i)$$

$$\text{PPMI}(\text{Tofu}, w_i) = \max(0, \log\left(\frac{P(\text{Tofu}, w_i)}{P(\text{Tomate})P(w_i)}\right))$$



Wenn die Wahrscheinlichkeit größer 1 ist, dann ist der **log** der Wahrscheinlichkeit größer 0

Wörter kommen **überdurchschnittlich oft zusammen** vor, wenn die gemeinsame Wahrscheinlichkeit im Verhältnis zu einzelnen Wahrscheinlichkeiten größer als 1 ist.

→ denn: Wenn zwei  $w_k$  unabhängig voneinander sind dann ist die gemeinsame  $w_k$  gleich deren Produkt.

$$P(A \text{ und } B) = P(A) * P(B)$$
$$P(A) / P(B) > 1$$

# Gewichtung der Vektoreinträge

Gewichtung durch Positive Pointwise Mutual Information:

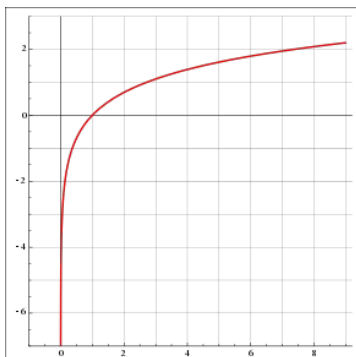
$$\vec{v}_i^{w_{\text{Tofu}}} = \text{PPMI}(\text{Tofu}, w_i)$$

$$\text{PPMI}(\text{Tofu}, w_i) = \max\left(0, \log\left(\frac{P(\text{Tofu}, w_i)}{P(\text{Tomate})P(w_i)}\right)\right)$$

Wörter kommen **überdurchschnittlich oft zusammen** vor, wenn die gemeinsame Wahrscheinlichkeit im Verhältnis zu einzelnen Wahrscheinlichkeiten größer als 1 ist.

→ denn: Wenn zwei  $w_k$  unabhängig voneinander sind dann ist die gemeinsame  $w_k$  gleich deren Produkt.

$$P(A \text{ und } B) = P(A) * P(B)$$
$$P(A) / P(B) > 1$$



Wenn die Wahrscheinlichkeit größer 1 ist, dann ist der **log** der Wahrscheinlichkeit größer 0

PPMI ist dann  $>0$ , wenn die Wörter überdurchschnittlich oft zusammen vorkommen

# Vektoren nach PPMI-Gewichtung

<div> <div>Ziel- wort</div> <div>Kontextwort</div> </div>	$\vec{v}^{\text{Tofu}}$	$\vec{v}^{\text{Tomate}}$	$\vec{v}^{\text{LKW}}$	$\vec{v}^{\text{Fahrrad}}$
und	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1.5 \end{bmatrix}$	$\begin{bmatrix} 0.3 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0.2 \\ 0 \end{bmatrix}$
Knoblauch	$\begin{bmatrix} 0.8 \\ 0.6 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0.4 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$
Schinken	$\begin{bmatrix} 1.0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0.6 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 1.1 \end{bmatrix}$
Gemüse	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0.6 \\ 1.5 \end{bmatrix}$	$\begin{bmatrix} 1.1 \\ 0 \end{bmatrix}$
schneiden				
fahre				
Autobahn				

$$\cos(\vec{v}^{\text{LKW}}, \vec{v}^{\text{Fahrrad}}) = 0.38$$

$$\cos(\vec{v}^{\text{Tofu}}, \vec{v}^{\text{Tomate}}) = 0.11$$

$$\cos(\vec{v}^{\text{Tofu}}, \vec{v}^{\text{Fahrrad}}) = 0.0$$

$$\cos(\vec{v}^{\text{LKW}}, \vec{v}^{\text{Tomate}}) = 0.0$$

...

# Mehr zu PPMI

- Pointwise Mutual Information:  $\text{PMI}(w_i, w_j) = \log \left( \frac{P(w_i, w_j)}{P(w_i)P(w_j)} \right)$

- Die PMI misst die Abweichung von der statistischen Unabhängigkeit
- Falls Auftreten der Wörter unabhängig ist, gilt

$$P(w_i, w_j) = P(w_i)P(w_j) \Leftrightarrow \text{PMI}(w_i, w_j) = 0$$

- Die PMI misst, wie sehr das Zusammen-Auftreten der Wörter von der durch die Einzel-Wahrscheinlichkeiten zu erwarteten Häufigkeit abweicht (positiv oder negativ)

- Warum *positiv*-PMI?

- Die positiven Korrelationen sind am ausgeprägtesten
- Es ist sehr schwierig, starke negativ-korrelierte Wörter zu finden

# Mehr zu PPMI

- PPMI mit Offset:  $\text{PPMI}_k(w_i, w_j) = \max(0, \text{PMI}(w_i, w_j) - k)$
- Warum ein *Offset*?
  - Schwache Korrelationen sind oft nur statistisches Rauschen und können entfernt werden

# Was gilt als Kontext?

- Weiter Kontext:
  - Alle Wörter im gleichen Dokument
  - Alle Wörter im gleichen Satz
- Enger Kontext:
  - Nur Wort direkt links/rechts von Zielwort
  - Nur Wörter, die syntaktisch im Satz verbunden sind (Argumente, Köpfe)
- Weite Kontexte: Ähnlichkeit eher thematisch (nicht strikte Ersetzung)
- Enge Kontexte: Ähnlichkeit nach syntaktischen Eigenschaften (Wortart, Kasus)



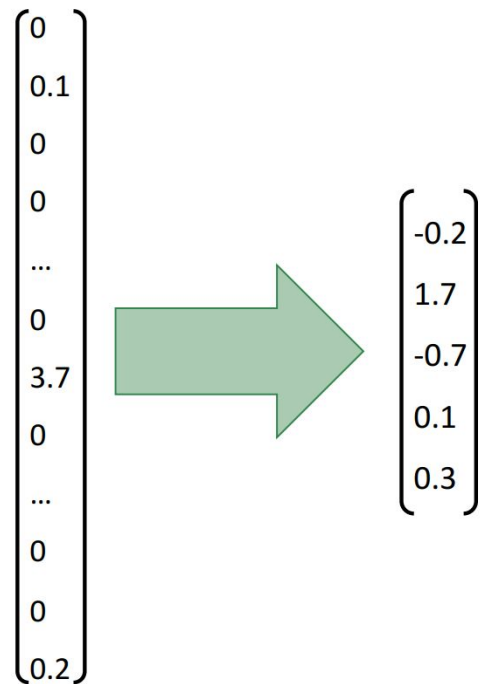
# In der Praxis

- 2-5 Wörter links und rechts vom Zielwort haben sich als gute Kontextgröße erwiesen
- Setzen Sie die PPMI-Werte auf 0, wenn das Wortpaar seltener als ca. 5-Mal vorgekommen ist

*This is a test sentence to **show** the context size for PPMI. Word-pairs with <5 occurrences can be set to 0.*

# Ausblick: “*Word embeddings*”

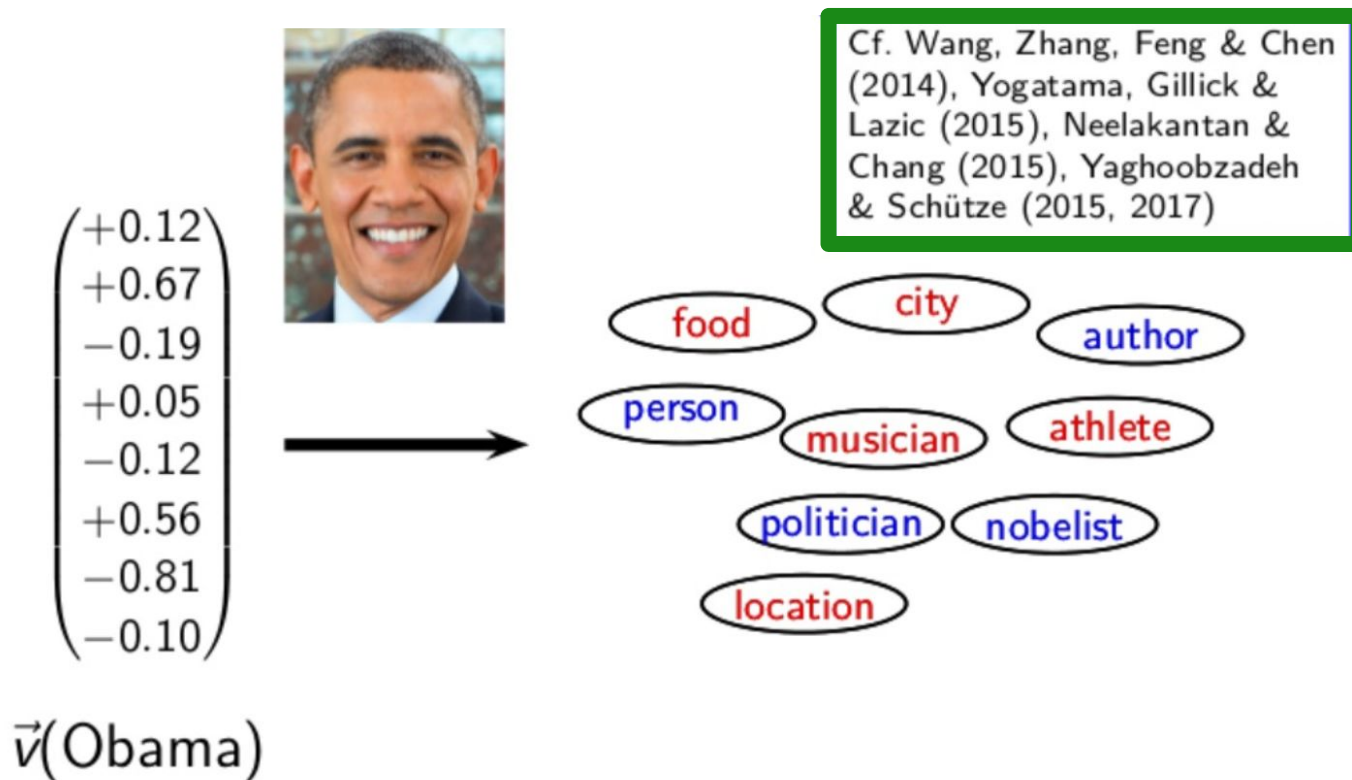
- WordSpace-Vektoren haben sehr viele Einträge (>10000)
  - Hoher Speicherverbrauch
  - Ähnlichkeit von 0 für viele Wortpaare
- Besser:
  - Vektoren mit weniger Einträgen (50-1000)
  - “*Word embeddings*”
- Embedding-Verfahren:
  - Information so gut wie möglich mit weniger, dafür informativen Einträgen darstellen
  - Singular Value Decomposition, Word2vec, GloVe, ...



# Wort-Vektoren als Merkmalsvektoren

- Wort-Vektoren können nicht nur verwendet werden, um Wortähnlichkeiten zu berechnen
- Die Wort-Vektoren können auch als Merkmalsvektoren verwendet werden, um Wort-Eigenschaften vorherzusagen, z.B.
  - Wortart (Adj, Noun, Verb, ...)
  - Sentiment des Wortes (POS, NEG, NEU)
  - Bei Entitäten: Typ der Entität (Person, Organisation, Ort, ...)

# Wort-Vektoren als Merkmalsvektoren



# Anwendungsbeispiel: Historischer Sprachwandel

# Historischer Sprachwandel

- Veränderung der Wortbedeutung kann durch Korpus-basierte Wortähnlichkeit nachvollzogen werden [Kim, 2014]
- Erstellen von Wort-Vektoren aus Korpora verschiedener Zeiträume
- Für welches Wort ändert sich der Vektor am meisten? (Kosinus)

Word	Neighboring Words in	
	1900	2009
<i>cell</i>	<i>closet</i> <i>dungeon</i> <i>tent</i>	<i>phone</i> <i>cordless</i> <i>cellular</i>
<i>checked</i>	<del><i>checking</i></del> <i>recollecting</i> <i>straightened</i>	<del><i>checking</i></del> <i>consulted</i> <i>check</i>

**Häufige Verwendung um 1900: “sich sammeln, zurückhalten”**

1900: “She was about to say something further, but she *checked* herself.”

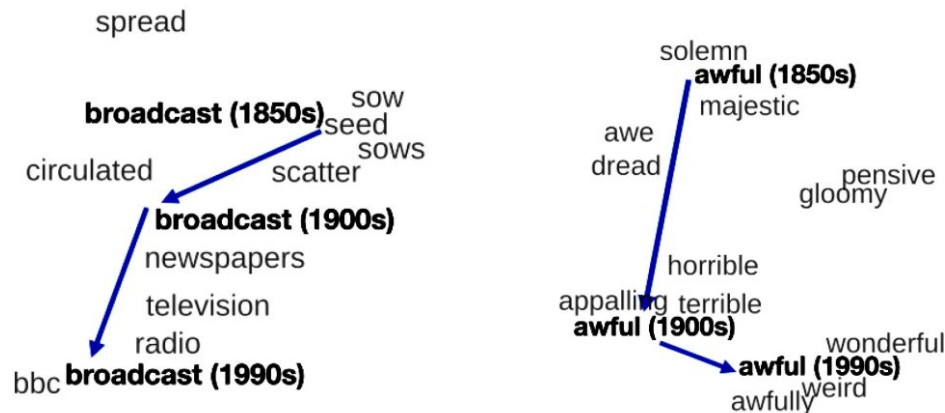
**Häufige Verwendung um 2009: “überprüfen, nachschauen”**

2009: “He’d *checked* his facts on a notepad from his back pocket.”

[Kim, 2014]

# Historischer Sprachwandel

- Beispiele aus [Hamilton, 2016]



- Analyse [Hamilton, 2016]: Seltene und mehrdeutige Wörter ändern sich besonders schnell

# Zusammenfassung

- Wortähnlichkeit
  - Wissensbasiert: Worteigenschaften
  - Korpus-basiert: Wortverwendung
- Lexikalische Hypothese
  - “You shall know a word by the company it keeps”
  - Wortbedeutung wird durch den Kontext bestimmt
- Methode: Wordspace mit Kosinus-Ähnlichkeit
- Anwendungsbeispiel: Historischer Sprachwandel



# Referenzen

- [Hamilton, 2016]: Cultural shift or linguistic drift? comparing two computational measures of semantic change
- [Kim 2014]: Temporal analysis of language through neural language models