

Proseminar Linguistische Annotation

Ines Rehbein und Josef Ruppenhofer

SS 2010



Vergleich verschiedener Annotationschemata

- **Feinkörnige versus grobkörnige Bedeutungsunterscheidungen**

- | | |
|-------------------|-------------------------------|
| ▶ Ausgang | SALSA: 9 Frames, GN: 3 Frames |
| ▶ ausgehen | SALSA: 9 Frames, GN: 9 Frames |

- Sind die Unterscheidungen klar verständlich?
- Sind die Bedeutungen klar voneinander abgegrenzt?
- Sind die Bedeutungsunterscheidungen vollständig?
- Können die Bedeutungen valide und reliabel annotiert werden?

- **Annotationsrichtlinien**

- ▶ Sind die Richtlinien klar verständlich?
- ▶ Werden die Bedeutungen gut beschrieben?
- ▶ Gibt es hinreichende Beispiele?
- ▶ Werden Kriterien genannt, die helfen, die Bedeutungen zu unterscheiden?

ausgehen SALSA - GermaNet

SALSA	GermaNet
Becoming	verben.Allgemein (Vorschlag)
Certainty	verben.Koerperfunktion (Haare)
Start, origin	verben.Kognition (PP auf)
Diminish	verben.Lokation
Cause_effect	verben.Schoepfung (Wirkung)
Going_out	verben.Veraenderung1 (Sache)
Hypothesize	verben.Veraenderung2 (Licht)
End_with_result	verben.Veraenderung3 (Straßen)
Emit_entity	verben.Veraenderung4 (Geduld)

Table: Wortbedeutungen für *ausgehen* in SALSA und GermaNet

ausgehen SALSA - Verteilung der Wortbedeutungen

Frame	Ben	Mel	B & M	Con	Joh	C & J	Sou
Becoming	1	1	1	1	3	1	0
Cause_effect	5	3	5	4	2	3	1
Certainty	11	34	31	22	23	9	24
Diminish	3	3	3	3	3	2	1
Emit_entity	3	3	3	4	5	4	1
End_with_result	2	2	2	2	1	2	3
Going_out	3	3	3	4	5	3	5
Hypothesize	26	2	5	14	45	27	12
Start	0	3	0	1	2	0	4

Table: Verteilung der Wortbedeutungen für *ausgehen* in den annotierten und adjudizierten Daten

ausgehen SALSA - Disagreement

#	B & M	C & J
1	Cause_effect	Unannotated
1	Certainty	Emit_entity
1	Diminish	Hypothesize
1	Emit_entity	Cause_effect
1	Emit_entity	Unannotated
1	Hypothesize	Certainty
2	Cause_effect	Emit_entity
22	Certainty	Hypothesize

Table: Unterschiede in der Annotation von *ausgehen* (adjudizierte Daten)

Ausgang GN - Disagreement

Frame	Con	Joh	C & J	Dog	Pat	D & P
Artefact	1	2	1	4	0	
Geschehen	35	35	37	35	37	
Ort	15	14	13	12	14	

Table: Verteilung der Wortbedeutungen für *Ausgang* in den annotierten und adjudizierten Daten

Ausgang GermaNet - Artefakt, Ort?

s41: Das Gedächtnis aber ist ein stetig verwildernder Garten am Haus, Ein- und **Ausgang** sind markiert von einer Metapher : einem schief in den Angeln hängenden Tor .

Ausgang GermaNet - Geschehen, Ort?

s47: Niemand weiß , was am **Ausgang** dieser kritischen Phase stehen wird .

Ausgang GermaNet - Artefakt, Geschehen, Ort?

s40: Im Warenverkehr nahmen Eingänge und **Ausgänge** etwa im gleichen Ausmaß zu , sodaß das Warenverkehrsdefizit nahezu stabil blieb .

Wareneingang - Artefakt, Geschehen, Ort?

Der LKW des Lieferanten oder der Paketdienst steht am **Wareneingang** und will seine Lieferung abgeben. Der Lagerist hat alle Informationen zu den erwarteten Lieferungen in seinem Rechner, der direkt am **Wareneingang** steht.

Über die Bestellnummer, den Lieferanten oder den angelieferten Artikel kann er aus der Liste der geplanten **Wareneingänge** den Richtigen herausfinden.

[...]

Die Inventur erfolgt für einzelne Lager-Einheiten.

Im **Warenausgang** werden verfügbare Lager- Einheiten aus dem Lager abgebucht und auf einen Lieferschein gebucht. Dadurch wird der Lagerbestand automatisch korrigiert. reneingang kann über seinen gesamten

Jeder **Wareneingang** kann über seinen gesamten Lebenszyklus im Lager verfolgt werden.

Gründe für mangelndes IAA

- Unklare Richtlinien
- Fehlende Bedeutungen
- Annotator-Bias
- Echte Ambiguitäten, fehlende Kontextinformationen
- Fehlendes Weltwissen
- Müdigkeit, mangelnde Konzentration
- “lumping” versus “splitting” (allgemeine vs. spezifischere Bedeutung)

Verbesserung des Annotationsprozesses

- Wie kann der Annotationsprozess unterstützt und die Qualität der Annotationen gewährleistet werden?
 - ▶ verständliche, eindeutige Richtlinien
 - ▶ umfassende Beispiele
 - ▶ ???