



Inducing German Semantic Verb Classes from Subcategorisation Information

Sabine Schulte im Walde
Institute for Natural Language Processing (IMS)
University of Stuttgart
Germany



Clustering Idea

- object \rightarrow cluster
- objects in the same cluster: as similar as possible
- objects in different clusters: as dissimilar as possible
- multivariate data analysis
- explores inherent data structure and natural groupings
- generalises over objects and their features



Clustering – Demarcation

- classification:
assigning objects to previously defined classes

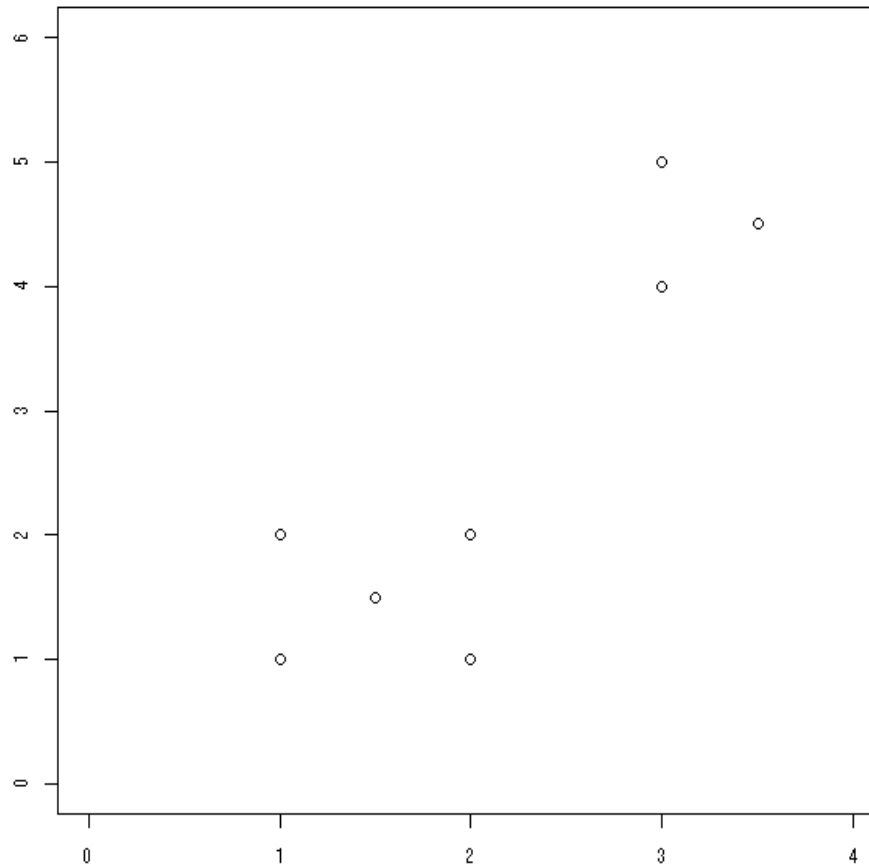
clustering:
no a priori knowledge about classes and members

- discriminant analysis:
improving a previously obtained classification by strengthening the class demarcations

clustering:
need to establish the class structure first

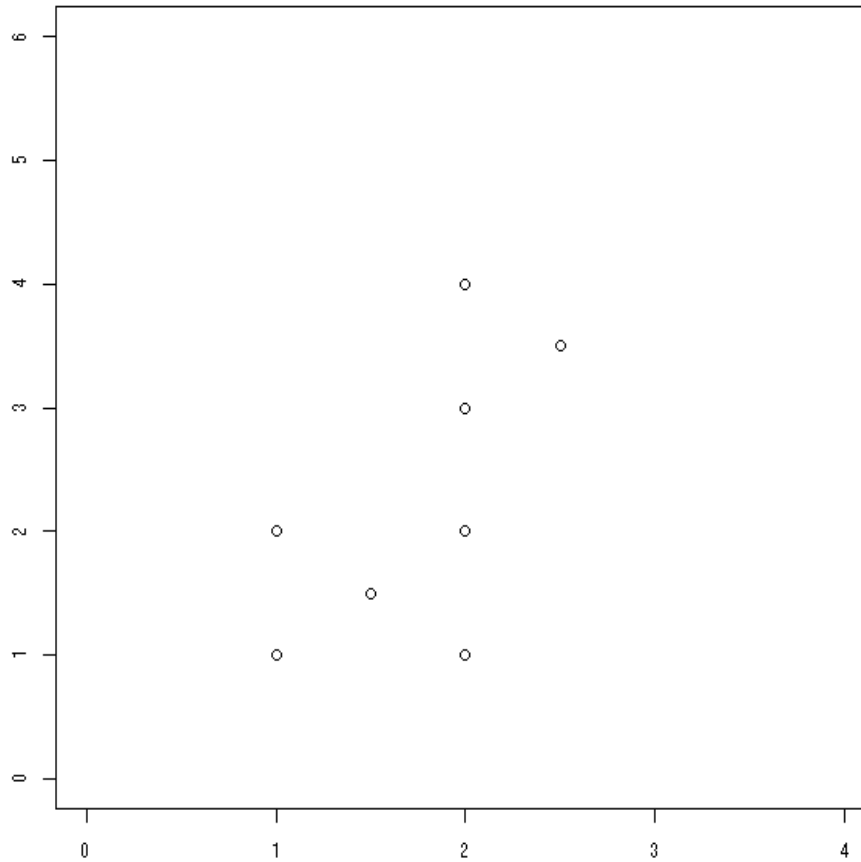


Clustering Example (1)



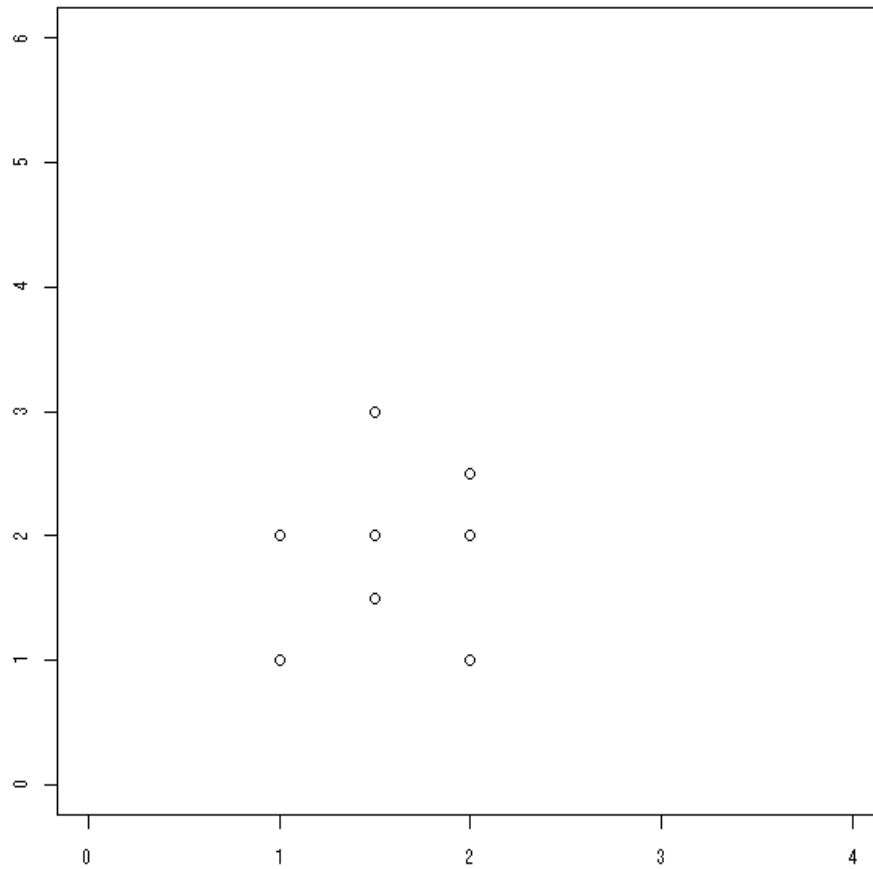


Clustering Example (2)





Clustering Example (3)





Challenges in Clustering

- object description
- multi-dimensional data
- ambiguities
- unknown cluster properties:
number, position, size, shape, density
- suitable clustering algorithm
- suitable similarity measure



Clustering Objects: German Verbs

anfangen
aufhören
beschreiben
denken
fliegen
freuen
geben
...



Object Features

- syntactic subcategorisation properties
- selectional preferences
- alternation behaviour
- morphological properties
- auxiliary choice
- voice



Feature Values

- frequency values
- probability values
- binary values



Clustering Goal: Semantic Verb Classes

- *Propositional Attitude*
ahnen, denken, glauben, vermuten, wissen
- *Aspect*
anfangen, beenden, beginnen
- *Description*
beschreiben, charakterisieren, darstellen, interpretieren
- *Manner of Motion*
fahren, fliegen, rudern, segeln
- *Position*
liegen, sitzen, stehen



Verb Classes

syntactic behaviour \leftrightarrow meaning components



Existing Verb Classes

- English (Levin 1993)
- Bangla, Korean (Jones et al. 1994)
- French (Saint-Dizier 1996)
- Spanish (Vázquez et al. 2000)



Related Work on English Verb Classes

- semantic class information based on syntactic signatures (Dorr/Jones 1996)
- verb alternation behaviour (Lapata 1999)
- verb classification based on argument structure (Stevenson/Merlo 1999)
- subcategorisation frames, selectional preferences, and alternation behaviour (McCarthy 2001)
- subcategorisation frames and selectional preferences (Schulte im Walde 2000)



Clustering Usage

- machine translation (Dorr 1997)
- document classification (Klavans/Kan 1998)
- word sense disambiguation (Dorr/Jones 1996)
- smoothing \rightsquigarrow machine translation (Prescher et al. 2000)
- probabilistic grammars (Riezler et al. 2000)



Clustering Approach

1. acquisition of lexical verb information
 - purely syntactic subcategorisation frames:
 - (a) coarse frame definition
 - (b) extension by prepositional phrase definition
 - selectional preferences
2. clustering by k-Means
 - find good values for parameters of clustering process
 - explore role of frame descriptions in verb classification
3. evaluation against manual verb classification



Subcategorisation Frame Elements

n	NP _{Nom}
a	NP _{Acc}
d	NP _{Dat}
r	reflexive pronoun
p	prepositional phrase
x	expletive <i>es</i>
i	subordinated non-finite clause
s-2	subordinated finite verb second clause
s-dass	subordinated finite <i>dass</i> -clause
s-ob	subordinated finite <i>ob</i> -clause
s-w	indirect <i>wh</i> -questions
k	copula constructions



Lexical Verb Information (1): Subcategorisation Frame Definition

Frame	Freq	Prob
ns-dass	1,929	0.27945
ns-2	1,888	0.27358
np	687	0.09951
n	608	0.08811
na	555	0.08046
ni	346	0.05015
nd	234	0.03392
nad	160	0.02325
nds-2	70	0.01011

probability distribution for *glauben* 'to think/believe'
(probability values >1%)



Lexical Verb Information (2): Subcategorisation Frame Definition + PPs

Refined Frame		Freq	Prob
np:Akk.über	acc / 'about'	480	0.11981
np:Dat.von	dat / 'about'	463	0.11568
np:Dat.mit	dat / 'with'	280	0.06983
np:Dat.in	dat / 'in'	81	0.02031

refined np probability distribution for *reden* 'to talk'
with total joint probability $p(\text{reden}, \text{np}) = 0.35820$
(probability values >1%)



Selectional Preference Acquisition – Approach

- basic approaches: Resnik (1997), Ribas (1995)
- input: nominal fillers, e.g. taken from $VP^{[verb]} \rightarrow V' NP.Akk^{[noun]}$
- preference concepts: WordNet synsets (Fellbaum 1998) \rightsquigarrow GermaNet synsets (Hamp/Feldweg 1997)
- preference strength: relative entropy between prior and posterior distribution (of verbs) over noun concepts with regard to a syntactic environment

$$SelPref(v, n) = p_{syn}(v, n) * \log \frac{p_{syn}(v, n)}{p_{syn}(n)} \quad (1)$$



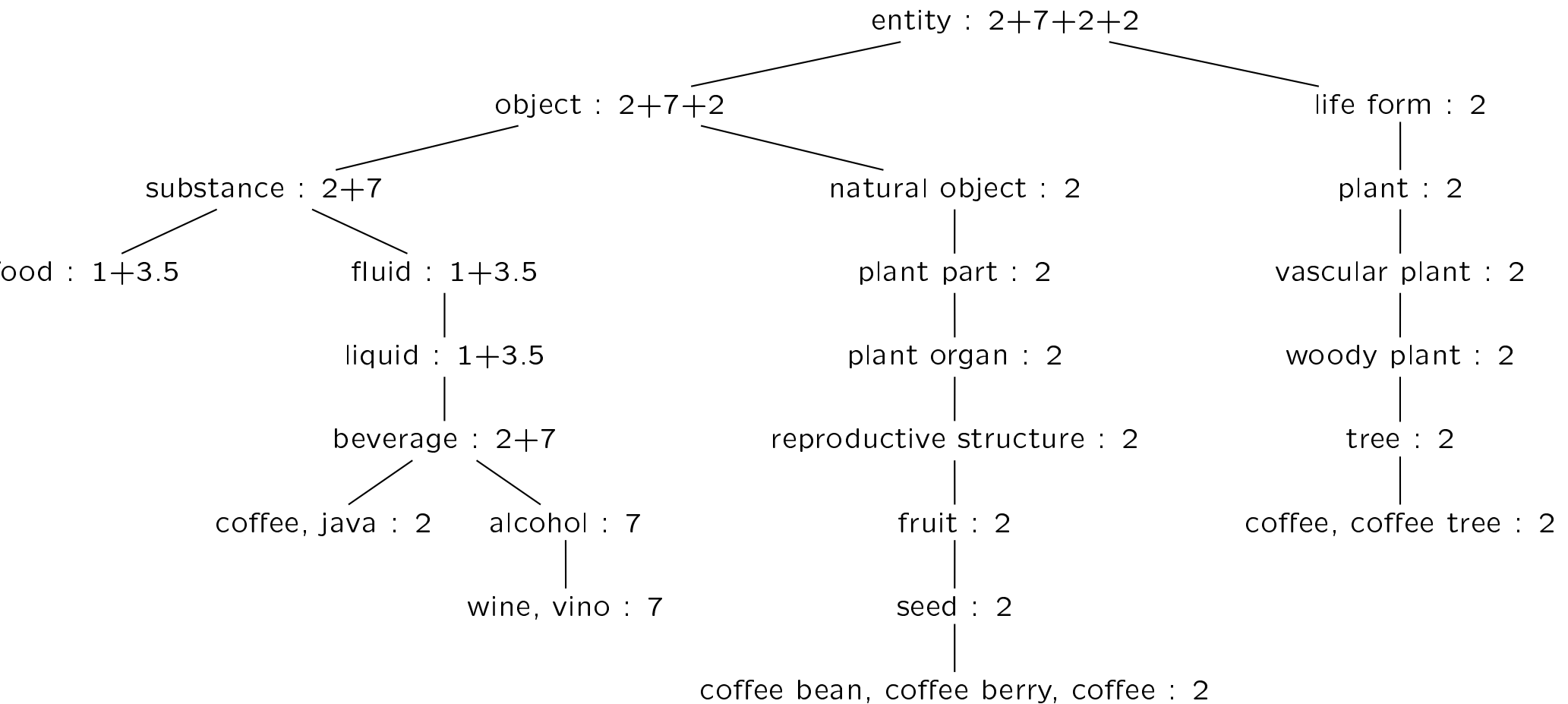
Selectional Preference Acquisition – Example

Noun		Freq
Fleisch	'meat'	34
Brot	'bread'	14
Fisch	'fish'	14
Suppe	'soup'	7
Eis	'ice-cream'	6
Sache	'thing'	6
Gemüse	'vegetables'	6
Kartoffel	'potato'	5
Kuchen	'cake'	5
Wurst	'sausage'	5
Tier	'animal'	4
Schokolade	'chocolate'	4
Pizza	'pizza'	4
Ei	'egg'	4
Würstchen	'sausage'	3
Wurm	'worm'	3
Tomate	'tomato'	3
Teller	'plate'	3

nominal direct object fillers for *essen* 'to eat'

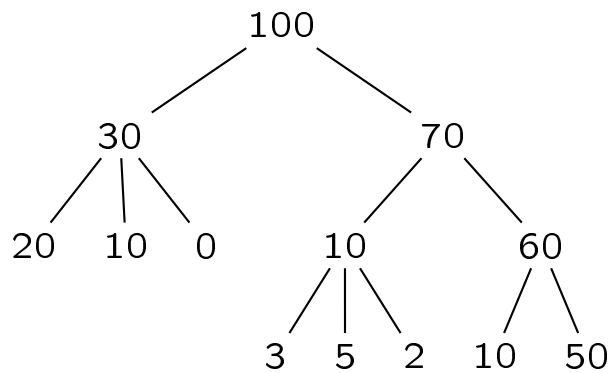


Selectional Preference Acquisition – Example

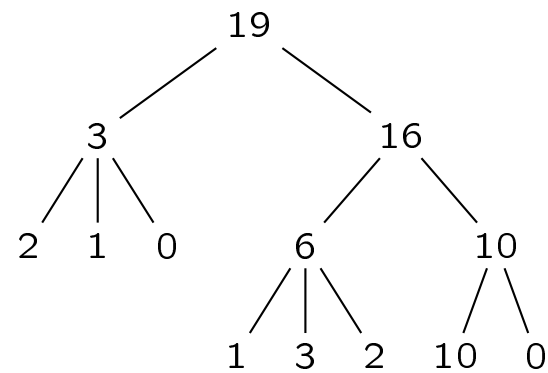




Selectional Preference Acquisition – Example



total freq = 100



total freq = 19



Selectional Preference Acquisition – Example

Noun Synset in NP _{Acc}		SelPref
Nahrung	'food'	0.13190
feste Nahrung	'solid food'	0.11777
Objekt	'object'	0.06189
Fleisch	'meat'	0.04813
Backware	'bakery products'	0.02898
Gericht	'food, dish'	0.02272
Grünzeug	'healthy food'	0.02247
Brot	'bread'	0.02008
Tier	'animal'	0.01903
Gemüse	'vegetables'	0.01711

selectional preferences for *essen* 'to eat'



Lexical Verb Information (3): Frame Definition + PPs + SelPrefs

Refined Frame		SelPref	Prob
na_NP _{Akk} :Nahrung, Lebensmittel	'food'	1.00985	0.19439
na_NP _{Akk} :Lebewesen, Kreatur, Wesen	'creature'	0.10324	0.01987
na_NP _{Akk} :Stoff, Substanz, Materie	'substance'	0.02659	0.00512
na_NP _{Akk} :Zeit	'time'	0.01247	0.00240
na_NP _{Akk} :Hilfsmittel, Mittel	'means'	0.00099	0.00019
na_NP _{Akk} :Körper, Physis	'body'	-0.00037	0.00000
na_NP _{Akk} :Struktur	'structure'	-0.00087	0.00000
na_NP _{Akk} :Attribut, Eigenschaft	'property'	-0.00161	0.00000
na_NP _{Akk} :?kognitiver_Prozess	'cognitive process'	-0.00288	0.00000
na_NP _{Akk} :Besitz	'property'	-0.00344	0.00000
na_NP _{Akk} :Stelle, Ort, Stätte	'place'	-0.00758	0.00000
na_NP _{Akk} :Zustand	'state'	-0.01000	0.00000
na_NP _{Akk} :Situation	'situation'	-0.01218	0.00000
na_NP _{Akk} :?kognitives_Objekt	'cognitive object'	-0.01301	0.00000
na_NP _{Akk} :Ding, Sache	'thing'	-0.10120	0.00000

refined na probability distribution for *essen* 'to eat'
with total joint probability $p(\text{essen}, \text{na}) = 0.22197$



German Semantic Verb Classes – Overview

reduced		all	
classes	verbs	classes	verbs
14	57	43	168

- relation to (Levin 1993)
- consistency with (Schumacher 1986)
- class size: 2-6/7
- high and low frequency verbs: $8 \leq \text{freq} \leq 31,710$
- basis: semantic intuition



German Semantic Verb Classes – Definition

1. *Aspect*: anfangen, aufhören, beenden, beginnen, enden
2. *Propositional Attitude*: ahnen, denken, glauben, vermuten, wissen
3. *Transfer of Possession (Obtaining)*: bekommen, erhalten, erlangen, kriegen
4. *Transfer of Possession (Supply)*: bringen, liefern, schicken, vermitteln, zustellen
5. *Manner of Motion*: fahren, fliegen, rudern, segeln
6. *Emotion*: ärgern, freuen
7. *Announcement*: ankündigen, bekanntgeben, eröffnen, verkünden
8. *Description*: beschreiben, charakterisieren, darstellen, interpretieren
9. *Insistence*: beharren, bestehen, insistieren, pochen
10. *Position*: liegen, sitzen, stehen
11. *Support*: dienen, folgen, helfen, unterstützen
12. *Opening*: öffnen, schließen
13. *Consumption*: essen, konsumieren, lesen, saufen, trinken
14. *Weather*: blitzen, donnern, dämmern, nieseln, regnen, schneien



German Semantic Verb Classes – Definition (1)

1. *Aspect*: anfangen, aufhören, beenden, beginnen, enden
2. *Propositional Attitude*: ahnen, denken, glauben, vermuten, wissen
3. *Desire*
 - (a) *Wish*: erhoffen, wollen, wünschen
 - (b) *Need*: bedürfen, benötigen, brauchen
4. *Transfer of Possession (Obtaining)*: bekommen, erhalten, erlangen, kriegen
5. *Transfer of Possession (Giving)*
 - (a) *Gift*: geben, leihen, schenken, spenden, stiften, vermachen, überschreiben
 - (b) *Supply*: bringen, liefern, schicken, vermitteln, zustellen
6. *Emotion*
 - (a) *Origin*: ärgern, freuen
 - (b) *Expression*: heulen, lachen, weinen
 - (c) *Objection*: ängstigen, ekeln, fürchten, scheuen



German Semantic Verb Classes – Definition (2)

7. *Manner of Motion*

- (a) *Manner of Locomotion*: gehen, klettern, kriechen, laufen, rennen, schleichen, wandern
- (b) *Rotation*: drehen, rotieren
- (c) *Rush*: eilen, hasten
- (d) *Means*: fahren, fliegen, rudern, segeln
- (e) *Flotation*: fließen, gleiten, treiben

8. *Manner of Look on the Face*: gähnen, grinsen, lachen, lächeln, starren

9. *Perception*: empfinden, erfahren, fühlen, hören, riechen, sehen, wahrnehmen

10. *Manner of Articulation*: flüstern, rufen, schreien

11. *Moaning*: heulen, jammern, klagen, lamentieren

12. *Communication*: kommunizieren, korrespondieren, reden, sprechen, verhandeln

13. *Observation*: bemerken, erkennen, erfahren, feststellen, realisieren, registrieren



German Semantic Verb Classes – Definition (3)

14. *Description*: beschreiben, charakterisieren, darstellen, interpretieren
15. *Presentation*: darstellen, demonstrieren, präsentieren, veranschaulichen, vorführen
16. *Statement*
 - (a) *Announcement*: ankündigen, bekanntgeben, eröffnen, verkünden
 - (b) *Constitution*: anordnen, bestimmen, festlegen
 - (c) *Promise*: versichern, versprechen, zusagen
17. *Speculation*: grübeln, nachdenken, phantasieren, spekulieren
18. *Insistence*: beharren, bestehen, insistieren, pochen
19. *Teaching*: beibringen, lehren, unterrichten, vermitteln
20. *Position*
 - (a) *Bring into Position*: legen, setzen, stellen
 - (b) *Be in Position*: liegen, sitzen, stehen
21. *Production*: bilden, erzeugen, herstellen, hervorbringen, produzieren



German Semantic Verb Classes – Definition (4)

- 22. *Renovation*: dekorieren, erneuern, renovieren, reparieren
- 23. *Support*: dienen, folgen, helfen, unterstützen
- 24. *(Active) Change of Quantum*: erhöhen, erniedrigen, senken, steigern, vergrößern, verkleinern
- 25. *Opening*: öffnen, schließen
- 26. *Existence*: bestehen, existieren, leben
- 27. *Consumption*: essen, konsumieren, lesen, saufen, trinken
- 28. *Elimination*: eliminieren, entfernen, exekutieren, töten, vernichten
- 29. *Basis*: basieren, beruhen, gründen, stützen
- 30. *Inference*: folgern, schließen
- 31. *Result*: ergeben, erwachsen, folgen, resultieren
- 32. *Weather*: blitzen, donnern, dämmern, nieseln, regnen, schneien



Clustering Methodology

- k-Means algorithm (MacQueen 1967)
- unsupervised hard clustering
- n objects $\rightarrow k$ clusters
- iterative re-organisation of cluster membership:
 1. initial cluster assignment
 2. calculation of cluster centroids
 3. determining closest cluster (centroid)
 4. re-arrangement of cluster membership
- locally optimal solutions by minimising sum-of-squared-error between objects and respective cluster centroid
- convergence proof (Selim/Ismail 1984)



Similarity Measures

- Kullback-Leibler divergence

$$d(v_1, v_2) = D(p \parallel q) = \sum_i p_i \log \frac{p_i}{q_i} \quad (2)$$

- information radius

$$d(v_1, v_2) = D(p \parallel \frac{p + q}{2}) + D(q \parallel \frac{p + q}{2}) \quad (3)$$

- skew divergence

$$d(v_1, v_2) = D(p \parallel w * q + (1 - w) * p) \quad (4)$$

with weight w set to 0.9



Cluster Initialisation

- random clusters
- agglomerative hierarchical clusters
 - single-linkage
 - complete-linkage
 - average verb distance
 - distance between cluster centroids
 - Ward's method



Clustering Evaluation – Basis

gold standard: hand-constructed verb classes

- evaluation of individual clusters
- evaluation of complete clustering:
similarity between two sets of equivalence relations



Clustering Evaluation – Methods

- linguistically intuitive desiderata based on pair-wise precision
- silhouette value (Kaufman/Rousseeuw 1990)
- mutual information (Strehl et al. 2000)
- class-wise precision/recall (Vilain et al. 1995)
- (adjusted) Rand index (Rand 1971), (Hubert/Arabie 1985)



Linguistically Intuitive Desiderata (1)

(1) completely correct clusters, different in size

geben

geben leihen

geben leihen schenken

geben leihen schenken spenden

geben leihen schenken spenden stiften

geben leihen schenken spenden stiften vermachen

geben leihen schenken spenden stiften vermachen überschreiben

(2) constant correct cluster part, step-wise impurity

ahnen vermuten wissen

ahnen vermuten wissen laufen

ahnen vermuten wissen laufen lachen

(2') decreasing correct cluster part, step-wise impurity

ahnen vermuten wissen

ahnen vermuten laufen

ahnen laufen lachen



Linguistically Intuitive Desiderata (2)

(3) step-wise impurity by consistent verbs

ahnen vermuten wissen

ahnen vermuten wissen laufen

ahnen vermuten wissen laufen rennen

ahnen vermuten wissen laufen rennen wandern

(4) correct merge of clusters

ahnen vermuten wissen

denken glauben

ahnen vermuten wissen denken glauben

(4') wrong merge of clusters

ahnen vermuten wissen

laufen rennen wandern

ahnen vermuten wissen laufen rennen wandern

(5) wrong clusters of different size

benötigen

benötigen lachen

benötigen lachen wissen



Linguistically Intuitive Desiderata (3)

- extreme cases
 - clustering result: one cluster with all verbs
 - clustering result: clustering consisting of singletons
- step-wise error introduction



Adjusted Pair-Wise Precision

- individual cluster evaluation:
based on linguistically intuitive desiderata

$$APP(A_i) = \frac{\text{number of correct pairs in } A_i}{\text{number of verbs in } A_i + 1} \quad (5)$$

- whole clustering evaluation:
weighted average over qualities of individual clusters

$$APP(C) = \frac{1}{|C|} \sum_j \frac{APP(C_j)}{|C_j|} \quad (6)$$



Silhouette Value

- how well does an object lie within a cluster?
- individual cluster evaluation: comparing average distance of object to same cluster objects and neighbour cluster objects; average over objects in cluster

$$a(o_i) = \frac{1}{|c_A| - 1} \sum_{o_j \in c_A, o_j \neq o_i} d(o_i, o_j) \quad (7)$$

$$b(o_i) = \min_{c_B \neq c_A} \frac{1}{|c_B|} \sum_{o_j \in c_B} d(o_i, o_j) \quad (8)$$

$$sil(o_i) = \frac{b(o_i) - a(o_i)}{\max\{a(o_i), b(o_i)\}} \quad (9)$$

- whole clustering evaluation: average silhouette value for individual clusters



Mutual Information

- no dependence on similarity measure or original dimensionality of input data
- individual cluster evaluation:
maximal number of verbs agreeing in automatic and manual clusters

$$\text{purity}(A_i) = \max_k \frac{|\{x_j \mid A[x_j] = i, B[x_j] = k\}|}{|\{x_j \mid A[x_j] = i\}|} \quad (10)$$

- whole clustering evaluation:
mutual information between automatic and manual clusterings

$$MI(A, B) = \frac{1}{n} \sum_i \sum_j AB_{ij} \frac{\log\left(\frac{AB_{ij} \cdot n}{\sum_i AB_{ij} \sum_j AB_{ij}}\right)}{\log(|A| \cdot |B|)} \quad (11)$$



Random Clustering Input – Example

- a) konsumieren, kriegen, vermuten
- b) anfangen
- c) ahnen, bekanntgeben, bestehen, **fahren, fliegen**, liefern, nieseln, pochen
- d) aufhören, **bekommen, erhalten**, essen, insistieren, regnen, segeln, vermitteln
- e) beginnen, freuen, interpretieren
- f) rudern, saufen, schneien, ärgern
- g) eröffnen, folgen, glauben
- h) zustellen
 - i) charakterisieren, dämmern, stehen
 - j) blitzen, verkünden, wissen
- k) beschreiben, **dienen**, donnern, schließen, **unterstützen**
- l) beenden, darstellen, **liegen, sitzen**
- m) ankündigen, denken, enden, lesen, schicken, öffnen
- n) beharren, bringen, erlangen, helfen, trinken



Clustering Result (ip) – Verb Classes

- a) **ahnen, vermuten, wissen** (0.75) *Propositional Attitude*
- b) **denken, glauben** (0.33) *Propositional Attitude*
- c) anfangen, aufhören, **beharren, bestehen, pochen**, rudern (0.57) *Insistence*
- d) beginnen, enden, **fahren, fliegen, liegen, segeln, sitzen, stehen** (0.78) *Manner of Motion and Position*
- e) **dienen, folgen, helfen** (0.75) *Support*
- f) **nieseln, regnen, schneien** (0.75) *Weather*
- g) **dämmern** (0.00) *Weather*
- h) **blitzen, donnern**, insistieren (0.25) *Weather*
- i) **freuen, ärgern** (0.33) *Emotion*
- j) lesen, saufen, schließen, öffnen (0.40) *Consumption and Opening*
- k) **essen, konsumieren**, kriegen, **trinken**, verkünden (0.50) *Consumption*
- l) ankündigen, beenden, bekanntgeben, bekommen, **beschreiben, bringen, charakterisieren, darstellen**, erhalten, erlangen, eröffnen, **interpretieren, liefern, schicken**, unterstützen, **vermitteln** (1.06) *Supply and Description*
- m) **zustellen** (0.00) *Supply*



Clustering Result (ip-pp) – Verb Classes

- a) **ahnen, vermuten, wissen** (0.75) *Propositional Attitude*
- b) **denken, glauben** (0.33) *Propositional Attitude*
- c) **anfangen, aufhören, beginnen, beharren, enden, insistieren, rudern** (0.88) *Aspect*
- d) **liegen, sitzen, stehen** (0.75) *Position*
- e) **dienen, folgen, helfen** (0.75) *Support*
- f) **nieseln, regnen, schneien** (0.75) *Weather*
- g) **dämmern** (0.00) *Weather*
- h) **blitzen, donnern, segeln** (0.25) *Weather*
- i) **bestehen, fahren, fliegen, pochen** (0.40) *Insistence, Manner of Motion*
- j) **freuen, ärgern** (0.33) *Emotion*
- k) **essen, konsumieren, saufen, trinken, verkünden** (1.00) *Consumption*
- l) **bringen, eröffnen, lesen, liefern, schicken, schließen, vermitteln, öffnen** (0.78) *Supply*
- m) **ankündigen, beenden, bekanntgeben, bekommen, beschreiben, charakterisieren, darstellen, erhalten, erlangen, interpretieren, kriegen, unterstützen** (1.00) *Description and Obtaining*
- n) **zustellen** (0.00) *Supply*



Clustering Result (ip) – Quality

		random	hierarchical				
			single	complete	average	centroid	ward
irad	APP	0.125	0.043	0.087	0.079	0.073	0.101
	MI	0.328	0.226	0.277	0.262	0.250	0.304
skew	APP	0.111	0.043	0.091	0.067	0.062	0.102
	MI	0.315	0.226	0.281	0.256	0.252	0.349

baseline (APP) = 0.017

baseline (MI) = 0.229



Clustering Result (ip-pp) – Quality

		random	hierarchical				
			single	complete	average	centroid	ward
irad	APP	0.144	0.107	0.123	0.118	0.081	0.151
	MI	0.357	0.229	0.319	0.298	0.265	0.332
skew	APP	0.114	0.104	0.126	0.118	0.081	0.159
	MI	0.320	0.289	0.330	0.298	0.265	0.372

baseline (APP) = 0.017

baseline (MI) = 0.229



Clustering Result (ip) – Feature Role p

- a) **glauben, vermuten** (0.33) *Propositional Attitude*
- b) **denken**, insistieren, **wissen** (0.25) *Propositional Attitude*
- c) **anfangen, aufhören, beginnen**, beharren, bestehen, rudern (0.57)
- d) enden, **liegen**, pochen, segeln, **sitzen, stehen** (0.43) *Position*
- e) **dienen, folgen** (0.33) *Support*
- f) **helfen** (0.00) *Support*
- g) **nieseln, regnen, schneien** (0.75) *Weather*
- h) **blitzen** (0.00) *Weather*
- i) **dämmern** (0.00) *Weather*
- j) donnern, **fahren, fliegen**, saufen (0.20) *Manner of Motion*
- k) **freuen, ärgern** (0.33) *Emotion*
- l) ahnen, ankündigen, beenden, bekanntgeben, bekommen, **beschreiben, charakterisieren, darstellen**, erlangen, **interpretieren**, kriegen, unterstützen (0.77) *Description*
- m) **bringen**, erhalten, eröffnen, **liefern, schicken, vermitteln, zustellen** (1.25) *Supply*
- n) **essen, konsumieren, lesen**, schließen, **trinken**, verkünden, öffnen (0.88) *Consumption*



Clustering Result (ip-pp) – Feature Role x

- a) **ahnen, vermuten, wissen** (0.75) *Propositional Attitude*
- b) **denken, glauben** (0.33) *Propositional Attitude*
- c) **anfangen, aufhören, beginnen, beharren, enden, nieseln**, rudern (0.75)
Aspect
- d) **liegen, sitzen, stehen** (0.75) *Position*
- e) **dienen, folgen, helfen, schneien** (0.60) *Support*
- f) **dämmern** (0.00) *Weather*
- g) **blitzen, donnern, insistieren, regnen**, saufen (0.50) *Weather*
- h) bestehen, **fahren, fliegen**, pochen, **segeln** (0.67) *Manner of Motion*
- i) **freuen, ärgern** (0.33) *Emotion*
- j) eröffnen, **essen, konsumieren**, schließen, **trinken**, verkünden (0.57)
Consumption
- k) **bringen, schicken** (0.33) *Supply*
- l) lesen, **liefern, vermitteln**, öffnen (0.20) *Supply*
- m) ankündigen, beenden, bekanntgeben, **bekommen, beschreiben, charakterisieren, darstellen, erhalten, erlangen, interpretieren, kriegen**, unterstützen (1.00) *Description and Obtaining*
- n) **zustellen** (0.00) *Supply*



Clustering Result (ip-pp) – Feature Roles

- a) ahnen, ankündigen, beenden, bekanntgeben, **bekommen**, **beschreiben**, **charakterisieren**, **darstellen**, **erhalten**, **erlangen**, eröffnen, **interpretieren**, konsumieren, **kriegen**, liefern, unterstützen, vermitteln, vermuten (0.89) *Description and Obtaining*
- b) **anfangen**, **aufhören**, **beginnen**, **beharren**, **bestehen**, **enden**, **insistieren**, **pochen**, rudern (1.20) *Aspect and Insistence*
- c) denken, folgen, glauben, helfen (0.80) *Support and Aspect*
- d) **dienen** (0.00) *Support*
- e) **nieseln**, **regnen**, **schneien** (0.75) *Weather*
- f) donnern, segeln (0.00)
- g) **dämmern** (0.00) *Weather*
- h) **fahren**, **fliegen** (0.33) *Manner of Motion*
- i) **freuen**, **ärgern** (0.33) *Emotion*
- j) **liegen**, **sitzen**, **stehen** (0.75) *Position*
- k) blitzen, saufen, verkünden, wissen (0.00)
- l) lesen, **schließen**, **öffnen** (0.25) *Opening*
- m) bringen, essen, schicken, trinken (0.40) *Supply and Consumption*
- n) **zustellen** (0.00) *Supply*



Feature Role – Quality

clustering with `information radius` on random reduced verb set

	ip		ip-pp		
	missing	all	all	missing	
	p			x	s
APP	0.101	0.125	0.144	0.123	0.093
MI	0.321	0.328	0.357	0.340	0.300



Clustering Result (ip-pp-pref) – Verb Classes

1. **ahnen**, bekanntgeben, **wissen** *Propositional Attitude*
2. **anfangen**, **aufhören** *Aspect*
3. ankündigen, anordnen, beenden, **beschreiben**, bestimmen, **charakterisieren**, **darstellen**, **interpretieren**, unterstützen *Description*
4. **basieren**, **beruhen** *Basis*
5. bedürfen, dämmern
6. **beginnen**, **enden**, korrespondieren, laufen *Aspect*
7. **beharren**, **insistieren** *Insistence*
8. beibringen, **leihen**, **schenken**, **vermachen** *Gift*
9. **bekommen**, **erhalten**, **kriegen** *Obtaining*
10. **bemerken**, **erfahren**, **feststellen**, fürchten, vermuten *Observation*
11. **benötigen**, **brauchen**, eliminieren *Need*
12. **bestehen**, **existieren**, gähnen, **liegen**, **stehen** *Existence and Position*
13. bilden, entfernen, verkleinern
14. blitzen, gehen, rudern *Manner of Motion*
15. bringen, geben, treiben *Giving*
16. dekorieren, schicken, töten, unterrichten



17. demonstrieren, hören, lehren, lesen, verkünden
18. **denken**, folgern, **glauben**, versichern *Propositional Attitude*
19. **dienen**, **folgen**, **helfen** *Support*
20. donnern, rotieren, segeln *Manner of Motion*
21. drehen, ekeln
22. eilen, flüstern, phantasieren, schleichen *Manner of Motion*
23. empfinden, erzeugen, exekutieren, scheuen
24. ergeben, **freuen**, fühlen, **ärgern** *Emotion:Origin*
25. **erhoffen**, **wünschen** *Wish*
26. **erhöhen**, **senken**, **steigern** *(Active) Change of Quantum*
27. **erkennen**, erlangen, **realisieren**, **sehen**, **wahrnehmen** *Observation and Perception*
28. erneuern, liefern, präsentieren, stellen
29. erniedrigen, ängstigen
30. **erwachsen**, **resultieren** *Result*
31. eröffnen, registrieren, reparieren, veranschaulichen
32. **essen**, **trinken** *Consumption*



33. fahren, **kommunizieren**, nachdenken, **reden**, spekulieren, **sprechen**, **verhandeln** *Communication*
34. festlegen, stützen, vergrößern
35. fliegen, fließen, **klettern**, **wandern** *Manner of Motion*
36. gleiten, kriechen *Manner of Motion*
37. **grinsen**, **heulen**, **jammern**, **klagen**, **lachen**, **lächeln**, rufen, schreien, **weinen** *Manner of Look on the Face, Emotion:Expression and Moaning*
38. grübeln, lamentieren
39. gründen, renovieren, **schließen**, vernichten, **öffnen** *Opening*
40. hasten, riechen
41. **herstellen**, **hervorbringen**, stiften, vermitteln *Production and Giving*
42. konsumieren
43. leben, saufen, sitzen
44. **legen**, produzieren, **setzen**, spenden *Bring into Position*
45. **nieseln**, **regnen**, **schneien** *Weather*
46. pochen, rennen, starren
47. **versprechen**, wollen, **zusagen** *Promise*
48. vorführen, zustellen, überschreiben *Giving*



Clustering Result (ip \rightarrow pp \rightarrow pref) – Quality

ward clustering with skew divergence on full verb set

	ip	ip-pp	ip-pp-pref					all
			n/na	+nd	+nad	+np	+ns-dass	
APP	0.041	0.065	0.070	0.064	0.074	0.070	0.075	0.059
MI	0.319	0.347	0.360	0.350	0.354	0.360	0.365	0.347

baseline (APP) = 0.006

baseline (MI) = 0.303



Relevance of Verb Features

- PPs improve clustering on syntactic features
- selectional preference choice improves clustering
- extracting feature groups illustrates relevance of individual features in clustering
- clustering with choice of (top) GermaNet nodes outperforms full node usage
- syntactic slots are of different relevance for verb distinction
- (reduced) verb set improves until limited amount of features
- optimisation of feature set shows that small amount of features is nearly as successful as full feature set



Future Work

- subcategorisation information for parser improvement
- feature variation: extension, crab
- clustering approach: soft, parametrical?, without distance measure
- classification approach: supervised learning
- classification task on learned classification
- human judgement task on clustering verbs
- reliability proof of clustering stability