

Multi-Prototype Vector-Space Models of Word Meaning (2010)

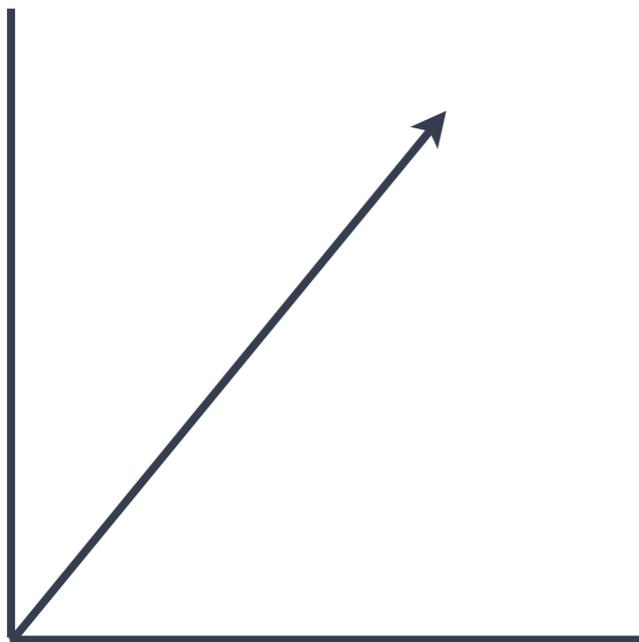
Joseph Reisinger & Raymond J. Mooney

Vortrag von Jonathan Poitz
Distributionelle Semantik - 12.12.2011

Vektorraummodelle

Kleine Wiederholung

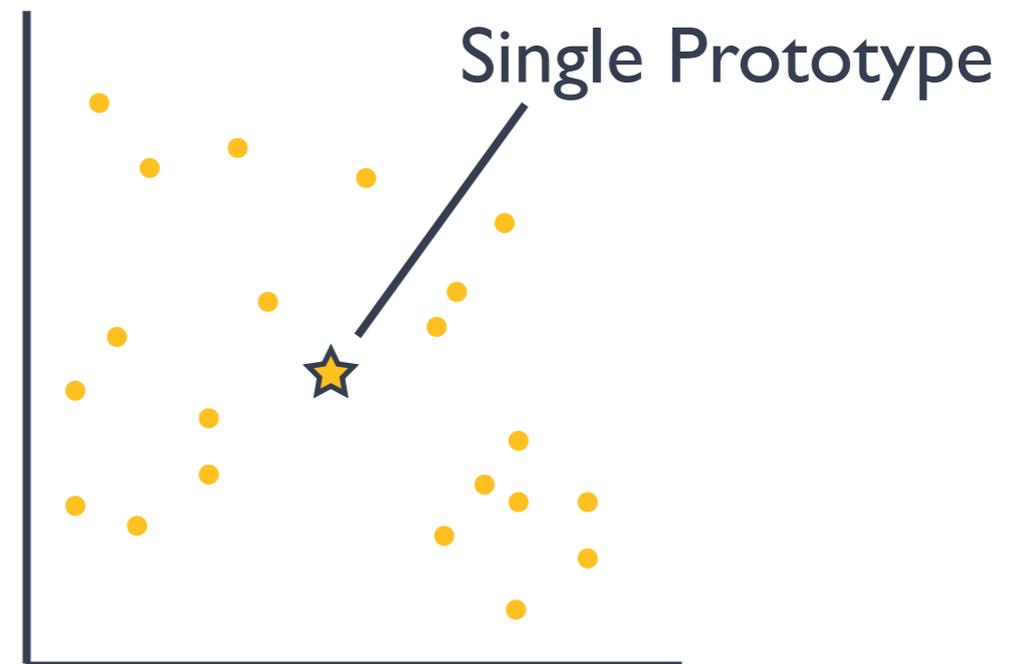
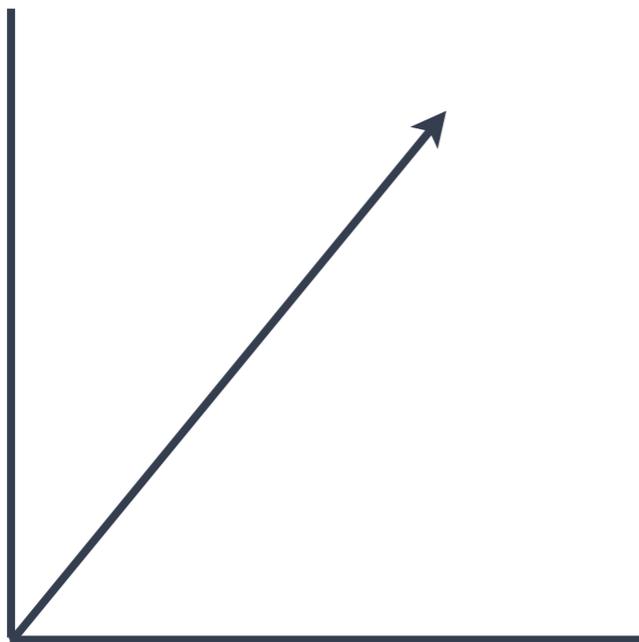
- Erstellung eines Vektors für ein Wort durch Kookkurrenzeninformationen aus Korpus (meist 10-Wort-Fenster oder Satzkontext)
- Vektor wurde generiert aus allen Auftreten des Wortes aus Korpus
- Keine Bedeutungsunterscheidung



Vektorraummodelle

Kleine Wiederholung

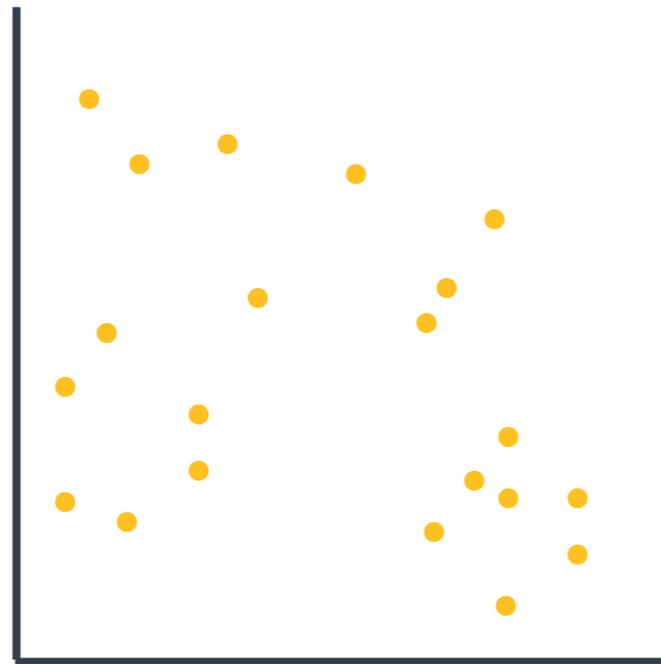
- Erstellung eines Vektors für ein Wort durch Kookkurrenzzinformationen aus Korpus (meist 10-Wort-Fenster oder Satzkontext)
- Vektor wurde generiert aus allen Auftreten des Wortes aus Korpus
- Keine Bedeutungsunterscheidung
- Diesen Vektor kann man auch Single Prototype nennen



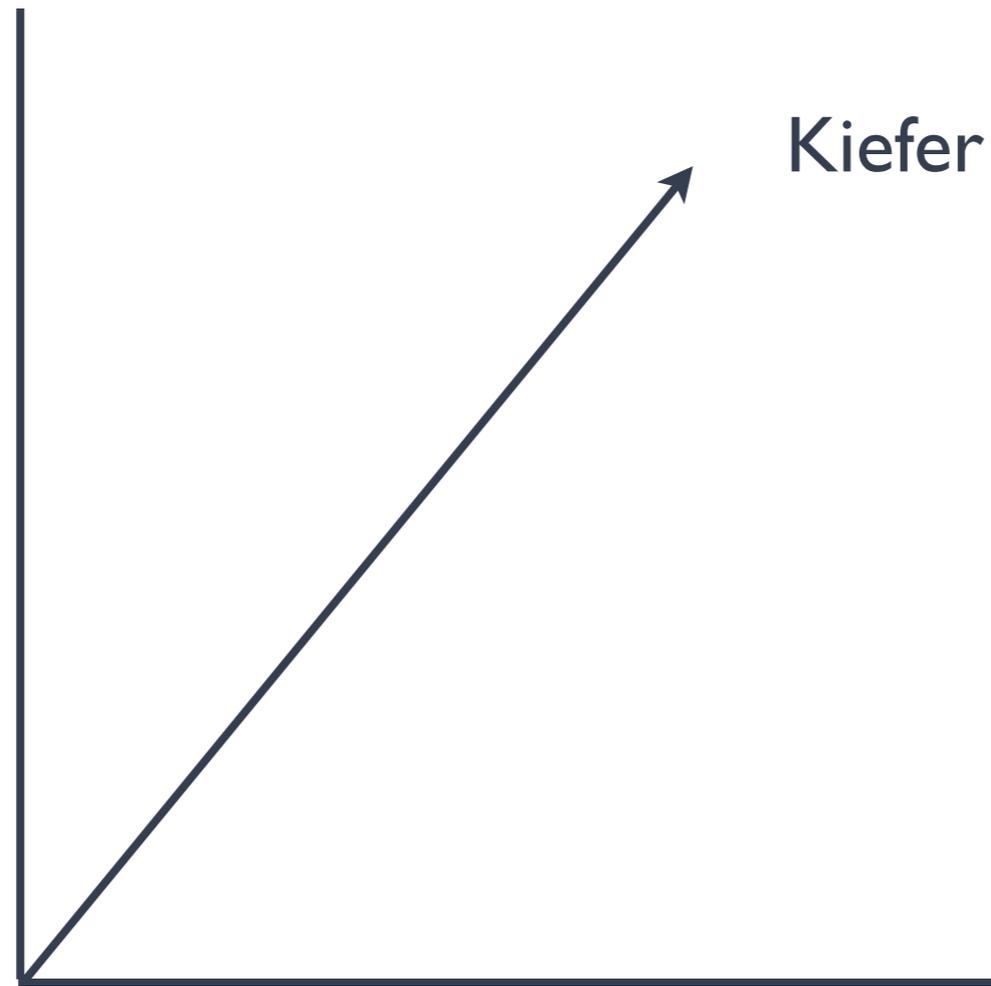
Am Rande ...

Exemplarmodelle

- Das komplette Gegenteil von den bisher gesehenen Modellen sind exemplarbasierte Modelle
- Jedes Auftreten des Wortes wird betrachtet!
- Exemplarmodell: Bei Ähnlichkeitsberechnung Vergleich aller Auftreten miteinander



Was haben wir bisher gesehen?



Was haben wir bisher gesehen?

Mitchell & Lapata '08

- Kompositionsmodell



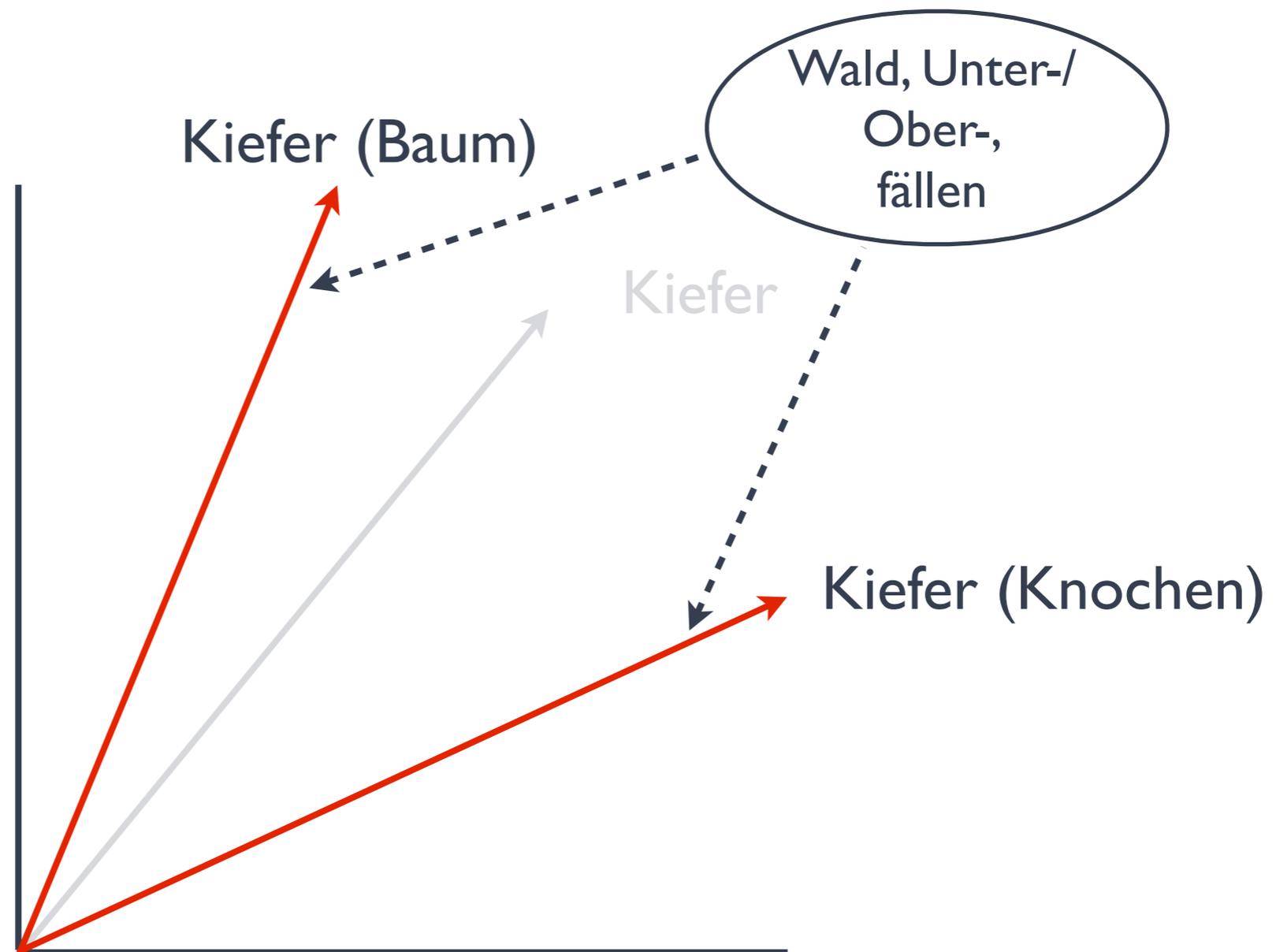
Was haben wir bisher gesehen?

Mitchell & Lapata '08

- Kompositionsmodell

Erk & Padó '08

- SelPref



Was haben wir bisher gesehen?

Mitchell & Lapata '08

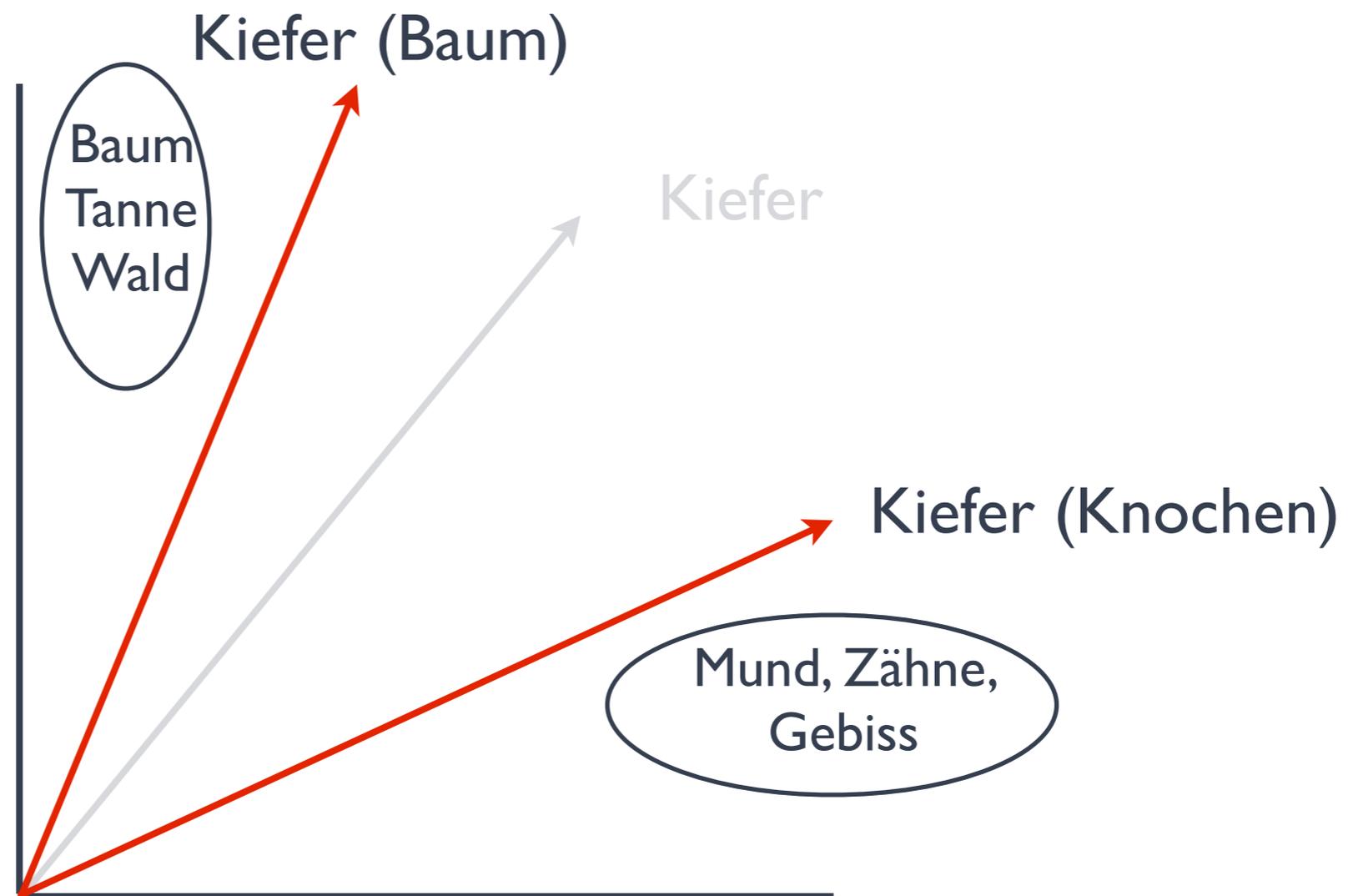
- Kompositionsmodell

Erk & Padó '08

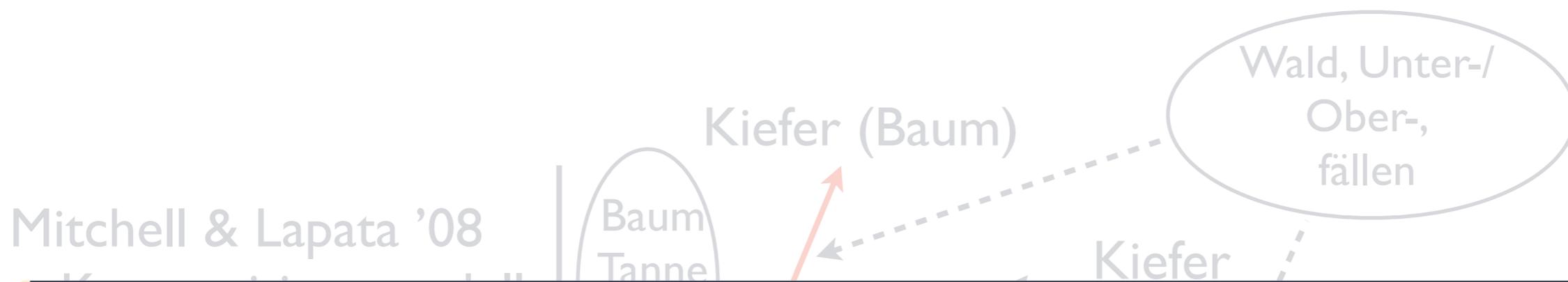
- SelPref

T, F & P '11

- „Kontextwolken“



Was haben wir bisher gesehen?

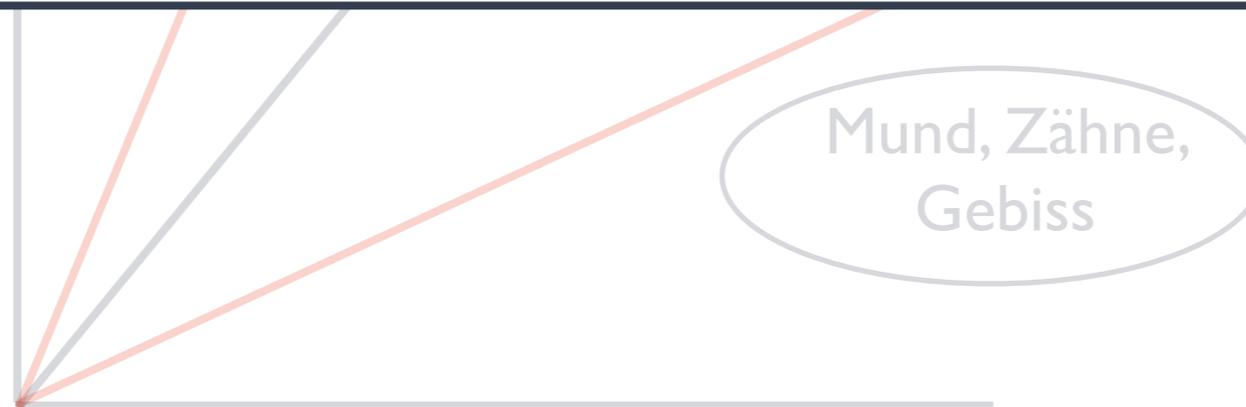


ABER:

- Zunächst: Generierung eines „unsauberen“ Vektorraumes
- Mehrere Wortbedeutungen in einem Vektor
- Wird erst später durch bestimmte Methoden bereinigt

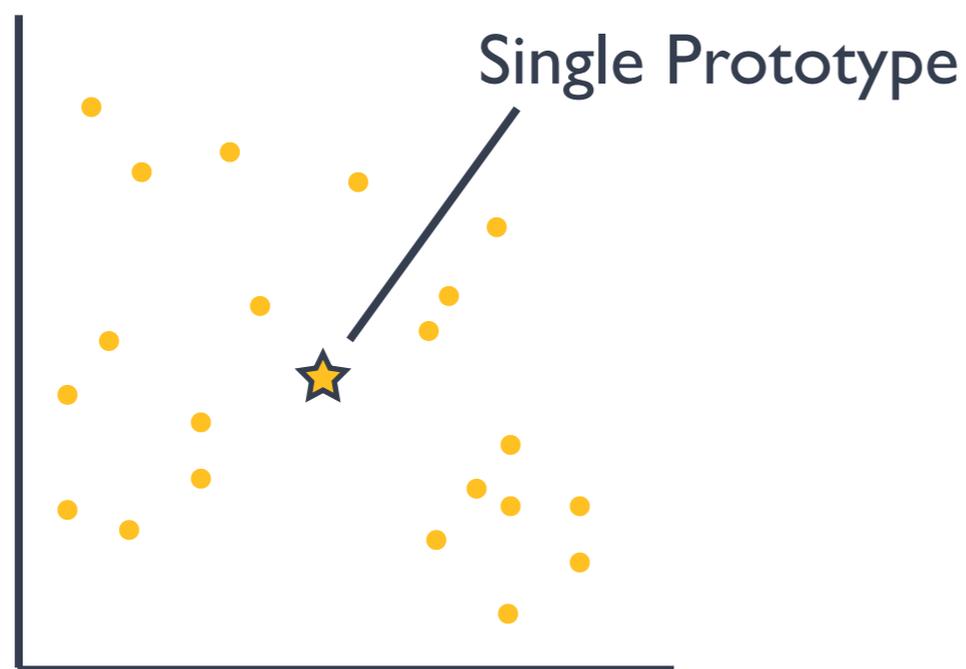
T, F & P '11

- „Kontextwolken“



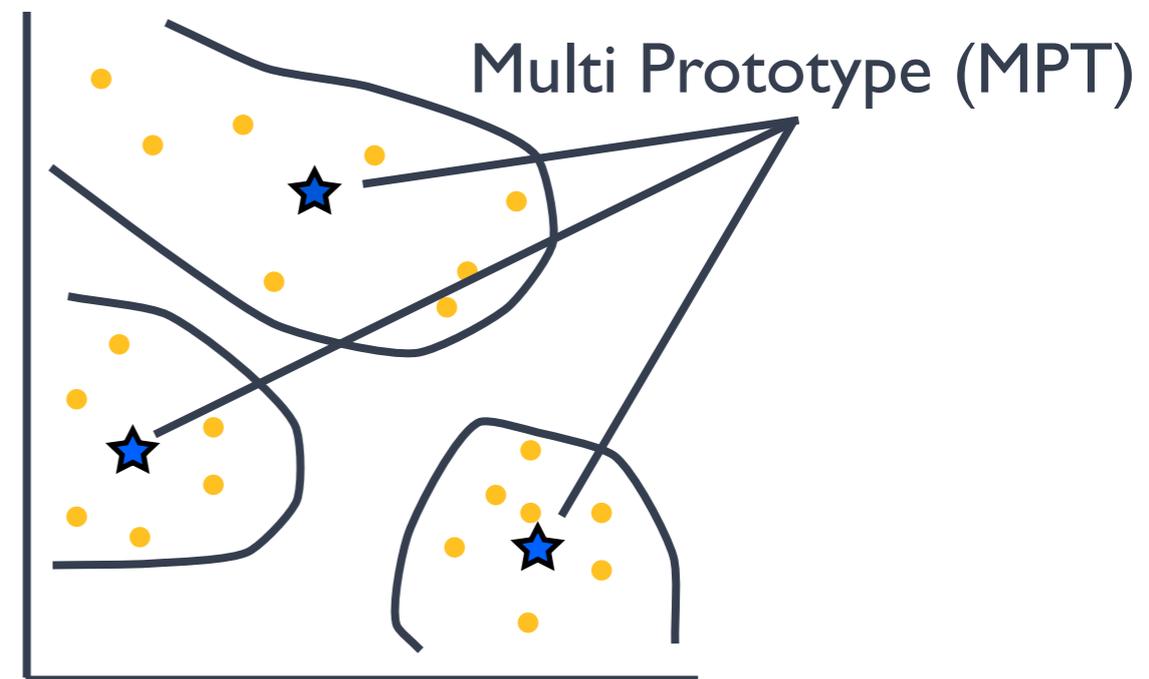
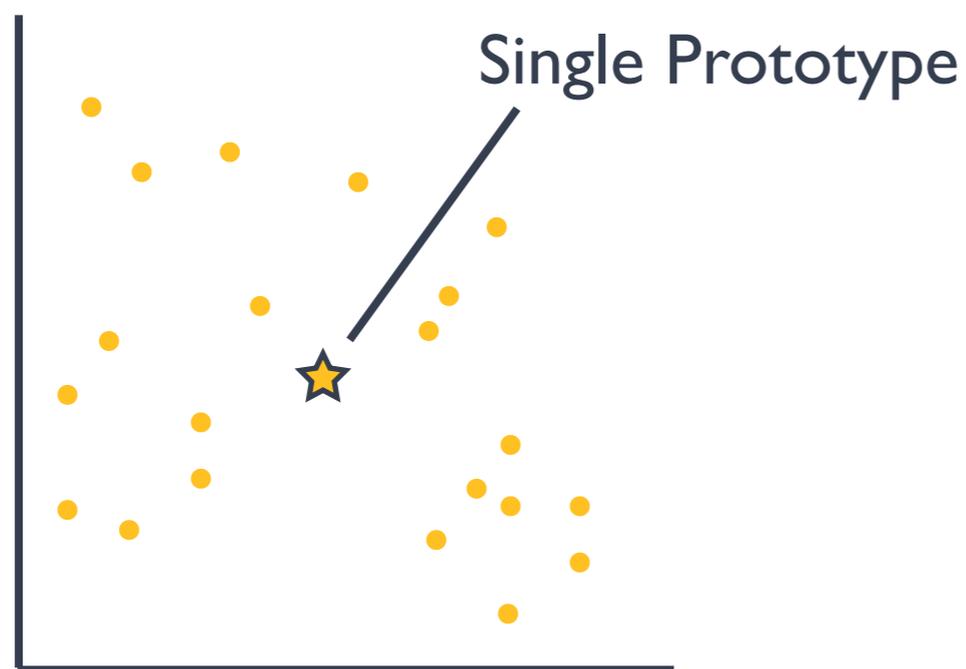
Eine neue Idee!

- Multi-Prototype-Modell versucht dies von Anfang an zu umgehen
- Eine neue Idee: Wort durch Menge distinktiver Vektoren dargestellt
- „Sauberer“ Vektorraum, wo Mehrdeutigkeit leichter modellierbar sein sollte



Eine neue Idee!

- Multi-Prototype-Modell versucht dies von Anfang an zu umgehen
- Eine neue Idee: Wort durch Menge distinktiver Vektoren dargestellt
- „Sauberer“ Vektorraum, wo Mehrdeutigkeit leichter modellierbar sein sollte



Wie geht das?

Mit Clustering!

- Clustering?

- Cluster - Gruppe von Objekten mit ähnlichen Eigenschaften
- Clustering - ordnet jedem Objekt ein Cluster zu

- Unterscheidung weiches und hartes Clustering

- keine klare Einteilung in Gruppen, sondern differenziert mit Wahrscheinlichkeiten

- strikte Einteilung in Gruppen

Der Algorithmus

Single-Prototype

... it is to have people **key** in, first of all, on the issues ...

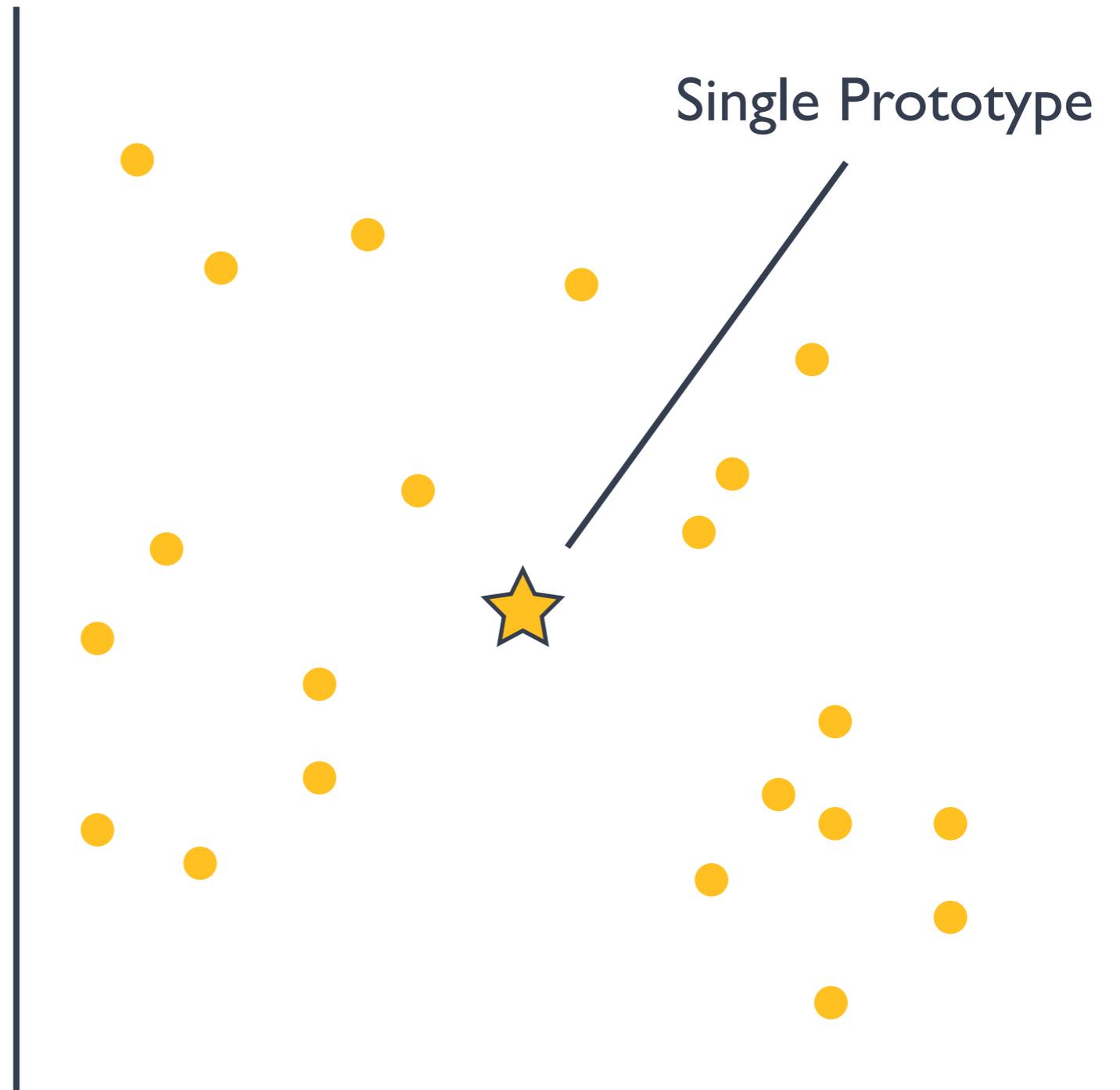
...And people may want to **key** in especially on things that ...

...Well, I think that's a **key** piece. Sharon is going to talk ...

... only a few apply the **key** points to their classrooms ...

... person who is considered the **key** -- the biggest opposition leader? And ...

... you please get me the **key** to my office here please ...



Der Algorithmus

Single-Prototype

... it is to have people **key** in, first of all, on the issues ...

...And people may want to **key** in especially on things that ...

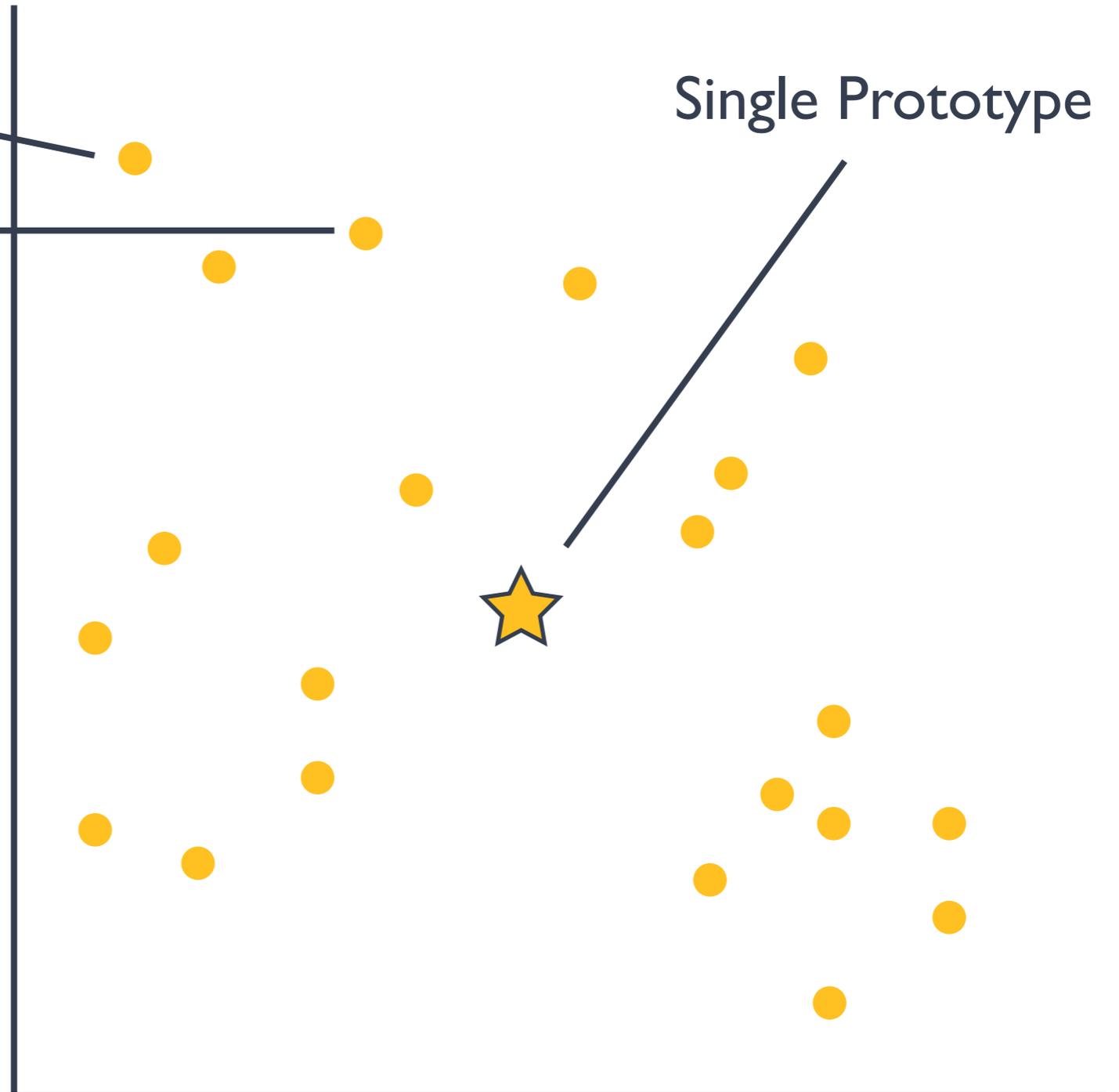
...Well, I think that's a **key** piece. Sharon is going to talk ...

... only a few apply the **key** points to their classrooms ...

... person who is considered the **key** -- the biggest opposition leader? And ...

... you please get me the **key** to my office here please ...

Single Prototype



Der Algorithmus

Single-Prototype

... it is to have people **key** in, first of all, on the issues ...

...And people may want to **key** in especially on things that ...

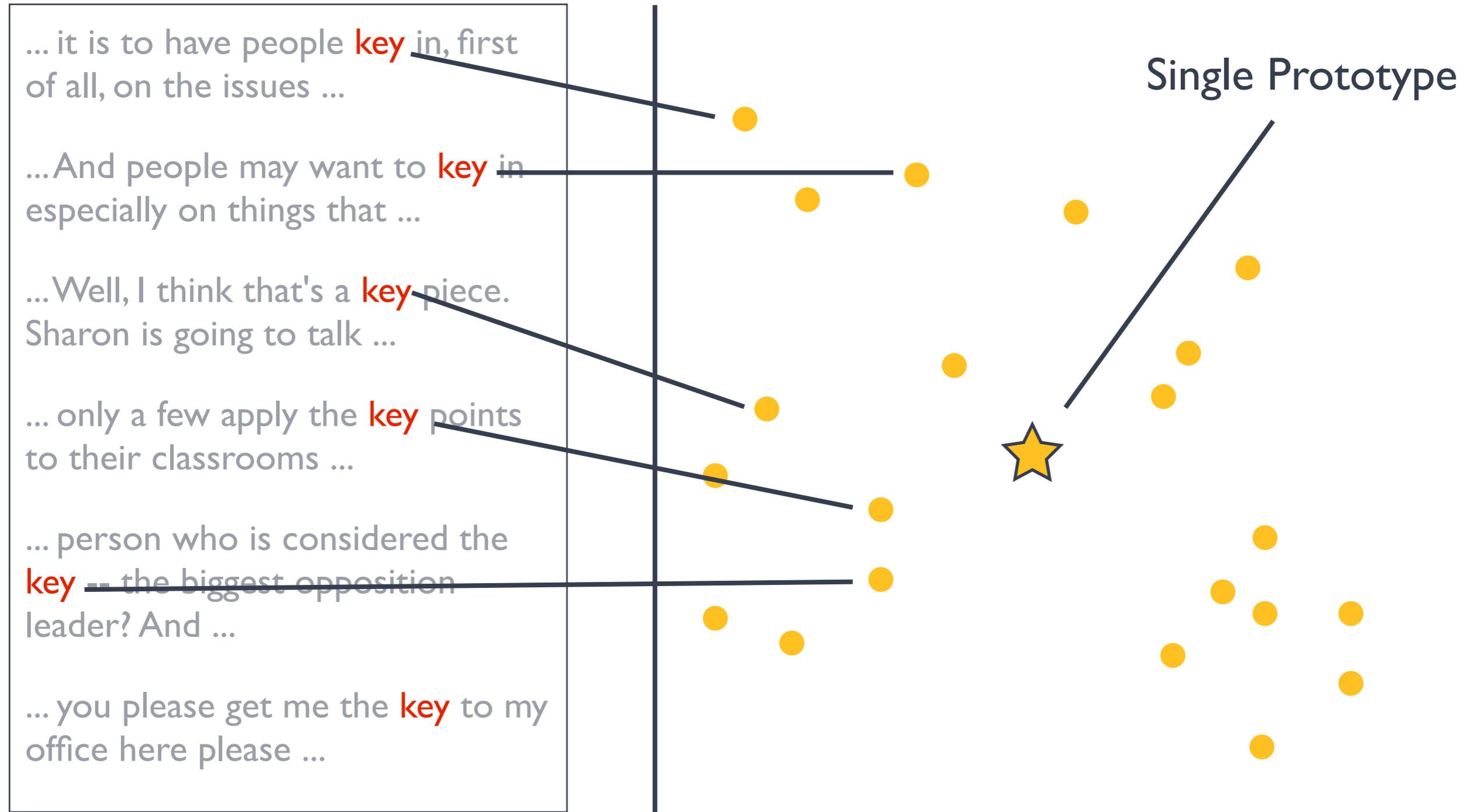
...Well, I think that's a **key** piece. Sharon is going to talk ...

... only a few apply the **key** points to their classrooms ...

... person who is considered the **key** -- the biggest opposition leader? And ...

... you please get me the **key** to my office here please ...

Single Prototype



Der Algorithmus

Single-Prototype

... it is to have people **key** in, first of all, on the issues ...

...And people may want to **key** in especially on things that ...

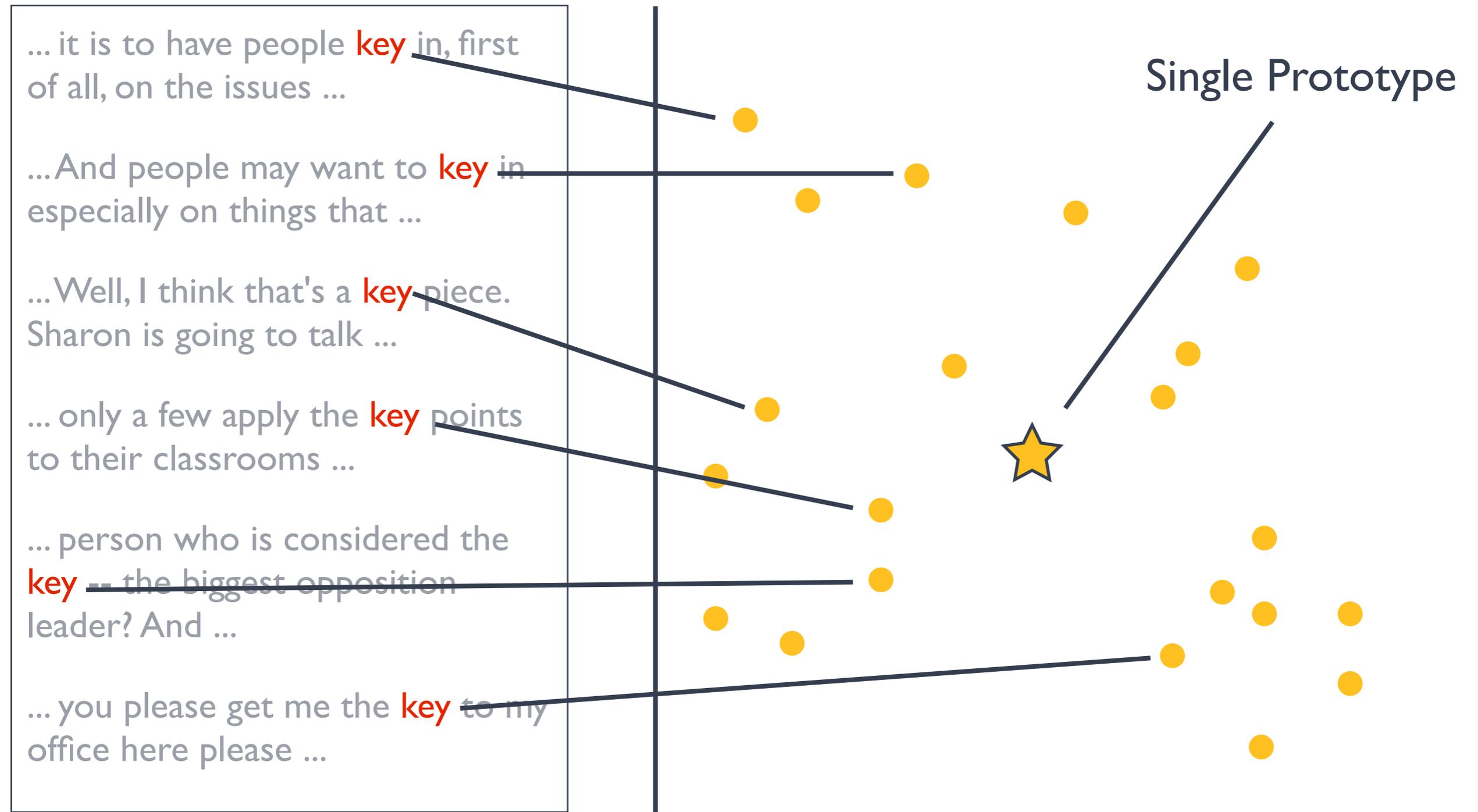
...Well, I think that's a **key** piece. Sharon is going to talk ...

... only a few apply the **key** points to their classrooms ...

... person who is considered the **key** -- the biggest opposition leader? And ...

... you please get me the **key** to my office here please ...

Single Prototype



Der Algorithmus

Multi-Prototype durch Clustering

... it is to have people key in, first of all, on the **key** issues ...

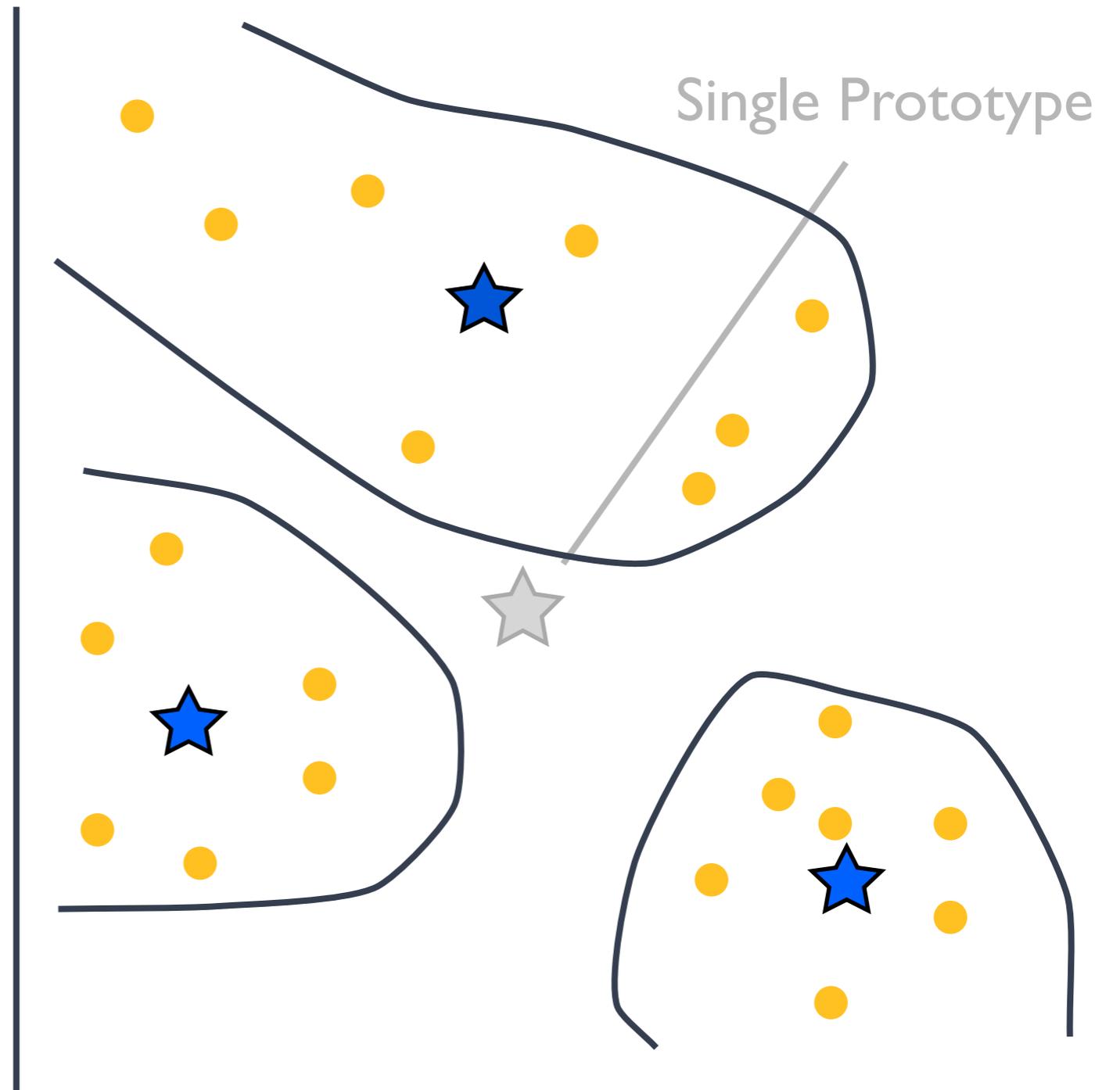
...And people may want to **key** in especially on things that ...

...Well, I think that's a **key** piece. Sharon is going to talk ...

... only a few apply the **key** points to their classrooms ...

... person who is considered the **key** -- the biggest opposition leader? And ...

... you please get me the **key** to my office here please ...



Der Algorithmus

Clustering



... it is to have people key in, first of all, on the key issues ...

...And people may want to key in especially on things that ...

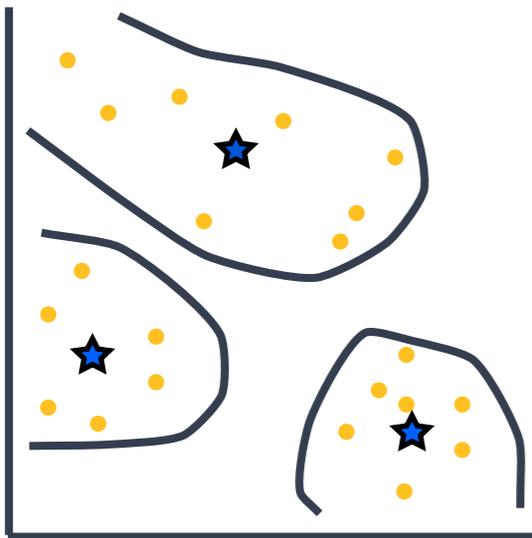
...Well, I think that's a key piece. Sharon is going to talk ...

... only a few apply the key points to their classrooms ...

... person who is considered the key -- the biggest opposition leader? And ...

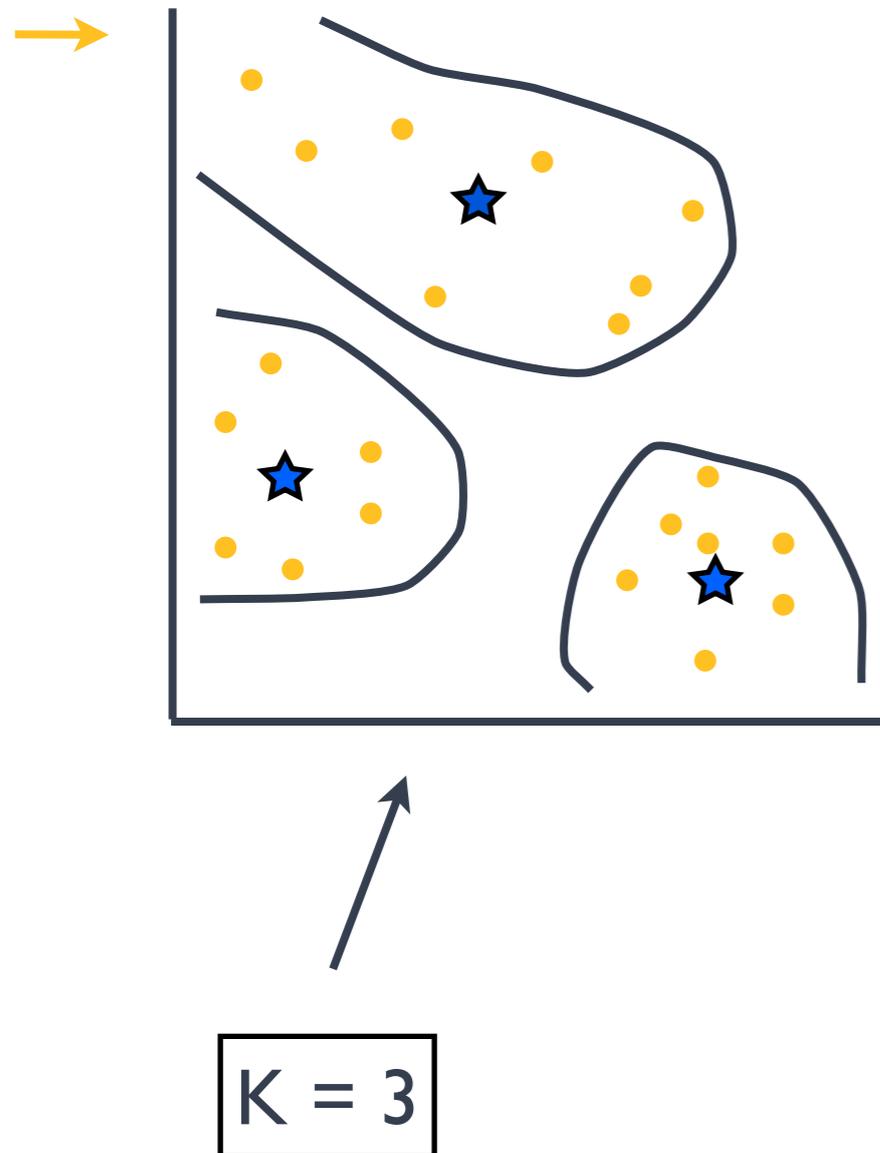
... you please get me the key to my office here please ...

- *Mixture of von Mises-Fisher distributions (movMF)*
- nutzt Cosinus-Ähnlichkeit zum Clustern
- ähnlich wie k-means (*Dhillon and Modha, 2001*)
- hartes Clustering



Der Algorithmus

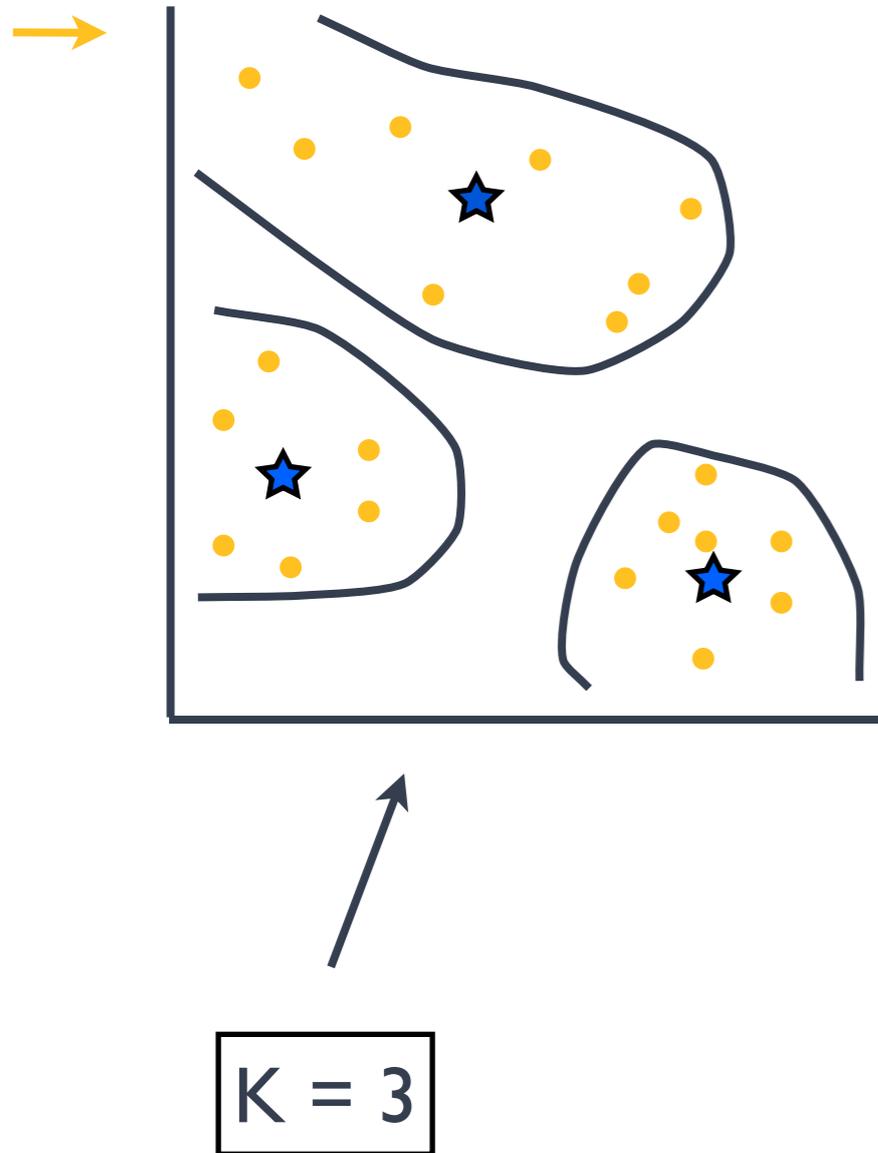
Clustering



- Anzahl der Cluster K wählbar - wird vorher fest ausgewählt
- K muss zwischen 1 und N liegen
- Bei $K = 1$ gibt es einen Cluster
→ Single-Prototype-Modell!
- Bei $K = N$ gibt es so viele Cluster wie Instanzen im Vektorraum → kein Clustering, Exemplarmodell!
- perfektes K schwer zu finden → siehe Evaluation

Der Algorithmus

Clustering



- Wichtig: Cluster entsprechen nicht zwingend distinktiven Wortbedeutungen, sondern fangen teilweise nur feine Unterschiede der Wortnutzung ein

Der Algorithmus

Ähnlichkeitsalgorithmus

Vier verschiedene Varianten

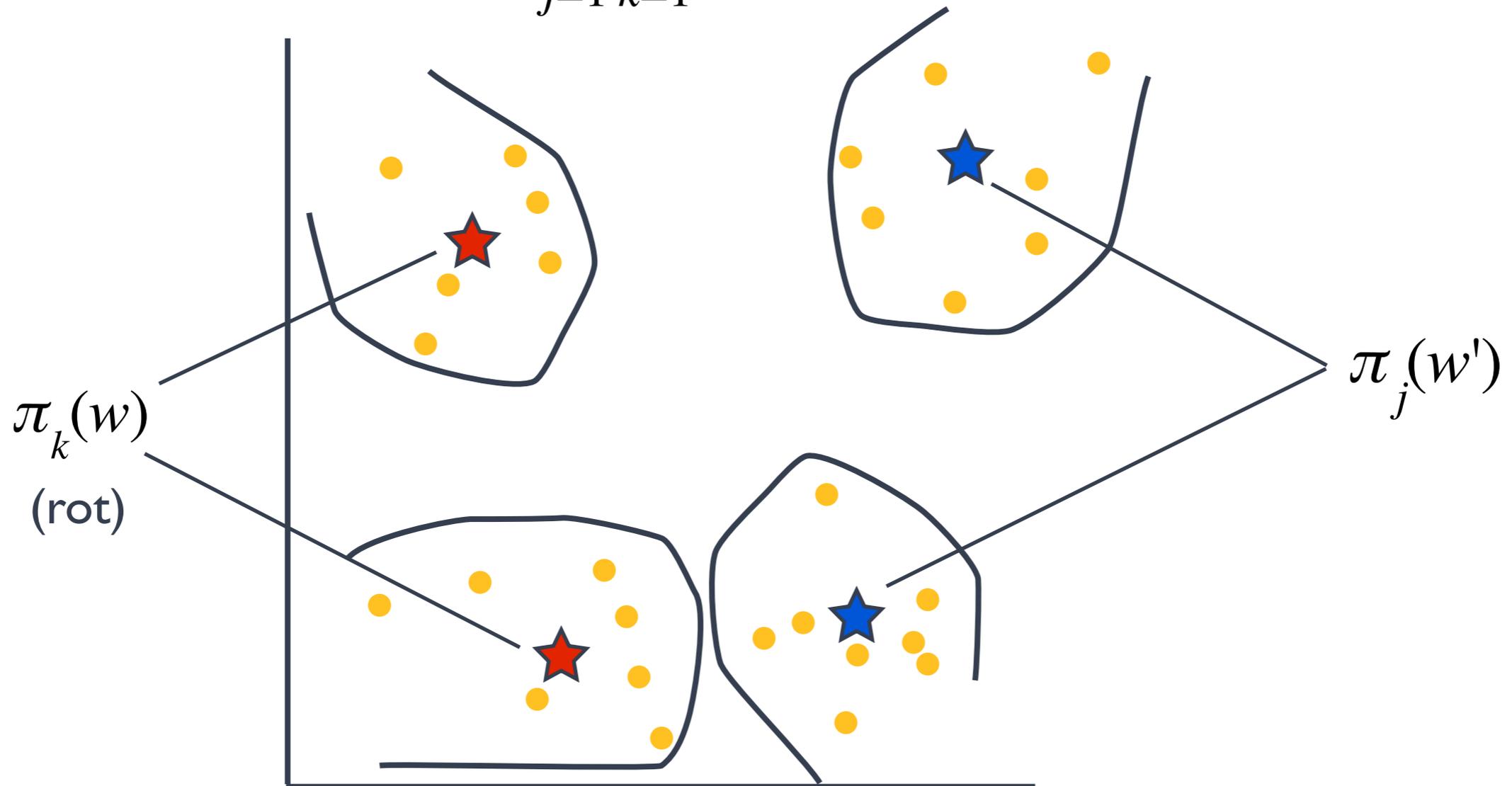
	durchschnittlich	maximal
ohne Kontext	<i>AvgSim</i> alle Clusterzentren werden betrachtet	<i>MaxSim</i> nur das ähnlichste Clusterpaar wird betrachtet
mit Kontext	<i>AvgSimC & MaxSimC</i> Gewichtung mit Kontextvektor (Wahrscheinlichkeit, dass Wort im Satzkontext auftritt)	

Der Algorithmus

Semantische Ähnlichkeit ohne Kontext

- $AvgSim(w, w')$ → durchschnittliche Ähnlichkeit aller Clusterzentren

$$AvgSim(w, w') := \frac{1}{K^2} \sum_{j=1}^K \sum_{k=1}^K sim(\pi_k(w), \pi_j(w'))$$

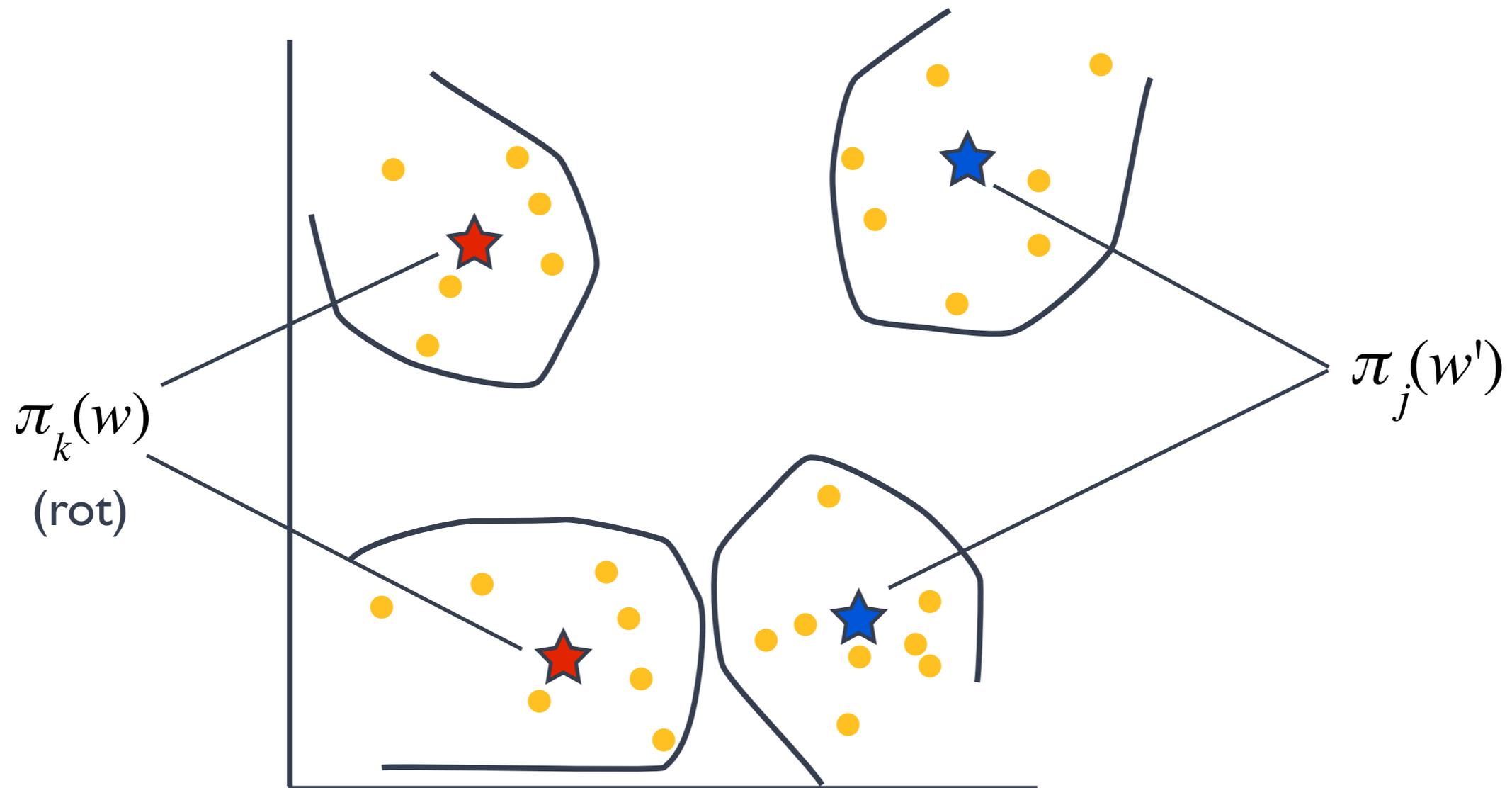


Der Algorithmus

Semantische Ähnlichkeit ohne Kontext

- $MaxSim(w, w')$ → Maximum der Ähnlichkeiten aller Clusterzentren

$$MaxSim(w, w') := \max(sim(\pi_k(w), \pi_j(w')))$$



Der Algorithmus

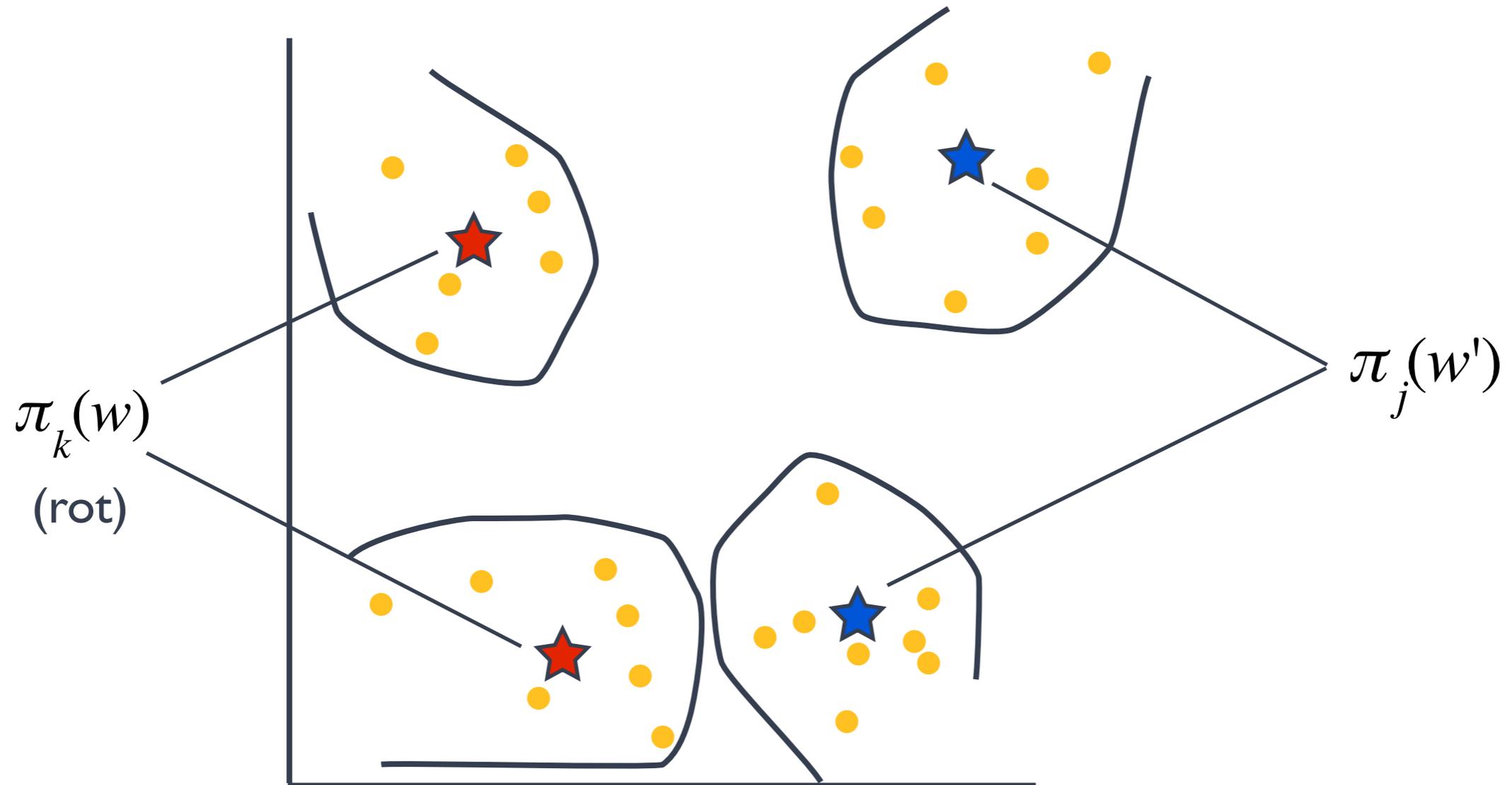
Semantische Ähnlichkeit mit Kontext

- Hierbei werden zwei Kontextvektoren hinzugezogen, die den Kontext repräsentieren.
- Diese dienen dann als Gewichtung für die jeweiligen Clusterzentren
- Beispiel:
 - *Der Förster ging in den Wald und fällte eine **Kiefer**, die sehr schöne Äste hatte.*
 - Satz wird als Kontextvektor modelliert, wahrscheinlich durch Addition der Vektoren jedes Wortes im Satz.

Der Algorithmus

Semantische Ähnlichkeit mit Kontext

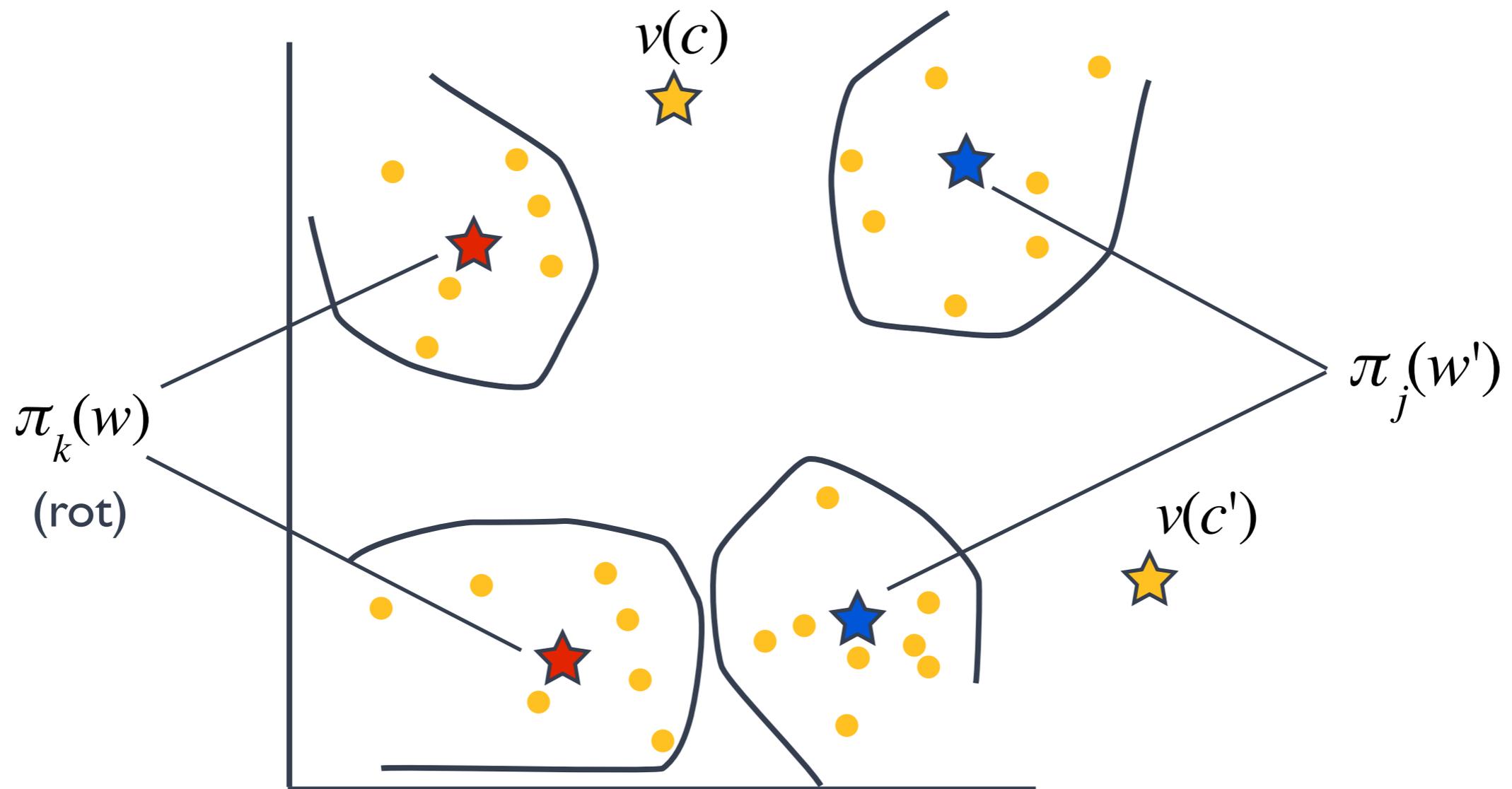
$$\text{AvgSim}(w, w') := \frac{1}{K^2} \sum_{j=1}^K \sum_{k=1}^K \text{sim}(\pi_k(w), \pi_j(w'))$$



Der Algorithmus

Semantische Ähnlichkeit mit Kontext

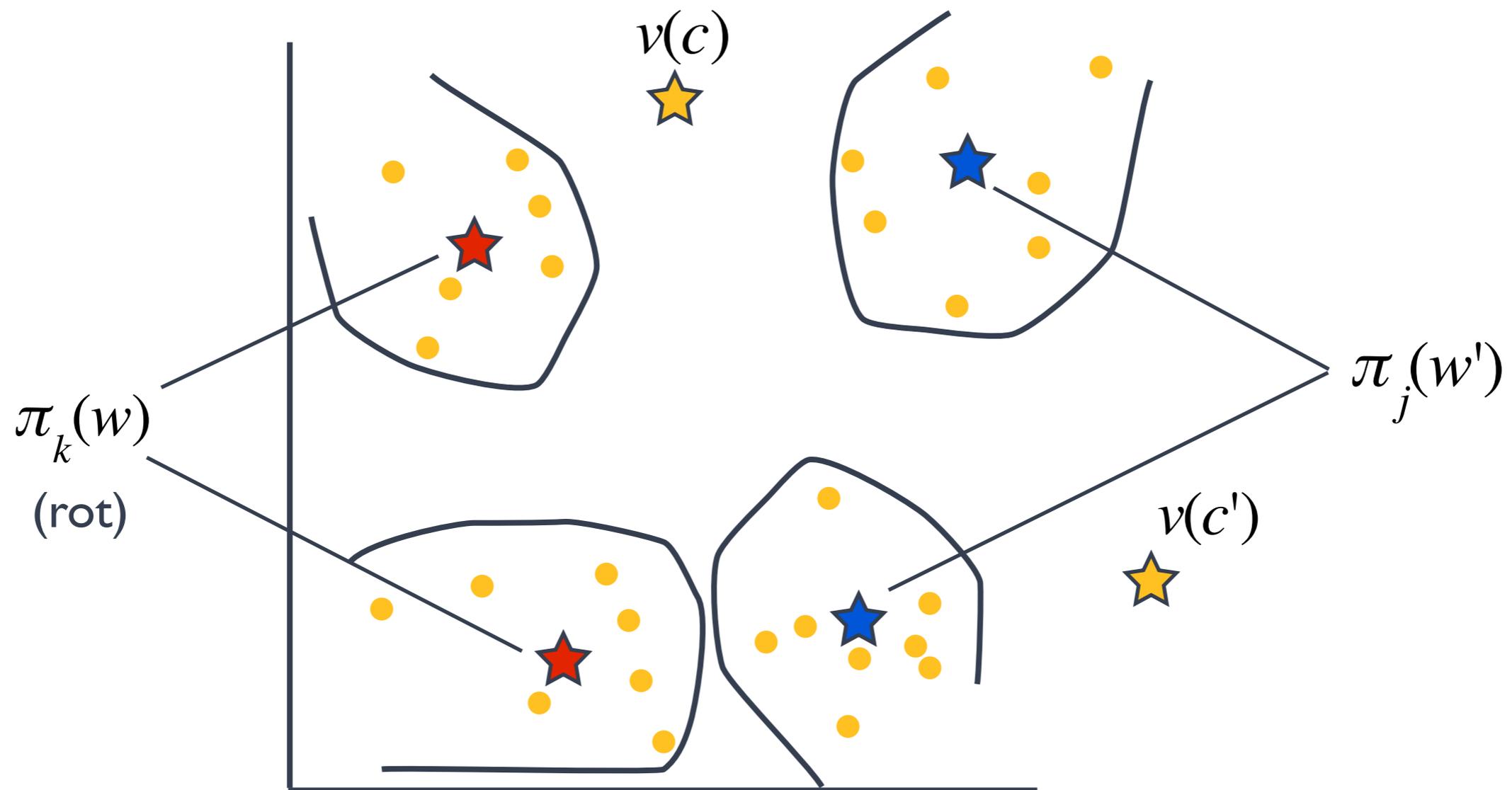
- $AvgSimC(w, w')$ → durchschnittliche Ähnlichkeit aller mit den Kontextvektoren gewichteten Clusterzentren



Der Algorithmus

Semantische Ähnlichkeit mit Kontext

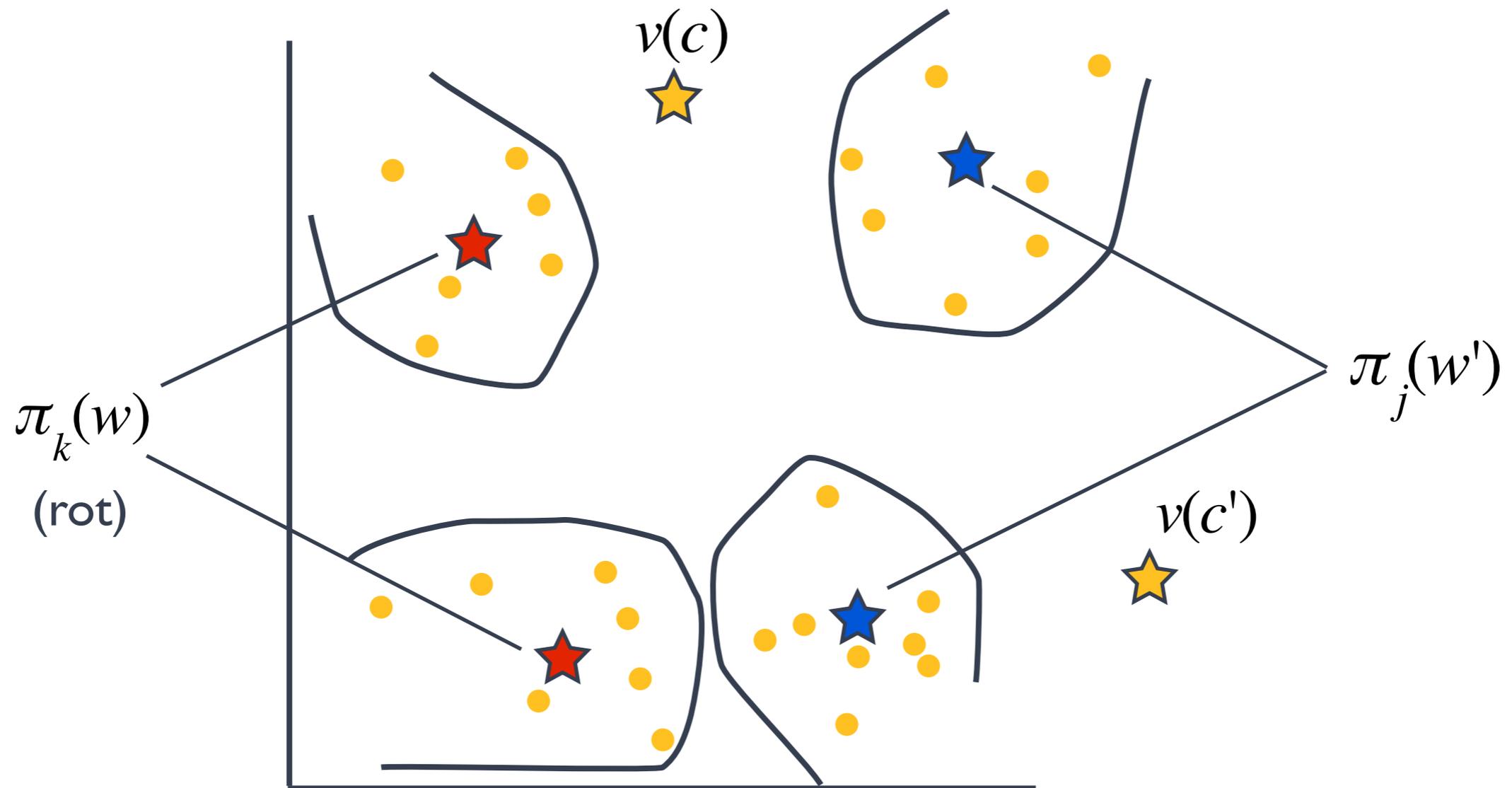
$$\text{AvgSimC}(w, w') := \frac{1}{K^2} \sum_{j=1}^K \sum_{k=1}^K \text{sim}(v(c), \pi_k(w)) \text{sim}(v(c'), \pi_k(w')) \text{sim}(\pi_k(w), \pi_j(w'))$$



Der Algorithmus

Semantische Ähnlichkeit mit Kontext

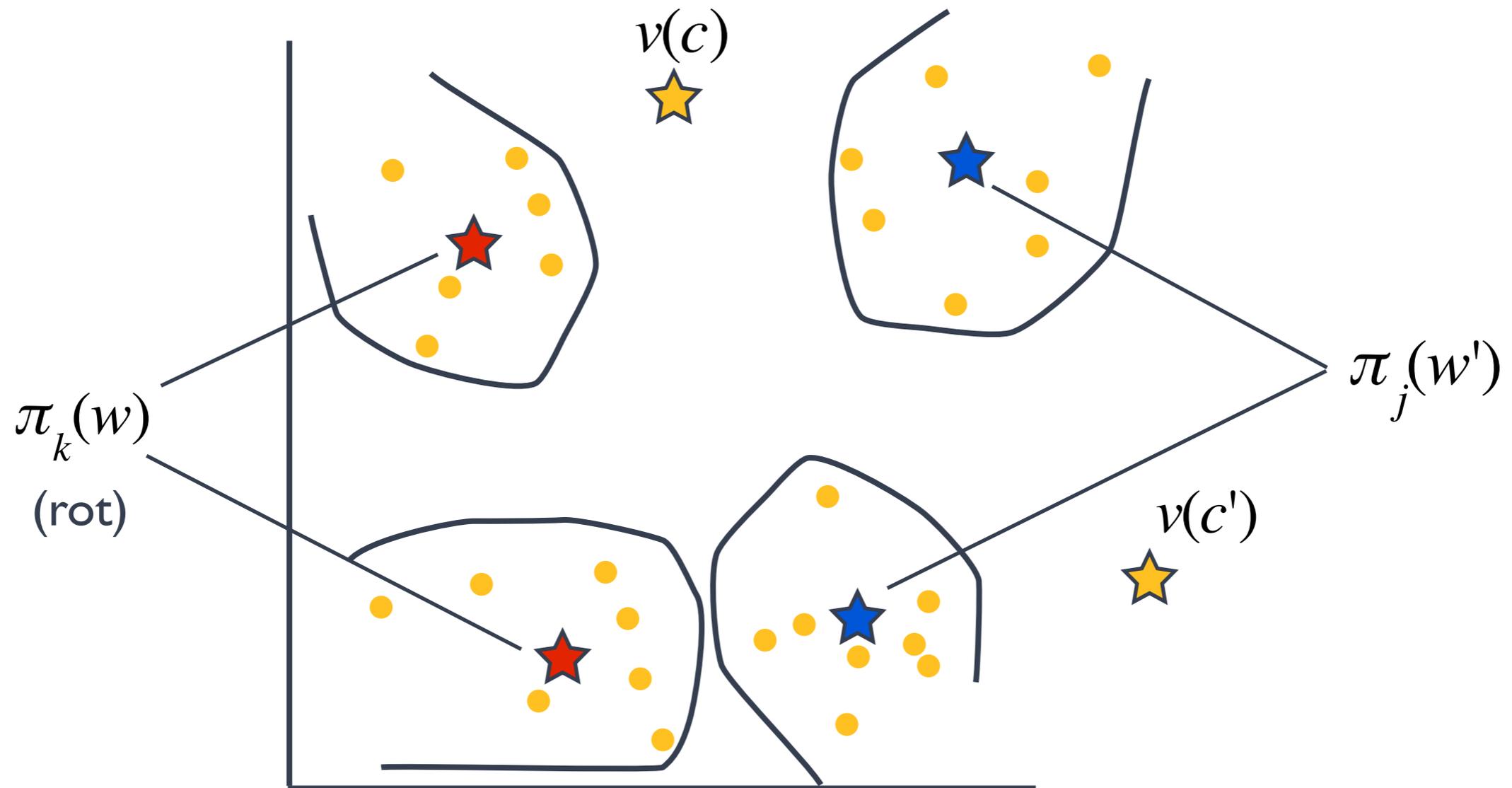
- $MaxSimC(w, w')$ → Maximum der mit den Kontextvektoren gewichteten Ähnlichkeiten aller Clusterzentren



Der Algorithmus

Semantische Ähnlichkeit mit Kontext

$$\text{MaxSimC}(w, w') := \text{sim}(\hat{\pi}_k(w), \hat{\pi}_j(w'))$$



Der Algorithmus

Zusammenhang der Formeln mit und ohne Kontext

$$\text{AvgSim}(w, w') := \frac{1}{K^2} \sum_{j=1}^K \sum_{k=1}^K \text{sim}(\pi_k(w), \pi_j(w'))$$

$$\text{AvgSimC}(w, w') := \frac{1}{K^2} \sum_{j=1}^K \sum_{k=1}^K \text{sim}(v(c), \pi_k(w)) \text{sim}(v(c'), \pi_k(w')) \text{sim}(\pi_k(w), \pi_j(w'))$$

$$\text{MaxSim}(w, w') := \max(\text{sim}(\pi_k(w), \pi_j(w')))$$

$$\text{MaxSimC}(w, w') := \text{sim}(\hat{\pi}_k(w), \hat{\pi}_j(w'))$$

$$\hat{\pi}(w) := \pi_{\text{argmax}_{1 \leq k \leq K} \text{sim}(v(c), \pi_k(w))}$$

Der Algorithmus

Zusammenhang der Formeln mit und ohne Kontext

$$AvgSim(w, w') := \frac{1}{K^2} \sum_{j=1}^K \sum_{k=1}^K sim(\pi_k(w), \pi_j(w'))$$

$$AvgSimC(w, w') := \frac{1}{\alpha} \sum_{j=1}^K \sum_{k=1}^K (sim(v(c), \pi_k(w)) \cdot sim(v(c'), \pi_j(w'))) \cdot sim(\pi_k(w), \pi_j(w'))$$

AvgSim und MaxSim sind Spezialfälle von AvgSimC und MaxSimC mit konstanter Clustergewichtung

→ Avg/MaxSimC können die Wahrscheinlichkeit für sowohl kontextabhängige als auch isolierte Wörter berechnen

$$MaxSimC(w, w') := sim(\hat{\pi}_k(w), \hat{\pi}_j(w'))$$

$$\hat{\pi}(w) := \pi_{\arg\max_{1 \leq k \leq K} sim(v(c), \pi_k(w))}$$

Was haben wir jetzt?

- variabler Clustering-Algorithmus. Kann verschiedenartige Vektorräume generieren
 - Exemplar: Anzahl der Cluster = Anzahl der Instanzen
 - Single-Prototype: ein Cluster
 - Multi-Prototype: Mittelding
- verschiedene Varianten des Ähnlichkeitsalgorithmus
 - Kann verschiedene semantische Relationen modellieren:
 - mit oder ohne Kontext
 - „durchschnittliche“ oder „maximale“ Ähnlichkeit

Experimentelle Evaluation

Überblick

- Vektor-Repräsentation
- Trainings- und Testmaterial
- Evaluation für
 - Semantische Ähnlichkeit zweier Wörter
 - Bestimmung von Synonymen ohne und mit Kontext

Evaluation

Mehrdeutigkeit?!

- *zur Erinnerung:*
- Homonymie: unterschiedliche Bedeutung + unterschiedliche Herkunft
 - z.B. Arm/arm
 - (Fuß-)Ball/(Tanz-)Ball
 - Elf/elf
- Polysemie: (leicht) unterschiedliche Bedeutung + gemeinsamer Ursprung
 - z.B. Strom (elektr./Wasser)
 - Mutter (Mensch/Schraube)

Experimentelle Evaluation

Vektor-Repräsentation

... it is to have people **key** in, first of all, on the issues ...

...And people may want to **key** in especially on things that ...

...Well, I think that's a **key** piece. Sharon is going to talk ...

... only a few apply the **key** points to their classrooms ...

... person who is considered the **key** -- the biggest opposition leader? And ...

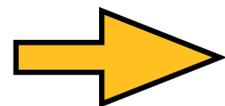
... you please get me the **key** to my office here please ...

- Im einfachsten Fall:
 - Auszählen der Kookkurrenz (Problem z. B. Funktionswörter)
- im Paper: tf-idf und χ^2
- Gewichtungsfunktionen, die die Wortfrequenz im Dokument mit einbeziehen
 - Die reine Termfrequenz (tf) wird normalisiert durch die inverse Dokumentfrequenz (idf)
 - seltene Wörter werden höher gewichtet als häufige
- auch möglich: PMI (Pointwise Mutual Information)
 - Information, die in einem Ereignis über ein anderes enthalten ist

Experimentelle Evaluation

Worauf trainiert wurde ...

- Zwei Trainingskorpora
 - ein Snapshot der englischen Wikipedia von 2009
 - ohne Wiki-Markup, ohne Artikel < 100 Wörter → ca. 2 Mrd. Wörter
 - die dritte Edition des englischen Gigaword Korpus
 - ohne Artikel < 100 Wörter → knapp 4 Mrd. Wörter
- Wikipedia deckt eine größere Auswahl von Bedeutungsvariationen ab
- Gigaword enthält nur Nachrichtentexte - weniger Bedeutungsvarianz



- Evaluation wird zeigen, dass Modell auf beiden Korpora besser ist
- schlägt Baseline-Methoden auch bei Gigaword

Experimentelle Evaluation

Womit getestet wurde ...

- Die Vergleichsdaten von Menschen wurden aus dem WordSim-353-Datenset genommen (*Finkenstein et al., 2001*)
- Testkorpus, das 353 Wortpaare enthält, deren Ähnlichkeit von Testpersonen auf einer 0-10 Integer-Skala bewertet wurde
- Korpus enthält sowohl ein- als auch mehrdeutige Wörter

Experimentelle Evaluation

Worauf trainiert wurde ...

...
tiger	tiger	10
vodka	gin	8,46
man	woman	8,3
vodka	brandy	8,13
stock	market	8,08
computer	keyboard	7,62
marathon	sprint	7,47
book	paper	7,46
tiger	cat	7,35
company	stock	7,08
student	professor	6,81
love	sex	6,77
fertility	egg	6,69
professor	doctor	6,62
...

...
brother	monk	6,27
journey	car	5,85
man	governor	5,25
monk	oracle	5
smart	student	4,62
lad	brother	4,46
stock	live	3,73
drink	car	3,04
crane	implement	2,69
cemetery	woodland	2,08
stock	phone	1,62
stock	CD	1,31
professor	cucumber	0,31
king	cabbage	0,23
...

Auszug aus WordSim 353

Experimentelle Evaluation

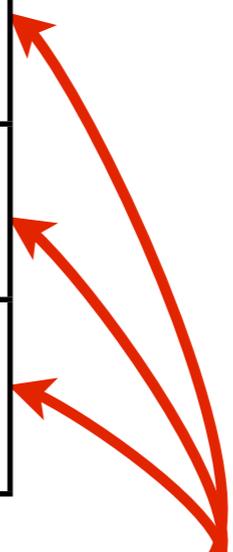
I. Semantische Ähnlichkeit zweier Wörter

Spearman's <i>Rho</i>	<i>multi-prototype (mit AvgSim)</i>		
	K = 5	K = 20	K = 50
Wikipedia tf-idf	0,69	0,76	0,76
Wikipedia chi ²	0,58	0,56	0,52
Gigaword tf-idf	0,64	0,61	0,61
Gigaword chi ²	0,32	0,35	0,33

Experimentelle Evaluation

I. Semantische Ähnlichkeit zweier Wörter

Spearman's Rho	multi-prototype (mit AvgSim)		
	K = 5	K = 20	K = 50
Wikipedia tf-idf	0,69	0,76	0,76
Wikipedia chi ²	0,58	0,56	0,52
Gigaword tf-idf	0,64	0,61	0,61
Gigaword chi ²	0,32	0,35	0,33



- Erhöhen der Clusterzahl → feinere Bedeutungsunterschiede, aber auch mehr Rauschen

Experimentelle Evaluation

I. Semantische Ähnlichkeit zweier Wörter

<i>Spearman's Rho</i>	<i>multi-prototype (mit AvgSim)</i>			<i>combined</i>
	K = 5	K = 20	K = 50	
Wikipedia tf-idf	0,69	0,76	0,76	0,77
Wikipedia chi ²	0,58	0,56	0,52	0,59
Gigaword tf-idf	0,64	0,61	0,61	0,62
Gigaword chi ²	0,32	0,35	0,33	0,34

- Deshalb: Kombinieren der verschiedenen Clustermengen
 - erzielt meist ähnlich gute Ergebnisse wie die Auswahl des besten K .

Experimentelle Evaluation

I. Semantische Ähnlichkeit zweier Wörter

Spearman's <i>Rho</i>	<i>prototype</i>	<i>exemplar</i>	<i>multi-prototype (mit AvgSim)</i>			<i>combined</i>
			K = 5	K = 20	K = 50	
Wikipedia tf-idf	0,53	0,6	0,69	0,76	0,76	0,77
Wikipedia chi ²	0,54	0,65	0,58	0,56	0,52	0,59
Gigaword tf-idf	0,49	0,48	0,64	0,61	0,61	0,62
Gigaword chi ²	0,25	0,41	0,32	0,35	0,33	0,34

Experimentelle Evaluation

I. Semantische Ähnlichkeit zweier Wörter

Spearman's <i>Rho</i>	<i>prototype</i>	<i>exemplar</i>	<i>multi-prototype (mit AvgSim)</i>			<i>combined</i>
			K = 5	K = 20	K = 50	
Wikipedia tf-idf	0,53	0,6	0,69	0,76	0,76	0,77
Wikipedia chi ²	0,54	0,65	0,58	0,56	0,52	0,59
Gigaword tf-idf	0,49	0,48	0,64	0,61	0,61	0,62
Gigaword chi ²	0,25	0,41	0,32	0,35	0,33	0,34

- tf-idf-Features zeigen generell eine deutlich höhere Korrelation als chi²-gewichtete-Vektoren

Experimentelle Evaluation

I. Semantische Ähnlichkeit zweier Wörter

Spearman's <i>Rho</i>	<i>prototype</i>	<i>exemplar</i>	<i>multi-prototype (mit AvgSim)</i>			<i>combined</i>
			K = 5	K = 20	K = 50	
Wikipedia tf-idf	0,53	0,6	0,69	0,76	0,76	0,77
Wikipedia chi ²	0,54	0,65	0,58	0,56	0,52	0,59
Gigaword tf-idf	0,49	0,48	0,64	0,61	0,61	0,62
Gigaword chi ²	0,25	0,41	0,32	0,35	0,33	0,34

- Der Exemplaransatz zeigt fast immer deutlich höhere Korrelationen als das Single-Prototyp-Modell

Experimentelle Evaluation

I. Semantische Ähnlichkeit zweier Wörter

Spearman's <i>Rho</i>	<i>prototype</i>	<i>exemplar</i>	<i>multi-prototype (mit AvgSim)</i>			<i>combined</i>
			K = 5	K = 20	K = 50	
Wikipedia tf-idf	0,53	0,6	0,69	0,76	0,76	0,77
Wikipedia chi ²	0,54	0,65	0,58	0,56	0,52	0,59
Gigaword tf-idf	0,49	0,48	0,64	0,61	0,61	0,62
Gigaword chi ²	0,25	0,41	0,32	0,35	0,33	0,34

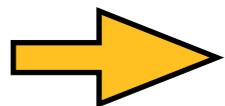
- Der Exemplaransatz zeigt fast immer deutlich höhere Korrelationen als das Single-Prototyp-Modell
- Das kombinierte Modell ist bei tf-idf wiederum signifikant besser und bei chi² nur unwesentlich schlechter

Experimentelle Evaluation

I. Semantische Ähnlichkeit zweier Wörter

Spearman's Rho	prototype	exemplar	multi-prototype (mit AvgSim)			combined
			K = 5	K = 20	K = 50	
Wikipedia tf-idf	0,53	0,6	0,69	0,76	0,76	0,77
Wikipedia chi ²	0,54	0,65	0,58	0,56	0,52	0,59
Gigaword tf-idf	0,49	0,48	0,64	0,61	0,61	0,62
Gigaword chi ²	0,25	0,41	0,32	0,35	0,33	0,34

- Resümee:
 - Worst Case: combined MPT mindestens so gut wie exemplar
 - Best Case (bei Wahl des passenden Korpus-Feature-Paares): deutliches Übertreffen der anderen Modelle



Experimentelle Evaluation

2. Bestimmung von Synonymen

- Vergleich von Single- und Multi-Prototyp-Modellen beim Bestimmen von Synonymen gegeben ein Ausgangswort
 - Ausgangswörter sind homonyme und polysemische englische Wörter
 - Homonyme: carrier, crane, cell, company, issue ...
 - Polyseme: cause, chance, journal, market, network ...

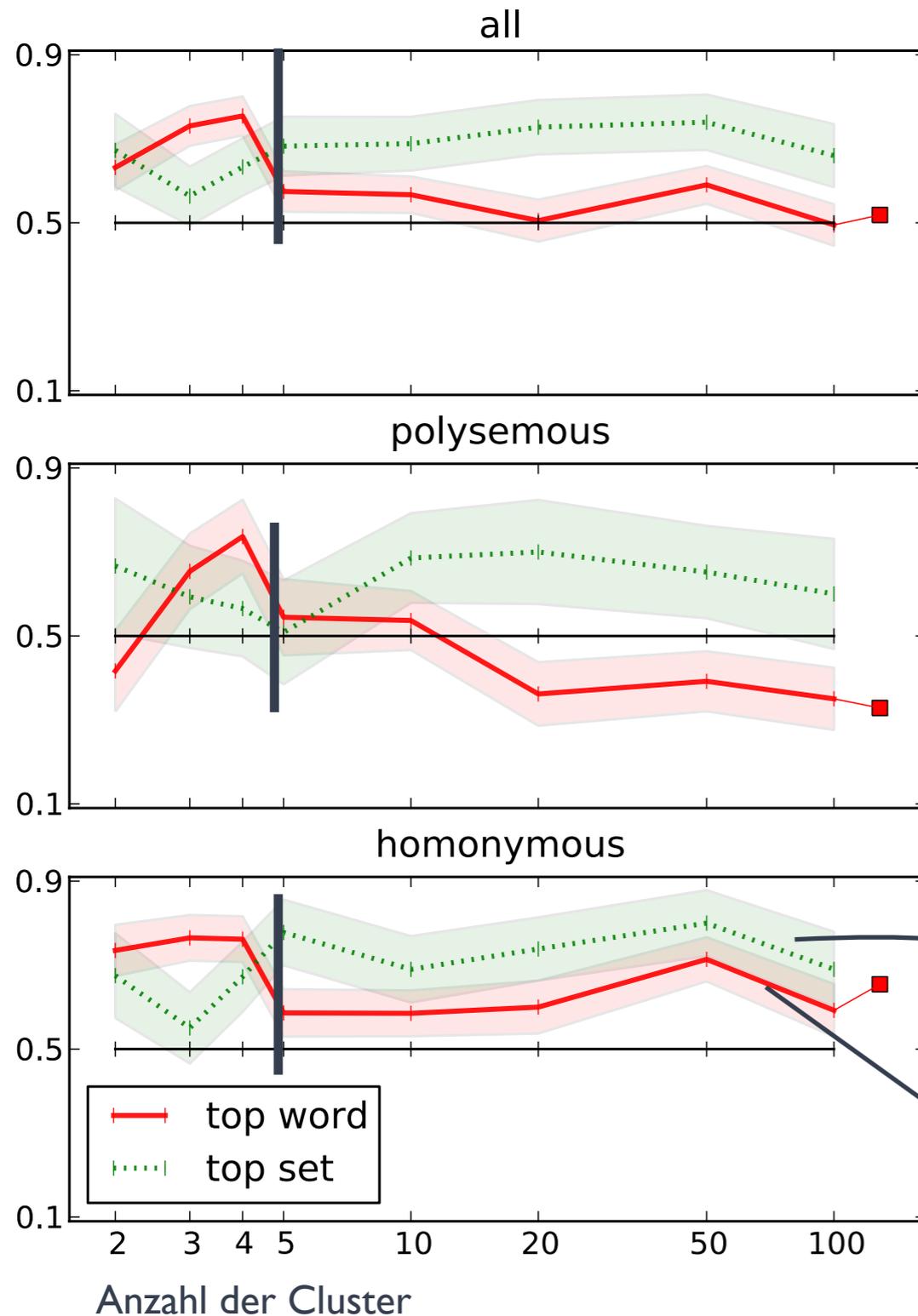
Experimentelle Evaluation

2. Bestimmung von Fast-Synonymen

- Aufbau des Experiments:
- Multi-Prototype → für jeden Prototypen jedes Ausgangswortes wurden die k ähnlichsten Worte bestimmt
- Single-Prototype → das ähnlichste Wort wurde ebenfalls bestimmt
- Testpersonen bekamen jeweils bestes Ergebnis
- Welches wird bevorzugt?
- Bei der Evaluation wurden die Ausgangswörter getrennt betrachtet
 - isoliert
 - in einem zufälligen Satzkontext

Experimentelle Evaluation

2a. Bestimmung von Fast-Synonymen ohne Kontext

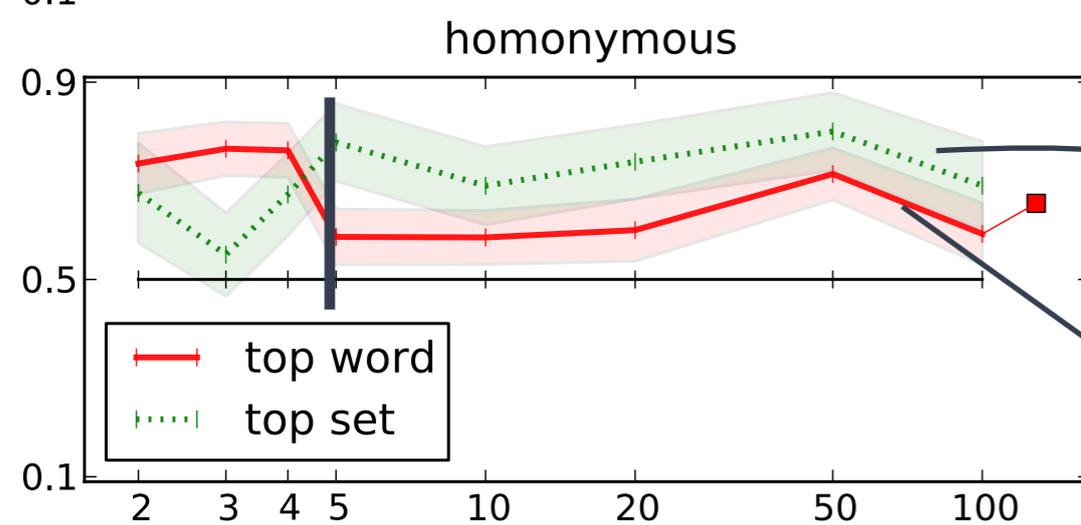
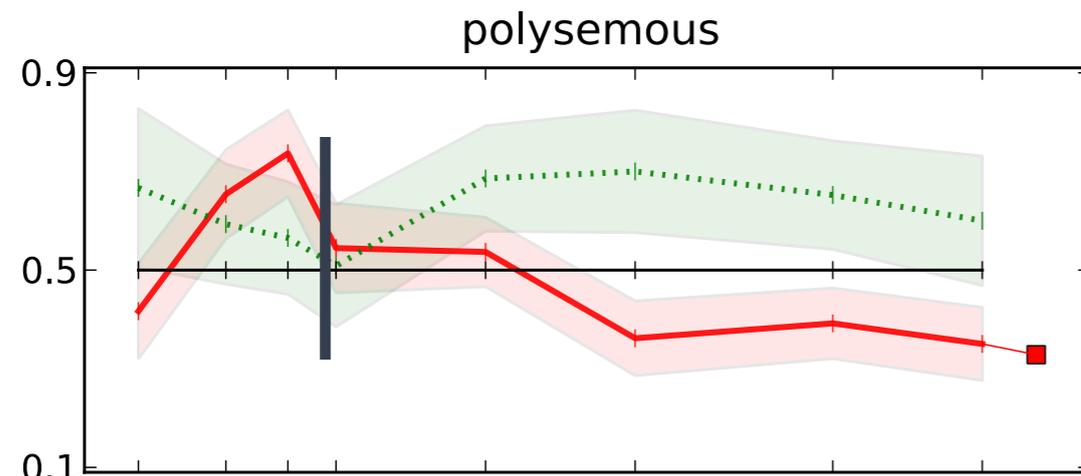
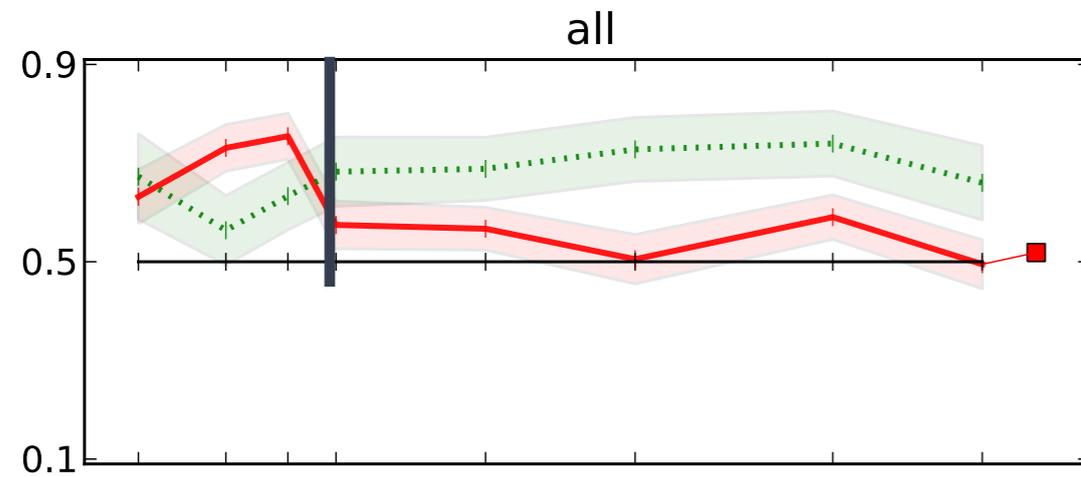


- Berechnung mit AvgSim
- bis 5 Cluster signifikant bessere Ergebnisse
- danach für homonyme Wörter auch, da Bedeutung dieser sehr distinkt
 - je mehr Cluster, desto klarer
- bei polysemischen Wörtern ist MPT auch besser, wenn man die richtige Anzahl Cluster wählt

(grün)
(rot)

Experimentelle Evaluation

2a. Bestimmung von Fast-Synonymen ohne Kontext



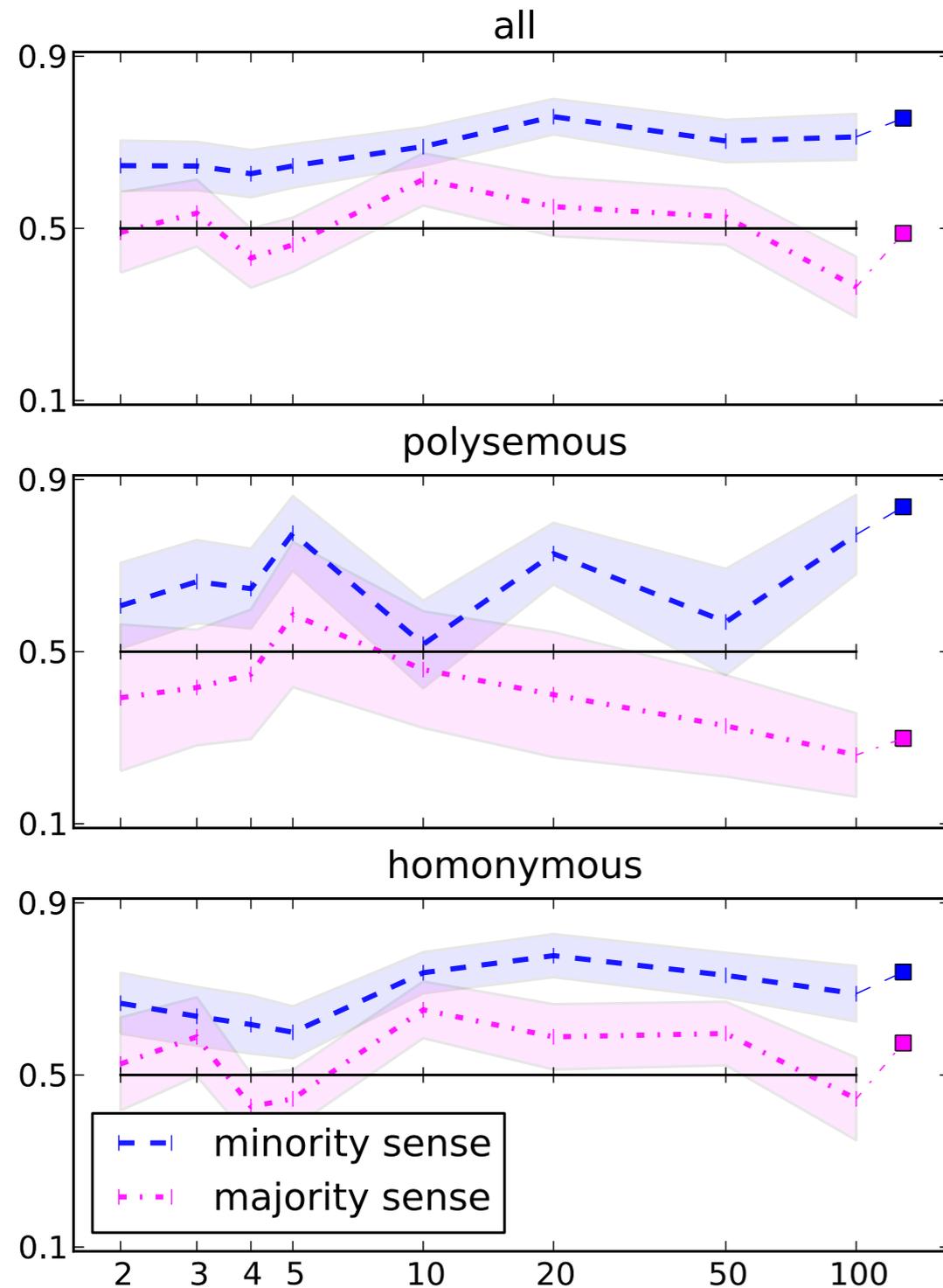
- Problem bei Synonymbestimmung: je mehr Cluster, desto mehr Rauschen
- Lösungsansatz: nicht top word, sondern top set (top 3)
- Dadurch signifikant besser!

(grün)
(rot)

Anzahl der Cluster

Experimentelle Evaluation

2b. Bestimmung von Fast-Synonymen mit Kontext



- Berechnung mit AvgSimC
- majority sense (Hauptbedeutung), kein signifikanter Unterschied von Single zu Multi-Prototype
- minority sense (Nebenbedeutung) ist MPT allerdings deutlich stärker!

Experimentelle Evaluation

Varianz bei menschlichen Bewertungen?

- Evaluation, wie Varianz bei Testpersonen zustande kommt
- Korrelation zwischen
 - Varianz von menschlichen Bewertungen für ein Wortpaar
 - Varianz der Ähnlichkeitsberechnungen des Modells
- Resultat: Negative Korrelation
 - Intuitiv: Wenn Varianz der Ähnlichkeitsberechnungen für ein Wortpaar sinkt (Cluster nicht leicht voneinander zu unterscheiden)
 - Dann steigt Varianz der menschlichen Bewertungen

Fazit der Evaluation

- Bei Ähnlichkeitsberechnung ist Multi-Prototype mit dem richtigen Korpus und den richtigen Features klar überlegen
- Bei Synonymbestimmung ohne Kontext klarer Sieger, da Rauschen durch größere Clusteranzahl durch Auswahl aus Top3 vermindert werden konnte
- Mit Kontext zeigt es vor allem seine Stärken beim Erkennen der Nebenbedeutungen von Worten, die ein einzelner Cluster nicht einfangen kann

Probleme & Unklarheiten

- Korpus, einfacher Bag of Words oder mit syntaktischen Informationen
 - z.Z. keine explizite Einschränkung → Verben und Nomen können bspw. zusammengefasst werden im Korpus

... it is to have people **key** in, first of all, on the issues ...

...And people may want to **key** in especially on things that ...

...Well, I think that's a **key** piece. Sharon is going to talk ...

... only a few apply the **key** points to their classrooms ...

... person who is considered the **key** -- the biggest opposition leader? And ...

... you please get me the **key** to my office here please ...

Ausblick in die Zukunft ...

Was kann noch verbessert werden?

- Repräsentation der Features kann noch verbessert werden
 - Latent/Explicit Semantic Analysis (Deerwester et al., 1990; Gabrilovich and Markovitch, 2007)
- Perfekte Clusteranzahl?
 - Variiert von Wort zu Wort → MPT nutzte den kombinierten Ansatz
 - anderes Clusteringverfahren, was dieses Problem umgeht
 - z. B. Dirichlet Process Mixture Model (Rasmussen, 2000)
- Die Bedeutungen, die in den Clustern eines Wortes stecken, stellen nicht zwangsweise echte semantische Unterschiede dar
 - man könnte die generierten Prototypen mit Bedeutungsdatenbanken abgleichen

Ausblick in die Zukunft ...

Was kann noch verbessert werden?

- Repräsentation der Features kann noch verbessert werden
 - Latent/Explicit Semantic Analysis (Deerwester et al., 1990; Gabrilovich and Markovitch, 2007)
- Perfekte Clusteranzahl?

Model ist komplett modular

→ Veränderungen gut zu realisieren

- Die Bedeutungen, die in den Clustern eines Wortes stecken, stellen nicht zwangsweise echte semantische Unterschiede dar
 - man könnte die generierten Prototypen mit Bedeutungsdatenbanken abgleichen

Das war's!

Fragen?