

German Semantic Verb Classes – Manual and Automatic Acquisition

Sabine Schulte im Walde

Brockhaus Duden Neue Medien GmbH
Mannheim

Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart

*Computational Linguistics Colloquium
January 22, 2004*

Overview

1. Semantic verb classes
2. Manual definition of German semantic verb classes
3. Automatic acquisition of German semantic verb classes
 - Statistical grammar model
 - Clustering methodology
 - Cluster analysis examples
4. FrameNet semantic verb classes for German
5. Comparison of German semantic verb classes

Motivation for Verb Classes in NLP

- Central role of verb for structure and meaning of sentence
- Lexical verb resources in Natural Language Processing
- Semantic verb classes
 - Lexical and conceptual semantic verb information
 - Redundancy reduction in verb description
 - Prediction and refinement of vague properties
- Clustering methodology for automatic class induction

Semantic Verb Classes

verb meaning components ↔ verb behaviour

[...] the behavior of a verb, particularly with respect to the expression and interpretation of its arguments, is to a large extent determined by its meaning.

Levin (1993, page 1)

Diathesis Alternation: Example

- *Der Wagen fährt in die Innenstadt.*
`The car drives to the city centre.'
- *Die Frau fährt nach Hause.*
`The woman drives home.'
- *Der Filius fährt einen blauen Ferrari.*
`The son drives a blue Ferrari.'
- *Der Junge fährt seinen Vater zum Zug.*
`The boy drives his father to the train.'

German Semantic Verb Classes: Overview

- German verbs → semantic classes
- Verb description on conceptual level
- Corpus examples at syntax-semantic interface
- Purposes:
 - a) Evaluate reliability and performance of clustering
 - Relationship verb meaning and behaviour
 - Clustering methodology for lexical resource
 - b) Usage in NLP applications

German Semantic Verb Classes: Examples

1. *Aspect*: anfangen, aufhören, beenden, beginnen, enden
Manner of Motion
8. *Locomotion*: gehen, klettern, kriechen, laufen, rennen, ...
9. *Rotation*: drehen, rotieren
10. *Rush*: eilen, hasten
11. *Vehicle*: fahren, fliegen, rudern, segeln
12. *Flotation*: fließen, gleiten, treiben
28. *Insistence*: beharren, bestehen₁, insistieren, pochen
34. *Support*: dienen, folgen, helfen, unterstützen
37. *Existence*: bestehen₂, existieren, leben

German Semantic Verb Classes: Constitution

- 168 German verbs → 43 semantic classes
- Class size: 2-7 verbs (Ø 3.9 verbs per class)
- Ambiguity: 8 verbs with 2 senses
- High and low frequency verbs: $8 \leq \text{freq} \leq 71,604$
- Basis: verb similarity in conceptual structure
- Class labels on two conceptual levels

Verb Class Scene Description: *Aspect* Verbs

Verbs:	<i>anfangen, aufhören, beenden, beginnen, enden</i>
Scene:	[E An event] begins or ends, either internally caused or externally caused by [I an initiator]. The event may be specified with respect to [T tense], [L location], [x an experiencer], or [R a result].
Frame Roles:	Initiator, E vent
Modification Roles:	T emporal, L ocal, e X periencer, R esult
Levin Class:	55.1 (Aspectual Verbs → Begin Verbs)
Schumacher Class:	22.1 (Verben der speziellen Existenz → Verben der Existenzsituierung) <i>beenden</i> is not classified.

Verb Class Frame Variants: *Aspect* Verbs (1)

Frame	Participating Verbs & Corpus Examples
n _E	<p>+ anfangen, aufhören, beginnen / +_{adv} enden / ¬ beenden</p> <p>Nun aber muß [_E der Dialog] anfangen.</p> <p>... bevor [_E der Golfkrieg] angefangen hatte ...</p> <p>Erst muß [_E das Morden] aufhören.</p> <p>[_E Der Gottesdienst] beginnt.</p> <p>[_E Das Schuljahr] beginnt [_T im Februar].</p> <p>[_X Für die Flüchtlinge] beginnt nun [_E ein Wettlauf gegen die Zeit].</p> <p>[_E Die Ferien] enden [_R mit einem großen Fest].</p> <p>[_E Druckkunst] endet [_R beim guten Buch].</p> <p>[_E Die Partie] endete [_R 0:1].</p> <p>[_E Der Informationstag] ... endet [_T um 14 Uhr].</p>

Verb Class Frame Variants: *Aspect* Verbs (2)

Frame	Participating Verbs & Corpus Examples
n_I	<p>+ anfangen, aufhören / \neg beenden, beginnen, enden</p> <p>[_I Die Hauptstadt] muß anfangen.</p> <p>... daß [_I er] [_T pünktlich] anfang.</p> <p>Jetzt können [_I wir] nicht einfach aufhören.</p> <p>Vielleicht sollte [_I ich] aufhören und noch studieren.</p>
$n_I a_E$	<p>+ anfangen, beenden, beginnen / \neg aufhören, enden</p> <p>Nachdem [_I wir] [_E die Sache] angefangen hatten.</p> <p>... [_I er] versucht, [_E ein neues Leben] anzufangen.</p> <p>[_I Die Polizei] beendete [_E die Gewalttätigkeiten].</p> <p>[_T Nach dem Abi] beginnt [_I Jens] [_L in Frankfurt] [_E seine Lehre].</p>

Verb Class Frame Variants: *Aspect* Verbs (3)

Frame	Participating Verbs & Corpus Examples
$n_I i_E$	<p>+ anfangen, aufhören, beginnen / \neg beenden, enden</p> <p>[_I Ich] habe angefangen, [_E Hemden zu schneiden].</p> <p>[_I Die Bahn] will [_T 1994] anfangen [_E zu bauen].</p> <p>... daß [_I der Alkoholiker] aufhört [_E zu trinken].</p> <p>... daß [_I die Säuglinge] einfach aufhören [_E zu atmen].</p> <p>[_I Tausende von Pinguinen] beginnen [_E dort zu brüten].</p>
$n_I p_E:mit_{Dat}$	<p>+ anfangen, aufhören, beginnen / \neg beenden, enden</p> <p>Erst als [_I der gesammelte Hofstaat] [_E mit Klatschen] anfang.</p> <p>Aber [_I wir] müssen endlich [_E damit] anfangen.</p> <p>... müßten [_I viel mehr Frauen] [_E mit der Arbeit] aufhören.</p> <p>Schließlich begann [_I er] [_E mit dem Selbstentzug].</p> <p>[_I Man] beginne [_E mit eher kartharsischen Werken].</p>

Clustering Methodology

1. Statistical acquisition of lexical verb information
2. Association of verbs with distributional frame vectors
3. Automatic verb clustering by standard technique *k-Means*
4. Clustering evaluation against manual verb classification
5. Clustering interpretation

Statistical Grammar Model

- Lexicalised probabilistic context-free grammar (Charniak, 1995; Carroll and Rooth, 1998)
- Unsupervised training by *EM-Algorithm* (Baum, 1972)
- Robust statistical parser *LoPar* (Schmid, 2000)
- 35 million words of German newspaper corpora
- Lexicalised statistical grammar parameters
- Corpus-based empirical lexical induction
 - Subcategorisation behaviour of verbs

Experiment Feature Choice

- D1** Purely syntactic definition of verb subcategorisation
- D2** Syntactico-semantic definition of subcategorisation with prepositional preferences
- D3** Syntactico-semantic definition of subcategorisation with prepositional and selectional preferences

Subcategorisation Frame Elements

n	noun phrase (case: nominative)
a	noun phrase (case: accusative)
d	noun phrase (case: dative)
r	reflexive pronoun
p	prepositional phrase
x	expletive <i>es</i>
i	non-finite clause
s-2	finite verb second clause
s-dass	finite <i>dass</i> -clause
s-ob	finite <i>ob</i> -clause
s-w	indirect <i>wh</i> -question
k	copula construction

Examples:

- na
- np
- npr
- nds-dass

Prepositional Phrase Types

- **Akk:** an, auf, bis, durch, für, gegen, in, ohne, um, unter, vgl, über
- **Dat:** ab, an, auf, aus, bei, in, mit, nach, seit, unter, von, vor, zu, zwischen, über
- **Gen:** wegen, während
- **Nom:** vgl

Examples: Akk.an, Dat.nach, Gen.wegen, Nom.vgl

GermaNet Top Level Nodes

- Lebewesen `creature`
- Sache `thing`
- Besitz `property`
- Substanz `substance`
- Nahrung `food`
- Mittel `means`
- Situation `situation`
- Zustand `state`
- Struktur `structure`
- Physis `physis`
- Zeit `time`
- Ort `space`
- Attribut `attribute`
- Kognitives Objekt `cognitive object`
- Kognitiver Prozess `cognitive process`

Subcategorisation Frame Distribution (D1)

glauben
`to think, to believe`

Frame Type	Freq	Prob	Bin
ns-dass	1,929	0.279	1
ns-2	1,888	0.274	1
np	687	0.100	1
n	608	0.088	1
na	555	0.080	1
ni	346	0.050	1
nd	234	0.034	1
nad	160	0.023	1
nds-2	70	0.010	1
nai	62	0.009	0

Subcategorisation Frame+PP Distribution (D2)

reden
`to talk`

Frame Type		Freq	Prob	Bin
np		1,121	0.455	1
np:Akk.über	`about`	480	0.153	1
np:Dat.von	`about`	463	0.148	1
np:Dat.mit	`with`	280	0.089	1
np:Dat.in	`in`	81	0.026	1
np:Nom.vgl	`as`	14	0.004	0
np:Dat.bei	`at` _{place}	13	0.004	0
np:Dat.über	`about`	13	0.004	0
np:Dat.an	`at` _{tense}	12	0.004	0
np:Akk.für	`for`	10	0.003	0
np:Dat.nach	`after`	8	0.003	0

Nominal Verb Arguments

beginnen

`to begin´

→ n

Noun		Freq
Uhr	`o'clock´	85
Prozess	`process´	77
Kampf	`fight´	70
Verhandlung	`negotiation´	66
Krieg	`war´	64
Tag	`day´	52
Zeit	`time´	52
Arbeit	`work´	47
Geschichte	`story´	46
Karriere	`career´	42
Spiel	`game´	38
Diskussion	`discussion´	31

GermaNet Synset Preferences

beginnen
 `to begin`
 → n

Synset		Freq
Situation	`situation`	1.102
Ereignis	`event`	1.060
Geschehen	`happenings`	971
Handlung, Tat, Aktivität, Tätigkeit	`action`	733
Objekt	`object`	436
Zustand	`state`	302
Kognitiver Zustand	`cognitive state`	295
Zeit	`time`	257
Beziehung, Verhältnis, Relation	`relation`	255
Ding, Sache, Gegenstand, Gebilde	`thing`	218
Artefakt, Werk	`artefact`	155
Zeiteinheit	`unit of time`	155

GermaNet Top Level Preferences

beginnen
`to begin`

→ n

Synset	Freq	Prob
Situation	1.102	0.425
Zustand	302	0.116
Zeit	257	0.099
Sache	222	0.086
Kognitives Objekt	148	0.057
Kognitiver Prozess	140	0.054
Ort	108	0.041
Attribut	101	0.039
Struktur	87	0.034
Lebewesen	81	0.031
Besitz	37	0.014
Physis	4	0.002
Substanz	4	0.001
Nahrung	3	0.001

$D1 \rightarrow D2 \rightarrow D3$ for *beginnen* 'to begin'

<i>D1</i>		<i>D2</i>		<i>D3</i>	
np	0.428	n	0.278	np:Akk.um	0.161
n	0.278	np:Akk.um	0.161	<u>n</u> (Situation)	0.118
ni	0.087	ni	0.087	ni	0.087
na	0.071	np:Dat.mit	0.082	np:Dat.mit	0.082
nd	0.036	na	0.071	np:Dat.an	0.056
nap	0.032	np:Dat.an	0.056	np:Dat.in	0.055
nad	0.019	np:Dat.in	0.055	<u>n</u> (Zustand)	0.032
nir	0.012	nd	0.036	<u>n</u> (Zeit)	0.027
ns-2	0.009	nad	0.019	<u>n</u> (Sache)	0.024
xp	0.005	np:Dat.nach	0.014	<u>na</u> (Situation)	0.023

$D1 \rightarrow D2 \rightarrow D3$ for *fahren* `to drive`

$D1$		$D2$		$D3$	
n	0.339	n	0.339	<u>n</u> (Sache)	0.118
np	0.285	na	0.193	<u>n</u> (Lebewesen)	0.095
na	0.193	np:Akk.in	0.054	<u>na</u> (Lebewesen)	0.082
nap	0.059	nad	0.042	<u>na</u> (Sache)	0.063
nad	0.042	np:Dat.zu	0.041	<u>n</u> (Ort)	0.057
nd	0.040	nd	0.040	np:Akk.in	0.054
ni	0.010	np:Dat.nach	0.039	<u>na</u> (Sache)	0.047
ns-2	0.008	np:Dat.mit	0.034	np:Dat.zu	0.041
ndp	0.008	np:Dat.in	0.032	np:Dat.nach	0.039
ns-w	0.004	np:Dat.auf	0.018	np:Dat.mit	0.034

k-Means Clustering Algorithm

- *k*-Means algorithm (Forgy, 1965)
- 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

k-Means Experiment Setup

- Number of clusters $k = 14$ (reduced) or $k = 43$ (full)
- Frame distribution:
 - frequencies / probabilities / binaries
 - original / strengthened / smoothed / noisy
- Initial cluster assignment:
 - Random clusters
 - Agglomerative hierarchical clusters
- Similarity measure

Cluster Analysis Examples on D3

C₁ **nieseln regnen schneien** - *Weather*

dämmern - *Weather* C₂

C₃ **kriechen rennen** - *Manner of Motion: Locomotion*
eilen - *Manner of Motion: Rush*
gleiten - *Manner of Motion: Flotation*
starren - *Facial Expression*

beginnen enden - *Aspect*
bestehen existieren - *Existence*
liegen sitzen stehen - *Position*
laufen - *Manner of Motion: Locomotion* C₅

C₄ **klettern wandern** - *Manner of Motion: Locomotion*
fahren fliegen segeln - *Manner of Motion: Vehicle*
fließen - *Manner of Motion: Flotation*

C₆ **festlegen** - *Constitution*
bilden - *Production*
erhöhen senken steigern vergrößern verkleinern - *Quantum Change*

C₇ **töten** - *Elimination*
unterrichten - *Teaching*

Comparison of Semantic Verb Classes

- Purpose of verb classification
- Verb class model
- Starting point: verbs vs. classes
- Common properties of verbs in classes
- Automatic acquisition

Semantic Fields (Schumacher, 1986)

- German second language teaching
- 1,000 German verbs and verbal expressions
- Syntactic usage and selectional descriptions
- Conceptual paraphrases
- Idiomatic expressions
- Morphological component: word formation

Top Semantic Fields

- Allgemeine Existenz `general existence`
- Spezielle Existenz `specific existence`
- Differenz `difference`
- Relation, geistiges Handeln `relation and mental act`
- Handlungsspielraum `scope of acting`
- Sprachlicher Ausdruck `language expression`
- Vitale Bedürfnisse `vital needs`

Verb Process Classification

- Ballmer and Brennenstuhl (1986)
- 13,000 common German verbs
- Basis: process models
- Initial situation, transition, end situation
- Preconditions, results, consequences
- Verb meaning is participation of verb in process
- Single argument construction as syntactic description
- Rough selectional preferences
- 44 models

Verb Process Model: Example

gezeugt werden	`to be conceived´
im Embryo-Stadium sein	`to be an embryo´
geboren werden	`to be born´
Kind sein	`to be child´
aufwachsen	`to grow up´
pubertieren	`to be pubescent´
sich einleben	`to settle in´
sich paaren	`to mate with someone´
Nachkommen haben	`to have descendants´
leben	`to live´
altern	`to age´
sterben	`to die´
tot sein	`to be dead´

Lebensmodell `life model´

aufwachsen	`to grow up´	jemand 1
werden	`to become´	jemand 1
ranken	`to etwine´	pflanzlich 1
wuchern	`to proliferate´	pflanzlich 1
sich entwickeln	`to develop´	jemand 1

SALSA for German FrameNet

- Fillmore's frame semantics
- Frame: conceptual structure describing a situation
- Frame element: local role in frame
- Range of combinatory valences for word senses
- Hierarchical classification structure
- Lexicographic database with conceptual structures, participating elements and annotated example sentences
- Database for English nouns, verbs and adjectives ↔ BNC
- Database for German nouns and verbs ↔ TIGER

SALSA Frame Examples

- *Motion*: bewegen, fahren, fliegen, gehen, gleiten, heraus-springen, huschen, kreisen, rollen, springen, ...
- *Motion_Directional*: fallen, kippen, kreisen, schlingern, ...
- *Motion_Noise*: krachen, quietschen, trampeln
- *Self-Motion*: anspringen, aufspringen, besteigen, bewegen, bummeln, eilen, eintauchen, fliegen, gehen, gleiten, hasten, hineinspringen, klettern, kreisen, ...
- *Transfer*: gehen
- *Transportation*: bewegen, fahren, fliegen, rollen, schleichen
- *Travel*: abreisen, fahren, fliegen, gehen, in Urlaub gehen, reisen, zurückziehen

German Semantic Verb Classes: Examples

Manner of Motion

Locomotion: gehen, klettern, kriechen, laufen, rennen, ...

Rotation: drehen, rotieren

Rush: eilen, hasten

Vehicle: fahren, fliegen, rudern, segeln

Flotation: fließen, gleiten, treiben

Relation Classification \leftrightarrow Human Beings

- How do we know whether a classification is reasonable?
- How do we define `reasonable`?
- Does the purpose determine the class model?
- Which are the essential aspects of the classes?
- Asking x human beings,
how many different classifications would they create?
- Which is the basis for the judgement of human beings?

Future Research

- Extension of verb classification
- Human judgement task on verb classes
- Classification technique on verb data
- Soft clustering technique on verb data
- Clustering of particle verbs
- NLP application for semantic verb classes