

Studies concerning the automatic classification of *an* particle verbs



Sylvia Springorum
sylvia.springorum@ims.uni-stuttgart.de
Institut für Maschinelle Sprachverarbeitung

Theory-based Gold Standard

The verb particle *an* has about 11 different readings (DRT-based theoretical analysis (Springorum, 2009))
4 readings serve as semantic classes for the experiments:

Topological reading:

Contact between direct object of the *an* particle verb and an implicit background

- (1) Maria *kettet* den Hund *an*.
Maria chains the dog.

Directional reading:

The verb event points from the subject to the direct object of the *an* particle verb

- (2) Der Junge *lächelt* die Mutter *an*.
The boy smiles at the mother.

Event Initiation reading:

an contributes a change from a non-progressive state to a progressive state

- (3) Der Schiedsrichter *pfeift* das Spiel *an*.
The referee starts the game by whistling.

Partitive reading:

The verb event is performed only on parts of the direct object

- (4) Der Wurm *frisst* den Apfel *an*.
The worm partially eats the apple.

Empirical Features

(i) Prepositional heads of subcategorized PPs

- (5) Maria *kettet* den Hund *an* dem Fahrradständer *an*.
'Maria chains the dog at the bicycle rack [an]'
Maria chains the dog at the bicycle rack.
- (6) Der Film *regt* die Zuschauer *zum* Denken *an*.
'the movie inspires the audience to think [an]'
The movie makes the audience think

(ii) Direct objects subcategorized by the particle verb:

Directional verbs: communication attempt frequently comes with persons as direct objects
Topological verbs: are likely to subcategorize physical objects because of their contact semantics

To reduce the data sparseness:

Semantic generalization of the nominal heads of the direct objects (hypernym relation of GermaNet v. 5.2)

(iii) Baseline: Verb subjects as classification features:

Expected to provide little support as many of our *an* particle verbs occur with agentive subjects across the classes.

Discussion

Gold standard is Topological

- Classified as Event Initiation:

ansiedeln (settle)

„Event“ objects: Film, history, action, ...

- (7) Ich würde den Film im Fantasygenre *ansiedeln*. (**Metaphor**)
I would place the film in the genre of fantasy

- Classified as Directional:

anmalen (paint), *anbinden* (attach), *anfassen* (touch)

„Higher life form“ as object

Gold Standard is Event Initiation

- Classified as Directional:

anstimmen (intone)

incorrect data: Freund (friend); Mönch (monk) and girl as object

anstiften (incite), *anspornen* (cheer on)

Object is „Higher life form“; Event is expressed with a zu-PP

- (8) Der Chef *spornt* seine Mitarbeiter zu Höchstleistungen *an*.
The boss incites his employees to work more efficiently.

- (9) Den Bruder zu Unfug *anstiften* (to incite the brother to rag)

Gold Standard is Directional

- Classified as Topological:

anstreben (strive)

incorrect data: PP with an

- Classified as Event initiation:

angucken (watch), *anvisieren* (aim for)

Object is Event: Spiel (game); Lehre (apprenticeship)

German particle verbs are a challenge to theoretical and computational linguistics.

Verb **particles** are highly ambiguous e.g. *ankleben* (glue on); *ansehen* (look at)

Particle verbs can also have many readings e. g. *anziehen* (attract, dress, activate, ...)

Goal: Classification of *an* particle verbs

Combination of **theoretical knowledge** with **empirical methods**:

How can automatic classification tasks profit from lexical semantic theory?

How can semantic theory profit from automatic classification tasks?

Judgment-based Gold Standard

Upper bound

8 persons classified the 40 verbs into 4 classes with 10 verbs each.
Given: One verb per class

Topological: *anketten* (chain) Agreement: $p_o = \frac{1}{N} * \sum_{i=1}^k n_{ii} = 0.79$
Event Initiation: *antreiben* (activate)
Directional: *anschreien* (scream at)
Partitive: *anknabbern* (nibble partially)

Experiments

Vector features:

Proportions based on corpus frequencies greater than 1

Exp 1: 25 PPs

Exp 2: 252 semantic classes of objects

Exp 3: Combined features

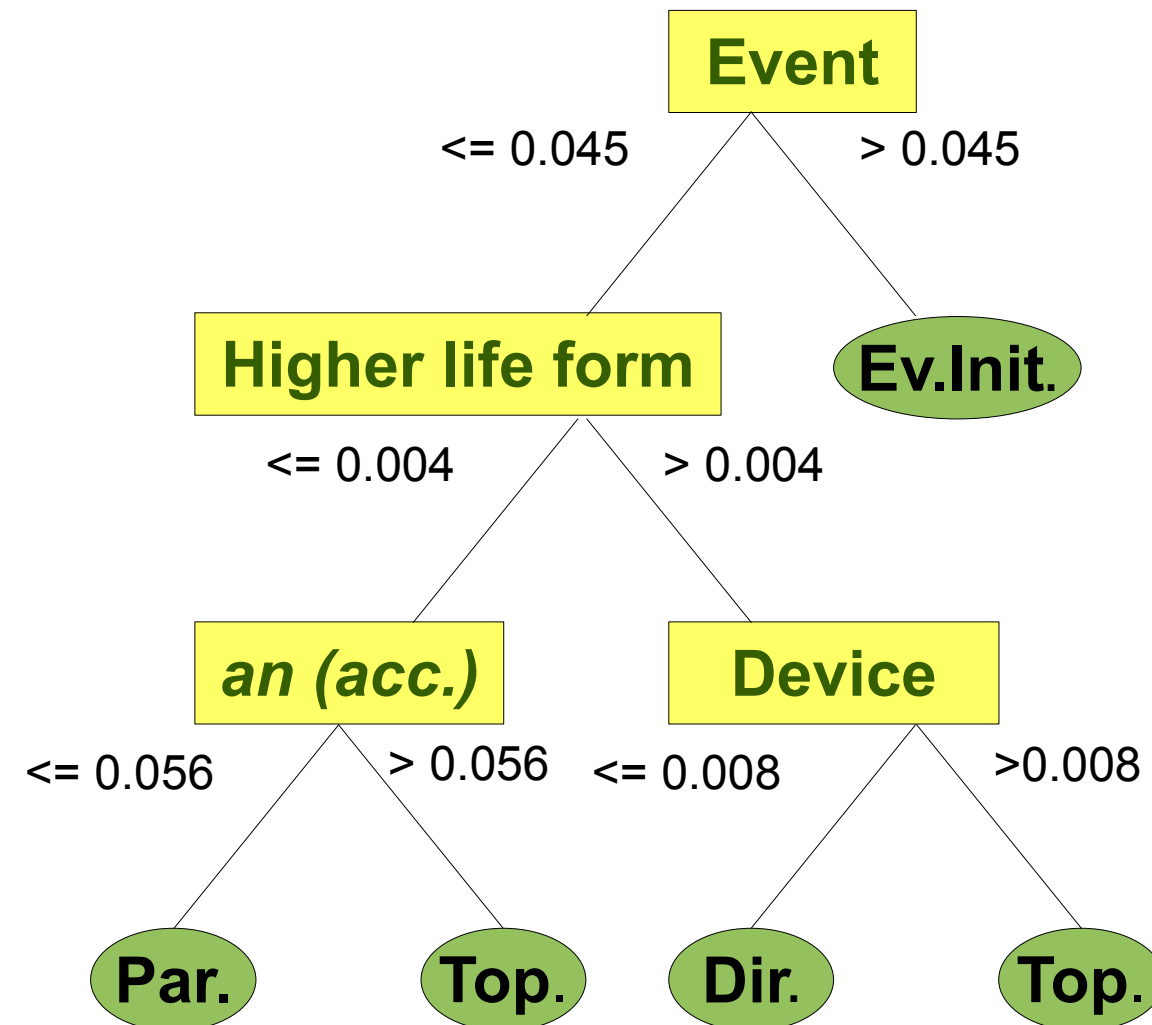
3.1: 25 PPs and 252 semantic object classes

3.2: Most successful PP with preposition *an(acc)* and 252 object classes

Results

Experiment	Features	+	%Top.	Ev. I.	Dir.	Par.
Baseline	Subject	13	32.50	0	3	1
Human judgement			79.06			
Exp. 1	PP	25	62.50	6	5	5
Exp. 2	Object Class	27	67.50	1	8	8
Exp. 3.1	PP + Obj. Class	27	67.50	5	5	7
Exp. 3.2	<i>an</i> + Obj. Class	28	70.00	4	7	7

Best experiment: PP with the preposition *an(acc)* and object classes



A	B	C	D	
anbauen anketten anlehnen anschließen	ansiedeln	anbinden anfassen anmalen	anschnallen anstreichen	A= Top.
	anheizen ankurbeln anpfeifen anregen anrichten antreiben anzetteln	anspornen anstiften anstimmen		B= Ev. I.
anstreben	angucken anvisieren	anblicken anlächeln anpeilen anreden anschreiben anschreien anstarren		C= Dir.
			all partitive verbs	D= Par.

Conclusion

How can automatic classification tasks profit from lexical semantic theory?

Different readings need different features

- an-* contributing a topological relation:
***an*-PPs** are reliable; Nominal indicators are subtle;
More GermaNet generalizations required: Object class: „Artifacts“ (instead for example “Device”)

- an-* contributing Event Initiation:
Nominal indicators are reliable; GermaNet class „Event“

- an-* contributing Direction
Nominal indicators are partially reliable; GermaNet class „Higher Life Form“ for communication attempt

- No feature yet for Partitive verbs

How can semantic theory profit from automatic classification tasks?

Cases in which the gold standard should be refined:

- Classified as Directional; gold standard is Event Initiation: *anspornen* (cheer on) and *anstiften* (incite)

Event initiation here: Communication to a person to make her act

Refinement: Reading with both event initiation and communication attempt meanings.

- Classified as Event Initiation ; gold standard is Directional:

Event-descriptions are Plan-descriptions:

Refinement: Directional sub reading with future plan meaning.

References

Gertrud Faaß, Ulrich Heid, and Helmut Schmid. 2001. Design and application of a standard for morphological analysis: Smor as an example of morphological . In Proceedings of LREC.

Mark Hall, Eibe Frank, Georey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The weka data mining software: An update. SIGKDD Explorations, 11(1).

Michael Schiehlen. 2003. A cascaded finite-state parser for German. In Proceedings of the 10th EACL.

Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In Proceedings of the 1st International Conference on New Methods in Language Processing.

Sylvia Springorum. 2009. Zur Semantik der Partikelverben mit an. Eine Studie zur Konstruktion ihrer Bedeutung im Rahmen der Diskursrepräsentationstheorie. Studienarbeit. Universität Stuttgart.

Sylvia Springorum. 2011a. DRT-based analysis of the German verb particle an. Leuvense Bijdragen 97. To appear.