# Ausgewählte Techniken der Maschinellen Übersetzung

Susanne J. Jekat

ZHW

E-mail: jes@zhwin.ch, Subject: MTZH

# Übersicht

Themenbereiche:

1. Erstellung und Standardisierung von Ressourcen für die Maschinelle Übersetzung (Offene Fragen)
2. **Grammatikformalismen für die Maschinelle Übersetzung**
3. Maschinelles Dolmetschen
4. Computergestützte Übersetzung
5. Evaluation von Systemen zur Maschinellen Übersetzung

# Offene Fragen

Unterschied zwischen SGML und XML?

Rhetorical Structure Theory?

# Unterschied zwischen SGML und XML

Grundlage für die Entwicklung von XML war die in den 80er-Jahren von IBM zu Dokumentationszwecken entwickelte Auszeichnungssprache SGML (Standard Generalized Markup Language). Der praktische Einsatz gestaltete sich für interessierte Anwender jedoch aufgrund der großen Komplexität von SGML äußerst schwierig. Als verschlankte Untermenge von SGML gibt XML dem Nutzer klar definierte Syntaxbausteine an die Hand, um Daten in elektronischen Dokumenten strukturiert beschreiben zu können. So lassen sich mit XML Daten abbilden, die in anderen Anwendungen weiterverarbeitet oder zum Beispiel mit einem Web-Browser auf dem Bildschirm angezeigt werden können. Durch die Möglichkeit bei XML flexibel eigene Steuerelemente (Tags) zu definieren, wurden gleichzeitig die starren Beschränkungen von HTML in Bezug auf die Beschreibungsvarianten (nur ca.70) aufgelöst

# Rhetorical Structure Theory

Formaler Rahmen für die Analyse der funktionalen Beziehungen zwischen Segmenten eines Textes und grundlegende Methode der Textgenerierung

Mann, W.C. & Thompson, S.A. (1987) Rhetorical Structure Theory. A Theory of Text Organisation. In Polanyi, L. (ed) The Structure of Discourse, Norwood, N.J..

# RST: Formale Elemente

Relationen werden definiert als Beziehungen zwischen nicht überlappenden Textsegmenten (meist Sätzen)

Nukleus: Hauptsegment

Satellit: Nebensegment

# Beispiel Restatement (Mann & Thompson, 1987:71)

**relation name**: RESTATEMENT

**constraints on N**: none

**constraints on S**: none

**constraints on the N + S combination**:

S restates N, where S and N are of comparable bulk

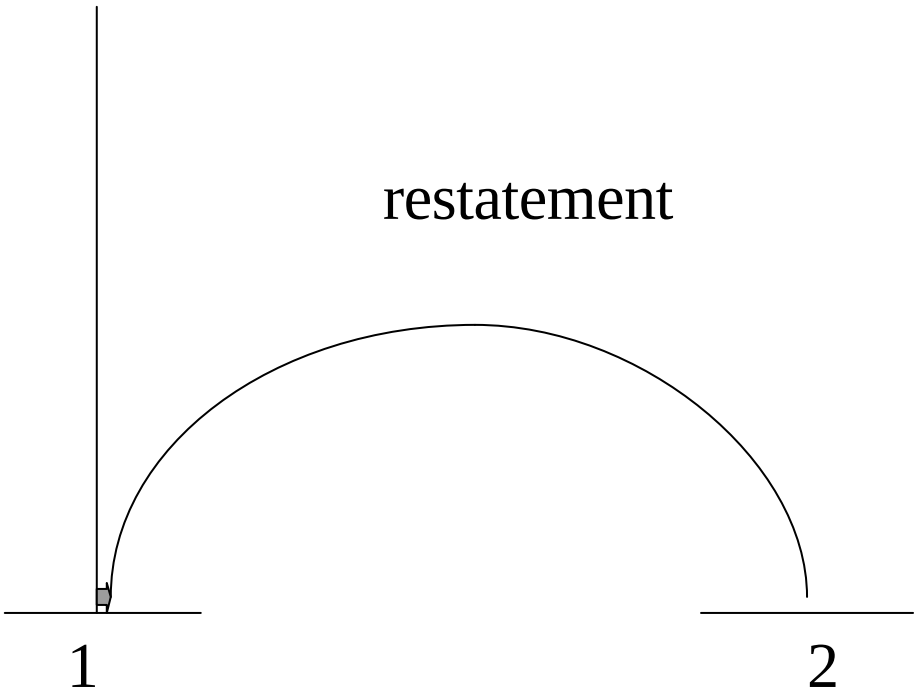
**the effect**: R(eader, Anm.S.J.) recognizes S as a restatement of N

**locus of the effect**: N and S

# RST, Restatement

1. A well-groomed car reflects its owner.
2. The car you drive says a lot about you.

1 -2



# Weitere Eigenschaften von Texten (Übergang zur Semantik auf Textebene)

Kohärenz: semantischer und pragmatischer Zusammenhang von aufeinanderfolgenden Sätzen, z. B. darstellbar in Form von Konzeptnetzen

Kohäsion: linguistische Mittel, die Zusammenhänge an der Textoberfläche wiederspiegeln, z.B. Pronomen oder parallele Satzstrukturen.

# Textqualität

Kohäsion + Kohärenz

ein kohärenter, nicht kohäsiver Text behandelt konsequent ein Thema, verwendet jedoch keine linguistischen Mittel, um den Zusammenhang zu verstärken

ein kohäsiver, nicht kohärenter Text erscheint zusammenhängend, hat aber keinen Sinnzusammenhang

# Thema 2: Grammatikformalismen für die Maschinelle Übersetzung

Übersicht Thema 2:

1. Einführung

2. Head Driven Phrase Structure Grammar

3. Lexical Functional Grammar

# Thema 2: Grammatikformalismen für die Maschinelle Übersetzung

Wissensgebiete zu Thema 2:

a) Syntax und Morphologie (Analyse des Quelltextes, zur Verfügung gestellt von Ressourcen, auf Satzebene)
b) Textlinguistik (Analyse des Quelltextes oberhalb der Satzebene)
c) Merkmalsstrukturen
d) Kontrastive Linguistik
e) MT (Grundlagen)
f) Translationswissenschaft? (eher Probleme der Translation)

# Syntax und Morphologie

„Syntax: System von Regeln, die beschreiben, wie aus einem Inventar von Grundelementen (Morphemen, Wörtern, Satzgliedern) durch spezifische syntaktische Mittel (Morphologische Markierung, Wort- und Satzgliedstellung, Intonation u.a.) alle wohlgeformten Sätze einer Sprache abgeleitet werden können. [...]. Die Grenzen zu Morphologie und Semantik sind fliessend, ihre Präzisierung daher theorieabhängig.“Bussmann (2002:676)

# Mindestens zu verarbeitende Phänomene auf Satzebene

- Identifikation der Teile eines Satzes
  - z.B. durch Konstituentenanalyse
- Beschreibung der Teile eines Satzes
  - z.B. Morphologie
- Beschreibung der Bedeutung eines Satzes
  - z.B. Frege-Prinzip

# Konstitutentenanalyse

Konstitutenten: in der strukturellen Satzanalyse Bezeichnung für jede sprachliche Einheit, die Teil einer grösseren Einheit ist

Identifikation durch Verschiebe- und Ersatzprobe: ist der zu analysierende Ausdruck im Satz frei verschiebbar, gilt er als Konstituente

# Konstituentenanalyse

Beispiel: Susanne hält einen Vortrag

*Susanne* ist ersetzbar durch *sie* (NP)

*hält einen Vortrag* ist ersetzbar durch *doziert* (VP)

die durch den jeweils ersten Zerlegungsschritt gewonnenen Elemente heissen unmittelbare Konstituenten

# Exkurs: Psychologische Realität der Konstituentenstruktur

Im Zweiten Weltkrieg
verfolgten die Nationen
sogar skurrile Pläne
wenn sie nur hoffen liessen
dass der Krieg bald endet.

Im Zweiten
Weltkrieg verfolgten die
Nationen sogar skurrile
Pläne wenn sie nur hoffen
liessen dass der Krieg bald
endet.

# Morphologie

(im Strukturalismus) Untersuchung der Form, inneren Struktur, Funktion und Vorkommen der Morpheme als kleinste bedeutungstragende Einheiten der Sprache.

Ermittlung durch operationale Verfahren: Verschiebe- und Ersatzproben (s. Folie 9)

(! Übergang zu Syntax fliessend: Den Hund beisst die Katze.)

# Frege-Prinzip

auch KOMPOSITIONALITÄTSPRINZIP

Ein meist G. Frege (1848-1925) zugeschriebenes Prinzip, demzufolge die Bedeutung eines komplexen Ausdrucks eine Funktion der Bedeutung seiner Teile und der Art ihrer syntaktischen Kombination ist. Hieraus ergibt sich das

Substitutionsprinzip: in einer komplexen Formel dürfen denotatgleiche Ausdrücke füreinander ersetzt werden, ohne dass sich das Denotat des Gesamtausdrucks ändert

# Anwendung des Substitutionsprinzips

Der Bundeskanzler hat die Richtlinienkompetenz.

Der Bundeskanzler ist Gerhard Schröder.

Gerhard Schröder hat die Richtlinienkompetenz.

(Intuitiv einwandfreier Schluss nach dem Substitutionsprinzip)

# Aber:

Der Bundeskanzler hat immer die Richtlinienkompetenz.

Gerhard Schröder ist Bundeskanzler.

??? Gerhard Schröder hat immer die Richtlinienkompetenz.

(Schluss nach dem Substitutionsprinzip ist mindestens fragwürdig)

=> kontextuelle Information notwendig für Interpretation

# Mindestens zu verarbeitende Phänomene auf Textebene

- Anapher
- Ellipse
- Pronominalisierung

# Anapher

hier: sprachliche Einheiten, die sich auf Elemente des vorangegangenen Kontextes beziehen (verknappte Wiederaufnahme)

(! weitere Definitionen)

# Anapher

vgl. MT DE ➔ FR

Michael kauft sich ein Fahrrad. Er will damit nach Frankreich fahren. (le vélo)

Michael kauft sich ein Auto. Er will damit nach Frankreich fahren. (la voiture)

# Ellipse

Aussparung von syntaktisch notwendigen sprachlichen Elementen, die aus dem sprachlichen Kontext (oder der Redesituation) rekonstruierbar sind.

Der Terrorist gestand endlich.

# Pronominalisierung

Ersetzung von referenzidentischen nominalen Syntagmen durch Personalpronomina.

Vorwärtspronominalisierung auf Satzebene:

Therese behauptet, dass sie eine gute Studentin ist.

Rückwartspronominaliserung auf Satzebene:

Bevor sie Beate besuchte, ging Susanne in den Zoo.

# Pronominalisierung auf Textebene

MT FR ➔ DE

Il tirait la balle. Elle disparait.

Er feuerte die Kugel ab. ***Sie*** verschwand.

Er trat den Ball. ***Er*** verschwand.

# Typen von Grammatikformalismen

Traditionelle Grammatik:

Beschreibung von Syntax und Morphologie durch präskriptive Regeln und Beispiele

„Präpositionen „regieren“ bestimmte Kasus. Meistens folgt einer Präposition immer der gleiche Kasus: Er wohnt ausserhalb des Dorfes (Genitiv) . Er befindet sich ausserhalb des Hauses (Genitiv).“ Grosser Duden Grammatik, 1959

# Phrasenstrukturgrammatik

auch Konstituentenstrukturgrammatik

Der syntaktische Aufbau von Sätzen wird in Form von hierarchisch geordneten Konstituenten beschrieben.

S ➔NP + VP

Bei Oberflächenorientiertheit syntaktisch-semantische Probleme: Susanne liest den Vortrag ab.

# Generalized Phrase Structure Grammar

- Grammatikmodell ohne Transformationen und mit einer einzigen Repräsentationsebene
- Die syntaktische Repräsentation erfolgt in Form eines Phrasenstrukturbaumes, syntaktische Kategorien in Form partiell spezifizierter Merkmalsstrukturen

# Merkmalsstrukturen

# Quellen


- Bussmann, Hadumod (2002)  Lexikon der Spradchwissenschaft. Stuttgart:Kröner.
- Butt, Miriam, Holloway King, Tracy, Nino, M. & Segond, F. (2000) A Grammar Writer`s Cookbook. Stanford: CSLI Publications
- Klabunde, Ralf et al. , ed, (2004) Computerlinguistik und Sprachtechnologie. Eine Eine Einführung. Heidelberg:Spektrum.
- Sag, I. & Wasow, T. (1999) Syntactic Theory: A formal Introduction. CSLI Publications