Architecture of the Lucy Translation System

Dr. Petra Gieselmann
Lucy Software and Services GmbH, Munich
Contents

- Rule-Based Translation Process
- Statistical Enhancements
- Discussion
Translation Method

Analysis
- Morpholog. Analysis
- Parsing
- Framing/ Multiwords
- Phrasal Analysis
- Anaphora Resolution

Transfer
- Structural Transfer
- contextual Transfer
- Lexical Transfer

Generation
- TLDepend. Transformations
- TL Word Order
- Morpholog. Generation

Chart
- SL Tree
- TL Tree
Input Sentence

“The ballet dancers performed very well”
Morphological Analysis

Position
Length

Allomorph Table (b-tree)

R E I S T A F E L

0 1 2 3 4 5 6 7 8

1

R ABB
r DET
Eis NST
eis VST
eis VAD
ta ABB
ta VST
fe ABB
l ABB

Reis NST Tafel NST
reis VST tafel VST
tafel VST

Transition Matrix

<table>
<thead>
<tr>
<th>start</th>
<th>ABB</th>
<th>DET</th>
<th>NST</th>
<th>VAD</th>
<th>VST</th>
<th>end</th>
</tr>
</thead>
<tbody>
<tr>
<td>start</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>ABB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>DET</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>NST</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>VAD</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>VST</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>end</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Analysis: Word Formation

“The ballet dancers performed very well”

Word Formation Rules
Analysis: Homography

"The ballet dancers performed very well"

Homography (Lexical Ambiguity)
"The ballet dancers performed very well"
Analysis: Proper Names

“John Major went down the hill”

Compounding & Proper Names
Analysis: Building up Phrases

Samples:

NP → DET NO
NP → DET AP NO
CLS → NP PRED NP PP
CLS → NP PRED ADVP

“The ballet dancers performed very well”
Analysis: Grammar Rules

Grammar Rule

- Tests
  - NP --> DET(1) NO (2)
  - (check-agreement 1 2)

- Constructors
  - Feature Traffic

Context-free Grammar enhanced with Tests and Transformations
Example of a Rule

Rule

ID  VB-NO-VB-1
    VB  NO  VB

Tests

TEST
(check-general-info?)
(check-compound-vb?)

Constructors

CONSTR
(update-general-info)
(feature-traffic-from-son 2)
(update-compound-vb)

(decide-quote-xp-parse-cat)

LEVEL  7
TYPE  WORD
AUTHOR  "eugenia"
EXAMPLE  "horse riding"
Analysis: Cross-categorical Phenomena

- Coordination
- Comparatives
- Complementizers
- Negation
- Commas
- Orthography
- Quotes
Monolingual Lexicon

- Morphological
- Syntactic
- Semantic

Morphological Analysis
Syntactic Analysis
## Monolingual Lexicon

<table>
<thead>
<tr>
<th>CAN</th>
<th>“leave”</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT</td>
<td>VST</td>
</tr>
<tr>
<td>TT</td>
<td>(I T DT)</td>
</tr>
<tr>
<td>PV</td>
<td>(“by”)</td>
</tr>
</tbody>
</table>
| ARGS | (((SSUBJ N1) (SDOBJ N1) OPT (ADV LOC) (SIOBJ N1 (TYPE P1) (PREP “for”)))
|      | ((SSUBJ N1) (SDOBJ N1) (SIOBJ N1 (TYN SOC PRO POT HUM ANI) (PREP “to”)))
|      | ((SSUBJ N1 NO (ICP-ING-SUBJ)) (SDOBJ N1) (SOCOMP ADJ N0 (ICP-ING)))
| ALO  | “leav”           |
| CL   | (G-ING I-E PR-ES1) |

---

## Monolingual Package

- **Canonical Information**
- **Morphological Information**
**Multiword Entries**

<table>
<thead>
<tr>
<th>CAN</th>
<th>“pillow case”</th>
<th>NO ⇒ NO+NO</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT</td>
<td>NST</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MW-HEAD</td>
<td>“case”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MW-TYPE</td>
<td>NST-NST</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MW-BODY</td>
<td>((NST “pillow” (NU SG))</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(HEAD))</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

00 “pillow case” NST ⇒ “Kopfkissenbezug” NST

00 “Kopfkissenbezug” NST ⇒ “pillow case” NST

<table>
<thead>
<tr>
<th>CAN</th>
<th>“leave of absence”</th>
<th>NO ⇒ NO+STRING</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT</td>
<td>NST</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MW-HEAD</td>
<td>“leave”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MW-TYPE</td>
<td>NST-STRING</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MW-BODY</td>
<td>((HEAD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(STRING “of absence”))</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

00 “leave of absence” NST ⇒ “Beurlaubung” NST

00 “Beurlaubung” NST ⇒ “leave of absence” NST
Process of Labelling the Constituents of a Clause with a Role Value:

- To see if the Element being attached is a possible Role
- To check that the Sentence is complete
- To filter Analysis Trees
- To translate better:
  - I saw him → Je l’ai vu.
  - I gave the book to him. → Je lui ai donné le livre.
Output of the syntactic Analysis:

- **Success:** 1 Interpretation

- **Failure:** Phrasal Analysis
Transfer

- Structural Transfer
- Contextual Transfer
- Lexical Transfer
Transformation of SL Structure into TL Structure

“Das gestern von ihm gekaufte Auto ist blau.”

“The car bought yesterday by him is blue.”
For all the Categories with Complements

The coach took these facts into account

These facts were taken into account by the coach

42203 take ➔ “bemächtigen”
(XX-VST-DOBJ :CAN “hold”)
42203 take ➔ ergreifen
(XX-VST-DOBJ :CAN “hold”)
(XX-VST-NONROLE :CAT “PP” :CAN “of” :TYN* HUM)
41203 take ➔ führen
(XX-VST-DOBJ :TYN HUM)
(XX-VST-POBJ :CAN “to” :TYN LOC)
41203 take ➔ berücksichtigen
(XX-VST-DOBJ)
(XX-VST-NONROLE :CAN into” :HEADCAN “account”)
10203 take ➔ dauern
(XX-VST-ADV :ADVTYPE TMP)
10000 take ➔ nehmen
Lexical Transfer

For all the Categories without Complements

00 “man” NST ⇐ “Mann” NST

(XLX)

<table>
<thead>
<tr>
<th>NST: Mann</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALO</td>
</tr>
<tr>
<td>CAN</td>
</tr>
<tr>
<td>CAT</td>
</tr>
<tr>
<td>CL</td>
</tr>
<tr>
<td>NU</td>
</tr>
<tr>
<td>OR</td>
</tr>
<tr>
<td>SL-CAN</td>
</tr>
<tr>
<td>SL-CAT</td>
</tr>
<tr>
<td>TL-CAN</td>
</tr>
<tr>
<td>TL-CAT</td>
</tr>
<tr>
<td>WORD#</td>
</tr>
</tbody>
</table>
Generation

- Morphological Generation

- Target Language dependent Operations
## Morphological Generation

<table>
<thead>
<tr>
<th>(“Mann”NST)</th>
<th>(“Mann”NST)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALO “Mann”</td>
<td>ALO “Männer”</td>
</tr>
<tr>
<td>CL (S-ES)</td>
<td>CL (P-ER)</td>
</tr>
<tr>
<td>DR (NP RD)</td>
<td>DR (NP RD)</td>
</tr>
<tr>
<td>GD M</td>
<td>GD M</td>
</tr>
<tr>
<td>KN CNT</td>
<td>KN CNT</td>
</tr>
<tr>
<td>SX (M)</td>
<td>SX (M)</td>
</tr>
<tr>
<td>TYN (HUM))</td>
<td>TYN (HUM))</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(“es2”N-FLEX)</th>
<th>(“er2”N-FLEX)</th>
<th>(“er2”N-FLEX)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALO “es”</td>
<td>ALO “er”</td>
<td>ALO “er”</td>
</tr>
<tr>
<td>CA (A N)</td>
<td>CA (G)</td>
<td>CA (N)</td>
</tr>
<tr>
<td>CL (S-3)</td>
<td>CL (P-E1)</td>
<td>CL (S-1)</td>
</tr>
<tr>
<td>NU (SG)</td>
<td>NU (PL)</td>
<td>NU (SG)</td>
</tr>
<tr>
<td>PLC (NI))</td>
<td>PLC (NI)</td>
<td>PLC (NI)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(“es2”N-FLEX)</th>
<th>(“er2”N-FLEX)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALO “es”</td>
<td>ALO “er”</td>
</tr>
<tr>
<td>CA (G)</td>
<td>CA (A G N)</td>
</tr>
<tr>
<td>CL (S-ES S-S/ES)</td>
<td>CL (P-ER)</td>
</tr>
<tr>
<td>NU (SG)</td>
<td>NU (PL)</td>
</tr>
<tr>
<td>PLC (NI))</td>
<td>PLC (NI)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(INFLECT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NST:Mann</td>
</tr>
<tr>
<td>ALO “Männer”</td>
</tr>
<tr>
<td>CAN “Mann”</td>
</tr>
<tr>
<td>CAT NST</td>
</tr>
<tr>
<td>CL (P-ER)</td>
</tr>
<tr>
<td>DR (NP RD)</td>
</tr>
<tr>
<td>KN CNT</td>
</tr>
<tr>
<td>NU PL</td>
</tr>
<tr>
<td>OR LC</td>
</tr>
<tr>
<td>SL-ALO “men”</td>
</tr>
<tr>
<td>SL-CAN “man”</td>
</tr>
<tr>
<td>SL-CAT “man”</td>
</tr>
<tr>
<td>SX (M)</td>
</tr>
<tr>
<td>TL-ALO “Männer”</td>
</tr>
<tr>
<td>TL-CAN “Mann”</td>
</tr>
<tr>
<td>TL-CAT NST</td>
</tr>
<tr>
<td>TYN (HUM)</td>
</tr>
<tr>
<td>WORD# 17</td>
</tr>
</tbody>
</table>
“you gave it to me”

“tu” “as donné” “le” “me”

“tu” “me” “le” “as donné”

“tu me l’as donné”

“give it to me”

“donne” “le” “moi”

“donne” “le” “moi”

“donne-le-moi”
Translation Process & Moduls

60%-70%
Analysis Skeleton

15%-20%
Transfer Skeleton
Transfer Skeleton
Transfer Skeleton
Transfer Skeleton
Transfer Skeleton
Transfer Skeleton
Transfer Skeleton
Transfer Skeleton
Transfer Skeleton
Transfer Skeleton

15%-20%
Generation Skeleton

Interface

Analysis Skeleton

Interface

Analysis Skeleton

Interface

Analysis Skeleton

Generation Skeleton

Generation Skeleton

Generation Skeleton

Generation Skeleton

Generation Skeleton

Generation Skeleton

Generation Skeleton

Generation Skeleton

Generation Skeleton

Generation Skeleton
Possible Statistical Enhancements

- SMT as automatic post-editor of RBMT output
  - Slightly better results

- Multi-Engine Approach
  - If the analysis fails?

- Stochastic CFPS grammar

- Probabilistic transfer lexicon

- Bilingual terminology extraction
Discussion

- Where are good places for statistical approaches to improve the rule-based system?