Limited compositionality phenomena in SALSA

Katrin Erk

Linguistics Department

University of Texas at Austin

joint work with: Aljoscha Burchardt, Anette Frank, Andrea Kowalski, Sebastian Pado, Manfred Pinkal

---

The SALSA project

- ☐ Manual annotation of
  - ■ 1.5 million words of syntactically annotated German newspaper text (TIGER Corpus)
  - ■ with lexical semantic information in the FrameNet paradigm: word senses (frames) and semantic roles (frame elements)
- ☐ Automatic word sense and semantic role analysis (Shalmaneser, Erk&Pado 06), and its use

- ☐ Researchers: Aljoscha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Pado, Marco Pennachiotti, Manfred Pinkal, Caroline Sporleder
- ☐ Funded by the Leibniz Program of DFG and DFG Normalverfahren (German Science Foundation)

Talk outline

- ☐ SALSA: FrameNet-annotation of German newspaper data
- ☐ *Limited compositionality phenomena* in SALSA: metaphor, idioms, support constructions
- ☐ Borderline cases
- ☐ Borderline cases in word sense annotation

The SALSA data

- ☐ Recent first release: http://www.coli.uni-saarland.de/projects/salsa/corpus/
- ☐ About 20.000 verb occurrences with their semantic roles
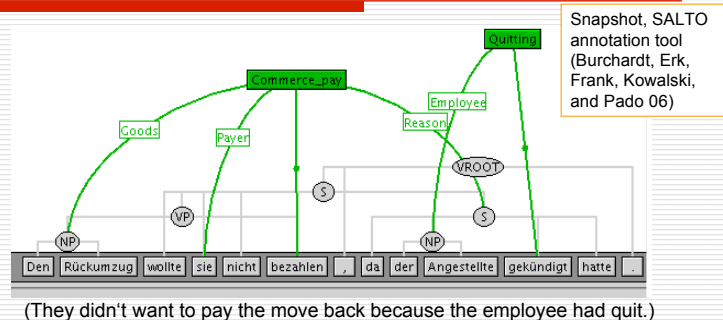- ☐ (Burchardt, Erk, Frank, Kowalski, Pado and Pinkal 06)

Frame Semantics and FrameNet

- Frame Semantics (Fillmore 1976, 1977, ..)
  - **Frame**: a conceptual structure or prototypical situation
  - **Frame elements**: participants of the situation
  - **Request**: A **Speaker** asks an **Addressee** for something, or to carry out some action. The content of the Request is the **Message**, ...

    [SPEAKER *SPD*] [FEE *asks*] [ADDRESSEE *coalition*] [MESSAGE *to talk about reforms*]
  - Frame evoking elements (targets): verbs, nouns, adjectives
- Berkeley FrameNet Project (Baker, Fillmore & Lowe 98)
  - Database of frames for a core lexicon of English
    - Frame definitions, lexical entries, corpus annotations
  - Current release: > 800 frames, around 8000 word senses

Basis of SALSA annotation:
the TIGER Corpus

- 1.5 million words / 80 K sentences of German newspaper text (Frankfurter Rundschau)
- Stuttgart/Potsdam/Saarbrücken, first release in 2003
- Surface-oriented, relatively theory-neutral annotation, grammatical functions
- (Brants, Dipper, Hansen, Lezius and Smith 02)

The SALSA Annotation Scheme

Snapshot, SALTO annotation tool (Burchardt, Erk, Frank, Kowalski, and Pado 06)

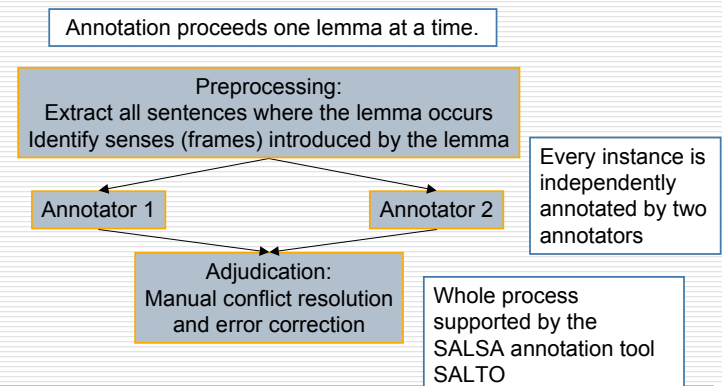(They didn't want to pay the move back because the employee had quit.)

Semantics:
- Word sense (frames) and semantic roles
- Independent frames
- Trees of depth one
- One edge points to target, others to frame elements

TIGER Syntax:
- Node labels: constituents
- Edge labels: gramm. functions
- Crossing edges
- POS

The annotation process

Annotation proceeds one lemma at a time.

Preprocessing:
Extract all sentences where the lemma occurs
Identify senses (frames) introduced by the lemma

Annotator 1   Annotator 2

Every instance is independently annotated by two annotators

Adjudication:
Manual conflict resolution and error correction

Whole process supported by the SALSA annotation tool SALTO

Some numbers

	Agreement (Frames)	Agreement (Roles)
Inter-Annotator	84.9%	85.7%
Inter-Adjudicator	97.0%	96.2%

Percentage agreement, no kappa:
Not all items get markup,
annotators may differ in extent of an annotated role

Word sense and semantic role annotation

☐ Predicate word sense and semantic role assignment interact
- ■ The train passed the station.
- ■ Can you pass the butter?

☐ How to deal with this?
- ■ Modularize annotation task as far as possible: Pred. sense distinctions only as far as necessary for role assignment
- ■ Integrated study of semantic frames with their participants=roles: This is the FrameNet choice.

Word sense and semantic role annotation

☐ "FrameNet and SALSA follow Fillmore (1985), which states that *Frame Semantics does not seek to draw an a priori distinction between semantics proper and (an idealized notion of) text understanding*" and consequently try to annotate what is actually understood. This makes the task more complex but should finally yield a more informative annotation." (Ellsworth et al, 2004)

☐ Consequence for metaphor, idioms and support constructions: Have to be tagged in the course of word sense and semantic role annotation

Talk outline

☐ SALSA: FrameNet-annotation of German newspaper data

☐ Limited compositionality phenomena in SALSA

☐ Borderline cases

☐ Borderline cases in word sense annotation

Annotation of limited compositionality phenomena

- ☐ Metaphor
  - ■ Just now work seemed the best antidote to the frustration that was <u>boiling</u> inside her
- ☐ Idioms
  - ■ Kerry <u>gave</u> them <u>the slip</u> and was found Sunday night in Kota Bahru.
- ☐ Support constructions
  - ■ Someone <u>made a statement</u> about my need two kerrect my shpelling.
- ☐ Meaning extension
  - ■ Die Temperaturfühler <u>sitzen</u> in der Thermoskanne
    The temperature sensors <u>sit</u> in the vacuum flask

Limited compositionality: Some figures

|	Sample of 246 Lemmas		Sub-corpus nehmen (to take)	
	Number	%	Number	%
Standard readings	**4638**	**85,7%**	**42**	**17,4%**
Metaphor	369	6,8%	38	15,8%
Support	326	6,0%	132	54,8%
Idiom	79	1,5%	29	12,0%
Non-literal use	**774**	**14,3%**	**199**	**82,6%**
Total	5412	100,0%	241	100,0%

(Burchardt, Erk, Frank, Kowalski, Pado and Pinkal 06)

Annotation criteria: support

- ☐ Prototypical properties of support constructions:
  - ■ Predicative/deverbal noun plus (support) verb
  - ■ Noun has its "usual" meaning
  - ■ Noun is the semantic predicate
  - ■ Semantic arguments of the noun are syntactic arguments of the verb
  - ■ Verb has no or little meaning contribution (may be causation, aspect …)
  - ■ Sometimes, the support verb construction can be paraphrased by a verb
    - ☐ Kritik üben ~ kritisieren (exercise criticism ~ criticize)

Annotation: Support

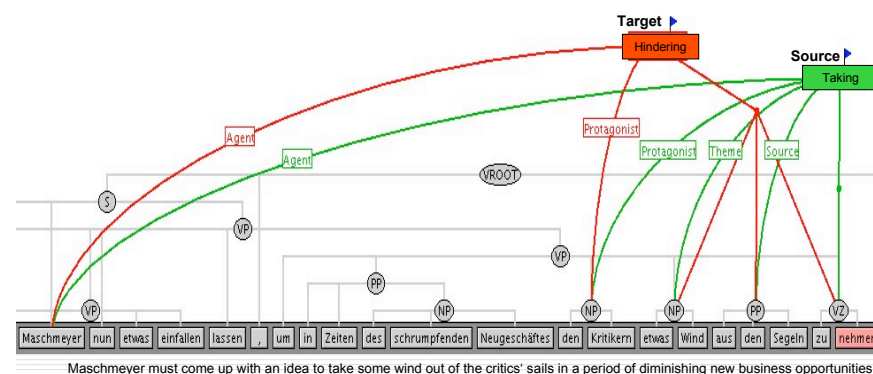## Pseudo-frame "Support", Pseudo-role "Supported"
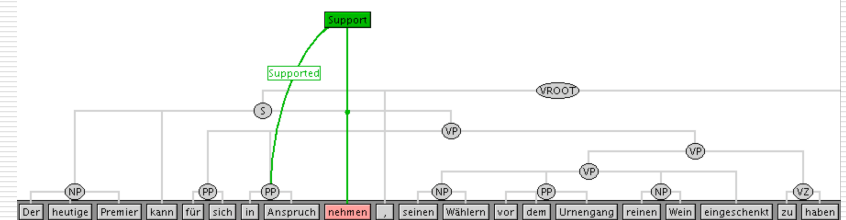
Today's prime minister can for himself in claim   take   his voters before the walk to the urn pure wine poured to have
Today's prime minister can lay claim to having poured his voters pure wine (told them the truth) before the vote

Annotation criteria: Metaphor

- ☐ Non-literal ("figurative") meaning
- ☐ Semi-compositional:
  - ▪ recoverable literal meaning
  - ▪ mapping from literal to non-literal meaning
- ☐ Recoverability of literal meaning sometimes subjective

Annotation: metaphor

Annotation of both source and target frame

Maschmeyer must come up with an idea to take some wind out of the critics' sails in a period of diminishing new business opportunities
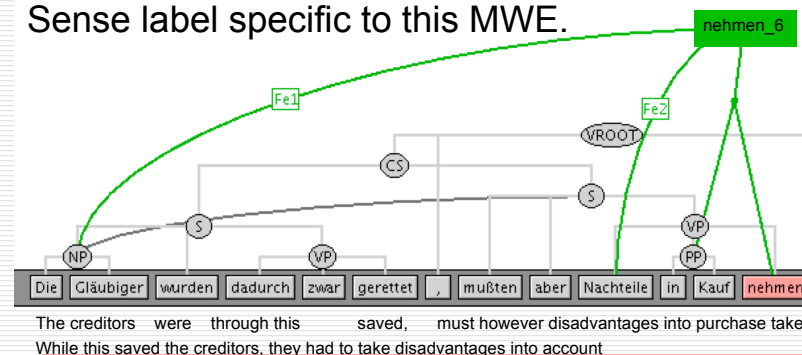
Annotation criteria: Idioms

- ☐ Meaning introduced by the whole construction (modulo variability)
- ☐ Sometimes, idiosyncratic syntactic construction (by and large)
- ☐ Non-compositional: literal meaning not recoverable
  - ▪ Again, this is sometimes subjective

Annotation: Idioms

Multiword expression introduces frame.

Sense label specific to this MWE.

The creditors were through this saved, must however disadvantages into purchase take

While this saved the creditors, they had to take disadvantages into account

Contrasting annotation of metaphors and idioms

- ☐ Metaphors:
  - ■ Literal (source) frame evoked by the verb
  - ■ Nonliteral (target) frame evoked by multiword expression
  - ■ Roles of literal reading may form part of target in nonliteral reading
  - ■ [Danny's corner]$_{Agent}$ refused to throw$_{CAUSE\_MOTION}$ [in]$_{Goal}$ [the towel]$_{Theme}$.
  - ■ [Danny's corner]$_{Capitulator}$ refused to throw in the towel$_{GIVING\_IN}$.
- ☐ Idioms:
  - ■ Frame evoked by multiword expression
  - ■ [Kerry]$_{Evader}$ gave [them]$_{Pursuer}$ the slip$_{EVADING}$ and was found Sunday night in Kota Bahru.

Meaning Extensions

- ☐ Die Temperaturfühler sitzen in der Thermoskanne
  The temperature sensors sit in the vacuum flask

- ☐ Generalized, more abstract sense
- ☐ Loss of inferences
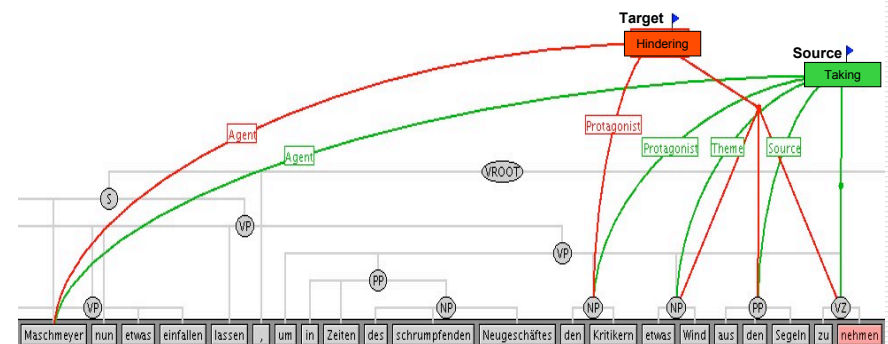- ☐ Violation of selectional constraints

Annotation: Meaning Extensions

(Definition and) annotation of frames with more abstract meaning: Being_located vs. Posture
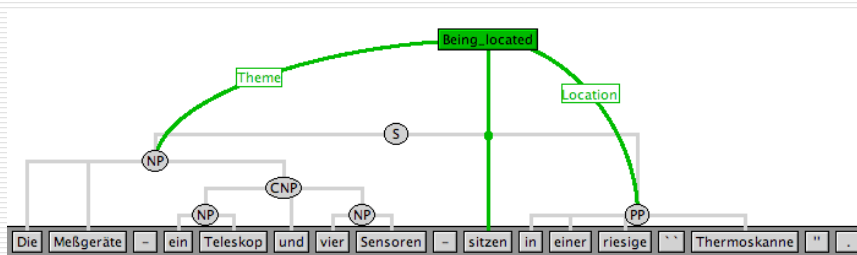
The instruments - a telescope and four sensors - sit inside a giant "thermos flask"

The relation between source and target annotation for metaphor

Maschmeyer must come up with an idea to take some wind out of the critics' sails in a period of diminishing new business opportunities
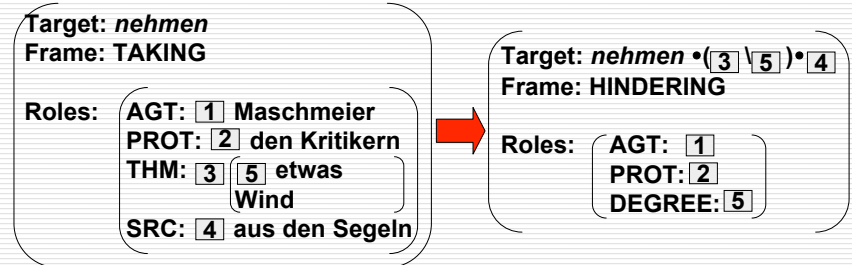
„take the wind out of someone's sails"

FRAME:	TAKING
TGT:	nimmt
AGENT:	Maschmeyer [1]
PROTAGONIST:	den Kritikern [2]
THEME:	den Wind [3]
SOURCE:	aus den Segeln [4]

FRAME:	HINDERING
TGT:	nimmt [3] [4]
AGENT:	[1]
PROTAGONIST:	[2]
ACTION:	-
HINDRANCE:	-

„take some wind out of someone's sails"

... nimmt Maschmeier den Kritikern etwas Wind aus den Segeln

Target: *nehmen*
Frame: TAKING

Roles:
AGT: [1] Maschmeier
PROT: [2] den Kritikern
THM: [3] [5] etwas Wind
SRC: [4] aus den Segeln

→

Target: *nehmen* •([3] \[5])• [4]
Frame: HINDERING

Roles:
AGT: [1]
PROT: [2]
DEGREE: [5]

Talk outline

- ☐ SALSA: FrameNet-annotation of German newspaper data
- ☐ Non-compositional phenomena in SALSA
- ☐ Borderline cases
- ☐ Borderline cases in word sense annotation

Delimitation Problems

- ☐ Lexicalized support verb construction versus idiom:
  - ■ in Kraft treten (to take effect)
- ☐ Support verb construction versus extended meaning:
  - ■ Noten geben/benoten (to give grades/to grade)
  - ■ eine Stelle geben/einstellen (to employ)

Delimitation Problems

- ☐ Metaphor versus lexicalization:
    - ■ Der "Pluralismus von Erklärungen" aus der CDU/CSU-FDP-Koalition zeige, dass die Einigkeit über die Pflegeversicherung nur "vorgetäuscht" worden sei, "um über die Sommerpause zu <u>kommen</u>", sagte Klose.

        (The "multiplicity of explanations" given by the coalition showed that they only "pretended" to agree on nursing care insurance "in order to <u>get through</u> the summer break", Klose said.)

- ☐ How strongly is the metaphor/the literal reading still perceived? How far has lexicalization progressed?

Delimitation Problems

- ☐ Metaphor vs. support vs. lexicalization:
    - ■ Zwar <u>liege</u> die Verantwortung allein bei der Bundesregierung…
      (While responsibility <u>lies</u> solely with the federal government…)
    - ■ Support
    - ■ Metaphor, with a Location source
    - ■ Separate word sense of lie: being located, even for abstract objects

Delimitation problems

- ■ nehmen-4: *einen Weg nehmen* ("to use/choose a route")
    - "Man braucht heutzutage nur einen Katalog mit Vorhangstoffen aufzublättern, um zu wissen, welchen Weg eine historisch gewordene […] Moderne mit so grandiosen Werken wie Monets `` Wasserlilien '' […] am Ende in die allgemeine Akzeptanz des bloß noch Modischen nahm ."

        You only need to open a draperies catalogue to know <u>which road</u> a Modernism that has become historical, with such grandiose pieces like Monet's Water Lilies, <u>has finally taken</u> into the general acceptance of what is simply fashionable.
- ■ Metaphorical occurrence of something that
    - ■ could be analyzed as a separate sense of "nehmen" (mixture of choose and use, only for routes),
    - ■ or a support construction,
    - ■ or simply the frame Using

Annotation format issues

- ☐ SALSA: Completely separate annotation schemes for support, idiom, metaphor, "normal" frames
- ☐ No easy way to mark borderline cases
- ☐ Would be good to mark them
    - ■ because I think they were frequent -- but no way to count them now in the finished Salsa corpus
    - ■ to make the annotation more informative

Talk outline

- ☐ SALSA: FrameNet-annotation of German newspaper data
- ☐ Non-compositional phenomena in SALSA
- ☐ Borderline cases
- ☐ Borderline cases in word sense annotation

Word sense overlap in SALSA

bemerken (to remark/to notice)

Becoming aware: "Ob die da drin uns überhaupt bemerken?"
   "(I wonder) if the people inside notice us at all."

Statement: "Der Befangene sieht nicht", bemerkt er.
   "A biased person does not see", he comments.

BOTH?: Kein Wunder, daß Gerhard Schäfer in seinem Buch derzeit eine "Renaissance der Verbindungen in den neuen Ländern" bemerkt.
(It is) not surprising that Gerhard Schäfer notices/comments on a "renaissance of fraternities in the new states".

Word sense overlap in SALSA

zeigen (to show) - one of the most difficult lemmas we did!
Among the senses:

- ☐ Reasoning: He showed that the earth was flat.
- ☐ [Entity shows property]: Schlechtes Orientierungsvermögen zeigten zwei dänische Seeleute beim Auffinden ihres Schiffes. (Two Danish seamen showed a bad sense of orientation in finding their ship.)
- ☐ Judgment_direct_address: … um dem Toten noch einmal , ein letztes Mal , ihre Achtung […] zu zeigen . (to show the deceased their respect once more, for the last time)
- ☐ [Phenomenon becomes apparent]: die Grundprobleme sozialer Ungleichheit - die zeigen sich etwa im Wohnen in den Grossstaedten (the basic problems of social injustice - they show, for example, in living conditions in large cities)

Exploring borderline cases of word sense

- ☐ Cases where more than one sense seems to apply:
  seem to occur with whichever lexicon resource
- ☐ But how frequent are they really?
  - ■ Multiple sense assignment in existing corpora: 0.3% in SemCor, 8% in OpenMindWordExpert
  - ■ In Salsa, I suspect bias against multiple sense assignment from both annotation tool and guidelines
- ☐ Suppose we take non-disjoint senses to be the rule rather than the exception: How can we model this?
  - ■ Psychology: prototype / exemplar models of concept representation; computational models available

An annotation/experiment on borderline cases of word sense

- Annotation setting:

  > Not to have made a will can only **add** to that stress.
  - On a range from 1-5, how much does each sense apply?
    - Sense "increase":   [1-5]
    - Sense "state, say, tell":   [1-5]
    - Sense "change, alter, modify":   [1-5]
    - Sense "calculate":   [1-5]
    - Sense "count, enumerate":   [1-5]
    - Sense "form, constitute":   [1-5]

- WordNet senses
- Multiple annotators (at least 5 per instance)
- Experiment-like setting

Summary

- Salsa: frame (word sense) and Frame-semantic role annotation of the German TIGER corpus
- FrameNet mindset: annotate what is understood
- Limited compositionality phenomena in Salsa:
  - Metaphor, idiom, support constructions, meaning extension
  - Much trouble with borderline cases. Question of appropriate annotation format
- Borderline cases in word sense annotation in general

Word sense overlap in SALSA: figures

- Individual annotator data:
  550 lemmas, 24,825 annotated predicates
- Adjudicated:
  502 lemmas, 18,671 annotated predicates
- Overall word sense overlap: quite rare
  - single-annotator data:
    1185 / 24,835 items: 4.77%
  - adjudicated data:
    924 / 18,671 items: 4.95%

Word sense overlap in SALSA: sample lemmas

Lemma	Senses most involved in overlap	items with overlap
besichtigen (visit)	Scrutiny, Perception_ active	100% (10 / 10)
treiben (drive/propel/float)	Causation, Cause_motion, Reason	71% (22 / 31)
abloesen (detach/displace)	Change_of_Leadership, Take_place_of	56% (14 / 25)
beobachten (watch/observe)	Becoming_aware, Perception_active	52% (17 / 33)
kennen (know)	Awareness, [familiarity]	41% (24 / 58)
warnen (warn)	Statement, [communicate danger])	41% (47 / 114)

nehmen: Support constructions

- in Anspruch nehmen   (to take advantage of sth)
- Bezug nehmen auf   (to refer to)
- Abschied nehmen   (to take leave of s.o.)
- Anlauf nehmen   (to take a running start)
- in Angriff nehmen   (to tackle)
- einen Verlauf nehmen (to progress)
- in Empfang nehmen   (to accept)
- …

nehmen: Metaphors

- den Wind aus den Segeln nehmen
    (to take the wind out of someones' sails / defuse s.o.'s arguments)
- eine Last auf sich nehmen
    (to take on a burden)
- den/seinen Hut nehmen
    (to take ones' hat/to quit)
- aufs Korn nehmen
    (to take sth in the crosshairs / to scrutinize, to attack)
- unter die Lupe nehmen
    (to put under a magnifying glass / to scrutinize)
- …

nehmen: Idioms

- in Kauf nehmen
    (to put up with / "take into purchase")
- zum Mann nehmen
    (marry / take as husband)
- eine Hürde/Barriere nehmen
    (to take a hurdle)
- das Heft in die Hand nehmen
    (to take control / "take the hilt")
- jemanden beim Wort nehmen
    (to take someone at their word)
- …

Note: each idiom gets its own, specialized frame