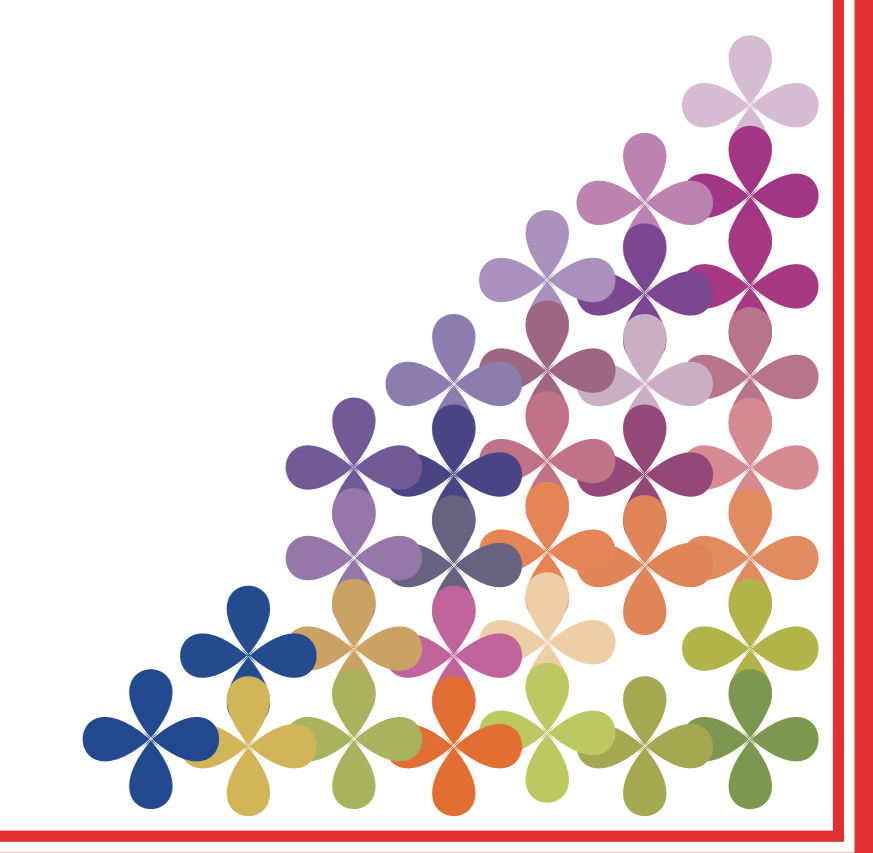




この翻訳，あなたなら何点あげますか？



～本当にBLEUでいいんですか？～

How Many Points Do You Give For This Translation?

- Why Do You Use BLEU? -

概要： 質の高い翻訳システムを実現するためには，自動評価が欠かせません．そこで，システム翻訳と正解訳との間で共通して出現する単語に着目し，それらの出現順を順位相関係数に基づき採点することで翻訳システムを評価する手法である RIBES (Rank-based Intuitive Bilingual Evaluation Score) を考案しました．

RIBES: Rank-based Intuitive Bilingual Evaluation Score

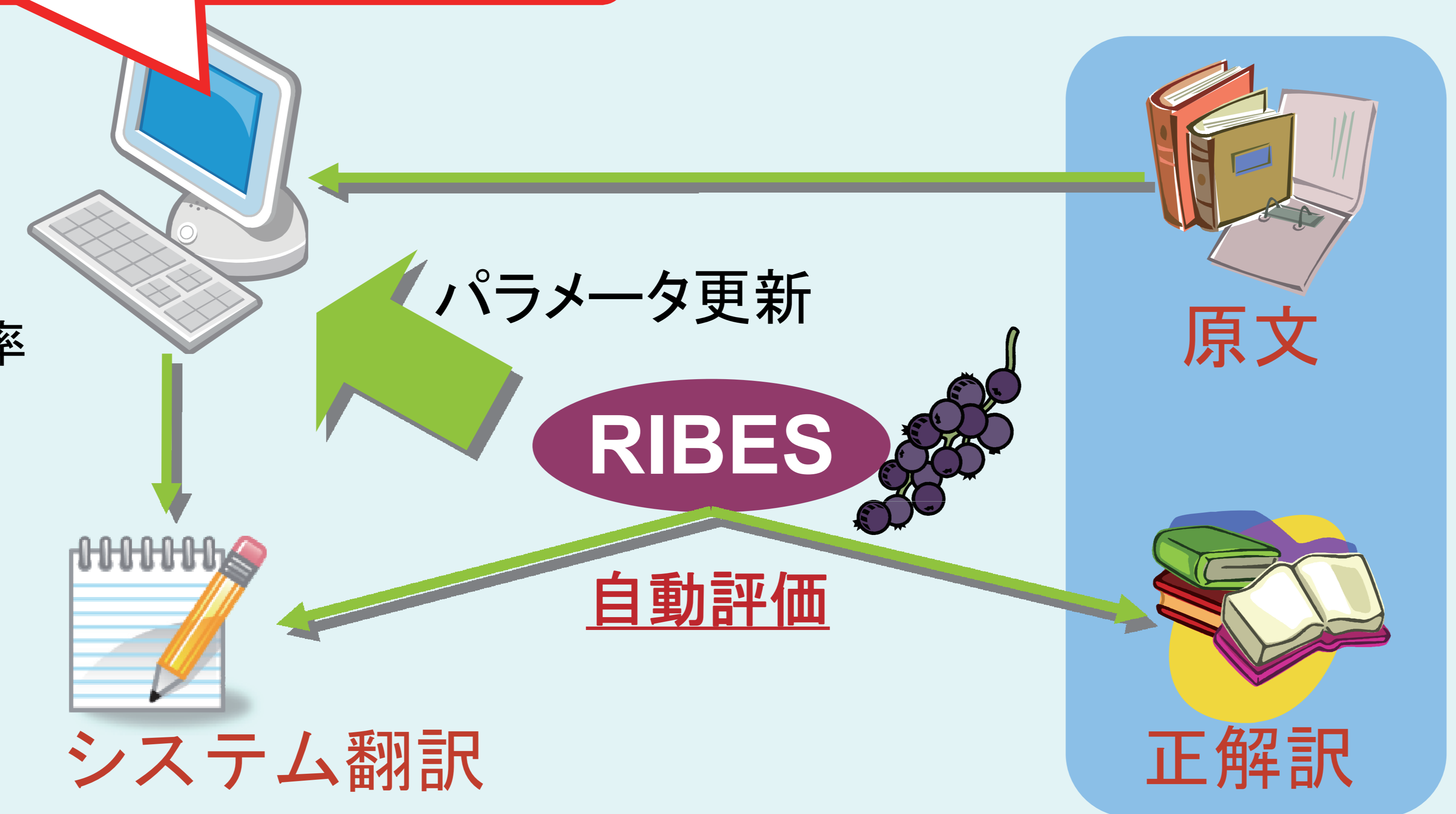
$$RIBES = NRC \times P^\alpha$$

システム翻訳と正解訳の
語順の近さ

単語単位での正解率

入力: John saw a girl yesterday.
出力: ジョンは昨日少女を見た.

開発用対訳データ



技術のポイント：
Nグラムのような局所的な単語の並びではなく，文全体での大域的な単語の並びに着目

翻訳システムのチューニング



	実例	人手評価		自動評価	
		忠実性	可読性	BLEU	RIBES
原文	彼は雨に濡れたので 風邪をひいた				
正解訳	he caught a cold because he got soaked in the rain				
機械訳 1	he caught a cold because he had gotten wet in the rain	○	○	0.53	0.93
機械訳 2	he got soaked in the rain because he caught a cold (彼は風邪をひいたので 雨に濡れた ...?)	×	○	0.74	0.38

連絡先 (Contact information)
平尾 努 (Tsutomu HIRAO)
 言語知能研究グループ
 (Linguistic Intelligence Research Group)