

高精度日本語自然言語解析研究紹介

後藤 智範 研究室

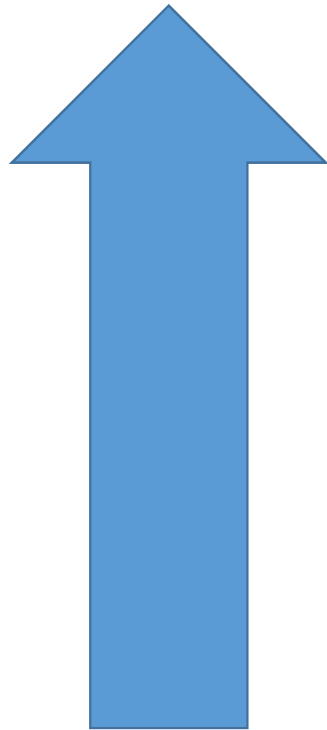
情報科学科 4年

小原 拓巳

目次

1. 日本語文解析の概要
2. 形態素解析
3. 係り受け解析
4. 表層格解析
5. 深層格解析
6. 他大学開発解析ツール
7. 当研究室での研究対象
8. 難しい日本語文の解析

1 日本語文解析の概要



談話解析

- 文脈解析
- 深層格解析

構文解析

- 表層格解析
- 係り受け解析
- 項解析

チャンキング

- 文節チャンキング
- 固有表現抽出

形態素解析

- 単語分割
- 品詞同定
- 活用形同定

1 日本語文解析の概要

形態素解析

文を単語ごとに分割し、各単語ごとに動詞や名詞といった品詞を割り当てる

用言の場合は活用形も同定する

動詞、形容詞、
形容動詞など

終止形、連体形、
假定形...

各単語 = 形態素
(文を構成する最小要素)

1 日本語文解析の概要

チャンキング

形態素解析で単語ごとに分割された文に対し、名詞や用言どうしを
接続(連結)させて、形態素を文の構成要素ごとにまとめる

1 日本語文解析の概要

構文解析

形態素解析で得られた単語の並びに対し、単語どうしの構文関係进行分析し、その文がどのような構造をしているのかを解析する

意味解析

単語の持つ意味や、単語どうしの意味的な関係についての知識を使って、その文がどのような意味を持つかという解釈の候補を生成し、その中のどれが最も意味的に正しいかを絞り込む

2 形態素解析

原文:勉強するために本を買った。

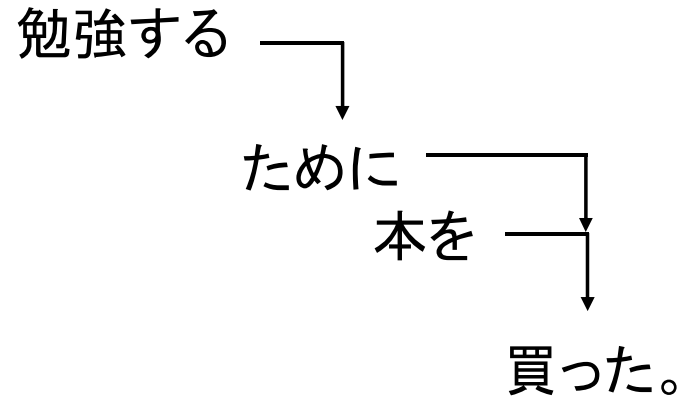
形態素解析の結果



| | | | | | | | |
|------|------|----|----|----|-----|-----|----|
| 勉強する | ため | に | 本 | を | 買った | 。 | |
| サ変名詞 | 形式名詞 | 助詞 | 名詞 | 助詞 | 動詞 | 助動詞 | 句点 |

文が単語ごとに分割され、各単語に品詞が付与される

3 係り受け解析



| | 「係り」属性値 | 「受け」属性値 |
|------|---------|-------------|
| 勉強する | 用言:連体形 | 文頭、用言:サ変名詞 |
| ために | 助詞「に」 | 体言 |
| 本を | 助詞「を」 | 体言 |
| 買った。 | | 文末、句点、用言:動詞 |

4 表層格解析

- 格

格 = 文法的・意味的役割

格の種類で単語間の役割がわかる

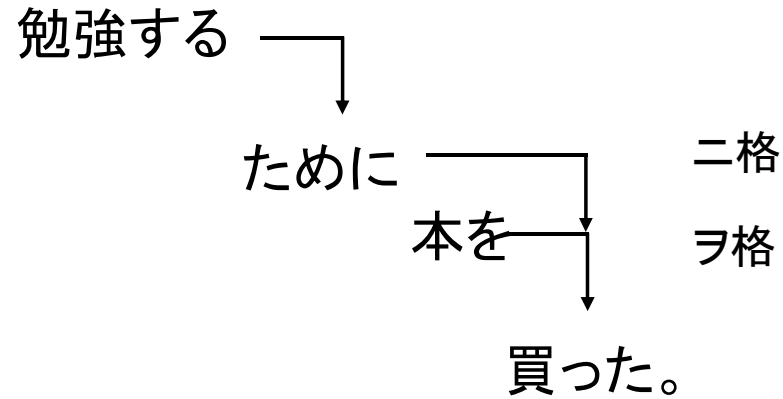
日本語の場合、格は格助詞によって表されるのが典型的

- 表層格

語の位置関係や格助詞などによって表層的に決まる格
ガ格(～が ～は)、ヲ格(～を ～の)、ニ格(～に) など

➡ 語を見れば格の種類が分かる

4 表層格解析



表層格を持つ語

ために
本を

表層格の種類

二格
ヲ格

5 深層格解析

・深層格

意味的に決まる格
語を見ただけでは分からない

<例>

| | |
|-------|----------------|
| 動作主格: | 動作を引き起こすもの |
| 対象格: | 動作が作用する対象となるもの |
| 時間格: | 動作が起こる時間を表すもの |


5 深層格解析

原文：勉強するために本を買った。



構文解析後

勉強するために
本を



買った。

深層格

目的格
対象格

6 他大学開発解析ツール

■形態素解析

▪ JUMAN

開発母体: 京都大学大学院 情報学研究科 知能情報学専攻 黒橋・河原研究室

リリース開始年 1994年、最終更新年 2014年

辞書用語数: 229209

解説ページ URL: <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

デモページ URL: <http://lotus.kuee.kyoto-u.ac.jp/nl-resource/cgi-bin/juman.cgi>

▪ Mecab

開発母体: 奈良先端科学技術大学院大学 情報科学研究科 自然言語処理学 松本研究室

リリース開始年 2006年、最終更新年 2013年

辞書用語数: 392126

解説ページ URL: <http://taku910.github.io/mecab/>

6 他大学開発解析ツール

■ チャンキング・係り受け解析

- ・ **KNP**(チャンキング機能無し)

開発母体: 京都大学大学院 情報学研究科 知能情報学専攻 黒橋・河原研究室

リリース開始年 1994年、最終更新年 2015年

解説ページ URL: <http://nlp.ist.i.kyoto-u.ac.jp/?KNP>

デモページ URL: <http://reed.kuee.kyoto-u.ac.jp/nl-resource/cgi-bin/knp.cgi>

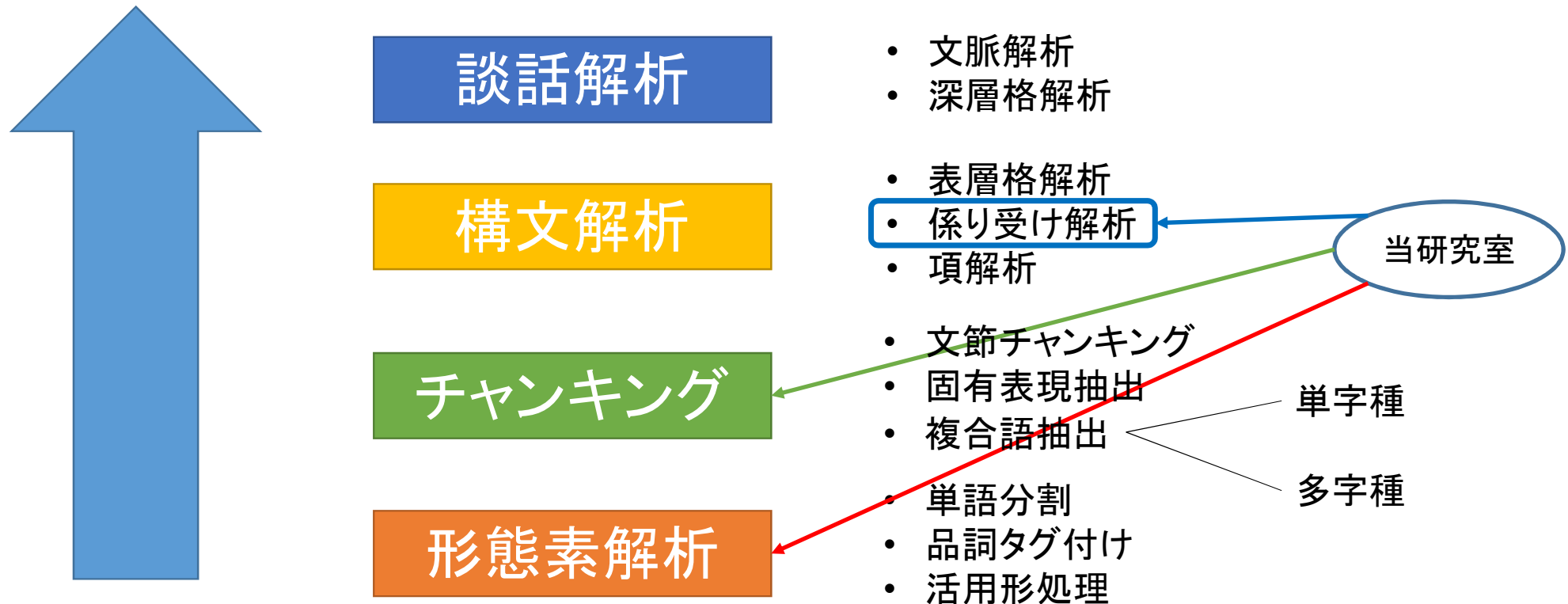
- ・ **Cabocha**(チャンキング機能有り)

開発母体: 奈良先端科学技術大学院大学 情報科学研究科 自然言語処理学 松本研究室

リリース開始年 2001年、最終更新年 2014年

解説ページ URL: <https://taku910.github.io/cabocha/>

7 当研究室での研究対象



8 難しい日本語文の解析

实例

- ・ソニーモバイルコミュニケーションズ株式会社は、世界最薄/最軽量を謳う10.1型Androidタブレット「Xperia Z2 Tablet」のWi-Fi版「SGP512JP」を5月31日に発売する。
- ・その結果、高分子濃厚系の微分貯蔵弾性率は $G'(\omega, \gamma; t) = G'(\omega, 0) \phi(\gamma; t)$ の形に書けることを見いだした。
- ・カチオン伝導性ポリマー(P(MAAM-co-MEO))にクラウンエーテル構造を導入した系について、発現されるイオン伝導度とキャリアカチオンのイオン半径との相関について検討を行った。



学術論文や特許抄録などに出現する文

語の特徴: 多字種複合語、長漢字熟語が頻出
文の特徴: 重文&複文が頻出、並列(名詞句、述語、節)構造が多々出現

8 難しい日本語文の解析

原文：ソニーモバイルコミュニケーションズ株式会社は、世界最薄/最軽量を謳う10.1型Androidタブレット「Xperia Z2 Tablet」のWi-Fi版「SGP512JP」を5月31日に発売する。

形態素解析

ソニーモバイルコミュニケーションズ株式会社は、世界最薄 / 最軽量を謳う
10.1型Androidタブレット「Xperia Z2 Tablet」のWi-Fi版「SGP512JP」を5月31日に発売する。

形態素解析結果（各単語の下のボックスは形態素の種類を示す）：
ソニーモバイルコミュニケーションズ株式会社は、世界最薄 / 最軽量を謳う
10.1型Androidタブレット「Xperia Z2 Tablet」のWi-Fi版「SGP512JP」を5月31日に発売する。

チャンキング(名詞)

ソニーモバイルコミュニケーションズ株式会社は、世界最薄/最軽量を謳う
10.1型Androidタブレット「Xperia Z2 Tablet」のWi-Fi版「SGP512JP」を5月31日に発売する。

8 難しい日本語文の解析

係り受け解析

ソニーモバイルコミュニケーションズ株式会社 は、
 世界最薄/最軽量 を
 謳う
 10.1型Androidタブレット
 「Xperia Z2 Tablet」 の
 Wi-Fi版
 「SGP512JP」 を
 5月31日 に

発売する。

「係り」属性値

助詞「は」
 助詞「を」
 助詞「の」
 助詞「を」
 助詞「に」

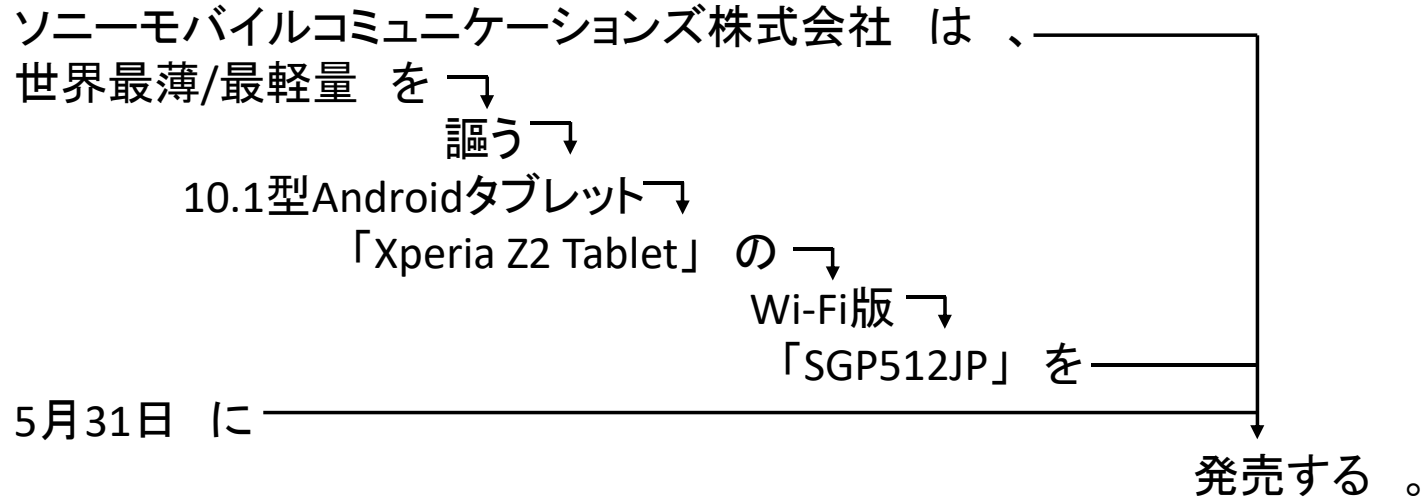
「受け」属性値

文頭、体言、読点
 体言
 体言
 体言
 体言
 文末、句点、用言：サ変名詞

ソニーモバイルコミュニケーションズ株式会社は、
 世界最薄/最軽量を
 「Xperia Z2 Tablet」の
 「SGP512JP」を
 5月31日に
 発売する。

8 難しい日本語文の解析

表層格解析



ソニーモバイルコミュニケーションズ株式会社は、
世界最薄/最軽量を
「Xperia Z2 Tablet」の
「SGP512JP」を
5月31日に

表層格

ガ格
ヲ格
ヲ格
ヲ格
ニ格

8 難しい日本語文の解析

深層格解析

原文： ソニーモバイルコミュニケーションズ株式会社は、世界最薄/最軽量を謳う10.1型Androidタブレット「Xperia Z2 Tablet」のWi-Fi版「SGP512JP」を5月31日に発売する。

