Improving Statistical Machine Translation by Means of Transfer Rules

Nurit Melnik

Hebrew to English Machine Translation

Language Technologies Institute

Carnegie Mellon University

Headed by
Alon Lavie

Computational Linguistics Group

University of Haifa

Headed by
Shuly Wintner

With Danny Shacham (Haifa U.) and Erik Peterson (CMU)

Hebrew-specific challenges for MT

- High lexical & morphological ambiguity
- Limited electronic linguistic resources
- Lack of comprehensive electronic open-source bilingual dictionaries

- Consequently:

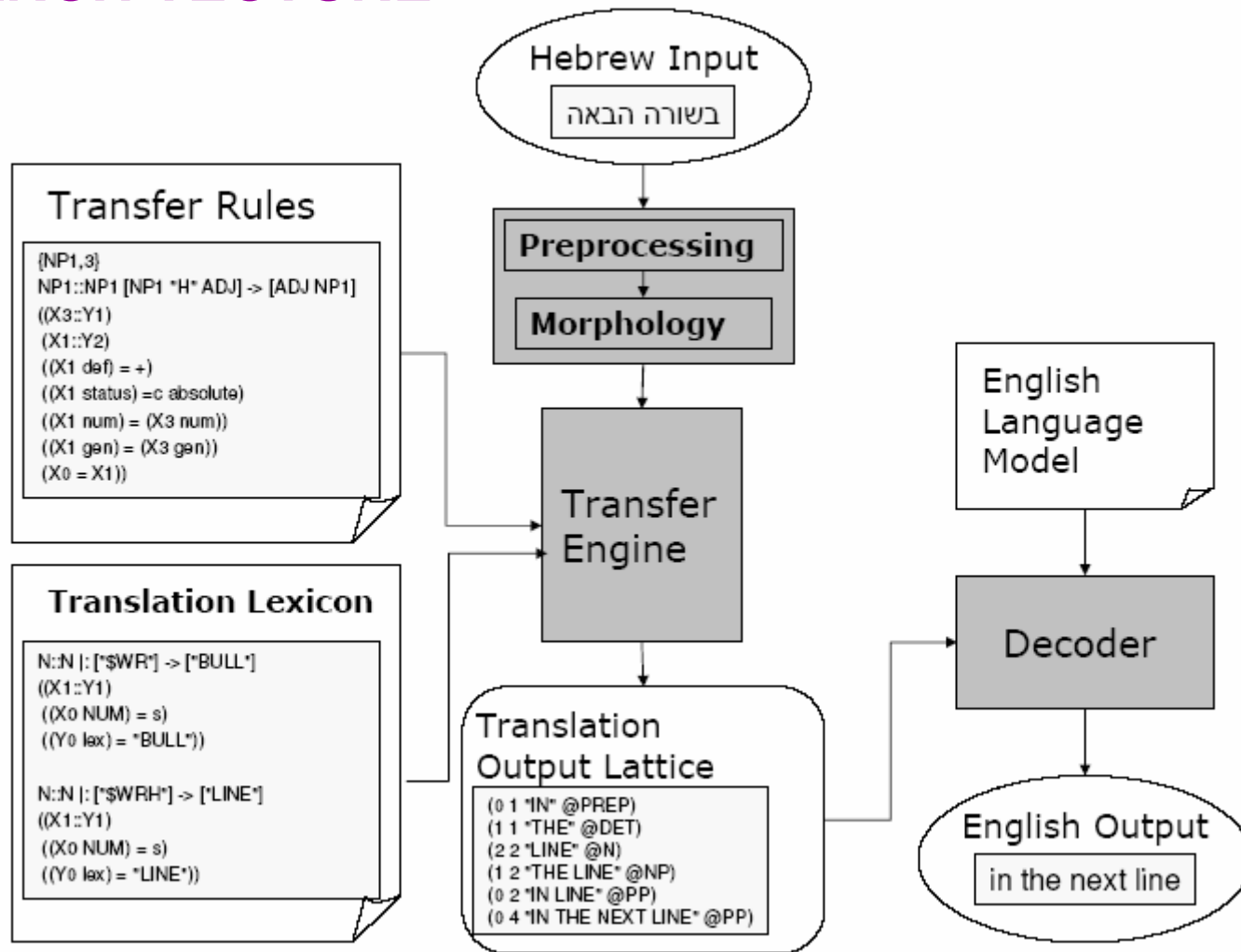 **State of the art technologies are not applicable to Hebrew.**

The AVENUE Project

Language Technologies Institute, CMU

- The goal
  - The design and rapid development of new MT methods for languages for which only limited resources are available
- Projects
  - Aymara (Bolivia)
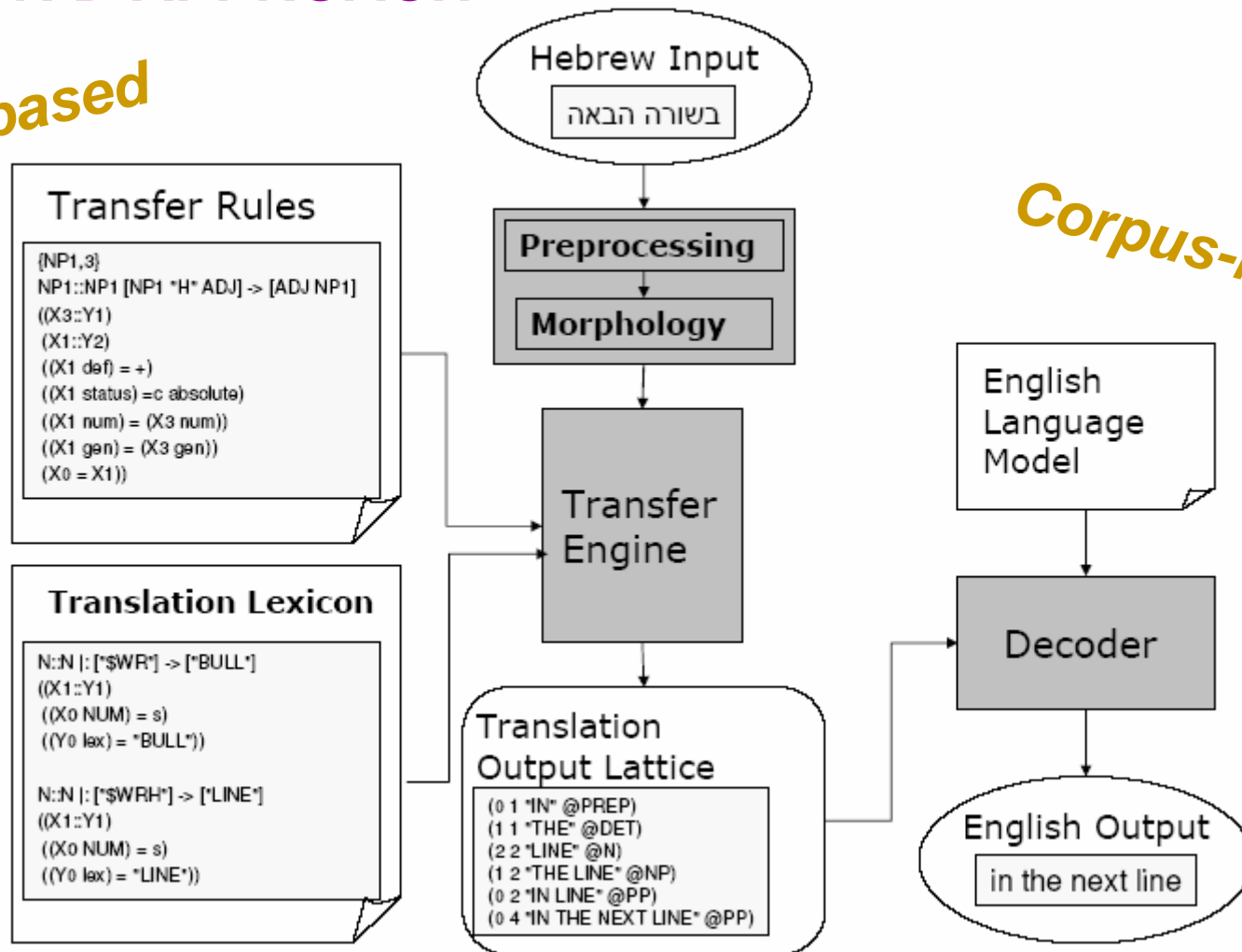 - Quechua (Peru)
  - Mapudungun (Chile)

THE ARCHITECTURE

Lavie, Peterson, Probst, Wintner and Eytani. 2004. **Rapid Prototyping of a Transfer-based Hebrew-to-English Machine Translation System**. *Proceedings of The 10th International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 1-10, Baltimore, MD, October 2004.

A HYBRID APPROACH

Rule-based

Corpus-based

Lavie, Peterson, Probst, Wintner and Eytani. 2004. **Rapid Prototyping of a Transfer-based Hebrew-to-English Machine Translation System**. *Proceedings of The 10th International Conference on Theoretical and Methodological Issues in Machine Translation*, pages 1-10, Baltimore, MD, October 2004.
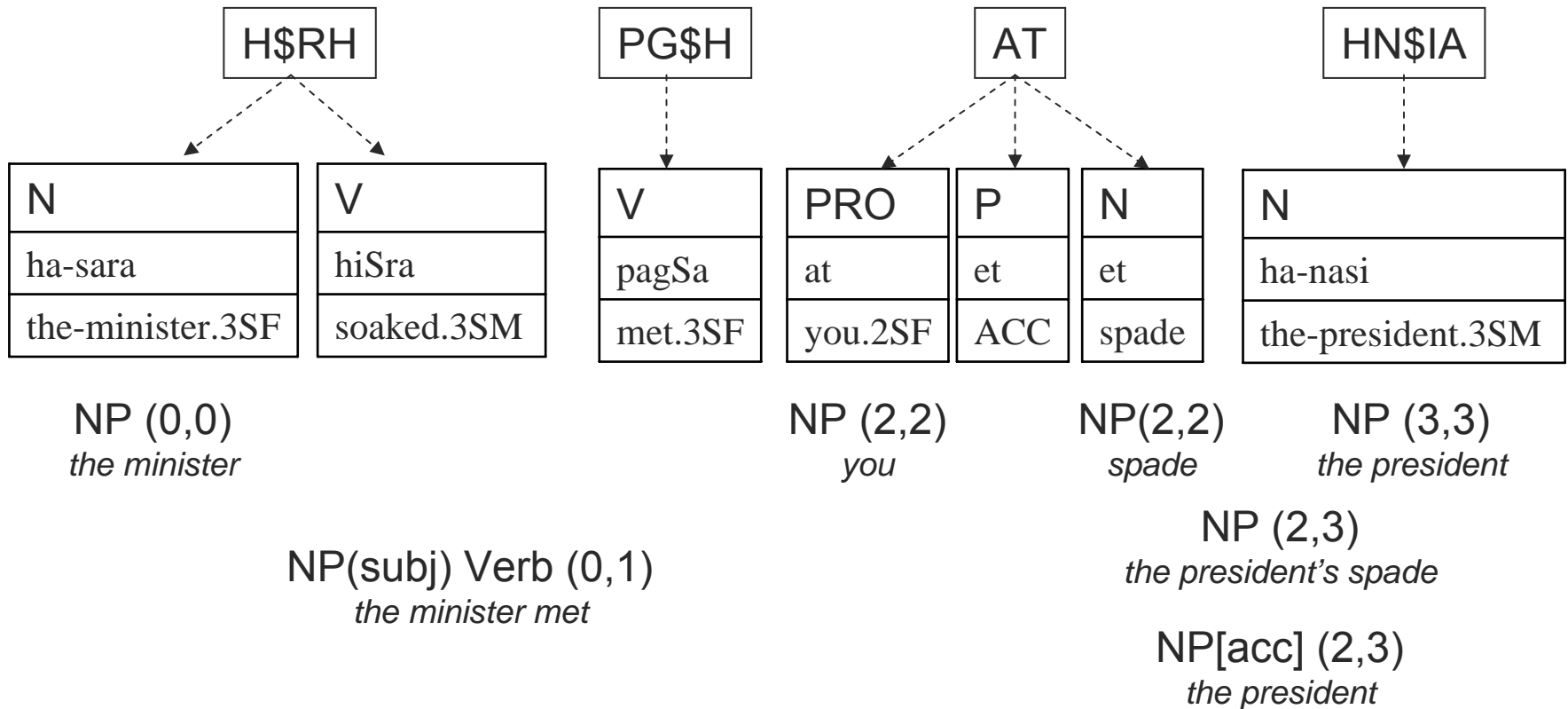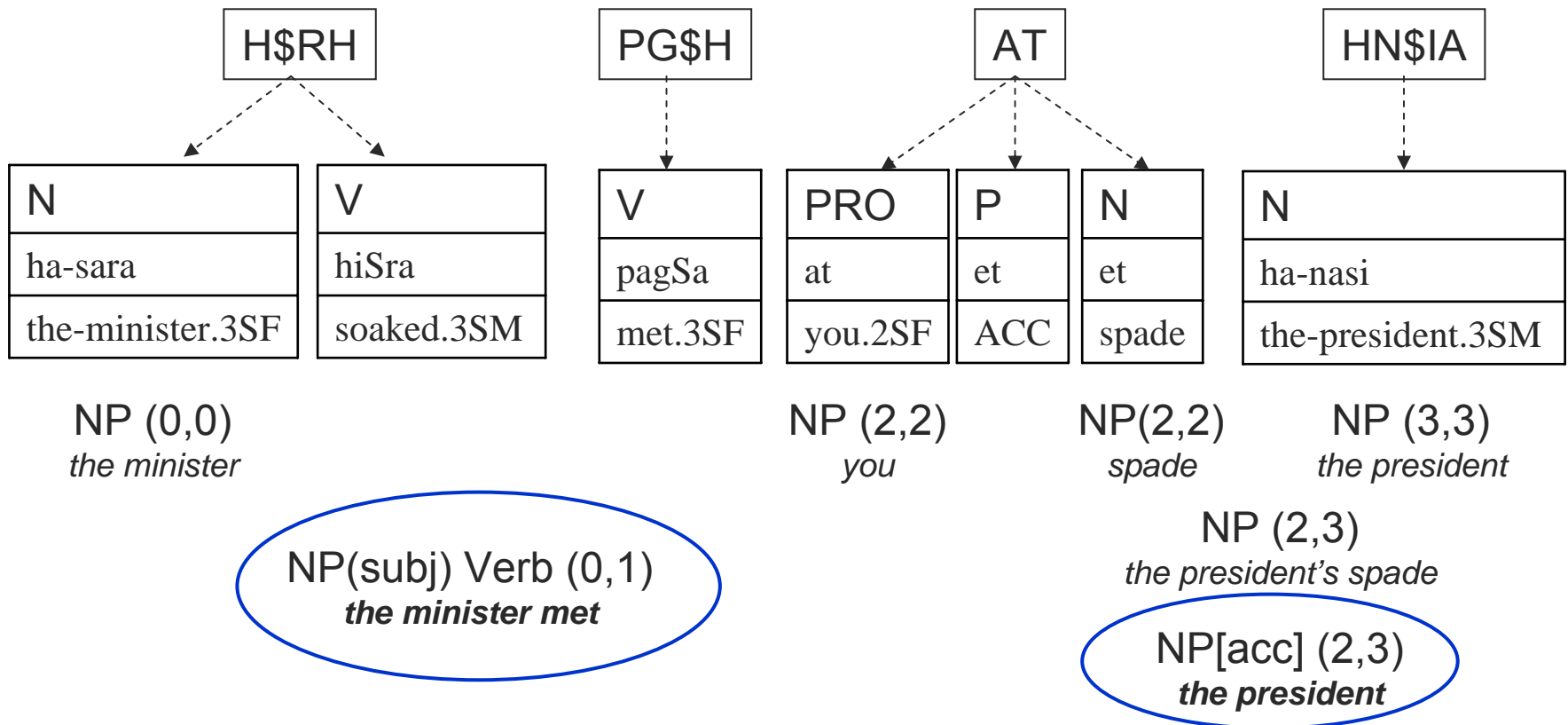
Syntactic Transfer Rules

- Transfer rules embody the 3 stages of translation
  - Analysis of source language
  - Transfer
  - Generation of target language
- Currently: 33 transfer rules
  (The original version written by Alon Lavie)

The Lattice

H$RH		PG$H	AT			HN$IA

N	V
ha-sara	hiSra
the-minister.3SF	soaked.3SM

NP (0,0)
the minister

V
pagSa
met.3SF

PRO	P	N
at	et	et
you.2SF	ACC	spade

NP (2,2)
you

NP(2,2)
spade

N
ha-nasi
the-president.3SM

NP (3,3)
the president

NP(subj) Verb (0,1)
the minister met

NP (2,3)
the president's spade

NP[acc] (2,3)
the president

The Decoder

The decoder uses the statistical Language Model of English to pick the most likely translation.

H$RH	

N	V
ha-sara	hiSra
the-minister.3SF	soaked.3SM

NP (0,0)
the minister

PG$H

V
pagSa
met.3SF

AT		

PRO	P	N
at	et	et
you.2SF	ACC	spade

NP (2,2)
you

NP(2,2)
spade

HN$IA

N
ha-nasi
the-president.3SM

NP (3,3)
the president

NP (2,3)
the president's spade

NP(subj) Verb (0,1)
the minister met

NP[acc] (2,3)
the president

Some Syntactic Challenges for Hebrew-English MT

- ## The structure of Noun Phrases

- ## Subject-Verb inversion

nistaymu	ha-bxirot
ended	the-elections

➡️ The elections ended.

- ## Pro-drop

hitsbati	ba-bxirot
voted.**1S**	in-the-elections

➡️ **I** voted in the elections.

- ## Argument Structure (valency)

Dani	himlits	**al**	ha-seret
Danny	recommended	**on**	the-movie

➡️ Danny recommended the movie.

Some Syntactic Challenges for Hebrew-English MT

- ## Possessor Dative Construction

| hitkalkela **la-nu** ha-mexonit | → | **Our** car broke down. |
| broke-down **to-us** the-car | | |

- ## Anaphor resolution

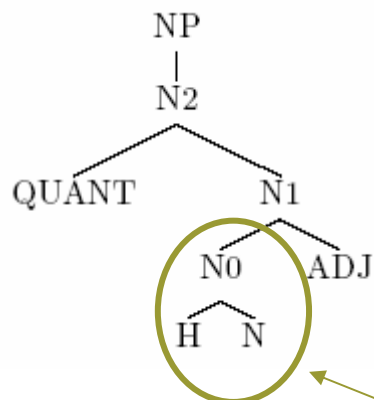| ha-memSala arxa et yeSivata ha-riSona | → | The government held |
| the-government held ACC **her**-meeting the-first | | **its** first meeting. |

Hebrew-English Syntactic Transfer

- Noun Phrases
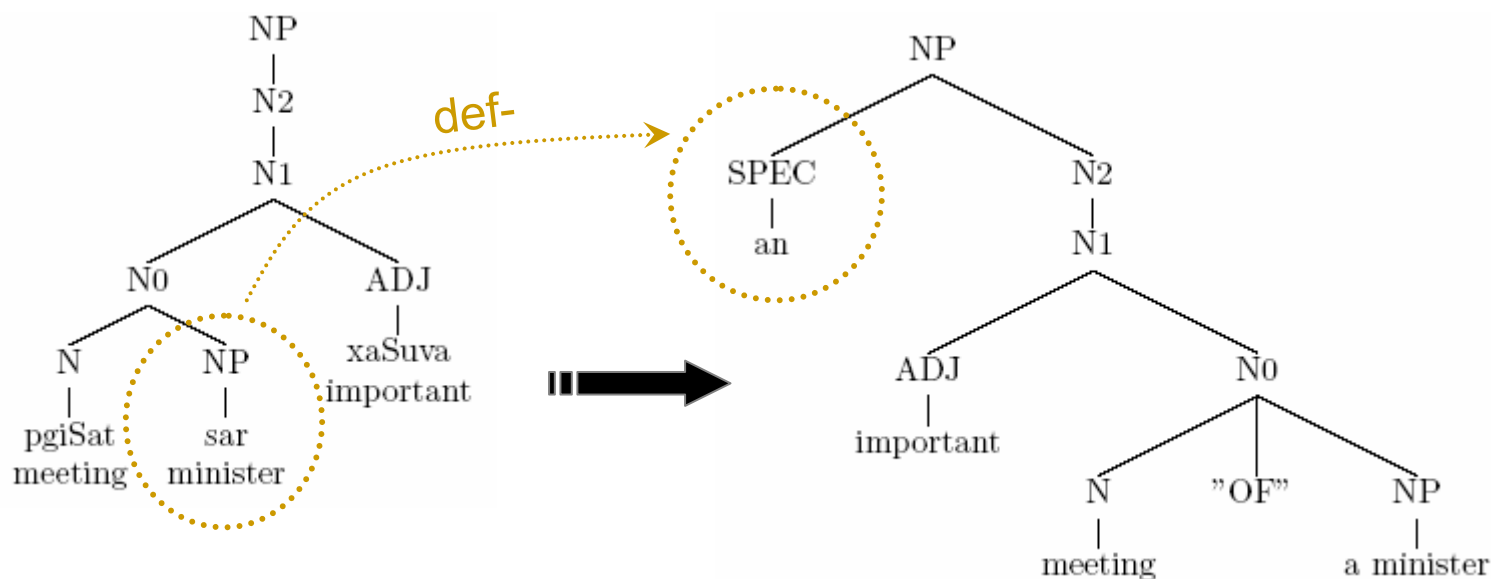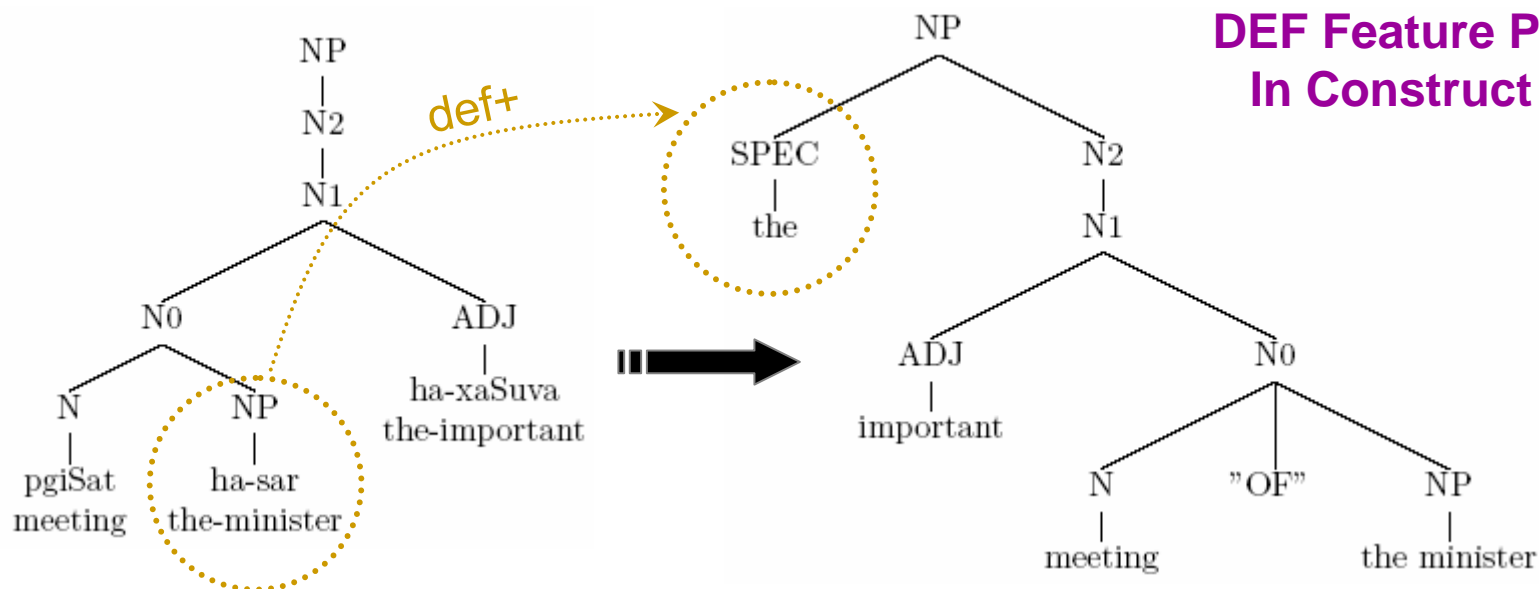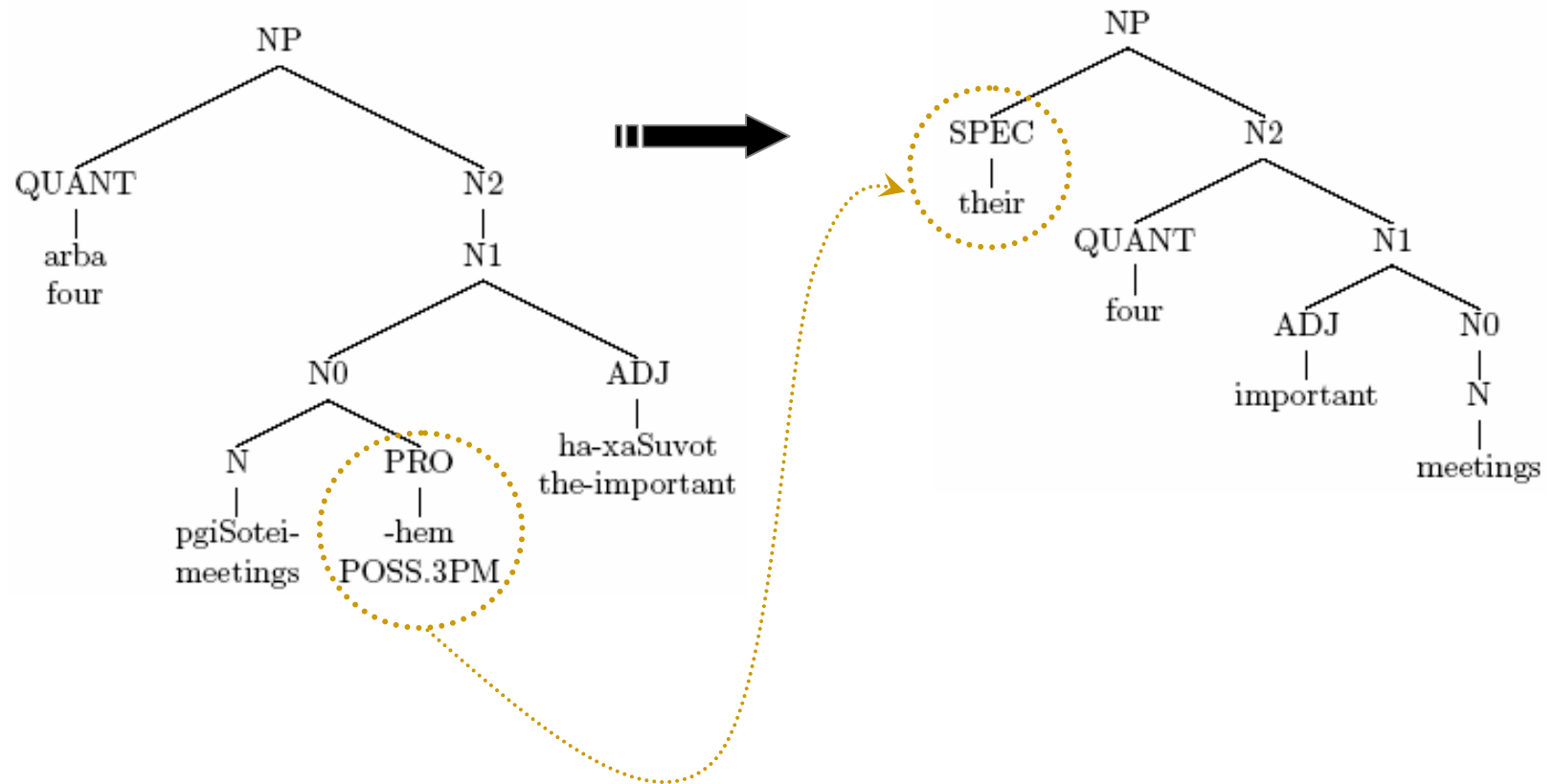- Subject-Verb inversion

Transfer Rules for NPs

Hebrew NP

English NP

syntactic specifiers (only English)

the morphological level (only Hebrew)

def+

NP — N2 — N1

N0 — ADJ (ha-xaSuva / the-important)

N (pgiSat / meeting) — NP (ha-sar / the-minister)

⟹

NP — SPEC (the) — N2 — N1

ADJ (important) — N0

N (meeting) — "OF" — NP (the minister)

def-

NP — N2 — N1

N0 — ADJ (xaSuva / important)

N (pgiSat / meeting) — NP (sar / minister)

⟹

NP — SPEC (an) — N2 — N1

ADJ (important) — N0

N (meeting) — "OF" — NP (a minister)

Possessor Feature Structure Percolation

Input

פגישתם

PGI$TM

pgiSat-am

meeting.3SF-POSS.3PM

Output

THEIR MEETING

Morph. Analysis

```
(
 ( SPANSTART     0 )
 ( SPANEND      1 )
 ( SCORE       1 )
 ( LEX      PGI$H )
 ( POS       N )
 ( GEN     feminine )
 ( NUM      singular )
 ( STATUS      absolute )
)

(
 ( SPANSTART     1 )
 ( SPANEND      2 )
 ( SCORE       1 )
 ( LEX     *PRO* )
 ( POS      PRO )
 ( TRANS     *PRO* )
 ( GEN     masculine )
 ( NUM      plural )
 ( PER      3 )
 ( CASE      possessive )
)
```

Transfer Rules

```
{NP0,2}
NP0::NP0 [N PRO] -> [N]
(
(X1::Y1)
((X2 case) = possessive)
((X0 possessor) = X2)
((X0 def) = +)
((Y1 num) = (X1 num))
(X0 = X1)
(Y0 = X0)
)
```

```
{NP,3}
NP::NP [NP2] -> [PRO NP2]
(
(X1::Y2)
((X1 possessor) =c *DEFINED*)
((Y1 case) = (X1 possessor case))
((Y1 per) = (X1 possessor person))
((Y1 num) = (X1 possessor num))
((Y1 gen) = (X1 possessor gen))
(X0 = X1)
(Y0 = Y2)
)
```

Noun Phrases – Construct State

החלטת הנשיא הראשון

HXL@T   [HNSIA   HRA$WN]
decision.3SF-CS the-president.3SM the-first.3SM

THE DECISION OF THE FIRST PRESIDENT

החלטת הנשיא הראשונה

[HXL@T   HNSIA]   HRA$WNH
decision.3SF-CS the-president.3SM the-first.3SF

THE FIRST DECISION OF THE PRESIDENT

CLG

Noun Phrases - Possessives

הנשיא הכריז שהמשימה הראשונה שלו תהיה למצוא פתרון לסכסוך באזורנו

HNSIA	HKRIZ	HMIMH	HRA$WNH	$LW	THIH
the-president	announced	that-the-task.3SF	the-first.3SF	of-him	will.3SF

LMCWA	PTRWN	LSKSWK	BAZWRNW
to-find	solution	to-the-conflict	in-region-POSS.1P

Without transfer grammar:
```
THE PRESIDENT ANNOUNCED THAT THE TASK THE BEST OF HIM
WILL BE TO FIND SOLUTION TO THE CONFLICT IN REGION OUR
```

With transfer grammar:
```
THE PRESIDENT ANNOUNCED THAT HIS FIRST TASK WILL BE
TO FIND A SOLUTION TO THE CONFLICT IN OUR REGION
```

Subject-Verb Inversion

אתמול הודיעה הממשלה שתערכנה בחירות בחודש הבא

ATMWL   HWDI&H   HMM$LH
yesterday announced.3SF   the-government.3SF

$T&RKNH   BXIRWT   BXWD$ HBA
that-will-be-held.3PF   elections.3PF   in-the-month   the-next

Without transfer grammar:
YESTERDAY ANNOUNCED THE GOVERNMENT THAT WILL RESPECT
OF THE FREEDOM OF THE MONTH THE NEXT

With transfer grammar:
YESTERDAY **THE GOVERNMENT ANNOUNCED** THAT ELECTIONS
WILL ASSUME IN THE NEXT MONTH

Subject-Verb Inversion

<div dir="rtl">

לפני כמה שבועות הודיעה הנהלת המלון שהמלון יסגר בסוף השנה

</div>

LPNI	KMH	$BW&WT	HWDI&H	HNHLT	HMLWN
before	several	weeks	announced.3SF	management.3SF.CS	the-hotel

$HMLWN	ISGR	BSWF	H$NH
that-the-hotel.3SM	will-be-closed.3SM	at-end.3SM.CS	the-year

Without transfer grammar:
```
IN FRONT OF A FEW WEEKS ANNOUNCED ADMINISTRATION THE
HOTEL THAT THE HOTEL WILL CLOSE AT THE END THIS YEAR
```

With transfer grammar:
```
SEVERAL WEEKS AGO THE MANAGEMENT OF THE HOTEL ANNOUNCED
THAT THE HOTEL WILL CLOSE AT THE END OF THE YEAR
```

CLG

Qualitative Evaluation

- **Error Types**
  - ○ Syntactic errors
  - ○ Lexical errors
  - ○ Language Model errors

Syntactic errors

Syntactic structures that are not covered by the current grammar

- Passive
- Pro-drop
- Participles
- Negation
- Copula-less constructions
- …

Lexical Errors

Complex lexical items that are missing from the lexicon

- Multi-word phrases
  - axar kax
    after like-this
    'later'
- (Semi-)fixed expressions
  - magi'a   lo   maskoret
    reaches.3SF   to-him   salary.3SF
    'he deserves a salary'
  - ha-yeled   ben   sheva
    the-boy   son   seven
    'the boy is seven years old'

Language Model Errors

- The English Language Model is used to pick the most likely translation from a set of options in the lattice.

- "LM errors" occur when the LM does not pick the best option.

Language Model Errors

- **Wrong lexical choices**
  - ...אני רוצה את השכר
    - **Selected**: ...I want the **charter**…
    - **Better**: ... I want the **salary**…

- **Wrong syntactic choices**
  - ...שמארגנת מנהלת ההגירה
    - **Selected**: ...that **the organizer** of the management of the immigration
    - **Better**: ...that the administration of the immigration **organizes**

Conclusion

- Purely statistical MT is not possible for languages with limited resources.
- The solution: A hybrid system
  - Transfer-rule-based methods for the resource-poor source language
  - Statistical methods for the resource-rich target language