

# SYNTACTIC REORDERING FOR JAPANESE-TO-ENGLISH TRANSLATION

JASON KATZ-BROWN  
([JASONKB@GOOGLE.COM](mailto:JASONKB@GOOGLE.COM))  
MICHAEL COLLINS  
([MCOLLINS@CSAIL.MIT.EDU](mailto:MCOLLINS@CSAIL.MIT.EDU))  
MIT CSAIL



# MIT SYSTEM OVERVIEW

- Japanese->English Moses system.
- Two techniques to reorder Japanese during preprocessing.
  1. Just **Reverse** every word
  2. Flatten a **Cabocha** dependency tree
- **Reverse** system gives highest human eval scores of statistical systems in Patent Translation Task.



# REORDERING IN PREPROCESSING

- Reorder Japanese words into an English word order during preprocessing.
- From: 私はケーキを食べました (I cake ate)
- To: は私たまし食べをケーキ (I ate cake)
- Then word alignment and decoding can be monotonic.
- Collins (2006) introduced for German->English.



# REVERSE REORDERING

- English is mostly head-initial.
- Japanese is mostly head-final.
- Let's reverse the word order!



# REVERSE REORDERING

- Keep punctuation marks where they are.
- Keep everything before topic marker は at the beginning.
- Reverse everything.
- 嬉しい私は ケーキを食べました  
は私い嬉し たまし食べをケーキ (Good!)
- 嬉しい私が ケーキを食べました  
たまし食べをケーキ が私い嬉し (Bad!)



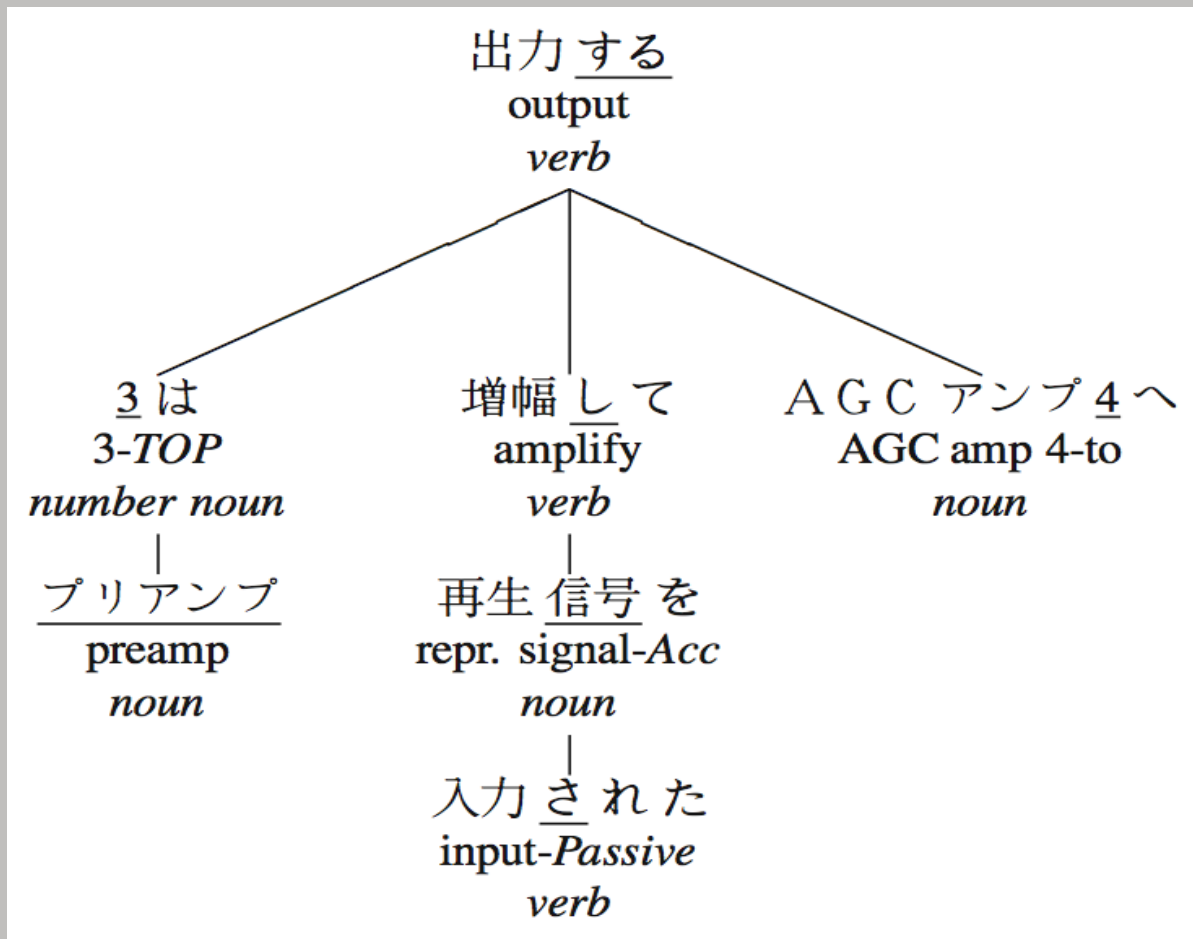
# REVERSE REORDERING

- From: プリアンプ 3 は入力された再生信号を増幅して AGC アンプ 4 へ出力する。
- To: は 3 プリアンプする出力へ 4 アンプ AGC てし増幅を信号再生たれさ入力。
- *The pre-amplifier 3 outputs to the agc amplifier 4 amplifies the input reproduction signal. (DistortionLimit=0)*
- *The preamplifier 3 and outputs it to the AGC amplifier 4 amplifies the input reproduction signal. (DistortionLimit=9)*



# CABOCHA REORDERING

- Flatten a Cabocha dependency tree.



は3プリアンプする出力てし増幅を  
信号再生たれさ入力へ4アンプAGC。

1. Flatten verbs after their subjects, before their objects.
2. Flatten all others before their children.
3. Reverse within chunks.



# REORDERING EXPERIMENTS

<i>DistortionLimit</i>	BASELINE	REV	CABOCHA	Seconds per sentence <sup>a</sup>
0	20.86	20.32	21.61	2.2
6	23.76	25.44	24.79	5.0
9	25.24	25.49	25.12	7.8
unlimited	26.07	25.08	24.58	37.2

- In monotonic translation (DistortionLimit=0), **Cabocha** reordering is best.
- With some distortion allowed in decoder (DistortionLimit=0), **Reverse** reordering is best.
- With unlimited distortion allowed in decoder (DistortionLimit= $\infty$ ), **no reordering** is best.





# NTCIR PATENT MT RESULTS

Formal Run ID	Preorder	BLEU	Adequacy	Fluency	Seconds per sentence
MIT (1)	REV	27.14	3.15	3.66	8.4
MIT (2)	CABOCHA	26.44	(N/A)	(N/A)	8.0
Moses	BASELINE	27.14	2.81	3.55	18.2
tsbmt	(N/A)	26.44	3.81	3.94	0.23

Our system compared to Moses baseline:

- Twice as fast decoding.
- Same BLEU score.
- Improved manual evaluation.

Our system compared to rule-based:

- Much worse :-)

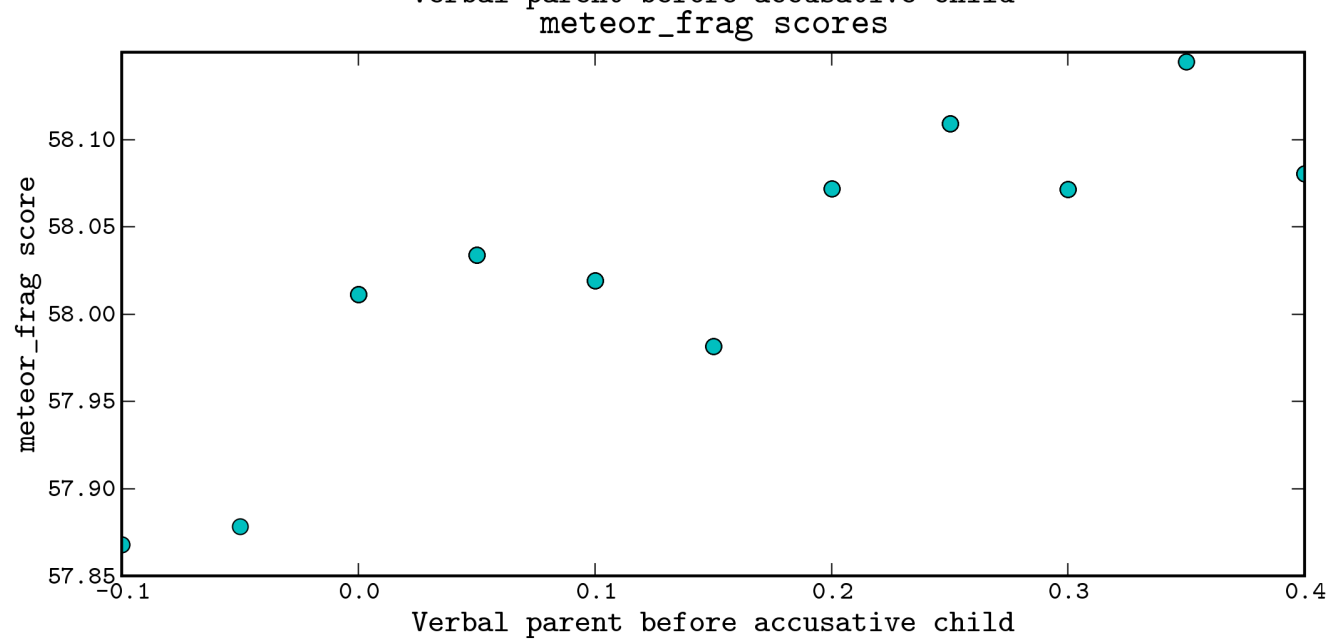
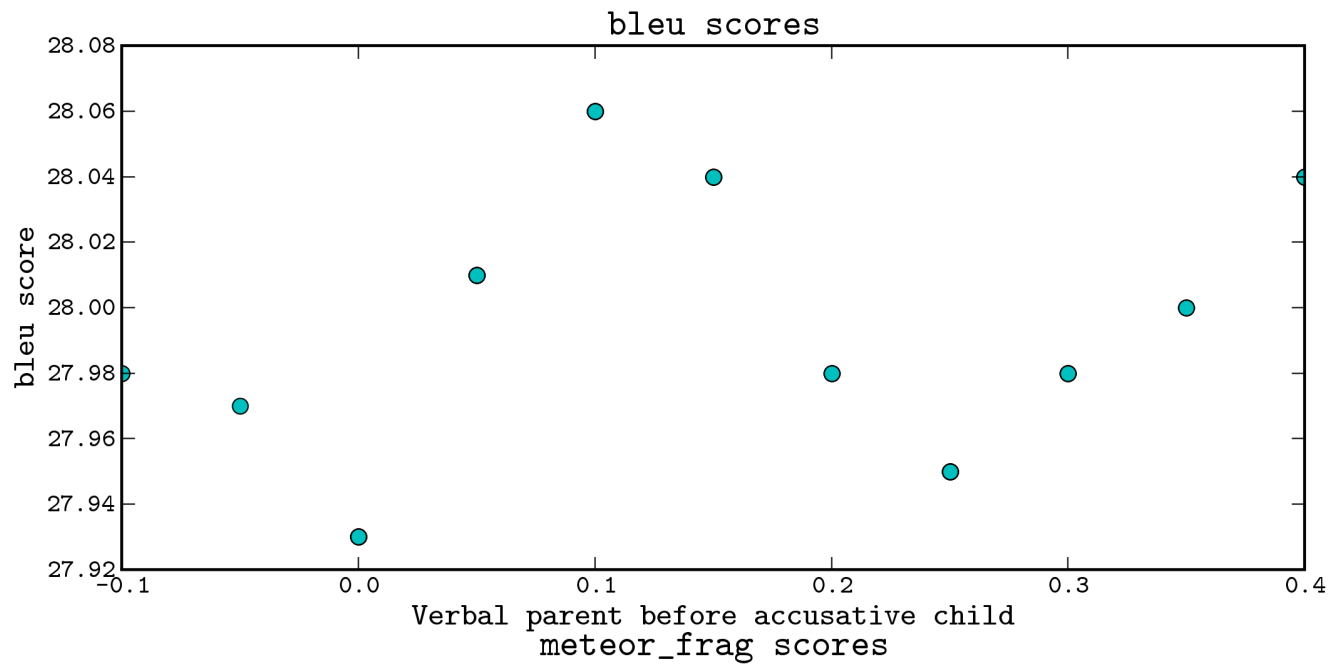


# VERB-BEFORE-OBJECT FEATURE

- Feature = 1 if a verb is translated before its object.
- Feature = 0 otherwise.



'Verbal parent before accusative child' against bleu, meteor\_frag



# CONTRIBUTIONS

- Quick and effective reordering strategy for preprocessing
- Better translation quality and faster than baseline Moses

