Empirical Methods in Natural Language Processing
Lecture 18
Machine translation (V): Syntax-Based Models

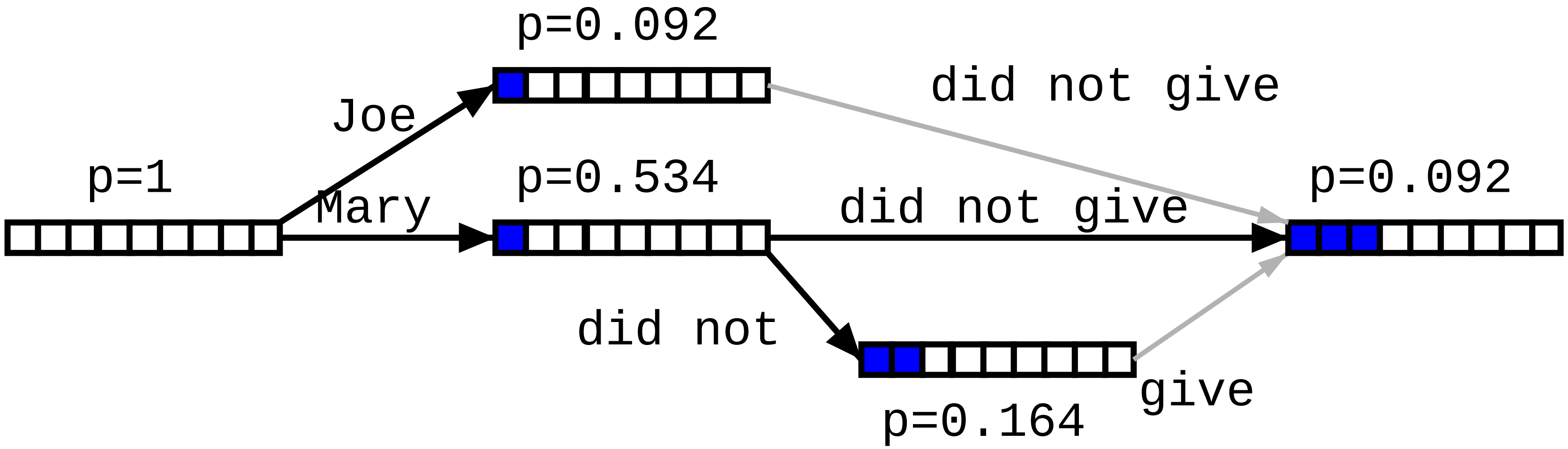
Philipp Koehn & Adam Lopez

6 March 2008

School of informatics

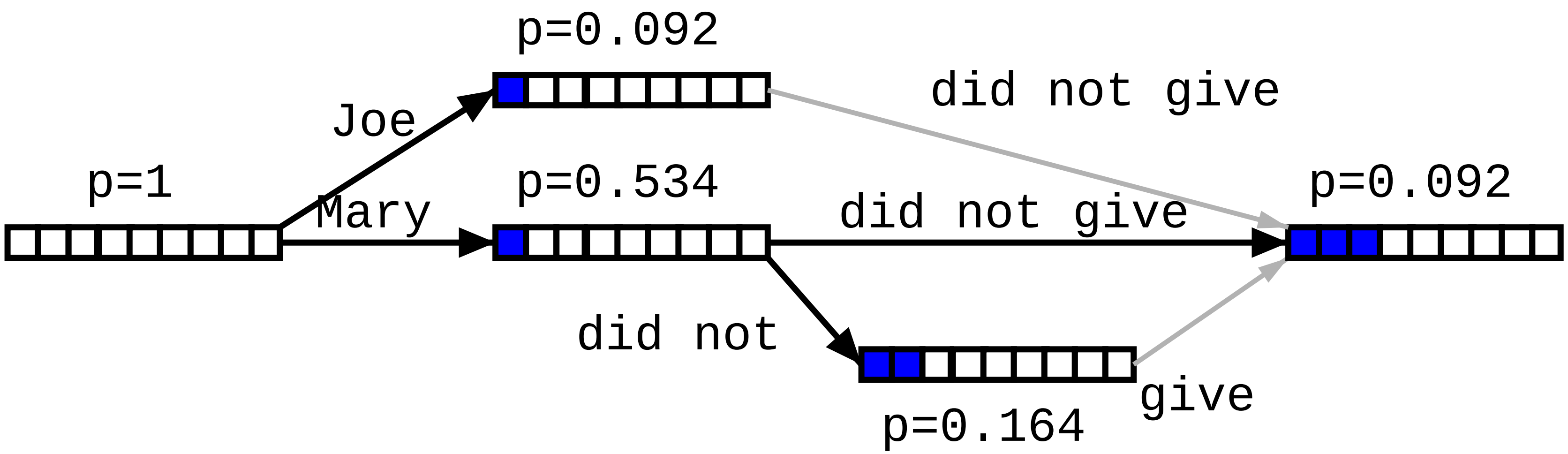
Phrase-based SMT

- Already works pretty well.
- Are there any problems that we need to solve here?



Phrase-based SMT

- *Computational*: computing all possible reorderings is NP-complete.
- *Linguistic*: language is not finite-state.



Syntax-based SMT

- What's going on here? A whole lot of things...
- Chiang (2005) makes a distinction between *formally* syntax-based and *linguistically* syntax-based.

		formally syntax-based
	phrase-based	hierarchical phrase-based [Chiang 2005]
linguistically syntax-based	reordering + phrase-based [Collins et al. 2005]	syntax-based SCFG [Yamada & Knight 2002]

Linguistic Advantages of Syntax-Based Translation

- *Generalized* reordering for syntactic reasons
  - e.g., move German object to end of sentence
- Better explanation for *function words*
  - e.g., prepositions, determiners
- Conditioning to *syntactically related words*
  - translation of verb may depend on subject or object
- Use of *syntactic language models*
  - ensuring grammatical output

Clause Level Restructuring [Collins et al.]

		formally syntax-based
	phrase-based	hierarchical phrase-based [Chiang 2005]
linguistically syntax-based	*reordering + phrase-based [Collins et al. 2005]*	syntax-based SCFG [Yamada & Knight 2002]

Clause Level Restructuring [Collins et al.]

- Why **clause structure**?
  - languages *differ vastly* in their clause structure (English: SVO, Arabic: VSO, German: fairly *free order*; a lot details differ: position of adverbs, sub clauses, etc.)
  - large-scale restructuring is a *problem* for phrase models

- **Restructuring**
  - *reordering* of constituents (main focus)
  - add/drop/change of *function words*

- Details see [Collins, Kucerova and Koehn, ACL 2005]

Clause Structure

```
S   PPER-SB   Ich     I
    VAFIN-HD  werde    will
    VP-OC     PPER-DA Ihnen   you
              NP-OA   ART-OA  die   the
                      ADJ-NK  entsprechenden    corresponding
                      NN-NK   Anmerkungen   comments
              VVFIN   aushaendigen     pass on
              $,      ,    ,
              S-MO    KOUS-CP damit    so that
                      PPER-SB Sie    you
                      VP-OC   PDS-OA  das   that
                              ADJD-MO eventuell    perhaps
                              PP-MO   APRD-MO  bei   in
                                      ART-DA   der   the
                                      NN-NK    Abstimmung  vote
                              VVINF   uebernehmen    include
                      VMFIN   koennen   can
$. .             .
```

MAIN CLAUSE

SUB-ORDINATE CLAUSE

- *Syntax tree* from German parser
  - statistical parser by Amit Dubay, trained on TIGER treebank

Reordering When Translating

```
S    PPER-SB  Ich                               I
     VAFIN-HD werde                             will
     PPER-DA  Ihnen                             you
     NP-OA    ART-OA  die                       the
              ADJ-NK  entsprechenden             corresponding
              NN-NK   Anmerkungen                comments
     VVFIN    aushaendigen                      pass on
$,   ,                                         ,
S-MO KOUS-CP  damit                             so that
     PPER-SB  Sie                               you
     PDS-OA   das                               that
     ADJD-MO  eventuell                         perhaps
     PP-MO    APRD-MO  bei                      in
              ART-DA   der                       the
              NN-NK    Abstimmung                vote
     VVINF    uebernehmen                       include
     VMFIN    koennen                           can
$. .                                           .
```

- *Reordering* when translating into English
  - tree is *flattened*
  - clause level constituents line up

Clause Level Reordering

```
S     PPER-SB  Ich                          1 I
      VAFIN-HD werde                        2 will
      PPER-DA  Ihnen                        4 you
      NP-OA    ART-OA  die                    the
               ADJ-NK  entsprechenden       5  corresponding
               NN-NK   Anmerkungen             comments
      VVFIN    aushaendigen                 3 pass on
$,    ,                                      ,
S-MO  KOUS-CP  damit                        1 so that
      PPER-SB  Sie                          2 you
      PDS-OA   das                          6 that
      ADJD-MO  eventuell                    4 perhaps
      PP-MO    APRD-MO  bei                   in
               ART-DA   der                 7  the
               NN-NK    Abstimmung             vote
      VVINF    uebernehmen                  5 include
      VMFIN    koennen                      3 can
$.  .                                        .
```

- Clause level reordering is a *well defined task*
  - label German constituents with their *English order*
  - done this for 300 sentences, two annotators, high agreement

Systematic Reordering German → English

- Many types of reorderings are **systematic**
  - *move verb group together*
  - *subject - verb - object*
  - *move negation in front of verb*

⇒ *Write rules by hand*

  - apply rules to test and training data
  - train standard *phrase-based* SMT system

System	BLEU
baseline system	25.2%
with manual rules	26.8%

Improved Translations

- we must also this criticism should be taken seriously .

$\longrightarrow$ we must also take this criticism seriously .

- i am with him that it is necessary , the institutional balance by means of a political revaluation of both the commission and the council to maintain .

$\longrightarrow$ i agree with him in this , that it is necessary to maintain the institutional balance by means of a political revaluation of both the commission and the council .

- thirdly , we believe that the principle of differentiation of negotiations note .

$\longrightarrow$ thirdly , we maintain the principle of differentiation of negotiations .

- perhaps it would be a constructive dialog between the government and opposition parties , social representative a positive impetus in the right direction .

$\longrightarrow$ perhaps a constructive dialog between government and opposition parties and social representative could give a positive impetus in the right direction .

Other Linguistically Syntax-Based Approaches

- *Reranking* phrase-based SMT output with syntactic features
  - create n-best list with phrase-based system
  - POS tag and parse candidate translations
  - rerank with syntactic features
  - see [Koehn, 2003] and JHU Workshop [Och et al., 2003]

- Incorporate syntax into decoder [Tillman and Ney, 2003]
  - Add finite-state control structure to allow long-distance movement of verbs in German-English translation.

Formal Advantages of Syntax-Based Translation

- Foundation in *well-understood* models from formal language theory (theoretical computer science).
  - Maybe they have some use after all

- *Computational complexity* is (in principle) just as much as we need to model linguistic phenomena, and no more.
  - Polynomial even with full reordering.
  - Caveat: no easy trick to speed it up as with phrase-based models.

- *Apply advances* made algorithms for statistical parsing.
  - Earley, CKY, etc.

Synchronous Context-Free Grammars

		formally syntax-based
	phrase-based	*hierarchical phrase-based [Chiang 2005]*
linguistically syntax-based	reordering + phrase-based [Collins et al. 2005]	syntax-based SCFG [Yamada & Knight 2002]

Synchronous Context-Free Grammars

- Finite-state transducers model *regular* language

- Regular tree transducers model *context-free* language

- Various guises of SCFG
  - Syntax-directed Transduction (Lewis and Stearns 1968)
  - Inversion Transduction Grammar (Wu 1995-1998)
  - Head Transducers (Alshawi et al. 2000)
  - Multitext Grammar (Melamed 2003)

Inversion Transduction Grammars

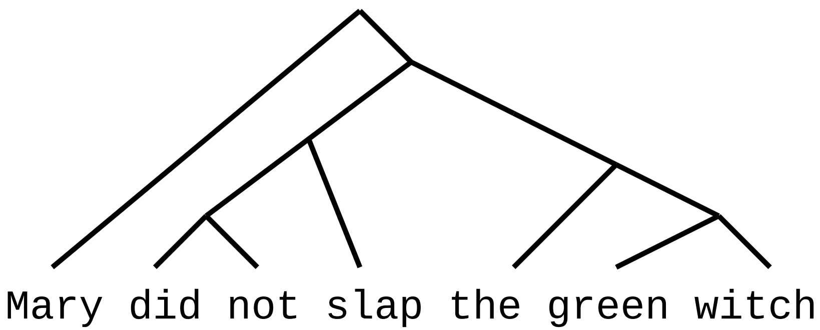
- Generation of *both* English and foreign trees [Wu, 1997]

- Rules (binary and unary)
  - $A \rightarrow A_1 A_2 \| A_1 A_2$
  - $A \rightarrow A_1 A_2 \| A_2 A_1$
  - $A \rightarrow e \| f$
  - $A \rightarrow e \| *$
  - $A \rightarrow * \| f$

$\Rightarrow$ *Common binary tree* required

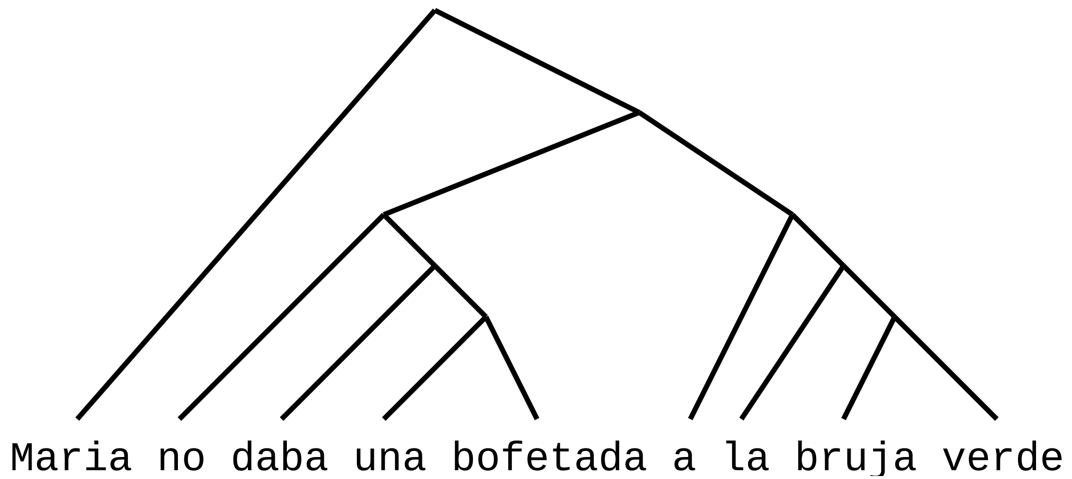
- limits the complexity of reorderings – polynomial in length, exponential in arity

Syntax Trees



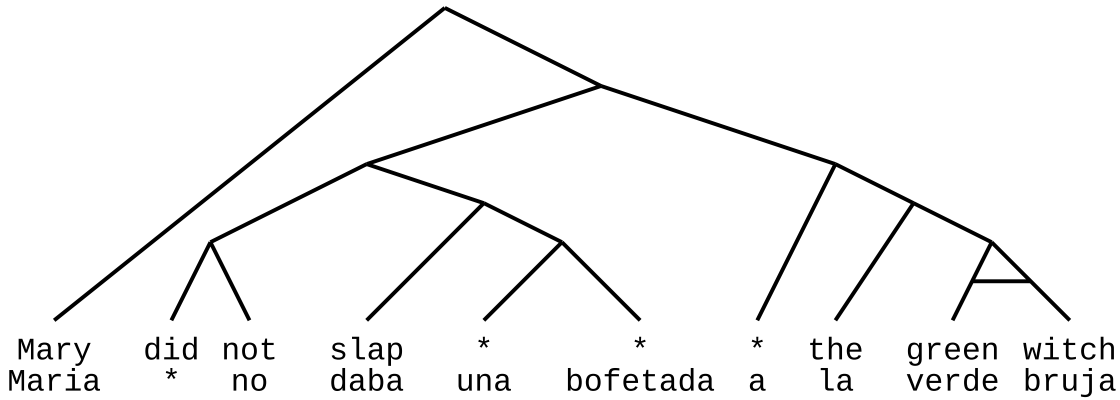
- English binary tree

Syntax Trees



- Spanish binary tree

Syntax Trees



- *Combined tree* with reordering of Spanish

Chiang: Hierarchical Phrase-based Model

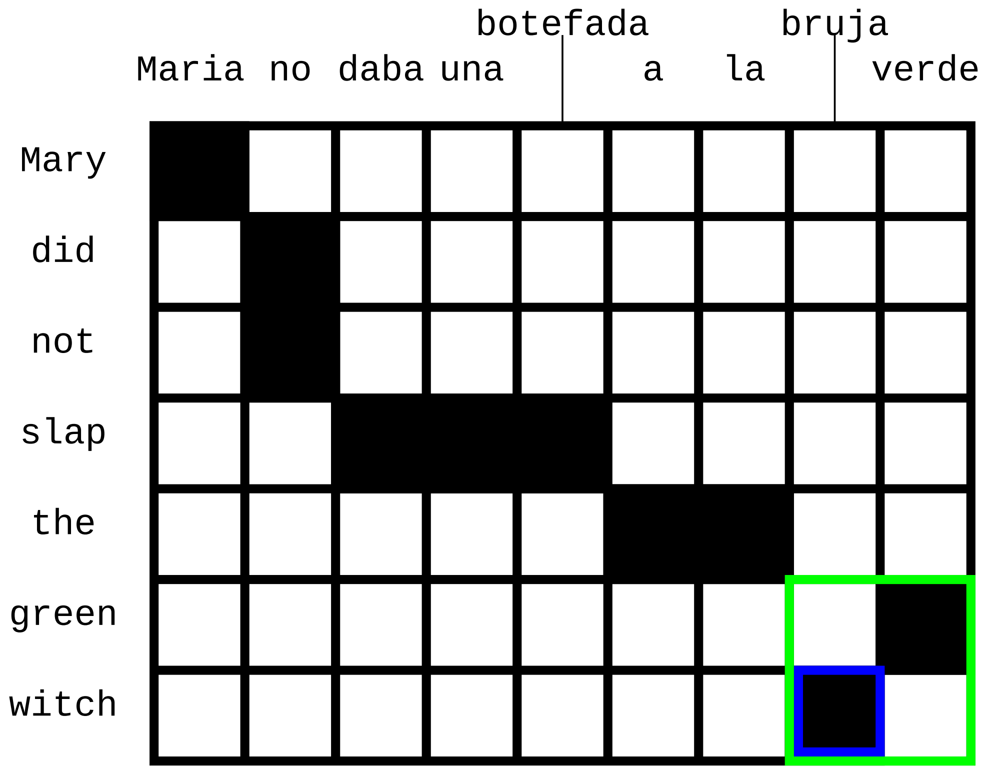
- **Chiang** [ACL, 2005] (best paper award!)
  - context free bi-grammar
  - *one non-terminal* symbol
  - right hand side of rule may include non-terminals and terminals

- *Competitive* with phrase-based models in 2005 DARPA/NIST evaluation

Types of Rules

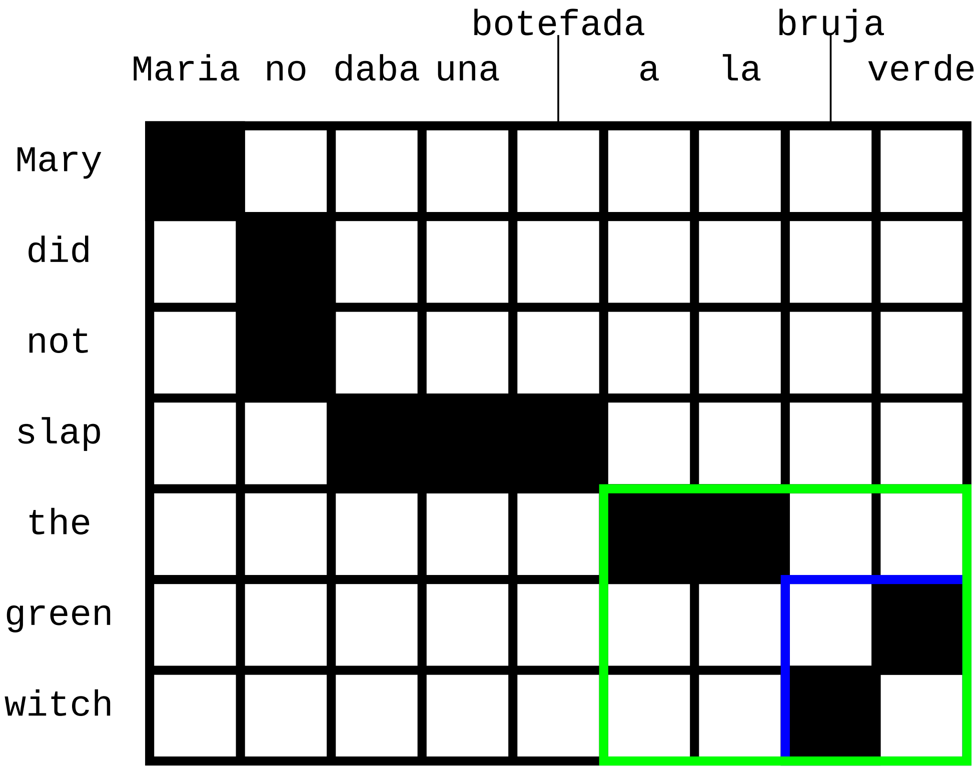
- *Word* translation
  - $X \rightarrow$ *maison* $\|$ *house*
- *Phrasal* translation
  - $X \rightarrow$ *daba una bofetada* $|$ *slap*
- Mixed non-terminal / terminal – *hierarchial phrases*
  - $X \rightarrow X_1$ *bleue* $\|$ *blue* X_1
  - $X \rightarrow$ *ne* X_1 *pas* $\|$ *not* X_1
  - $X \rightarrow X_1\ X_2 \| X_2$ *of* X_1
- Technical rules
  - $S \rightarrow S_1\ X_2 \| S_1\ X_2$
  - $S \rightarrow X_1 \| X_1$

Learning Hierarchical Rules



$X \rightarrow X$ verde || green X

Learning Hierarchical Rules



$$X \rightarrow \text{a la X} \parallel \text{the X}$$

Details of Chiang's Model

- Too many rules

$\rightarrow$ *filtering* of rules necessary

- *Efficient* parse decoding possible
  - hypothesis stack for each span of foreign words
  - only *one non-terminal* $\rightarrow$ hypotheses comparable
  - *length limit* for spans that do not start at beginning
  - m-gram language model integration increases complexity by $O(n^{2m})$

Language is not Context-Free!

- Maybe it's *mildly context-sensitive*?
  - Synchronous Tree-Adjoining Grammar [Shieber 1992, others]
  - Generalized Multitext Grammar [Melamed 2004]

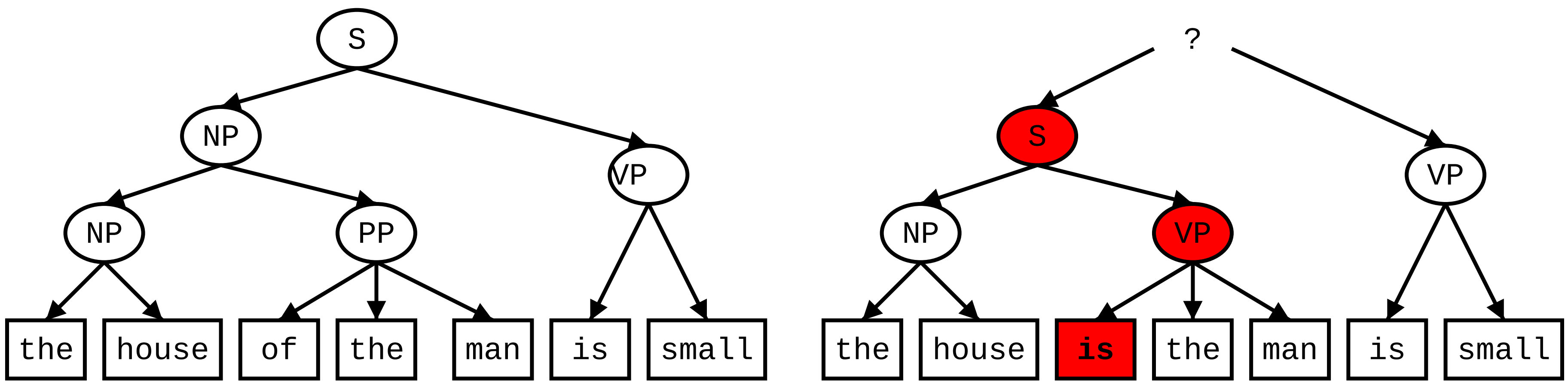
- Various transducer formalisms – [Knight & Graehl 2005] for overview.

Syntactic Language Model

		formally syntax-based
	phrase-based	hierarchical phrase-based (Chiang 2005)
linguistically syntax-based	reordering + phrase-based (Collins et al. 2005)	*syntax-based SCFG (Yamada & Knight 2002)*

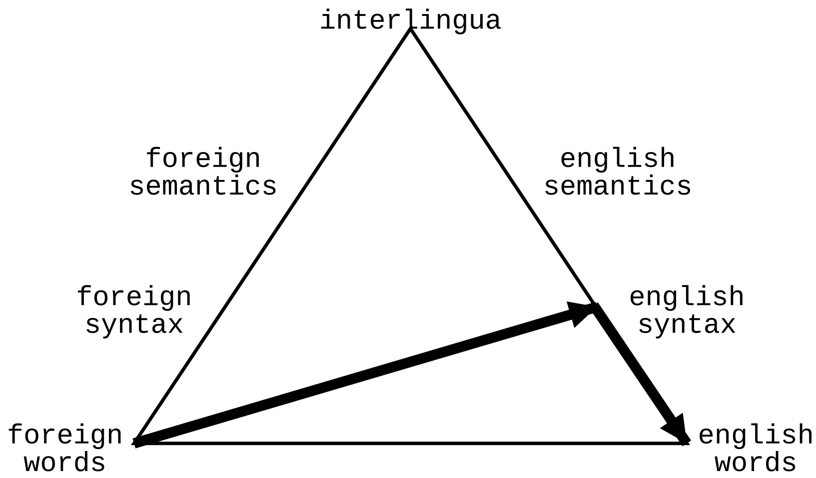
Syntactic Language Model

- *Good syntax tree* → good English
- Allows for *long distance constraints*



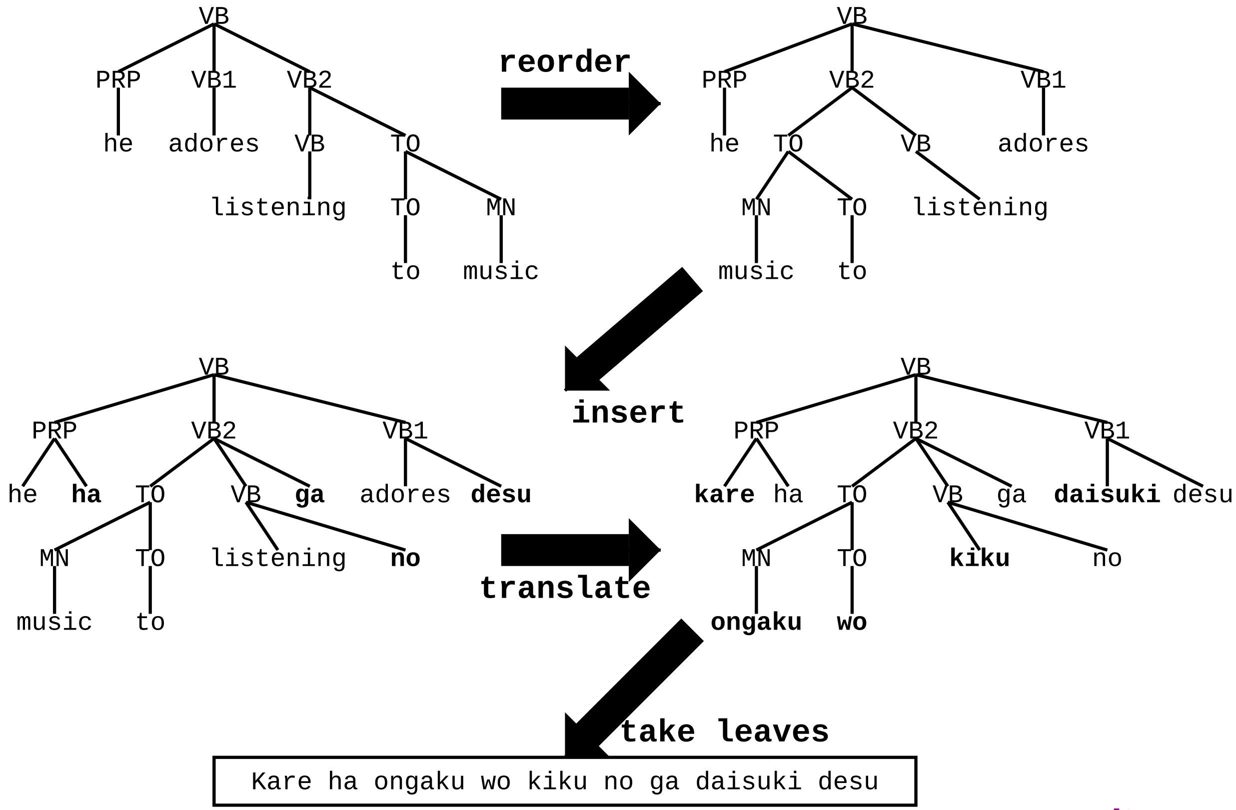
- Left translation preferred by syntactic LM

String to Tree Translation



- Use of English *syntax trees* [Yamada and Knight, 2001]
  - exploit *rich resources* on the English side
  - obtained with statistical parser [Collins, 1997]
  - *flattened tree* to allow more reorderings
  - works well with syntactic language model

Yamada and Knight [2001]



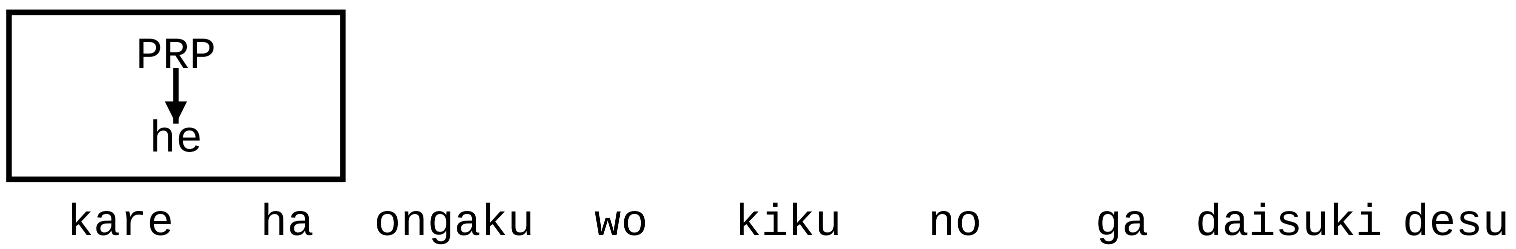
[from Yamada and Knight, 2001]

Reordering Table

Original Order	Reordering	p(reorder|original)
PRP VB1 VB2	PRP VB1 VB2	0.074
PRP VB1 VB2	**PRP VB2 VB1**	**0.723**
PRP VB1 VB2	VB1 PRP VB2	0.061
PRP VB1 VB2	VB1 VB2 PRP	0.037
PRP VB1 VB2	VB2 PRP VB1	0.083
PRP VB1 VB2	VB2 VB1 PRP	0.021
VB TO	VB TO	0.107
VB TO	**TO VB**	**0.893**
TO NN	TO NN	0.251
TO NN	**NN TO**	**0.749**

Decoding as Parsing

- **Chart Parsing**

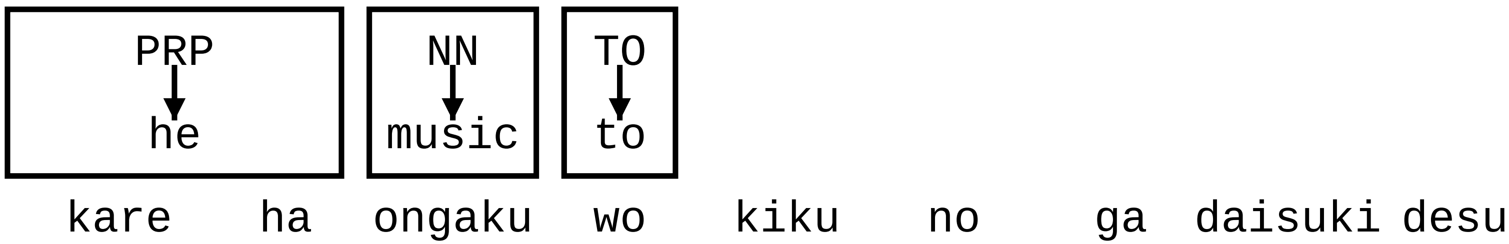


- Pick Japanese *words*

- Translate into *tree stumps*

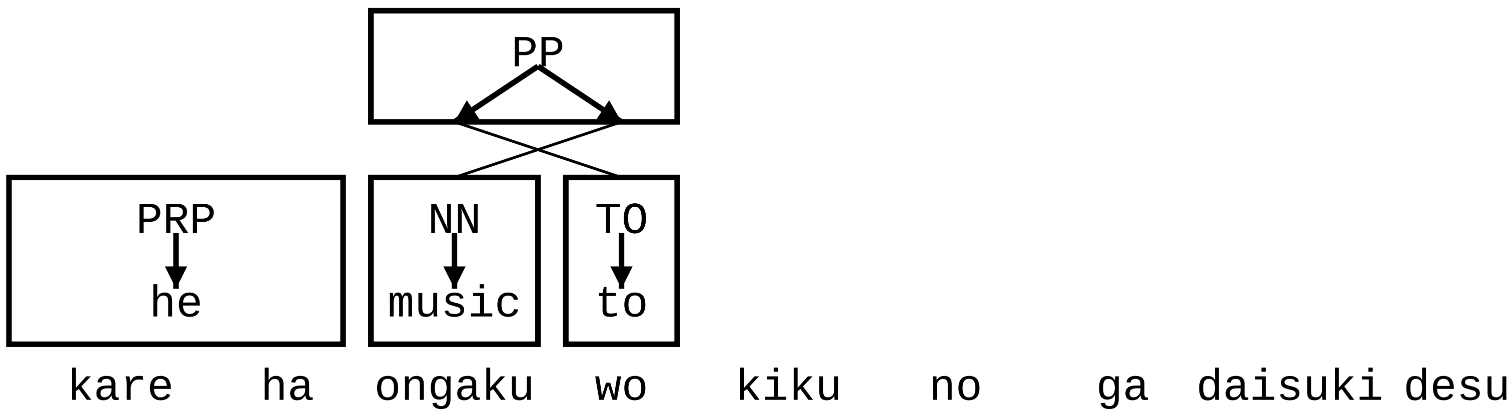
Decoding as Parsing

- Chart Parsing



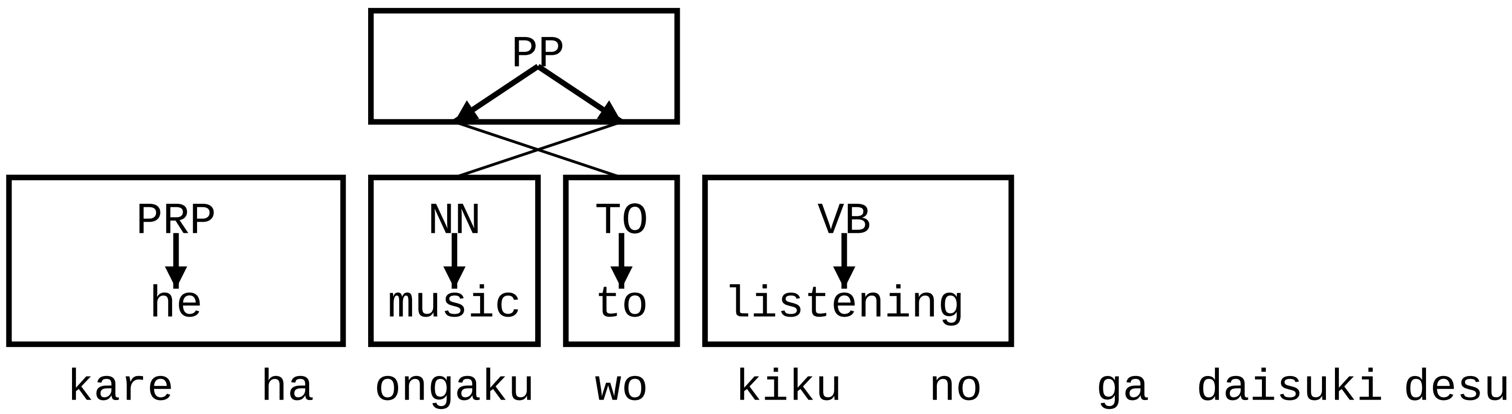
- Pick Japanese words
- Translate into tree stumps

Decoding as Parsing



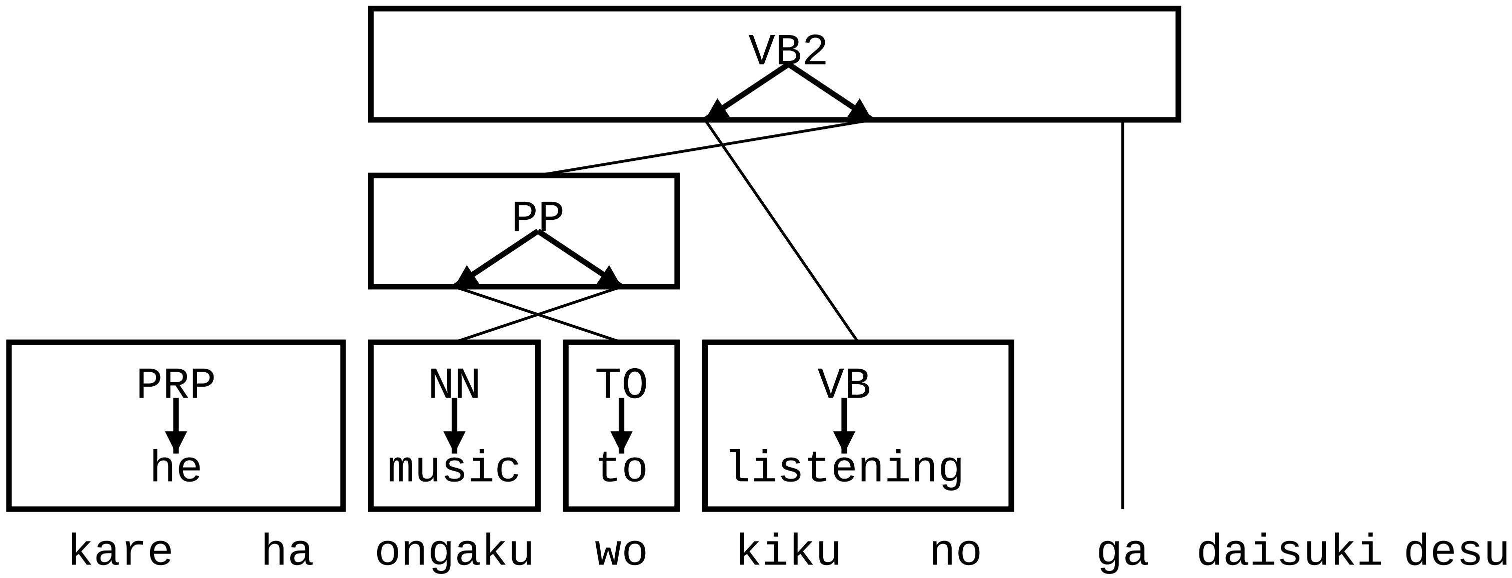
- Adding some *more entries*...

Decoding as Parsing

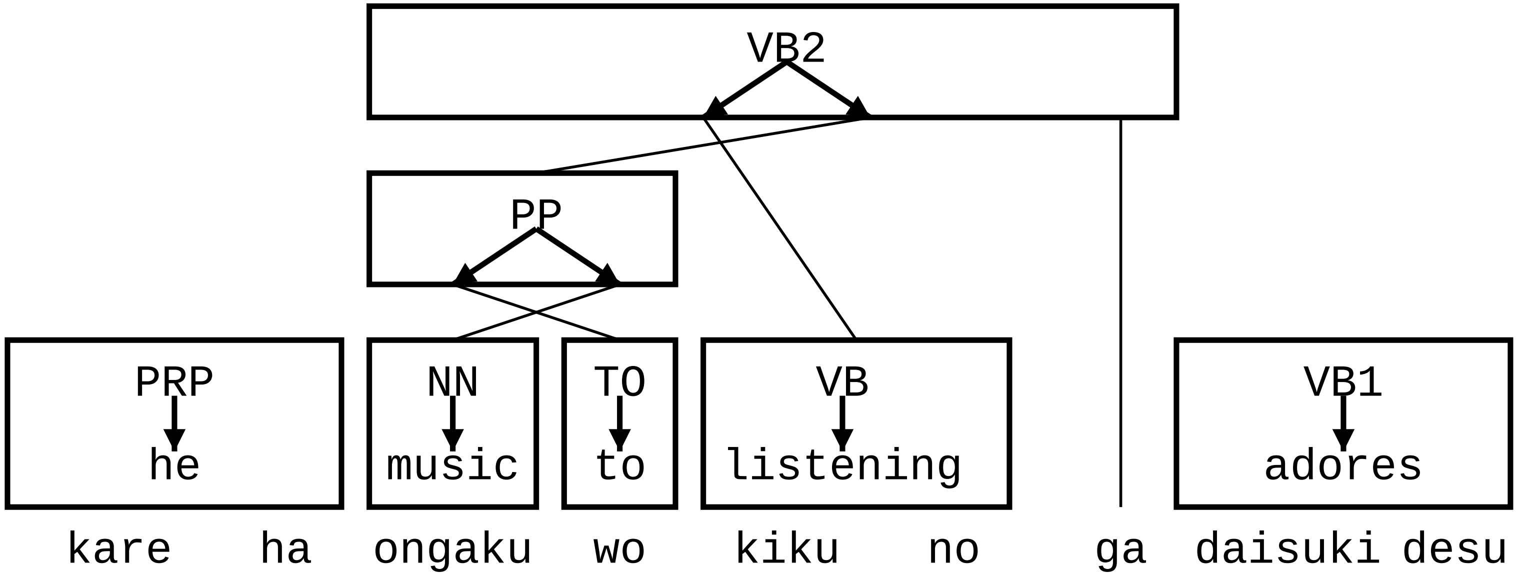


- *Combine entries*

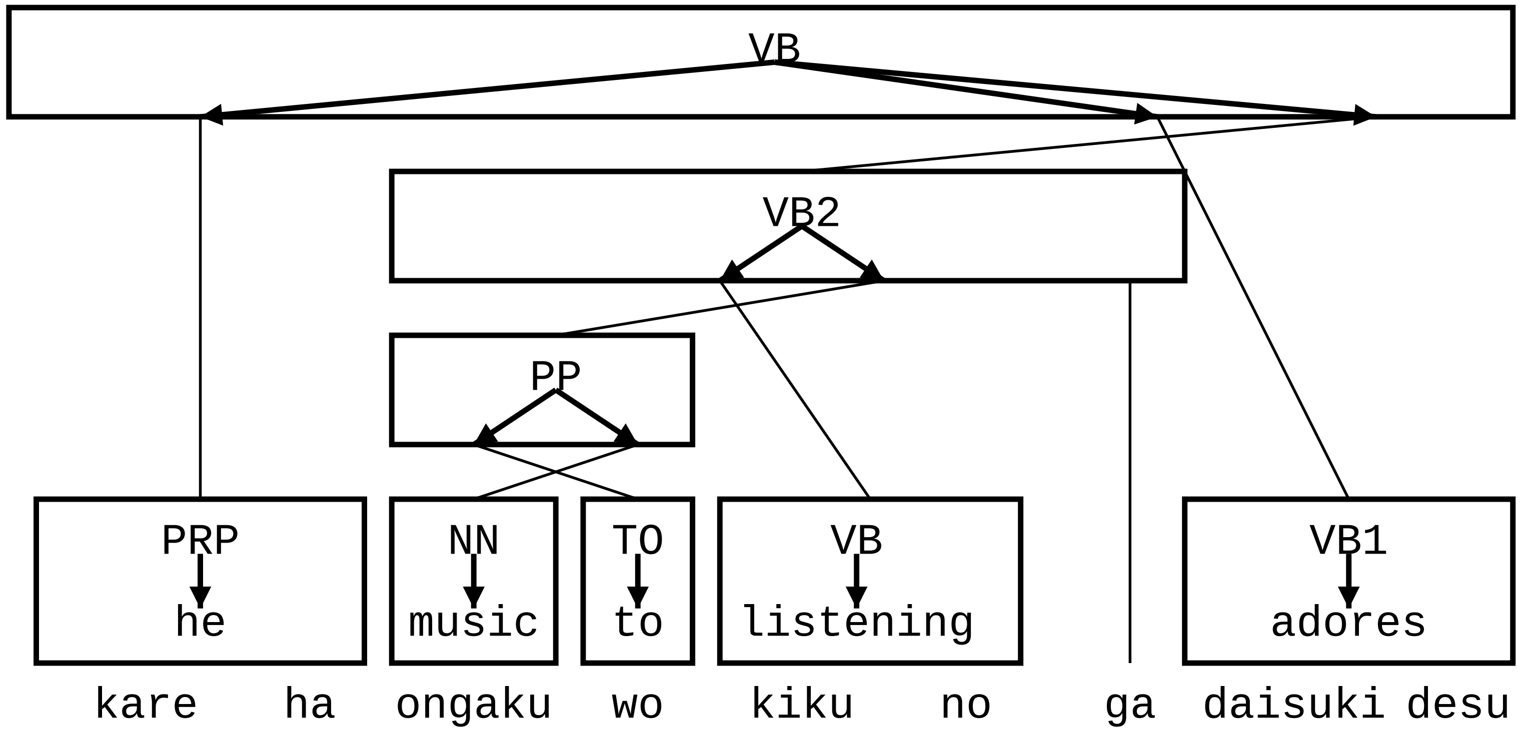
Decoding as Parsing



Decoding as Parsing



Decoding as Parsing



- *Finished* when all foreign words covered

Yamada and Knight: Training

- *Parsing* of the English side
  - using Collins statistical parser

- *EM training*
  - translation model is used to map training sentence pairs
  - EM training finds low-perplexity model

  $\rightarrow$ *unity of training and decoding* as in IBM models

Is the Model Realistic?

- Do English trees *match* foreign strings?
- Crossings between French-English [Fox, 2002]
  - 0.29-6.27 per sentence, depending on how it is measured
- Can be reduced by
  - *flattening tree*, as done by [Yamada and Knight, 2001]
  - detecting *phrasal* translation
  - *special treatment* for small number of constructions
- Most coherence between **dependency structures**

Other Syntax-Based Approaches

- ISI: extending work of Yamada/Knight
  - more *complex rules*
  - performance approaching phrase-based
- Prague: Translation via *dependency structures*
  - parallel Czech–English dependency treebank
  - tecto-grammatical translation model [EACL 2003]
- U.Alberta/Microsoft: *treelet translation*
  - translating from English into foreign languages
  - using dependency parser in English
  - project *dependency tree* into foreign language for training
  - map parts of the dependency tree ("treelets") into foreign languages

Syntax: Does it help?

- *Getting there*
  - for some languages competitive with best phrase-based systems
- *Some evidence*
  - work on reordering German
  - ISI: better for short sentences Chinese–English
  - automatically trained tree transfer systems promising
- Why not yet?
  - if real syntax, we need *good parsers* — are they good enough?
  - syntactic annotations add a level of *complexity*
  - $\rightarrow$ difficult to handle, slow to train and decode
  - few researchers good at statistical modeling and syntactic theories