

# Dependency-Based Construction of Semantic Space Models ( Padó, Lapata 2007)

Distributionelle Semantik WS 11/12

21.11.2011

# Was ist das Problem?

- Wortbasierte Vektormodelle betrachten nur einzelne Wörter :

- Beispiel

Hans liebt Maria = Maria liebt Hans

→ Nur einzelne Wörter betrachtet

→ unterschiedlicher Bedeutung aber gleiche Berechnung

→ Syntaktisches Wissen fehlt

→ Syntaxbasiertes Vektormodell

# 1. Wortbasierte Modelle

- Lowe (2001)
- Betrachtet Kookkurrenz
- Korpus: a lorry might carry sweet apples
- T (Zielwörter) = {lorry, carry, sweet, fruit}
- B = {lorry, might, carry, sweet, apples}
- Zielwort durch 5-dimensionalen Vektor dargestellt

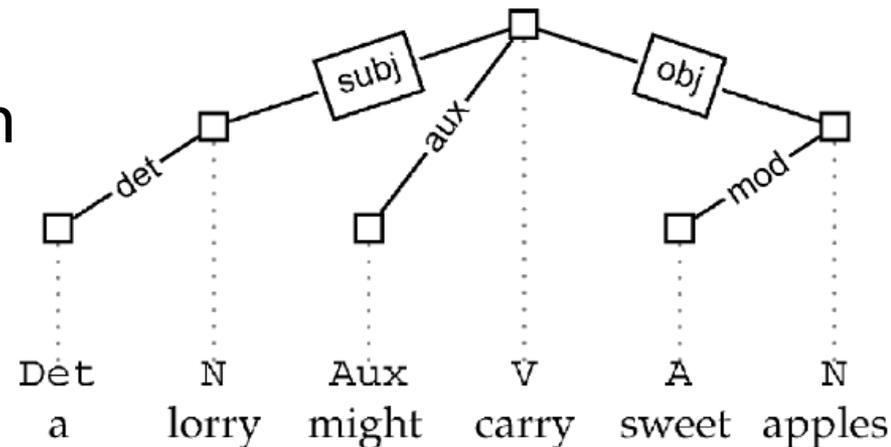
	lorry	might	carry	sweet	apples
lorry	0	1	1	0	0
carry	1	1	0	1	1
sweet	0	1	1	0	1
fruit	0	0	0	0	0

# Inhalt

- Überblick Syntaxbasierter Modelle
- Neues Modell vorgestellt und Vergleich zu vorangehenden Modellen
- Erklärung der verwendeten Parameter
- Priming Experiment
- TOEFL Synonym Experiment
- Sense-ranking Experiment

# Syntaxbasiertes Raummodell

- Grefenstette, Lin
- Mehr als Kookkurrenz :  
betrachten syntaktische Relationen  
zwischen Wörtern
- Basiselemente mit Tupel  $(r,w)$  = Attribut/Features  
Beispiel: (subj, Lorry)
- Betrachtet Dependenzgraph



# Beispiel

- Basiselemente= (subj, lorry), (aux, might),  
(mod, sweet), (obj, apples)

- A = binäre Verbindungsfunktion :

→ besteht Relation?

- Lin's Ähnlichkeitsmaß:

$$sim_{lin}(t_1, t_2) = \frac{\sum_{(r,w) \in T(t_1) \cap T(t_2)} I(t_1, r, w) + I(t_2, r, w)}{\sum_{(r,w) \in T(t_1)} I(t_1, r, w) + \sum_{(r,w) \in T(t_2)} I(t_2, r, w)}$$

Matrix bekommt numerischen Wert

(subj,lorry)

(aux, might)

(mod, sweet)

(obj, apples)

carry

1

1

0

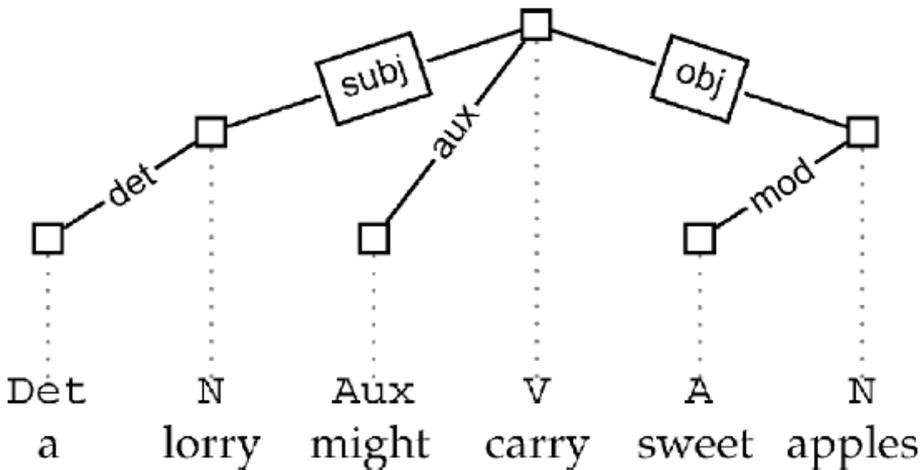
1

# Vergleich der Modelle

- Vergleiche der Modelle an der selben Anwendung sind sehr selten
- Grefenstette : Vergleich automatischer Thesaurus Extraktion
  - Syntaxbasiertes Modell liefert bessere Resultate
- Aber: Information Retrieval weniger erfolgreich
- Allgemein: Syntaxbasierte Modelle unterstützen Analyse für Satzstruktur und Bedeutung
  - gebräuchlicher als Wortbasierte Modelle

# Rahmen für Semantisches Raummodell

## ■ Dependenzgrammatik



Dependenzkennung:

- a [ det ] lorry
- lorry [ subj ] carry
- might [ aux ] carry
- apples [ obj ] carry
- sweet [ mod ] apples

## ■ Hauptverb: carry

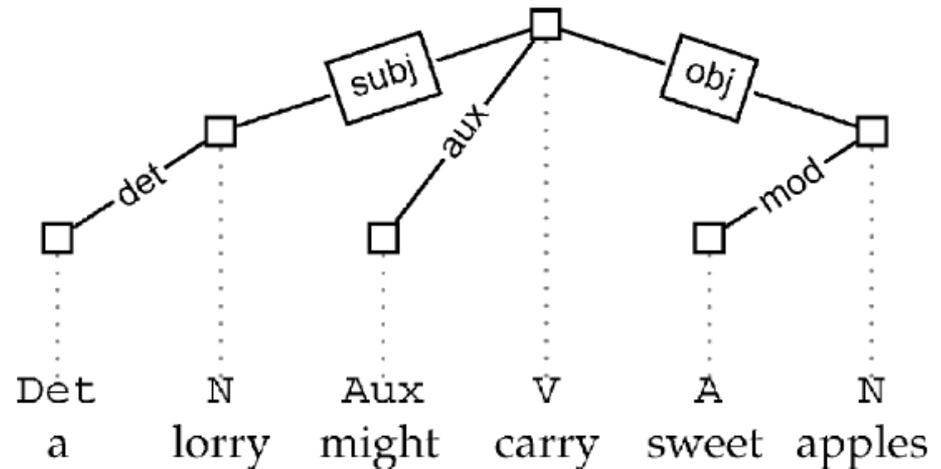
## ■ Bedeutung des Satzes wird durch Syntax deutlich:

Ein Lastwagen(**Subjekt**) könnte(Hilfsverb) süße Äpfel(**Objekt**)  
tragen(**Prädikat**)

# Semantisches Raummodell: 3 Parameter

## 1) Auswahlfunktion (context selection function):

- Entscheidet welche Pfade zur Repräsentation des Zielwortes  $t$  beitragen
  - Minimum (Länge 1)
  - Medium (Länge  $\leq 3$ )
  - Maximum (Länge  $\leq 4$ )



# Semantisches Raummodell: 3 Parameter

## 2) Wertungsfunktion (path value function) $v$

- weist Pfaden Gewichtung zu
  - Plain  
gleicher Wert für jeden Pfad
  - Length  
kleiner Pfad bekommt höheren Wert
  - Gram-rel  
Pfade nach Hierarchie gewertet:  
Subjekt > Objekt > Obliques > Genitiv > sonstige

# Semantisches Raummodell: 3 Parameter

## 3) Basiselementfunktion (basis mapping function) $\mu$

- Auswahl der Basis
- Können Wörter, syntaktische Relationen oder Kombination beider sein

# Evaluation der Parameter

## Mit Cosinusähnlichkeitsmaß

	plain	length	gram-rel
minimum	0,45	0,45	0.43
medium	0,45	0,45	0,44
maximum	0,47	0,46	0,45

## Mit Lin's Ähnlichkeitsmaß

	plain	length	gram-rel
minimum	0,58	0,58	0.58
medium	0,60	0,62	0,59
maximum	0,56	0,59	0,55

# Semantisches Raummodell: 3 Parameter

## Mit Cosinusähnlichkeitsmaß

	plain	length	gram-rel
minimum	0,45	0,45	0.43
medium	0,45	0,45	0,44
maximum	0,47	0,46	0,45

## Mit Lin's Ähnlichkeitsmaß

	plain	length	gram-rel
minimum	0,58	0,58	0.58
medium	0,60	<b>0,62</b>	0,59
maximum	0,56	0,59	0,55

## Optimales Dependenzmodell:

- Medium  
Auswertungsfunktion
- Länge basierte  
Wertungsfunktion
- Lin's Ähnlichkeitsmaß
- Basiselemente =  
2000 häufigste Wörter

# Semantisches Raummodell: 3 Parameter

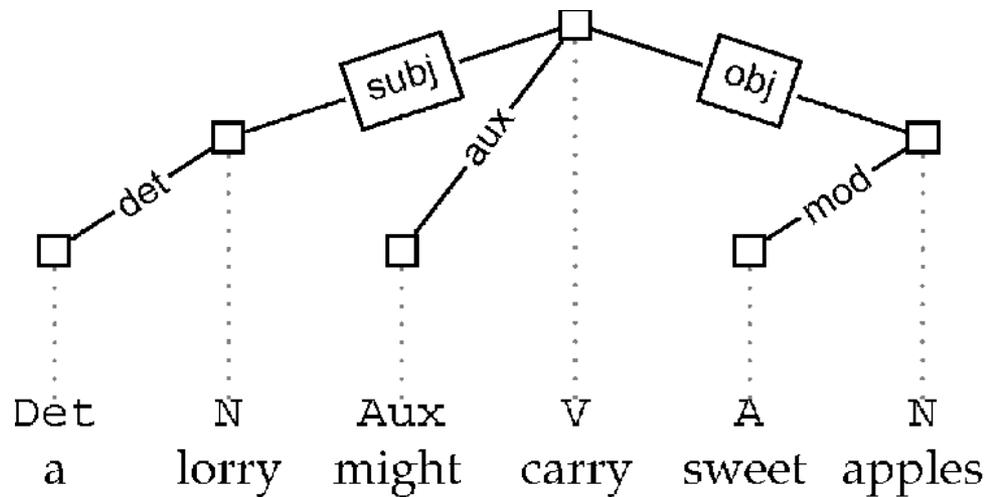
- Durch die 3 Parameter erhält man ein Gerüst für vektorbasierte Modelle
- 1. Schritt: Erstellung des Kontextes
- 2. Schritt: Erstellung der Dimensionen
- 3. Schritt: Wertung der Abhängigkeitspfade
- Letzter Schritt:  
Konstruktion des semantischen Raums

# Semantisches Raummodell: Erstellung des Kontext

- 1. Schritt: Definition eines angebrachten Kontextes für Zielwörter
- Kontext = verankerte Pfade im Dependenzgraphen
- Vermutung: Menge der verankerten Pfade von T ist eine Teilmenge der Pfade, die wichtige Verteilungsinformationen über T beinhalten
- Teilmenge = syntaktischer Kontext

# Semantisches Raummodell: Erstellung des Kontext

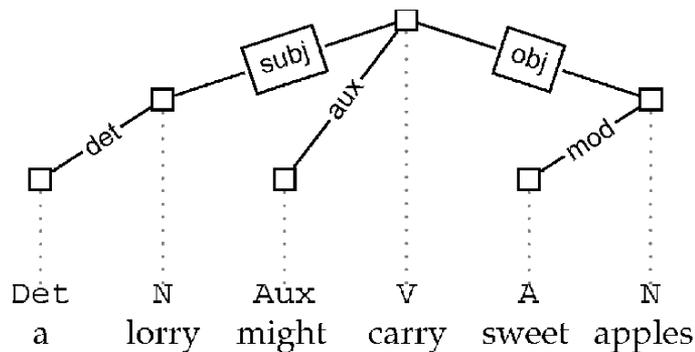
## ■ Verankerte Pfade:



- $\langle \text{lorry, carry} \rangle, \langle \text{lorry, a} \rangle = 2$  Pfade der Länge 1
- $\langle \text{lorry, carry, apples} \rangle, \langle \text{lorry, carry, might} \rangle = 2$  Pfade der Länge 2
- $\langle \text{lorry, carry, apples, sweet} \rangle = 1$  Pfad der Länge 3

# Semantisches Raummodell: Erstellung des Kontext

- Man kann den syntaktischen Kontext auf die Pfade begrenzen, die spezielle Eigenschaften aufweisen.
- Beispiel: bestehen nur aus Subjekt und Objekt



→ <lorry, carry> und <lorry, carry, apples>  
fallen weg: <lorry, **a**>, <lorry, carry, **might**>  
<lorry, carry, apples, **sweet**>

# Semantisches Raummodell: Bestimmung der Dimensionen

- 2. Schritt: Bestimmung der Basiselemente
- Basiselementfunktion erlaubt es, zugrunde liegende Beziehungen verschiedener Pfade zu erschließen
- Äquivalente Pfade können auf dasselbe Basiselement abgebildet werden
  - Pfade lassen sich in äquivalente Klassen (durch Basiselement bezeichnet) aufteilen

# Semantisches Raummodell: Wertung der Abhängigkeitspfade

- 3. Schritt: Spezifikation des relativen Wertes der verschiedenen Pfade
- Definition der lokalen Kookkurrenz Frequenz  
= Summe aller Pfadwerte im syntaktischen Kontext
- Globale Kookkurrenz Frequenz  
= Summe aller lokalen Kookkurrenz Frequenzen  
→ Messung für Kookkurrenz im gesamten Korpus

# Semantisches Raummodell: Konstruktion des semantischen Raumes

*A lorry might carry sweet apples*

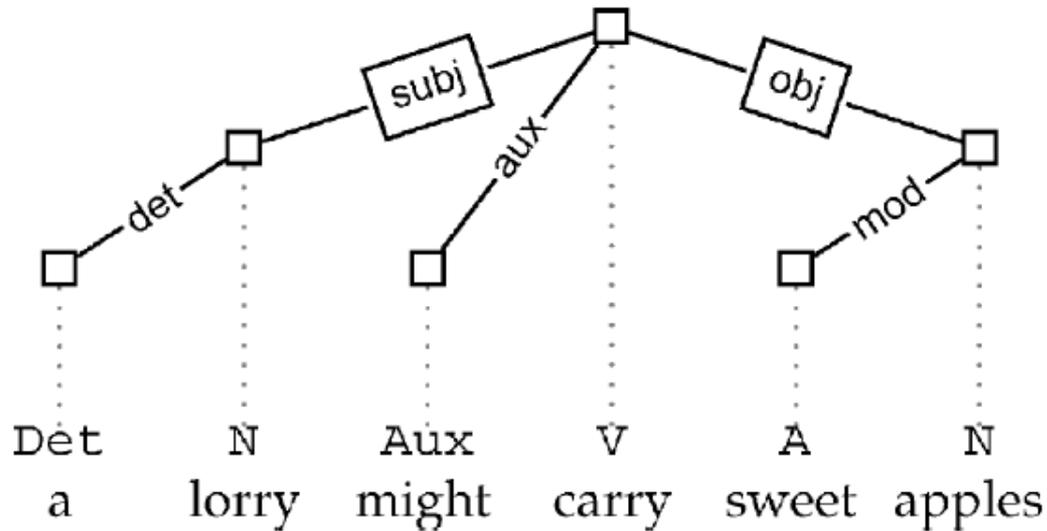
- Um Vektoren für die Zielwörter

$T = \{ \text{lorry, might, carry, sweet, fruit} \}$

darzustellen benutzt man die 3 Parameter :

- Auswahlfunktion, die nur Subjekt und Objekt betrachtet
- Länge basierte Wertefunktion
- Wortbasierte Basiselementfunktion

# Semantisches Raummmodell: Konstruktion des semantischen Raumes



	lorry	might	carry	sweet	apples
lorry	0	0	1	0	0,5
migth	0	0	0	0	0
carry	1	0	0	0	1
sweet	0	0	0	0	0
fruit	0	0	0	0	0

# Zwischenfazit

- Konstruktion eines dependenzbasierenden semantischen Raummodells
- Einbeziehung von linguistischem Wissen
- Flexibles Modell:
  - Abhängig von Parameter kann der Raum über Wörter, syntaktische Relationen oder einer Kombination dieser bestehen
- Präsentation indirekter Dependenzrelationen

# Zwischenüberischt

- Wortbasierte Modelle betrachten nur einzelne Wörter
  - Syntaxbasierte Modelle betrachte syntaktische Informationen der Wörter
- Experimente um die Ergebnisse der beiden Modelle zu vergleichen

# Experimente

- Erfolg von dependenzbasierten Modellen in der Praxis?
  - Bessere Ergebnisse als Wortbasierte Modelle?
- Verschiedene Experimente durchgeführt

# Experimente: Korpus und Parser

- British National Corpus
  - 100 Mio. Wörter
  - geschriebenes + gesprochenes Englisch
  - Zeitschriften, Zeitungen, Briefe, etc.
- Um Abhängigkeitsrelationen zu konstruieren wird BNC mit MINIPAR geparst

# Experimente: Parameter

- Dependenzkontext definiert über 14 Dependenzrelationen
- Jede über 500,000 mal in BNC
- Amod (Adjektiv Modifikation), fc (finite Ergänzung), rel (Relativsatz), etc.

# Experimente

## 2 Hypothesen:

- 1. Das dependenzbasierte Modell kann
  - Einzelwort Priming (Single-Word Priming)
  - Synonymerkennung
  - Sense Rankingsimulieren
- 2. liefert bessere Ergebnisse als übliche wortbasierte Modelle

# 1. Experiment: Einzelwort Priming

- konzentriert auf Semantische Ähnlichkeit bzw. Unähnlichkeit zwischen Wörter
- Die Präsentation eine Hauptwortes erleichtert Aussprache oder lexikalische Entscheidung über ein Zielwort

Tiger Löwe vs. Tiger Wand

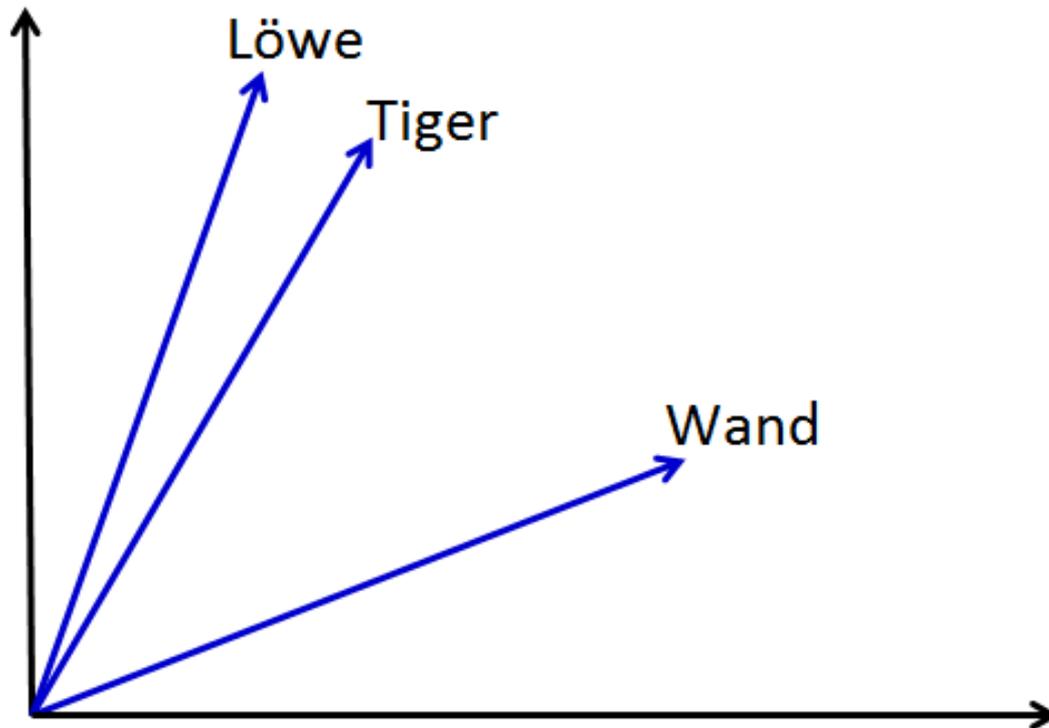
→ Antworten schneller und genauer, wenn Wörter semantisch verwandt ( = **priming effect** )

# 1. Experiment: Einzelwort Priming

- Modellierung der Lesezeit für Haupt- und Zielwort für ihre Distanz im semantischen Raum
- Priming Effekt durch Vergleich der Distanzen
- 500 häufigsten Wörter des gesprochenen Teils des BNC

# 1. Experiment: Einzelwort Priming

- Annahme: Distanz niedriger bei verwandten Haupt- und Zielwörter



# 1. Experiment: Einzelwort Priming

- $\eta^2$  = Maß zum Messen der Erfolgsintensität  
(wie gut simuliert Modell Mensch?)
- Vergleich Dependenz Modell und Wortbasiertes Modell
  - Dependenzbasiertes Modell :  $\eta^2 = 0,332$
  - Wortbasiertes Modell :  $\eta^2 = 0,284$
- Hypothese bestätigt

## 2 . Experiment: Synonymerkennung

- TOEFL ( = Test of English as a Foreign Language)
- Testet Wissen von nicht-Muttersprachler bzgl. der englischer Sprache

You will find the office at the main **intersection**  
a) place *b) crossroad* c) roundabout d) building

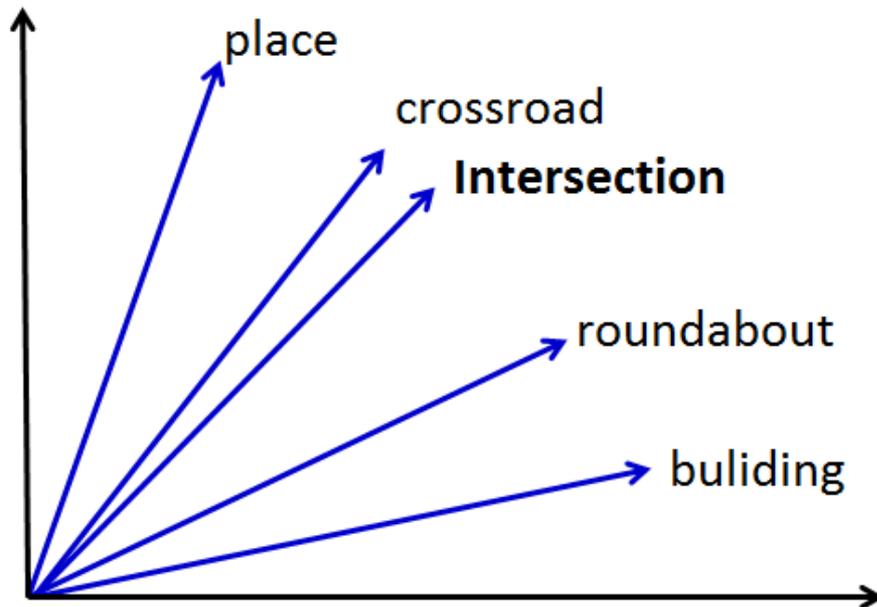
- Zielwort mit möglichen Synonymen gegeben

## 2 . Experiment: Synonymerkennung

- TOEFL Datenset
- Abstand zwischen Vektor des Zielwortes und den jeweiligen Vektoren der möglichen Synonymen betrachtet
  - kleinster als Synonym gewählt
- Genauigkeit = Prozentsatz der richtigen Antworten

## 2 . Experiment: Synonymerkennung

- Vektorbasiertes Modell soll richtiges Synonym finden



- Synonym: crossroad , da kleinste Distanz

## 2 . Experiment: Synonymerkennung

Vergleich der Modelle

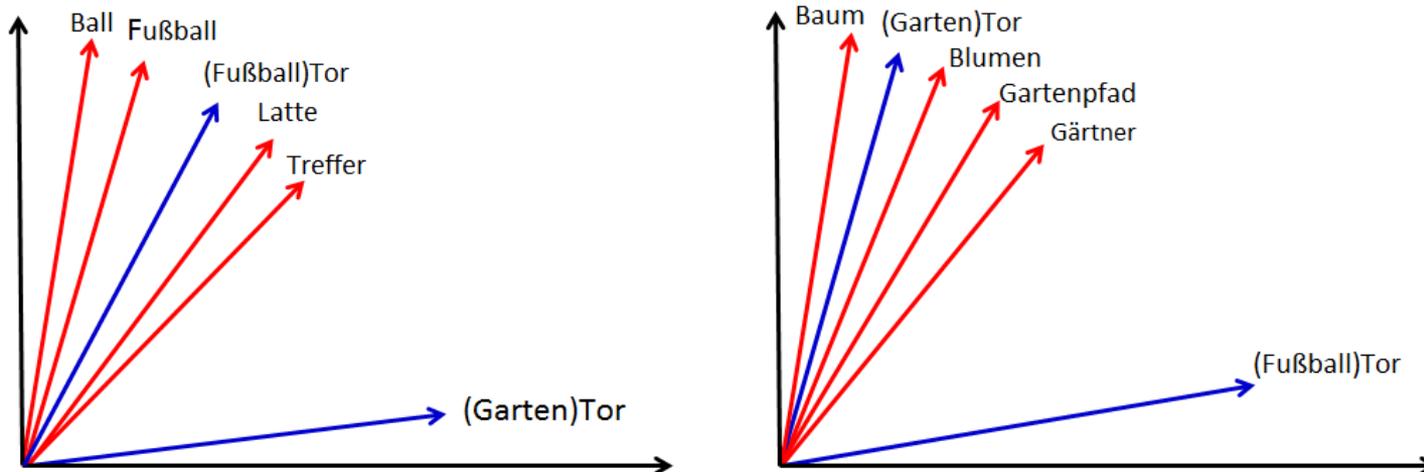
Wortbasiertes Modell : Genauigkeit = 61,3 %

Abhängigkeitsbasiertes Modell: Genauigkeit = 73,0 %

→ Hypothese bestätigt

# 3. Experiment: Sense Ranking

- Quantifizierung der einzelnen Ähnlichkeitsgrade zwischen ähnlichen Nachbarn und der Bedeutungen eines mehrdeutigen Wortes
- Modell wählt Bedeutung mit geringster Distanz
- Ähnlichste Bedeutung = 1. Bedeutung (first sense)



# 3. Experiment: Sense Ranking

Predominant sense score

Methode:

$$PS(ws_i) = \sum_{n_j \in N(w)} sim_{distr}(w, n_j) \times \frac{sim_{sem}(ws_i, n_j)}{\sum_{ws_{i'} \in S(w)} sim_{sem}(ws_{i'}, n_j)}$$

- „Predominant sense score“ um first sense aller mehrdeutigen Nomen aus SemCor zu berechnen
- (SemCor = Untermenge des Brown Korpus, die mit Wortbedeutung annotiert ist)
- Nur Nomen mit Ambiguitätsrate  $> 2$  ( $\rightarrow$  2750 Nomen)
- Durchschnittliche Bedeutungsambiguität = 4,55
- Optimale Dependenzmodell
- 50 Nachbarn betrachtet

# 3. Experiment: Sense Ranking

Auswertung Sense Ranking

- Identifizierung der prädominanten Bedeutung für ein Wort

$$Acc_{sr} = \frac{|\{w \in W_{ps} \mid ps_{sc}(w) = ps_r(w)\}|}{|W_{ps}|}$$

Auswertung WSD-Potenzial (Wortbedeutung Disambiguierung)

- Man ordnet jedem Nomen die prädominante Bedeutung zu
- Dann misst man das Verhältnis der Token, deren prädominante Bedeutung mit denen der SemCor identisch ist

$$Acc_{wsd} = \frac{\sum_{w \in W} |\{t \in T(w) \mid ps_r(w) = sense_{sc}(t)\}|}{\sum_{w \in W} |T(w)|}$$

# 3. Experiment: Sense Ranking

McCarthy et al. :

- Abhängigkeitspfad der Länge 1
- Betrachtet nur kleine Menge an Abhängigkeitsrelationen (Verb – Subjekt, Verb-Objekt, Nomen-Nomen, Adjektiv-Nomen)
- Lin's Ähnlichkeitsmaß

# 3. Experiment: Sense Ranking

Ergebnis:

Modelle	$ACC_{sr}$	$ACC_{wsd}$
Wortbasiertes Modell	49,3	49,9
Dependenzbasiertes Modell	54,3	54,3
McCarthy	54,0	46,0

→ Hypothese bestätigt

# Zusammenfassung

- Wortbasierte Modelle betrachten nur einzelne Wörter
- Syntaxbasierte Modelle betrachte syntaktische Informationen der Wörter
- Experimente zeigen, dass syntaxbasierte Modelle die besseren Ergebnisse liefern

# Fazit

- Erfolgreicher Rahmen für Vektorbasiertes Modell aufgestellt
  - Einbeziehung von Linguistischem Wissen und Abhängigkeitsrelationen
    - Dependenzbasiertes Modell arbeitet besser als Wortbasiertes

Danke für die Aufmerksamkeit!

Fragen?