

高精度日本語自然言語解析研究紹介

後藤 智範 研究室

理学専攻 博士前期課程1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

日本語自然言語解析



文脈解析

- 談話構造解析
- 照応解析
- 述語項解析

意味解析

- ゼロ代名詞推定
- 深層格解析

構文解析

- 表層格解析
- 係り受け解析

チャンキング

- 固有表現同定
- 複合語(単字種、多字種)同定

形態素解析

- 活用形同定
- 品詞同定
- 単語分割

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を分析し、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

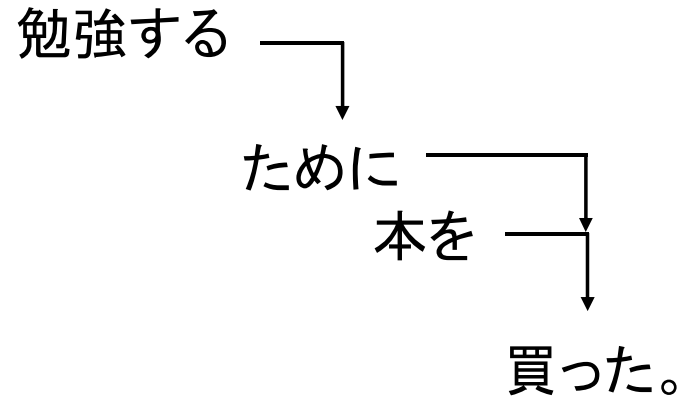
動詞

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

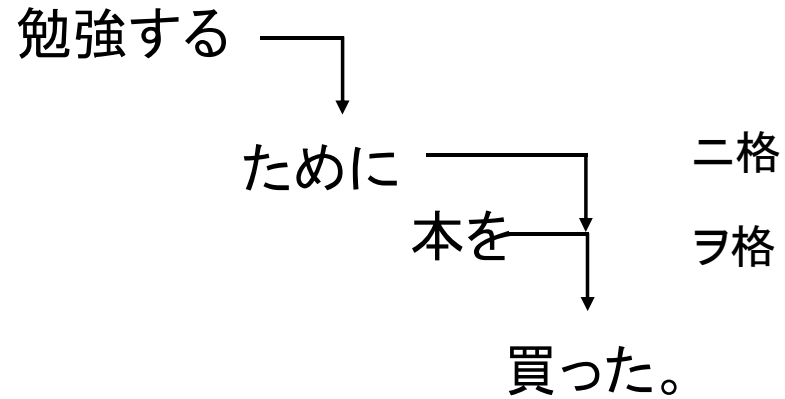
日本語の場合、格は格助詞によって表されるのが典型的

- 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>

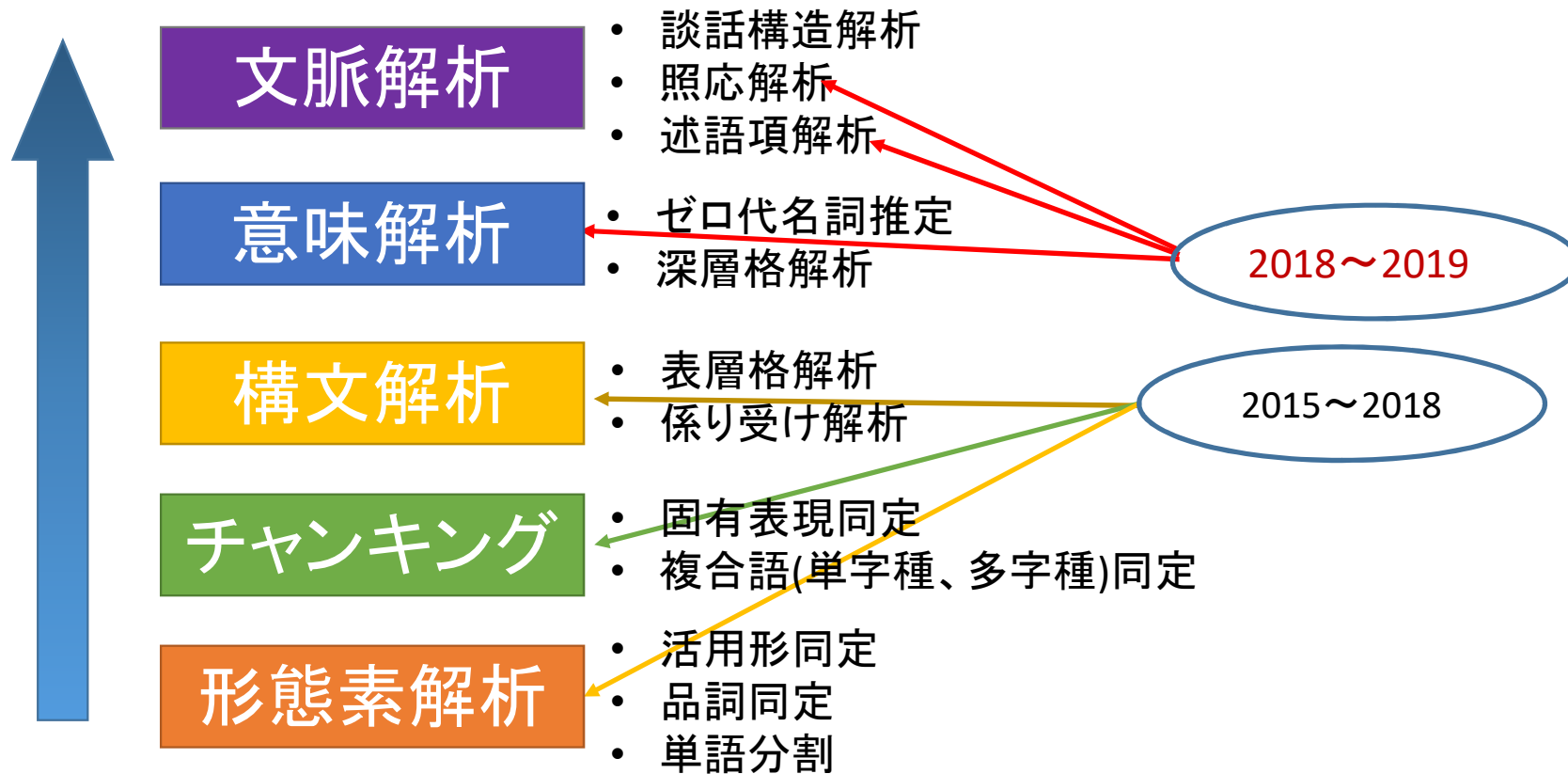
- ・ **Cabocha**(チャンキング機能有り)

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/>

本研究室での研究内容



8 難しい日本語文の解析

実例

- ・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

もっと難しい日本語文: **ゼロ代名詞**

前の文で出現した主語がその後の文では省略されている → 主語の無い文

(3つ前の文)

三相誘導電動機の等価回路をT形で表わし、無負荷試験,拘束試験などからJEC-37の方法で求めた定数を使って特性計算すると、実測値とずれを生ずる。

(2つ前の文)

この誤差の大きな原因として、二次導体の表皮効果による二次抵抗 R_2 と二次漏れリアクタンス X_2 の変化が考えられる。

(1つ前の文)

そこで、他の定数 R_1, X_1, R_M, X_M などはJEC-37で求め、 R_2 と X_2 の2定数のみを調整して、定格出力時(任意の出力も可)の電流、入力が計算値と実験値で一致するような定数の求め方を示す。

(対象の文)

出力が1.5~900KWまで、2次のスロット形状が全閉,半閉,二重全閉,二重半閉などの6機種にこの方法を適用し、実験結果と比較検討する。

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（各単語の品詞）：
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。

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

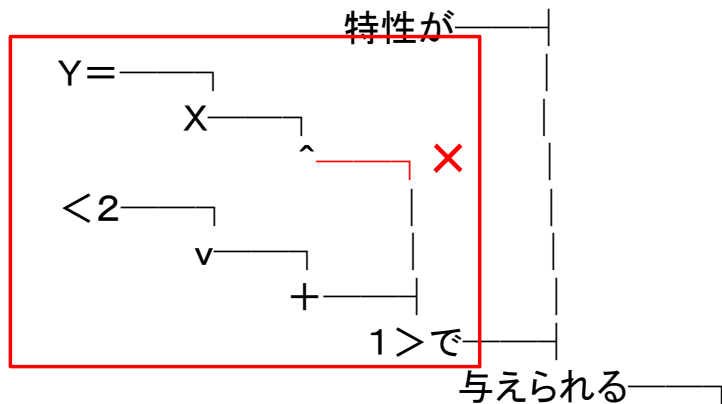
文頭、体言、読点
 体言
 体言
 体言
 体言
 文末、句点、用言：サ変名詞

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。

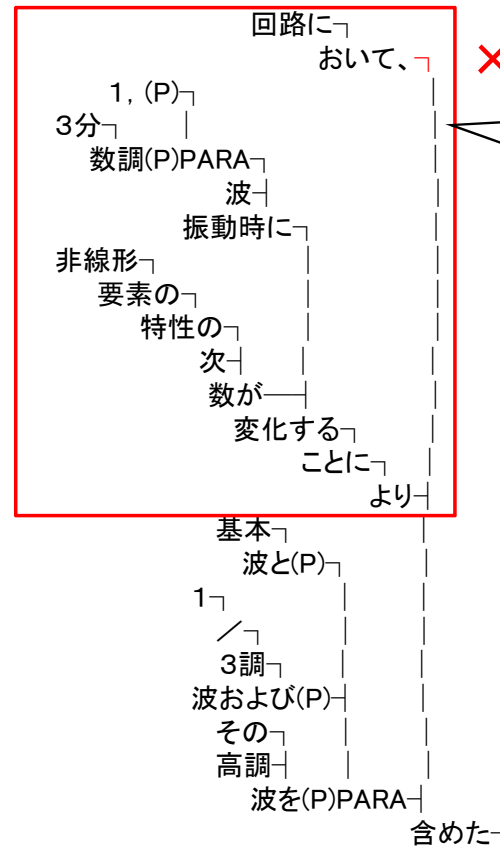
8 難しい日本語文の解析 — KNPとの比較

例文「本稿は離散フーリエ変換を用いた調和項比較法により非線形要素の特性が $Y=X^{<2v+1>}$ で与えられる非線形回路において、1,3分数調波振動時に非線形要素の特性の次数が変化することにより基本波と1/3調波およびその高調波を含めた分数調波のそれぞれの振幅と位相に及ぼす影響について考察する。」

KNP(京都大学 黒橋・河原研究室)



「X^」の正しい係り先は「<2」であるため、誤解析



他の従属節の用言を係り先としている

「回路において」の正しい係り先は「変化する」であるため、誤解析

8 難しい日本語文の解析 — KNPとの比較

例文「本稿は離散フーリエ変換を用いた調和項比較法により非線形要素の特性が $Y=X^{2v+1}$ で与えられる非線形回路において、1,3分数調波振動時に非線形要素の特性の次数が変化することにより基本波と1/3調波およびその高調波を含めた分数調波のそれぞれの振幅と位相に及ぼす影響について考察する。」

本研究室開発

本稿は

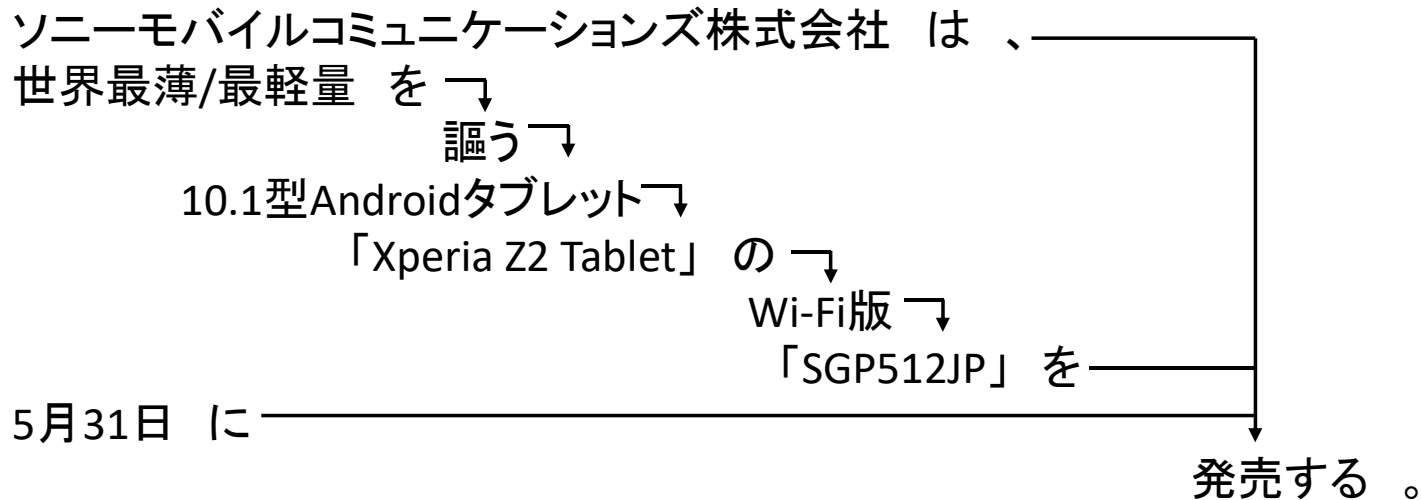
離散フーリエ変換を用いた調和項比較法により非線形要素の特性が $Y=X^{2v+1}$ で与えられる非線形回路において、1,3分数調波振動時に非線形要素の特性の次数が変化することにより

「回路において」が正しい係り先に係っている

基本波と1/3調波およびその高調波を含めた分数調波のそれぞれの振幅と位相に及ぼす影響について考察する。

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

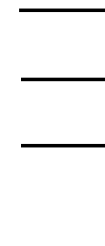
8 難しい日本語文の解析

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。

↓ 構文解析後

ソニーモバイルコミュニケーションズ株式会社は、
Wi-Fi版「SGP512JP」を
5月31日に



発売する。

深層格

動作主格

対象格

時間格