JOHNS HOPKINS
UNIVERSITY
CENTER FOR LANGUAGE
& SPEECH PROCESSING

human language technology
center of excellence

Learning Sentential Paraphrases from Bilingual Parallel Corpora for Text-to-Text Generation

Juri Ganitkevitch, Chris Callison-Burch,
Courtney Napoles, and Benjamin Van Durme

What are Paraphrases...

Differing textual expressions of the same meaning:

cup	mug
the king's speech	His Majesty's address
X_1 talks to X_2	X_1 converses with X_2
NN devoured NP	NP was eaten by NN

...good for?

Anything that deals with **text** and **meaning**, i.e. automatic...

...summarization, translation, question answering, compression, simplification, natural language generation, entailment recognition, etc.

Data-Driven Paraphrase Extraction

Where do paraphrases come from? **Data**.

Approaches group nicely by type of text corpora they rely on.

Types of Corpora

Monolingual parallel:	English – English
Monolingual comparable:	English ~ English
Plain monolingual:	English
Bilingual parallel:	English – French

What a scene! Seized by the tentacle and **glued to** its suckers, the unfortunate man was **swinging in the air** at the **mercy** of this enormous appendage. He gasped, he choked, he yelled: "Help! Help!" I'll hear his **harrowing plea** the rest of my life!
The **poor fellow** was **done for**.

What a scene! The unhappy man, seized by the tentacle and **fixed to** its suckers, was **balanced in the air** at the **caprice** of this enormous trunk. He rattled in his throat, he was stifled, he cried, "Help! help!" That **heart-rending cry**! I shall hear it all my life.
The **unfortunate man** was **lost**.

Monolingual Parallel Data

Barzilay & McKeown ’01

Emma **cried** and he tried to **console** her.

Emma **burst into tears** and he tried to **comfort** her.

Pang, Knight & Marcu ’03

Syntactic alignments on parallel reference translations from MT

Monolingual Parallel Data

Drawbacks:

Corpora are scarce and small

Amount and coverage of paraphrases extracted are often lacking

A staggering 5 million Americans **have been victims of identity theft** in the last five years, according to a federal trade commission survey out this week.

In the last year alone, 1 million people **have had their identity purloined**.

At a press conference in the Reagan Presidential Library, the popular rotund governor from the great state of New Jersey **disappointed** the Republican base saying he won't run.

Many Republicans' **hearts were broken** by Chris Christie **reiterating** his refusal to **run for presidency.**

Last night, the Garden State governor **stated once again** that he will not seek the **presidential nomination.**

Monolingual Comparable Data

Dolan, Quirk & Brockett ’04

Identify near-parallel sentence pairs using an edit distance metric

Resulting sentence pairs are often too similar

Monolingual Comparable Data

Drawbacks:

- Data becomes noisy
- Rich paraphrases are hard to extract
- Depending on the method, may learn only near-trivial paraphrases

Monolingual Data

- E.g. any & all English text
- Vast quantities easily available
- Use distributional characteristics to identify paraphrases

Monolingual Data

Lin & Pantel ’01

duty | responsibility

Modified by: additional, administrative, assigned, assumed, collective, congressional, constitutional

Object of: accept, articulate, assert, assign, assume, attend to, avoid, become, breach

Bhagat & Ravichandran ’08

Scale to large corpora using LSH

Monolingual Data

Drawbacks:

Lacks alignment information, relies on distributional similarity measures

No sentential paraphrases

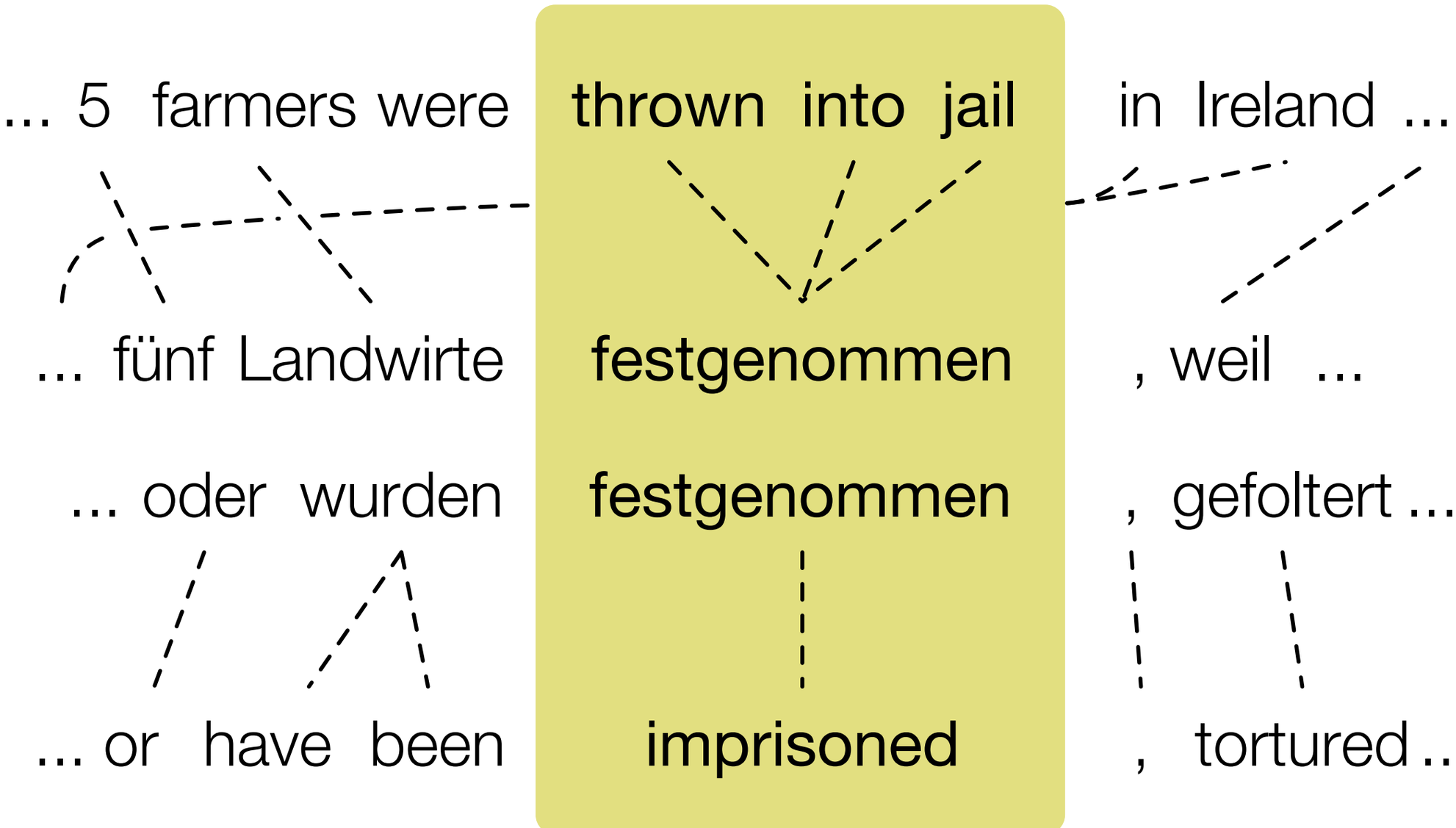
Prone to cousin & antonym errors:

boy | girl rise | fall

Bilingual Parallel Data

- E.g. sentence-aligned corpora in English and any foreign language (Hansards, EuroParl, etc.)
- Available in large quantities
- But: how do they yield paraphrases?

Pivoting


... 5 farmers were
thrown into jail
in Ireland ...
... fünf Landwirte
festgenommen
, weil ...
... oder wurden
festgenommen
, gefoltert ...
... or have been
imprisoned
, tortured ...

Pivoting

$$
\begin{aligned}
p(e_2|e_1) &= \sum_f p(e_2, f|e_1) \\
&= \sum_f p(e_2|f, e_1)p(f|e_1) \\
&\approx \sum_f p(e_2|f)p(f|e_1)
\end{aligned}
$$

Pivot-Based Paraphrase Extraction

- Solid body of research in this direction
- Builds on experience in MT
- But: unclear whether extraction of **syntactically informed** paraphrases possible

Translation to Paraphrasing

cup | Tasse

Callison-Burch & Bannard ’05

the king’s speech | die Rede des Königs

$X \rightarrow$ X_1 talks to X_2 | X_1 redet mit X_2

Madnani et al. ’07

VP → NN devoured NP | NN aß NP

Translation to Paraphrasing

cup	Tasse
the king's speech	die Rede des Königs

Callison-Burch & Bannard '05

$X \rightarrow$ X_1 talks to X_2	X_1 redet mit X_2
$VP \rightarrow$ NN devoured NP	NN aß NP

Madnani et al. '07

Translation to Paraphrasing

cup	|	Tasse
the king's speech	|	die Rede des Königs
$X \rightarrow$ X_1 talks to X_2	|	X_1 redet mit X_2
VP $\rightarrow$ NN devoured NP	|	NN aß NP

Callison-Burch & Bannard '05

Madnani et al. '07

Translation to Paraphrasing

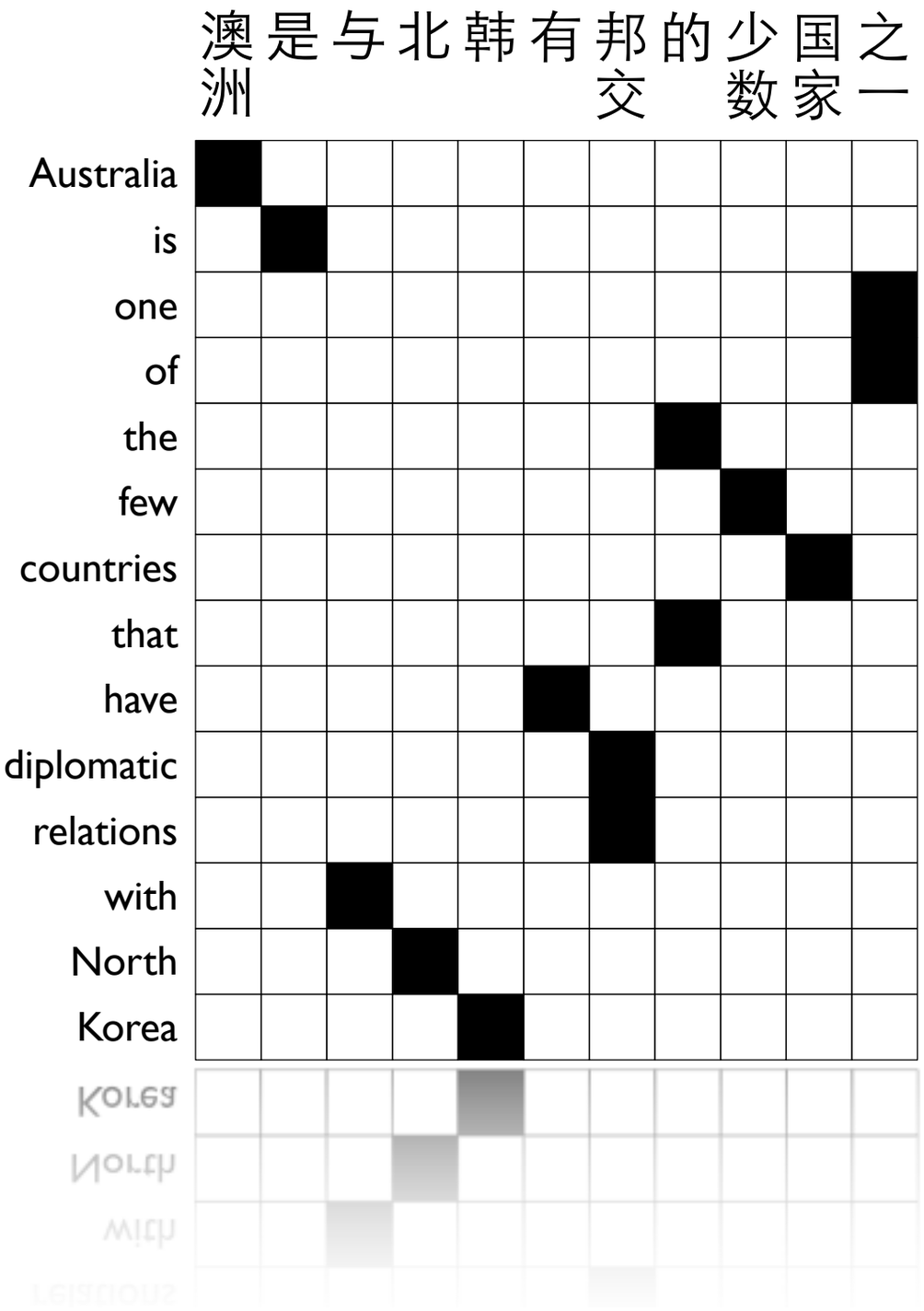
cup	Tasse	
the king's speech	die Rede des Königs	Callison-Burch & Bannard '05
$X \rightarrow$ X_1 talks to X_2	X_1 redet mit X_2	Madnani et al. '07
VP → NN devoured NP	NN aß NP	New!

Phrase Extraction

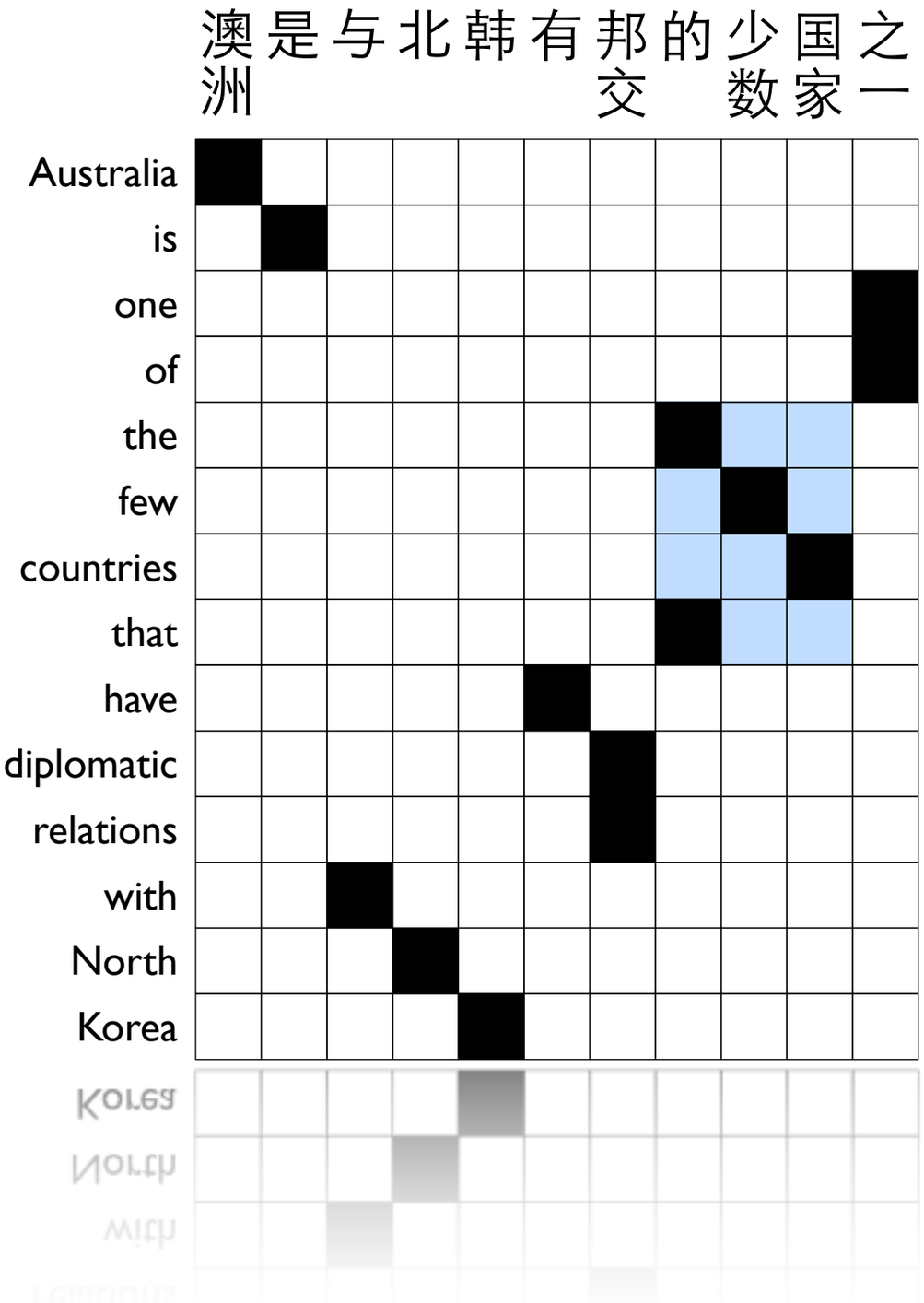
澳洲 是 与 北 韩 有 邦交 的 少数 国家 之一

Australia
is
one
of
the
few
countries
that
have
diplomatic
relations
with
North
Korea

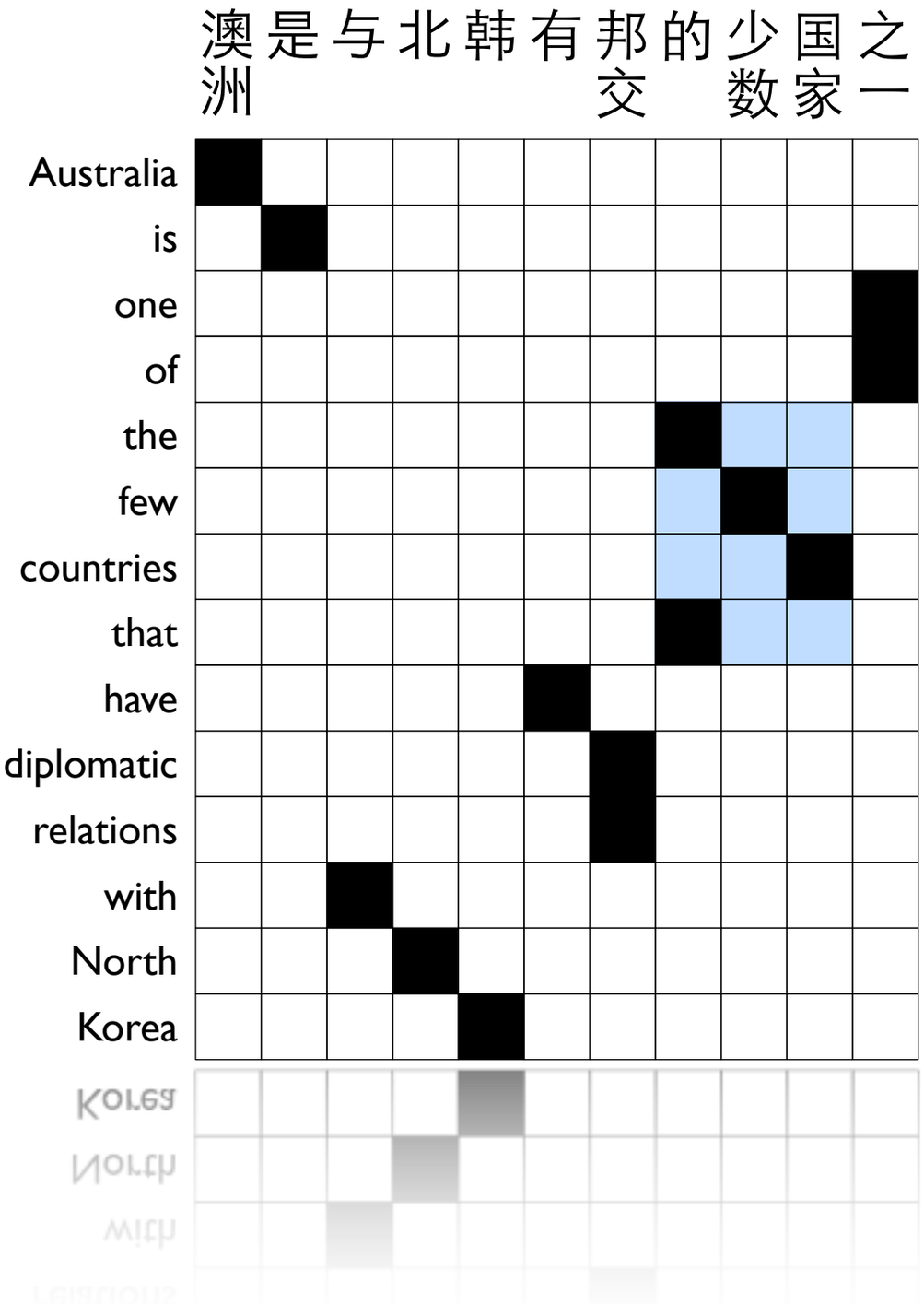
Phrase Extraction



Phrase Extraction



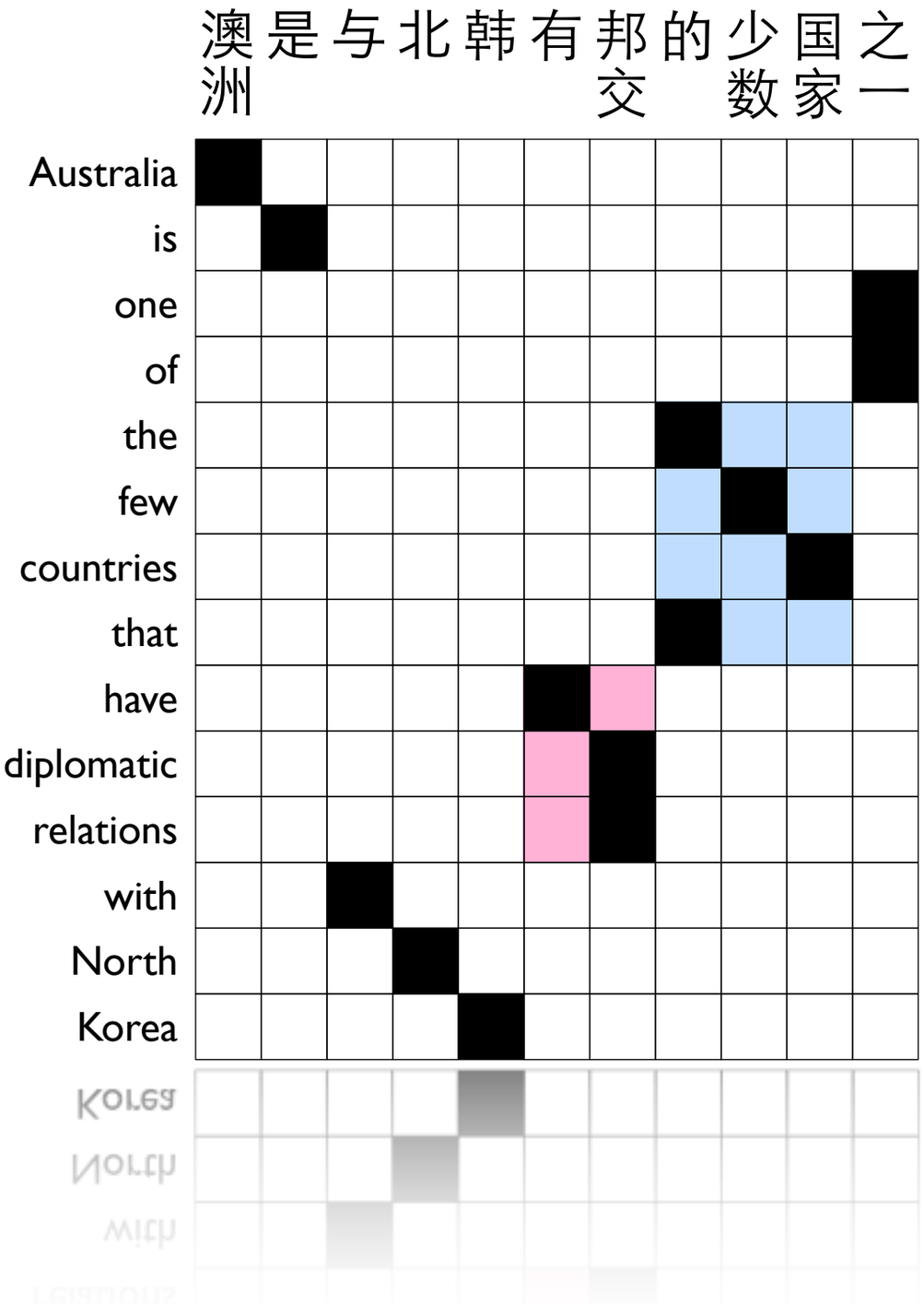
Phrase Extraction



的少数国家

the few countries that

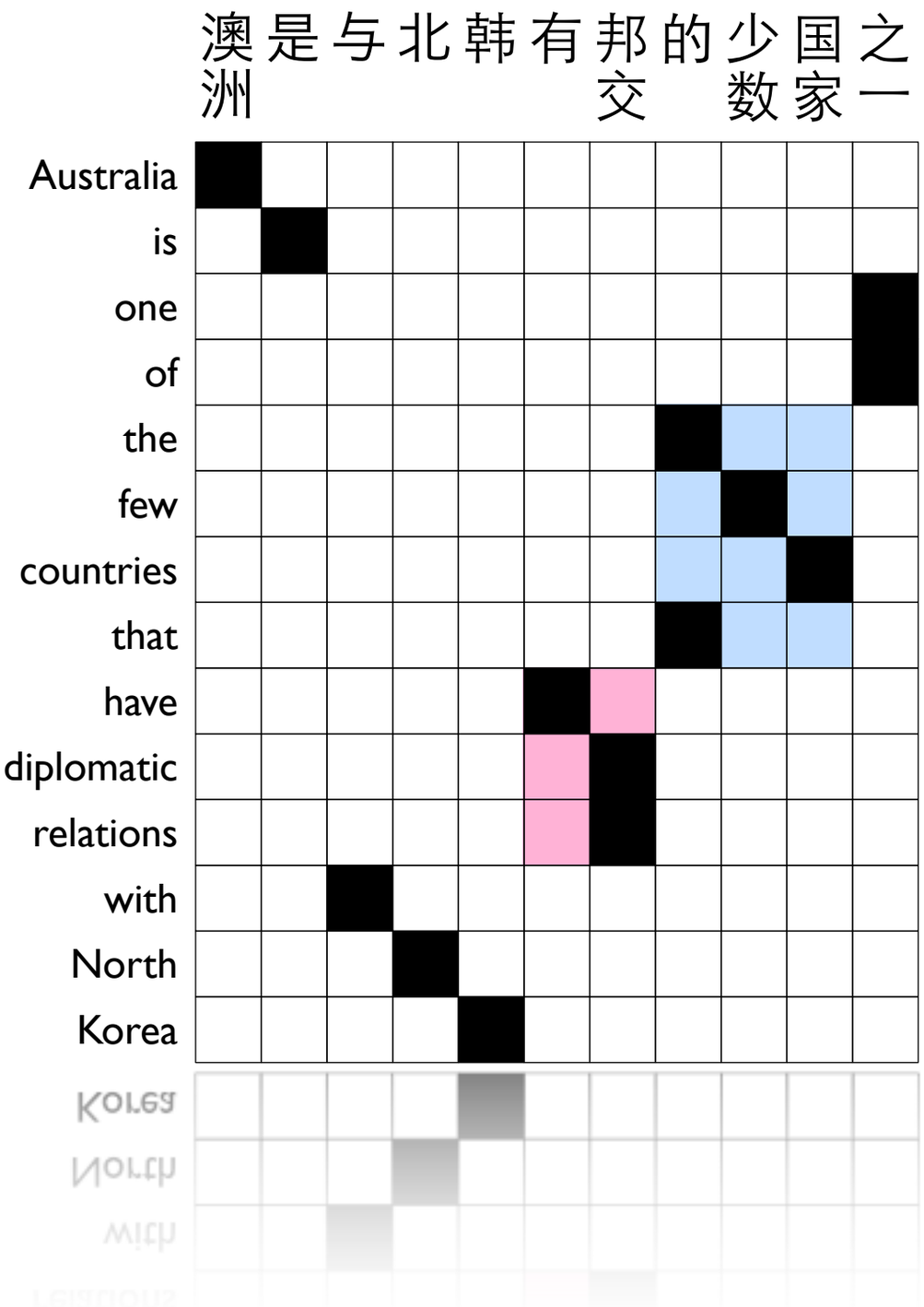
Phrase Extraction



的少数国家

the few countries that

Phrase Extraction



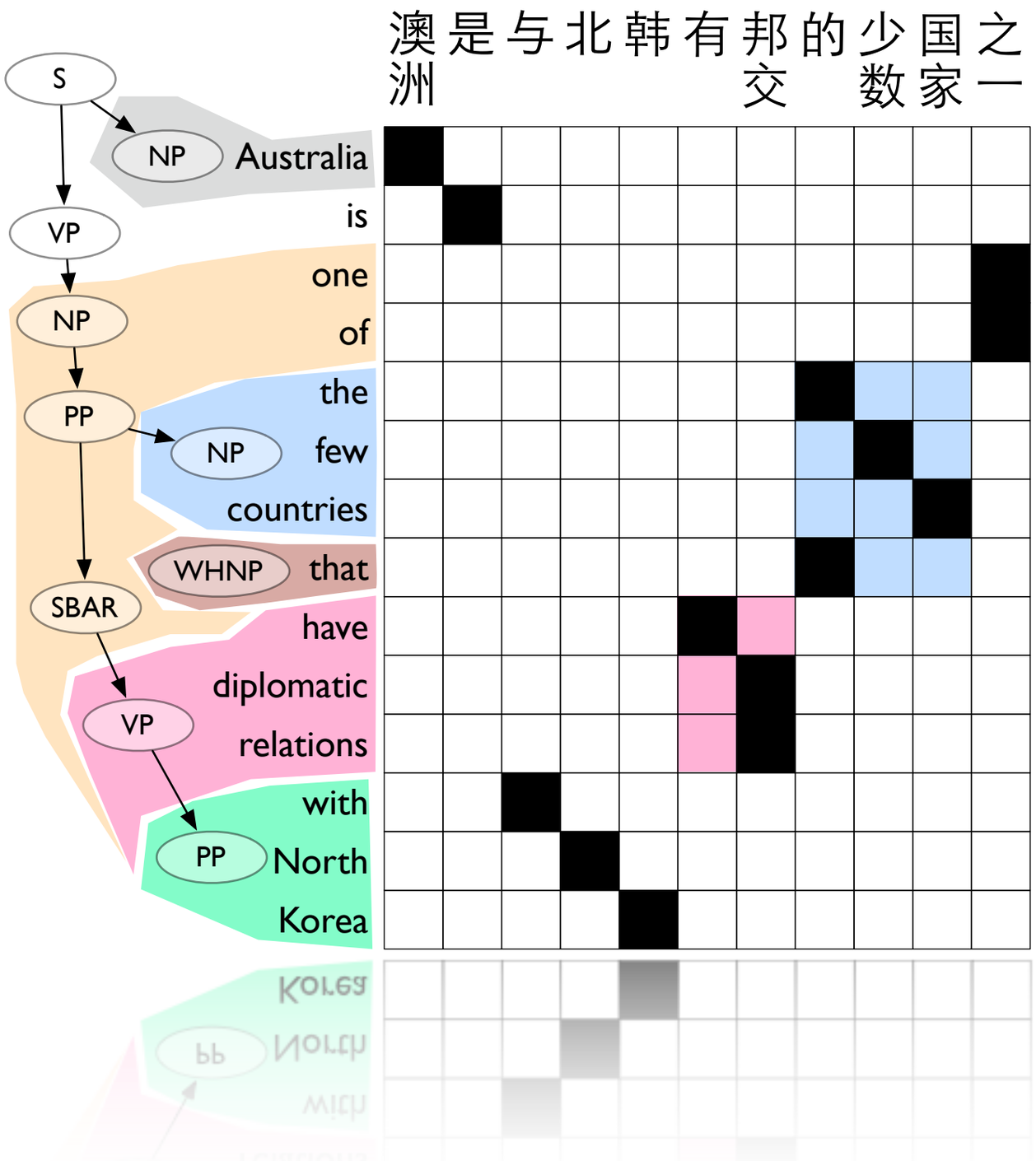
的少数国家

the few countries that

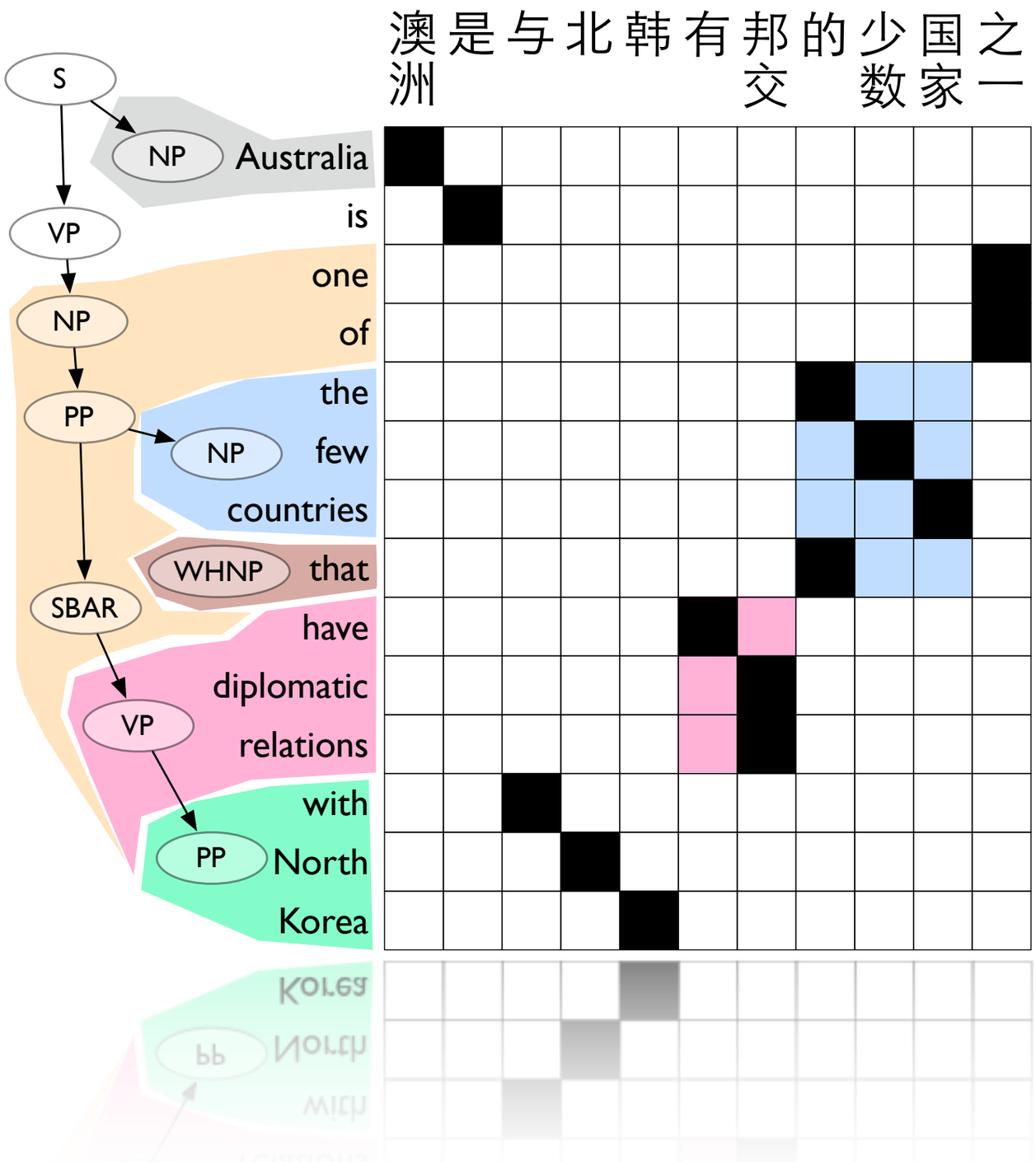
有邦交

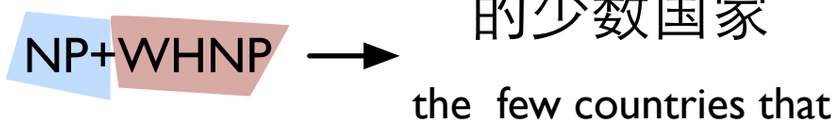
have diplomatic relations

SCFG Extraction

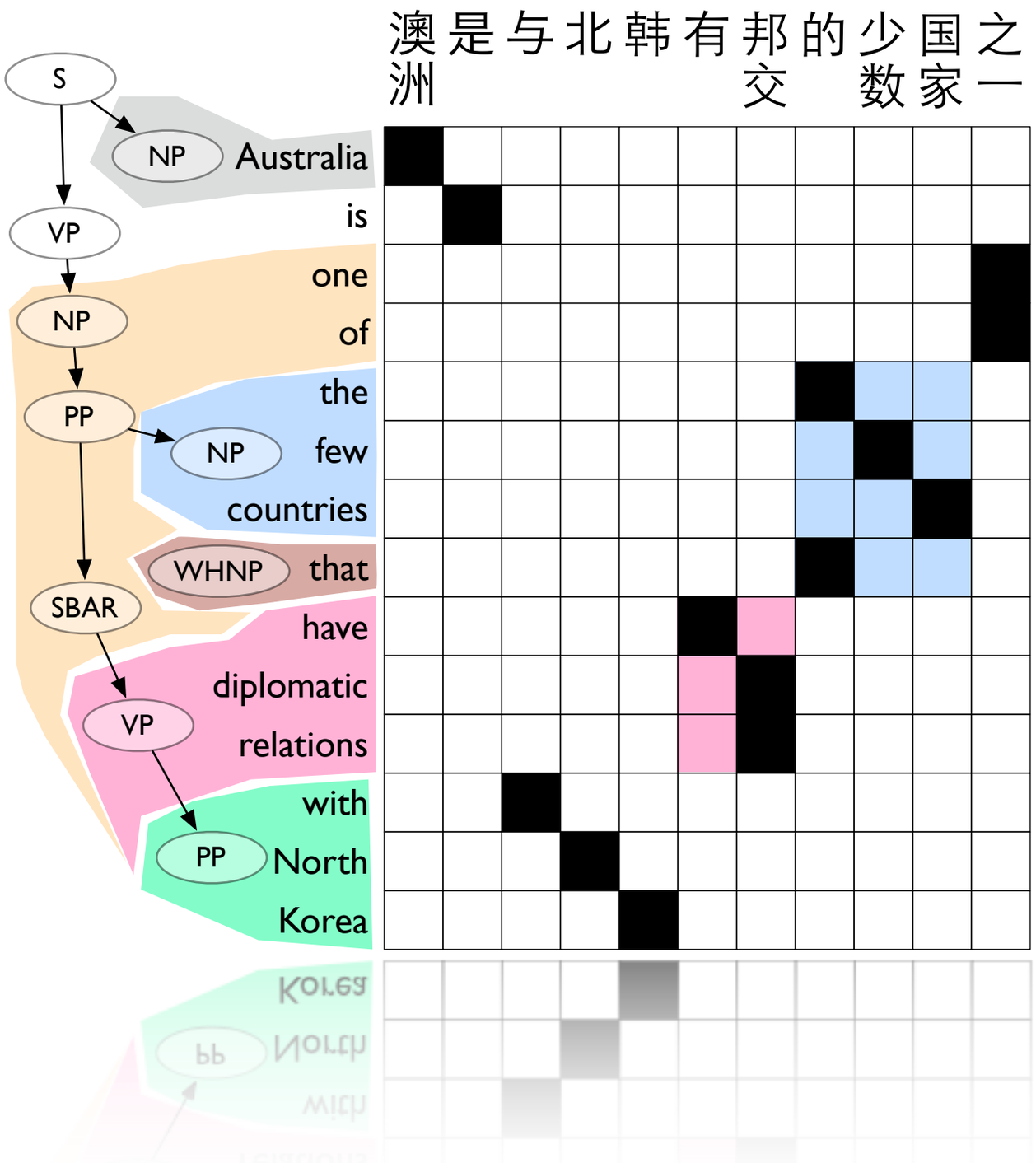


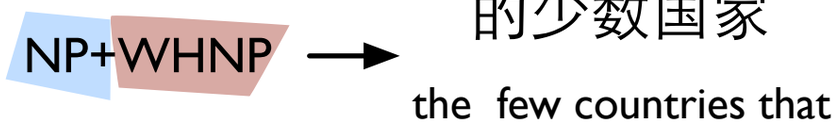
SCFG Extraction

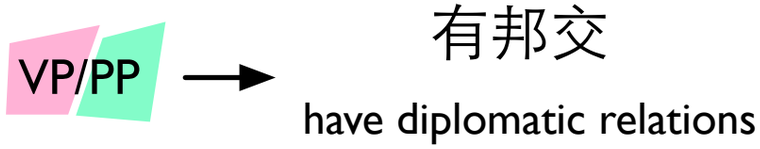




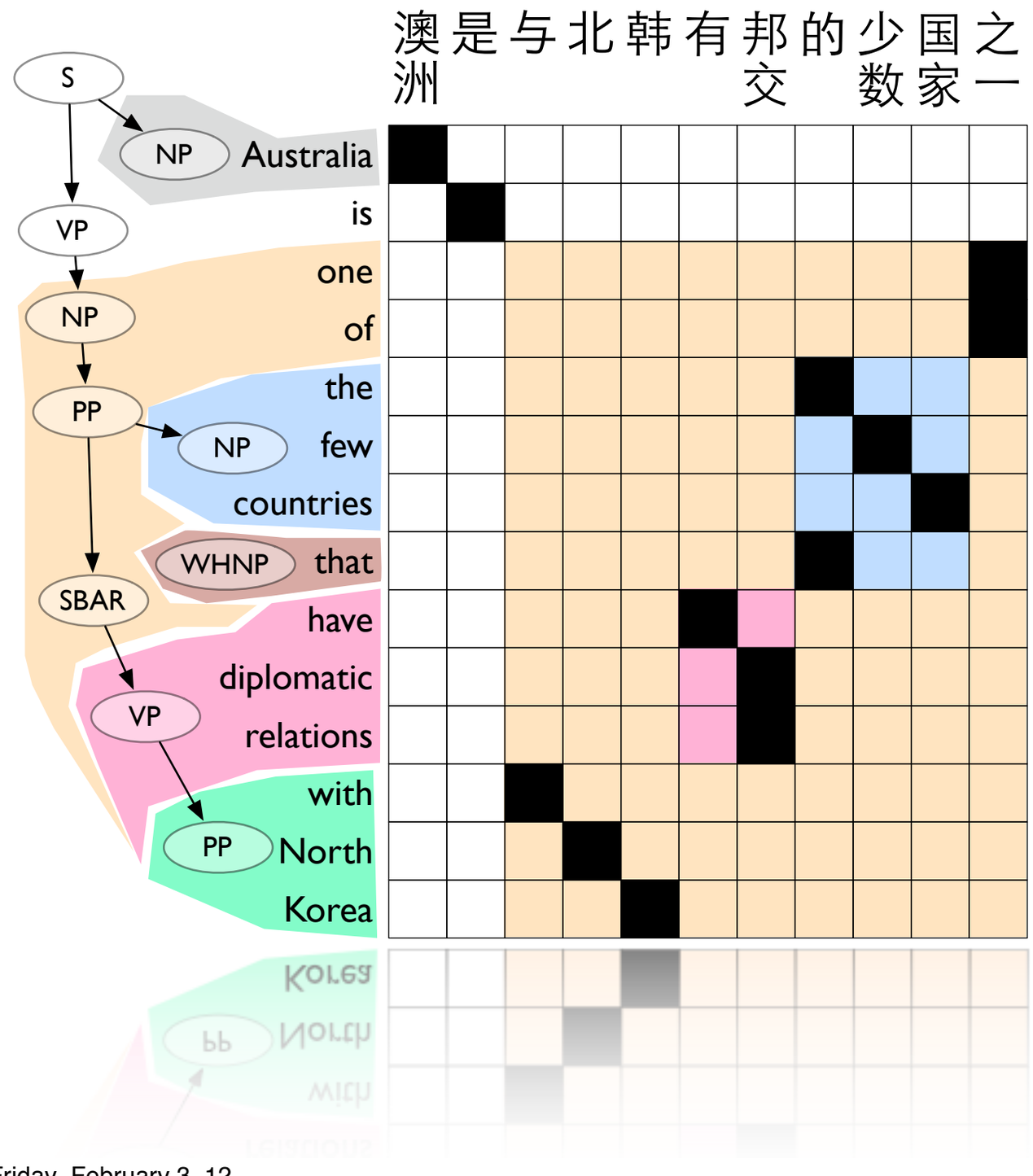
SCFG Extraction

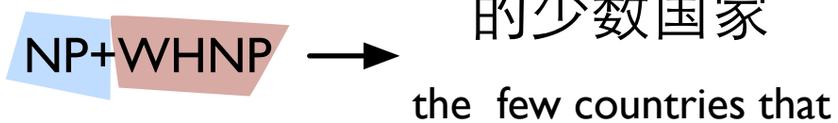


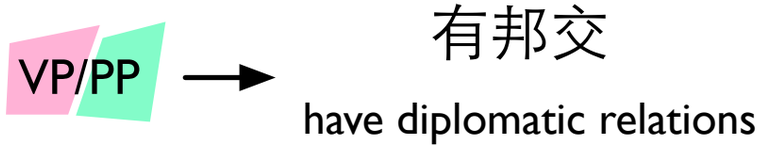




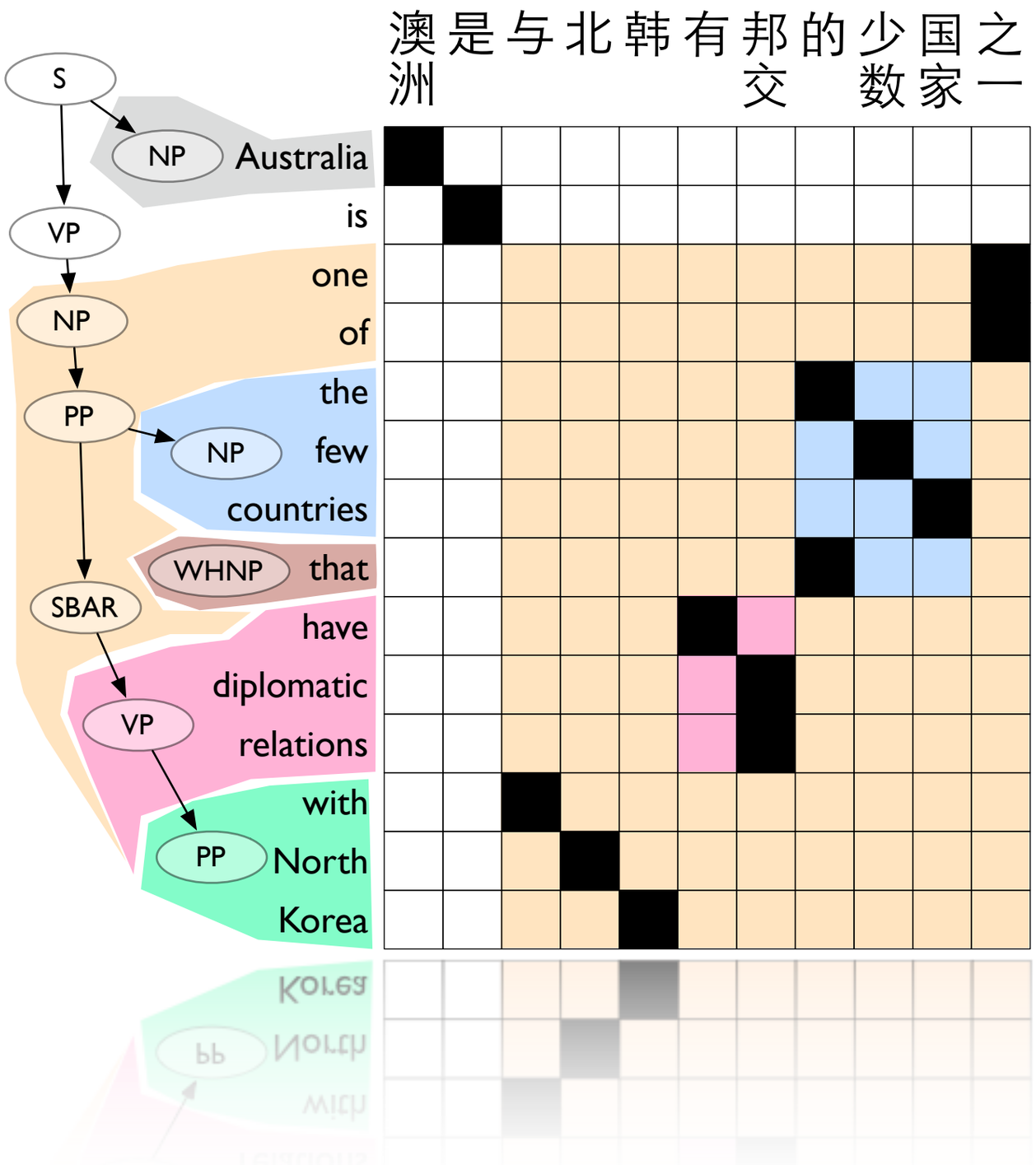
SCFG Extraction







SCFG Extraction



NP+WHNP → 的少数国家

the few countries that

VP/PP → 有邦交

have diplomatic relations

NP → 与北韩 VP/PP NP+WHNP 之一

one of NP+WHNP VP/PP with North Korea

Possessive Pivoting

NP → NP 's NN | le NN de NP

NP → the NN of NP | le NN de NP

Possessive Pivoting

NP → NP 's NN | le NN de NP

NP → the NN of NP | le NN de NP

Possessive Pivoting

NP → NP 's NN | le NN de NP

NP → the NN of NP | le NN de NP

Possessive Pivoting

NP → NP 's NN | le NN de NP

NP → the NN of NP | le NN de NP

Possessive Pivoting

NP → NP 's NN | le NN de NP

NP → the NN of NP | le NN de NP

Possessive Pivoting

NP → NP 's NN | le NN de NP

NP → the NN of NP | le NN de NP

combine to

Possessive Pivoting

NP → NP 's NN | le NN de NP

NP → the NN of NP | le NN de NP

combine to

NP → NP 's NN | the NN of NP

Our Machines!

Thrax

Toolkit for distributed extraction of translation and paraphrase grammars

Joshua

SCFG-based decoder for translation and paraphrasing

Goal: Expressiveness

Generalization vs. memorization

Capture paraphrastic transformations

Prefer general patterns over memorized phrases

Possessive rule	NP →	the NN of the NNP	|	the NNP's NN
	NP →	the NNS_1 made by the NNS_2	|	the NNS_2's NNS_1
Dative shift	VP →	give NN to NP	|	give NP the NN
	VP →	provide NP_1 to NP_2	|	give NP_2 NP_1
Adv./adj. phrase move	S/VP →	ADVP they VBP	|	they VBP ADVP
	S →	it is ADJP VP	|	VP is ADJP
Verb particle shift	VP →	VB NP up	|	VB up NP
Reduced relative clause	SBAR/S →	although PRP VBP that	|	although PRP VBP
	ADJP →	very JJ that S	|	JJ S
Partitive constructions	NP →	CD of the NN	|	CD NN
	NP →	all DT\NP	|	all of the DT\NP
Topicalization	S →	NP, VP.	|	VP, NP.
Passivization	SBAR →	that NP had VBN	|	which was VBN by NP
Light verbs	VP →	take action ADVP	|	to act ADVP
	VP →	to take a decision PP	|	to decide PP

Limitations

The SCFG formalism's power is limited:

Verb **morphology** is hard to generalize

Morphosyntactic paraphrases have to be memorized

Asynchronous effects cannot be generally represented

Text-to-Text Applications

Claim:

> Paraphrasing is suitable to tackle sentential text-to-text tasks

However:

> **Naive** application will fail, **adaptation** is necessary

Task Adaptation

SMT	T2T
Naive application of the MT machinery to the task	Task-specific adaptations

- Development data
- Objective function
- Feature set
- Grammar augmentations

Example Task: Sentence Compression

Reduce length of a sentence (#tokens) while retaining the meaning

Compression ratio: $\varphi = \frac{length_{compression}}{length_{original}}$

Example Task: Sentence Compression

Reduce length of a sentence (#tokens) while retaining the meaning

Compression ratio: $\varphi = \frac{length_{compression}}{length_{original}}$

Paraphrasing as a task and problem is of paramount importance to a multitude of applications in the field of NLP.

Example Task: Sentence Compression

Reduce length of a sentence (#tokens) while retaining the meaning

Compression ratio: $\varphi = \frac{length_{compression}}{length_{original}}$

Paraphrasing ~~as a task and problem is of paramount importance to a multitude of applications in the field of NLP~~.

Example Task: Sentence Compression

Reduce length of a sentence (#tokens) while retaining the meaning

Compression ratio: $\varphi = \frac{length_{compression}}{length_{original}}$

is awesome

Paraphrasing ~~as a task and problem is of paramount importance to a multitude of applications in the field of NLP~~.

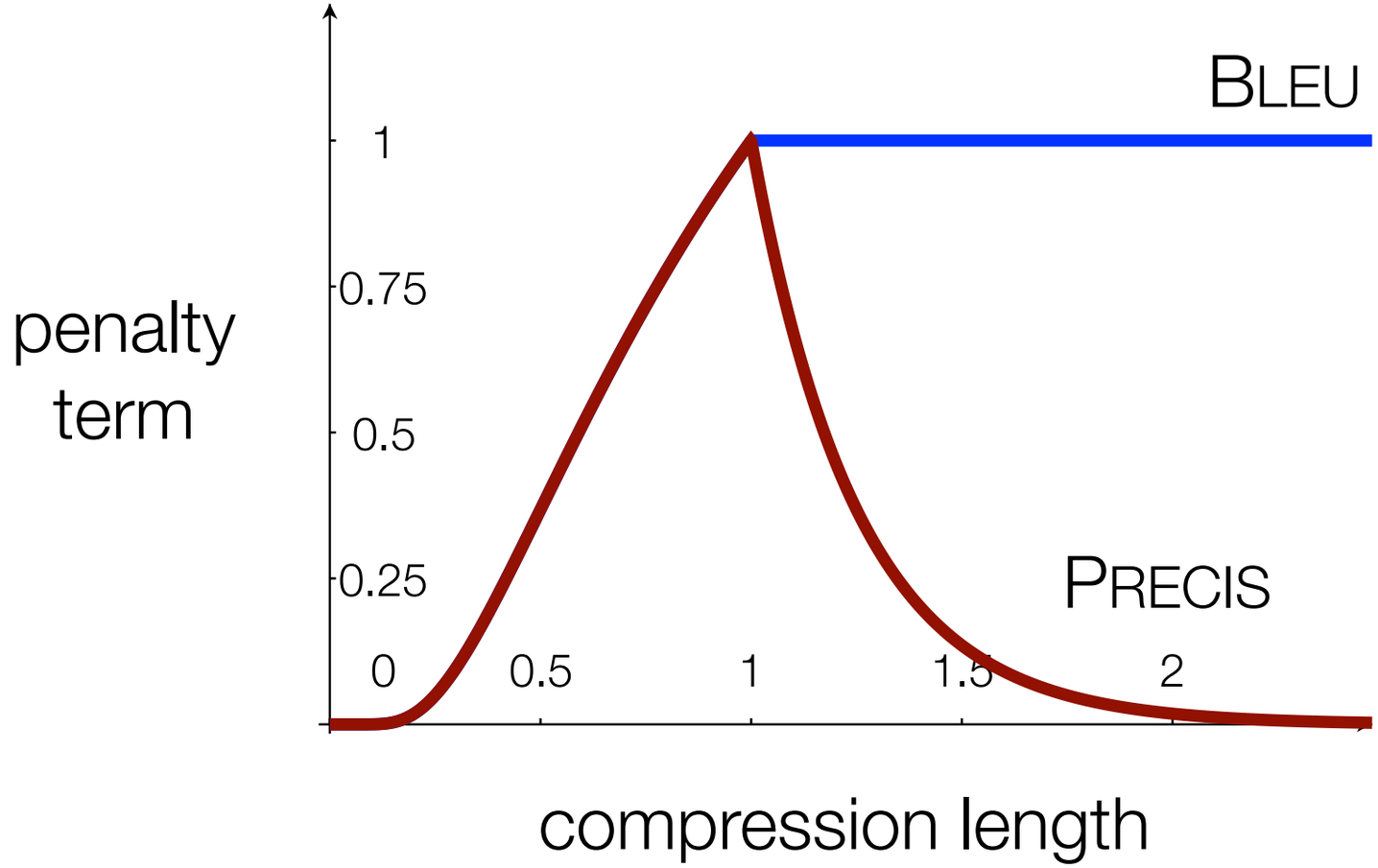
Development Data

SMT	T2T
English reference translations that are used to calculate BLEU for SMT.	Selected pairs of reference translations that significantly differ in length.

and he said that the project **will cover** the needs of the region **in the long term.**	17
he said the project **includes** all the district's **long-term** needs.	10

Objective Function

SMT	T2T
Optimized for English-to-English BLEU score. Causes self-paraphrasing.	Add a “verbosity penalty” to BLEU that allows a target compression ratio to be set.



Features

SMT	T2T
Phrasal and lexical probabilities quantify general paraphrase quality.	Features counting number of source and target words and the difference between them.

$$\text{VP} \rightarrow \text{NP was eaten by NN} \mid \text{NN ate NP}$$

$$p(e_1|e_2) = 0.1$$

$$c_{e_1} = 3$$

$$c_{e_2} = 1$$

$$c_{diff} = -2$$

Augmentations

SMT	T2T
It is not typical for additional task-specific rules to be added in the standard SMT pipeline.	Augment the grammar with deletion rules for specific POS (JJ, RB, DT) allowing for shorter compressions.

$$\text{JJ} \rightarrow \text{superfluous} \mid \varepsilon$$

$$\text{RB} \rightarrow \text{redundantly} \mid \varepsilon$$

$$\text{DT} \rightarrow \text{the} \mid \varepsilon$$

Comparing Compressions Is Tricky

Higher compression ratios are strongly correlated with better scores

Fair system comparison requires matching up compression ratios

Evaluation

- Human evaluation on MTurk
- Question: “How well do these sentences retain the meaning of the original.”
- Judging grammaticality and meaning on a 1-5 scale

Baseline Systems

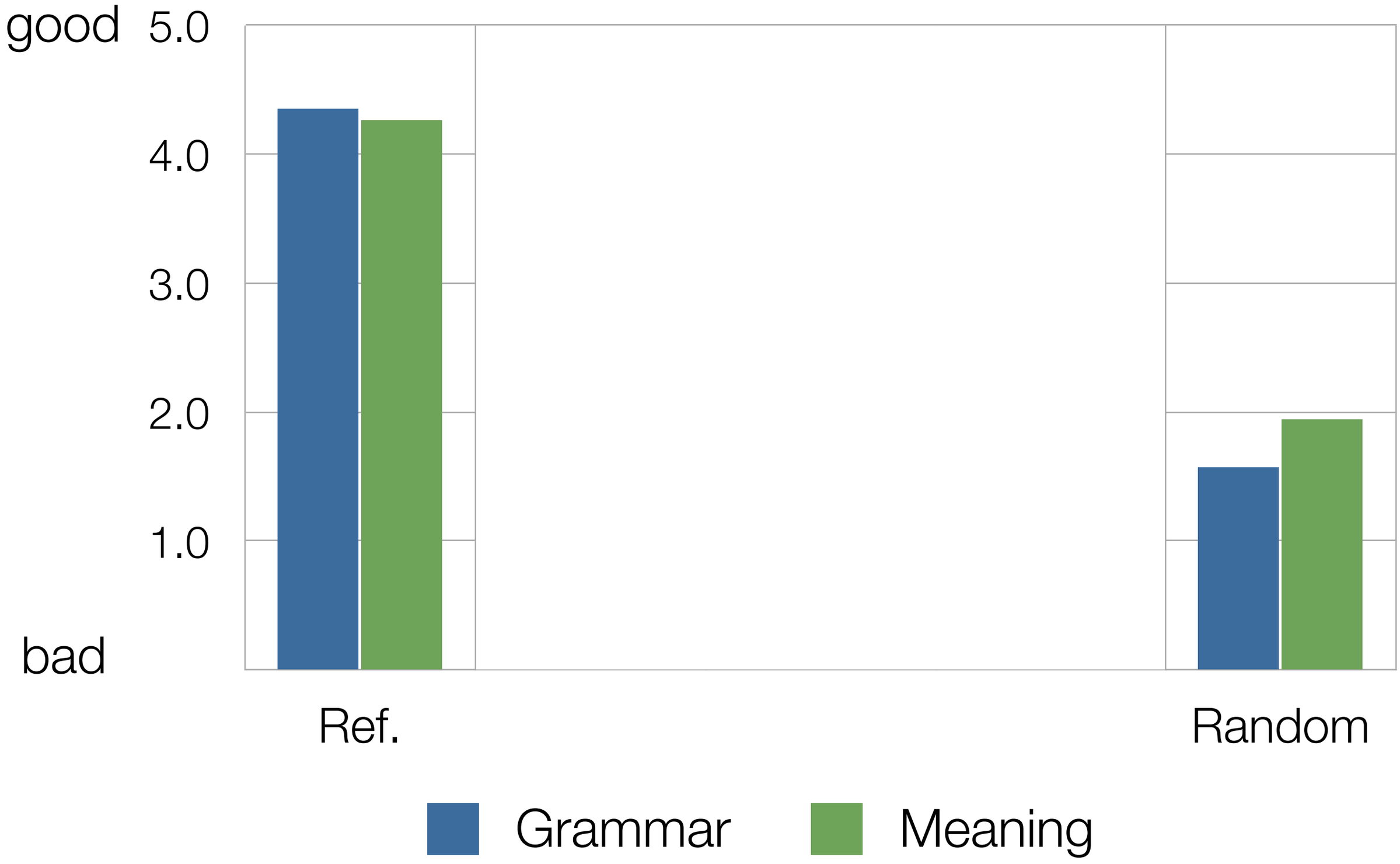
ILP, Clarke & Lapata ’08

Uses a set of constraint features to find best deletions via ILP

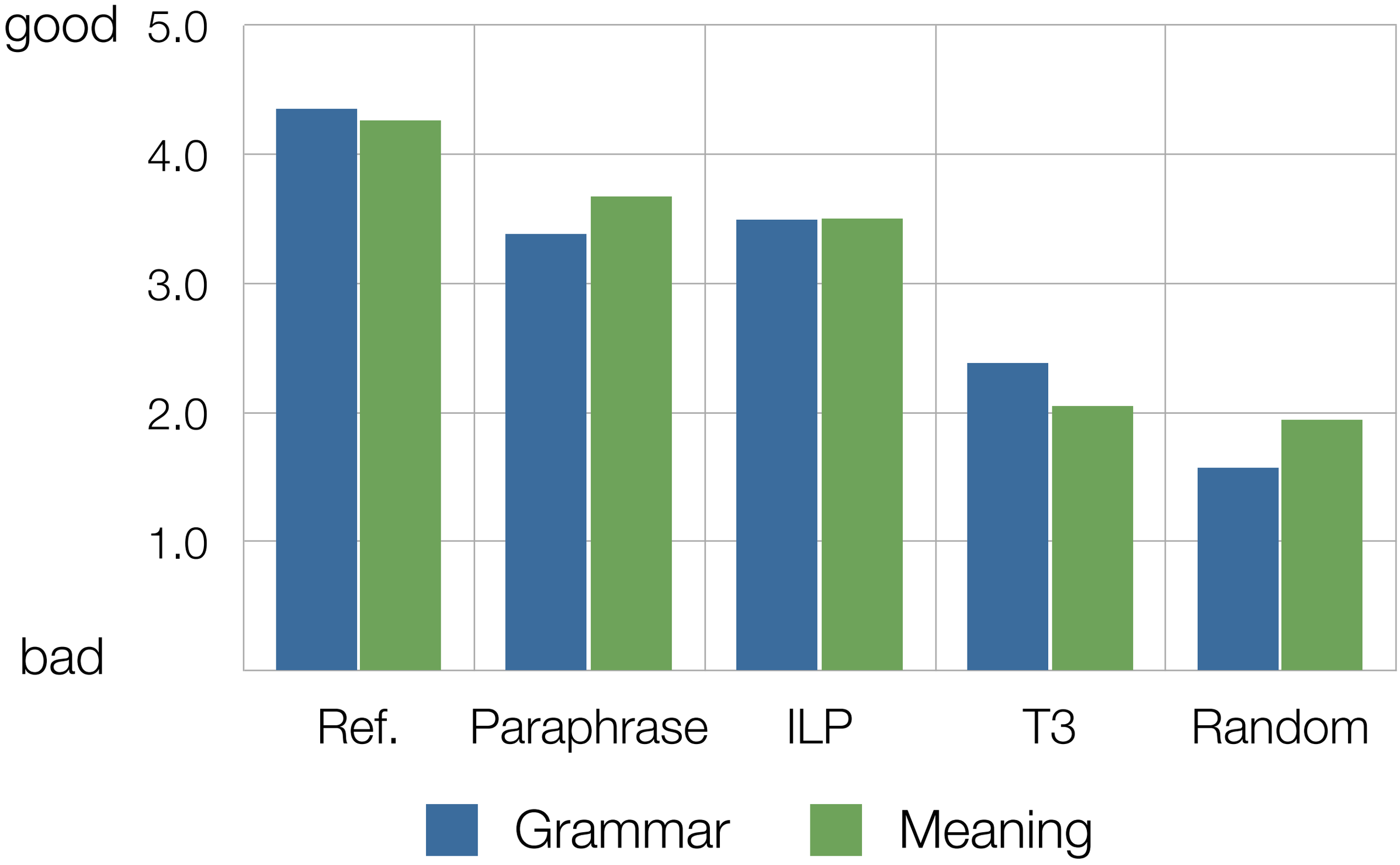
T3, Cohn & Lapata ’07

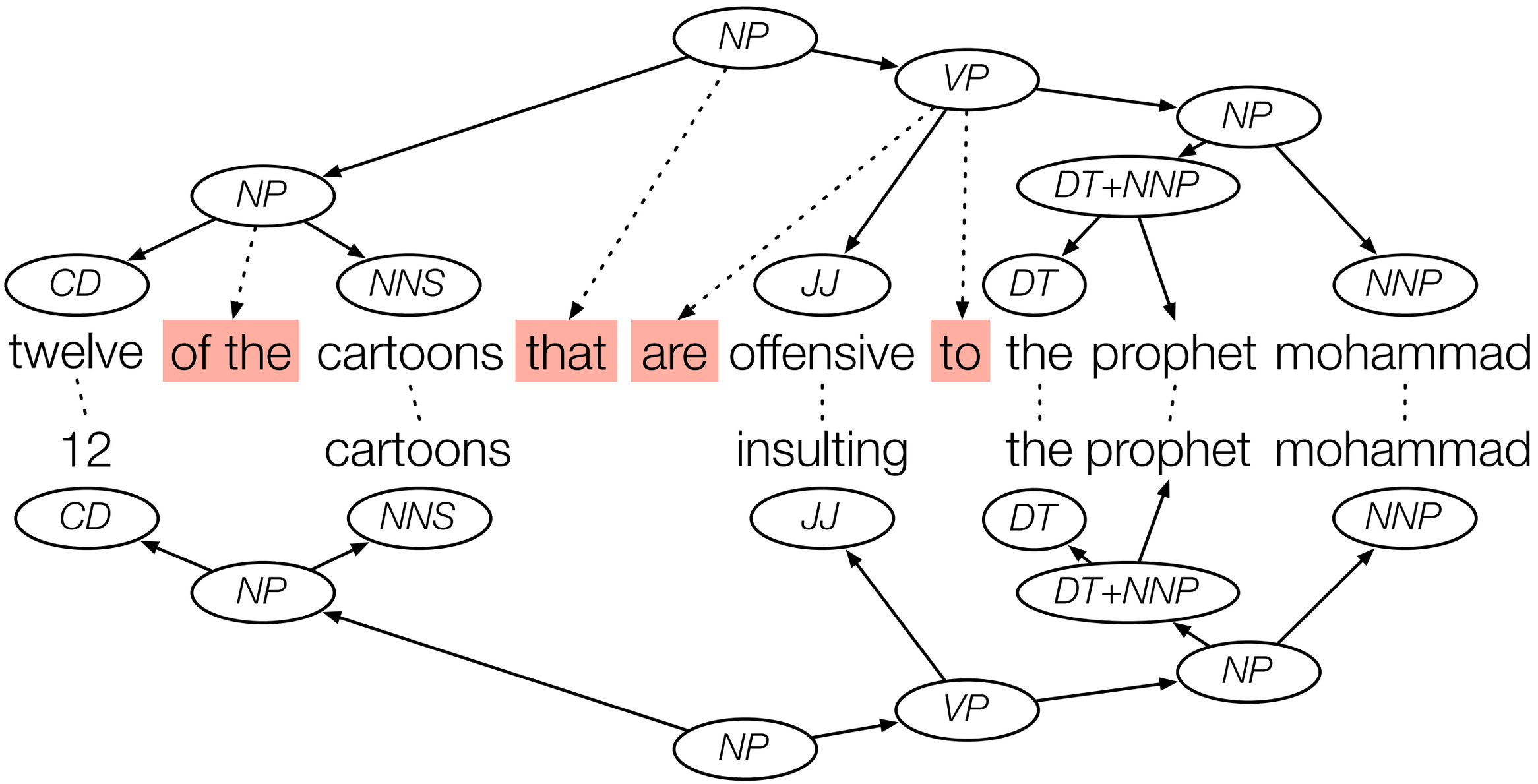
Uses an STSG to delete constituents

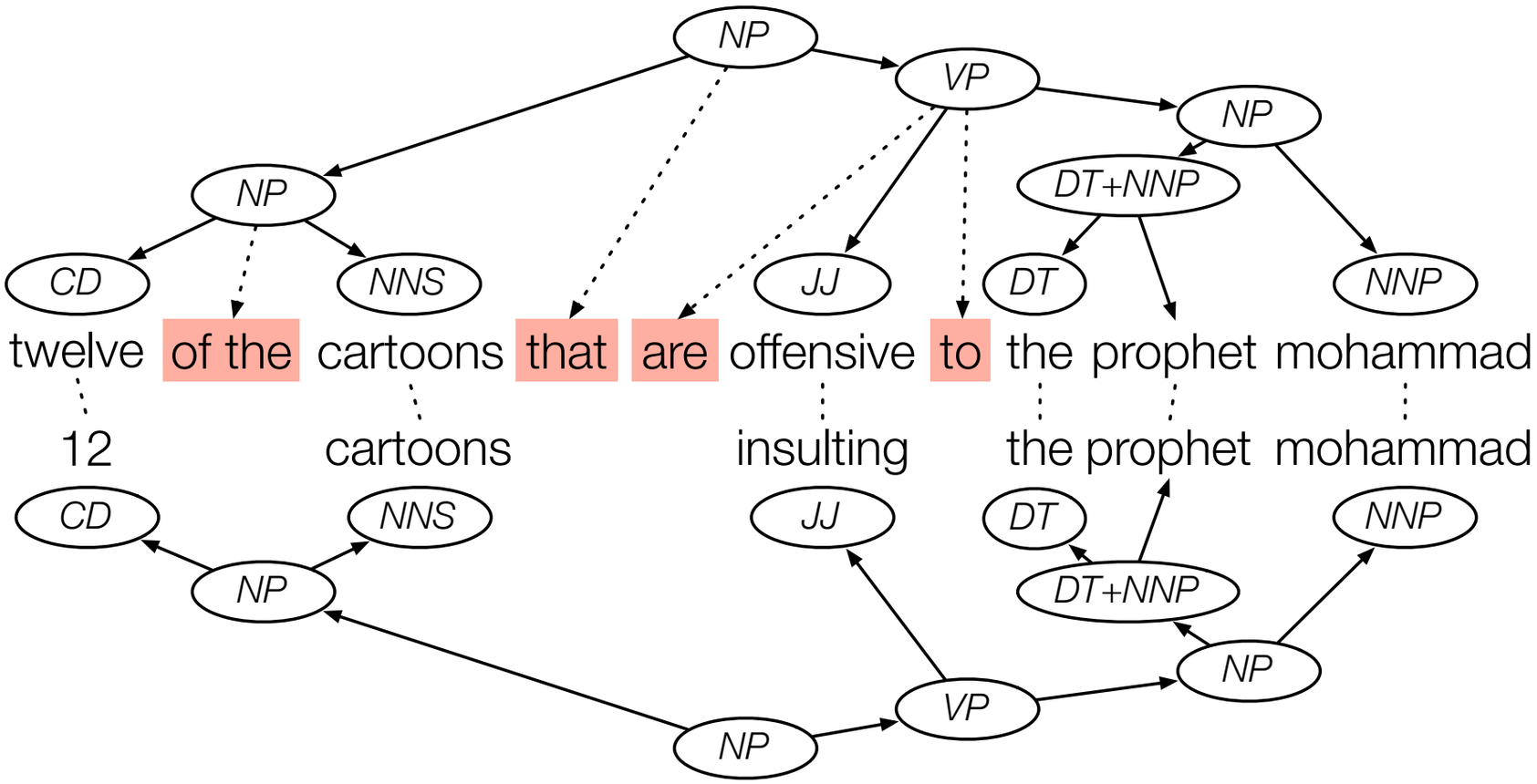
Results



Results


good
5.0
4.0
3.0
2.0
1.0
bad
Ref.
Paraphrase
ILP
T3
Random
Grammar
Meaning


NP
VP
NP
NP
DT+NNP
CD
NNS
JJ
DT
NNP
twelve
of the
cartoons
that
are
offensive
to
the
prophet
mohammad
12
cartoons
insulting
the prophet
mohammad
CD
NNS
JJ
DT
NNP
NP
DT+NNP
NP
NP
VP



Lexical paraphrase:

$\text{JJ} \rightarrow \text{offensive} \mid \text{insulting}$

Reduced relative clause:

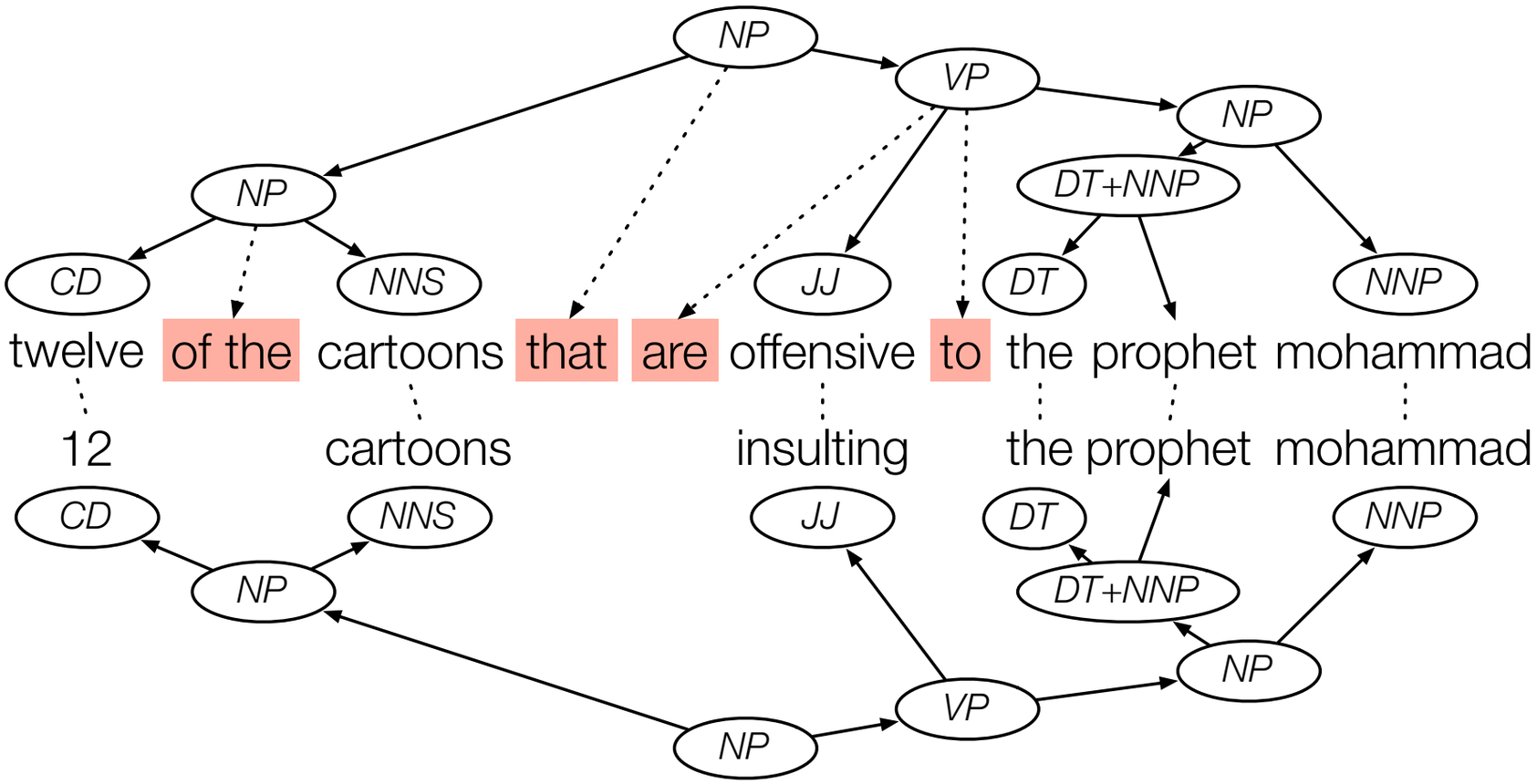
$\text{NP} \rightarrow \text{NP that VP} \mid \text{NP VP}$

Pred. adjective copula deletion:

$\text{VP} \rightarrow \text{are JJ to NP} \mid \text{JJ NP}$

Partitive construction:

$\text{NP} \rightarrow \text{CD of the NNS} \mid \text{CD NNS}$



JJ → beleidigend | offensive
JJ → beleidigend | insulting

NP → NP die VP | NP VP
NP → NP die VP | NP that VP

VP → sind JJ für NP | are JJ to NP
VP → sind JJ für NP | JJ NP

NP → CD der NNS | CD of the NNS
NP → CD der NNS | CD NNS

Four Pillars

Extraction - extract paraphrases from data

Representation - learn compact and rich paraphrase grammars

Recognition - detect semantic identity and entailment

Generation - use paraphrases to generate text

This Work

Extraction & Representation

Extended large-scale paraphrase acquisition from bitexts to syntactic paraphrases

Generation

Introduced a straightforward and effective adaptation framework

Ongoing Work

Extraction & Generation

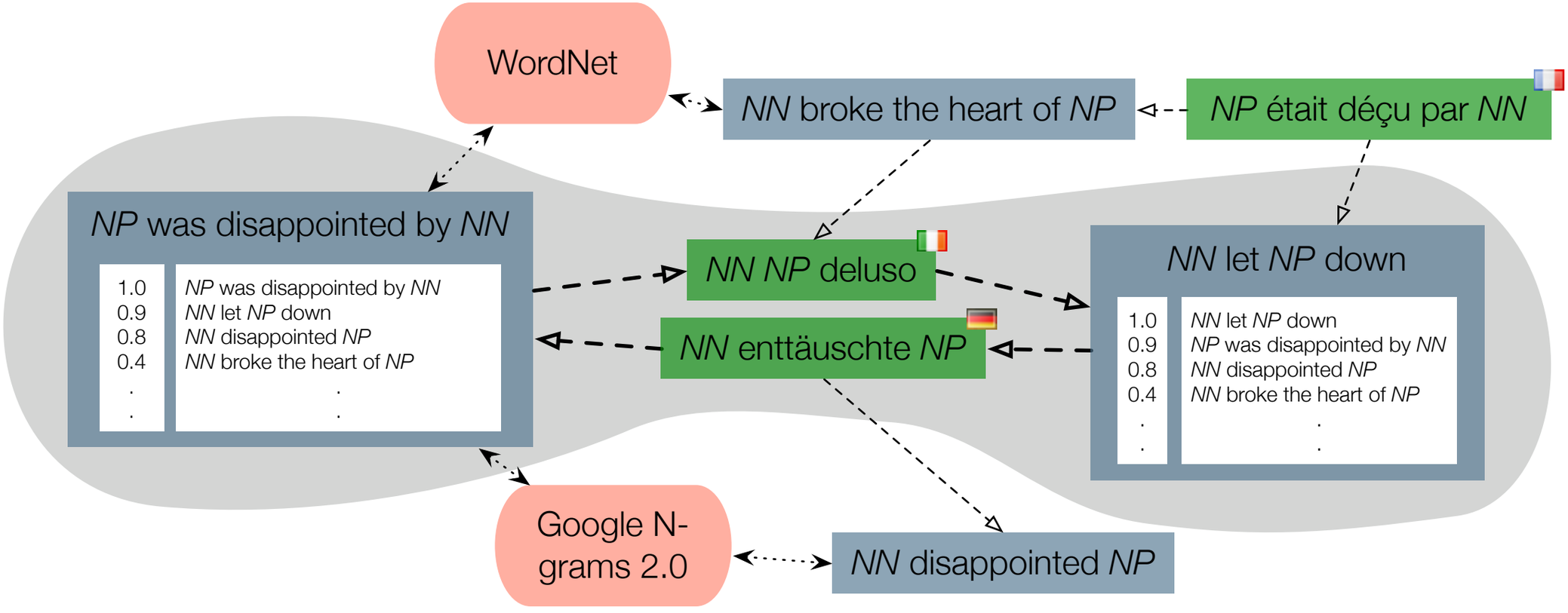
Include **distributional similarity** features in decoding process:

- Allows for scoring across phrase boundaries
- Contextualization in rule applications

Thank you.

Future Work - Extraction

Combine alignments and distributional similarity in a belief-propagation approach



Future Work - Representation

Move towards indirect “semantic” grammars and a two-stage decoding process

Monolingual Synchronous Grammar

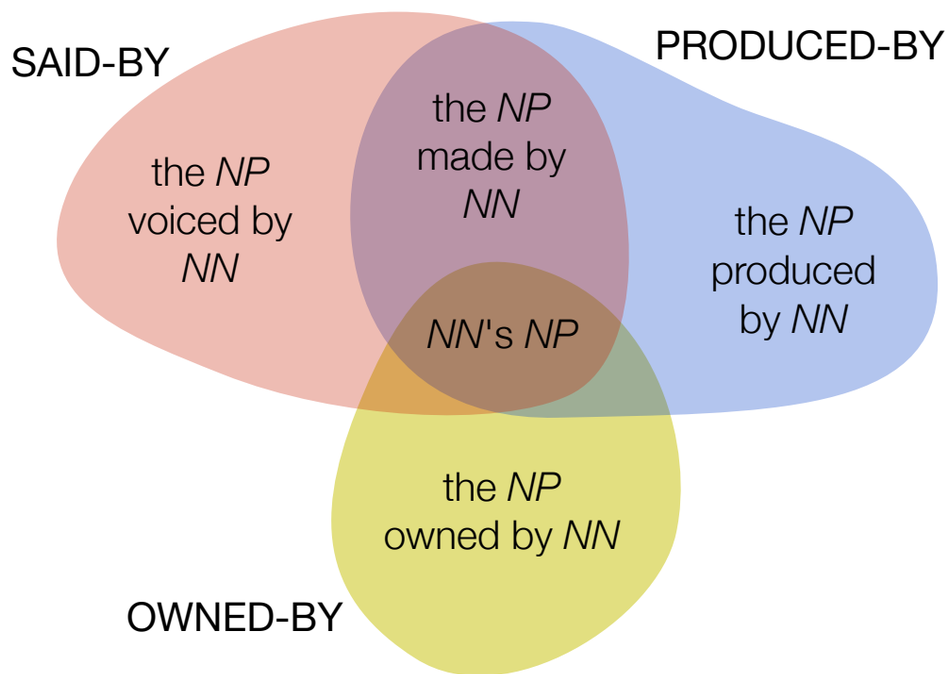
the *NP* voiced by *NN* → the *NP* made by *NN*
the *NP* voiced by *NN* → *NN*'s *NP*

the *NP* made by *NN* → the *NP* voiced by *NN*
the *NP* made by *NN* → *NN*'s *NP*
the *NP* made by *NN* → the *NP* produced by *NN*

the *NP* produced by *NN* → the *NP* made by *NN*
the *NP* produced by *NN* → *NN*'s *NP*

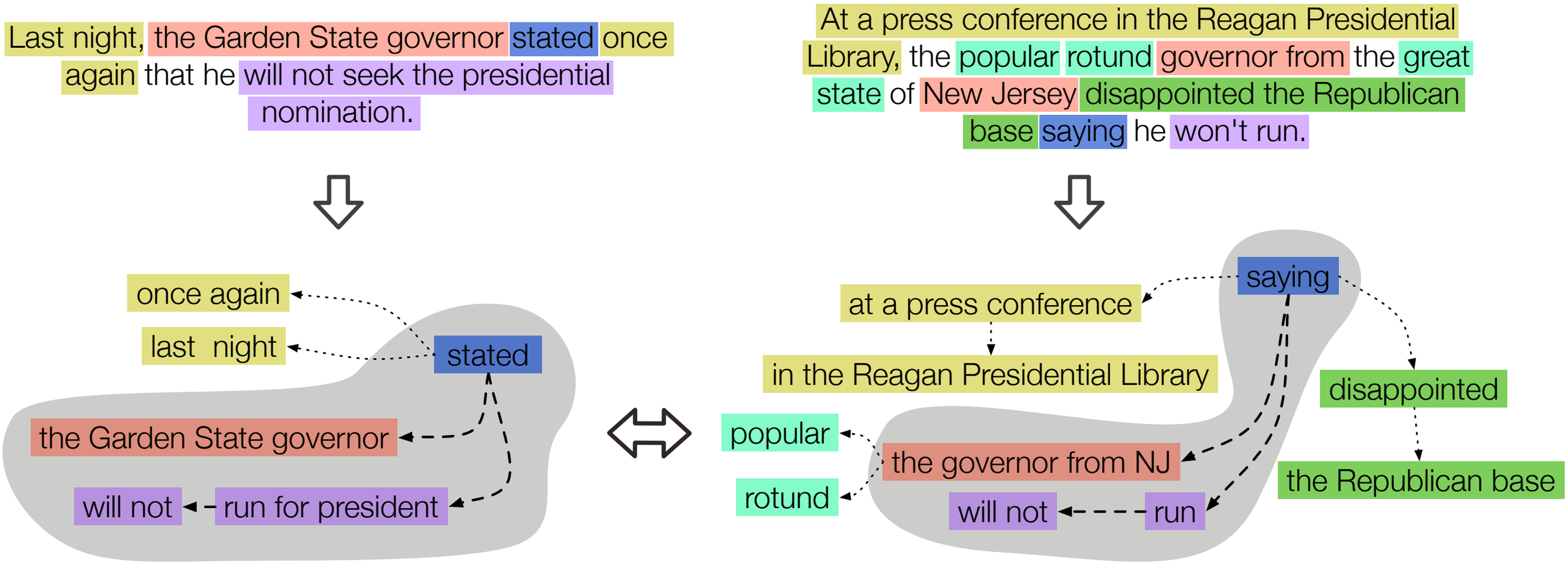
NN's *NP* → the *NP* made by *NN*
NN's *NP* → the *NP* owned by *NN*
NN's *NP* → the *NP* produced by *NN*

Semantic Grammar



Future Work - Recognition

Cast testing for partial semantic overlap as a parse intersection problem



Future Work - Generation

Document-level parsing and incremental generation