

第1回学生向け！ナレッジグラフ推論チャレンジ2021 応募シート

※本応募シートの内容を記載した PDF ファイルを電子メールにて提出して下さい。
(シート作成に使用するソフトウェアは問いません)

送付先：kgrc@knowledge-graph.jp

1. 応募者に関する情報

- 氏名またはチーム名：新井律希
- 所属：大阪電気通信大学情報工学部情報工学科
- メールアドレス（代表）：gp18a012@oecu.jp

2. 推論・推理過程の説明

-私は係り受け解析を用いた推理小説ナレッジグラフの構築支援ツールの開発を行った。このツールはナレッジグラフの構築手順の中でも SVO 分解の部分を自動化するもので、短文化された文章から主語・述語・目的語の三つ組みを取り出すことを目的としている。本ツールの手法として、形態素解析と係り受け解析という自然言語処理を利用している。形態素解析とは図1のように文章を言葉が意味を持つ最小単位の単語に分割し、品詞を取得する処理である。係り受け解析とは図1のように形態素解析で分割した単語同士の依存関係（以下構文従属関係とする）を構造化する処理である。また、係り受け解析で得た主要な構文従属関係を表1に示す。

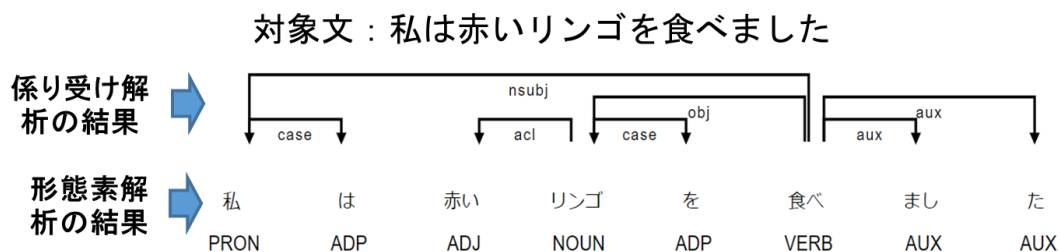


図1 係り受け解析と形態素解析の具体例

表 1 主要な構文従属関係

分類	構文従属関係	意味
主語	nsubj	主語名詞
	csbj	主部
	dislocated	転置
目的語	obj	目的語
	obl	斜格名詞
述語	無し（構文従属関係の始点）	
助詞	case	各表示

また、図 2 のように係り受け解析で得られた構文従属関係の nsubj や obj をそのまま主語や目的語にしてしまうと主語や目的語に掛かる修飾語が抜けてしまう。

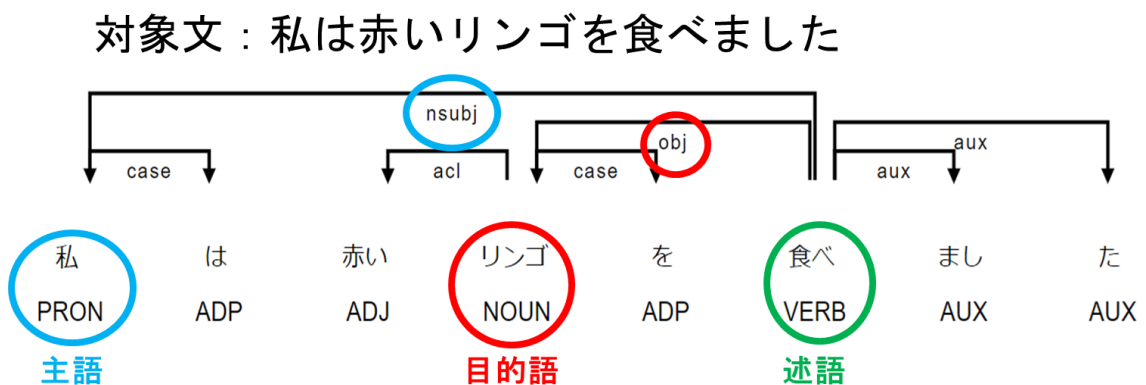


図 2 主語・述語・目的語となる単語のみを取得した例

そこで、図 3 のように主語・述語・目的語となる単語の修飾語を取得するために各単語の係り受け先を辿り、それぞれ係り受け関係の中で先頭の単語と末尾の単語の情報を取得する。

対象文：私は赤いリンゴを食べました

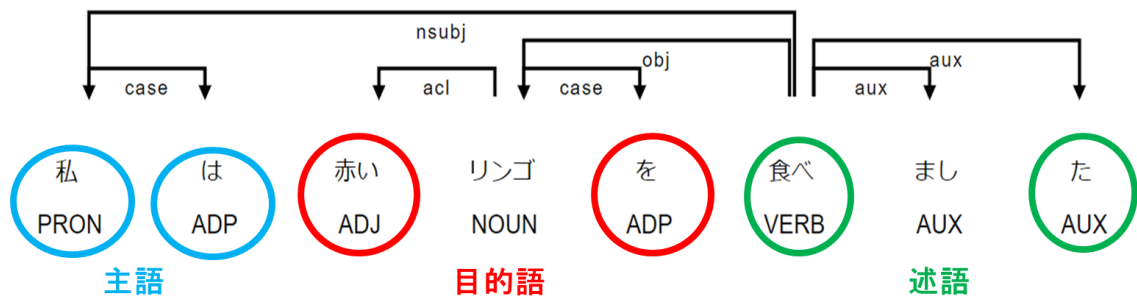


図3 主語・述語・目的語の係り受け先の中で先頭の単語と末尾の単語を取得した例

図3で取得した情報を利用して図4のように先頭の単語から末尾の単語から順番に単語の結合を行う。この時、末尾の単語の構文従属関係が case (助詞) の場合は末尾の単語は結合を行わない。

対象文：私は赤いリンゴを食べました

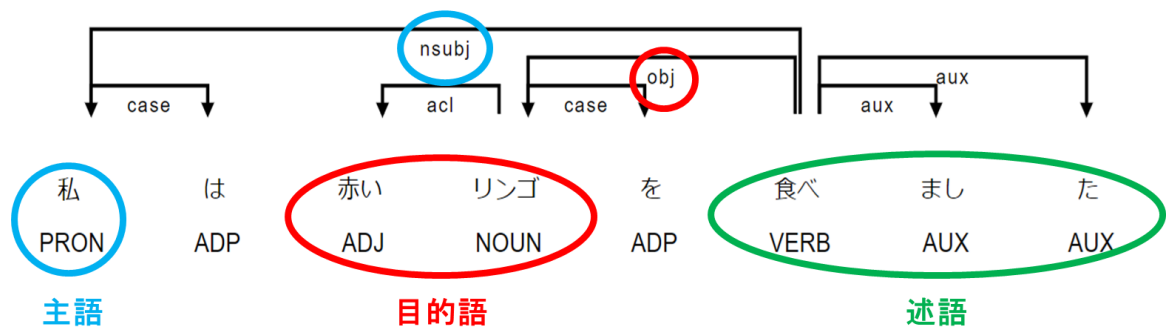


図4 先頭から末尾まで単語を結合した例 (完成形)

基本的にはこのような手法で主語・述語・目的語を取得している。この手法を用いて推理小説6作品に対してSVO分解を行った。表2は各小説の分類別の正解率を示したものである。

またSVO分解を行った結果の詳細は、

https://github.com/rikk-2/-KG_construction_support-/tree/main/係り受け解析を用いた推理小説KG構築支援ツール/experiment

に下記のファイル名で公開している。

- ・<小説名>.tsv が各小説のナレッジグラフの元データ (TSV形式)
- ・<小説名>SVO結果.tsv が本ツールを用いてSVO分解した結果
- ・<小説名>SVO結果評価.tsv がSVO分解したと元データを比較した評価結果

表 2 6 小説の分類別正解率

小説	主語	述語	目的語	総正解率	場面正解率
花婿失踪事件	0.8051	0.5319	0.6646	0.6728	0.3121
僧坊荘園	0.6875	0.5286	0.3644	0.5399	0.2201
踊る人形	0.9000	0.7591	0.6364	0.7818	0.4298
入院患者	0.9571	0.7024	0.6489	0.8192	0.5864
白銀号事件	0.8538	0.6981	0.6119	0.7355	0.4232
悪魔の足	0.9104	0.7043	0.6189	0.7894	0.4785

表 2 の結果から各小説ごとにおける主語の正答率平均は 85 % (最小 68 %、最大 95 %)、各小説ごとにおける述語の正答率平均は 65 % (最小は 52 %、最大は 75 %)、各小説ごとにおける目的語の正答率は 59 % (最小は 36 %、最大は 66 %)ということが分かった。このことから更なる正答率向上のためには目的語と述語の改良が必要といえる。

-対象とした作品と使用した知識の範囲(ID)

花婿失踪事件、僧坊荘園、踊る人形、入院患者、白銀号事件、悪魔の足

-パフォーマンス情報 (計算機のスペック, 実行時間, 使用するメモリなど)

3. 実行プログラム

https://github.com/rikk-2/-KG_construction_support/

- 実行方法の説明

実行環境として下記を使用した。

- Jupyter Notebook 6.3.0(anaconda3) 無償
- Python 3.8.8 無償

また、自然言語処理を用いるために下記を使用した

- ja-ginza-electra 5.0.0 無償

4. 資料の共有について

応募いただいた「作品応募シート」および「プログラム等」について、可能であれば、ナレッジグラフ推論チャレンジのサイトで公開させていただきたいと考えております※。下記のフォームにご入力の上お知らせください。

※ナレッジグラフ推論チャレンジ関連イベントで応募資料の一部を使用させていただくこともあります。

・応募フォーム

-公開の可否：

- ☐ 公開してよい
- ☐ 非公開とする

-公開形式：

- ☐ ナレッジグラフ推論チャレンジのサイトで公開
- ☐ 独自のサイトで公開してリンクを希望

→公開先 URL (※)：

・応募したプログラム、データ等

-公開の可否：

- ☐ 公開してよい
- ☐ 非公開とする

-公開形式：

- ☐ ナレッジグラフ推論チャレンジのサイトで公開
- ☐ 独自のサイトで公開してリンクを希望

→公開先 URL (https://github.com/rikk-2/-KG_construction_support/)：

※公開先 URL については、後日、公開先が決定してからご連絡いただいても結構です。