

日本語述語項構造解析器 YuCha「夕茶」

奈良先端科学技術大学院大学
林部 祐太

公式サイト
<http://hayashibe.jp/yucha/>

述語項構造解析とは？

文書中の各述語に対して、「項」となる名詞句等を当てること
=>「語間の意味的な関係」を捉える!

私は今日、トマトカレーを(1)食べました。
明日は太郎とインドカレーを(2)食べに
ショッピングセンターに(3)行く予定です。

	ガ	ヲ	ニ
(1)食べる	私	トマトカレー	
(2)食べ	私	インドカレー	
(3)行く	私		ショッピングセンター

※(2)(3)については「ト格=太郎」

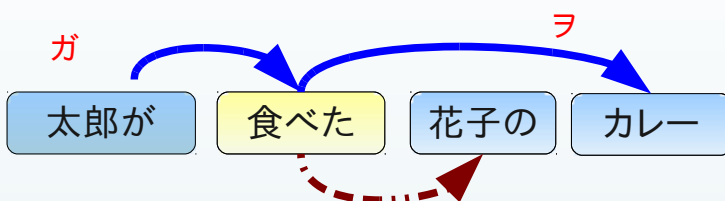
他解析器との比較

日本語の述語項構造を、別途言語資源を入手せずとも、だれでも簡単に解析できるようにすることを目標としています。

	KNP	Syncha	Yucha(夕茶)
開発元	京大黒橋研	NAIST松本研(飯田,小町)	NAIST松本研(林部)
Version	3.01 (2009.10)	1.4 (2007/4/20)	0.13 (現在開発中)
License	修正BSD	LGPL	GPL
実装	C	perl	C++
API	×	×	対応予定
並列解析	○	×	未定
前文照応解析	×	○	○
事態性名詞	×	○	対応予定
学習データ	京大コーパス	Naist Text Corpus	Naist Text Corpus等
必要データ	なし	日本語語彙体系	なし

どのように役に立つの？

★構文解析の精度向上



↑ Cabochaでは誤って解析

意味的な関係を捉えることで、正しく係り受け解析ができる

★表面的な記述に囚われない自動翻訳

「カレーを食べたいなあ。そういえば、
あのお店が美味しかったなあ。」
≠「That shop was delicious.」

(2010年9月現在のエキサイト翻訳の結果。
正しくは「カレー」が美味しい)

★情報検索・要約・言い換えなどの 精度向上

「意味」まで踏み込んで
処理できる!

現在の開発状況

NaistTextCorpusをTest:Train=1:9で使用して実験

	ガ格	ヲ格	ニ格
Acc	52.24%	78.30%	83.97%
F1	68.63%	68.70%	46.02%

今後の課題

- 項探索モデルの改良
 - 現在はトーナメントモデル[lida03]
 - 各格の項探索を協調してできるようにする
- 文外項の同定の精度向上
- ガヲニ格以外の探索

今後の予定

- 2010年度中に正式リリースを目指します