Architecture of the Lucy Translation System

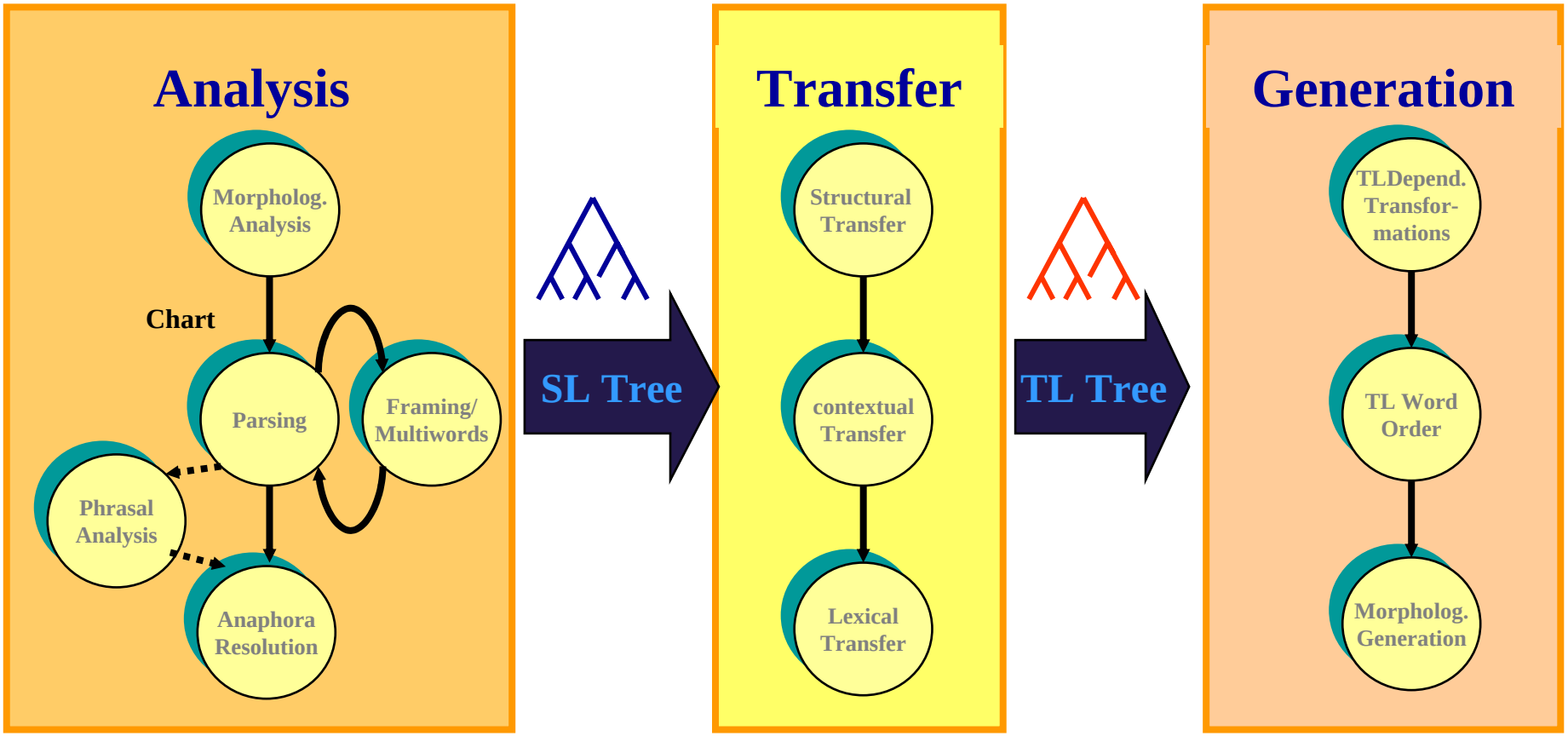
Dr. Petra Gieselmann

Lucy Software and Services GmbH, Munich

Contents

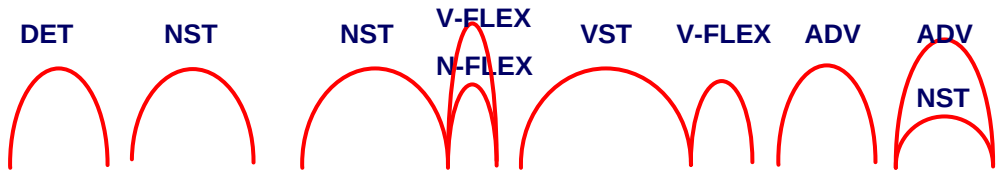
- **Rule-Based Translation Process**
- **Statistical Enhancements**
- **Discussion**

Translation Method



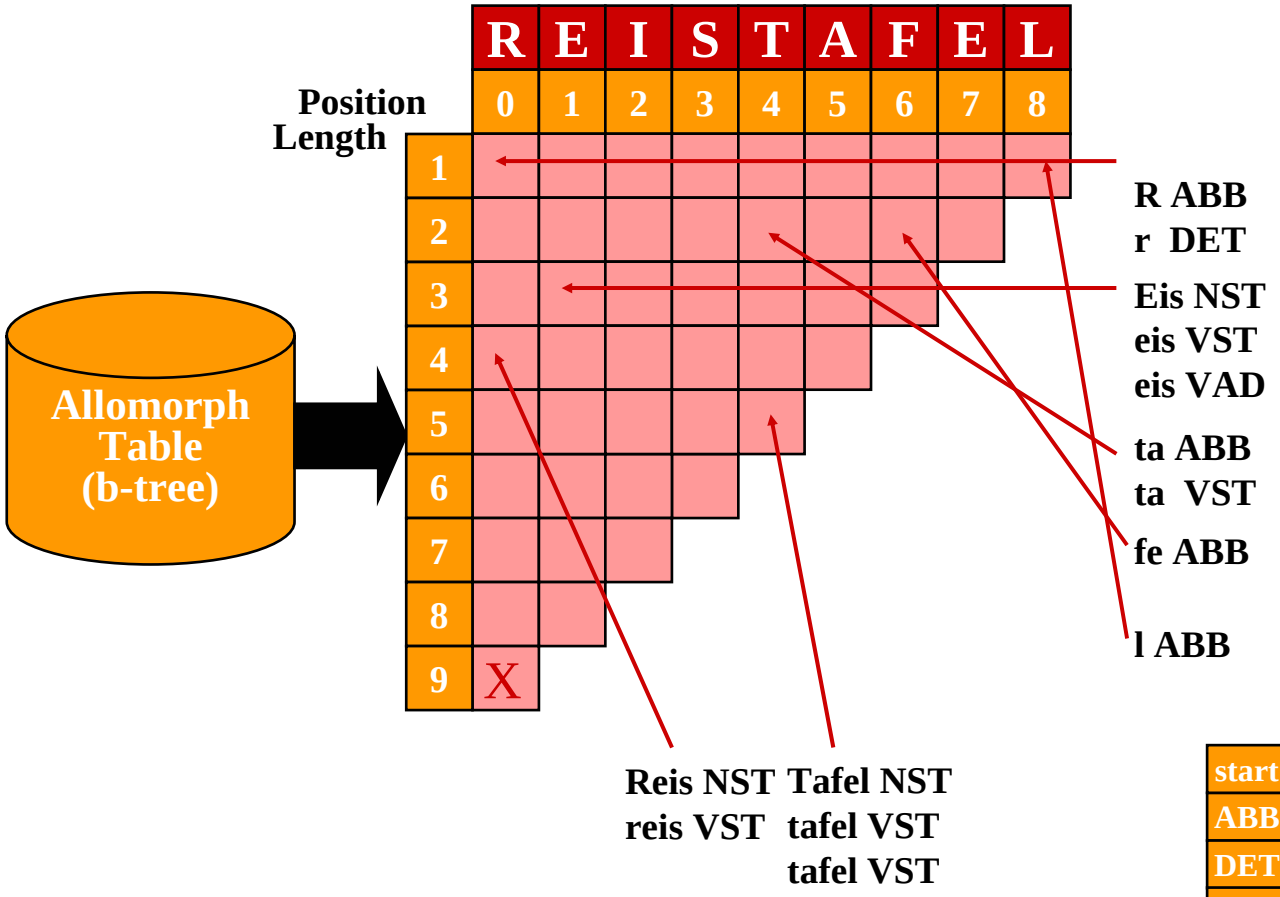
Analysis: Morphology

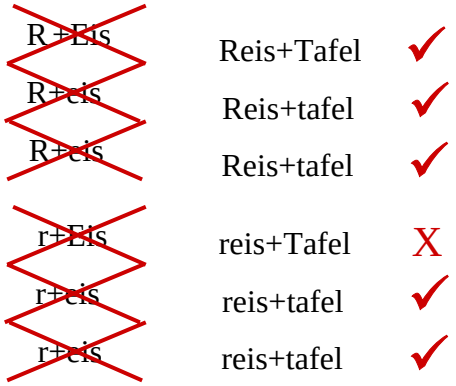
Input Sentence



“The ballet dancers performed very well” ← **Morphology**

Morphological Analysis

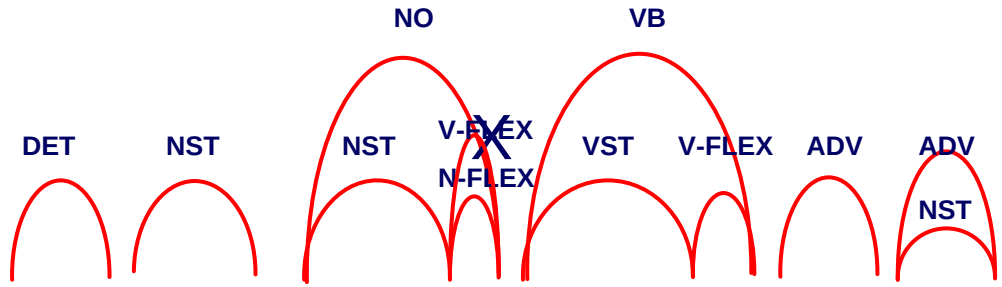




Transition Matrix

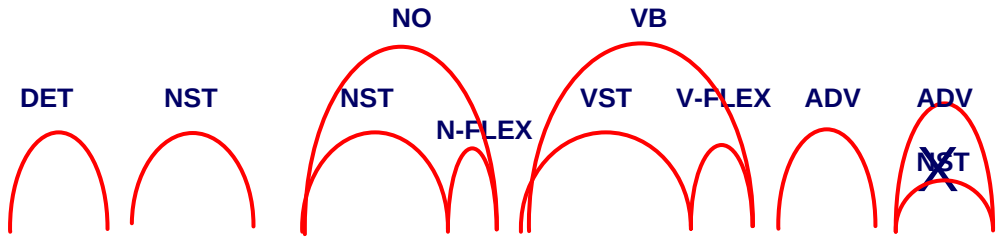
	start	ABB	DET	NST	VAD	VST	end
start	✓	✓	✓	✓	✓		
ABB						✓	
DET						✓	
NST			✓		✓	✓	
VAD					✓	✓	
VST					✓	✓	
end							

Analysis: Word Formation



“The ballet dancers performed very well” ← **Word Formation Rules**

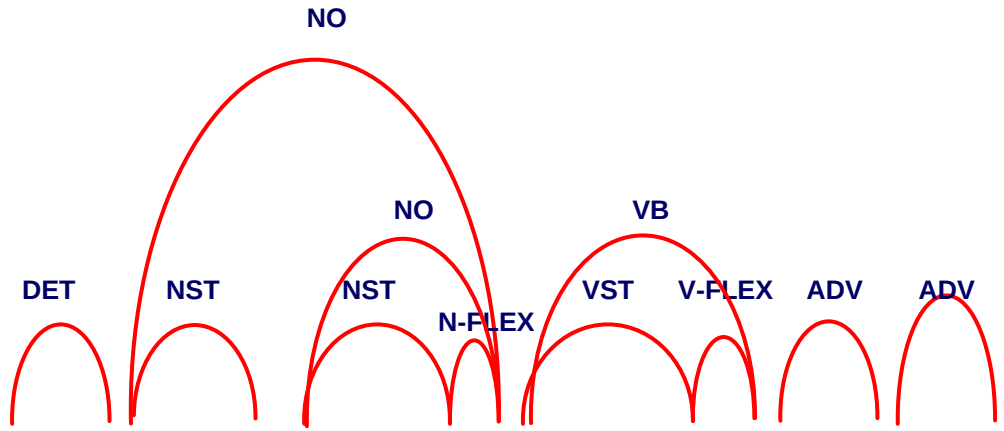
Analysis: Homography



“The ballet dancers performed very well”

Homography (Lexical Ambiguity)

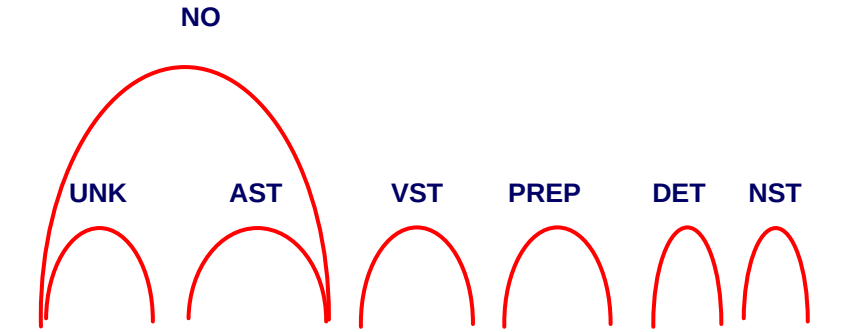
Analysis: Compounding



“The ballet dancers performed very well”

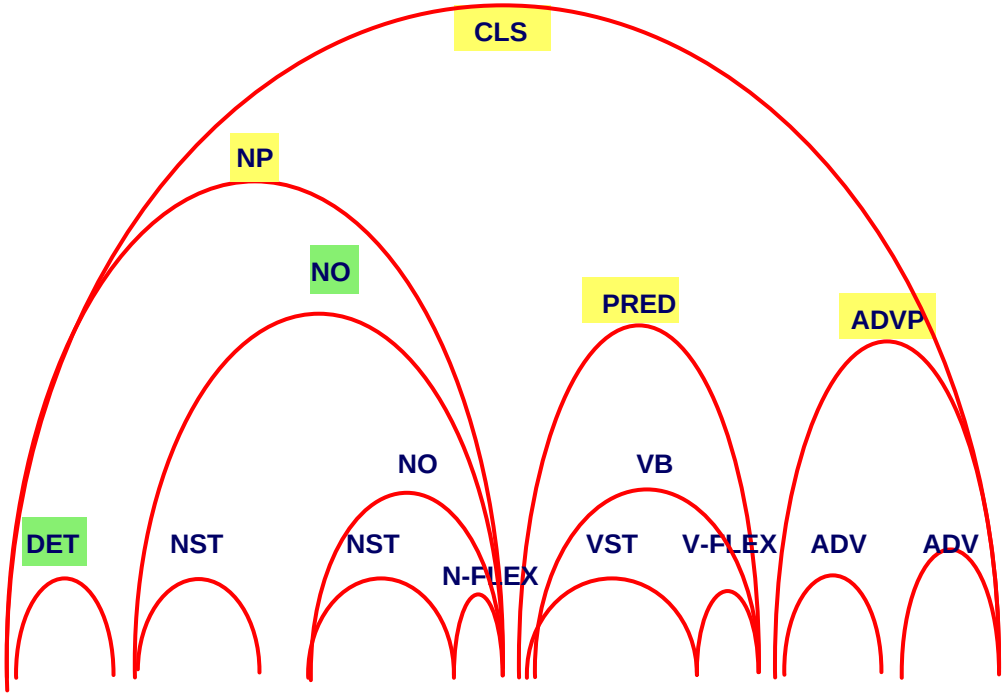
Compounding & Proper Names

Analysis: Proper Names

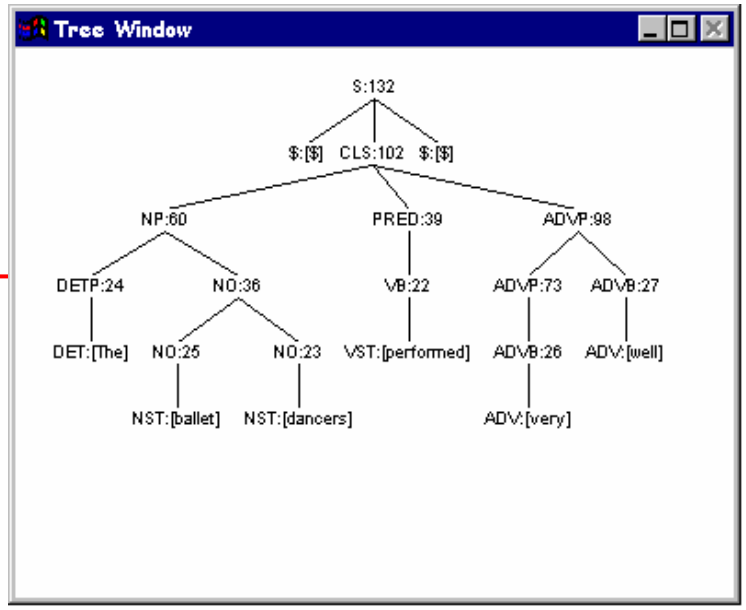


Compounding & Proper Names

Analysis: Building up Phrases



“The ballet dancers performed very well”



Samples:

NP --> DET NO
NP --> DET AP NO
CLS --> NP PRED NP PP
CLS --> NP PRED ADVP

Analysis: Grammar Rules

Grammar Rule → **NP --> DET(1) NO (2)**

- **Tests** → **(check-agreement 1 2)**
- **Constructors** → **Feature Traffic**

→ **Context-free Grammar enhanced with Tests and Transformations**

Example of a Rule

Rule

Tests

Constructors

```
ID    VB-NO-VB-1

VB    NO    VB

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

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-categorial Phenomena

- **Coordination**
- **Comparatives**
- **Complementizers**
- **Negation**
- **Commas**
- **Orthography**
- **Quotes**

Analysis: Lexicons

Monolingual Lexicon

- **Morphological** → **Morphological Analysis**

- **Syntactic** → **Syntactic Analysis**
- **Semantic**

Monolingual Lexicon

```
CAN     "leave"
CAT     VST
TT      (I T DT)
PV      ("by")
ARGS    ((($SUBJ N1) ($DOBJ N1) OPT ($ADV LOC) ($IOBJ N1 (TYPE P1) (PREP "for")))
         (($SUBJ N1) ($DOBJ N1) ($IOBJ N1 (TYN SOC PRO POT HUM ANI) (PREP "to")))
         (($SUBJ N1 NO (ICP ING-SUBJ)) ($DOBJ N1) ($OCOMP ADJ N0  (ICP-ING)))
         (($SUBJ N1) OPT ($ADV TMP LOC)))
ALO     "leav"
CL      (G-ING I-E PR-ES1)
```

```
CAN     "leave"
CAT     VST
TT      (I T DT)
PV      ("by")
ARGS    ((($SUBJ N1) ($DOBJ N1) OPT ($ADV LOC) ($IOBJ N1 (TYPE P1) (PREP "for")))
         (($SUBJ N1) ($DOBJ N1) ($IOBJ N1 (TYN SOC PRO POT HUM ANI) (PREP "to")))
         (($SUBJ N1 NO (ICP ING-SUBJ)) ($DOBJ N1) ($OCOMP ADJ N0  (ICP-ING)))
         (($SUBJ N1) OPT ($ADV TMP LOC)))
ALO     "left"
CL      (P-0 PA-0)
```

Monolingual Package

- Canonical Information
- Morphological Information

Multiword Entries

CAN "pillow case"
CAT NST
MW-HEAD "case"
MW-TYPE NST-NST
MW-BODY ((NST "pillow" (NU SG)) (HEAD))

NO ➲ NO+NO

pillow case

NO

pillow case

00 "pillow case" NST ➲ "Kopfkissenbezug" NST

00 "Kopfkissenbezug" NST ➲ "pillow case" NST

CAN "leave of absence"
CAT NST
MW-HEAD "leave"
MW-TYPE NST-STRING
MW-BODY ((HEAD) (STRING "of absence"))

NO ➲ NO+STRING

leave of absence

NO

leave of absence

00 "leave of absence" NST ➲ "Beurlaubung"NST

00 "Beurlaubung" NST ➲ "leave of absence" NST

Analysis: Framing of Clause

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.

Translation Process: Analysis

Output of the syntactic Analysis:

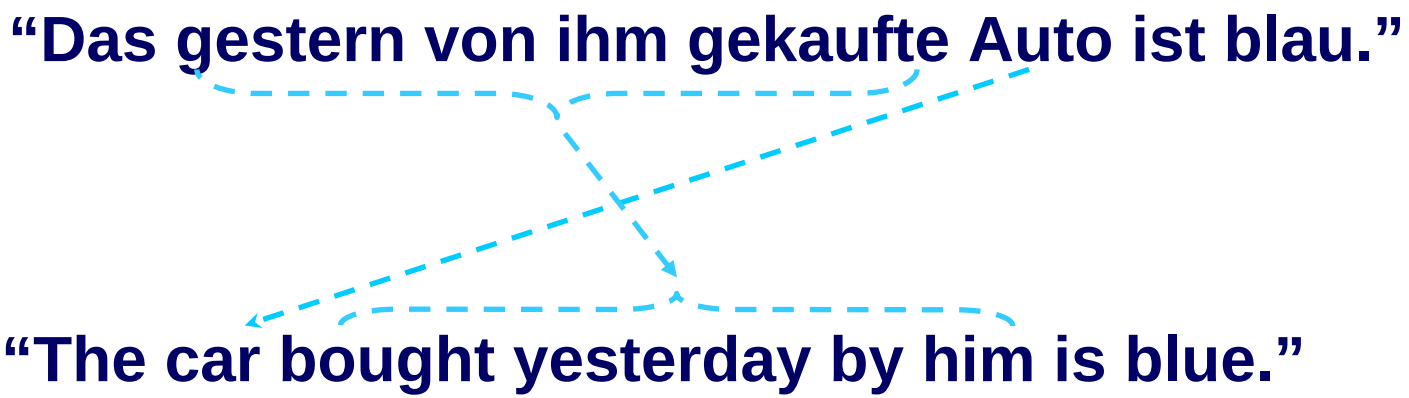
- **Success:** → **1 Interpretation**
- **Failure:** → **Phrasal Analysis**

Transfer

- **Structural Transfer**
- **Contextual Transfer**
- **Lexical Transfer**

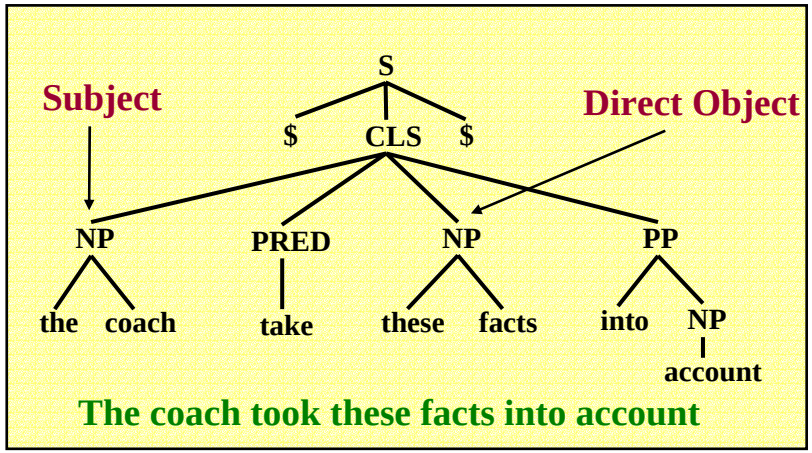
Structural Transfer

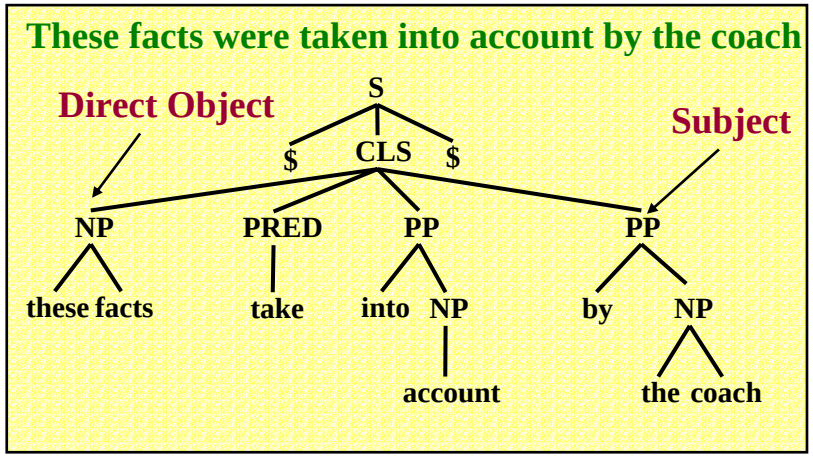
Transformation of SL Structure into TL Structure



Contextual Transfer

For all the Categories with Complements





42203 take ➔ "bemächtigen"
(XX-VST-DOBJ :CAN "hold")
(XX-VST-NONROLE :CAT "PP" :CAN "of" :TYN HUM)
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)

NST:Mann	
ALO	“men”
CAN	“Mann”
CAT	NST
CL	(P-0)
NU	PL
OR	LC
SL-CAN	“man”
SL-CAT	NST
TL-CAN	“Mann”
TL-CAT	NST
WORD#	17

Generation

- **Morphological Generation**
- **Target Language dependent Operations**

Morphological Generation

("Mann" NST
ALO "Mann"
CL (S-ES)
DR (NP RD)
GD M
KN CNT
SX (M)
TYN (HUM))

("Mann" NST
ALO "Männ"
CL (P-ER)
DR (NP RD)
GD M
KN CNT
SX (M)
TYN (HUM))

("es2" N-FLEX
ALO "es"
CA (A N)
CL (S-3)
NU (SG)
PLC (NI))

("es2" N-FLEX
ALO "es"
CA (G)
CL (S-ES S-S/ES)
NU (SG)
PLC (NI))

("er2" N-FLEX
ALO "er"
CA (G)
CL (P-E1)
NU (PL)
PLC (NI))

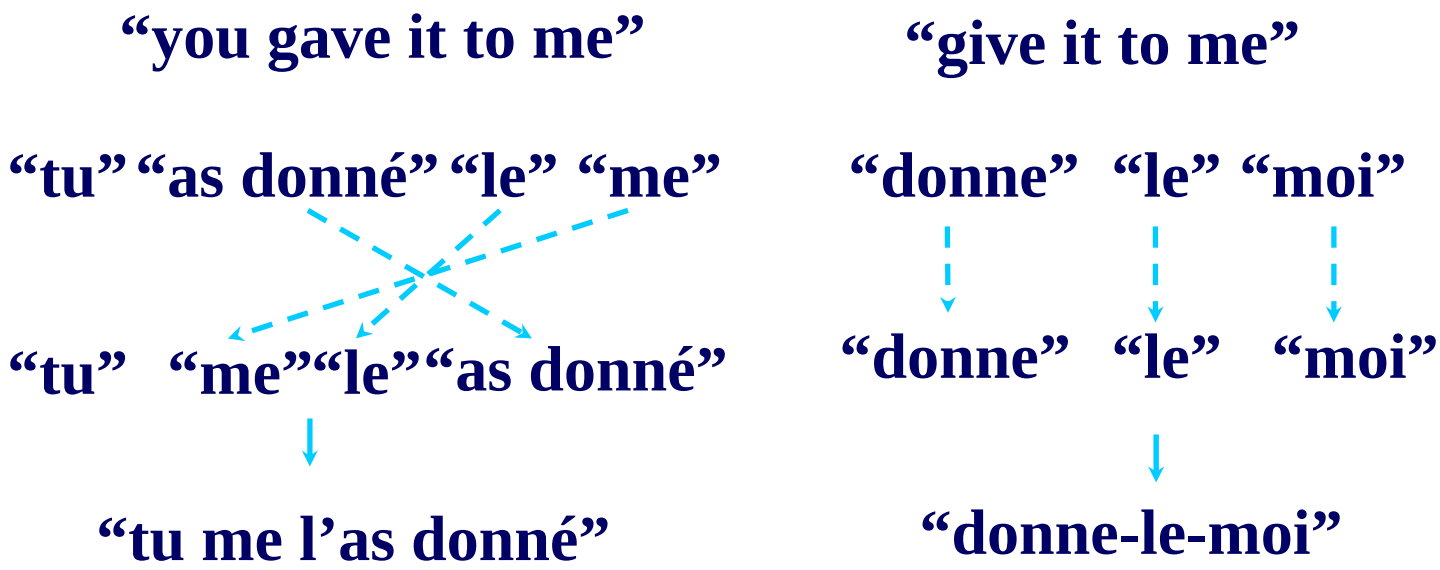
("er2" N-FLEX
ALO "er"
CA (A G N)
CL (P-ER)
NU (PL)
PLC (NI))

("er2" N-FLEX
ALO "er"
CA (N)
CL (S-1)
NU (SG)
PLC (NI))

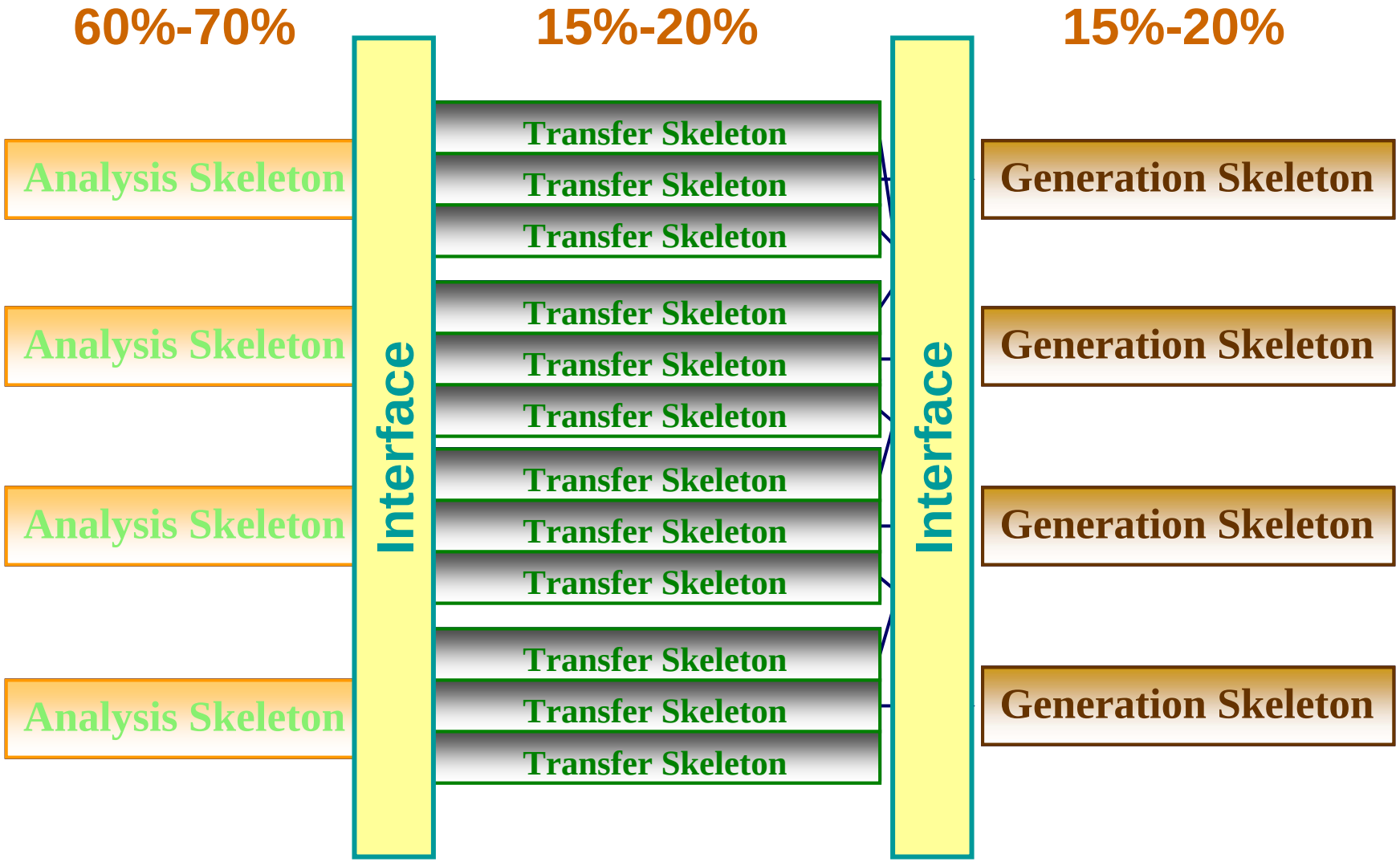
(INFLECT)

NST:Mann	
ALO	"Männer"
CAN	"Mann"
CAT	NST
CL	(P-ER)
DR	(NP RD)
KN	CNT
NU	PL
OR	LC
SL-ALO	"men"
SL-CAN	"man"
SL-CAT	"man"
SX	(M)
TL-ALO	"Männer"
TL-CAN	"Mann"
TL-CAT	NST
TYN	(HUM)
WORD#	17

Generation


“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



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?**