Advances in Statistical Machine Translation: Phrases, Noun Phrases and Beyond

Philipp Koehn

koehn@isi.edu

Information Sciences Institute

Department of Computer Science

University of Southern California

Outline

- Statistical Machine Translation
- Phrase-Based Methods
- Syntactic Structure
  - Noun Phrase Translation
  - Clause Structure

Machine Translation

- Translating text in a foreign language into English
- One of the oldest problems in Artificial Intelligence
- AI-hard: reasoning and world knowledge required
- State of the art:

The United States and India May Will Be Held in the Past 40 Years the First Joint Military Exercises

(Afp report from new Delhi) India and U. S. will be held in the past 39 years the first joint military exercises in the world's two biggest democracies the cooperative relationship between making milestone.

The Defense Ministry said in a class Indian paratrooper Brigade mid-May and the US Pacific Command of the special units in the well-known far and near the Thai women Maha tomb near joint military exercises.

The two countries will provide air support.

(Chinese-English, statistical machine translation system)

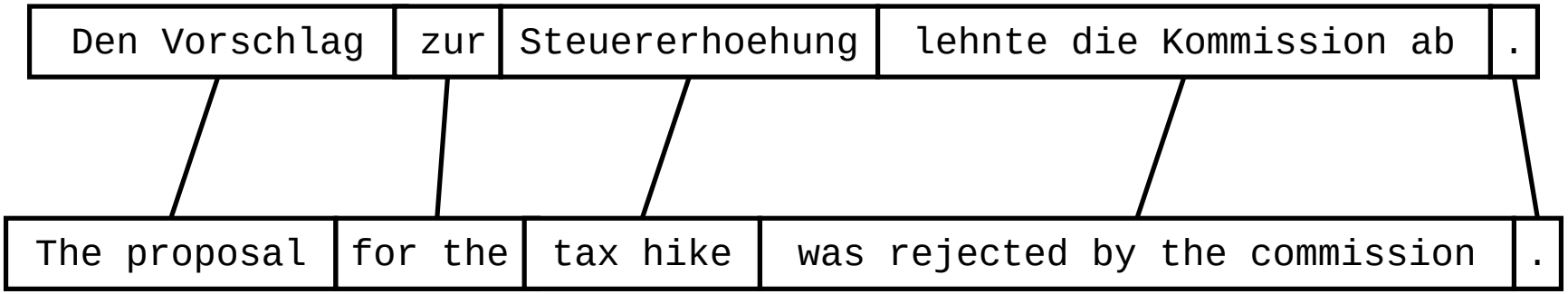
Statistical Machine Translation

- Learn translation from parallel text

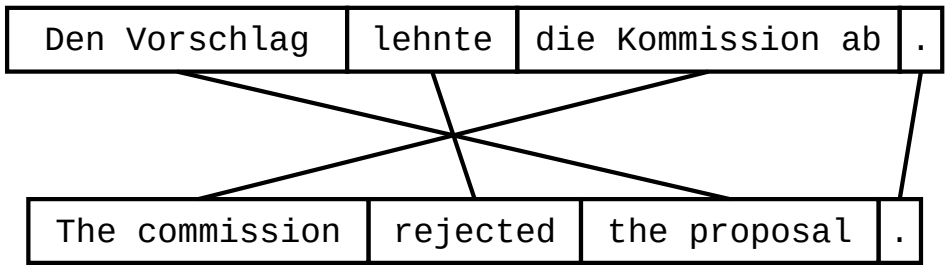
- Currently corpora for many language pairs available
  - up to 20-200 million words
  - e.g., Europarl `http://www.isi.edu/~koehn/europarl/`,
    11 European languages, 20 million words each

Phrase-Based Methods

- Currently best performing methods map phrases



- "Phrases"
  - any sequences of words
  - reordering of phrases possible



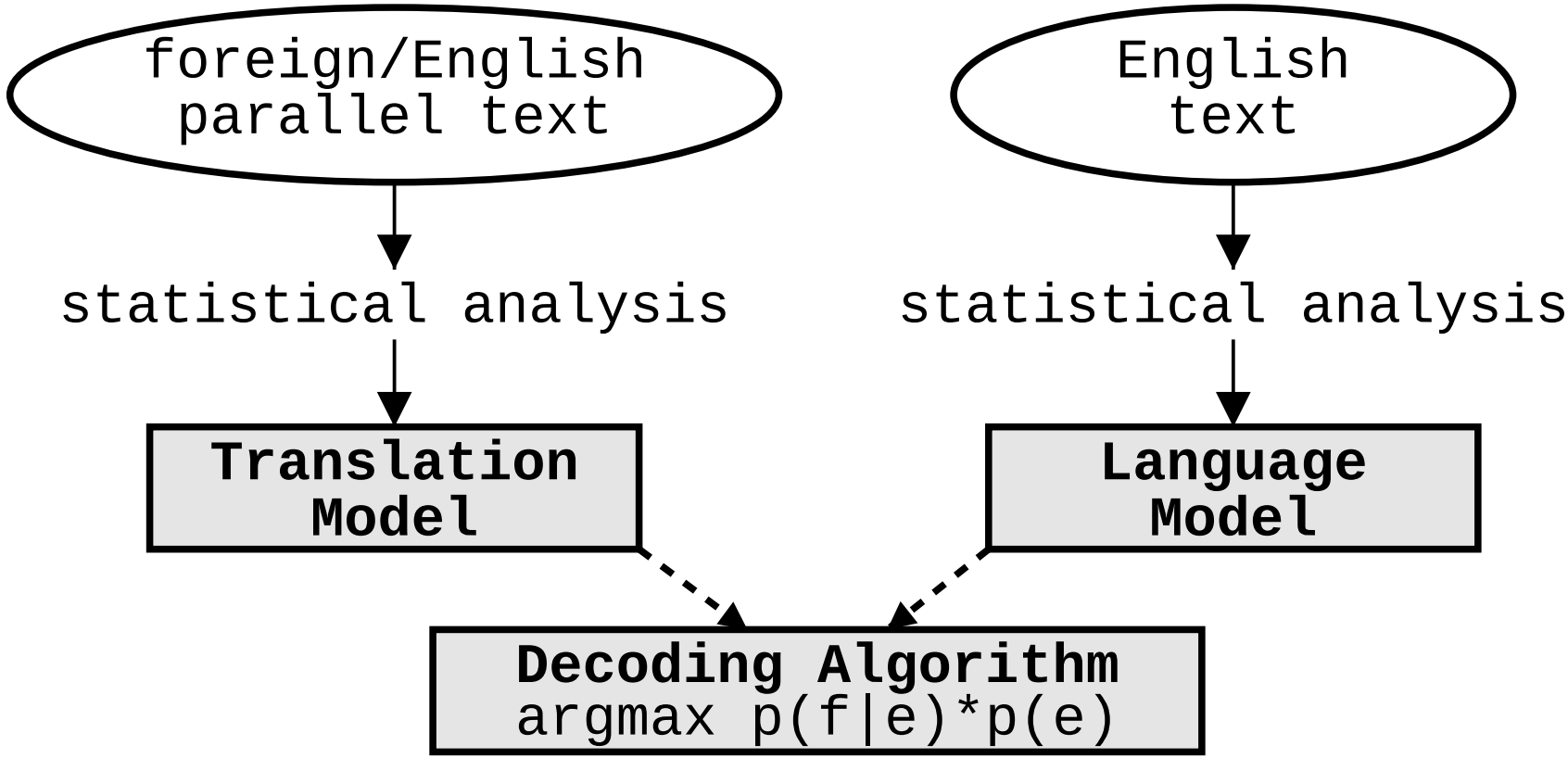
Phrase Translation Table

- Phrase Translations for “den Vorschlag”:

English	$\phi(e\|f)$	English	$\phi(e\|f)$
the proposal	0.6227	the suggestions	0.0114
’s proposal	0.1068	the proposed	0.0114
a proposal	0.0341	the motion	0.0091
the idea	0.0250	the idea of	0.0091
this proposal	0.0227	the proposal ,	0.0068
proposal	0.0205	its proposal	0.0068
of the proposal	0.0159	it	0.0068
the proposals	0.0159	...	...

Noisy Channel Model

- Bayes rule: $p(e|f) \sim p(f|e) * p(e)$



Decoding

- Translation ("decoding") is NP-complete [Knight, 1999]

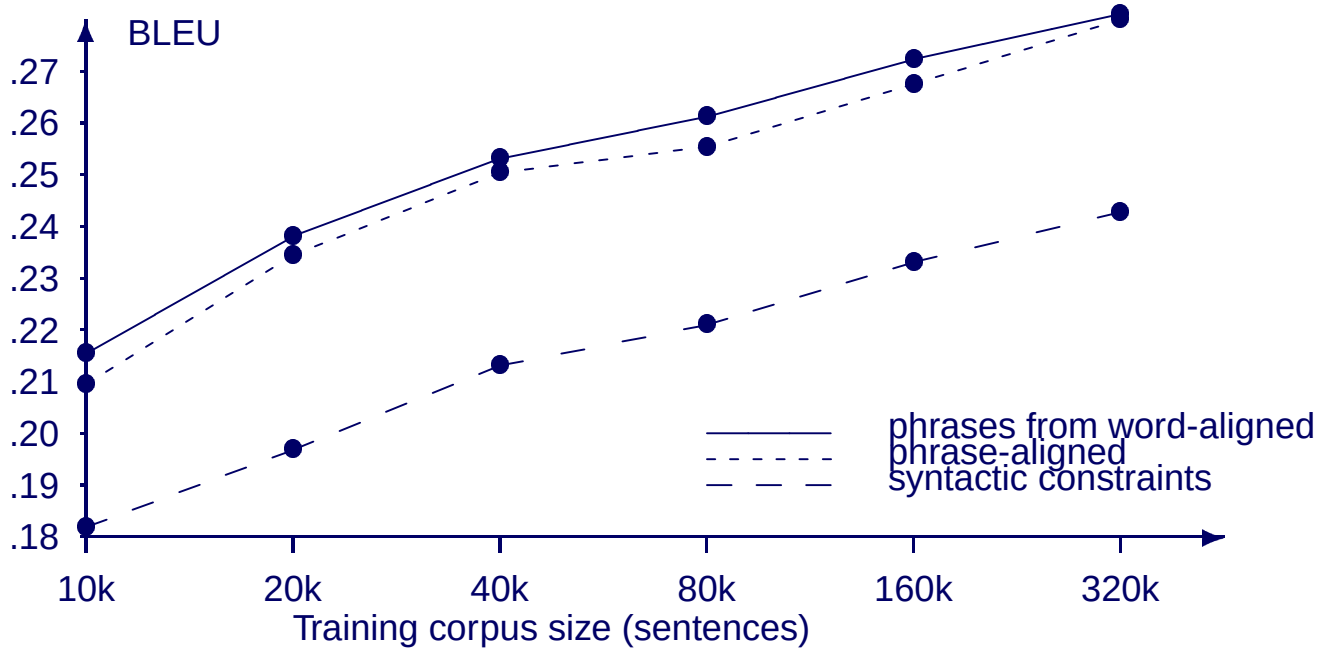
$\Rightarrow$ Various heuristic search methods

  - dynamic programming beam search [Och et al., 2001]
  - greedy search [Marcu and Wong, 2002]
  - finite state transducers [Kumar and Byrne, 2003]

- My decoder "Pharaoh" freely available (soon)

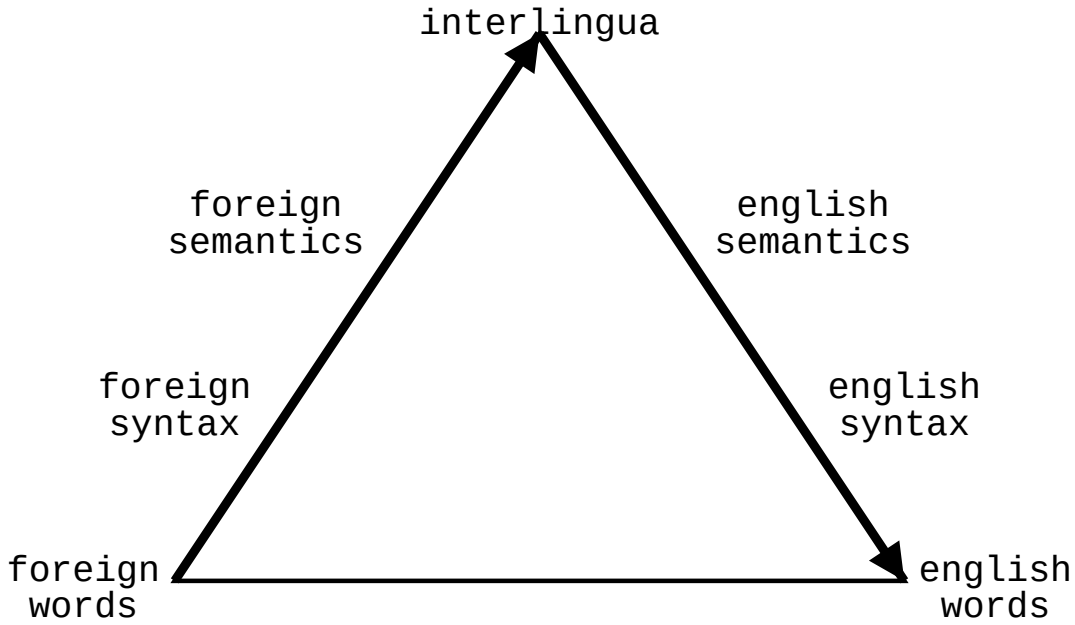
Learning Phrase Translation Tables

- Comparison of various methods [Koehn et. al, 2003]
  - directly aligning phrases in parallel corpora
  - aided by word alignments
  - syntactic constraints



Syntax and Statistical MT

- Why syntax?
  - many transformations can be best explained in syntactic terms
  - syntactic annotation on the foreign input adds additional knowledge
  - syntactic annotation on the English output aids grammatical output
- Machine Translation Pyramid



Previous Syntax-Based Transfer Models

- Various attempts as using syntax:
  - tree-based transfer process [Alshawi, 1996] [Wu, 1997]
  - string to tree translation [Yamada and Knight, 2001]
  - using syntactic chunks [Schafer and Yarowsky, 2003]
  - loosely tree-based [Gildea, 2003]
  - syntactic features [Koehn and Knight, 2003] [Och et al., 2003]
- None showed significant improvement over phrase models

Syntactically Structured Statistical MT

Focus on two main syntactic categories:

- Noun phrases
  - contain most of vocabulary
  - can be translated in separation
  - subject of my PhD thesis [Koehn, 2003]
- Clauses
  - syntactic restructuring not well captured by phrase models
  - ongoing work

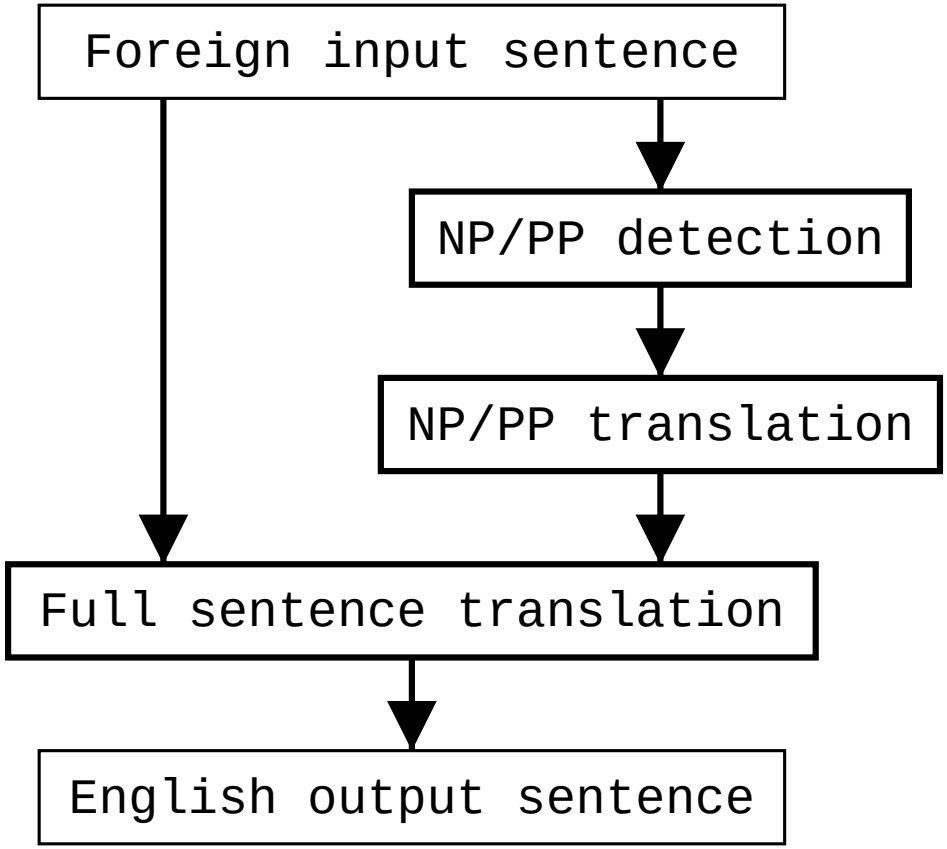
Noun Phrase Translation

- Definition NP/PP
  - the maximal noun phrases and prepositional phrases attached to the clause level
  - do not contain relative clauses
- Examples
  - *(The proposal for the tax hike)* was rejected *(by the commission)* .
- Cover roughly half of all words, all nouns, most of vocabulary

Translatability

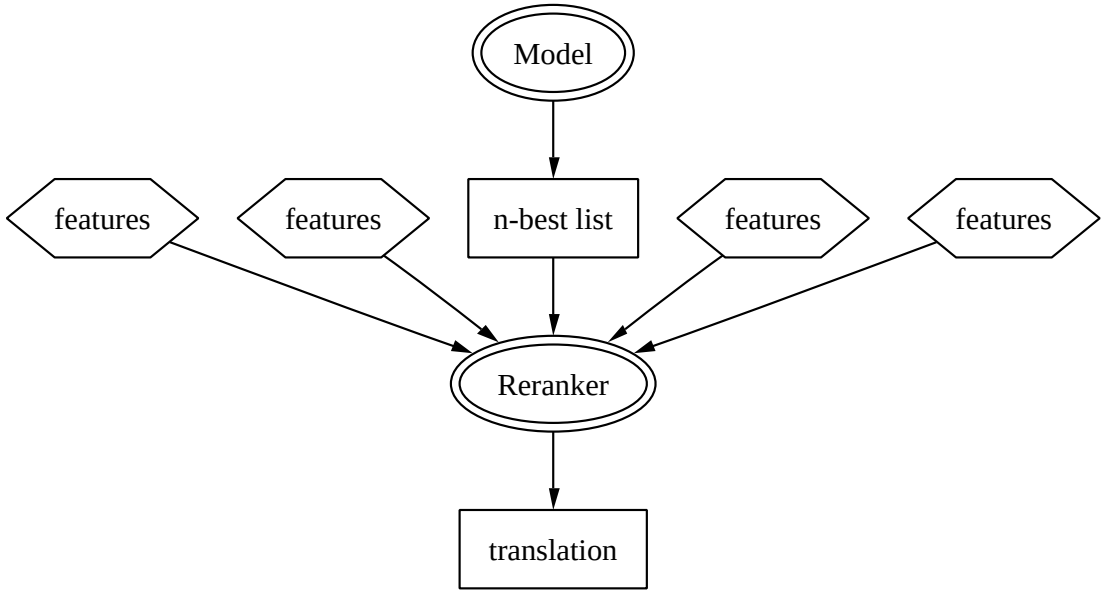
- Study on German-English Europarl corpus
- Are NP/PPs translated as NP/PPs?
  - 75% are translated, 98% can be
  - exceptions
    - merge with verb: *make an observation*
    - PP translated as adverb: *in der Hauptsache = mainly*
- Translation in Isolation?
  - human translation w/o sentence context
  - 89% NP/PPs correctly translated
  - 9% wrong leading preposition
  - 2% wrong content word

Framework



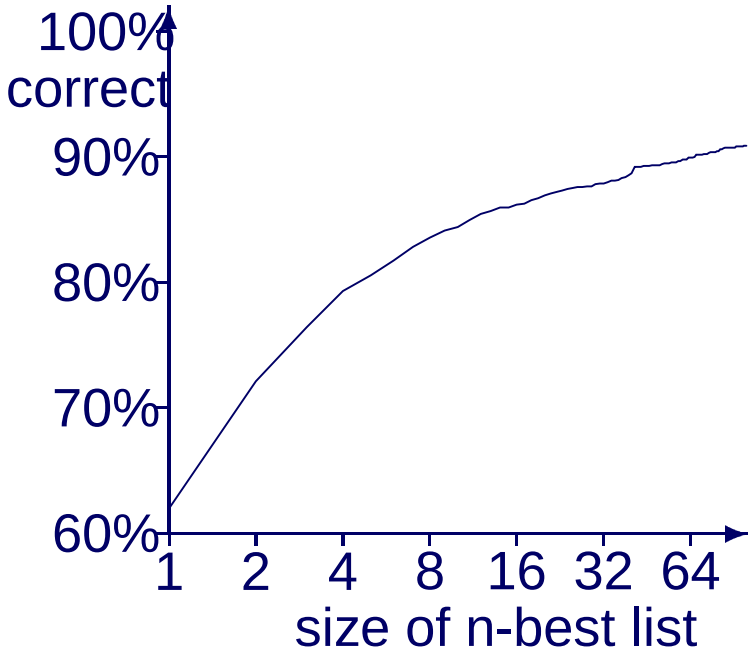
- NP/PPs translated by modular subsystem

Translation as Reranking



- Base model proposes candidate
- Reranking with additional features
  - maximum entropy
  - similar to [Och and Ney, 2002]

Translation as Reranking: Why Possible?



- 60% of NP/PPs translated correctly
- 90% of NP/PPs have correct translation in 100-best list
- Advantage of reranking: global features

Error Analysis: Not in n-Best List

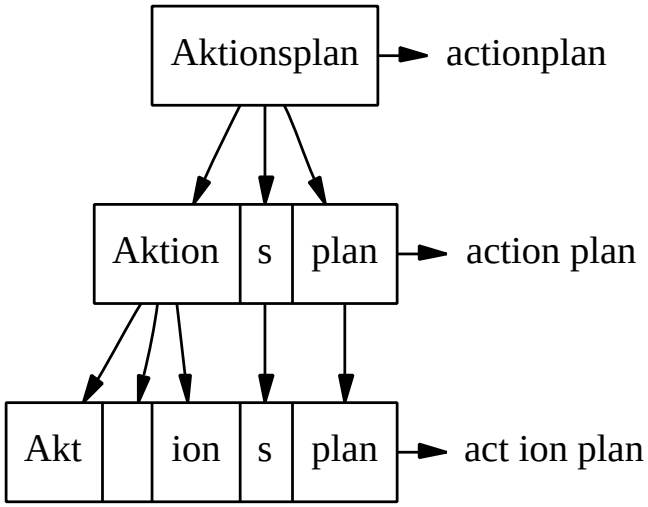
Error	Frequency
Unknown Word	34%
Tagging or parsing error	28%
Unknown translation	14%
Complex syntactic restructuring	7%
Too long	6%
Untranslatable	2%
Other	9%

- 10% of NP/PPs: no acceptable translation in list
- Main problem: unknown words, translations

Special Modeling for NP/PP Translation

- Compound splitting
- Web n-Grams
- Syntactic features

Compound Splitting

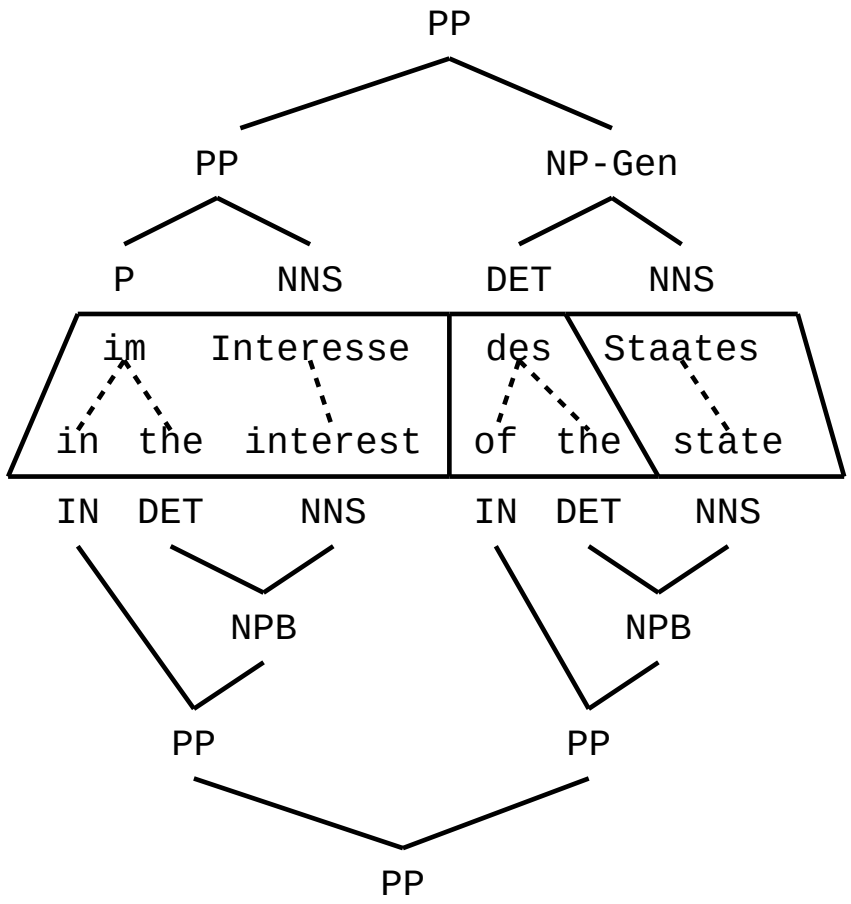


- Compounding occurs in German, Finnish, Greek, ...
  - increased vocabulary size
  - leads to sparse data problems and unknown words
- Frequency based method for compound splitting
  - break up, if parts are more frequent than whole
  - geometric mean: $S_{\text{best}} = \text{argmax}_S(\prod_{p_i \in S} \text{count}(p_i))^{\frac{1}{n}}$
- More detail in [Koehn and Knight, EACL 2003]

Web n-Grams

- Web as language model
  - 3 billion documents indexed by Google
  - 97% of trigrams seen on web
  - 30% of 7-grams seen on web
- Occurrence on web as feature
  - does phrase occur on web?
  - do all its n-grams occur on web? (n=2...7)
  - at least once? at least ten times?
  - using Google to collect counts

Syntactic Features



given foreign parse tree

phrase translation with word alignments

POS tags from tagger

syntax tree from parser

- Keep foreign syntactic parse tree
- Annotate English candidate translation with syntax

Syntactic Features (2)

- Given two syntactic parse trees

$\Rightarrow$ Any computable property between pair can be feature

- We implemented three features
  - preservation of number of a noun (singular stays singular)
  - preservation of preposition (no dropping of preposition, except if there is movement)
  - number agreement in baseNPs (not: *this nice green flowers*)
- Many more conceivable

Evaluation

- German-English translation task
- Europarl corpus
  - extracted NP/PP pairs using Giza++ for word alignment, Collins' parser for English and Lopar parser for German
- Training
  - base model trained on 743,370 aligned NP/PPs
  - feature values trained on 683 NP/PPs
- Test set
  - 1362 NP/PPs from 534 sentences

Accuracy (Human Judgment)

System	NP/PP Correct	
Word-Based Model	724	53.2%
Phrase-Based Model	800	58.7%
Compound Splitting	838	61.5%
Re-Estimated Parameters	858	63.0%
Web Count Features	881	64.7%
Syntactic Features	892	65.5%

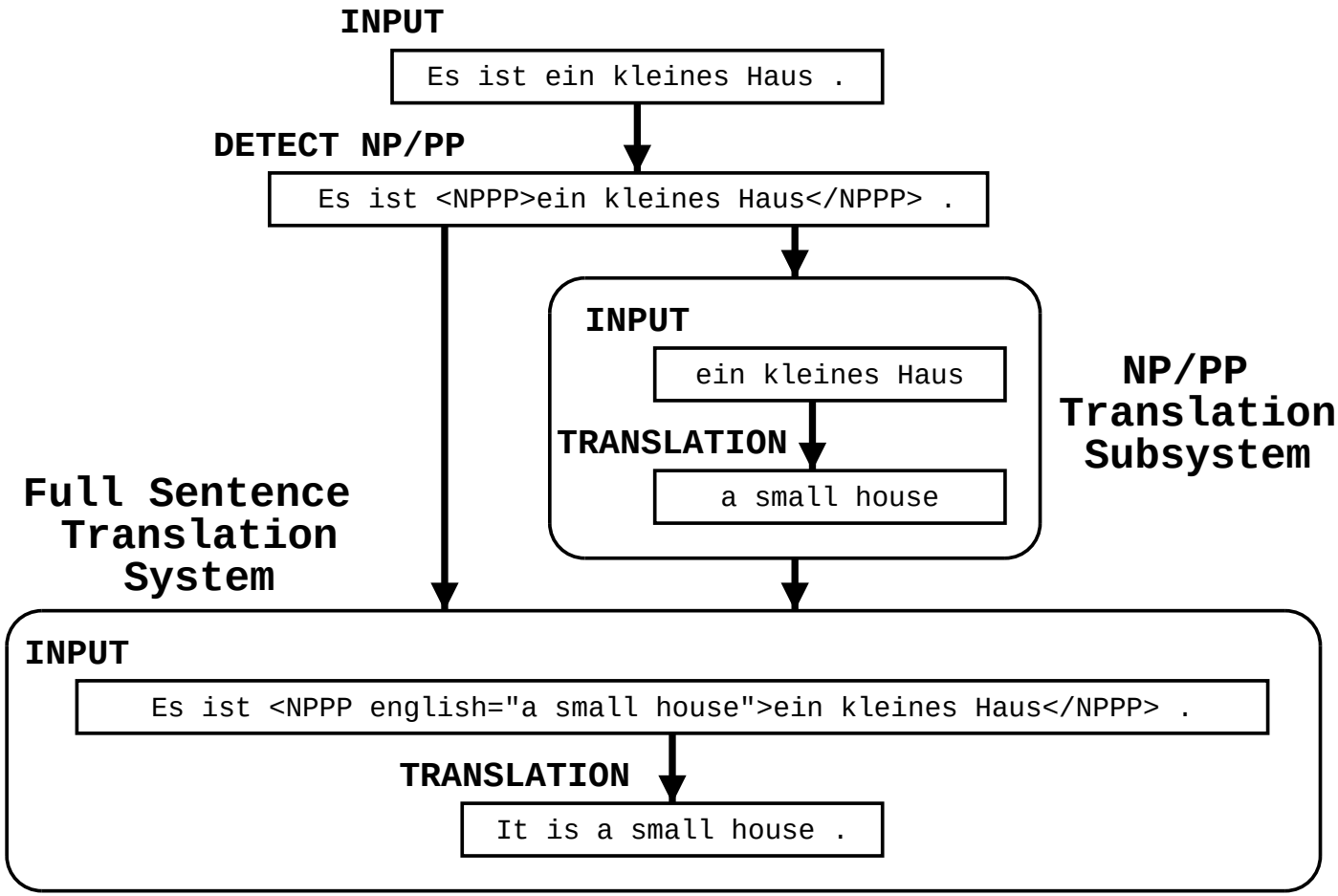
- Overall +12.3% improvement
- 95% Statistical significance interval 2.5%

Error Analysis: Reranking Failure

Error	Freq.	Error	Freq.
Involving content words	**44%**	**Involving only function words**	**56%**
Wrong word choice	16%	Wrong phrase start	38%
Content word mistranslated	4%	Internal preposition choice	4%
Wrong phrase choice	3%	Pronoun / anaphora	4%
Content dropped	13%	Pronoun added or dropped	3%
Content added	2%	Determiner added, dropped, wrong	2%
Number of noun wrong	2%	Function word phrase choice	2%
Other	**2%**	Function word mistranslated or dropped	2%
Reordering wrong	1%	Preposition dropped	1%
Other	1%		

- 25% of NP/PPs: acceptable translation in list, not picked
- Main problems: wrong phrase start, word choice, dropping of content

Integration



- Translations passed to full sentence translation system
  - using XML markup
  - allow passing of reranked list (with probabilities)

Evaluation of Integration

- Performance on full-sentence translation (BLEU score)

System	Word-Based MT	Phrase-Based MT
baseline system	0.176	0.220
with NP/PP subsystem	0.199	0.224

- Why little improvement for phrase-based MT?
  - cuts around NP/PP disable overlapping phrase translations
  - parsing errors force hard decisions

Conclusions on NP/PP Translation

- It is possible to separate out NP/PP translation
- Improved NP/PP translation performance
- Improved overall sentence translation performance
  - still needs better integration
  - still needs better conditioning on sentence context

Clause Level

- Phrase-based system mistranslated the example:
  - German: *Den Vorschlag zur Steuererhoehung lehnt die Kommission ab .*
  - English: *The proposal for the tax hike is rejected by the commission .*
  - Phrase-MT: *The tax increase proposal opposes the commission .*
- Why?
  - semantic reasons: proposals usually don't reject, they get rejected
  - syntactic reason: *Den Vorschlag* is accusative case, therefore object
- Syntactic information is ignored by phrase-based system
  - only indicated by the determiner *den*

Clause Level Transformations

- Required: clause level transformation

  `G: object - verb - subject`

  $\Rightarrow$ `subject - verb - object`

  $\Rightarrow$ `object - passive verb - subject`

- There are more such transformations
  - some systematic due to different syntax in foreign and English
  - some driven by verbs whose translation has different subcategorization
- How do we add this to statistical machine translation?
  - future work...

Conclusion

- Reviewed phrase-based MT
- Syntactic structure
  - improved translation quality by special handling of noun phrases
  - clause level transformations as research challenge

Thank You!

- Questions?