“Efficient Top-down BTG Parsing for Machine Translation Preordering”

Tetsuji Nakagawa (Google), ACL-IJCNLP 2015

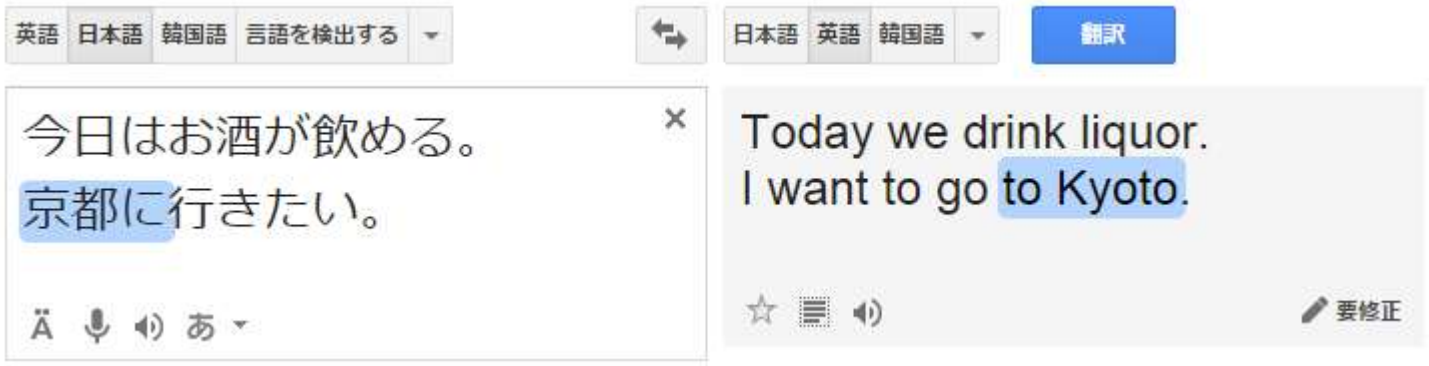
29 August 2015

Sho Hoshino (D2, NII Miyao lab.)

hoshino@nii.ac.jp

Reordering for Machine Translation

- A task to resemble/predict target word orders given a source input:

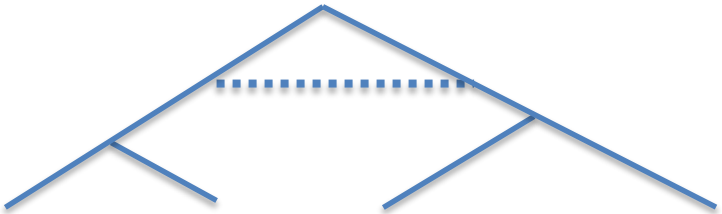


(by http://translate.google.com)

- Why do we say *"**I want to go to Kyoto**"*, but not *"Kyoto to go to want I"* ?

Take-home Messages

- **Reordering ＝ bilingual parsing** (*biparsing*)



大判　焼き　を　食べたい

reordering

を　食べたい　大判　焼き

- Machine translation ≒ (almost) parsing

monotonic translation

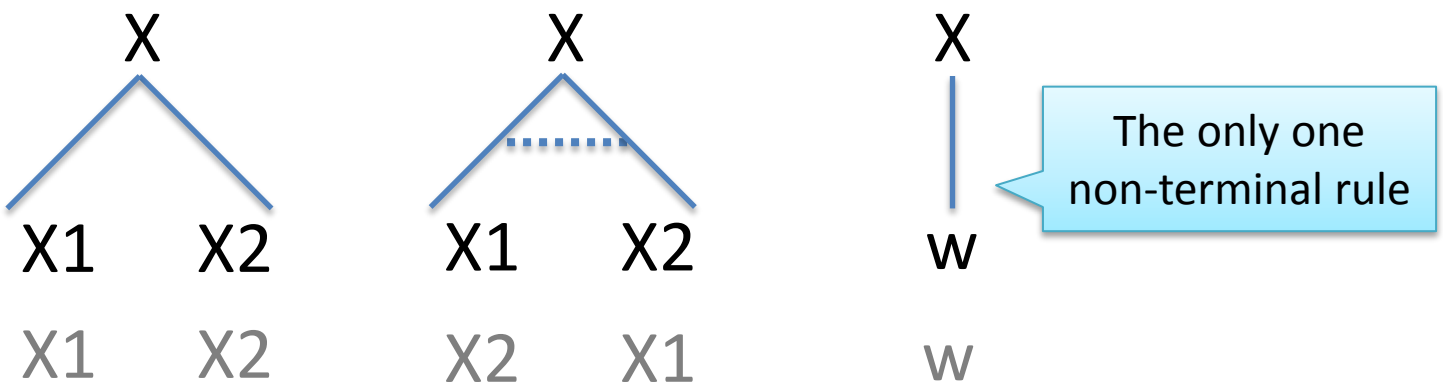
I　want　to　eat　oobanyaki

This Work

- "Supervised biparsing with inexact search"
  - BTG parsing without syntactic parsers [Neubig+ 2012]
- **Two contributions:**
  - Proposed a top-down algorithm for BTG parsing
    - Replaced CYK: **10 times faster!**
  - Reduced complexity from $\mathrm{O}(n^5)$ to $\mathrm{O}(kn)$
    - Perceptron + early update + beam search [width: *k*]
- Better BLEU scores for various language pairs

BTG: Bracketing Transduction Grammars [Wu 1997]

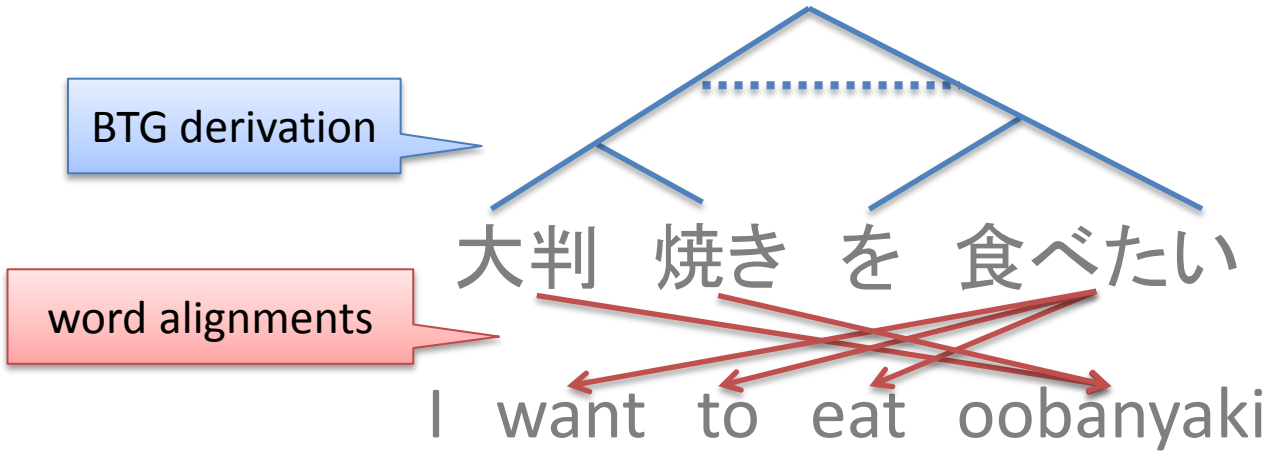
- A minimal ITG (Inversion Transduction Grammars)
  - One of the simplest synchronous grammars



  - Binary decisions express {**reversed** or **not**} reordering information

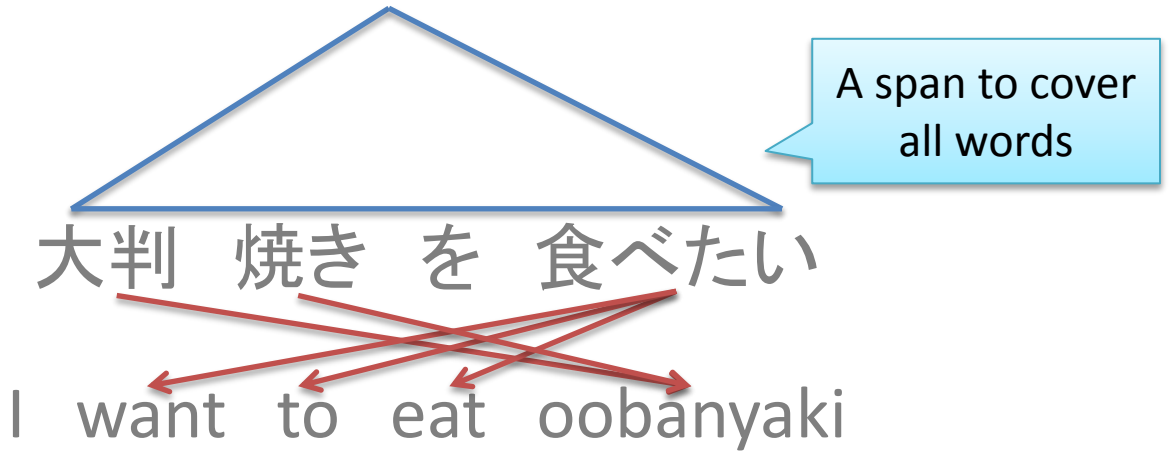
How to create a BTG tree?

- An incremental top-down algorithm
  - Similar to previous bottom-up [Huang+ 2009]



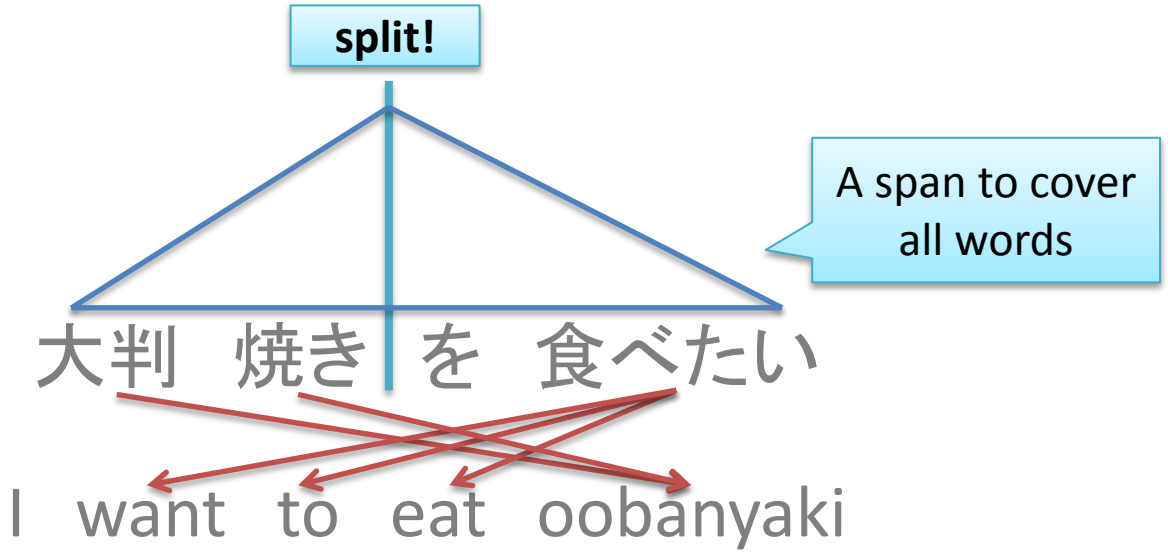
How to create a BTG tree?

- An incremental top-down algorithm
  - Similar to previous bottom-up [Huang+ 2009]



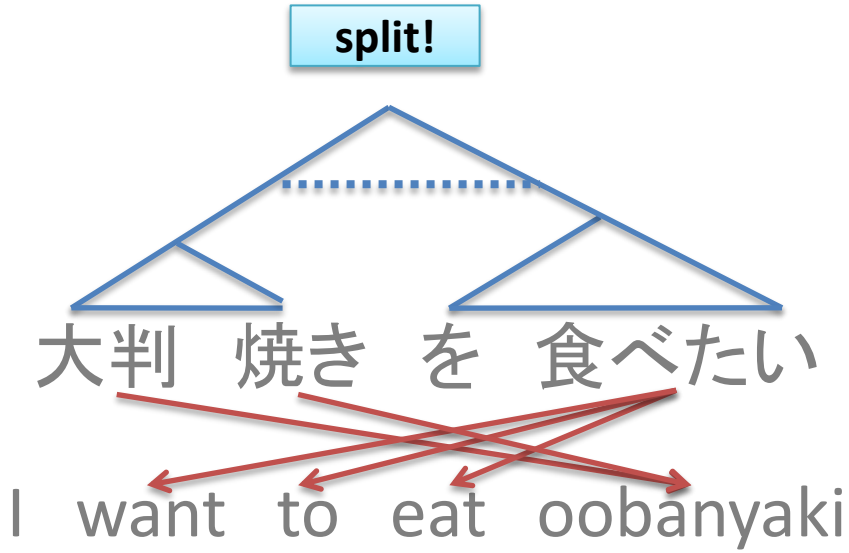
How to create a BTG tree?

- An incremental top-down algorithm
  - Similar to previous bottom-up [Huang+ 2009]



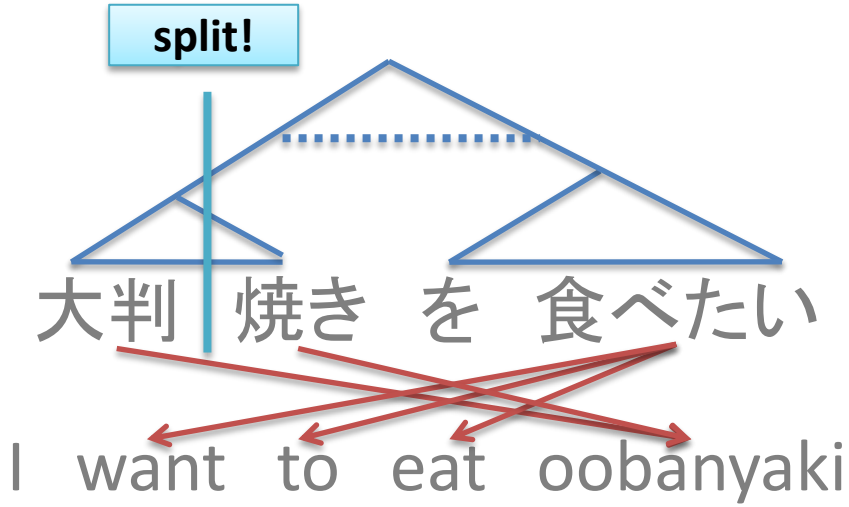
How to create a BTG tree?

- An incremental top-down algorithm
  - Similar to previous bottom-up [Huang+ 2009]



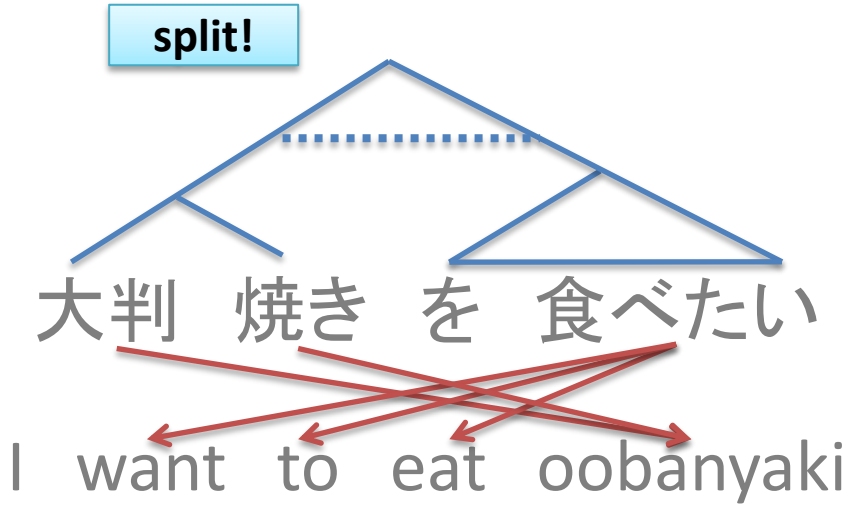
How to create a BTG tree?

- An incremental top-down algorithm
  - Similar to previous bottom-up [Huang+ 2009]



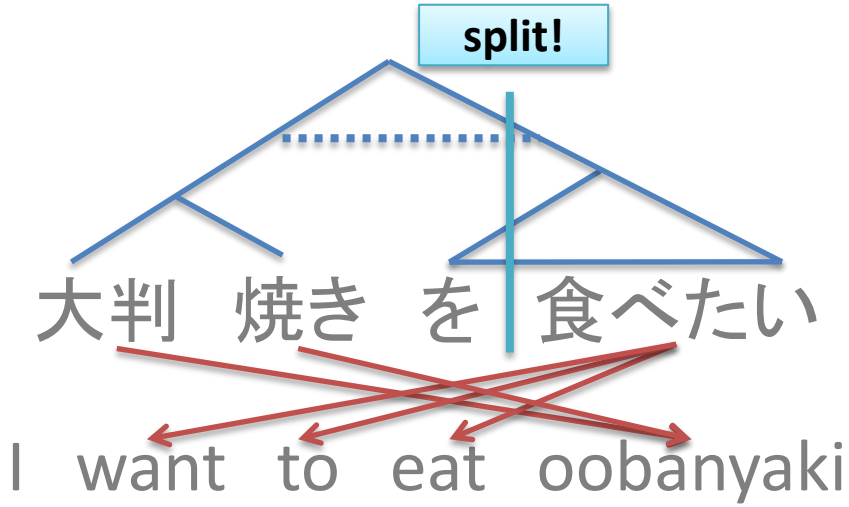
How to create a BTG tree?

- An incremental top-down algorithm
  - Similar to previous bottom-up [Huang+ 2009]



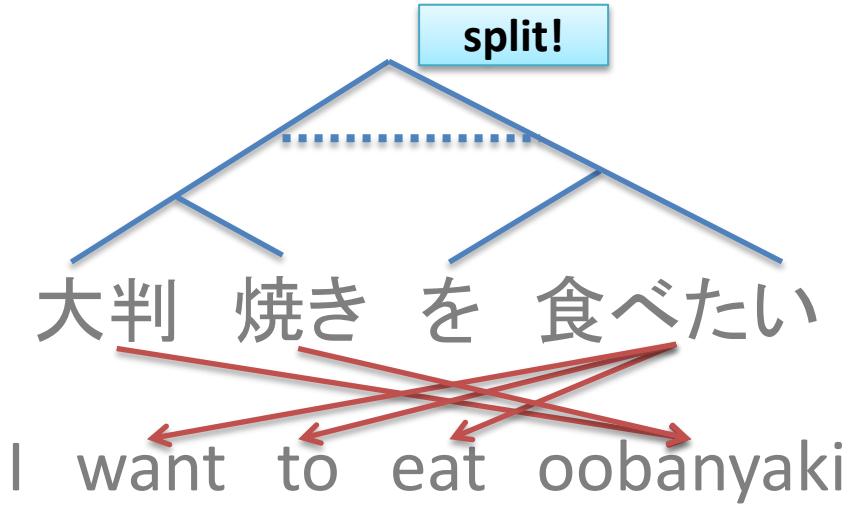
How to create a BTG tree?

- An incremental top-down algorithm
  - Similar to previous bottom-up [Huang+ 2009]



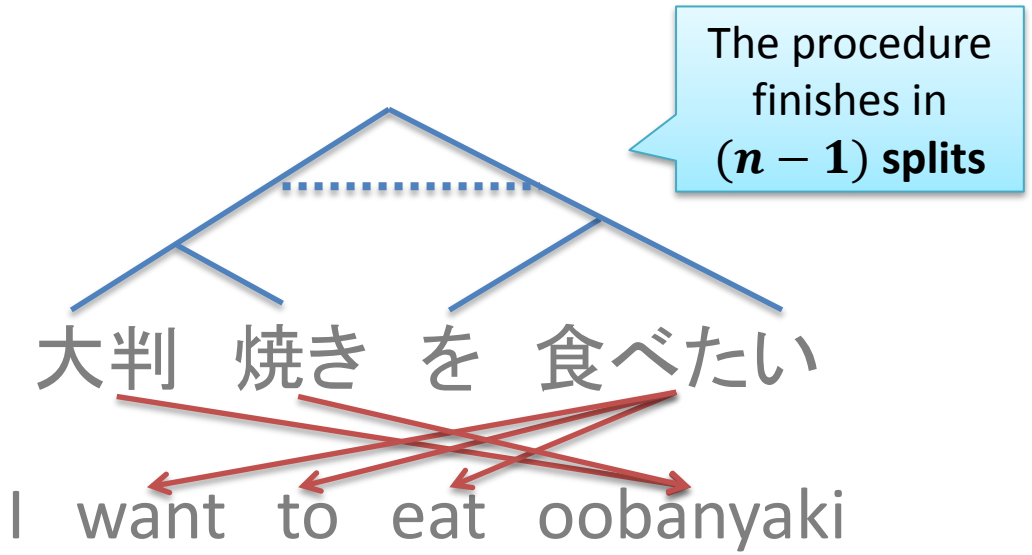
How to create a BTG tree?

- An incremental top-down algorithm
  - Similar to previous bottom-up [Huang+ 2009]



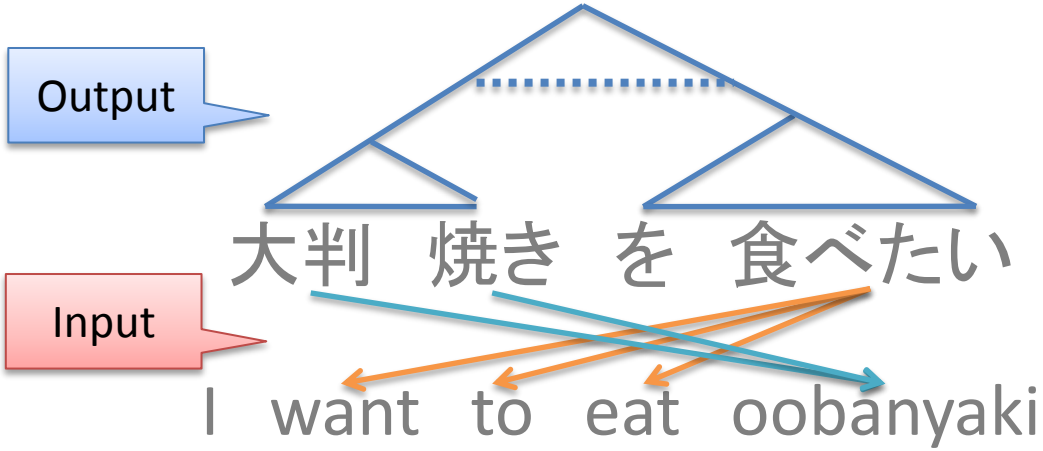
How to create a BTG tree?

- An incremental top-down algorithm
  - Similar to previous bottom-up [Huang+ 2009]



Oracle {**reversed** or **not**} reordering

- Defined by using word alignments
- **Reversed** if left-hand side > right-hand side



- Ambiguous examples are removed from data
  - because alignments are not totally ordered set

Learning BTG parser from scratch

- Supervised learning using oracle reordering
  - Latent variable perceptron [Sun+ 2009]
    - x: input sentence, $\acute{x}$: reordered sentence
    - **latent** z: BTG tree, m: BTG node , $\hat{z}$: the best BTG tree

$$\hat{z} = \underset{z \in Z(x)}{\operatorname{argmax}} \sum_{m \in Nodes(z)} \lambda \cdot \theta(m)$$

$$\acute{x} = \mathrm{Proj}(\hat{z})$$

parameter

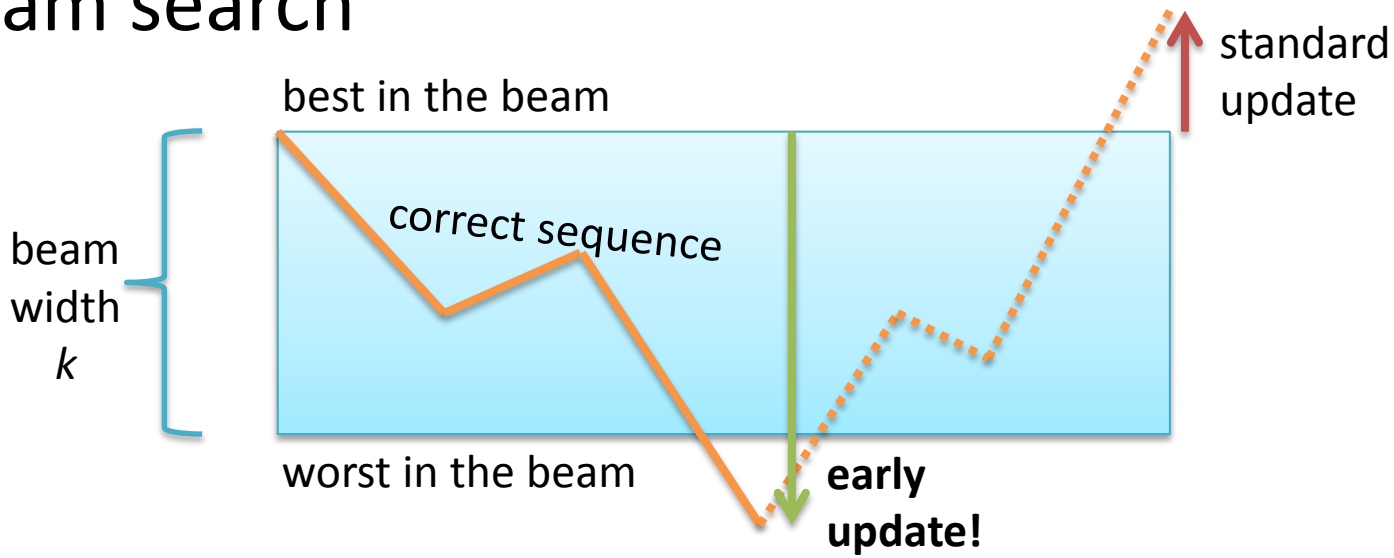
feature function

  - With early update technique [Collins and Roark 2004]

Perceptron with Inexact Search

[Collins and Roark 2004; Huang+ 2012]

- Update perceptron at the first violation of beam search



- Good performance in POS tagging [F1 97.35] and dependency parsing [F1 92.09]

Features

- Basic template [Neubig+ 2012]

$o(q-p), oBalance(p,q,r),$
$ox^w_{p-1}, ox^w_p, ox^w_{r-1}, ox^w_r, ox^w_{q-1}, ox^w_q, ox^w_p ox^w_{q-1}, ox^w_{r-1} ox^w_r,$
$ox^p_{p-1}, ox^p_p, ox^p_{r-1}, ox^p_r, ox^p_{q-1}, ox^p_q, ox^p_p ox^p_{q-1}, ox^p_{r-1} ox^p_r,$
$ox^c_{p-1}, ox^c_p, ox^c_{r-1}, ox^c_r, ox^c_{q-1}, ox^c_q, ox^c_p ox^c_{q-1}, ox^c_{r-1} ox^c_r,$

word surface form
part-of-speech tag
word class
Brown cluster=256

- Additional template

$o\min(r-p,5)\min(q-r,5), oó, oód,$
$ox^w_{p-1}ox^w_p, ox^w_p ox^w_{r-1}, ox^w_p ox^w_r, ox^w_{r-1}ox^w_{q-1}, ox^w_r ox^w_{q-1}, ox^w_{q-1}ox^w_q,$
$ox^w_{r-2}ox^w_{r-1}ox^w_r, ox^w_p ox^w_{r-1}ox^w_r, ox^w_{r-1}ox^w_r ox^w_{q-1}, ox^w_{r-1}ox^w_r ox^w_{r+1},$
$ox^w_p ox^w_{r-1}ox^w_r ox^w_{q-1}, oódx^w_p, oódx^w_{r-1}, oódx^w_r, oódx^w_{q-1}, oódx^w_p oódx^w_{q-1},$
$ox^p_{p-1}ox^p_p, ox^p_p ox^p_{r-1}, ox^p_p ox^p_r, ox^p_{r-1}ox^p_{q-1}, ox^p_r ox^p_{q-1}, ox^p_{q-1}ox^p_q,$
$ox^p_{r-2}ox^p_{r-1}ox^p_r, ox^p_p ox^p_{r-1}ox^p_r, ox^p_{r-1}ox^p_r ox^p_{q-1}, ox^p_{r-1}ox^p_r ox^p_{r+1},$
$ox^p_p ox^p_{r-1}ox^p_r ox^p_{q-1}, oódx^p_p, oódx^p_{r-1}, oódx^p_r, oódx^p_{q-1}, oódx^p_p oódx^p_{q-1},$
$ox^c_{p-1}ox^c_p, ox^c_p ox^c_{r-1}, ox^c_p ox^c_r, ox^c_{r-1}ox^c_{q-1}, ox^c_r ox^c_{q-1}, ox^c_{q-1}ox^c_q,$
$ox^c_{r-2}ox^c_{r-1}ox^c_r, ox^c_p ox^c_{r-1}ox^c_r, ox^c_{r-1}ox^c_r ox^c_{q-1}, ox^c_{r-1}ox^c_r ox^c_{r+1},$
$ox^c_p ox^c_{r-1}ox^c_r ox^c_{q-1}, oódx^c_p, oódx^c_{r-1}, oódx^c_r, oódx^c_{q-1}, oódx^c_p oódx^c_{q-1},$

more and more combinations

Experimental Results (in BLEU,k=20)

	No-preordering	Manual rules	[Neubig+ 2012]	This Work
nl-en 蘭英	34.01	-	33.83	**35.49**
en-nl 英蘭	25.33	-	25.30	**25.82**
en-fr 英仏	25.86	-	26.50	**26.75**
en-ja 英日	13.80	**18.68**	17.40	17.66
en-es 英西	29.50	-	29.70	**30.26**
fr-en 仏英	32.33	-	32.43	**33.00**
hi-en 印英	19.86	-	24.24	**24.98**
ja-en 日英	10.31	14.02	14.59	**14.84**
ko-en 韓英	14.13	-	18.65	**19.67**
tr-en 土英	18.26	-	22.80	**23.91**
ur-en ウ英	14.48	-	16.62	**17.65**
cy-en ウェ英	41.68	-	41.79	**41.95**

Experimental Results (in BLEU, k=20)

ほぼ全ての言語対で最高精度

	No-preordering	Manual rules	[Neubig+ 2012]	This Work
nl-en 蘭英	34.01	-	33.83	**35.49**
en-nl 英蘭	25.33	-	25.30	**25.82**
en-fr 英仏	25.86	英日は例外	26.50	**26.75**
en-ja 英日	13.80	**18.68**	17.40	17.66
en-es 英西	29.50	-	29.70	**30.26**
fr-en 仏英	32.33	-	32.43	**33.00**
hi-en 印英	19.86	-	24.24	**24.98**
ja-en 日英	10.31	14.02	14.59	**14.84**
ko-en 韓英	14.13	-	18.65	**19.67**
tr-en 土英	18.26	-	22.80	**23.91**
ur-en ウ英	14.48	-	16.62	**17.65**
cy-en ウェ英	41.68	-	41.79	**41.95**

Summary

- BTG parsing (biparsing) with perceptron
- A top-down algorithm to replace CYK
- 10 timers faster & better translation accuracy

- Google Translate will be improved (already?)