Morphology, Syntax and What's In Between

Towards Paradigmatic Syntax

Reut Tsarfaty
The Department of Linguistics and Philology
Uppsala University

Machine Translation and Morphologically Rich Languages
University of Haifa
January 23-27, 2011

Syntax, MT and MRLs

Some Brief Historical Overview

Syntax, MT and MRLs

Some Brief Historical Overview

- What Kevin said

Some Brief Historical Overview

- What Kevin said
- What Yoav said
- What all linguists say all the time

Syntax, MT and MRLs

Some Brief Historical Overview

- What Kevin said
- What Yoav said
- What all linguists say all the time

$\rightsquigarrow$ Morphology and Syntax Interact!

Syntax, MT and MRLs

Some Brief Historical Overview

- What Kevin said
- What Yoav said
- What all linguists say all the time

$\rightsquigarrow$ Morphology and Syntax Interact!

No, pre/post-processing won't work

Today

An Answer to Kevin (Sunday):

How to model a morphosyntactic tree

A Question to Jason (Thursday):

How to learn bilingual paradigms

Today

The Task:
Statistical Parsing

The Challenge:
Complex Form-Function Correspondence

The Method:
Following the footsteps of Morphology

The Proposal:
A Relational-Realizational Approach

$\Rightarrow$ A Stepping Stone
Towards computational typology

Part 1: The Task

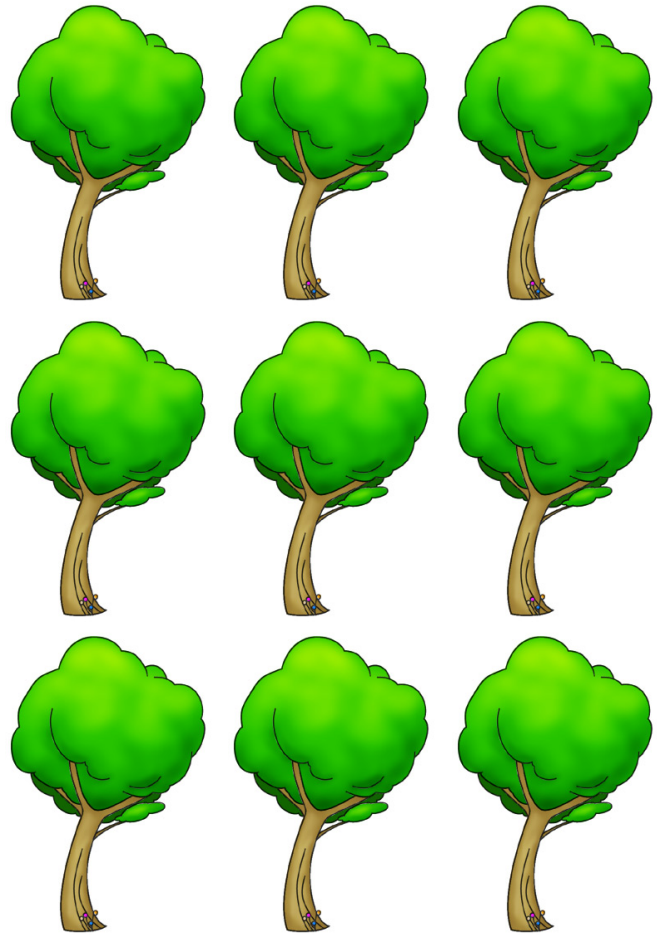
Statistical Parsing

Statistical Parsing

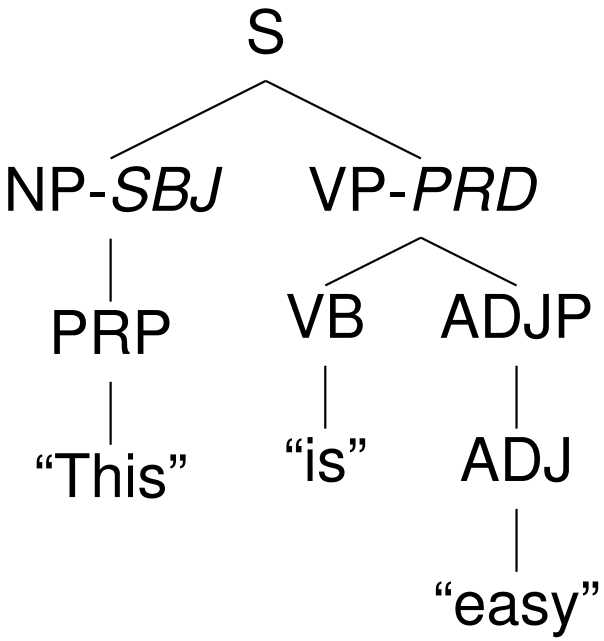
”This is easy”

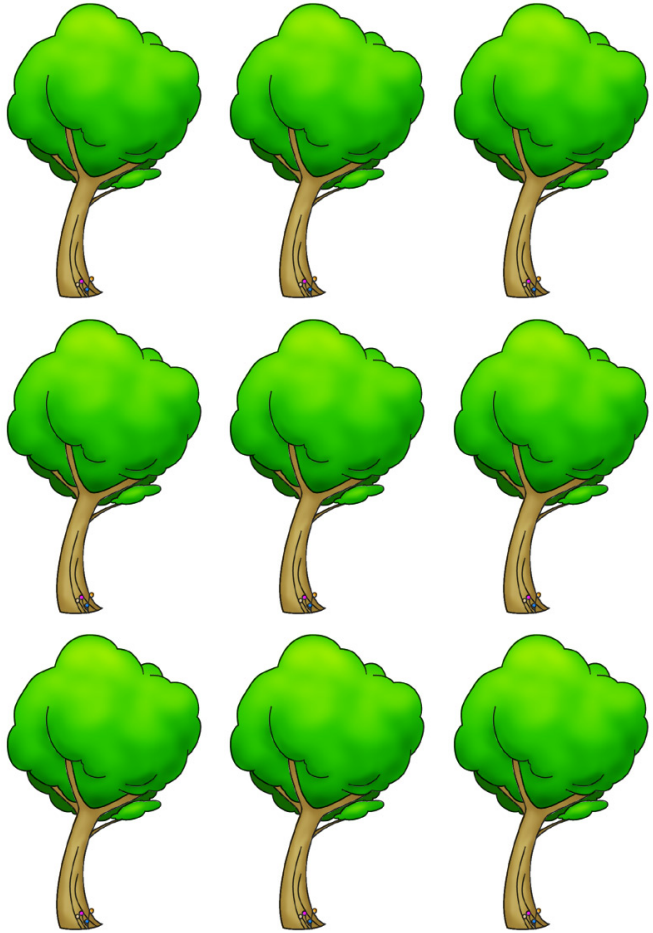
Statistical Parsing

”This is easy”

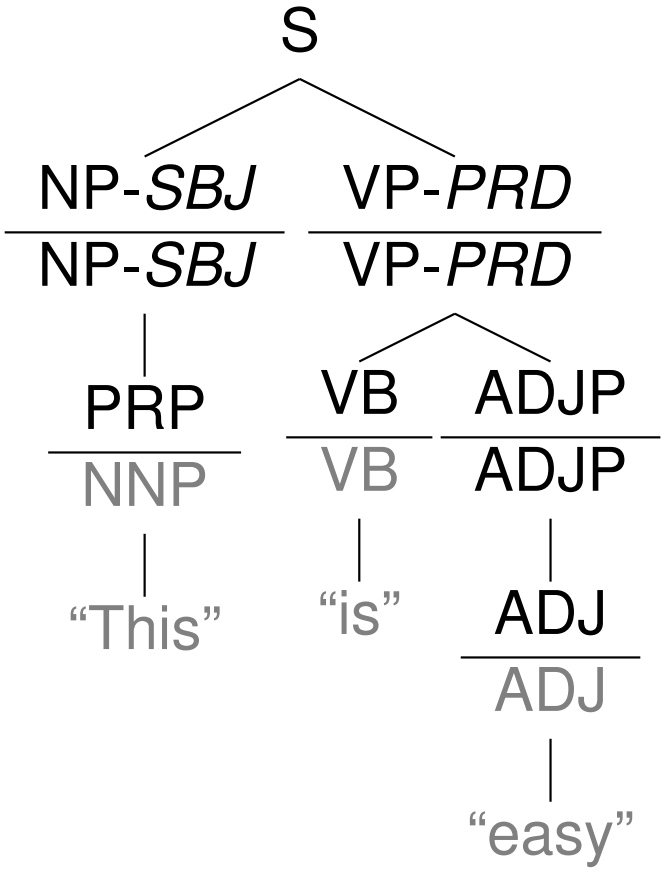


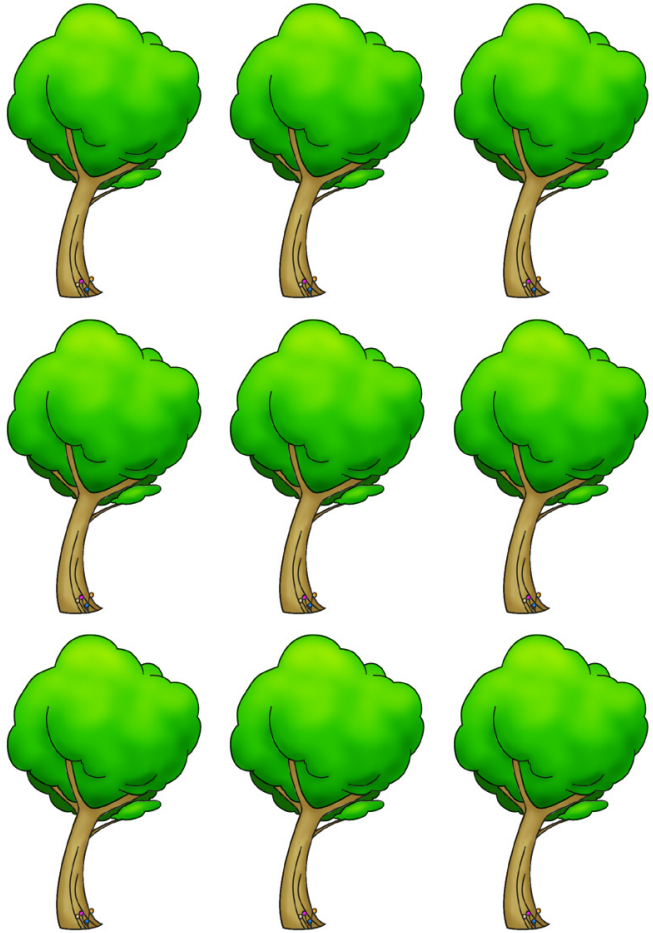
Statistical Parsing



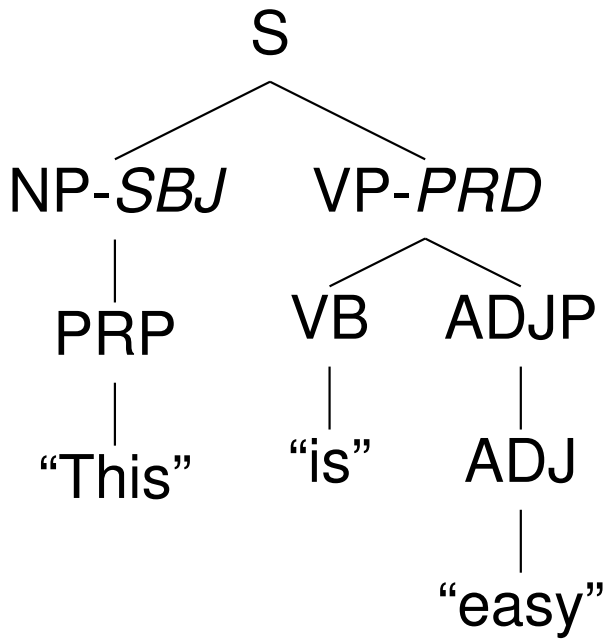


Statistical Parsing





Works Great for English



Model	Study	F-Score
Treebank Grammar	Charniak 1996	75
Head-Driven	Collins 1997	88.6
Discriminative Reranking	Collins 2000	89.7
Discriminative-Reranking	Johnson & Charniak 2005	91.0
Self-Training	McClosky 2006	92.1
State-Splits	Petrov et al 2007	90.1
Forest Reranking	Liang Huang 2008	91.7

Not As Well For Other Languages

And what about this?

將水煮開後才使用。

And this?

إغلي الماء قبل استعماله

And this?

יש להרתיח את המים לפני השימוש.

And? ...

Language	Parser	F-Score
German	Rafferty & Manning 2008	79.2
Czech	Collins et al. 1999	79.3
Chinese	Levy & Manning 2003	78.8
Arabic	Maamouri, Bies & Kulick 2008	78.1
Hebrew	Tsarfaty & Sima'an 2007	74.4

Not So Well For MRLs

Why?

Homework Assignment

Read:

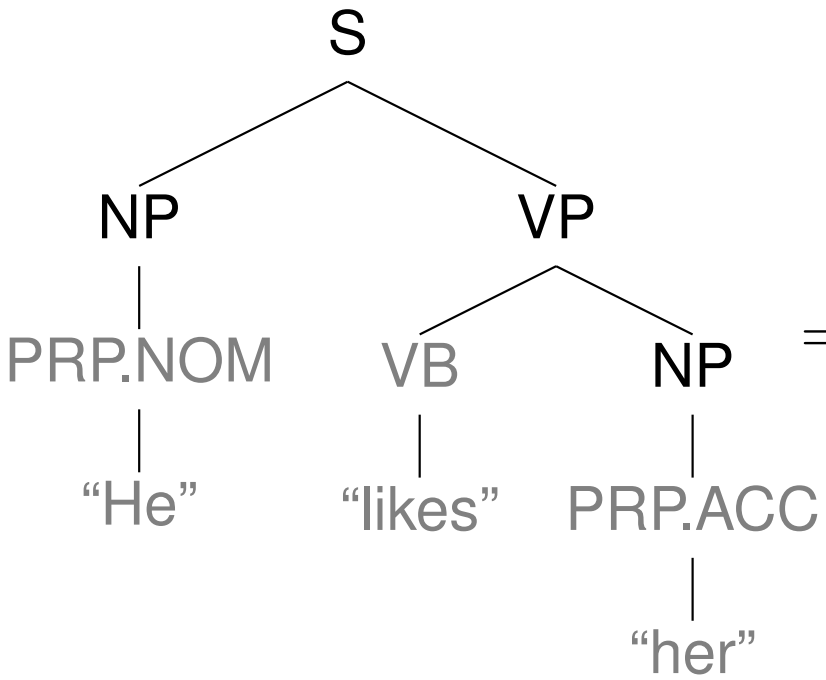
- Tsarfaty et al. "*Statistical Parsing From Morphologically Rich Languages: What, How and Whither*" Proceedings of the First Workshop on Statistical Parsing of MRLs 2010

Answer:

- What are the three main challenges in parsing of MRLs?
- How are these going to affect your MT system?

English Pronouns

English Pronouns

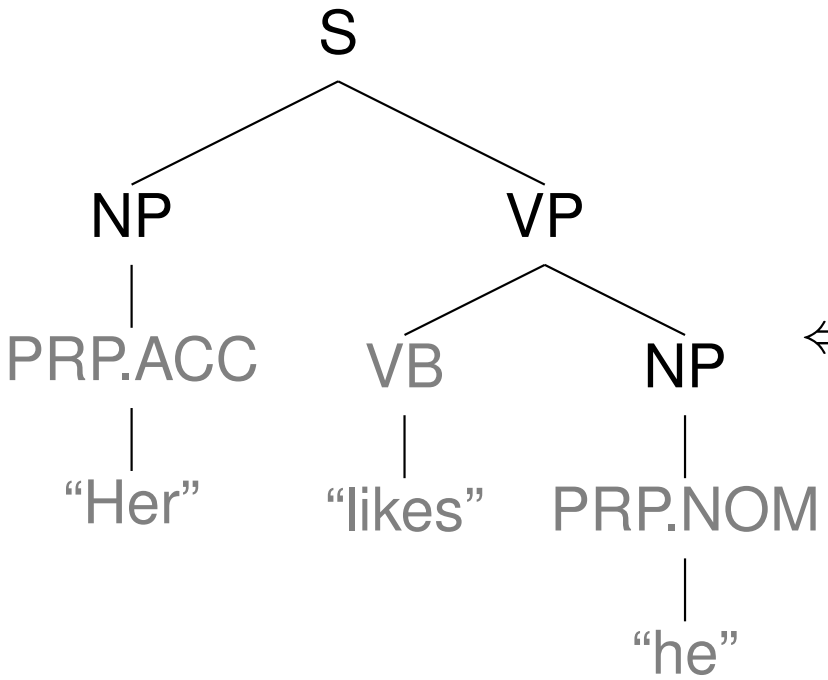


$\Rightarrow$

P(NP VP|S)	1
P(PRP.NOM|NP)	0.5
P(PRP.ACC|NP)	0.5
P(VB NP|VP)	1
P(”He”|PRP.NOM)	1
P(”likes”|VB)	1
P(”her”|PRP.ACC)	1

P(“He likes her”)$=$P(NP VP|S)$\times$...$\times$P(“her”|PRP.ACC)$=$ 0.25

English Pronouns

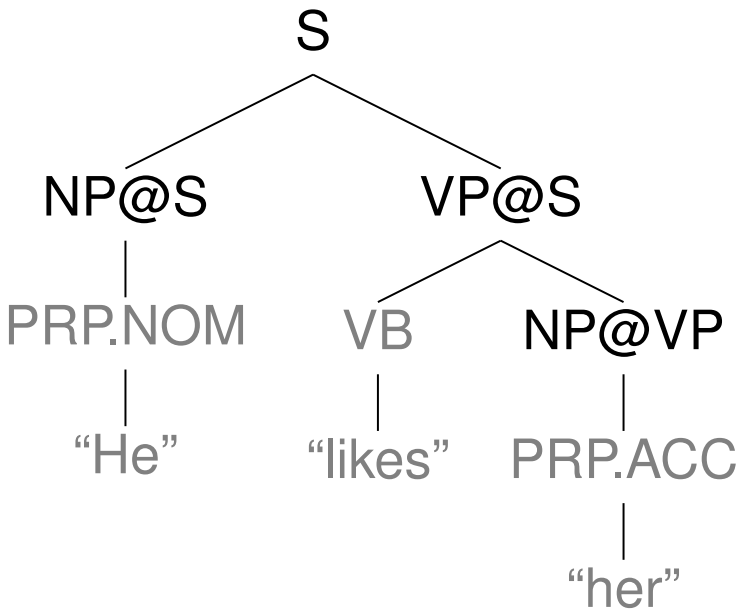


$\Leftarrow$

P(NP VP|S)	1
P(PRP.NOM|NP)	0.5
P(PRP.ACC|NP)	0.5
P(VB NP|VP)	1
P(”He”|PRP.NOM)	1
P(”likes”|VB)	1
P(”her”|PRP.ACC)	1

P(“Her likes he”)=P(NP VP|S)$\times$...$\times$P(“her”|PRP.ACC)= 0.25

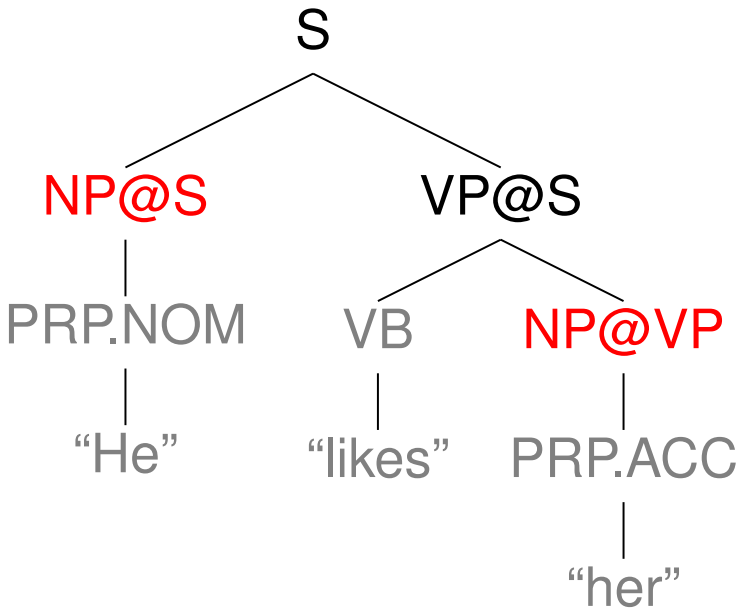
Example 1: Parent Encoding (Johnson 1998)



$\Rightarrow$

P(NP@S VP@S|S)	1
P(PRP.NOM |NP@S)	1
P(PRP.ACC |NP@VP)	1
P(VB NP@VP|VP@S)	1
P(”He”|PRP.NOM)	1
P(”likes”|VP)	1
P(”her”|PRP.ACC)	1

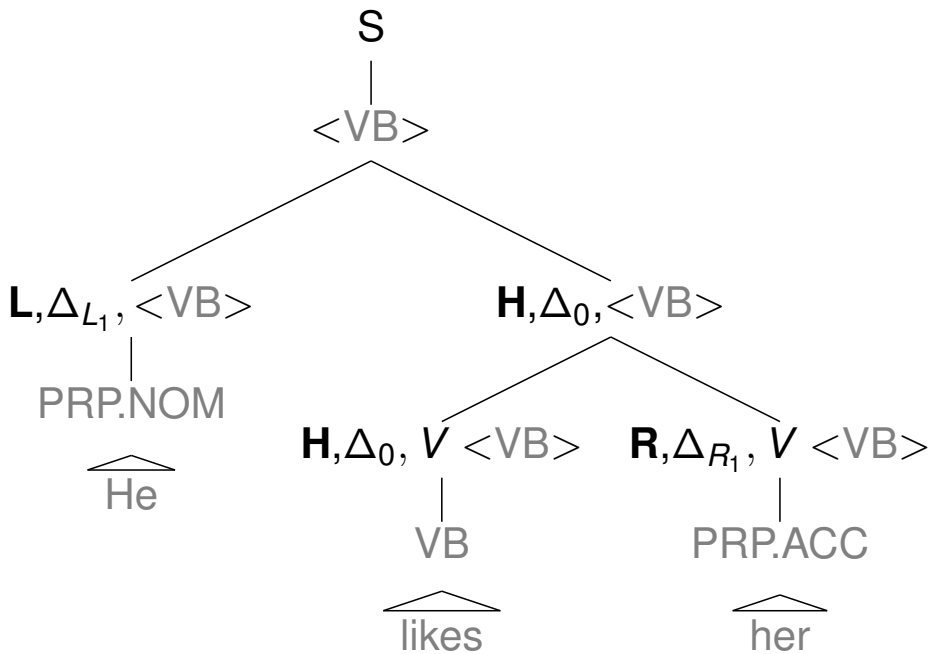
Example 1: Parent Encoding (Johnson 1998)



$\Rightarrow$

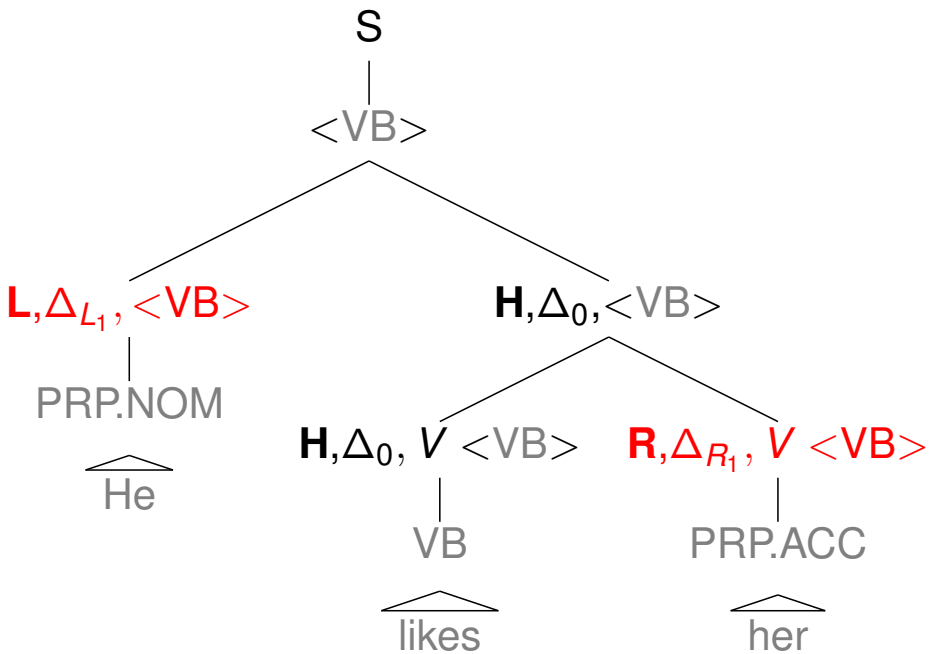
P(NP@S VP@S|S)	1
P(PRP.NOM |NP@S)	1
P(PRP.ACC |NP@VP)	1
P(VB NP@VP| VP@S)	1
P(”He”|PRP.NOM)	1
P(”likes”|VP)	1
P(”her”|PRP.ACC)	1

Example 2: Head-Driven Processes (Collins 1999)



$\Rightarrow$	
P(<VB>|S)	1
P(**L**Δ_{L_1},**H**Δ_0| <VB>,S)	1
P(PRP.NOM|**L**,Δ_{L_1},<VB>,S)	1
P(VP|**H**,Δ_0,<VB>,S)	1
P(<VB>|VP)	1
P(PRP.ACC|**R**,Δ_{R_1},<VB>,S)	1
P(VB|**H**,Δ_0,<VB>,S)	1
P("He"|PRP.NOM)	1
P("likes"|VB)	1
P("her"|PRP.ACC)	1

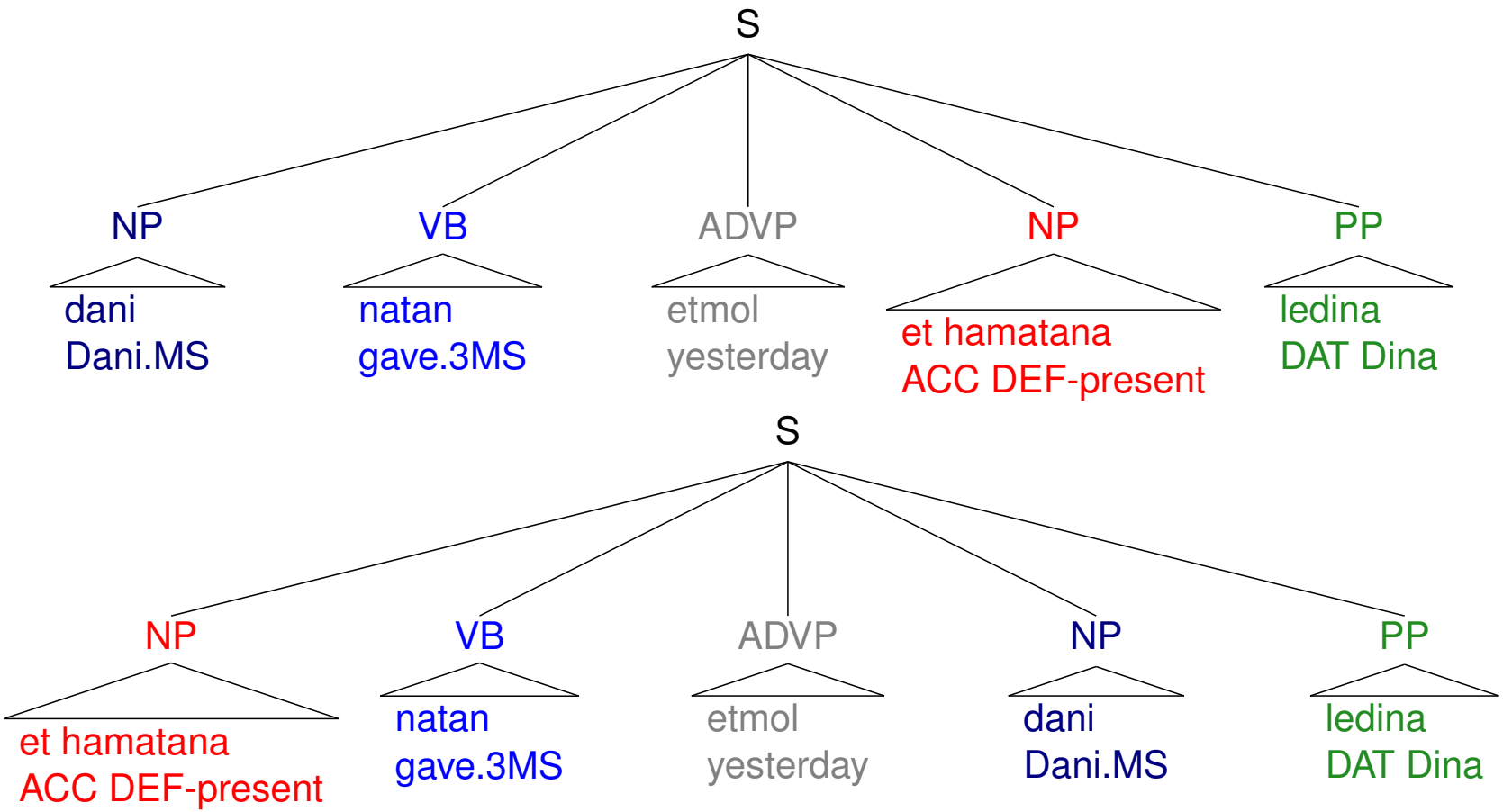
Example 2: Head-Driven Processes (Collins 1999)



⇒	
P(<VB>|S)	1
P(**L**Δ_{L_1},**H**Δ_0| <VB>,S)	1
P(PRP.NOM|**L**,Δ_{L_1},<VB>,S)	1
P(VP|**H**,Δ_0,<VB>,S)	1
P(<VB>|VP)	1
P(PRP.ACC|**R**,Δ_{R_1},<VB>,S)	1
P(VB|**H**,Δ_0,<VB>,S)	1
P(”He”|PRP.NOM)	1
P(”likes”|VB)	1
P(”her”|PRP.ACC)	1

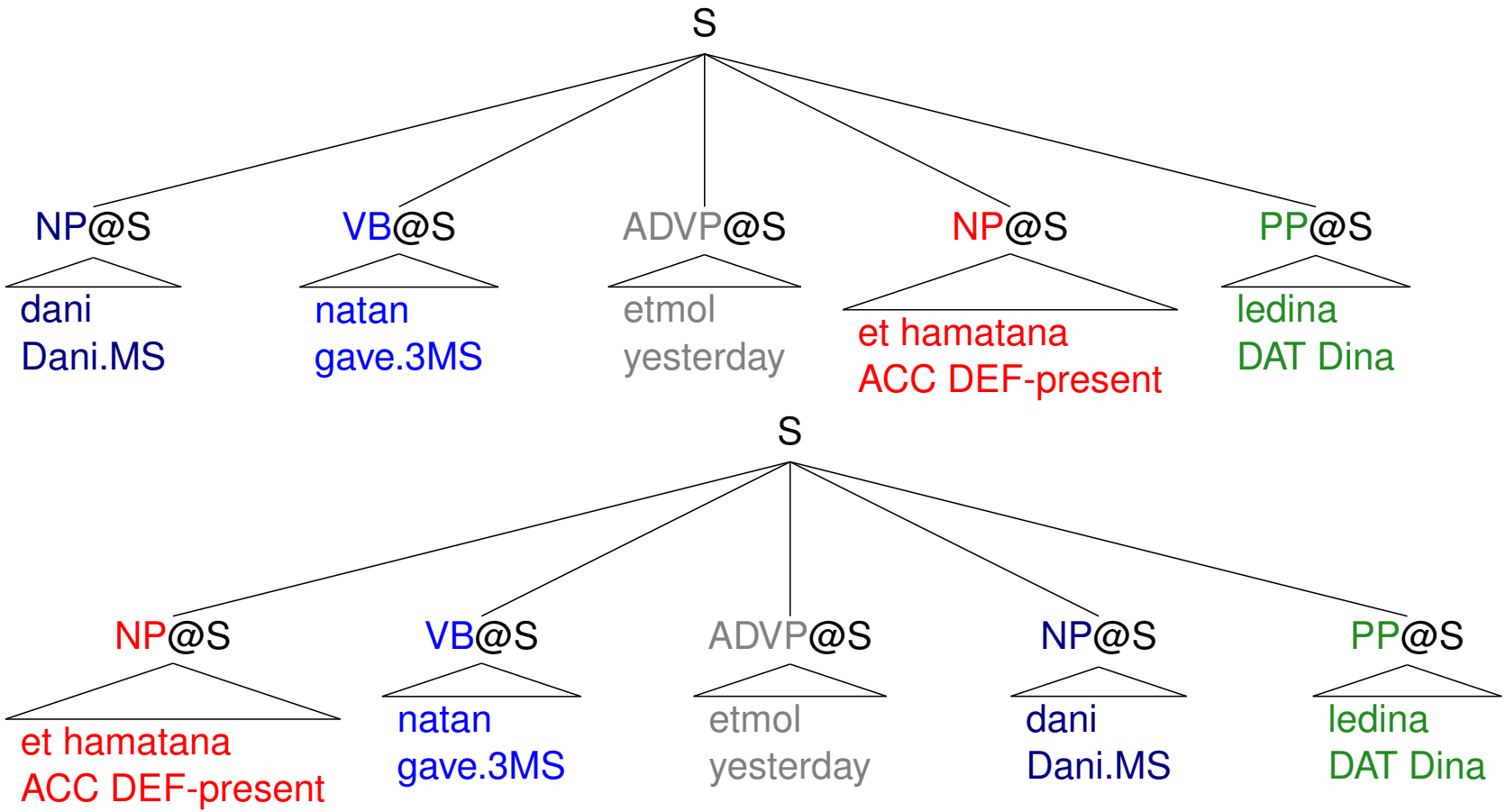
What if we take a different treebank

The Modern Hebrew Treebank



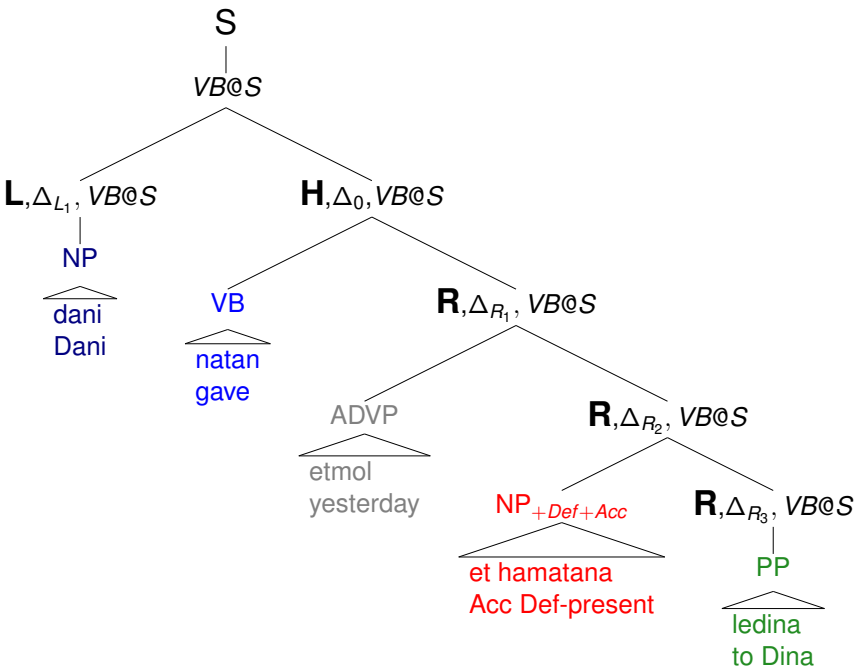
Example 1:

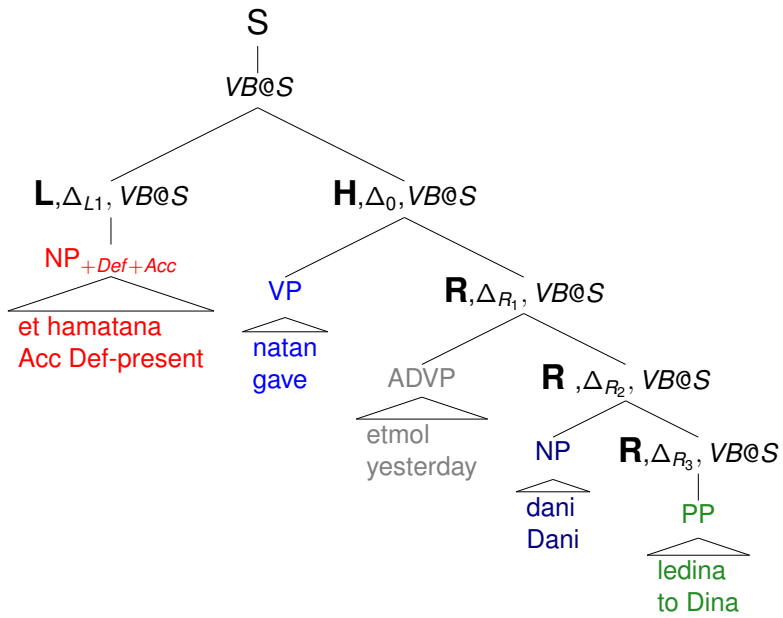
Parent Encoding



Example 2:

The Head-Driven (HD) Approach





So What Is Going On?

An Observation

- Configurational languages:
  $\rightsquigarrow$ Tree configurations approximate grammatical functions
- Less-configurational languages:
  $\rightsquigarrow$ Not.

So What Is Going On?

An Observation

- Configurational languages:
  $\rightsquigarrow$ Tree configurations approximate grammatical functions
- Less-configurational languages:
  $\rightsquigarrow$ Not.
  $\rightsquigarrow$ No regularities to statistically learn from data

A Question

What kind of form-function correspondence patterns
our parser needs to learn from the data?

Part 2: The Challenge

Modeling Form-Function Correspondence

The Data

The Data

Typological Dimensions of Variation

Basic Word-Order Typology

(Greenberg 1966, Mithun 1992)

Morphological Typology

(Sapir 1921, Greenberg 1954)

Nonconfigurationality

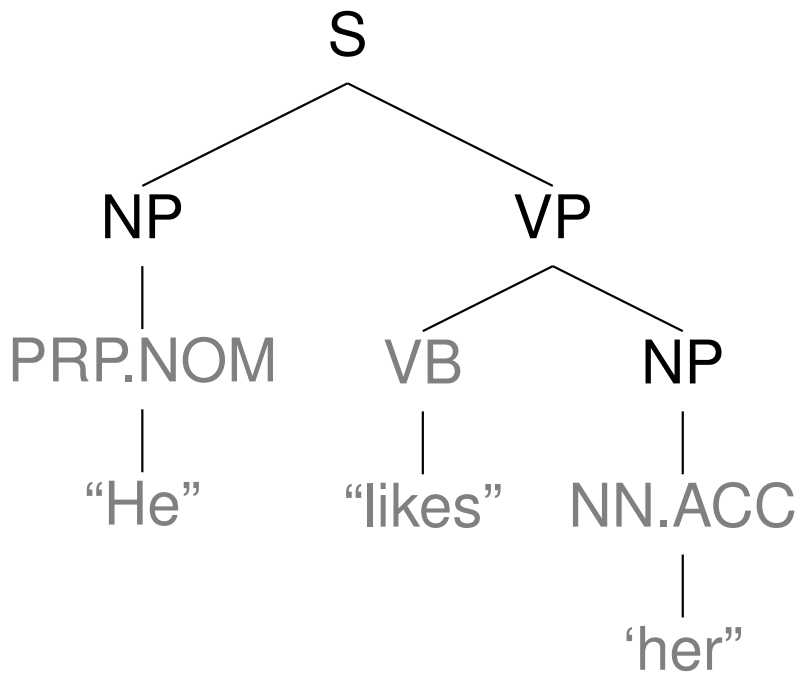
(Hale 1983, Austin and Bresnan 1996)

Nonconfigurationality as Misalignment

Predicate-Argument Relations

‘SBJ’ did ‘PRD’ to ‘OBJ’

Syntactic Configuration

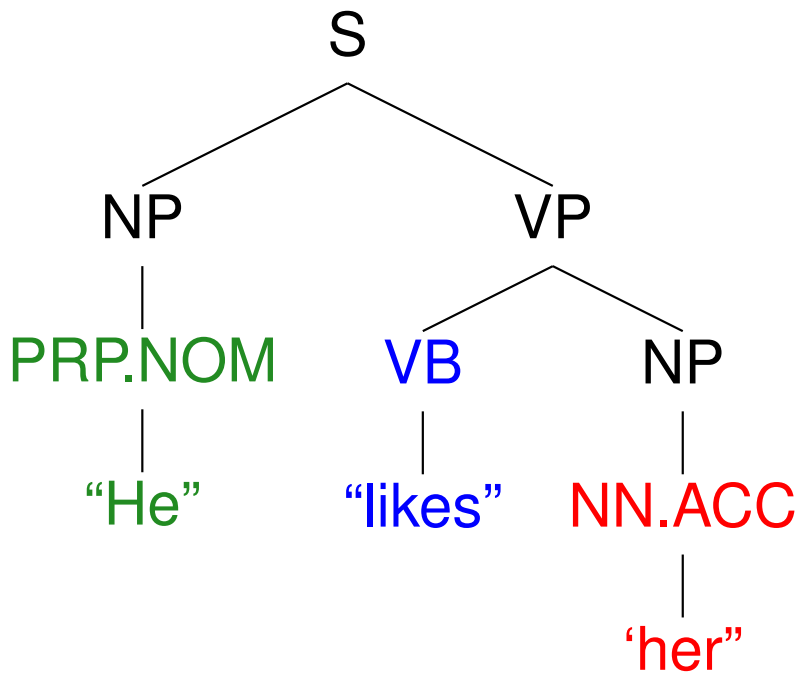


Nonconfigurationality as Misalignment

Predicate-Argument Relations

‘SBJ’ did ‘PRD’ to ‘OBJ’

Configurational Languages

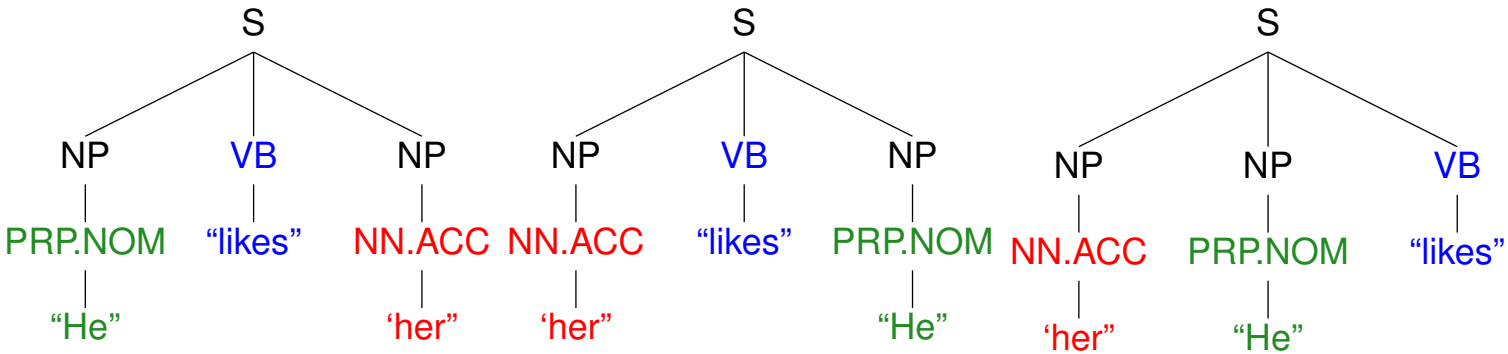


Nonconfigurationality as Misalignment

Predicate-Argument Relations

‘SBJ’ did ‘PRD’ to ‘OBJ’

Less-Configurational Languages



Morphosyntactic Exponence in Hebrew

Word-Order

(1) a. dani natan et hamatana ledina
Dani gave ACC the-present to-Dina
"Dani gave the present to Dina" (SVO)

b. et hamatana natan dani ledina
ACC the-present gave Dani to-Dina
"Dani gave the present to Dina" (OVS)

c. natan dani et hamatana ledina
gave Dani ACC the-present to-Dina
"Dani gave the present to Dina" (VSO)

d. ledina natan dani et hamatana
to-dina gave Dani ACC the-present
"Dani gave the present to Dina" (VSO)

Exponence Relations in Hebrew (1:1)

Case-Assigning Prepositions

(2) a. dani natan et hamatana ledina
Dani gave ACC DEF-present DAT-Dina

b. et hamatana natan dani ledina
ACC DEF-present gave Dani DAT-Dina

c. natan dani et hamatana ledina
gave Dani ACC DEF-present DAT-Dina

d. ledina natan dani et hamatana
DAT-dina gave Dani ACC DEF-present

Exponence Relations in Hebrew (1:many)

Differential Object-Marking

(3) a. dani natan et hamatana ledina
Dani gave ACC DEF-present to-Dina

b. et hamatana natan dani ledina
ACC DEF-present gave Dani to-Dina

c. natan dani et hamatana ledina
gave Dani ACC DEF-present to-Dina

d. ledina natan dani et hamatana
to-dina gave Dani ACC DEF-present

Exponence Relations in Hebrew (1:many)

Feature Spreading (Danon, 2007)

(4) a. dani natan [et matnat yom hahuledet] ledina
Dani gave [ACC present day DEF-birth] to-Dina

b. [et matnat yom hahuledet] natan dani ledina
[ACC present day DEF-birth] gave Dani to-Dina

c. natan dani [et matnat yom hahuledet] ledina
gave Dani [ACC present day DEF-birth] to-Dina

d. ledina natan dani [et matnat yom hahuledet]
to-dina gave Dani [ACC present day DEF-birth]

Exponence Relations in Hebrew (1:many)

Agreement

(5) a. dani natan et hamatana ledina
Dani.MS gave.3MS ACC DEF-present DAT-Dina

b. et hamatana natan dani ledina
ACC DEF-present gave.3MS Dani.MS DAT-Dina

c. natan dani et hamatana ledina
gave.MS Dani.3MS ACC DEF-present DAT-Dina

d. ledina natan dani et hamatana
DAT-dina gave.3MS Dani.MS ACC DEF-present

Exponence Relations in Hebrew (many:1)

Clitics and Null Anaphors

(6) a. dani natan et hamatana ledina
Dani.MS gave.3MS ACC DEF-present DAT-Dina
"Dani gave the present to Dina"

b. natati et hamatana ledina
gave.1S ACC DEF-present DAT-Dina
"I gave the present to Dina"

c. natatiha ledina
gave.1S.ACC.3FS DAT-Dina
"I gave it to Dina"

Language Types and Morphosyntactic Exponence

Recap:

CONFIGURATIONAL ——— NONCONFIGURATIONAL
1:1 ——————————————— many : many

- Exponence relations relate grammatical functions to the formal means that realize them in the syntactic structure
- Configurationality is a special case of a 1:1 mapping between grammatical functions to configurational positions

Question:

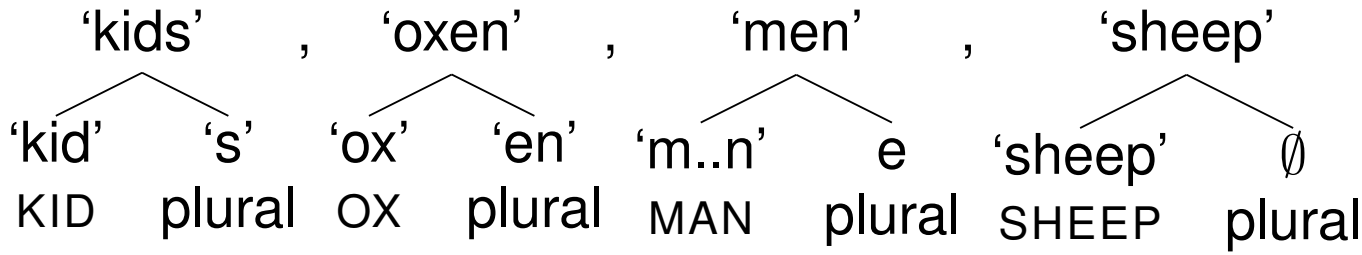
How can we model and statistically learn generally complex, many-to-many, form-function correspondence in syntax?

Part 3: The Proposal

Following the footsteps of morphology

Modeling Morphology (i): Terminology

Morpheme-Based Morphology (Bloomfield, 1933)



Morphological Exponence (Matthews 1991)

- Simple Exponence (1:1)
- Cumulative Exponence (many:1)
- Extended/Distributed Exponence (1:many)

Modeling Morphology (II): Assumptions (Stump 2001)

LEXICAL vs. INFERENTIAL Approaches

- LEXICAL:
  morphemes are primary, properties stored in the lexicon
- INFERENTIAL:
  properties are primary, forms are computed

INCREMENTAL vs. REALIZATIONAL Approaches

- INCREMENTAL:
  morphemes/properties are accumulated incrementally
- REALIZATIONAL:
  property-bundles are pre-condition for spell-out

Modeling Morphology (III): A Taxonomy

	LEXICAL	INFERENTIAL
INCREMENTAL	Item & Arrangement (Bloomfield 1933) (Lieber 1992)	Item & Processes (Hocket 1954) (Steele 1995)
REALIZATIONAL	Distributed Morphology (Halle and Marantz 1993) Lexical Phonology	(Extended) Word & Paradigm (Matthews 1972), (Anderson 1992) (Stump 2001), (Blevins 2006)

Table: A Taxonomy of Models for Morphology (Stump 2001)

The Strategy (IV): (Extended) Word-and-Paradigm

Paradigmatic Organization

/EAT/	1Sing	2Sing	3Sing	1Pl	2Pl	3Pl
Past	1SingPast	2SingPast	3SingPast	1PlPast	2PlPast	3PlPast
Present	1SingPres	2SingPres	3SingPres	1PlPres	2PlPres	3PlPres
Perfect	1SingPerf	2SingPerf	3SingPerf	1PlPerf	2PlPerf	3PlPerf

Realization Rules

/EAT/ +1SingPast | ‘ate’ , /EAT/ +3SingPast | ‘ate’ , /EAT/ +1SingPres | ‘eats’ , /EAT/ +3SingPres | ‘eat’

The Proposal (I): “Lifting” the Terminology

Morphological Exponence

- Simple (1:1, e.g., PL$\rightsquigarrow$‘s’ in ‘cats’)
- Cumulative (many:1, e.g., 3RD,SG$\rightsquigarrow$‘s’ in ‘eats’)
- Distributed/Extended (1:many PL$\rightsquigarrow$‘i’,‘ren’ in ‘children’)

The Proposal (I): “Lifting” the Terminology

Morphological Exponence : Properties $\rightsquigarrow$ Words

- Simple (1:1, e.g., PL$\rightsquigarrow$‘s’ in ‘cats’)
- Cumulative (many:1, e.g., 3RD,SG$\rightsquigarrow$‘s’ in ‘eats’)
- Distributed/Extended (1:many PL$\rightsquigarrow$‘i’,‘ren’ in ‘children’)

(Morpho)Syntactic Exponence : Relations $\rightsquigarrow$ Positions

- Simple (1:1, e.g., SBJ $\rightsquigarrow$ nominative)
- Cumulative (many:1, e.g., PRD,OBJ $\rightsquigarrow$ clitics)
- Distributed/Extended (1:many, e.g., SBJ $\rightsquigarrow$ agreement)

The Proposal (II): Modeling Assumptions

Configurational vs. Relational Approaches

- Configurational:
  configurations are primary, relations are derived
- Relational:
  relations are primary, configurations are computed

Incremental vs. Realizational Approaches

- Incremental:
  constructive operations,
  individually/incrementally define/add relations
- Realizational:
  interpretive operations,
  sets of relations are precondition to realization

The Proposal (III): A Taxonomy (Tsarfaty 2010)

	CONFIGURATIONAL	RELATIONAL
INCREMENTAL	X-Bar Theory	Dependency Grammar
REALIZATIONAL	Tree Adjoining Grammar	

Table: A Taxonomy of Generative Syntactic Frameworks

The Proposal (IV): Relational-Realizational Modeling

The *Relational* Assumption

- Paradigms organize the syntactic domain
- Cells in paradigms define sets of relations
- Sets of relations are realized in different configurations

The *Realizational* Assumption

- Sets of relations (Arg-St) are primitives
- Rules interpret sets of relations as surface orms
- Rules can refer to multiple relations and span clauses

Realization in Syntax is Recursive!

Realization of a cells refers to function cells in other paradigms

The Proposal (IV): Relational-Realizational Modeling

S⟨PRED⟩ FEATS ARG-ST	Affirmative	Interrogative	Imperative
intransitive	S_{affirm}+{SBJ,PRD}	S_{inter}+{SBJ,PRD}	S_{imper}+{SBJ,PRD}
transitive	S_{affirm}+{SBJ,PRD,OBJ}	S_{inter}+{SBJ,PRD,OBJ}	S_{imper}+{SBJ,PRD,OBJ}
ditransitive	S_{affirm}+{SBJ,PRD,OBJ,COM}	S_{inter}+{SBJ,PRD,OBJ,COM}	S_{imper}+{SBJ,PRD,OBJ,COM}

Figure: Paradigmatic Organization

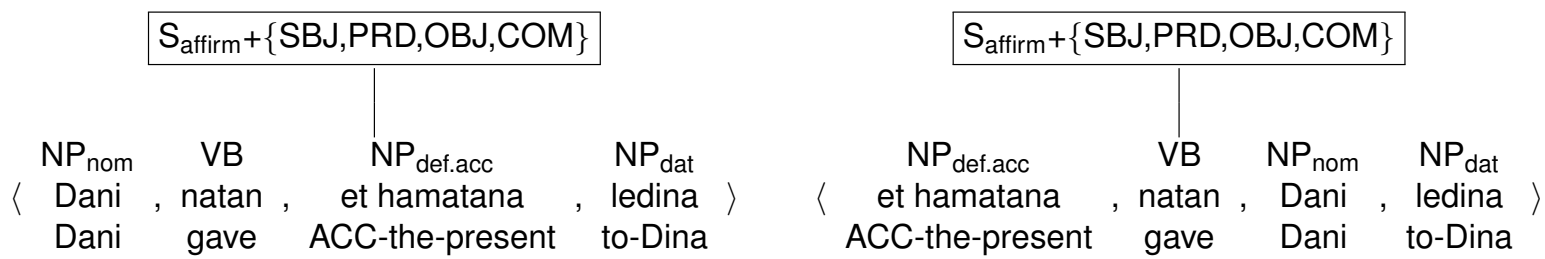
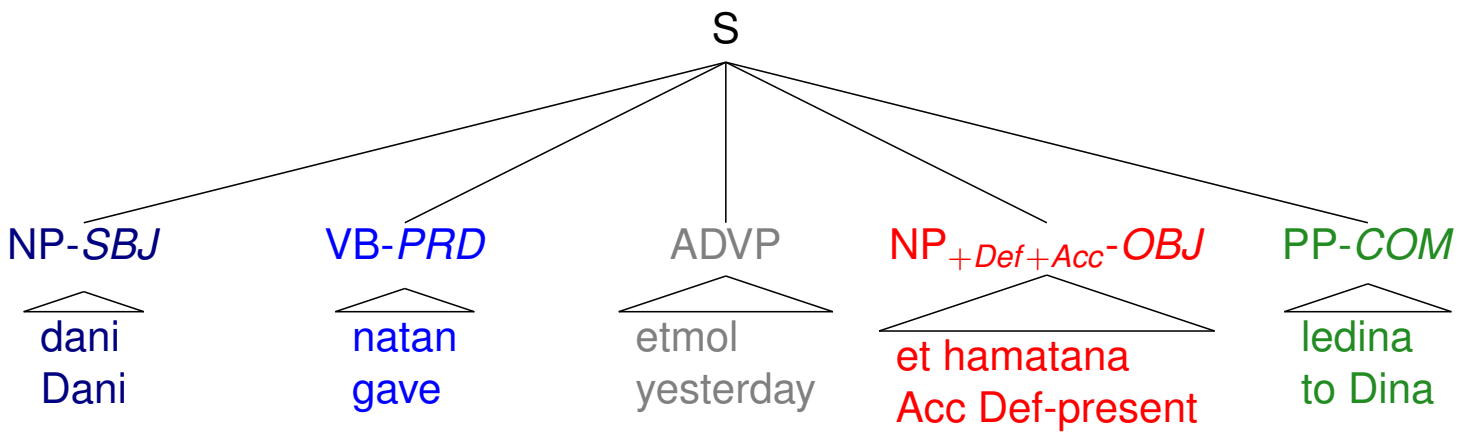


Figure: Realization Rules

Realization Rules

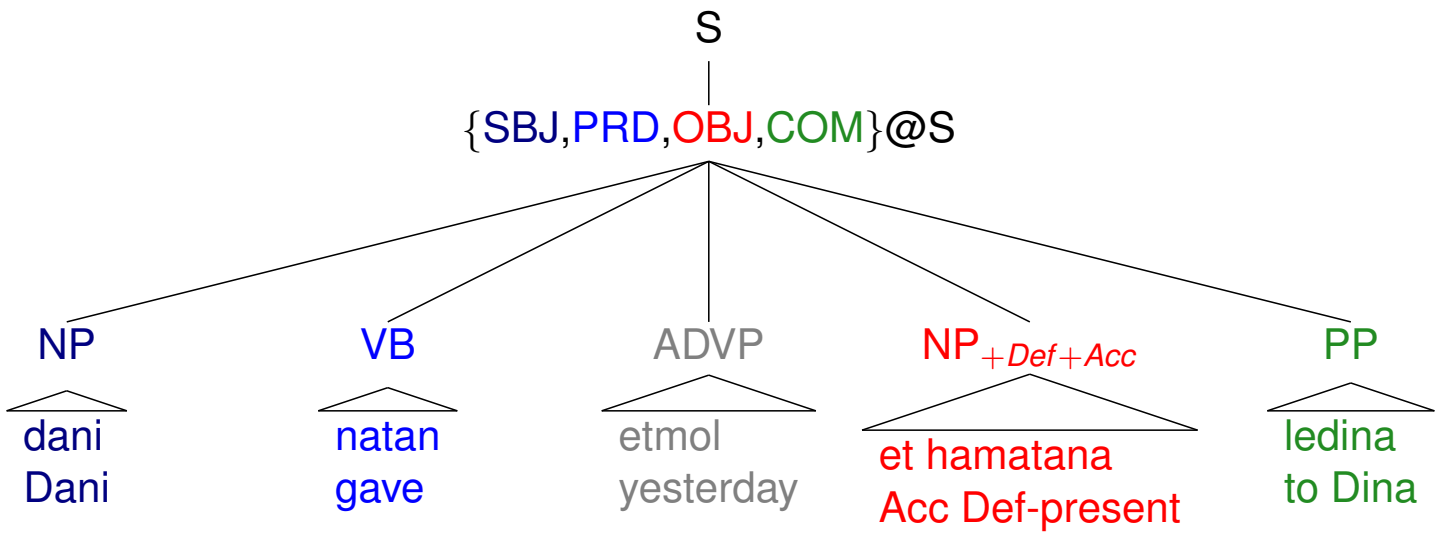
Realization Rules

A PCFG Approximation



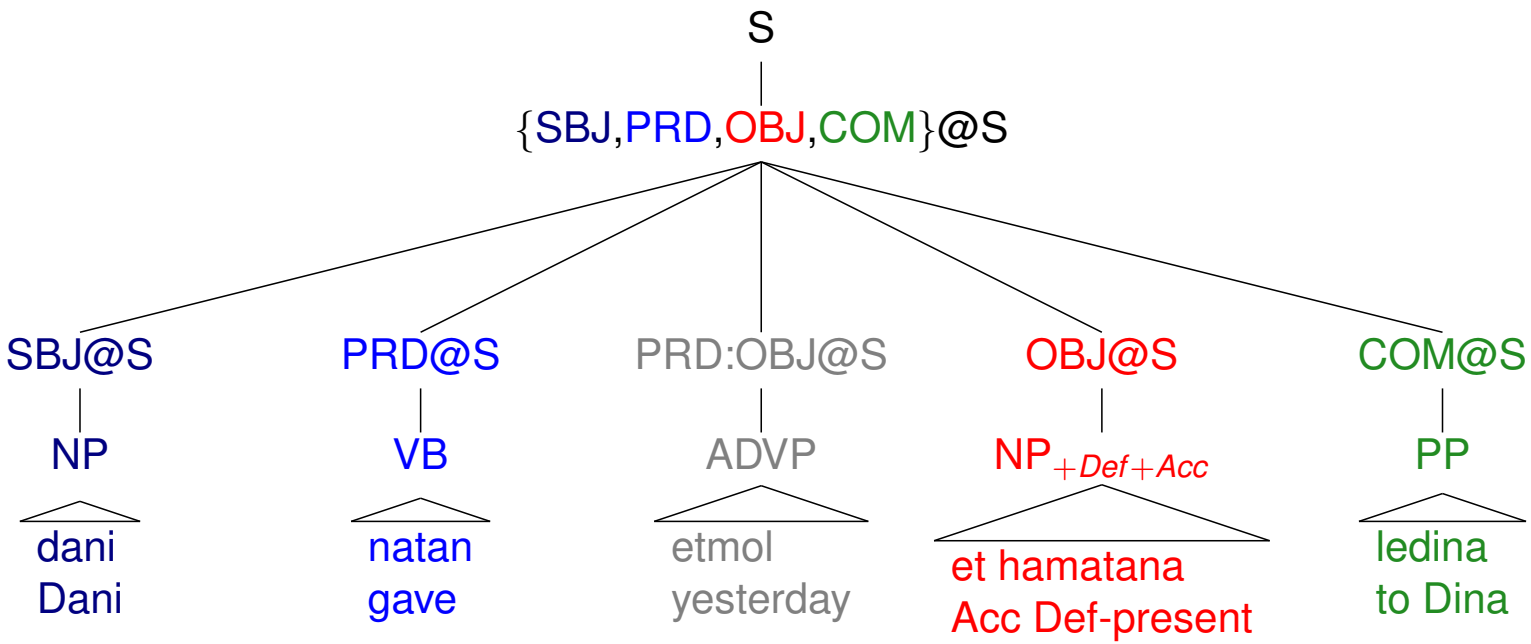
Realization Rules

Form-Function Separation

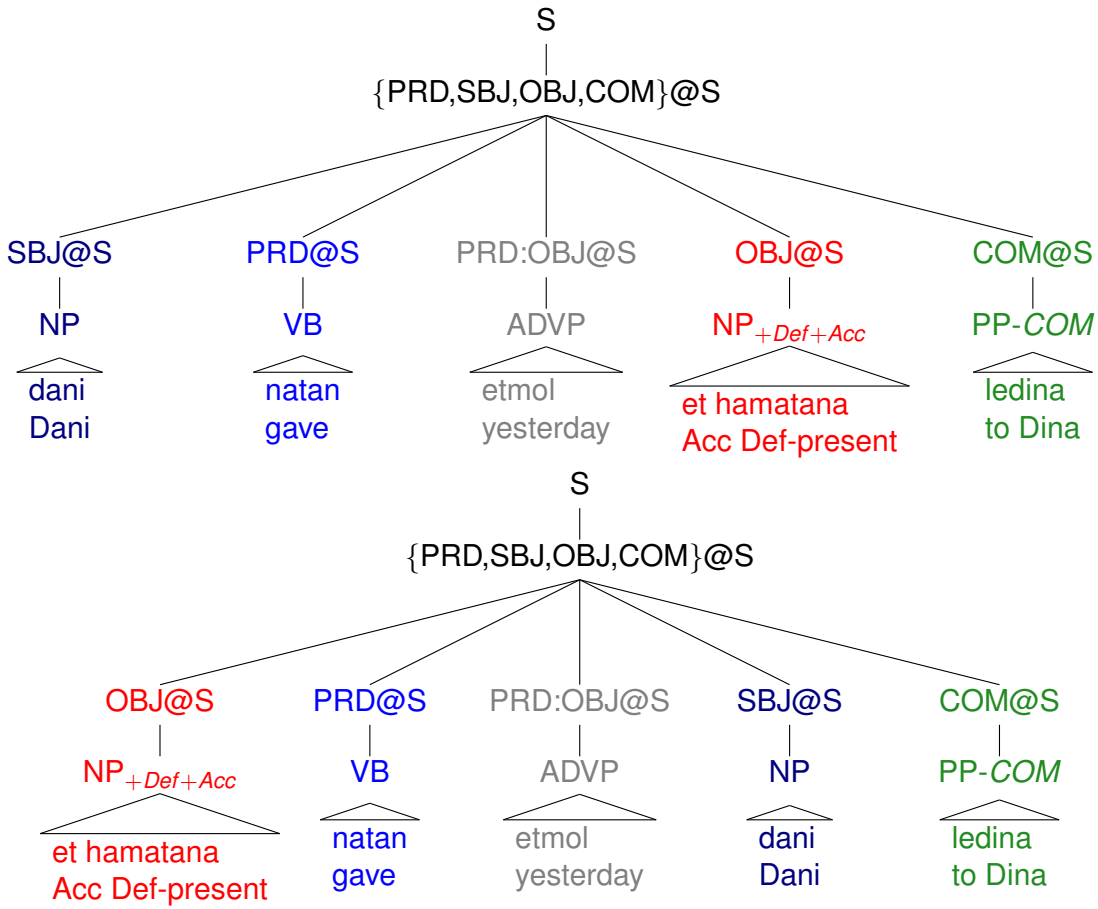


Realization Rules

Morphological and Syntactic Realization

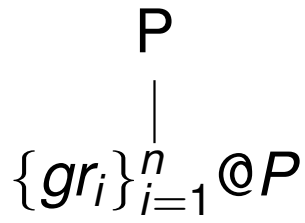


Realization Rules: Economy and Generalization

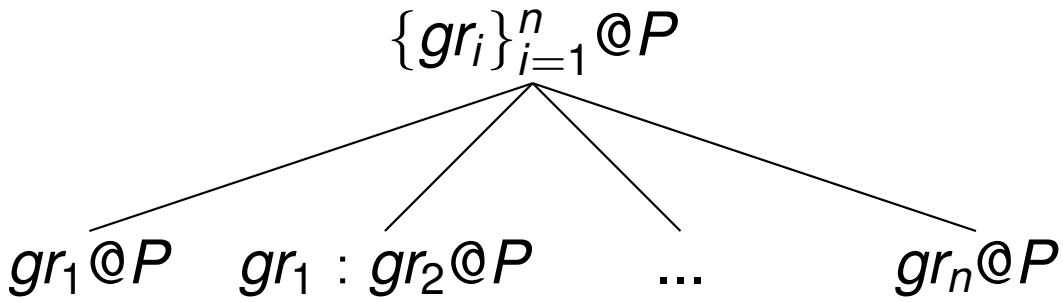


The Generative Model

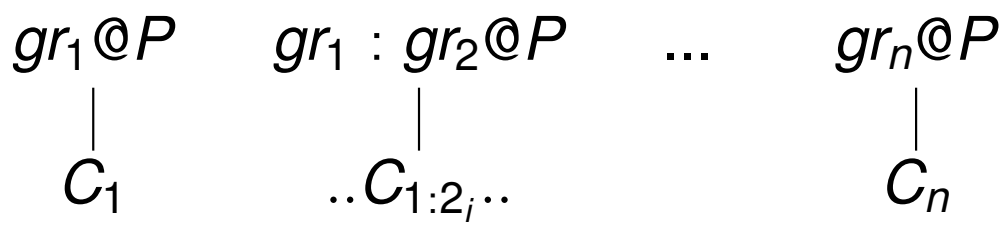
Projection:



Configuration:



Realization:



The Probabilistic Model

The RR Probabilities:

$$
\begin{array}{ll}
\mathbf{P_{RR}}(r) = & \\
\textit{Projection} & \mathbf{P_p}(\{gr_i\}_{i=1}^{n}|P) \times \\
\textit{Configuration} & \mathbf{P_c}(\langle gr_0 : gr_1, g_1, \ldots\rangle|\{gr_i\}_{i=1}^{n}, P) \times \\
\textit{Realization} & \prod_{i=1}^{n} \mathbf{P_{r_1}}(C_i|gr_i, P) \times \\
 & \mathbf{P_{r_2}}(\langle C_{0_1}, \ldots, C_{0_{m_0}}\rangle|gr_0 : gr_1, P) \times \\
 & \prod_{i=1}^{n} \mathbf{P_{r_2}}(\langle C_{i_1}, \ldots, C_{i_{m_i}}\rangle|gr_i : gr_{i+1}, P)
\end{array}
$$

The RR Parser:

$$\pi^* = argmax_\pi P(\pi) = argmax_\pi \prod_{r\in\pi} \mathbf{P_{RR}}(r)$$

Part IV: Applications

- Cross-Linguistic Parsing
- Cross-Linguistic Variation

Application I: Parsing Modern Hebrew

Data

The Modern Hebrew Treebank v2, head annotated.
6500 sentences, 500/5500/500 dev/train/test split

Models

- Grammatical Functions: PRD, SBJ, OBJ, COM, CNJ
- Morphological Splits: **PoS/Def/Acc/Gender**

Estimation

Relative Frequency + Simple Unknown Words Smoothing

Parsing

Exhaustive Viterbi Parsing (using BitPar, Schmid 2004)

Evaluation

PARSEVAL (i) Overall, and (ii) Per Category Evaluation

A Taxonomy of PCFG-based Parsers

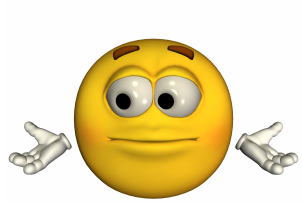
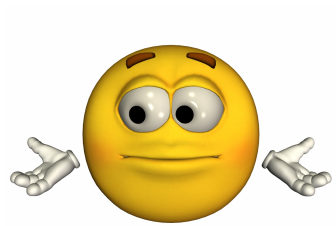
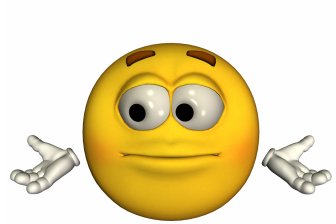
	CONFIGURATIONAL	RELATIONAL
INCREMENTAL	Head-Driven Parsing (Collins 1999) 	
REALIZATIONAL		Relational-Realizational (Tsarfaty et al. 2009) 

Table: A Taxonomy of PCFG-Based Parsing Frameworks

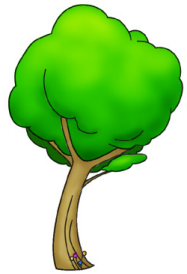
Overall Results

		
74.66/74.35 (7385)	73.52/74.84 (21399)	**76.32/76.51 (13618)**

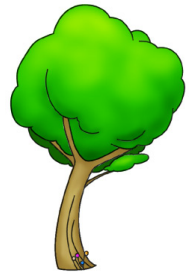
Overall Results

		
74.66/74.35 (7385)	73.52/74.84 (21399)	**76.32/76.51 (13618)**

Hebrew Parsing Results Using Gold Standard Input

		
∅	83.06	**83.49**
	(5914)	**(6688)**
gender	82.18	**83.70**
	(10765)	**(10063)**
case/*def*	79.53	**83.66**
	(12700)	**(12386)**
gender/*case*/*def*	80.89	**84.13**
	(13028)	**(13618)**

Swedish Parsing Results Using Gold Standard Input

		
∅	78.65	**77.71**
	(8696)	**(10099)**
gender	73.20	**78.09**
	(11382)	**(12593)**
case/*def*	74.90	**79.09**
	(11239)	**(13912)**
gender/*case*/*def*	68.97	**77.89**
	(13347)	**(14991)**

Application II: Cross-Linguistic Variation

Application II: Cross-Linguistic Variation

1: Apply the model to different languages, e.g.,

- **Hebrew:** a Semitic Language
- **Swedish:** a Germanic Language

2: Learn the distribution of model parameters

- RR-Projection
- RR-Configuration
- RR-Realization

3: Instantiate typological parameters for UG

Parameter 1: Basic Word-Order (Greenberg 1963)

Basic Word-Order Parameter in Hebrew: P(< *configuration* >|{SBJ,PRD,OBJ}@S)

Probability	*Configuration*
35.3%	SBJ PRD OBJ □
15.6%	SBJ PRD □ OBJ □
12.3%	□ PRD SBJ OBJ □
10.3%	SBJ □ PDR OBJ □
6.5%	□ SBJ PRD OBJ □
4.1%	SBJ □ PRD □ OBJ □
3.7%	□ PRD SBJ □ OBJ □
3%	OBJ PRD SBJ □
1.7%	□ SBJ PRD □ OBJ □
1.7%	□ PRD OBJ SBJ □
1.3%	SBJ □ PRD OBJ □
1 %	□ PRD □ SBJ OBJ □

Parameter 1: Basic Word-Order (Greenberg 1963)

Basic Word-Order Parameter in Swedish: P(< *configuration* >|{SBJ,PRD,OBJ}@S)

Probability	*Configuration*
35.5%	SBJ PRD OBJ □
18.9%	SBJ PRD □ OBJ □
13.9%	□ PRD SBJ PBJ □
8.1%	SBJ PRD OBJ
4.7%	□ PRD SBJ □ OBJ
3.5%	OBJ PRD SBJ
2.6%	SBJ PRD OBJ □
1.7%	OBJ PRD SBJ COM □
1.6%	PRD SBJ OBJ □
1.6%	□ PRD SBJ OBJ
1%	□ PRD SBJ □ OBJ

Parameter 2: Inflectional Systems

The Object-Marking Parameter in Hebrew: P(< *morphosyntactic representation* >| OBJ@S)

Probability	Realization
43.5%	NP.⟨*NN*⟩
14.7%	NP.DEF.ACC⟨*NN*⟩
8.8%	NP⟨*NNT*⟩
7.4%	NP.DEF.ACC⟨*NNP*⟩
6.7%	NP.DEF.ACC⟨*NN.DEF*⟩
6.5%	NP.DEF.ACC⟨*NNT*⟩
5.8%	NP.DEF.ACC⟨*PRP*⟩

Parameter 2: Inflectional Systems

The Object-Marking Parameter in Swedish: P(< *morphosyntactic representation* >| OBJ@S)

Probability	Realization
46%	NP.IND.NOM
20%	NP.DEF.NOM
13.4%	S
7.3%	NP.DEF.NOM-OBJ
4.9%	VP
3.6%	NP.IND
2.8%	NP.NOM

Towards Computational Typology and Statistical UG

We can potentially use the RR parameters to...

- Quantify Intra-Language Variation
- Quantify Cross-Linguistic Variation
- Quantify Nonconfigurationality
- Learn Probabilistic P&P

Conclusion

We presented a Relational-Realizational Architecture for Specifying and Learning Morphosyntactic Descriptions

- Simple
- Formal
- Robust
- Implementable
- Interpretable
- Explanatory

Paradigms augmented with realization rules may constitute a useful and powerful modeling strategy also for **Statistical MT**

Homework Assignment II

Read:

Tsarfaty, R. *Relational-Realizational Parsing*
PhD Thesis, University of Amsterdam, 2010

Implement:

- **Transfer rules** to model *morphosyntactic variation*
- **Alignment tools** to align *paradigm cells*

Thank You!

Questions?

Thank You!

Questions?

Now a postdoc at Uppsala, looking for a tenure track position :)

Acknowledgements

Thanks to. My PhD Committee: Remko, Khalil, Joakim, Mark, Yoad, Henk, Michiel. My morphology teachers: Jim, Greg. Lots of Colleagues and co-authors. Dutch and Swedish funding agencies.