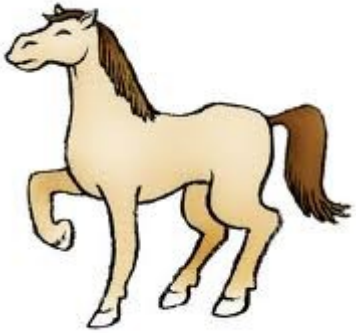
Von Saskia Reifers

Vector-based Models of Semantic Composition

Von Jeff Mitchell und Mirella Lapata

Proseminar: Distributionelle Semantik, 14.11.2011

Einleitung



← das Pferd läuft

die Farbe läuft →

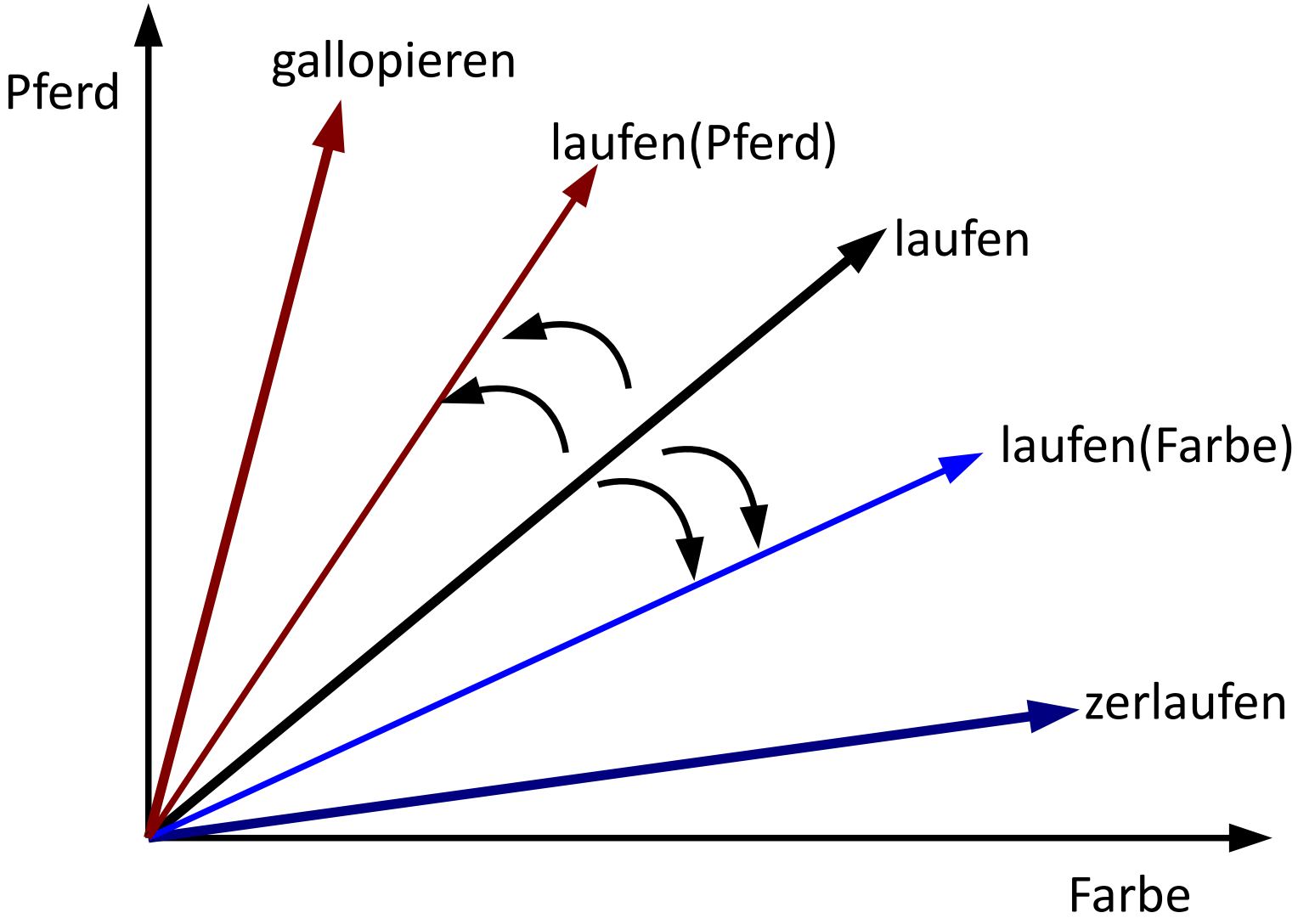
das Pferd gallopiert

die Farbe gallopiert??

das Pferd zerläuft??

die Farbe zerläuft

Einleitung


Pferd
gallopieren
laufen(Pferd)
laufen
laufen(Farbe)
zerlaufen
Farbe

Inhaltsverzeichnis

Worum geht es?

Mitchell&Lapata

Problem:

- Bisherige Modelle betrachten nur einzelne Worte.
- Modelle, die ganze Phrasen oder Sätze betrachten, haben bislang nur wenig Beachtung in der Literatur gefunden.
- Vektormittelwertbildung: $p = ½ (u + v)$
    - Unsensibel für syntaktische Struktur

It was not the sales manager who hit the bottle that day, but the office worker with the serious drinking problem.

That day the office manager, who was drinking, hit the problem sales worker with a bottle, but it was not serious.

Worum geht es?

Wieso ist das interessant?

- Ein Vergleich von verschiedenen Modellen hat bislang noch nicht so oft stattgefunden.
  - interessant, welches am besten abschneidet
- Vektor Ähnlichkeit im semantischen Raum stimmt im wesentlichen mit der menschlichen Beurteilung überein.
  - Deswegen ein Test mit Personen, um die menschliche Beurteilung mit einzubeziehen
  - empirisch begründet

Vorstellung der Modelle

Allgemeine Einführung

		animal	stable	village	gallop	jokey
Vektor u →	horse	0	6	2	10	4
Vektor v →	run	1	8	4	4	0
	gallop	3	9	1	0	6

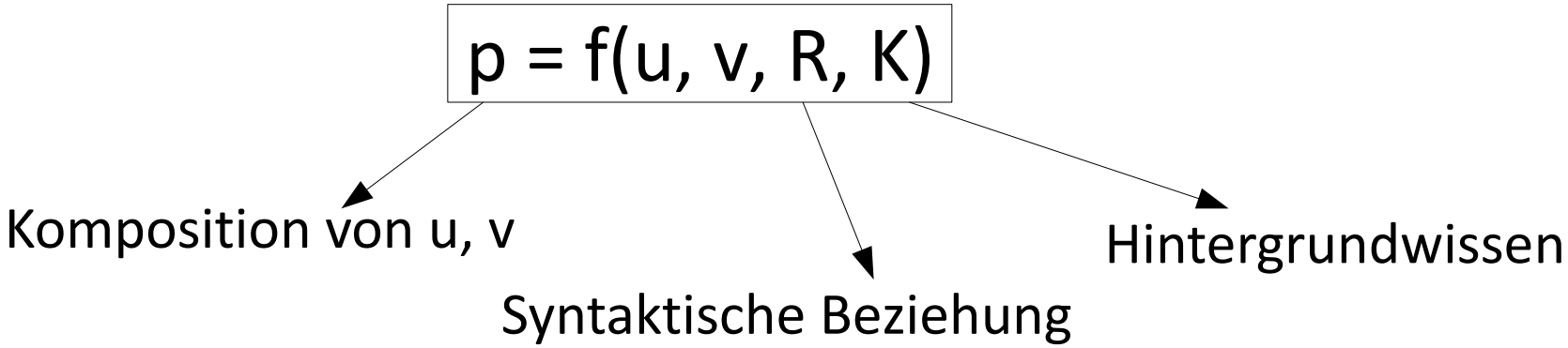
Semantischer Raum

- Vektor eines Wortes stellt Kookkurenz mit Nachbarwörtern dar
  - Kookkurrenz: gemeinsames Auftreten zweier Wörter

    → Kontext
- Semantische Komposition(p) wird durch zwei Vektoren(u, v) dargestellt

Vorstellung der Modelle

Rahmen für die Modelle



- Änderungen:
  - Weglassen des Hintergrundwissens K
  - R fixieren auf eine linguistische Struktur (hier Subjekt-Verb)

    → p = f(u, v)
    - p liegt im gleichen Vektorraum wie u und v
    - f: lineare Funktion, der Einfachheit halber

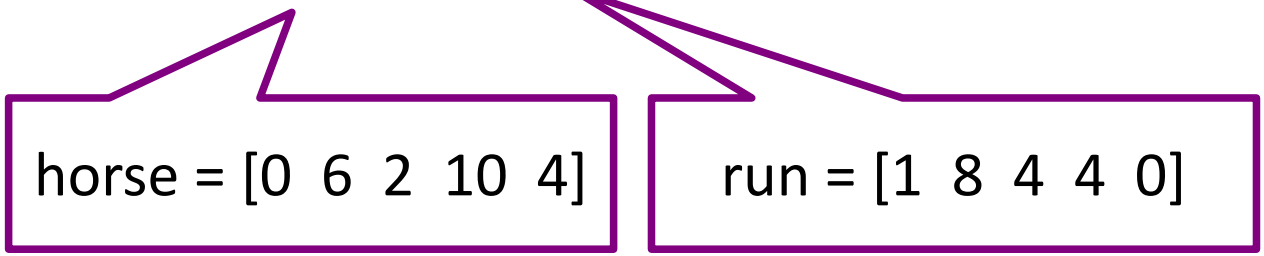
Vorstellung der Modelle

Addition

- keine Beachtung von Bedeutungsunterschieden durch Wortfolge
  - Beispiel: **Hans liebt Maria.** ↔ **Maria liebt Hans.**

    → unterschiedliche Bedeutung, aber gleiche Berechnung
- p = u + v
  - Beispiel:

    → horse + run = [1 14 6 14 4]



Vorstellung der Modelle

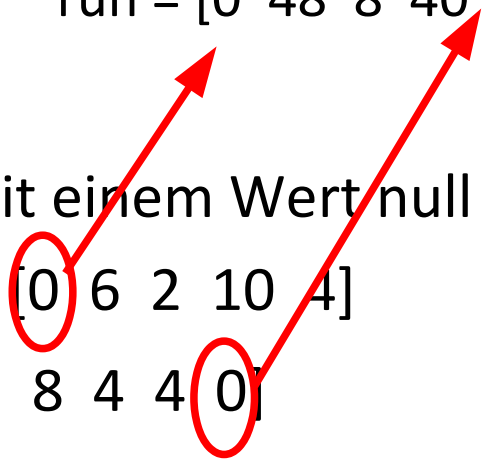
Multiplikation

- die Vektoren interagieren miteinander
  - heben wichtigen Inhalt besser hervor
- p = u * v
  - Beispiel:

    → horse * run = [0 48 8 40 0]
- Probleme
  - Vektoren mit einem Wert null

    horse = [0 6 2 10 4]

    run = [1 8 4 4 0]
  - Wortreihenfolge



Vorstellung der Modelle

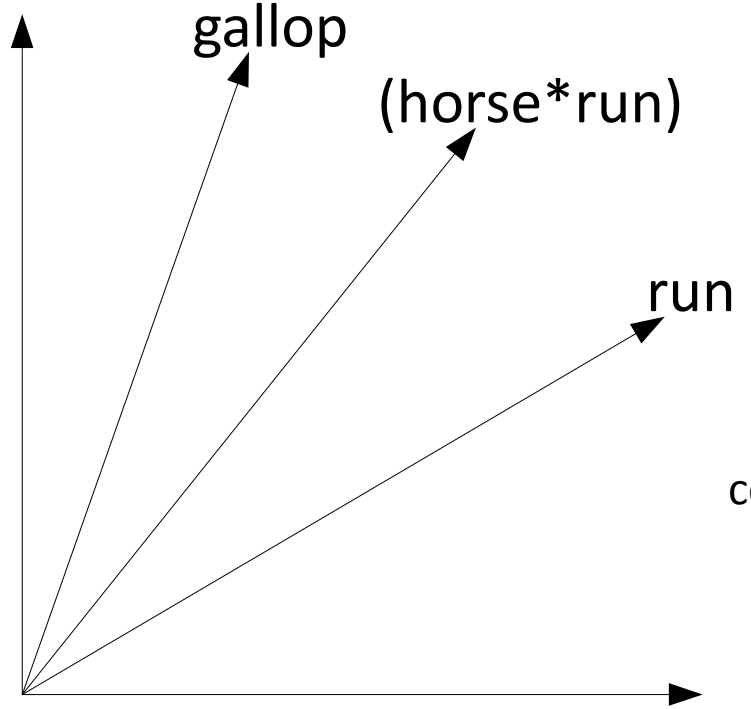
Beispiel für Multiplikation mit Zahlen

$$\cos = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}}$$

horse * run = [0 48 8 40 0]

run = [1 8 4 4 0]

gallop = [3 9 1 0 6]



$$\cos(\text{run}, \text{gallop}) = \frac{\sum \text{run} * \text{gallop}}{\sqrt{\sum \text{run}^2} * \sqrt{\sum \text{gallop}^2}} = \frac{79}{127} = 0{,}62$$

$$\cos(\text{horse} * \text{run}, \text{gallop}) = \frac{\sum (\text{horse} * \text{run}) * \text{gallop}}{\sqrt{\sum (\text{horse} * \text{run})^2} * \sqrt{\sum \text{gallop}^2}} = \frac{440}{490{,}5} = 0{,}897$$

Vorstellung der Modelle

Kintschs Modell

- Einbeziehung von Nachbarwörtern
- sensibel für syntaktische Struktur
- p = u + v + ∑ n
  - n: ein oder mehrere Vektoren der Nachbarwörter
  - Beispiel:
    - horse + run + ride = [3 29 13 23 5]

	animal	stable	village	gallop	jokey
horse	0	6	2	10	4
run	1	8	4	4	0
ride	2	15	7	9	1

Vorstellung der Modelle

Gewichtete Addition

- unterschiedliche Gewichtung
  - syntaktische Struktur wird bewusster
  - wichtige Informationen werden mehr betont
- $p = \alpha u + \beta v$
  - Beispiel:
    - $\alpha = 0{,}4$ und $\beta = 0{,}6$
    - α * horse + β * run = [0,6 5,6 3,2 6,4 1,6]

Vorstellung der Modelle

Kombination von Addition und Multiplikation

- das Beste von beiden Modellen
  - syntaktische Struktur
  - kein Problem mit Nullen
  - Inhalt wird besser aufgegriffen
- $p = \alpha u + \beta v + \gamma uv$
  - Beispiel:
    - $\alpha = 0{,}3$ und $\beta = 0{,}5$ und $\gamma = 0{,}2$
    - α * horse + β * run + γ * horse * run = [0,5 15,4 4,2 13 1,2]

Testreihe mit Personen

Warum?

- Brauchbare Testdaten für den Vergleich der Modelle
  - Datensatz
  - Test-Ergebnisse zum Vergleich
- Datensatz soll bestätigt sein durch Menschen
  - Nicht nur Daten, die genau zum Vergleich passen
  - Allgemein gültig
    - Menschen können Ähnlichkeit gut feststellen

→ empirisch begründet

Testreihe mit Personen

Testdaten

	Nomen	Referenz	Hoch	Tief
Das	Pferd	läuft	gallopiert	zerläuft
Die	Farbe	läuft	zerläuft	gallopiert
Das	Feuer	glüht	brennt	strahlt
Das	Gesicht	glüht	strahlt	brennt
Das	Handy	brummt	vibriert	knurrt
Der	Bär	brummt	knurrt	vibriert

Testreihe mit Personen

Material und Design

- Zusammensetzung des Tests
  - Subjekt und intransitives Verb
- Daten aus dem British National Corpus(BNC)
- für jedes Subjekt-Verb Tupel

  → zwei Synonyme des Verbs
  - eins komaptible mit dem Referenz-Verb, eins nicht
    - aus WordNet
- Anfangs-Set bestand aus 20 Verben, jedes gepaart mit 10 Nomen und 2 Synonymen→ 400 Paare
  - Vortest: Getest, welche dieser Paare die größte Variation gezeigt haben → 120 Paare

Testreihe mit Personen

Ablauf und Aufgaben Beispiel

- Teilnehmer bekamen Satz-Paare
  - einen mit dem Referenz-Verb
  - einen mit einem Synonym
- Bewertung der Ähnlichkeit
  - auf einer Skala von 1 – 7

- Aufgaben-Beispiel

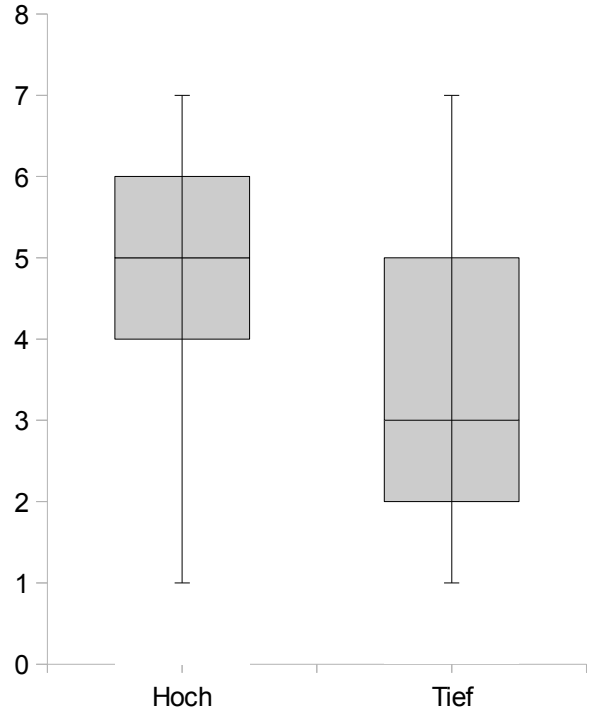
Das Feuer glüht.	Das Pferd rennt.
Das Feuer strahlt.	Das Pferd gallopiert.

- Teilnehmer-Zahl: 49 unbezahlte Freiwillige, Muttersprache Englisch

Testreihe mit Personen

Statistische Auswertung

- Wie zuverlässig sind die gesammelten Beurteilungen?
  - Verschiedene Tests
- Haben Teilnehmer semantisch korrekte Bewertungen gemacht?



- Sätze mit einem ähnlichen Synonym wurden als ähnlicher zum Referenz-Satz wahrgenommen und umgekehrt

Vergleich der Modelle

- Kontext-Fenster → 5 Wörter
- Vektor erstellt durch Kookkurrenz mit anderen Worten
- Auswertung der Modelle auf zwei Wegen
  - Kosinus-Ähnlichkeit zwischen Referenz-Satz und den Synonym-Sätzen
    - Erwartung: bessere Modelle kommen näher an die menschliche Bewertung ran
  - Spearmans ρ(rho)
    - Stellt den Zusammenhang von zwei Variabeln dar
    - Werte zwischen -1 und 1
      - je näher an 1 oder -1, desto größer der Zusammenhang

Vergleich der Modelle

- Drei Modelle, die zusätzliche Parameter haben
  - Gewichtete Addition
    - bestes Modell: 80% Verb und 20% Nomen
  - Kombination aus Addition und Multiplikation
    - bestes Modell: 95% Verb, 0% Nomen und 5% Multiplikation beider
  - Kintschs Modell
    - zwei extra Parameter
      - m Nachbarn, die am ähnlichsten zum Wort sind und davon k, die am ähnlichsten sind
    - m = 20 und k = 1

Resultat/Fazit

Modell	Hoch	Tief	ρ
BaseLine	0,27	0,26	0,08
Addition	0,59	0,59	0,04
Gew. Addition	0,35	0,34	0,09
Kintsch	0,47	0,45	0,09
Multiplikation	0,42	0,28	0,17
Kombination	0,38	0,28	0,19
Personentest	4,94	3,25	0,40

- Addition, gewichtete Addition, Kintsch
  - Keine Unterscheidung zwischen Hoch und Tief
  - Kein großer Unterschied zur Baseline

- BaseLine: Ähnlichkeit zwischen Referenz-Verb und dessen Synonymen
  - p = v

Resultat/Fazit

Fortsetzung

Modell	Hoch	Tief	ρ
BaseLine	0,27	0,26	0,08
Addition	0,59	0,59	0,04
Gew. Addition	0,35	0,34	0,09
Kintsch	0,47	0,45	0,09
Multiplikation	0,42	0,28	0,17
Kombination	0,38	0,28	0,19
Personentest	4,94	3,25	0,40

- Multiplikation und Kombination
  - sind näher an der menschlichen Bewertung
  - unterschied ist statistisch nicht signifikant
- Multiplikation hat im Vergleich zur Kombination keine freien Paramter und muss nicht optimiert werden

→ Fazit: Multiplikations-Modelle sind besser

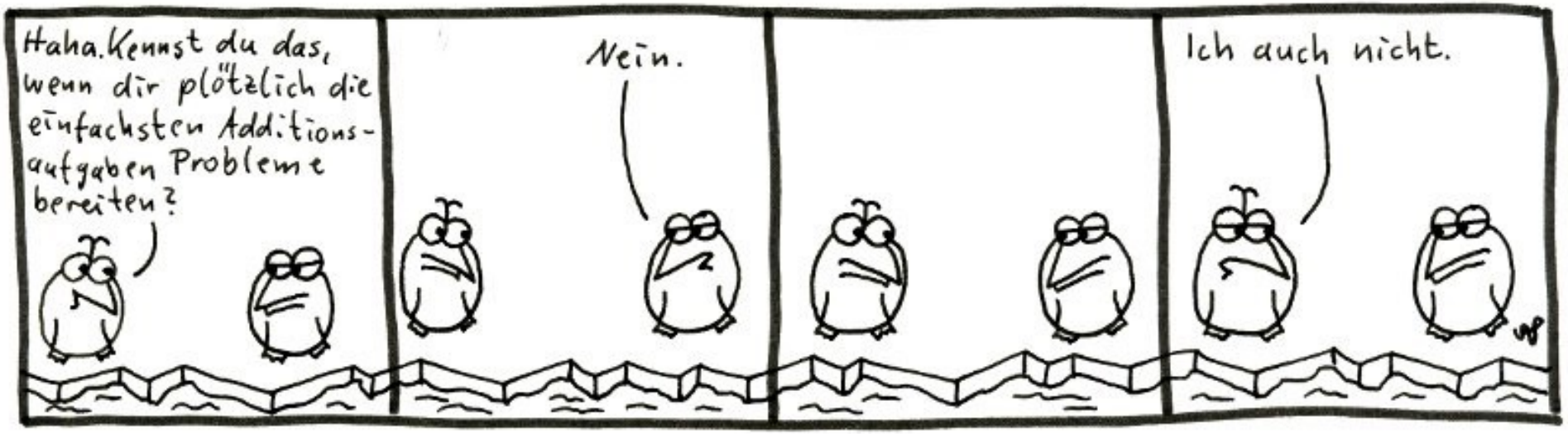
Ausblick

Was kann noch getan werden?

- Datensatz ausweiten
  - Größere Datensätze auswerten
  - Längere Phrasen/Sätze evaluieren
- Modelle verbessern
  - Multiplikation p = Cuv
    - Freien Parameter C genauer definieren

Kritik

- Je größer der Datensatz, desto mehr Null-Werte
  - Stimmt die Auswertung noch?
- Ein Vektor als Ergebnis ist zu wenig
  - Es kann nur ein kleiner Teil von den Informationen im Satz dargestellt werden
    - Aber: Sätze können beliebig lang sein, wichtige Informationen werden weggelassen
- Syntaktische Relationen werden ignoriert
  - Beispiel: a horse draws (das Pferd zieht etwas)
    draw a horse (zeichne ein Pferd)


Haha. Kennst du das, wenn dir plötzlich die einfachsten Additionsaufgaben Probleme bereiten?
Nein.
Ich auch nicht.

Vorstellung der Modelle

Zusammenfassung

	Vorteile	Nachteile	Formel
Addition	Einfach, häufige Verwendung	Keine Beachtung von Wortfolge oder Bedeutung	$p = u + v$
Gewichtete Addition	Syntax-Struktur ist bewusster	Zusätzliche Parameter	$p = \alpha u + \beta v$
Kintschs Modell	Sensibel für syntaktische Struktur	Zusätzliche Paramter	$p = u + v + \sum n$
Multiplikation	Hebt wichtige Inhalte hervor	Keine Beachtung von Wortfolge	$p = u * v$
Kombination Add. + Mul.	Syntax-Struktur, wichtige Inhalte, keine Nullwerte	Zusätzliche Parameter	$p = \alpha u + \beta v + \gamma uv$
BaseLine(NonComp)	Basis für Vergleiche	sehr simpel, keine Komposition	$p = v$