

nl-1. 形態素解析と構文解析

(自然言語処理入門)

URL: <https://www.kkaneko.jp/cc/nlp/index.html>

金子邦彦



自然言語処理

- **自然言語処理**では、コンピュータが、人間の言葉（英語や日本語）を処理する。
- **人間の言葉**を理解できる能力をもつアプリやサービスの制作に役立つ
 - （用途の例）
インターネット検索，対話システム，音声合成，
翻訳，かな漢字変換での予測変換，
迷惑メールフィルタ

自然言語処理の技術

- 文を単語に分割 (**単語の切り出し**)
- 単語の種類 (**品詞**) の判定
- 係り受けの関係などの分析 (**構文解析**)
- **意味解析**

など

Windowsなどで、単語を切り出し、品詞を自動判定する機能を持つ mecab コマンド



白い雲と青い空が美しい

!echo '白い雲と青い空が美しい' | mecab

白い	形容詞,*,イ形容詞アウオ段,基本形,白い,しろい,代表表記:白い/しろい 名詞派生:白/しろ
雲	名詞,普通名詞,*,*,雲,くも,代表表記:雲/くも 漢字読み:訓 カテゴリ:自然物
と	助詞,格助詞,*,*,と,と,連語
青い	形容詞,*,イ形容詞アウオ段,基本形,青い,あおい,代表表記:青い/あおい 名詞派生:青/あお
空	名詞,普通名詞,*,*,空,から,代表表記:空/から 漢字読み:訓 カテゴリ:抽象物
が	助詞,格助詞,*,*,が,が,連語
美しい	形容詞,*,イ形容詞イ段,基本形,美しい,うつくしい,代表表記:美しい/うつくしい 反義:形容詞:醜い/みにくい
EOS	

Google Colaboratory の実行画面

単語に切り分けられ、**単語ごとの読み仮名**、**品詞**などが、得られる

日本国民は、正当に選挙された国会における代表者を通じて行動し

!echo '日本国民は、正当に選挙された国会における代表者を通じて行動し' | mecab

日本	名詞,地名,*,*,日本,にほん,代表表記:日本/にほん 地名:国
国民	名詞,普通名詞,*,*,国民,こくみん,代表表記:国民/こくみん カテゴリ:人 ドメイン:政治
は	助詞,副助詞,*,*,は,は,連語
,	特殊,読点,*,*,、,、,*
正当に	形容詞,*,ナノ形容詞,ダ列基本連用形,正当だ,せいとうに,代表表記:正当だ/せいとうだ 反義:形容詞:不当だ/ふとうだ
選挙	名詞,サ変名詞,*,*,選挙,せんきょ,代表表記:選挙/せんきょ カテゴリ:抽象物 ドメイン:政治
さ	動詞,*,サ変動詞,未然形,する,さ,代表表記:する/する 付属動詞候補(基本) 自他動詞:自:成る/なる
れた	接尾辞,動詞性接尾辞,母音動詞,タ形,れる,れた,代表表記:れる/れる
国会	名詞,普通名詞,*,*,国会,こっかい,代表表記:国会/こっかい カテゴリ:組織・団体 ドメイン:政治
に	助詞,格助詞,*,*,に,に,連語
おける	動詞,*,母音動詞,基本形,おける,おける,代表表記:置ける/おける 可能動詞:置く/おく
代表	名詞,サ変名詞,*,*,代表,だいひょう,代表表記:代表/だいひょう 人名末尾 カテゴリ:人;抽象物
者	接尾辞,名詞性名詞接尾辞,*,*,者,しゃ,代表表記:者/しゃ 内容語 カテゴリ:人
を	助詞,格助詞,*,*,を,を,連語
通じて	動詞,*,サ変動詞,タ系連用テ形,通ずる,つうじて,代表表記:通ずる/つうずる 同義:動詞:通じる/つうじる
行動	名詞,サ変名詞,*,*,行動,こうどう,代表表記:行動/こうどう カテゴリ:抽象物
し	動詞,*,サ変動詞,基本連用形,する,し,代表表記:する/する 付属動詞候補(基本) 自他動詞:自:成る/なる
EOS	

Google Colaboratory の実行画面

Google Colaboratory での実行手順



① Google Colaboratory でノートブックを作成

② 次のコードセルを実行

```
!apt install -y mecab
```

③ 次のコードセルを実行

```
!echo '日本国民は、正当に選挙された国会における代表者を通じて行動し' | mecab
```

```
!echo '日本国民は、正当に選挙された国会における代表者を通じて行動し' | mecab
日本 名詞,地名,*,*,*日本,にほん,代表表記:日本/にほん 地名:国
国民 名詞,普通名詞,*,*,*国民,こくみん,代表表記:国民/こくみん カテゴリ:人 ドメイン:政治
は 助詞,副助詞,*,*,*は,は,連語
、 特殊,読点,*,*,*、,*
正当に 形容詞,*,*ナノ形容詞,タ列基本連用形,正当だ,せいとうに,代表表記:正当だ/せいとうだ 反義:形容詞:不当だ/ふとうだ
選挙 名詞,サ変名詞,*,*,*選挙,せんきょ,代表表記:選挙/せんきょ カテゴリ:抽象物 ドメイン:政治
された 動詞,*,*サ変動詞,未然形,する,さ,代表表記:する/する 付属動詞候補(基本) 自他動詞:自:成る/なる
国会 接尾辞,動詞性接尾辞,母音動詞,タ形,れる,れた,代表表記:れる/れる
国会 名詞,普通名詞,*,*,*国会,こっかい,代表表記:国会/こっかい カテゴリ:組織・団体 ドメイン:政治
に 助詞,格助詞,*,*,*に,に,連語
における 動詞,*,*母音動詞,基本形,おける,おける,代表表記:置ける/おける 可能動詞:置く/おく
代表者 名詞,サ変名詞,*,*,*代表,だいひょう,代表表記:代表/だいひょう 人名末尾 カテゴリ:人;抽象物
を 接尾辞,名詞性名詞接尾辞,*,*,*を,を,代表表記:者/しゃ 内容語 カテゴリ:人
を通じて 助詞,格助詞,*,*,*を,を,連語
し 動詞,*,*サ変動詞,タ系連用テ形,通ずる,つうじて,代表表記:通ずる/つうじて 同義:動詞:通じる/つうじて
し 名詞,サ変名詞,*,*,*行動,こうどう,代表表記:行動/こうどう カテゴリ:抽象物
し 動詞,*,*サ変動詞,基本連用形,する,し,代表表記:する/する 付属動詞候補(基本) 自他動詞:自:成る/なる
EOS
```

Python プログラム内で MeCab を使う

自動化のため → プログラム

```
import sys
import MeCab
m = MeCab.Tagger ("-Ochasen")
print(m.parse ("海は広い"))
```

```
>>> import sys
>>> import MeCab
>>> m = MeCab.Tagger ("-Ochasen")
>>> print(m.parse ("海は広い"))
海      ウミ      海      名詞 -一般
は      ハ      は      助詞 -係助詞
広い    ヒロイ    広い    形容詞 -自立      形容詞・アウオ段      基本形
EOS
```

- **構文解析**は、**係り受け**の解析を行うこと。

主語，述語，修飾語などの，単語間の関係の分析

- **構文解析** KNP のデモサイト

<http://lotus.kuee.kyoto-u.ac.jp/nl-resource/cgi-bin/knp.cgi>



The screenshot shows the KNP demo website interface. The browser address bar displays the URL: `lotus.kuee.kyoto-u.ac.jp/nl-resource/cgi-bin/knp.cgi`. The page title is "KNPを試してみる".

The first example shows the input sentence "空は青い" (The sky is blue). The parsing result is as follows:

```
# S-ID:1 KNP:4.2-407aed4a DATE:2019/06/07 SCORE:-5.37833
空は— <体言>
青い<用言:形><格解析結果:ガ/空;ニ/->
EOS
```

The second example shows the input sentence "白い雲と青い空が美しい" (The white clouds and blue sky are beautiful). The parsing result is as follows:

```
# S-ID:1 KNP:4.2-407aed4a DATE:2019/06/07 SCORE:-25.24505
白い— <用言:形><格解析結果:ガ/雲>
  雲と(P)— <体言>
青い— <用言:形><格解析結果:ガ/空;ニ/->
  空が(P)—PARA— <体言>
    美しい<用言:形><格解析結果:ガ/雲;ガ/空;ニ/-;デ/-;カラ/-;時間/-;ガ2/->
EOS
```


意味解析



- **意味解析**は、単語の意味の判定を行う
- それぞれの言葉の意味を、コンピュータが理解しながら、他の言葉の意味を判断するなど

駅前のコンビニは、**油を売っていた**

あの人は、仕事中に、**油を売っていた**

まとめ



- **単語** 単一の意味のまとめ
- **品詞** 単語の種別
- **係り受け** 主語, 述語, 修飾語などの, 単語間の関係
- **形態素解析** 自然言語の文から**単語**を切り出し,
品詞を推定すること
- **構文解析** **係り受け**の関係などを分析すること