

# Satz Umstrukturierung für statistisch maschinelle Übersetzung

Anna Schiffarth

Dozentin: Miriam Kaeshammer

Fortgeschrittene Methoden der  
statistisch maschinellen Übersetzung

# Einführung

- Beschreibung einer Methode für die Einbeziehung syntaktischer Informationen in statistisch maschinellen Übersetzungs- Systemen
- 1. Schritt: Analyse der Quell-Sprache die zu übersetzen ist
- 2. Schritt: Anwendung einer Reihe von Transformationen bezogen auf den Analyse Baum
- Ziel: korrekte Wortordnung der Zielsprache in Anlehnung an die Quellsprache

# Einführung

- Forschungen der statistisch maschinellen Übersetzung führten zu der Entwicklung des *phrase-based-systems*
- Erlaubt *multi-word units* (Phrasen) einer Sprache direkt in eine andere zu übersetzen
- Einschränkung ist, dass das System so gut wie keine Rücksicht auf syntaktische Informationen nimmt
- Syntaktische Informationen sind wichtig
- Aufschluss über verschiedene Wortordnungen/-stellungen einer Sprache

# Einführung

- Ansatz für den Umgang mit syntaktischen Informationen und dem phrase-based-system
  1. Analyse der Quell-Sprache, die übersetzt werden soll
  2. Reihe von Transformationen werden durch einen Parse Tree dargestellt → so kann man das Naheliegendste am besten raus filtern
  3. Ziel: herausfinden der besten Wortstellung der Zielsprache unter Berücksichtigung der Quellsprache

# Beispiel

- Quell-Satz:

Ich werde Ihnen die entsprechenden Anmerkungen aushändigen, damit Sie das eventuell bei der Abstimmung übernehmen können.

- Englische Übersetzung:

I will to you the corresponding comments pass on, so that you them perhaps in the vote adopt can.

# Beispiel

- Umgeordneter Satz:

Ich werde aushändigen Ihnen die entsprechenden Anmerkungen , damit Sie können übernehmen das eventuell bei der Abstimmung.

- Englische Übersetzung:

I will pass on to you the corresponding comments, so that you can adopt them perhaps in the vote.

# Phrase-based SMT

- Phrase-based Systems stützen sich auf IBM Modelle indem mehrere Wörter einer Sprache mehrere Wörter einer anderen Sprachen entsprechen
- Bsp: I will in Englisch = Ich werde in Deutsch
- Stützt sich zunächst auf word-to-word Übersetzungen
- Bei Eingabe eines Satzes führt das aber zu falschen Ergebnissen, Wortübersetzungen sind korrekt aber die Wortreihenfolge ändert sich in Übersetzungen nicht sinngemäß

# Syntax-based SMT

- Hierbei wird beim Übersetzungsprozess auf die syntaktische Repräsentation der Quell- und Zielsprache geachtet
  1. Ansatz: Verwendung von `bitext` Grammatiken, welche die Quell- und die Zielsprache analysiert
  2. Ansatz: nimmt Gebrauch von den syntaktischen Informationen der Zielsprache und transformiert das Übersetzungsproblem in ein Analyseproblem

Es soll die beste Ausgabe generiert werden!

# Vorverarbeitender Ansatz

- Enthält einen vorbereitenden Schritt
- Zu übersetzende Sätze einer Sprache werden modifiziert bevor sie an ein phrase-based Übersetzungssystem gegeben werden
- Der Satz wird in seine syntaktischen Eigenschaften zerlegt mit Hilfe eines Parse tree
- Deutsch hat die meisten unterschiedlichsten Wortstellungsmöglichkeiten
- Hat eine relativ freie Wortordnung im Gegensatz zu Französisch und Englisch
- Sowie hohe Flexibilität welche Phrase an den Satzanfang kommt

# Parse Tree

```
S PPER-SB Ich
  VAFIN-HD werde
    VP PPER-DA Ihnen
      NP-OA ART die
        ADJA entsprechenden
        NN Anmerkungen
      VVINF-HD aushaendigen
    ' '
  S KOUS damit
    PPER-SB Sie
      VP PDS-OA das
        ADJD eventuell
        PP APPR bei
          ART der
          NN Abstimmung
        VVINF-HD uebernehmen
      VMFIN-HD koennen
```

# Parse Tree

## Erläuterungen:

**Figure 1:** An example parse tree. Key to non-terminals: PPER = personal pronoun; VAFIN = finite verb; VVINF = infinitival verb; KOUS = complementizer; APPR = preposition; ART = article; ADJA = adjective; ADJD = adverb; -SB = subject; -HD = head of a phrase; -DA = dative object; -OA = accusative object.

# Deutsche Satz Struktur

## Beispiel

**Clause 1:** Ich/*I* werde/*will* Ihnen/*to you* die/*the*  
entsprechenden/*corresponding*  
Anmerkungen/*comments* aushaendigen/*pass\_on*

**Clause 2:** damit/*so that* Sie/*you* das/*them*  
eventuell/*perhaps* bei/*in* der/*the* Abstimmung/*vote*  
uebernehmen/*adopt* koennen/*can*

# Neuordnung mit phrase-based SMT

- Durch die Neuordnung wird die Übersetzung von Phrasen wie *werde aushändigen* in *will pass on* möglich
- Man kann Phrasen verschiedener Längen festlegen → vereinfacht die Übersetzung
- „Wir machen die Tür auf“
  - machen ... auf kommt von *auf machen* im deutschen und ist im englischen komplexer zu übersetzen: auf = to open; machen = to make
  - Das deutsche Verb auf machen ist im englischen durch zwei Verben dargestellt

# Satz Restrukturierung

- Methode für die Neuordnung deutscher Sätze
  1. Analyse des Satzes
  2. Anwendung einer Sequenz von Regeln welche den Satz umordnen stützend auf dem parse tree
- Es folgen 6 Restrukturierungsschritte

# 1 Verb initial

- Verb phrase hat das label -VP
  - Der Kopf (head) der phrase hat das label –HD  
diese muss in die initial Position innerhalb der  
verb phrase gerückt werden
- *Übernehmen* wurde so bewegt das es dem  
*das* vorausgeht (VP-OC)

# 1 Verb initial

S-MO KOUS-CP damit  
PPER-SB Sie  
VP-OC VVINF-HD uebernehmen  
PDS-OA das  
ADJD-MO eventuell  
PP-MO APPR-DA bei  
ART-DA der  
NN-NK Abstimmung  
VMFIN-HD koennen

## 2 Verb 2nd

- In einem untergeordnetem Satz markiert mit S-..., mit einem Komplementizer KOUS, PREL, PWS oder PWAV muss man den Kopf des Satzes finden und dorthin bewegen das es dem Komplementizer direkt folgt

## 2 Verb 2nd

S-MO KOUS-CP damit  
VMFIN-HD koennen  
PPER-SB Sie  
VP-OC VVINP-HD uebernehmen  
PDS-OA das  
ADJD-MO eventuell  
PP-MO APPR-DA bei  
ART-DA der  
NN-NK Abstimmung

## 2 Verb 2nd

- Kopf des Satzes ist *können*
- Wird so bewegt das dem Komplementizer *damit* folgt

# 3 Move Subject

- Für jeden Satz der mit dem label S markiert ist
- Bewegt man das Subjekt so das es dem Kopf vorrausgeht
- Subjekt ist definiert als left-most child mit dem label ...-SB oder PPER-EP und dem Kopf ..-HD

# 3 Move Subject

S-MO KOUS-CP damit  
PPER-SB Sie  
VMFIN-HD koennen  
VP-OC VVINF-HD uebernehmen  
PDS-OA das  
ADJD-MO eventuell  
PP-MO APPR-DA bei  
ART-DA der  
NN-NK Abstimmung

# 4 Particles

- Bei Verb Partikel Konstruktionen muss das Partikel so bewegt werden das es dem Verb vorrausgeht
- Label für finite Verben = VVFIN
- Label für Partikel = PTKVZ

```
S PPER-SB    Wir
  VVFIN-HD   fordern
  NP-OA ART  das
             NN  Praesidium
  PTKVZ-SVP  auf
```

```
S PPER-SB    Wir
  PTKVZ-SVP  auf
  VVFIN-HD   fordern
  NP-OA ART  das
             NN  Praesidium
```

# 5 Infinitive

- Infinitive Verben sind oft nicht in der korrekten Position nach der Transformation 1-4
  - Deswegen zufügen eines zweiten Schrittes, welcher Infinitive berücksichtigt
    1. Entfernen aller internen VP Knoten innerhalb des Parse Baumes
    2. Infinitiv so bewegt das es direkt dem finitem Verb folgt, wenn der Satz ein finites und ein Infinitives Verb enthält zudem muss ein Argument zwischen den beiden bestehen
- BSP: infinitiv *einreichen* finites Verb *konnten* mit dem direkten Objekt *es*

# 5 Infinitive

S PPER-SB Wir  
VMFIN-HD konnten  
PPER-OA es  
PTKNEG-NG nicht  
VP-OC VVINF-HD einreichen  
AP-MO ADV-MO mehr  
ADJD-HD rechtzeitig

The transformation removes the VP-OC, and moves the infinitive, giving:

S PPER-SB Wir  
VMFIN-HD konnten  
VVINF-HD einreichen  
PPER-OA es  
PTKNEG-NG nicht  
AP-MO ADV-MO mehr  
ADJD-HD rechtzeitig

# 6 Negation/Verneinung

- Negative Partikel haben das Label PTKNEG
- Sie folgen dem finitem Verb

```
S PPER-SB Wir
  VMFIN-HD konnten
  PTKNEG-NG nicht
  VVINF-HD einreichen
  PPER-OA es
  AP-MO ADV-MO mehr
           ADJD-HD rechtzeitig
```

Transformation	Example
Verb Initial	<p>Before: Ich werde Ihnen die entsprechenden Anmerkungen <u>aushändigen</u>, ...</p> <p>After: Ich werde <u>aushändigen</u> Ihnen die entsprechenden Anmerkungen, ...</p> <p>English: I shall be <u>passing on</u> to you some comments, ...</p>
Verb 2nd	<p>Before: ... damit Sie uebernehmen das eventuell bei der Abstimmung <u>koennen</u>.</p> <p>After: ... damit <u>koennen</u> Sie uebernehmen das eventuell bei der Abstimmung.</p> <p>English: ... so that could you adopt this perhaps in the voting.</p>
Move Subject	<p>Before: ... damit koennen <u>Sie</u> uebernehmen das eventuell bei der Abstimmung.</p> <p>After: ... damit <u>Sie</u> koennen uebernehmen das eventuell bei der Abstimmung.</p> <p>English: ... so that you could adopt this perhaps in the voting.</p>
Particles	<p>Before: <u>Wir</u> fordern das Praesidium <u>auf</u>, ...</p> <p>After: <u>Wir</u> <u>auf</u>fordern das Praesidium, ...</p> <p>English: <u>We</u> ask the Bureau, ...</p>
Infinitives	<p>Before: Ich werde der Sache <u>nachgehen</u> dann, ...</p> <p>After: Ich werde <u>nachgehen</u> der Sache dann, ...</p> <p>English: I will <u>look into</u> the matter then, ...</p>
Negation	<p>Before: <u>Wir</u> koennten einreichen es <u>nicht</u> mehr rechtzeitig ...</p> <p>After: <u>Wir</u> koennten <u>nicht</u> einreichen es mehr rechtzeitig ...</p> <p>English: <u>We</u> could not hand it in in time, ...</p>

# Experiment

- 100 Sätze
- Satzlänge von 10-20 Worten
- 2 Kommentatoren
- 3 Übersetzungsmöglichkeiten
  1. Menschliche Übersetzung
  2. Ausgabe durch ein Baseline System
  3. Ausgabe durch ein System stützend auf die Neuordnung

# Experiment

→ Neuordnungssystem gibt eine bessere Übersetzung als ein Baseline System

$$f(x) = + \text{ If } fr(x) > fb(x)$$

→ Baseline System produziert eine bessere Übersetzung als ein Neuordnungssystem

$$f(x) = + \text{ If } fb(x) > fr(x)$$

→ Beide Systeme produzieren eine gleiche Übersetzung für den gegebenen Satz

$$f(x) = 0 \text{ If } fr(x) = fb(x)$$

# Experiment

- Die Übersetzung mit den Neuordnungssystemen ergab die besten Ergebnisse

# Quellen

Clause Restructuring for Statistical Machine Translation, M. Collins et al., ACL 2005