# Computational Approaches to Semantic Relatedness

Antrittsvorlesung
Universität des Saarlandes

PD Dr. Sabine Schulte im Walde

Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart

30. November 2009

# Semantic Relatedness

- Words and larger linguistic units may be semantically related:

# Semantic Relatedness

- Words and larger linguistic units may be semantically related:
  - *toll–spitze–mies–verbesserungswürdig* (adj, great → bad)
  - *Amsel–Vogel* 'blackbird–bird'
  - *auftrinken–seinen Becher leeren* 'drink up'

# Semantic Relatedness

- Words and larger linguistic units may be semantically related:
  - *toll–spitze–mies–verbesserungswürdig* (adj, great $\longrightarrow$ bad)
  - *Amsel–Vogel* 'blackbird–bird'
  - *auftrinken–seinen Becher leeren* 'drink up'
- Lexical semantics is the study of how and what the words of a language denote $\rightsquigarrow$ word meaning and word relatedness

# Semantic Relatedness

- Words and larger linguistic units may be semantically related:
  - *toll–spitze–mies–verbesserungswürdig* (adj, great $\longrightarrow$ bad)
  - *Amsel–Vogel* 'blackbird–bird'
  - *auftrinken–seinen Becher leeren* 'drink up'
- Lexical semantics is the study of how and what the words of a language denote $\rightsquigarrow$ word meaning and word relatedness
- Words: single words, multi-words, word senses

# Semantic Relatedness

- Words and larger linguistic units may be semantically related:
  - *toll–spitze–mies–verbesserungswürdig* (adj, great $\longrightarrow$ bad)
  - *Amsel–Vogel* 'blackbird–bird'
  - *auftrinken–seinen Becher leeren* 'drink up'
- Lexical semantics is the study of how and what the words of a language denote $\rightsquigarrow$ word meaning and word relatedness
- Words: single words, multi-words, word senses
- Ambiguity: multiple word senses, e.g.,
  - *abnehmen* $\rightsquigarrow$ 'lose weight' (*Diät*) vs. 'support' (*übernehmen*) vs. 'take' (*nehmen*) vs. 'believe' (*glauben*)

# Semantic Relatedness

- Instances of semantic relatedness:

# Semantic Relatedness

- Instances of semantic relatedness:
  - underspecified semantic relatedness across a set of words:
    *toll–spitze–mies–verbesserungswürdig*
    *kaufen–verkaufen–kosten–bezahlen*

# Semantic Relatedness

- Instances of semantic relatedness:
  - underspecified semantic relatedness across a set of words:
    *toll–spitze–mies–verbesserungswürdig*
    *kaufen–verkaufen–kosten–bezahlen*
  - explicit semantic relations between words:
    *Amsel–Vogel*; *joggen–rennen*; *zuschließen–schließen*

# Semantic Relatedness

- Instances of semantic relatedness:
  - underspecified semantic relatedness across a set of words:
    *toll–spitze–mies–verbesserungswürdig*
    *kaufen–verkaufen–kosten–bezahlen*
  - explicit semantic relations between words:
    *Amsel–Vogel*; *joggen–rennen*; *zuschließen–schließen*
  - explicit semantic relations between the parts of multi-words:
    *Küchenmesser, Brotmesser*

# Semantic Relatedness

- Instances of semantic relatedness:
  - underspecified semantic relatedness across a set of words:
    *toll–spitze–mies–verbesserungswürdig*
    *kaufen–verkaufen–kosten–bezahlen*
  - explicit semantic relations between words:
    *Amsel–Vogel*; *joggen–rennen*; *zuschließen–schließen*
  - explicit semantic relations between the parts of multi-words:
    *Küchenmesser, Brotmesser*
  - degree of semantic relatedness between multi-words and their parts:
    *zuschließen–zu+schließen* vs. *anfangen–an+fangen*
    *Brotmesser–Brot+Messer* vs. *Klatschmohn–Klatsch+Mohn*

# Semantic Relatedness

- Why is this interesting?

  $\rightsquigarrow$ theoretical investigations (linguistic and cognitive):

  how do humans perceive and express semantic relatedness?

  $\rightsquigarrow$ semantic relatedness in computational tasks and applications:

  lexicography (thesaurus construction; compositionality);
  question answering, e.g., *Which company produces wooden doors?*

- How difficult is the task?

  $\rightsquigarrow$ human judgements, cf.

  *belästigen–nerven*; *abhalten–anhalten*; *einsetzen–setzen*;
  *sentence–decision*

# Synonymy

- Synonymy: words with identical or similar meaning
- Absolute synonymy occurs rarely
- Near-synonymy is pervasive
- Shades of meaning differences between two near-synonyms, e.g.,
  - emphasis: *Kleinkrieg–Fehde*; *enemy–foe*
  - style: *hacke–betrunken*; *pissed–drunk*
  - subcategorisation: *helfen–unterstützen*; *give–donate*
- Paraphrases: alternative ways to convey the same meaning, including and going beyond the word/phrase level, e.g.,
  *Das Glas ist halb voll.* ↔ *Das Glas ist halb leer.*
  *The glass is half full.* ↔ *The glass is half empty.*

# Antonymy

- Antonymy: words with opposite meaning
  - complementary/contradictory: negation of one predication entails its contradiction
    *succeed–fail/gelingen–misslingen; true–false/richtig–falsch*
  - reversive: positions or motion in opposite dimensions
    *appear–disappear/auftauchen–verschwinden;*
    *forwards–backwards/vorwärts–rückwärts*
  - contrary: assertion of one predicate entails denial of its contrary, but both contraries may be false
    *red–green/rot–grün; love–hate/lieben–hassen*
  - conversive: same situation seen from different perspectives
    *buy–sell/kaufen–verkaufen; teacher–pupil/Lehrer–Schüler*

# Hypernymy & Troponymy

- Hypernymy/Hyponymy: words with super-/sub-ordinated meaning
- Hyponym *is a kind of a* hypernym, e.g.
  *robin–bird; coffee–beverage; beetle–car*
  *Amsel–Vogel; Kaffee–Getränk; Golf–Auto*
- Co-Hyponymy: words with the same hypernym, e.g.
  *cat–dog; coffee–tea; Katze–Hund; Kaffee–Tee*
- Troponymy is hypernymy for verbs:
  *to* $v_{troponym}$ *is to* $v_{superordinate}$ *in some particular manner*
  *jog–run; cook–create; joggen–rennen; kochen–kreieren*
- Recursive application of hypernymy generates a hierarchy

# Scripts and Frames

- Script: sequence of events in a particular context, such as everyday situations (Schank & Abelson)
- Specialisation of a frame (Minsky)
- Related to frame semantics (Fillmore): frames as the background and situational knowledge needed for understanding a word or expression
- Relating participants in a situation to each other
- Abstracting over semantic relations such as synonymy, antonymy, causality, etc.

# Scripts and Frames

- Example of a script: *restaurant*
  - Scene 1: Entering: guest goes into restaurant, guest looks at tables, guest thinks where to sit, guest goes to table
  - Scene 2: Ordering: guest gets menu, guest chooses food, guest gives signal to waiter, waiter comes to table, guest orders food
  - Scene 3: Eating: waiter brings food, guest eats food
  - Scene 4: Exiting: waiter brings check, waiter gives check to guest, guest gives money to waiter, guest leaves restaurant

- Example of a frame: *cooking creation*

  *A cook creates a produced food from (raw) ingredients. The heating instrument and/or the container may also be specified.*

  lexical units: *bake, baker, concoct, cook up, cook, make, prepare, put together, whip up*, etc.

# Multi-Word Expressions

- Multi-word expression: any phrase that *may not be* entirely predictable on the basis of standard rules and lexica
- Compound nouns: head noun and modifier noun
  - examples: *Küchenmesser, Taschenmesser, Brotmesser, Gradmesser*
  - semantic relations between compound parts (e.g., purpose)
  - degree of compositionality with respect to head and modifier
- Particle verbs: particle and base verb
  - word class of particle: preposition, adverb, noun, adjective, etc.
  - preposition-based examples: *abholen, einsetzen, anfangen*
  - degree of compositionality with respect to particle and base verb

# Compositionality

- Degree of compositionality: to which extent is the meaning of the multi-word related to the meanings of its parts?
- Semantic relatedness as index for degree of compositionality
- Distributional identification:
  comparison of properties of parts and whole

# Summary: Semantic Relatedness

- Paradigmatic semantic relations between word senses: synonymy, antonymy, hypernymy
- Paradigmatic vs. syntagmatic: vertical, set-based substitutability wrt. elements of the same class (*Katze/Hund/Tier; cat/dog/animal*) vs. horizontal, syntactic sequence of elements (*Eis essen; aus und vorbei; eat ice-cream*)
- Situation-based semantic relatedness across word classes, e.g., *backen, Küche, Koch, Herd; bake, kitchen, cook, oven*, etc.
- Words vs. word senses vs. phrases: *abnehmen–zunehmen/halten/glauben; Maus–Computer/Katze; ins Gras beißen–sterben; auftrinken–seinen Becher leeren*
- Semantic relatedness of multi-word expressions: between and to parts, e.g., *Küchenmesser vs. Klatschmohn; abnehmen vs. anfangen*

# Resources and Approaches: Overview

- How do we know (automatically) the semantic relation between two word senses?
- Lexical acquisition: (automatic) definition of linguistic information on lexical items
- Sources for lexical acquisition:
  - (hand-coded) language resources, such as dictionaries, thesauri, taxonomies, ontologies, etc.
  - (annotated) corpus data

# Corpus-based Approaches: Overview

- Distributional semantics via corpus co-occurrence
- Distributional models:
  - Vector space models
  - Syntagmatic pattern models
  - Graph models
  - Semantic classification

# Distributional Semantics

- Contexts of a linguistic unit tell us something about the meaning of the linguistic unit
- Example: corpus can tell us that one can *buy, peal, and eat an apple*
- Distributional hypothesis:

  *You shall know a word by the company it keeps.* (Firth, 1957)

  *Each language can be described in terms of a distributional structure, i.e., in terms of the occurrence of parts relative to other parts.* (Harris, 1968)
- Problem: lack of commonsense knowledge
  inferential (i.e., how to use language distributionally) vs. referential (incorporating world knowledge) abilities (Marconi, 1997), e.g.,
  *Ananas–gelb; auftauen–Wasser; ananas–yellow; defrost–water*

# Co-Occurrence

- Context refers to corpus co-occurrence
- Paradigmatic semantic relatedness: words appear in similar contexts ($\rightarrow$ similar corpus co-occurrence)
  *Daddy pealed the {apple, pear, banana, fruit} for Jan.*
- Syntagmatic semantic relatedness: words appear together ($\rightarrow$ joint corpus co-occurrence)
  *He explained patiently until she understood.*
- The two do not exclude each other
  *He likes all fruit but oranges.*
- Procedure: identify corpus co-occurrence features and feed into computational model to induce semantic relatedness

# Examples: Co-Occurrence Features

Co-occurrence matrix for vector space model; $sim(w_i, w_j) = f(\vec{v_{w_i}}, \vec{v_{w_j}})$:

| | collect | imported | peal | read | ripe | rotten | salad |
|---|---|---|---|---|---|---|---|
| apple | 91 | 2 | 358 | 0 | 97 | 201 | 96 |
| banana | 3 | 135 | 221 | 5 | 133 | 89 | 54 |
| book | 188 | 74 | 2 | 415 | 0 | 44 | 1 |
| fruit | 1 | 62 | 43 | 0 | 54 | 199 | 286 |
| leaf | 23 | 0 | 0 | 11 | 21 | 157 | 1 |
| orange | 0 | 111 | 299 | 2 | 111 | 32 | 25 |
| pear | 32 | 18 | 37 | 1 | 143 | 34 | 65 |

Syntagmatic pattern for hypernymy (Hearst, 1992):
$Noun_i$ *such as* $Noun_1, Noun_2, ..., Noun_n$, e.g.,
*fruit such as apples, bananas, pears, oranges*

# Distributional Models

Choice and aspects of distributional models:

- Selection of features (bag-of-words, syntax-based, human-based)
- Selection of learning algorithm
- Problem: ambiguity

Example approaches:

1. Semantic classification of German verbs
2. Distributional model of German particle verbs
3. Cognitive aspects

# Semantic Classification

- Groupings of words according to semantic properties
- Classes refer to general semantic level;
  idiosyncratic lexical semantic properties are underspecified;
  semantic relations are underspecified
- Intuitive examples:
  *motion with a vehicle: drive, fly, row, etc.*
  *animal: bird, robin, penguin, dog, dachshund, etc.*
  *temperature: hot, cold, tepid, etc.*

# Automatic Classification

- Objects → Classes
- Objects in common classes: as similar as possible
- Objects in different classes: as dissimilar as possible
- Difficulty: selection of classification algorithm and parameter setting (incl. features)

# Overview: Verb Classification

- Hypothesis: verb behaviour $\leftrightarrow$ verb meaning aspects
- Distributional verb descriptions:
  syntactic frames, prepositional phrases, selectional preferences
- Clustering with k-Means algorithm
- Result: semantic verb classes

# Subcategorisation Frame Distribution

*glauben*
*'think, believe'*

| Frame Type | Freq |
|---|---|
| $NP_{nom}$-$S_{dass}$ | 1,929 |
| $NP_{nom}$-$S_2$ | 1,888 |
| $NP_{nom}$-$PP$ | 687 |
| $NP_{nom}$ | 608 |
| $NP_{nom}$-$NP_{acc}$ | 555 |
| $NP_{nom}$-$INF$ | 346 |
| $NP_{nom}$-$NP_{dat}$ | 234 |
| $NP_{nom}$-$NP_{acc}$-$NP_{dat}$ | 160 |
| $NP_{nom}$-$NP_{dat}$-$S_2$ | 70 |
| $NP_{nom}$-$NP_{acc}$-$INF$ | 62 |

# Nominal Preference Distributions

*reden über$_{Acc}$*
*'talk about'*

| Noun | | Freq |
|---|---|---|
| Geld | 'money' | 19 |
| Politik | 'politics' | 14 |
| Problem | 'problem' | 13 |
| Thema | 'topic' | 10 |
| Inhalt | 'content' | 9 |
| Koalition | 'coalition' | 6 |
| Ding | 'thing' | 5 |
| Freiheit | 'freedom' | 5 |
| Kunst | 'art' | 5 |
| Film | 'movie' | 5 |

# k-Means Algorithm

- Unsupervised hard clustering
- $n$ objects $\longrightarrow$ $k$ 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;
     go to step 2
- Parameters: number of clusters, similarity measures

# Clustering Example: Random Input

- 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

# Clustering Example: Output

- 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

# Summary: Semantic Classification

- Considerable agreement between clustering results and manual classification → successful linguistic and technical parameters
- Difficult compromise between general and idiosyncratic properties
- Ambiguity not modelled implicitely
- Underspecified semantic relations
- Classes cover synonymy, antonymy, hypernymy, etc.
- Relation of classes to situations?
  *ermorden* *'assassinate'*, *erschießen* *'shoot'*, *töten* *'kill'*, *festnehmen* *'arrest'*, *verhaften* *'arrest'*, *befragen* *'interrogate'*, *entlassen* *'release'*
- Semantic classes per se vs. with respect to task or application

# German Particle Verbs (PVs)

- Focus on preposition particles: *ab, an, auf, aus, bei, durch, ein, los, nach, über, um, unter, vor, wider, zu*
- Examples: *abholen, anfangen, einführen*
- Syntax:
  - PVs may change behaviour of base verbs
  - changes are quite regular (Stiebels, 1996; Aldinger, 2004)
- Semantics: transparent vs. opaque PV senses
  - *abholen* 'fetch' ↔ *holen* 'fetch'
  - *anfangen* 'begin' ↔ *fangen* 'catch'
  - *einsetzen* 'insert, begin' ↔ *setzen* 'put/sit (down)'

# Example: Syntactic Change (Addition)

- Sie lächelt.
  'She smiles.'
- * Sie lächelt [$NP_{acc}$ ihre Mutter].
  'She smiles her mother.'
- Sie lächelt [$NP_{acc}$ ihre Mutter] an.
  'She smiles at her mother.'

# Distributional Model of German Particle Verbs

- Problem: change of behaviour at the syntax-semantics interface
  (a) *einsetzen* ↔ *anfangen, beginnen* (opaque)
  (b) *einsetzen* ↔ *setzen* (transparent)
- Goal: predict meaning and compositionality of particle verbs
- Strategy 0: avoid frame-incorporating distributional features
- Strategy 1: syntax-semantic change of arguments is surprisingly regular, even for opaque particle verbs → model and exploit syntax-semantic transfer patterns
- Strategy 2: combine distributional models

# Strategy 0

- Model verbs (PV and BV) by nominal features, including and excluding syntax
- Predict meaning by determining nearest neighbours of PVs, based on distributional measures: $sim(w_i, w_j) = f(\vec{v_{w_i}}, \vec{v_{w_j}})$
- Relevant information in the distributions are nouns; references to argument structure (functions) are minor but important
- Window features are worse than nominal argument features
- Types of semantic relations (according to gold standard variation):
  GermaNet: 70% hypernymes, 22% synonyms, 1% antonyms;
  synonym/antonym dictionary: 43% synonyms, 48% antonyms;
  associations: variable picture, e.g., backward presupposition (*abstürzen/fliegen* 'crash'/'fly'), cause (*einbrocken/auslöffeln* 'get into/out of trouble'), script (*einschenken/trinken* 'pour'/'drink')

# Strategy 1

- Model syntax-semantic transfer patterns
- Predict degree of compositionality
- Gold standard: human judgements on compositionality, e.g., *umbringen*: 1.625; *abbestellen*: 6.750; *nachdrucken*: 9.250
- Experiment 1 (baseline): divergence of PV subcategorisation from average particle-based frame distribution → fails, because transfer applies to opaque as well as transparent PVs, plus ambiguity
- Experiment 2: overlap of nominal fillers in frame-argument combinations, e.g.

  *hängen*/NPnom-PP ↔ *aufhängen*/NPnom-NPacc-PP: *Bild, Plakat, Mörder*

  → fails, for data sparseness, frequency and ambiguity

# Summary: Distributional Model of PVs

- Plain distributional model (nearest neighbours) in comparison to abstraction by classification
- German particle verbs are a challenge to distributional models
- Selectional preferences are key to induce PV meaning
- Problem: pervasive ambiguity of particles and base verbs
- Side effect of gold standard definitions: variety of semantic relations among nearest neighbours

# Human Data on Semantic Relatedness

## Semantic memory and computational semantic relatedness

1. Human judgements on semantic relatedness/relations;
   ⇝ gold standard for evaluation purposes, e.g., compositionality
   ⇝ measure of difficulty of the task, i.e., wrt ambiguity
   - degree of compositionality of *einsetzen* (1-10)?
   - *belästigen–nerven* : synonymy? hypernymy?

2. Human asssociations reflect meaning components of words; psycholinguistic research on semantic memory
   ⇝ improved choice of salient distributional features?
   ⇝ gold standard of relatedness for evaluation or as model seeds
   - *herausfinden* ⇝ *erkennen* (6), *entdecken* (6), *forschen* (5), etc.
   - *Pfeife* ⇝ *Tabak* (17), *Rauch* (14), *Opa* (9), *Zigarette* (2), etc.

# Human Associations and Feature Choice

Associations reflect linguistic and conceptual features and therefore model verb meaning aspects. $\rightsquigarrow$ Assumption: If we can model associations by distributional features, we can build salient models of word meaning and word relatedness.

| Class | Features |
|---|---|
| *bedauern* 'regret', *heulen* 'cry', *jammern* 'moan', *klagen* 'complain, moan, sue', *verzweifeln* 'become desperate', *weinen* 'cry' | *Trauer* 'mourning', *weinen* 'cry', *traurig* 'sad', *Tränen* 'tears', *jammern* 'moan', *Angst* 'fear', *Mitleid* 'pity', *Schmerz* 'pain', etc. |
| *abnehmen* 'lose weight', *abspecken* 'lose weight', *zunehmen* 'gain weight' | *Diät* 'diet', *Gewicht* 'weight', *dick* 'fat', *abnehmen* 'lose weight', *Waage* 'scale', *Essen* 'food', *essen* 'eat', *Sport* 'sports', *dünn* 'thin', *Fett* 'fat', etc. |

# Summary: Cognitive Aspects

- Human associations provide data on semantic relatedness:
  a priori: underspecified; annotated: specified
- Exploit associations as feature indicators, model seeds, for evaluation
- Human judgements provide data on semantic relatedness
- Exploit judgements as compositionality rankings and relation rankings for evaluation

# Overall Summary

- Paradigmatic vs. situation semantic relations
- Distributional semantics:
  - Vector space models and syntagmatic patterns
  - Vector similarity (nearest neighbours) and generalisation (classification)
- Distributional co-occurrence features to model semantic relatedness
- Studies: semantic classification and particle verb compositionality
  - semantic classification: successful but underspecified relations
  - particle verb compositionality: first insights; need to model syn/sem transfer or highly ambiguous semantic classes of particles and bases
- Human associations provide keys to distributional knowledge and point to its commonsense limits

# Future Work

- Distinguish and combine distributional feature groups:
  - direct co-occurrence (words, syntax-based)
  - abstraction over co-occurrence (frame types, selectional preferences)
  - higher-order co-occurrence
  - syntagmatic patterns

$\rightsquigarrow$ Model underspecified semantic relatedness across a set of words to distinguish synonymy, antonymy and situational relations

*kaufen–erwerben* vs. *kaufen–verkaufen* vs. *kaufen–verkaufen–kosten–bezahlen*

$\rightsquigarrow$ Model explicit semantic relations between words

*Amsel–Vogel*; *joggen–rennen*

# Future Work

- Model the degree of semantic relatedness between multi-words and their parts:
  - compositionality of German particle verbs:
    - adjust features to reduce syntactic information
    - supervised approach on particle-sense-annotated data to learn transfer patterns from annotated data and to predict compositionality
    - combination of distributional descriptions or unsupervised classifications of particles and base verbs
  - compositionality of German compound nouns: combine evidence from part-whole pairings:
    - variability of compound parts by related nouns (e.g., co-hyponyms), cf. *Handwerk* vs. *Fußwerk*, and *Kunsthalle* vs. *Kunstgebäude*
    - featural similarity of compound parts, cf. *Kunst* and *Werk* appear in art domain
    - featural similarity of compound and head noun, e.g., *Kunstwerk* and *Werk*

# Future Work

- Interaction between human data, feature selection, and classification approaches
- Selection and combination of distributional co-occurrence models
- Use (annotated) human data as seeds and gold standard
- Computational approaches: nearest neighbours, semantic classification, ensemble classification, clustering with constraints
- Vary sources: web corpora and encyclopedias (i.e., wikis)
- Role of frequency and ambiguity