Statistical Machine Translation
Lecture 5
Syntax-Based Models

Philipp Koehn

pkoehn@inf.ed.ac.uk

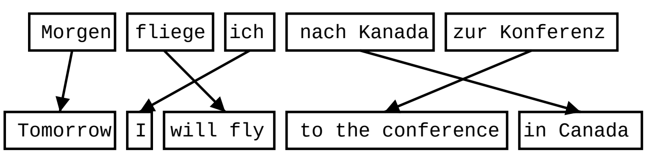
School of Informatics

University of Edinburgh

Outline

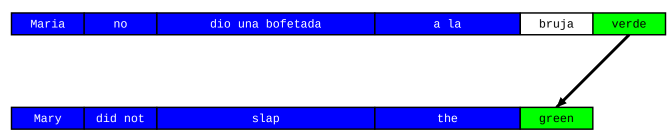
- Reminder: Modeling and Decoding
- Why Syntax?
- Yamada and Knight: translating into trees
- Wu: tree-based transfer
- Chiang: hierarchical transfer
- Koehn: clause structure
- Other approaches

Phrase-Based Translation Model



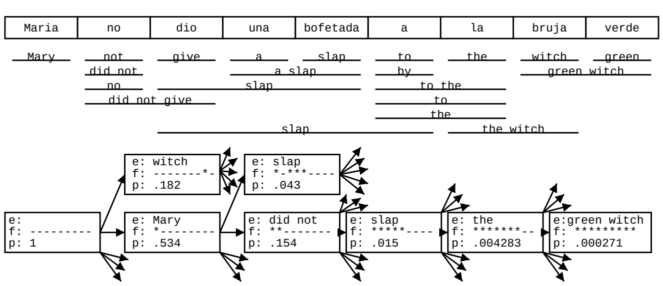
- Foreign input is segmented in phrases
  - any sequence of words, not necessarily linguistically motivated
- Each phrase is translated into English
- Phrases are reordered

Decoding



- Decoding process builds an English translation left to right, by picking foreign phrases to translate into English phrases

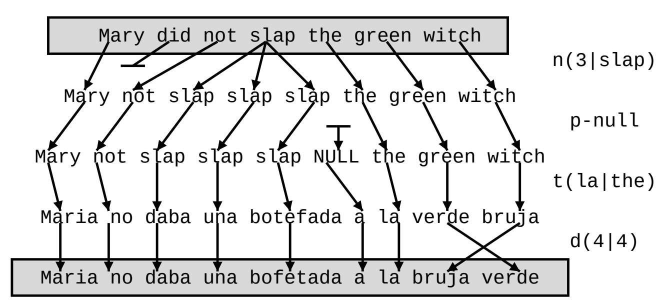
Search Space for Decoding Too Big



- Explosion of search space

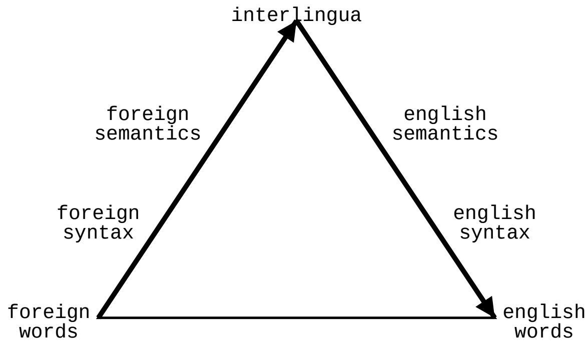
⇒ Pruning, Beam Search

Word-Based Translation Model



- Translation process is broken up into small step: word translation, reordering, duplication, insertion
- Decoding can be done similarly to phrase-based decoding

The Challenge of Syntax



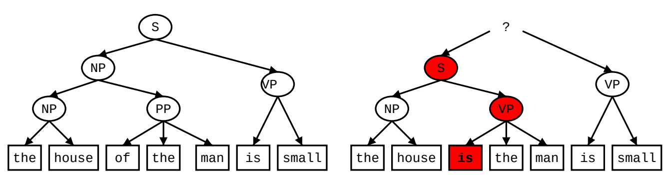
- The classical machine translation pyramid

Advantages of Syntax-Based Translation

- 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

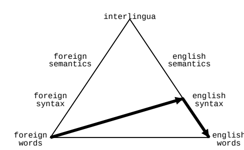
Syntactic Language Model

- Good syntax tree $\rightarrow$ good English
- Allows for long distance constraints



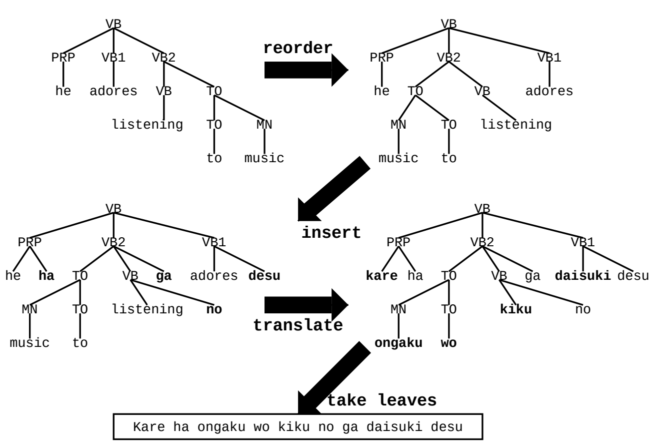
- Left translation preferred by syntactic LM

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]

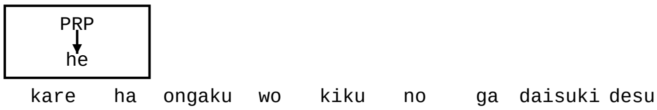


Reordering Table

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

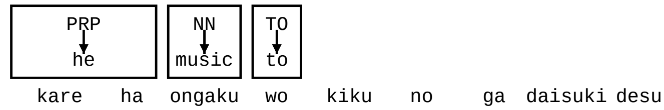
- Chart Parsing



- Pick Japanese words
- Translate into tree stumps

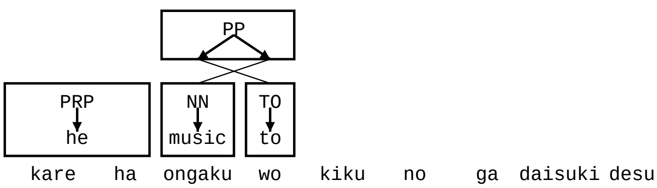
Decoding as Parsing

- Chart Parsing



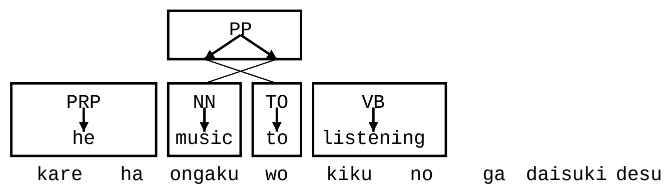
- Pick Japanese words
- Translate into tree stumps

Decoding as Parsing



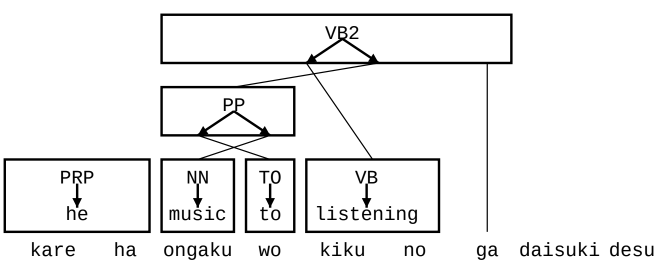
- Adding some more entries...

Decoding as Parsing

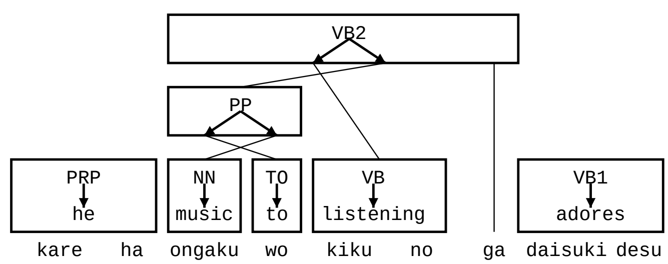


- Combine entries

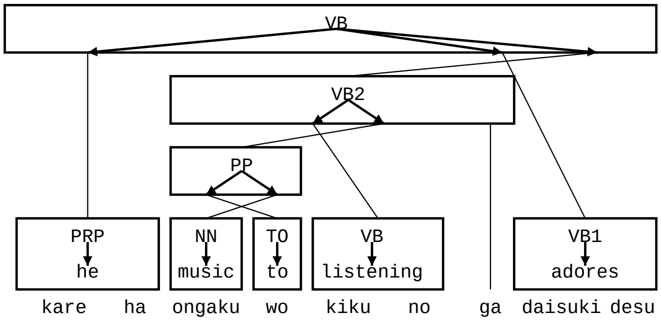
Decoding as Parsing



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
  - → unity of training and decoding as in IBM models

Is the Model Realistic?

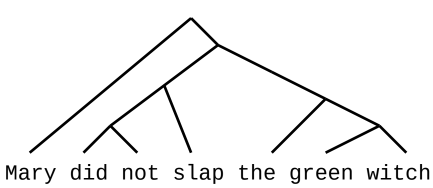
- Do English trees match foreign strings?
- Crossings between French-English [Fox, 2002]
  - 0.29-6.27 per sentence, depending on how it is measured
- Can be reduced by
  - flattening tree, as done by [Yamada and Knight, 2001]
  - detecting phrasal translation
  - special treatment for small number of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

Inversion Transduction Grammars

- Generation of both English and foreign trees [Wu, 1997]
- Rules (binary and unary)
  - $A \rightarrow A_1 A_2 \| A_1 A_2$
  - $A \rightarrow A_1 A_2 \| A_2 A_1$
  - $A \rightarrow e \| f$
  - $A \rightarrow e \| *$
  - $A \rightarrow * \| f$

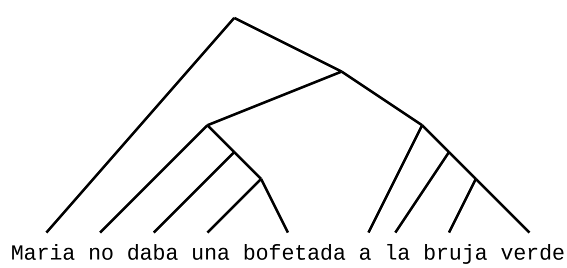
⇒ Common binary tree required
  - limits the complexity of reorderings

Syntax Trees



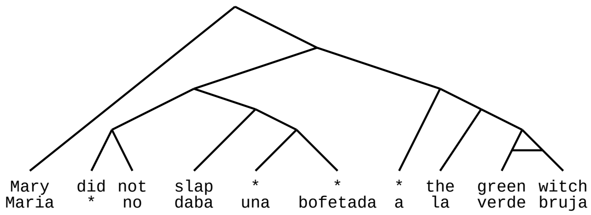
- English binary tree

Syntax Trees (2)



- Spanish binary tree

Syntax Trees (3)



- Combined tree with reordering of Spanish

Inversion Transduction Grammars

- Decoding by parsing (as before)
- Variations
  - may use real syntax on either side or both
  - may use multi-word units at leaf nodes
- Reordering constraints of ITG used in phrase-based systems

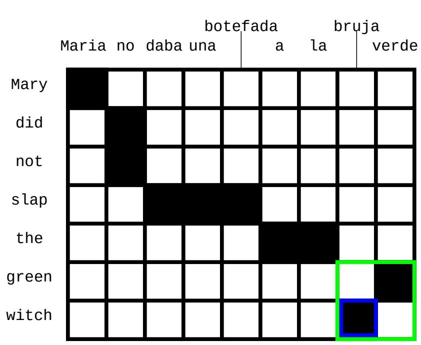
Chiang: Hierarchical Phrase Model

- 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

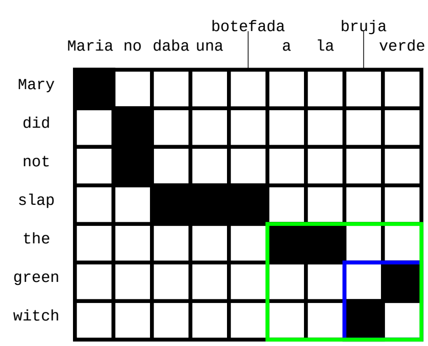
- Word translation
  - X → maison || house
- Phrasal translation
  - X → daba una bofetada | slap
- Mixed non-terminal / terminal
  - X → X bleue || blue X
  - X → ne X pas || not X
  - X → X1 X2 || X2 of X1
- Technical rules
  - S → S X || S X
  - S → X || X

Learning Hierarchical Rules



X → X verde || green X

Learning Hierarchical Rules



X → a la X || the X

Details

- Too many rules
  - → filtering of rules necessary
- Efficient parse decoding possible
  - hypothesis stack for each span of foreign words
  - only one non-terminal → hypotheses comparable
  - length limit for spans that do not start at beginning

Syntax-Aided Phrase-Based MT [Koehn]

- Approach:
  - stick with phrase-based system
  - special treatment for special syntactic problems
- Noun Phrase Translation
- Clause Level Restructuring

Clause Level Restructuring

- 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
- ACL 2005 paper [Collins, Koehn, Kucerova]

Clause Structure

```
S   PPER-SB  Ich     I                                           MAIN
    VAFIN-HD werde    will                                       CLAUSE
    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                     SUB-
                     PPER-SB Sie     you                         ORDINATE
                     VP-OC   PDS-OA  das    that                 CLAUSE
                             ADJD-MO eventuell     perhaps
                             PP-MO   APRD-MO  bei    in
                                     ART-DA   der    the
                                     NN-NK    Abstimmung  vote
                             VVINF   uebernehmen    include
                     VMFIN   koennen    can
$.  .                .
```

- 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 $\longrightarrow$ English

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

$\Rightarrow$ 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 .

$\rightarrow$ 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 .

$\rightarrow$ 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 .

$\rightarrow$ 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 .

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

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

Other Syntax-Based Approaches (2)

- 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]
- JHU Summer workshop 2005
  - final presentations this week
  - tools for syntax-based SMT

Syntax: Does it help?

- Not yet
  - best systems still phrase-based, treat words as tokens
- Well, maybe...
  - work on reordering German
  - 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 understand syntactic theories