Recent research in multilingual natural language processing

Philip Resnik
University of Maryland

Using parallel bilingual text for WSD

The Case of Parallel Translations

observable *surface representation*

I got a wedding gift for my brother

implicit process

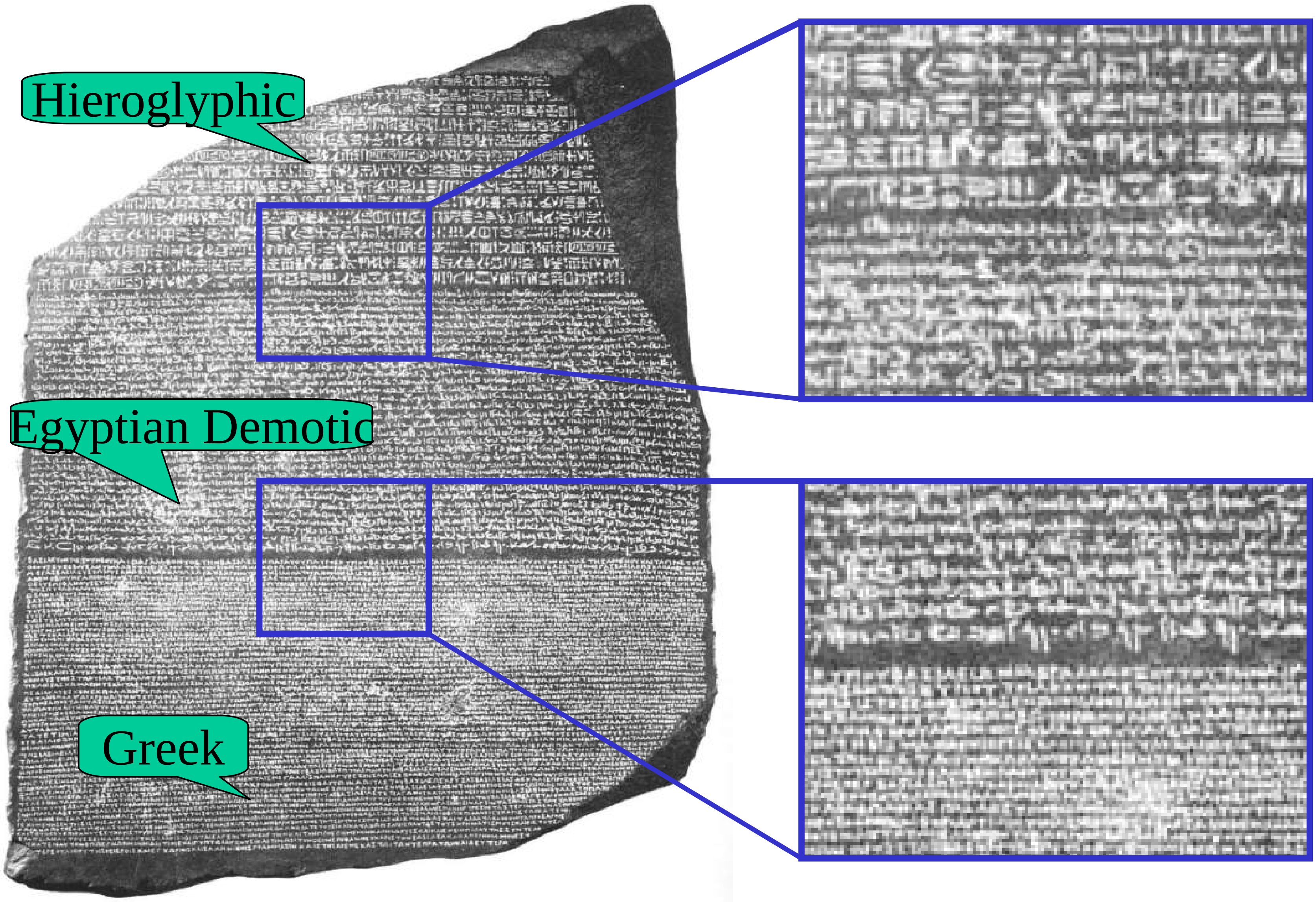
meaning

implicit process

nik	***nire***	***anaiari***	***ezkontza***	***opari***	***bat***	***erosi nion***
I-erg	MY	BROTHER-dat	WEDDING	GIFT	a	BUY-past

observable *surface representation*

This idea is not without precedent.



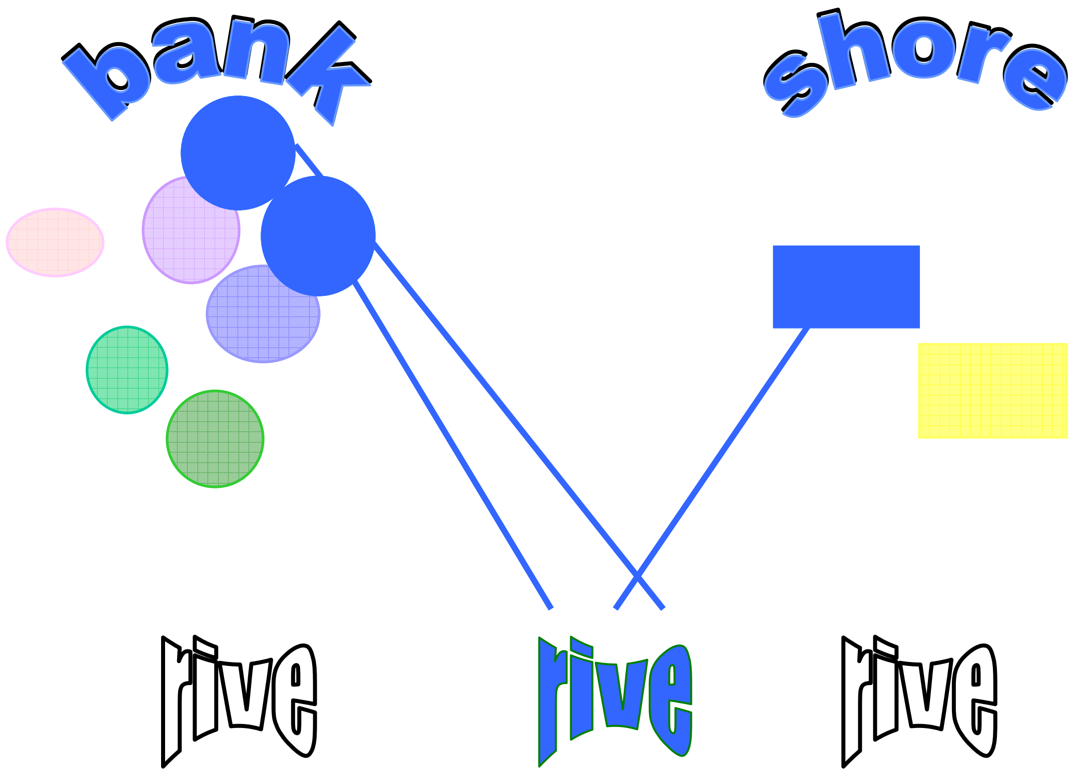
Key Claim

- Can we *recover* the hidden common meaning?
  - Probably not.

- Can we *exploit* the hidden common meaning?
  - Yes. And this will let us take supervised approaches to naturally occurring, *unannotated* data, helping to solve *monolingual* problems.

Sense Foregrounding

Observation: If two or more words are translated into the same word in a second language, then they often share some element of meaning



WSD Using Bilingual Text

- Collect English words sharing hidden meaning
- Identify senses closest to the shared meaning
- Label the words with those explicit senses

WSD Approach

Token Aligned Parallel Corpora

He has a house by the river bank.

Il a une maison par la rive de fleuve.

Create ***source sets***

RIVE : { BANK, SHORE, RIVERSIDE }

BANQUE : { BANK, BROKERAGE }

Assign senses to source type sets

RIVE : { BANK$_3$, SHORE$_1$, RIVERSIDE$_2$ }

BANQUE : { BANK$_2$, BROKERAGE$_1$ }

Assign senses to corpora tokens

He has a house by the river bank$_3$

Il a une maison par la rive de fleuve.

Note: French example is from MT output.

Collecting Words Sharing Hidden Meaning

I walked barefoot by the shore

J'ai marché nu-pieds par la rive

He has a house by the river bank

Il a une maison par la rive de fleuve

Target Set

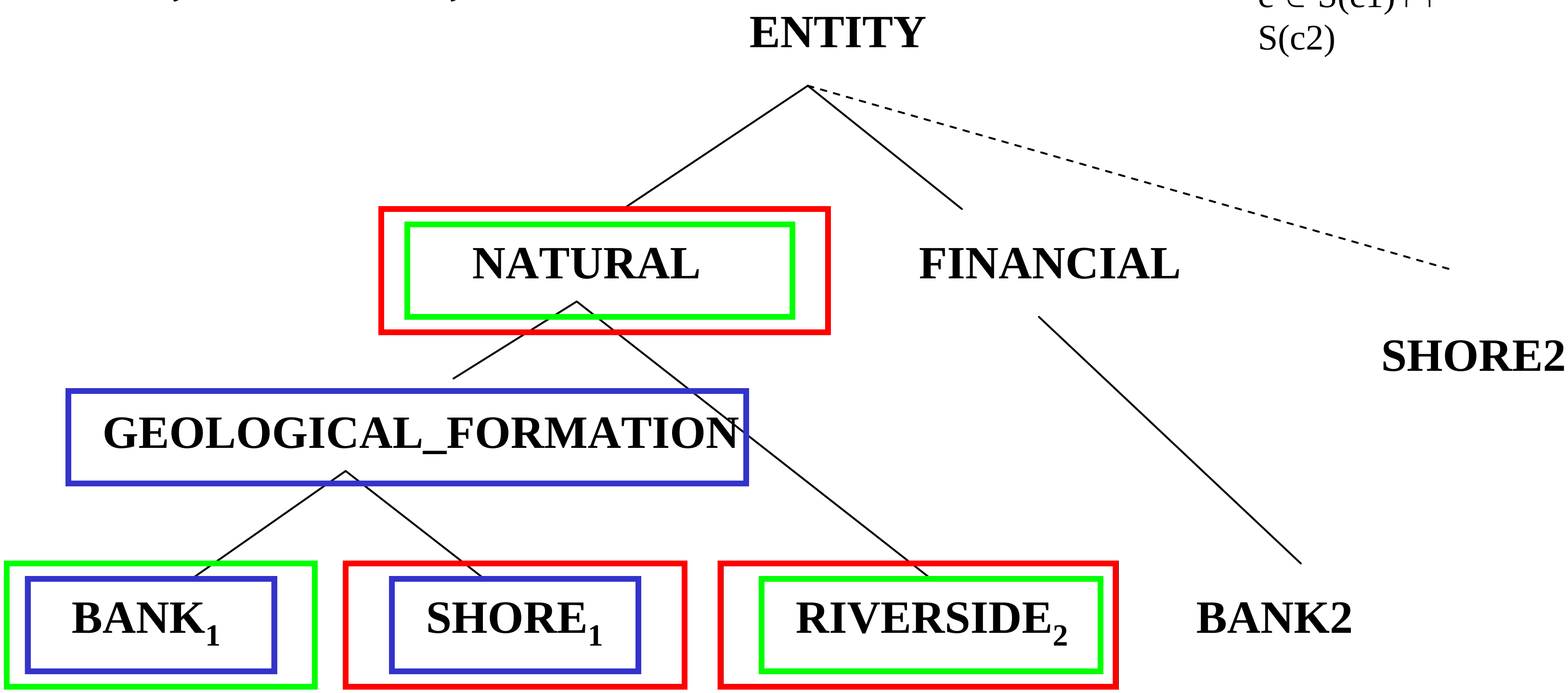
RIVE : { BANK , SHORE , RIVERSIDE }

Reminder to French speakers: this is machine-translated text

Identifying Senses

{BANK, SHORE, RIVERSIDE}

$$sim(c1,c2) = \operatorname*{argmax}_{c \in S(c1) \cap S(c2)} -\log Pr(c)$$



{$BANK_1$, $SHORE_1$, $RIVERSIDE_2$}

Labeling with Explicit Senses

RIVE : { BANK$_1$, SHORE$_1$, RIVERSIDE$_2$ }

I walked barefoot by the shore$_1$

J'ai marché nu-pieds par la rive

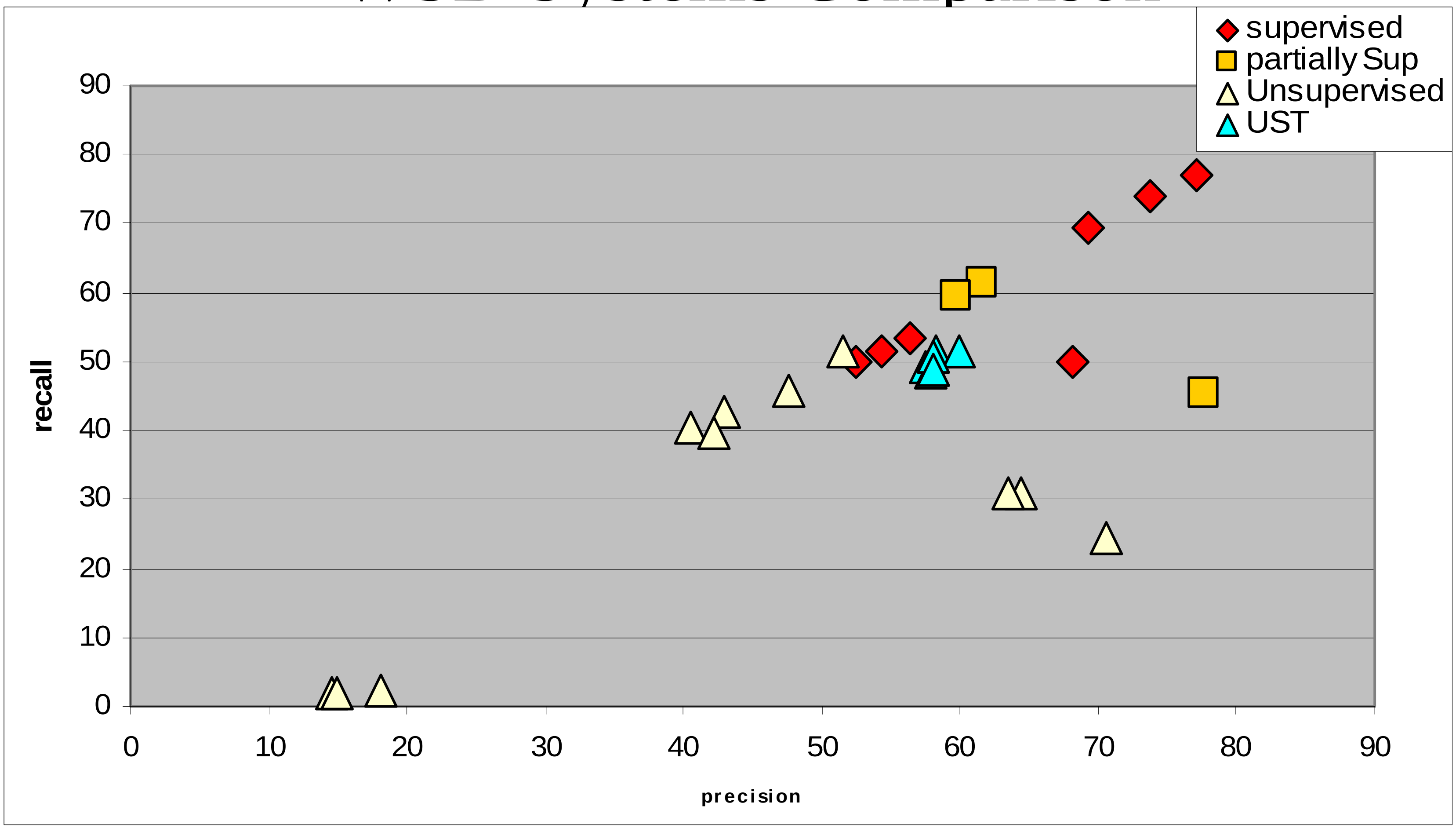
He has a house by the river bank$_1$

Il a une maison par la rive de fleuve

Did I mention that the French here is <u>machine translation</u> output?

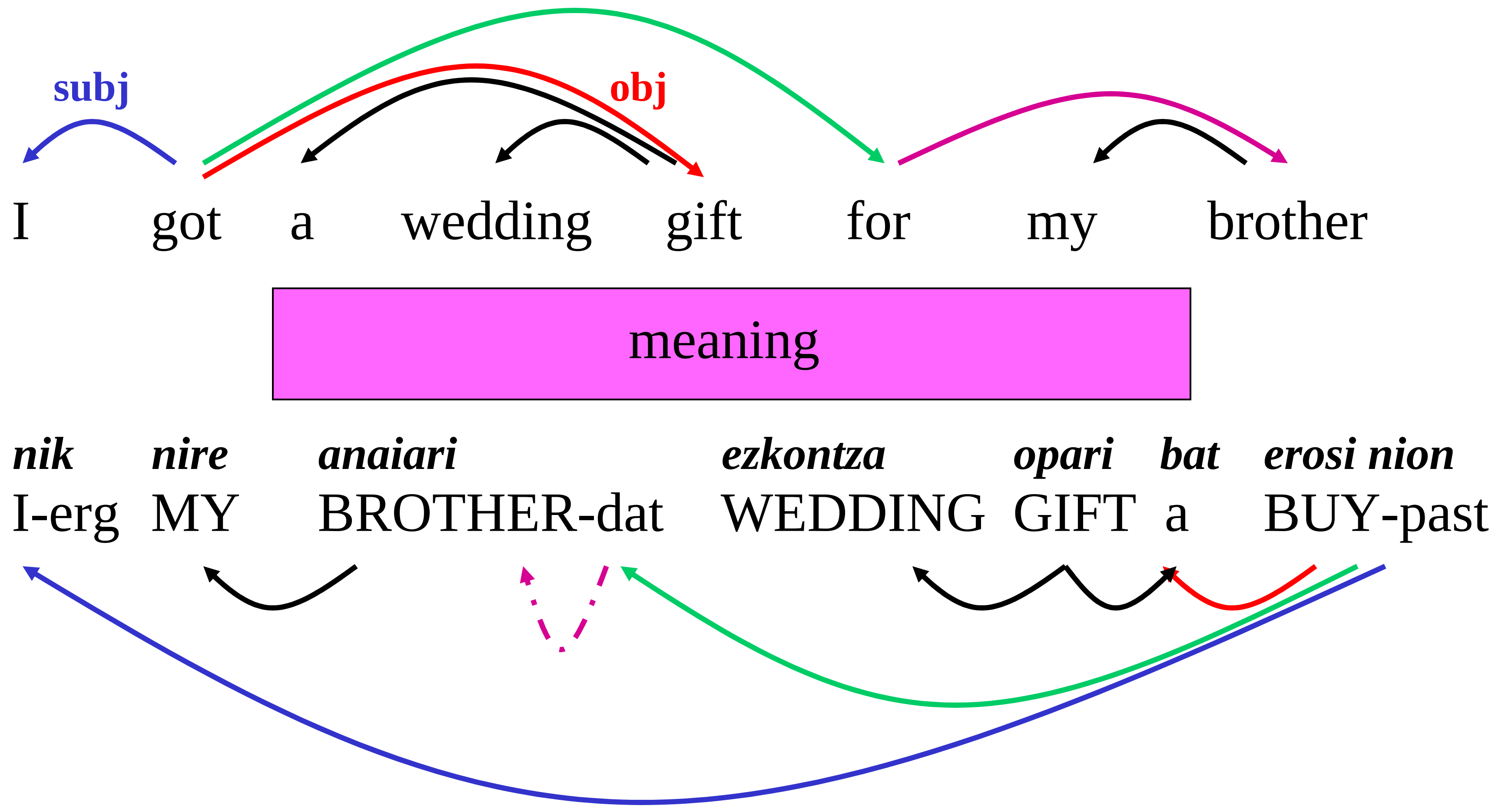
UST Evaluation

WSD Systems Comparison

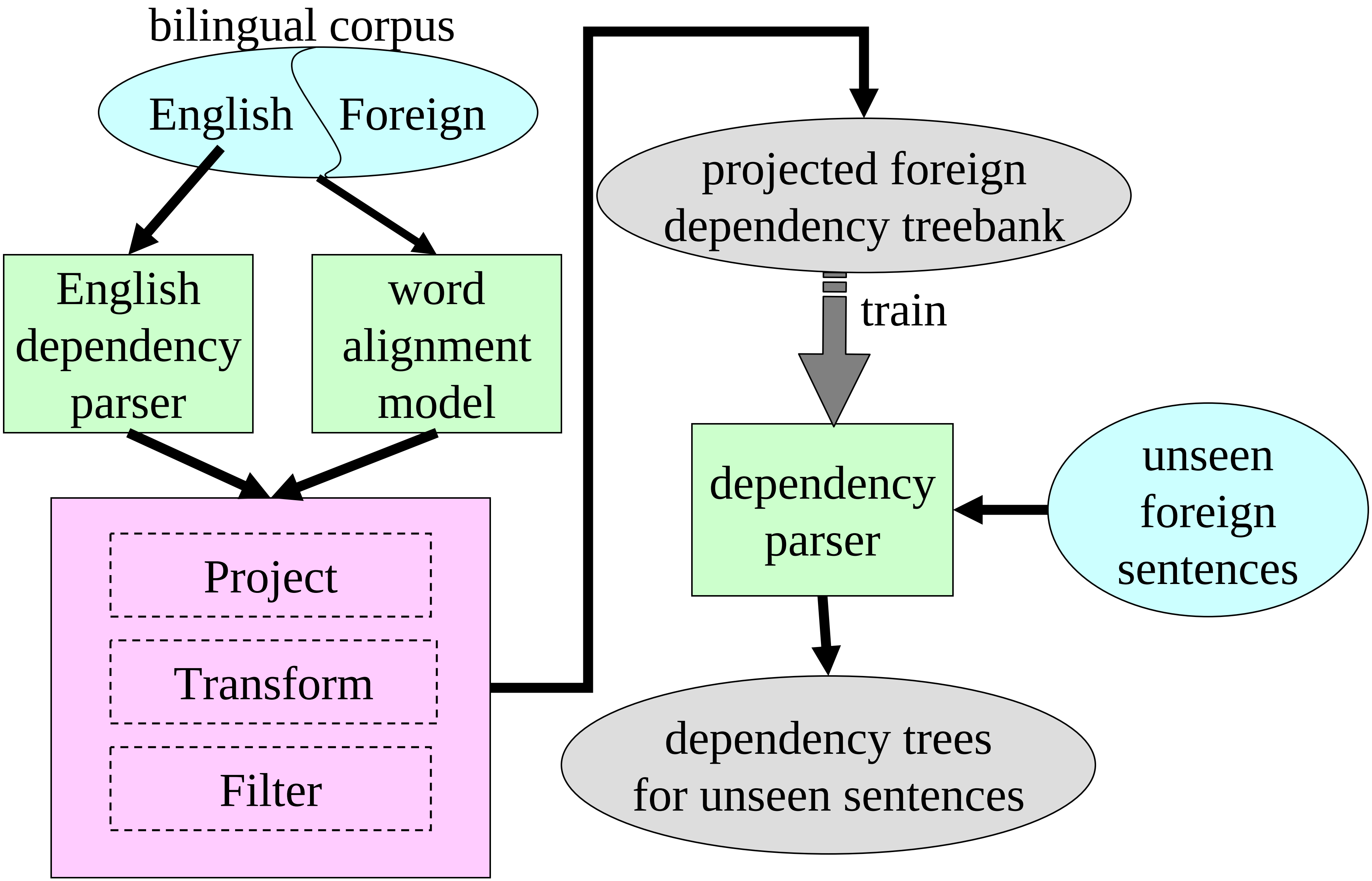


Using parallel text to bootstrap monolingual parsers for low-resource languages

Annotation Projection with the Direct Correspondence Assumption (DCA)



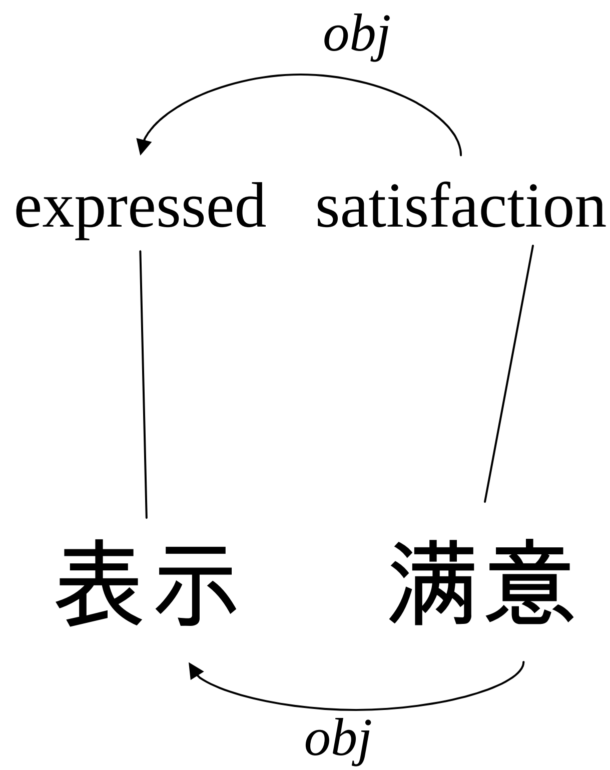
Dependency Projection Framework



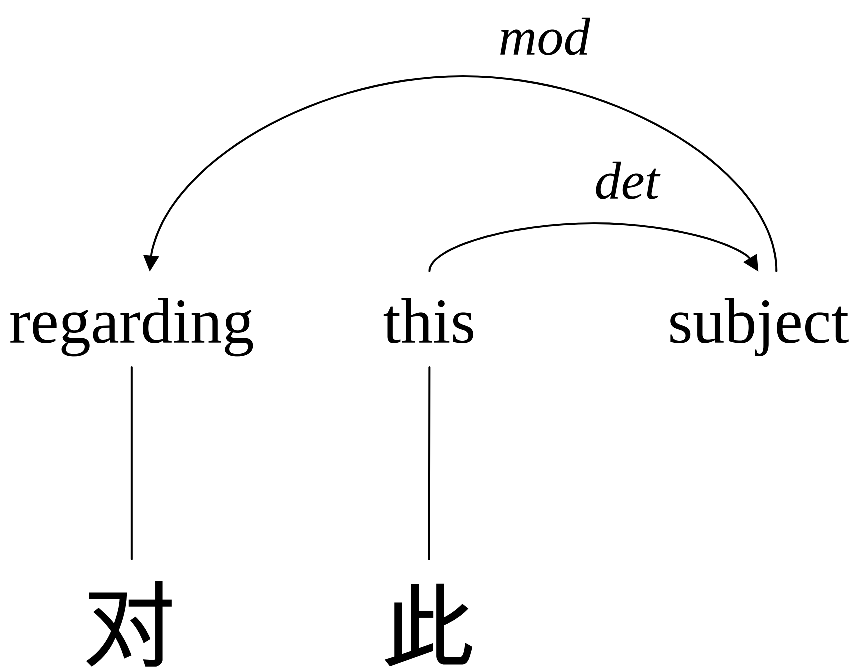
Direct Projection Algorithm

- If there is a syntactic relationship between two English words, then ensure that the same syntactic relationship also exists between their corresponding words in the second language.

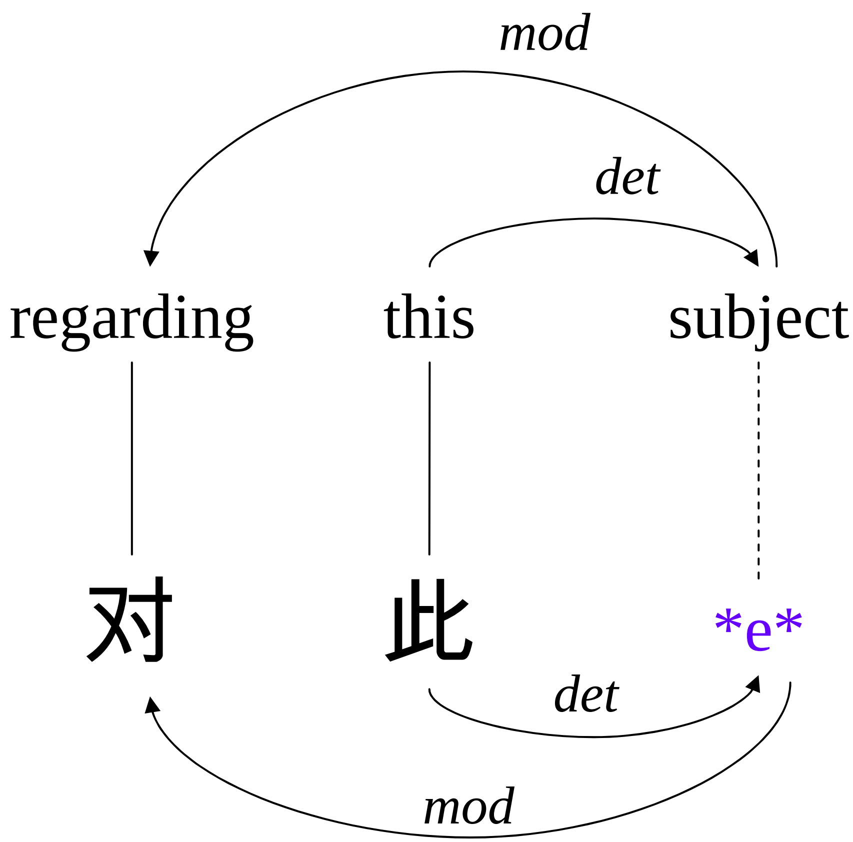
Unproblematic Cases



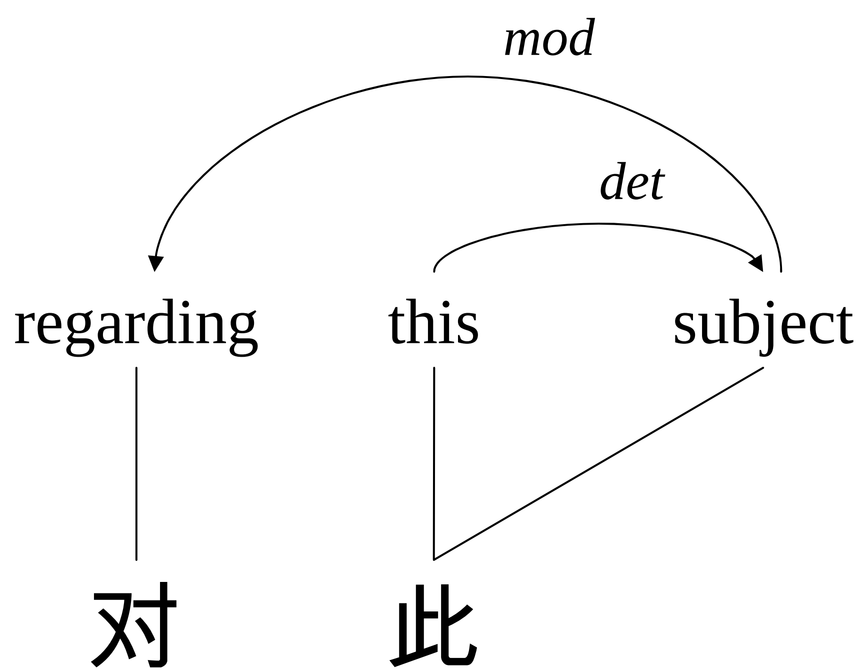
Problematic Case: Unaligned English



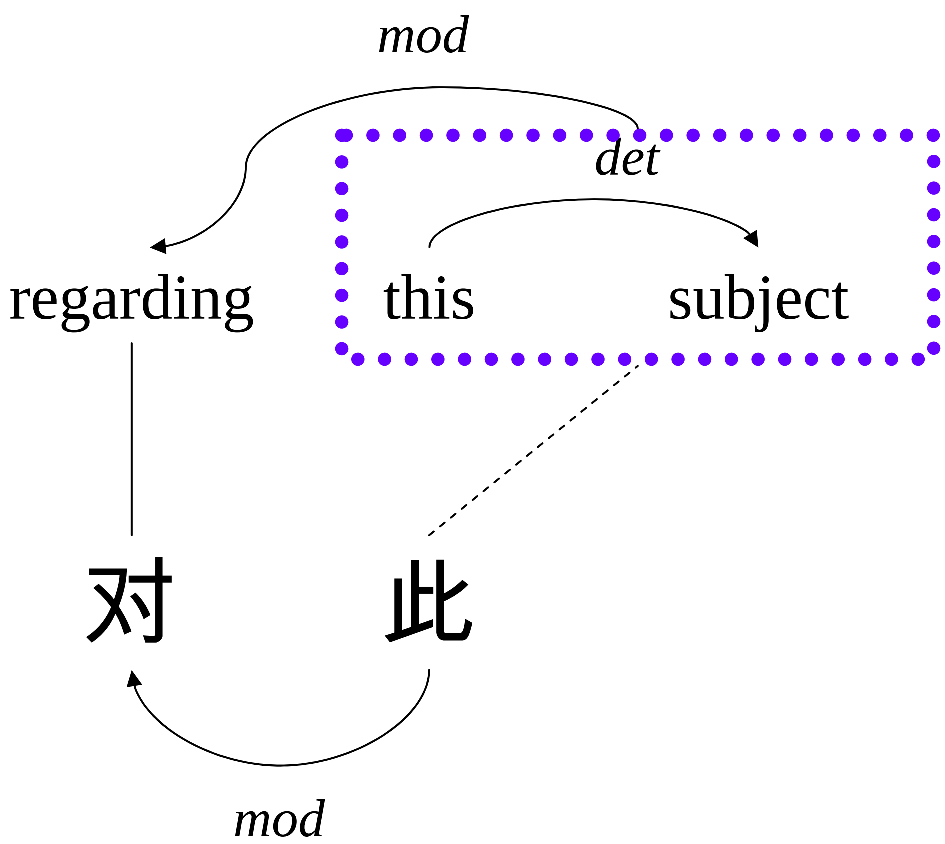
Problematic Case: Unaligned English



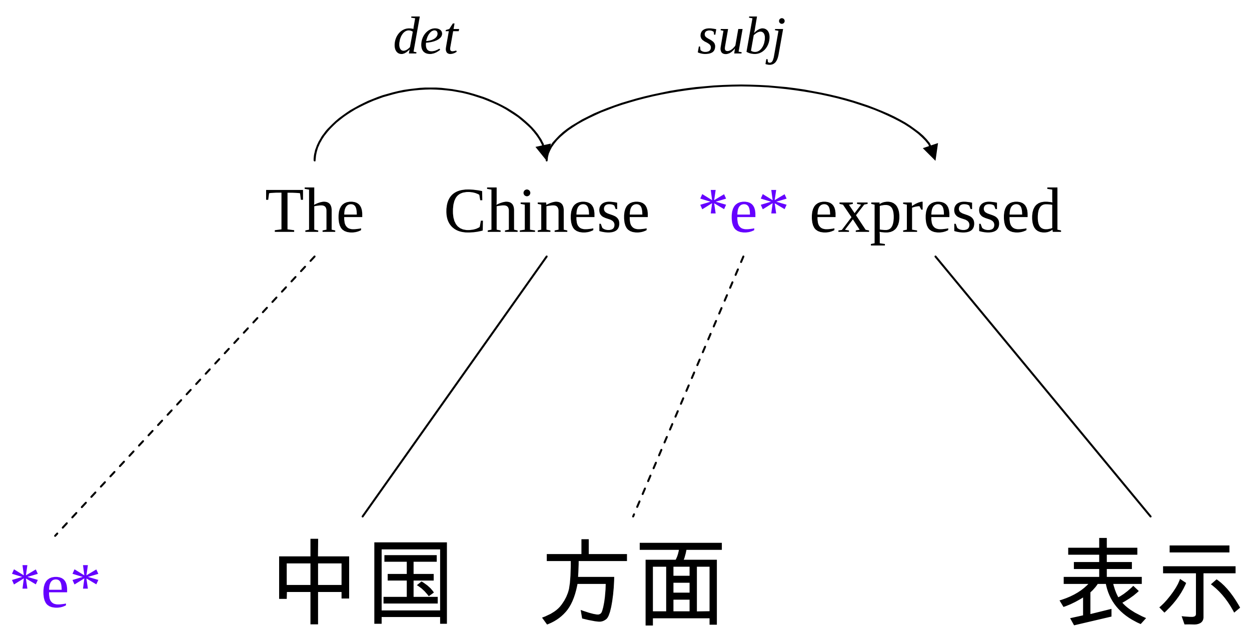
Problematic Case: many-to-1



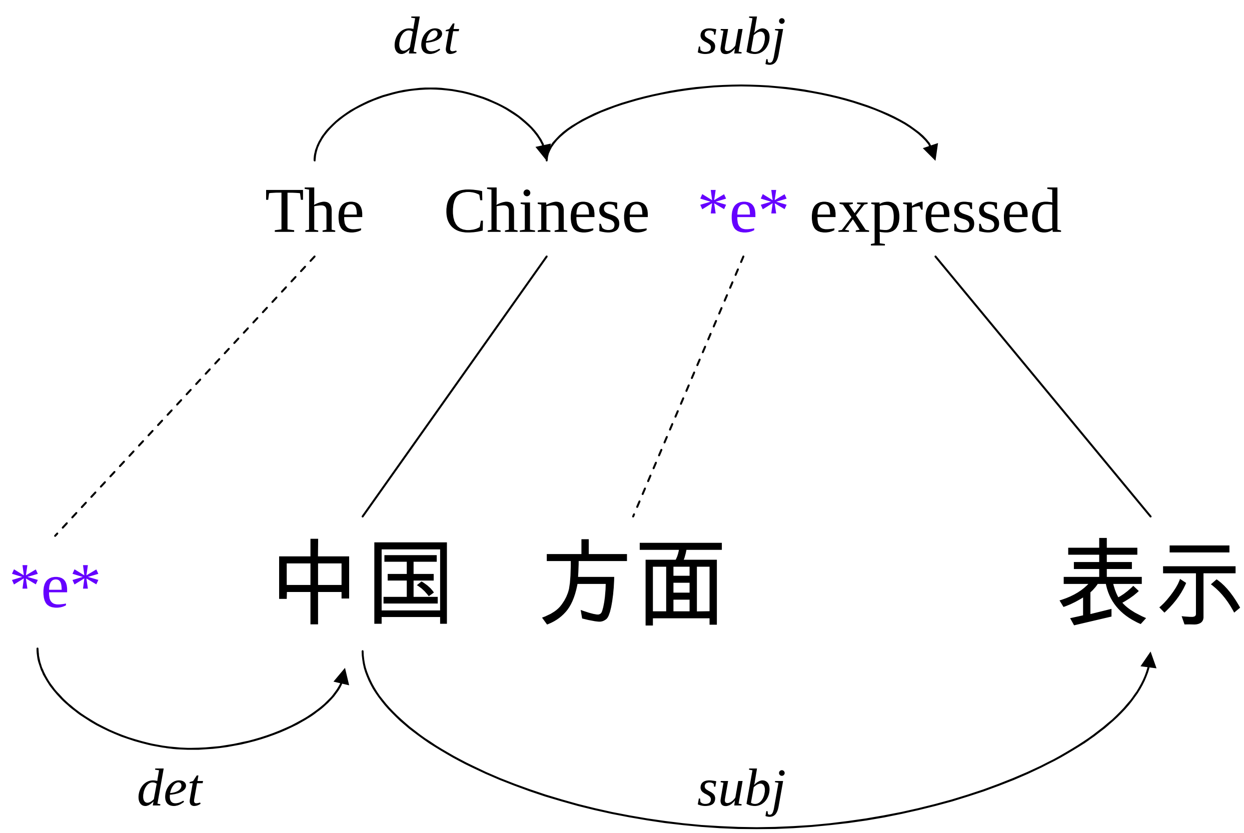
Problematic Case: many-to-1



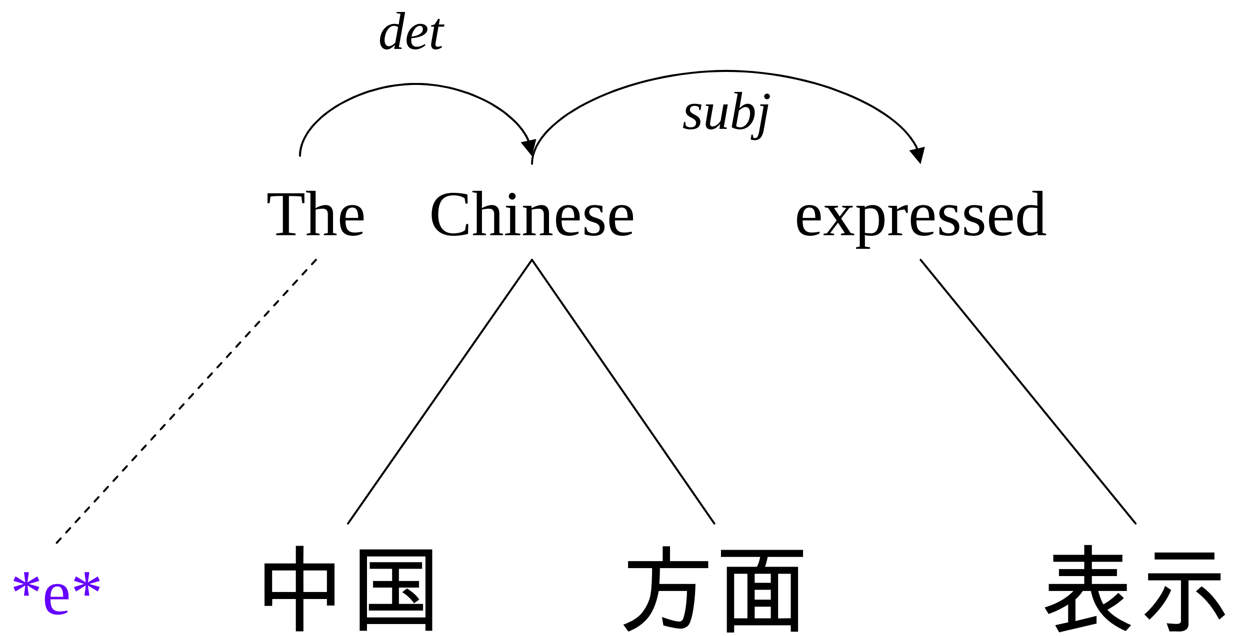
Problematic Case: Unaligned Chinese



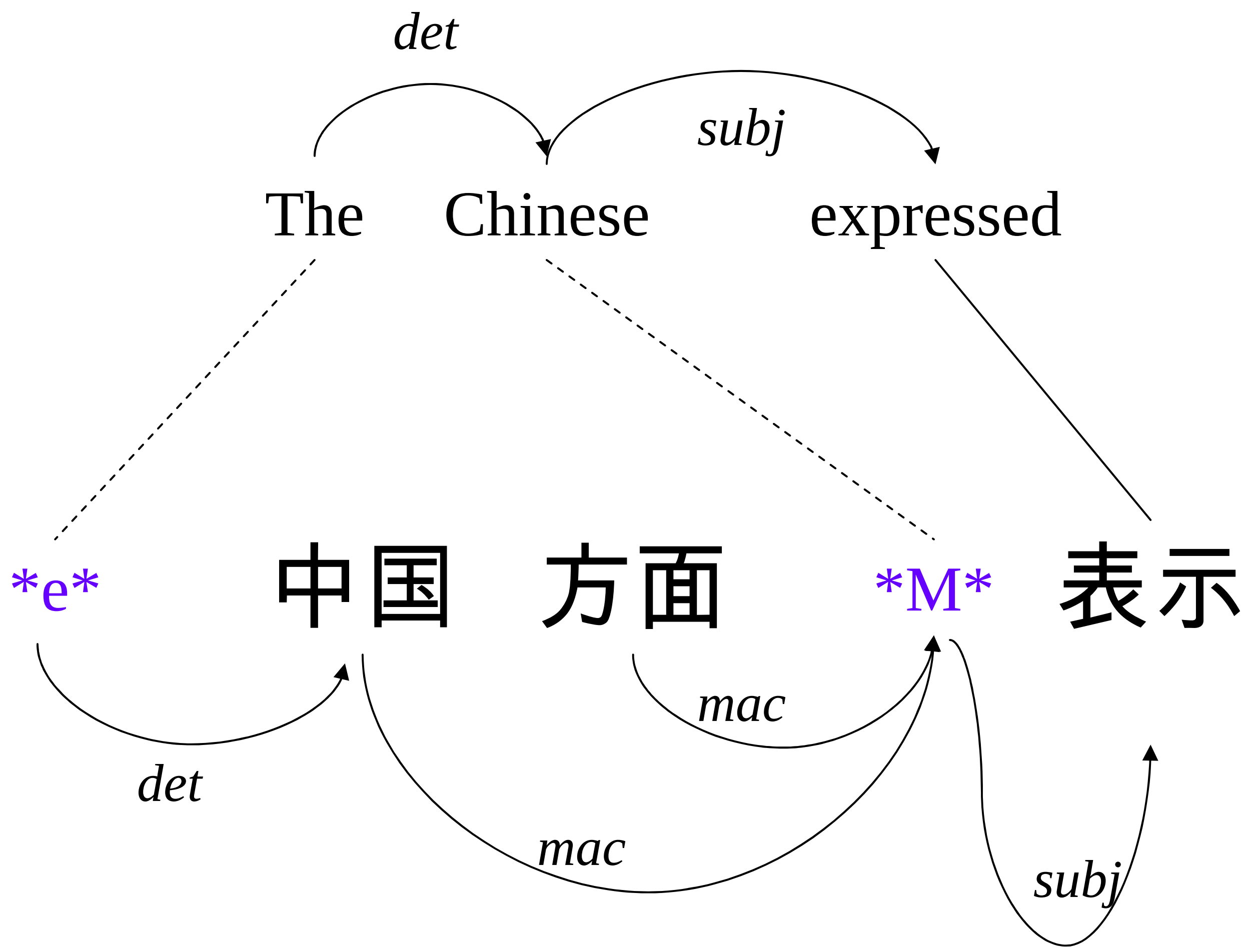
Problematic Case: Unaligned Chinese



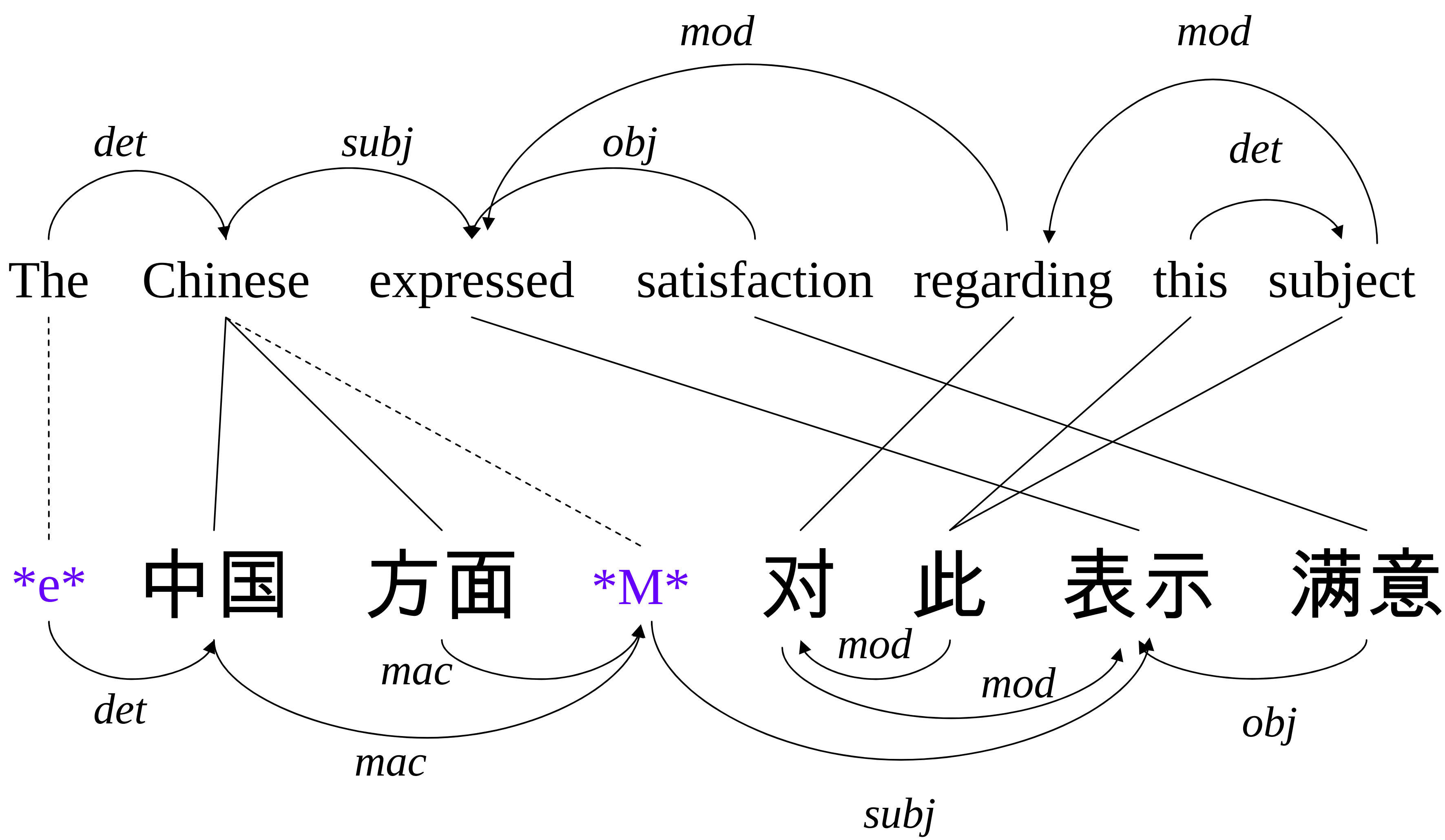
Problematic Case: 1-to-many



Problematic Case: 1-to-many



Output of the Direct Projection Algorithm

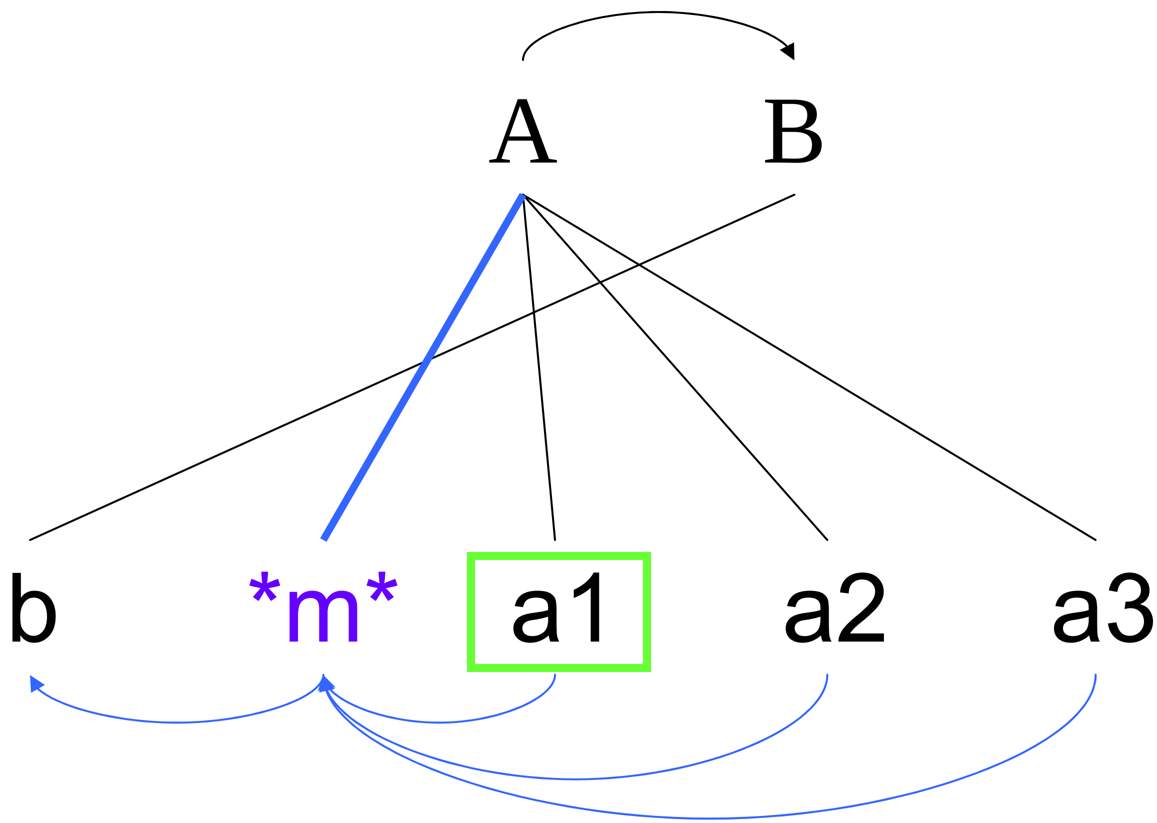


Rule-Based Post-Projection Cleanup

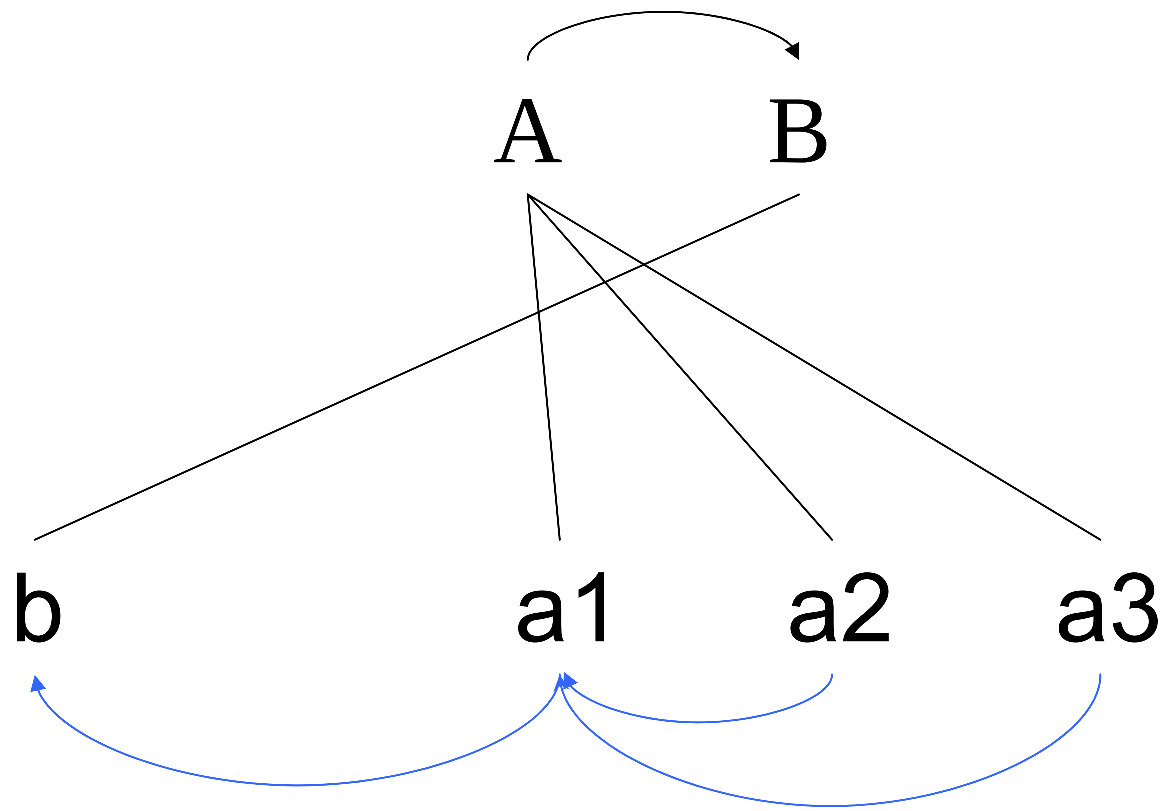
- Exploitation of general linguistic principles
  - Headness: Chinese is generally head-initial

- Development of post-processing rules
  - Functional/enumerated categories (closed class)
  - Projected parts of speech
  - Cf. *tsed* (Blaheta 2002)

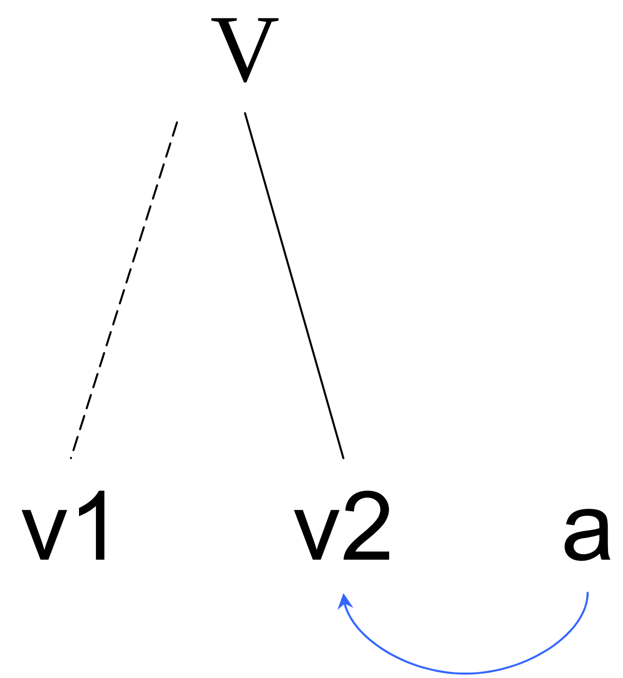
Head-Initial Promotion



Head-Initial Promotion



Aspectual Marker Attachment

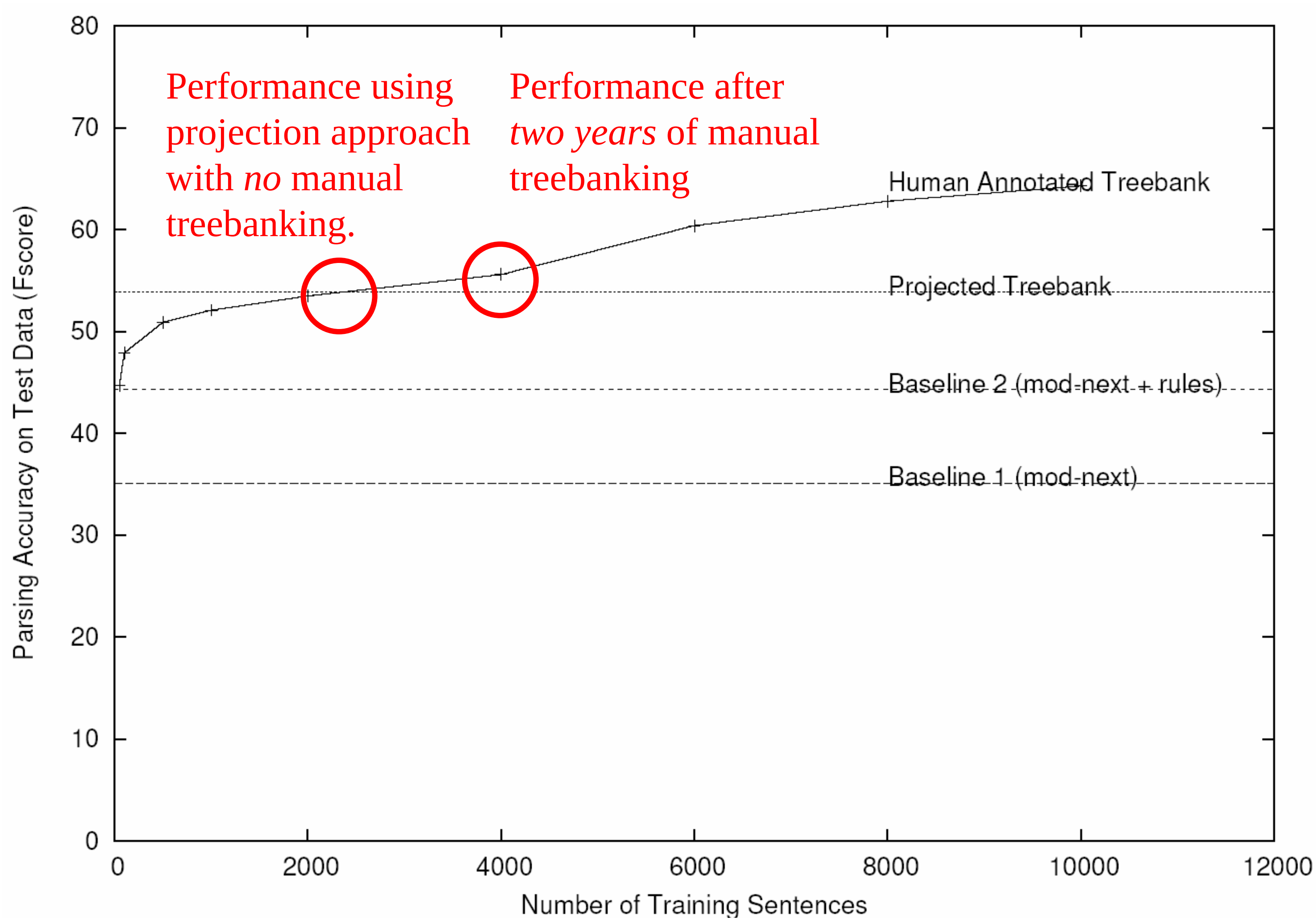


Quality of Automatically Annotated Chinese Data

Method	**Precision**	**Recall**	**F-measure**
Direct projection	34.5	42.5	38.1
Head-initial promotion rule	59.4	59.4	59.4 (+55.9%)
Rules	68.0	66.6	67.3 (+76.6%)

Filtering the Induced Treebank

- Projected treebank is noisy
  - Projection mismatch
  - Cascading component errors
- Automatically filter out bad training examples from projected treebank
  - Too many words were unaligned
  - Too many words are aligned to the same word
  - Projected tree has too many crossing dependencies.



Training a parser using the automatically projected treebank yielded almost the same level of parser performance as a parser trained on *4000* manually created trees from the Penn Chinese Treebank.

Training a Spanish Parser from Projected Treebank

Method	Training Corpus	Corpus Size	Parser Accuracy (100 test sent.)
Modify Prev (baseline)	-	-	34%
Stat. Parser	UN/FBIS/Bible (no filter)	98,000	67%
Stat. Parser	UN/FBIS/Bible (w/ filter)	20,000	72%
Commercial Parser	-	-	69%

One-Week Parser Results (Hindi)

- Post projection transformation: largely focused on case markers, light verbs
- Sentence filtering: don't use sentence pair if
  - There is a high percentage of alignment mismatches
  - Any English word aligns to 5 or more Hindi words

	Training sentences	Prec / Rec / F (Hindi)
Baseline: attach prev word	N/A	29.1 / 29.1 / 29.1
Baseline: attach next word	N/A	19.4 / 19.4 / 19.4
Statistical parser trained on projected, transformed trees	~14,700	44.1 / 43.9 / 44.0
Statistical parser using filtered training	~3,600	48.4 / 48.2 / 48.3

General Observations

- Limitations of assuming direct correspondence
  - Linguistic divergences literature (e.g. Dorr 1994)
  - Transfer based MT (e.g. Han et al. 2000)

- **But**: the DCA works to a surprising extent!

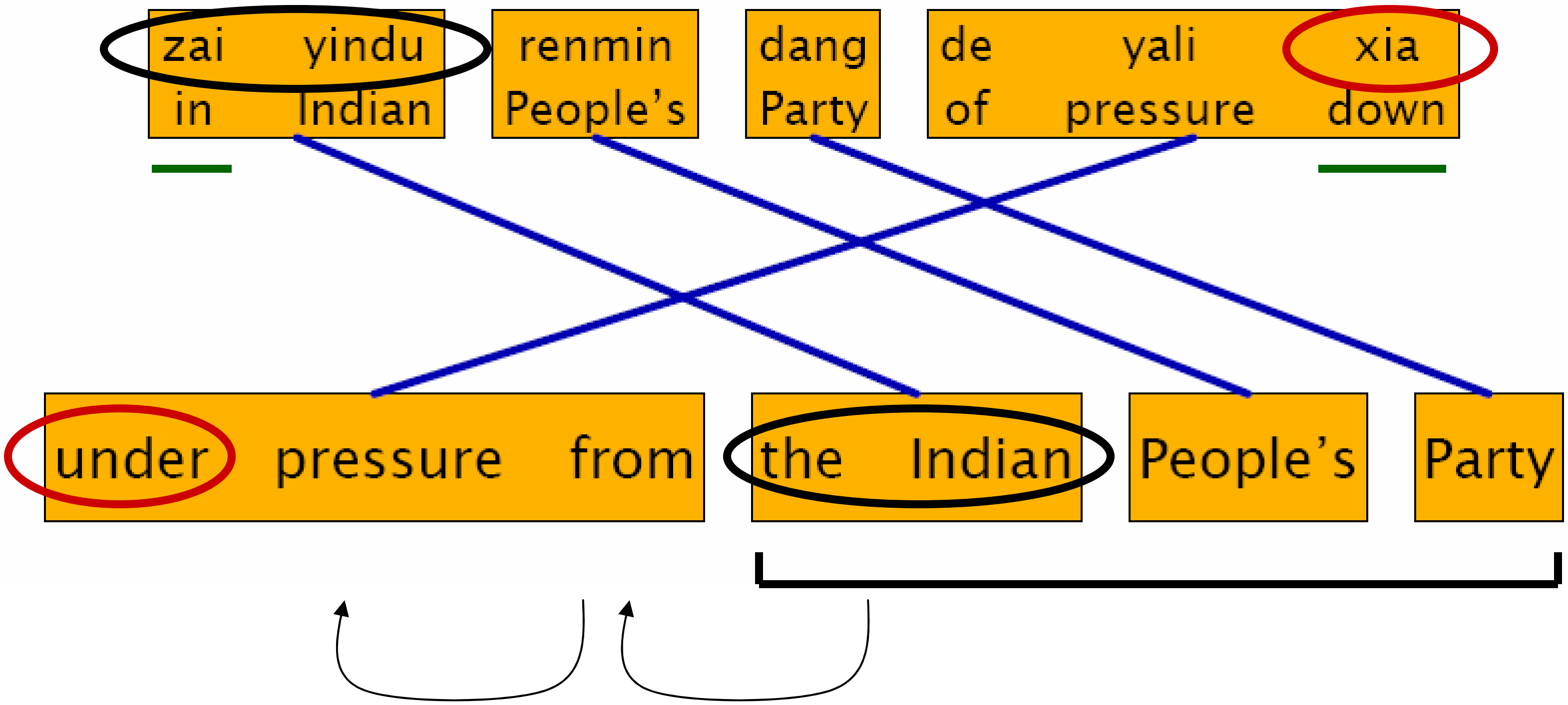
- Need better learning from noisy representations
  - Cf. Yarowsky and Ngai (2001), learning via annotation projection of POS tags, phrase bracketing, etc.

Hierarchical modeling for statistical machine translation

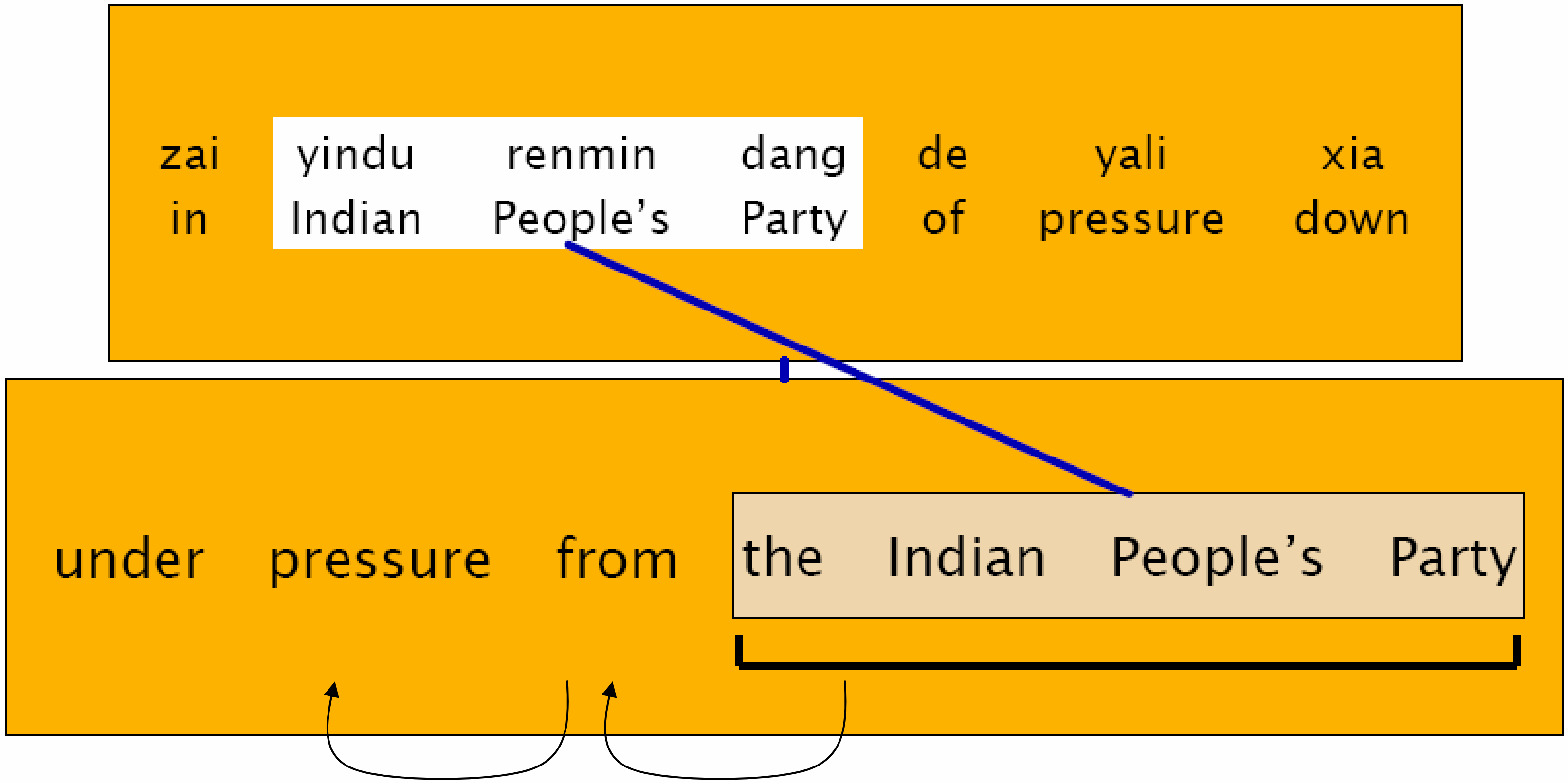
Hiero Statistical MT Framework

- Preserving meaning requires hierarchical structure, hence “parsing”.
  - David Chiang, “A hierarchical phrase-based model for statistical machine translation.” In *Proceedings of ACL 2005*, pages 263–270.
  - David Chiang, Adam Lopez, Nitin Madnani, Christof Monz, Philip Resnik, and Michael Subotin, "The Hiero Machine Translation System: Extensions, Evaluation, and Analysis", HLT/EMNLP 2005, Vancouver, October 2005.

Non-Hierarchical Phrases



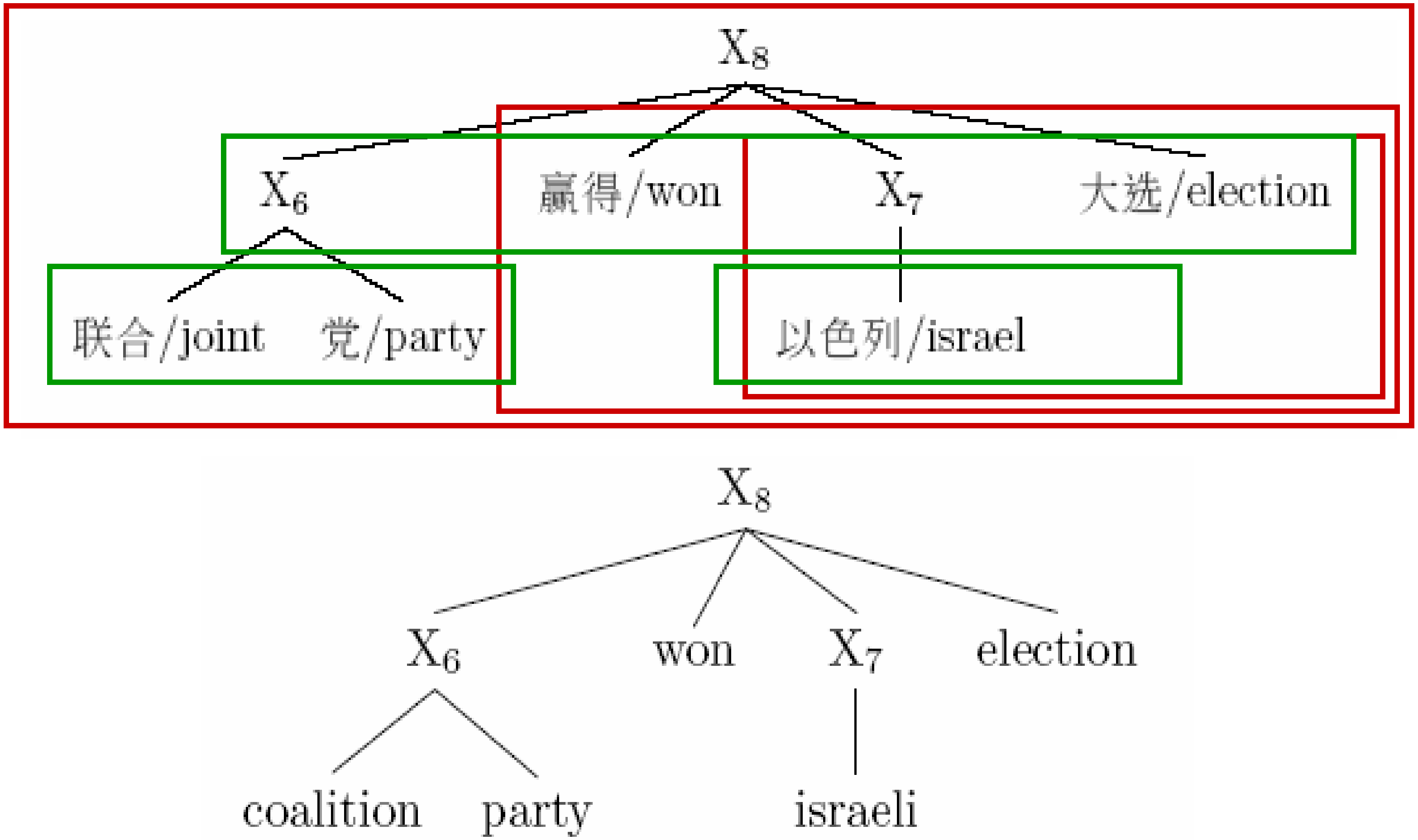
Hierarchical Modeling



Rank	Chinese	English
1	,	,
2	.	.
3	"	"
4	de	the
5	,	and
1710	X zongtong	president X
2097	$X_{\boxed{1}}$ de $X_{\boxed{2}}$	the $X_{\boxed{2}}$ of $X_{\boxed{1}}$
2850	jingnian X	X this year
10781	zai X xia	under X
32738	zai X nei	within X
218421	X de yali	pressure from X
300091	zai X yali xia	under pressure from X

Hiero Statistical MT Framework

- Preserving meaning requires hierarchical structure, hence “parsing”.
- The structures you want for good monolingual parsing are not always the same structures you want for good MT.


X8
X6
赢得/won
X7
大选/election
联合/joint
党/party
以色列/israel
X8
X6
won
X7
election
coalition
party
israeli

NIST MTEVAL 2005, Arabic	
Site	BLEU-4 Score
GOOGLE	0.5131
ISI	0.4657
IBM	0.4646
UMD	0.4497
JHU-CU	0.4348
EDINBURGH	0.3970
SYSTRAN	0.1079
MITRE	0.0772
FSC	0.0037

UMD TM used a fraction of the training data (1.5M words, no Ummah or UN); **Important** given limited data for new dialects, low-density language scenarios

LM trained on 365M words.

Hardware scale-up imminent.

NIST MTEVAL 2005, Chinese	
Site	BLEU-4 Score
GOOGLE	0.3531
ISI	0.3073
UMD	0.3000
RWTH	0.2931
JHU-CU	0.2827
IBM	0.2571
EDINBURGH	0.2513
ITCIRST	0.2445
NRC	0.2323
NTT	0.2321
ATR	0.1822
SYSTRAN	0.1471
SAAR	0.1310
MITRE	0.0542

UMD TM used 30M words.

LM trained on 168M words.