

Die Repräsentation und Auflösung von ambigen Wortbedeutungen in der Computerlinguistik

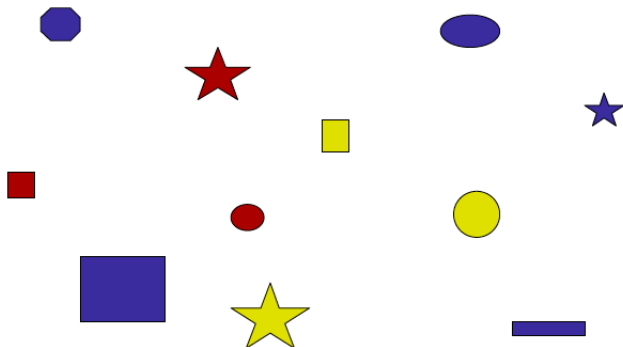
Klassifikation

PD Dr. Sabine Schulte im Walde

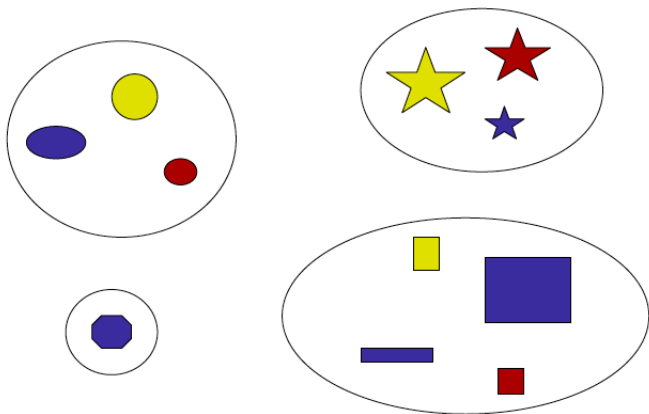
Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart

26. März 2010

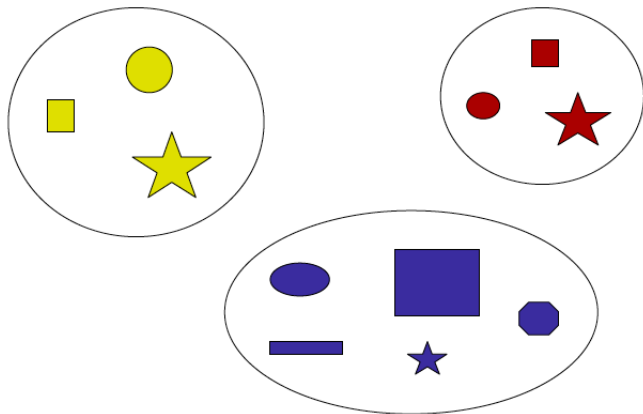
Illustration



Illustration



Illustration



Klassifikation

- Objekte → Klassen (Prozess und Ergebnis)
- Objekte in **gemeinsamen** Klassen: so **ähnlich** wie möglich
- Objekte in **unterschiedlichen** Klassen: so **verschieden** wie möglich
- Ziele:
 - Übersicht über Objekte, Erkennen von Datenstrukturen
 - Vergleichbarkeit von Objekten
 - Generalisierung über Objekte
 - Einordnung von einzelnen Objekten (Klassierung)

Klassifikationsbeispiele

- topographische Karten
- Größe/Haarfarbe von Menschen
- Beurteilung eines Kinofilms

Klassifikationsbeispiele

- topographische Karten
- Größe/Haarfarbe von Menschen
- Beurteilung eines Kinofilms

- Biologie: Systematik von Pflanzen/Tieren
- Medizin: Symptomklassifikation → Krankheiten erkennen
- Soziologie: Rollenverhalten in Gruppen

Klassifikationsbeispiele

- topographische Karten
- Größe/Haarfarbe von Menschen
- Beurteilung eines Kinofilms

- Biologie: Systematik von Pflanzen/Tieren
- Medizin: Symptomklassifikation → Krankheiten erkennen
- Soziologie: Rollenverhalten in Gruppen

- Part-of-Speech Tagging
- PP-Attachment
- semantische (Verb-/Nomen-/Adjektiv-/etc.)Klassen

Klassifikation in der Computerlinguistik

- Vielzahl von Klassifikationsproblemen
- Algorithmen der Künstlichen Intelligenz
- Schwierigkeit: Auswahl eines Klassifikationsalgorithmus und Bestimmung der Parameter
- Schwierigkeitsgrad abhängig von theoretischer Definition der Aufgabe (Bsp.: Wortarten vs. semantische Klassen)
- Orientierung: linguistische Kriterien

Parameter

- Ziel und Zweck der Klassifikation
- Objekte: relevante und repräsentative Daten
- Eigenschaften der Objekte
- Ähnlichkeitsmaße zum Vergleich der Objekte
- Algorithmus für die Klassifizierung und Klassierung
- Interpretation, Evaluierung und Anwendung

Semantische Klassifikation

- Klassifikation von Wörtern/Wortbedeutungen auf Basis ihrer semantischen Eigenschaften
- Unterschiedliche Arten von semantischen Klassen gemäß der semantischen Eigenschaften
- Klassen beziehen sich auf allgemeine semantische Ebene; idiosynkratische lexikalische Eigenschaften sind unterspezifiziert; semantische Relationen sind unterspezifiziert
- Intuitive Beispiele:
 - **Eigenbewegung**: laufen, rennen, schwimmen, klettern etc.
 - **Tier**: Vogel, Möwe, Meise, Hund, Dackel etc.
 - **Temperatur**: heiß, kalt, lau, warm etc.

Beispielansätze semantischer Klassifikationen

- 1 Hindle (1990): Nomen-Klassifikation (englisch)
- 2 Merlo & Stevenson (2001): Verb-Klassifikation (englisch)
- 3 Schulte im Walde (2006): Verb-Klassifikation (deutsch)

Nomen-Klassifikation

- Automatische semantische Klassifikation von englischen Nomen
- Basis: distributionelle Hypothese
- Information: Prädikat-Argument-Strukturen
 - ↪ Syntax-Semantik-Schnittstelle
 - ↪ Verben, deren Subjekte und/oder Objekte die Nomen sind
- Daten: 6 Millionen Wörter der *Associated Press news stories*, gepart
- Datensätze: 4.789 Verben in 274.613 Sätzen mit 26.742 Kopfnomen

Nomen-Ähnlichkeit

- Variante von **Mutual Information** berechnet die Stärke des Zusammenhalts von Verb v und Nomen n in Bezug auf eine bestimmte Funktion $func$ (Subjekt oder Objekt)

- Mutual Information allgemein:

$$I(x, y) = \log_2 \frac{p(x, y)}{p(x) p(y)}$$

- Mutual Information unter Einbezug der Anzahl Sätze N , in denen die Vorkommen auftreten:

$$C_{func}(n, v) = \log_2 \frac{\frac{f(n, v)}{N}}{\frac{f(n)}{N} \frac{f(v)}{N}}$$

Nomen-Ähnlichkeit

- Basis: Vergleich von MI-Werten von je zwei Nomen über alle Verb-Nomen-Paare
- Funktions-Ähnlichkeit (Subjekt, Objekt):

$$SIM_{func}(v_i, n_j, n_k) = \min(C_{func}(v_i, n_j), C_{func}(v_i, n_k)), \\ \text{falls } C_{func}(v_i, n_j) > 0 \text{ und } C_{func}(v_i, n_k) > 0$$

$$SIM_{func}(v_i, n_j, n_k) = \text{abs}(\max(C_{func}(v_i, n_j), C_{func}(v_i, n_k))), \\ \text{falls } C_{func}(v_i, n_j) < 0 \text{ und } C_{func}(v_i, n_k) < 0$$

$$SIM_{func}(v_i, n_j, n_k) = 0 \text{ sonst}$$

- Nomen-Ähnlichkeit:

$$SIM(n_1, n_2) = \sum_{i=0}^N SIM_{subj}(v_i, n_1, n_2) + SIM_{obj}(v_i, n_1, n_2)$$

Ergebnis: Beispiel

Noun	f(n)	verbs	SIM	Verbs
<i>boat</i>	153	79	370.16	_cruise, keel_, _plow, sink_, drift_, step off_, step from_, dock_, right_, submerge_, _near, hoist_, _intercept, charter_, stay on_, buzz_, stabilize_, _sit on, intercept_, hijack_, park_, _be from, rock_, get off_, board_, miss_, stay with_, _catch, yield_, bring in_, seize_, pull_, grab_, _hit, exclude_, weigh_, _issue, demonstrate_, _force, _cover, supply_, _name, _attack, damage_, launch_, _provide, appear_, _carry, _go to, look at_, attack_, _reach, _be on, watch_, use_, return_, _ask, destroy_, fire_, be on_, describe_, charge_, include_, be in_, report_, identify_, expect_, cause_, 's_, 's, _take, _make, _be, _say, _give, see_, _be, _have_, _get _near, charter_, hijack_, get off_, buzz_, intercept_, board_, damage_, sink_, seize_, _carry, attack_, _have_, be on_, hit, destroy_, watch_, _go to, _give_, _ask, _be, be on_, _say_, identify_, see_
<i>ship</i>	353	25	79.02	hijack_, intercept_, charter_, board_, get off_, _near, _attack, _carry, seize_, _have_, be on_, _catch, destroy_, hit, be on_, damage_, use_, _be, _go to, _reach, _say_, identify_, provide_, expect_, cause_, see_
<i>plane</i>	445	26	68.85	step off_, hijack_, park_, get off_, board_, _catch, seize_, _carry, attack_, be on, be on_, charge_, expect_, _have_, take, _say_, _make, include_, be in_, _be
<i>bus</i>	104	20	64.49	charter_, intercept_, hijack_, park_, board_, _hit, seize_, _attack, _force, _carry, use_, describe_, include_, be on_, _be, _make, _say_
<i>jet</i>	153	17	62.77	right_, dock_, _intercept, sink_, seize_, _catch, attack_, carry, attack_, _have_, describe_, identify_, use_, report_, _be, _say_, expect_, _give_
<i>vessel</i>	172	18	57.14	

Ergebnis: Beispiel

Noun	f(n)	verbs	SIM	Verbs
<i>table</i>	66	30	181.43	hide beneath_, convolute_, memorize_, sit at_, sit across_, redo_, structure_, sit around_, litter_, _carry, lie on_, go from_, _hold, wait_, come to_, return to_, turn_, approach_, cover_, be on_, share_, publish_, claim_, mean_, go to_, raise_, leave_, have_, do_, be_
<i>floor</i>	94	6	30.01	litter_, lie on_, cover_, be on_, come to_, go to_
<i>farm</i>	80	8	22.94	_carry, be on_, cover_, return to_, turn_, go to_, leave_, have_
<i>scene</i>	135	10	20.85	approach_, return to_, mean_, go to_, be on_, turn_, come to_, leave_, do_, be_
<i>America</i>	156	7	19.68	go from_, come to_, return to_, claim_, go to_, have_, do_
<i>experience</i>	129	5	19.04	structure_, share_, claim_, publish_, be_
<i>river</i>	95	4	18.73	sit across_, mean_, be on_, leave_
<i>town</i>	195	6	18.68	litter_, approach_, go to_, return to_, come to_, leave_
<i>side</i>	327	8	18.57	lie on_, be on_, go to_, _hold, have_, cover_, leave_, come to_
<i>hospital</i>	190	7	18.10	go from_, come to_, cover_, return to_, go to_, leave_, have_
<i>House</i>	453	6	17.84	return to_, claim_, come to_, go to_, cover_, leave_

Zusammenfassung

- Prädikat-Argument-Struktur ist nützliche empirische Basis für Nomen-Ähnlichkeit
- Eigenschaften können von subkategorisierenden Verben auf andere syntaktische Funktionen ausgeweitet werden, z.B. modifizierende Adjektive
- Ambiguität von Nomen ist nicht berücksichtigt

Verb-Klassifikation

- Automatische Klassifikation von intransitiven Verben in drei Klassen:
 - unergativ
 - unakkusativ
 - intransitiv verwendetes transitives Verb (*object drop*)
- Schwierigkeit: ähnliche Subkategorisierungsrahmen aber verschiedene thematische Rollen
- Basis: linguistisch motivierte Verb-Eigenschaften als Indikatoren für thematische Rollen
- Hypothese: Verb-Verhalten ↔ Verb-Bedeutung
- Klassifikation: überwachte Entscheidungsbäume

Verbklassen: Beispiele

- Unergative Verben:

The horse raced past the barn.

The jockey raced the horse past the barn.

- Unakkusative Verben:

The butter melted in the pan.

The cook melted the butter in the pan.

- Intransitiv verwendete transitive Verben:

The boy played.

The boy played soccer.

Thematische Rollen der Verbklassen

Klasse	intransitiv	transitiv	
	Subjekt	Subjekt	Objekt
unergativ	Agens	Agens, kausal	Agens
unakkusativ	Thema	Agens, kausal	Thema
transitiv	Agens	Agens	Thema

Vorgehensweise und Ergebnis

- Vorgehensweise:
 - ① Analyse der thematischen Muster der drei Klassen
 - ② Entwicklung von statistischen Indikatoren zur Beschreibung und Abgrenzung der Klassen
 - ③ Klassifikation durch überwachtetes Verfahren: Entscheidungsbäume
- Typ-basiert (nicht: Token-basiert)
- Ergebnis:
 - Genauigkeit: 69.5%
 - Baseline (Zufall): 34%
 - obere Grenze (Experten): 86.5%

Empirische Indikatoren für Eigenschaften der Verbklassen

Indikatoren

(Annahmen:

Komplexität entspricht Markiertheit;

Markiertheit und Frequenz sind gegenläufig)

① Transitivität (TRANS):

kausaler Agens bedeutet Kausalität;

kausative Konstruktionen beinhalten zwei Ereignisse (*causing event, core event*);

komplexer als ein Ereignis;

↪ Markiertheit: unergativ, unakkusativ > transitiv;

↪ Frequenz (TRANS): unergativ, unakkusativ < transitiv

unergative Verben beinhalten zwei Agenten;

sprachübergreifend seltener;

↪ Markiertheit: unergativ > unakkusativ

↪ Frequenz (TRANS): unergativ < unakkusativ

Empirische Indikatoren für Eigenschaften der Verbklassen

② Kausalität (KAUS):

bei kausativer Alternation (unergativ, unakkusativ) entsprechen sich intransitives Subjekt und transitives Objekt;

↪ Überlappung von Nomen in diesen grammatischen Funktionen;

↪ Frequenz (KAUS): unergativ, unakkusativ > transitiv

kausativ bedeutet transitiv, und unergative Verben sind selten transitiv;

↪ Frequenz (KAUS): unakkusativ > unergativ

③ Belebtheit (BELEBT):

unergative und transitive Verben haben agentivische transitive und intransitive Subjekte;

häufiger intransitiver Gebrauch unakkusativer Verben;

↪ Frequenz agentivisches Subjekt: unakkusativ < unergativ, transitiv

Agenten sind häufiger belebt als Themen;

↪ Frequenz Subjekt (BELEBT): unakkusativ < unergativ, transitiv

Empirische Indikatoren für Eigenschaften der Verbklassen

④ Passivität (PASS):

transitive Konstruktion \rightsquigarrow Passivität;
 \rightsquigarrow Frequenz aktive vs. passive Konstruktionen

⑤ Zeitform (ZEIT):

transitive Konstruktion \rightsquigarrow Partizip Perfekt;
 \rightsquigarrow Frequenz Partizip Perfekt

Indikatoren und erwartetes Verhalten

The features and expected behavior.

Feature	Expected Frequency Pattern	Explanation
Transitivity	Unerg < Unacc < ObjDrop	Unaccusatives and unergatives have a causative transitive, hence lower transitive use. Furthermore, unergatives have an agentive object, hence very low transitive use.
Causativity	Unerg, ObjDrop < Unacc	Object-drop verbs do not have a causal agent, hence low "causative" use. Unergatives are rare in the transitive, hence low causative use.
Animacy	Unacc < Unerg, ObjDrop	Unaccusatives have a Theme subject in the intransitive, hence lower use of animate subjects.
Passive Voice	Unerg < Unacc < ObjDrop	Passive implies transitive use, hence correlated with transitive feature.
VBN Tag	Unerg < Unacc < ObjDrop	Passive implies past participle use (VBN), hence correlated with transitive (and passive).

Daten

- Korpora:
65 Millionen Wörter *Wall Street Journal (WSJ)* getaggt;
29 Millionen Wörter WSJ geparkt
- Verben: je 20 Verben aus drei Klassen; Basis: Levin (1993)
- Korpora ohne Bedeutungsannotation
- Frequenzen für transitiven vs. intransitiven Gebrauch, aktiven vs. passiven Gebrauch, Gegenwart vs. Vergangenheit, kausativen vs. nicht-kausativen Gebrauch, belebtes vs. nicht-belebtes Subjekt
⇒ **distributionelle Schätzwerte**

Verb-Auswahl

Verbs used in the experiments.

Class Name	Description	Selected Verbs
Unergative	manner of motion	<i>jumped, rushed, marched, leaped, floated, raced, hurried, wandered, vaulted, paraded (group 1); galloped, glided, hiked, hopped, jogged, scooted, scurried, skipped, tiptoed, trotted (group 2).</i>
Unaccusative	change of state	<i>opened, exploded, flooded, dissolved, cracked, hardened, boiled, melted, fractured, solidified (group 1); collapsed, cooled, folded, widened, changed, cleared, divided, simmered, stabilized (group 2).</i>
Object-Drop	unexpressed object alternation	<i>played, painted, kicked, carved, reaped, washed, danced, yelled, typed, knitted (group 1); borrowed, inherited, organized, rented, sketched, cleaned, packed, studied, swallowed, called (group 2).</i>

Schätzwerte

Schätzwerte, Korpus getaggt:

- **Transitivität:** Verb, das unmittelbar einer Form von *be* folgt oder unmittelbar von einem potentiellen Objekt (Zahl, Pronomen, Artikel, Adjektiv, Nomen) gefolgt wird, gilt als transitiv
- **Passivität:** Vollverb, das im Imperfekt erscheint, oder Verb, das im Partizip Perfekt erscheint und dessen nächstes vorhergehendes Hilfsverb *have* ist, gilt als aktiv; Verb, dessen nächstes vorhergehendes Hilfsverb *be* ist, gilt als passiv
- **Zeitform:** Frequenz der Wortarten VBN (Partizip Perfekt) und VBD (Vollverb im Imperfekt)

Schätzwerte

Schätzwerte, Korpus gepart:

- **Kausalität:** Überlappung von Nomen in Subjekt- vs. Objektfunktion
- **Belebtheit:** Anzahl von Pronomen (*I, we, you, she, he, they*; nicht: *it*) in Subjektposition
(Annahme: Pronomen sind meistens belebt; manueller Test von *they*: 76% belebt)

Daten-Analyse

- Anzahl Datenpunkte:
 - **Transitivität**: 27.403
 - **Passivität**: 20.481
 - **Zeitform**: 36.297
 - **Kausalität**: 11.307
 - **Belebtheit**: 7.542
- beobachtete Distributionen entsprechen zu großem Anteil Vorhersagen
- Vergleich automatische vs. manuelle Distributionen bei sechs Verben (je zwei aus den drei Klassen): Ungenauigkeiten vor allem bei unakkusativen und transitiven Verben, aber Muster entsprechen den linguistischen Annahmen

Daten-Analyse

Manually (Man) and automatically (Aut) calculated features for a random sample of verbs.
T = TRANS, P = PASS, V = VBN, C = CAUS, A = ANIM.

	Unergative				Unaccusative				Object-Drop			
	hopped		scurried		folded		stabilized		inherited		swallowed	
	Man	Aut	Man	Aut	Man	Aut	Man	Aut	Man	Aut	Man	Aut
T	0.21	0.21	0.00	0.00	0.71	0.23	0.24	0.18	1.00	0.64	0.96	0.35
P	0.00	0.00	0.00	0.00	0.44	0.33	0.19	0.13	0.39	0.13	0.54	0.44
V	0.03	0.00	0.10	0.00	0.56	0.73	0.71	0.92	0.56	0.60	0.64	0.79
C	0.00	0.00	0.00	0.00	0.54	0.00	0.24	0.35	0.00	0.06	0.00	0.04
A	0.93	1.00	0.90	0.14	0.23	0.00	0.02	0.00	0.58	0.32	0.35	0.22

Automatische Klassifikation

- Frequenz-Verteilungen entsprechen Vektor-Beschreibungen:

Schablone: [Verb, TRANS, PASS, ZEIT, KAUS, BELEBT, Klasse]

Template: [verb, TRANS, PASS, TENSE/VBN, CAUS, ANIM, class]

Beispiel: [opened, .69, .09, .21, .16, .36, unacc]

- Klassifikator: Entscheidungsbaum (Decision Tree)
- Methoden: 10-fold cross-validation; single hold-out

Ergebnisse

Percent accuracy and standard error of the verb classification task using each feature individually, under a training methodology of 10-fold cross-validation repeated 50 times.

Feature	%Accuracy	%SE
CAUS	55.7	.1
VCN	52.5	.5
PASS	50.2	.5
TRANS	47.1	.4
ANIM	35.3	.5

Percent accuracy and standard error of the verb classification task using features in combination, under a training methodology of 10-fold cross-validation repeated 50 times.

Features Used	Feature Not Used	%Accuracy	%SE
1. TRANS PASS VBN CAUS ANIM		69.8	.5
2. TRANS VBN CAUS ANIM	PASS	69.8	.5
3. TRANS PASS VBN ANIM	CAUS	67.3	.6
4. TRANS PASS CAUS ANIM	VBN	66.5	.5
5. TRANS PASS VBN CAUS	ANIM	63.2	.6
6. PASS VBN CAUS ANIM	TRANS	61.6	.6

Ergebnisse

F score of classification within each class, under a single hold-out training methodology.

Features Used	Feature Not Used	F score (%) for Unergts	F score (%) for Unaccs	F score (%) for Objdrops
1. TRANS PASS VBN CAUS ANIM		73.9	68.6	64.9
2. TRANS VBN CAUS ANIM	PASS	76.2	75.7	61.6
3. TRANS PASS VBN ANIM	CAUS	65.1	60.0	62.8
4. TRANS PASS CAUS ANIM	VBN	66.7	65.0	51.3
5. TRANS PASS VBN CAUS	ANIM	72.7	47.0	60.0
6. PASS VBN CAUS ANIM	TRANS	78.1	51.5	61.9

Zusammenfassung

- Argumentstruktur ist ein zentraler Faktor bei semantischer Klassifikation von Verben
- empirische Distributionen können linguistische Eigenschaften modellieren
- Korpus-basierte, grobe Heuristiken sind nützlich als distributionelle Beschreibungen
- Ambiguität von Verben ist nicht berücksichtigt

Verb-Klassifikation

- Automatische Induktion von semantischen Verb-Klassen
- Mögliche Quellen: manuelle Ressourcen vs. Korpora
- Syntax-semantische Schnittstelle nutzbar machen:
Bedeutung ↔ Verhalten
- Potentiell saliente Merkmale für Verb-Modellierung:
 - syntaktische Subkategorisierung
 - Präpositionalphrasen
 - Argumente: Lexeme oder Generalisierungen
 - Adverbien etc.
- Granularität: unklar

Vorgehensweise

- Hypothese: **Verb-Verhalten** ↔ **Verb-Bedeutung**
- Distributionelle Verb-Beschreibungen:
syntaktische Subkategorisierungsrahmen, Präpositionalphrasen, Selektionspräferenzen
- Algorithmus: *k*-Means-Algorithm
- Ergebnis: semantische Verbklassen

Beispiel: Subkategorisierungsrahmen

	Rahmen	Freq
<i>glauben</i>	$NP_{nom}-S_{dass}$	1,929
	$NP_{nom}-S_2$	1,888
	$NP_{nom}-PP$	687
	NP_{nom}	608
	$NP_{nom}-NP_{akk}$	555
	$NP_{nom}-INF$	346
	$NP_{nom}-NP_{dat}$	234
	$NP_{nom}-NP_{akk}-NP_{dat}$	160
	$NP_{nom}-NP_{dat}-S_2$	70
	$NP_{nom}-NP_{akk}-INF$	62

Beispiel: Subkategorisierungsrahmen mit PP-Information

	Rahmen	Freq
<i>reden</i>	<i>NPnom-PP</i>	1,121
	<i>NPnom-PPakk.uber</i>	480
	<i>NPnom-PPdat.von</i>	463
	<i>NPnom-PPdat.mit</i>	280
	<i>NPnom-PPdat.in</i>	81
	<i>NPnom-PPnom.vgl</i>	14
	<i>NPnom-PPdat.bei</i>	13
	<i>NPnom-PPdat.uber</i>	13
	<i>NPnom-PPdat.an</i>	12
	<i>NPnom-PPakk.fur</i>	10
	<i>NPnom-PPdat.nach</i>	8

Beispiel: Selektionspräferenzen durch Nomen

reden über_{akk}

Nomen	Freq
Geld	19
Politik	14
Problem	13
Thema	10
Inhalt	9
Koalition	6
Ding	5
Freiheit	5
Kunst	5
Film	5

Beispiel: Selektionspräferenzen durch Nomen

	Nomen	Freq
<i>verfolgen – NPakk</i>	Ziel	86
	Strategie	27
	Politik	25
	Interesse	22
	Konzept	17
	Entwicklung	16
	Kurs	14
	Spiel	12
	Plan	11
	Spur	11

GermaNet Top-Level-Knoten (veraltete Version)

- Lebewesen
- Sache
- Besitz
- Substanz
- Nahrung
- Mittel
- Situation
- Zustand
- Struktur
- Physis
- Zeit
- Ort
- Attribut
- Kognitives Objekt
- Kognitiver Prozess

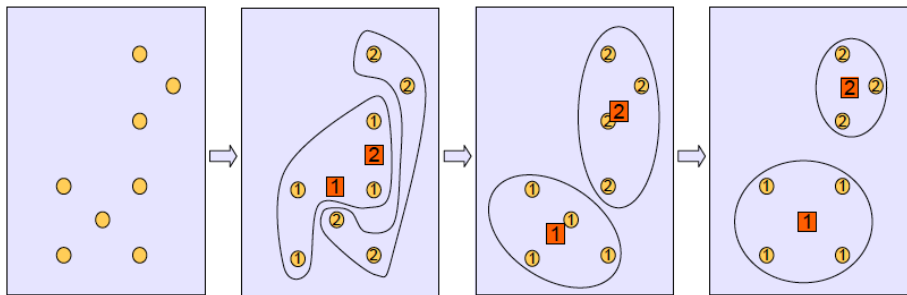
Beispiel: Selektionspräferenzen durch GermaNet

	Synset (Top-Level-Knoten)	Freq
<i>verfolgen – NPakk</i>	Situation	141
	Kognitives Objekt	110
	Zustand	81
	Sache	61
	Attribut	53
	Lebewesen	47
	Ort	46
	Struktur	14
	Kognitiver Prozess	12
	Zeit	5
	Besitz	3
	Substanz	2
	Nahrung	2
	Physis	1

k-Means-Algorithmus

- k-Means-Algorithmus (Forgy, 1965)
- Unsupervised Hard Clustering (Clustering vs. Klassifikation)
- n Objekte \rightarrow k Cluster
- Iterative Reorganisation von Cluster-Zugehörigkeit:
 - ① Anfangs-Klassierung jedes Objekts
 - ② Berechnung der Cluster-Schwerpunkte
 - ③ Ermitteln des nächsten Clusters (Schwerpunkt)
 - ④ Reorganisation der Cluster-Zugehörigkeit;
weiter mit Schritt 2
- Einfluss von Initialisierung und Ähnlichkeitsmaß

k-Means-Algorithmus: Illustration



Beispiel: Input-Klassifikation

- konsumieren kriegen vermuten
- anfangen
- ahnen bekanntgeben bestehen **fahren fliegen** liegen nieseln pochen
- aufhören **bekommen erhalten** essen insistieren regnen segeln vermitteln
- beginnen freuen interpretieren
- rudern saufen schneien ärgern
- eröffnen folgen glauben
- zustellen
- charakterisieren dämmern stehen
- blitzen verkünden wissen
- beschreiben **dienen** donnern schließen **unterstützen**
- beenden darstellen **liegen sitzen**
- ankündigen denken enden lesen schicken öffnen
- beharren bringen erlangen helfen trinken

Beispiel: Ergebnis-Klassifikation

- ahnen vermuten wissen – Propositional Attitude
- denken glauben – Propositional Attitude
- anfangen aufhören beginnen beharren enden insistieren rudern – Aspect
- liegen sitzen stehen – Position
- dienen folgen helfen – Support
- nieseln regnen schneien – Weather
- dämmern
- blitzen donnern segeln – Weather
- bestehen fahren fliegen pochen – Insistence, Manner of Motion
- freuen ärgern – Emotion
- essen konsumieren saufen trinken verkünden – Consumption
- bringen eröffnen lesen liefern schicken schließen vermitteln öffnen – Supply
- ankündigen beenden bekanntgeben bekommen beschreiben charakterisieren darstellen erhalten erlangen interpretieren kriegen unterstützen – Description, Obtain
- zustellen

Zusammenfassung

- Übereinstimmung zwischen Klassifikations-Ergebnis und manueller Klassifikation → erfolgreiche linguistische und technische Parameter
- Schwieriger Kompromiss zwischen idiosynkratischen und generellen Eigenschaften
- Ambiguität wird nicht explizit modelliert
- Was ist die Semantik der Klassen? Wie passen Situationen ins Bild, z.B. *ermorden, erschießen, töten, festnehmen, verhaften, befragen, entlassen*?
- Semantische Klassen per se vs. in Bezug auf eine Anwendung

Referenzen



Donald Hindle.

Noun Classification from Predicate-Argument Structures.

In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, pages 268–275, 1990.



Paola Merlo and Suzanne Stevenson.

Automatic Verb Classification Based on Statistical Distributions of Argument Structure.

Computational Linguistics, 27(3):373–408, 2001.



Sabine Schulte im Walde.

Experiments on the Automatic Induction of German Semantic Verb Classes.

Computational Linguistics, 32(2):159–194, 2006.