

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

大阪電気通信大学

情報通信工学部情報工学科4年

新井律希

背景・目的

- AIにおける説明可能性を高めるコンテストとして、人工知能学会セマンティックウェブとオントロジー研究会が開催する「ナレッジグラフ推論チャレンジ」がある。
- 「ナレッジグラフ推論チャレンジ」では推理小説をグラフ構造にしたナレッジグラフ(以下KG)を利用して機械的に推論を行う。
- しかし、現状では推理小説のKG化作業はすべて手作業で行われおり、作成に時間がかかる。そこで、KG化作業の自動化手法を開発する。

KG構築の作業手順

KGの構築は主に下記の作業に分かれる。

1. KG化する部分の抽出
2. SVO分解 & 短文化

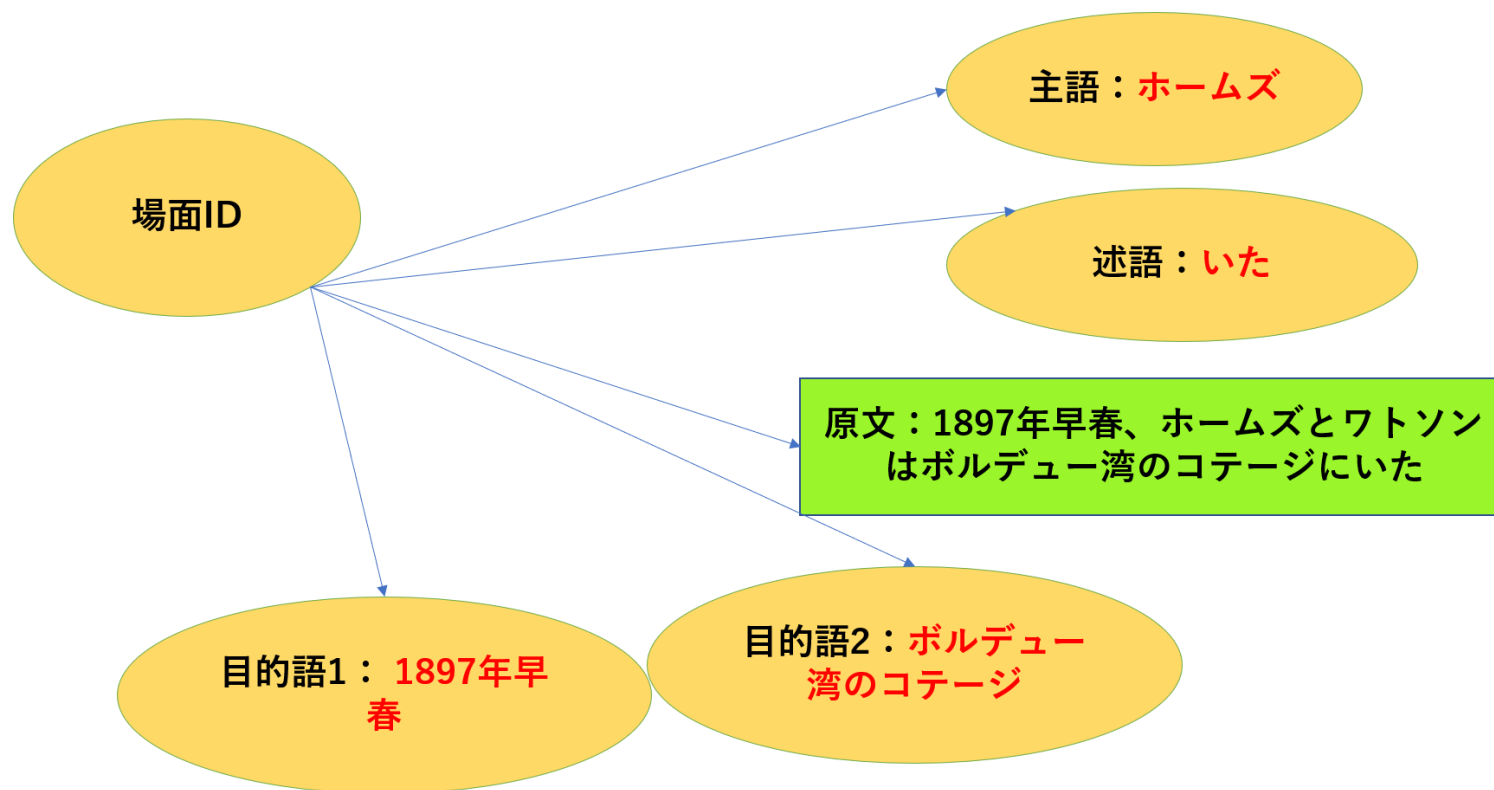
KG化する部分の抽出（例）

あれは、そう、1897年の春、ホームズの鋼の体にも翳りが見られるようになった。きわめて厄介な依頼が連続的にやってきたこともあるし、おそらくは時折やかす無茶の数々に拍車をかけられたのだろう。その年の3月、ハーレー街のドクター・ムーア・アガー——この人とホームズとの劇的な出会いについてはそのうちお話することがあるかもしれない——はこの有名な私立探偵に、全事件を凍結して完全な休息に身を委ねよ、さもなくば確実に体を壊すであろう、という強い勧告を与えた。ホームズは絶対的な知的独立性をもっていたから、健康状態など興味をよせるに値する問題ではなかったが、結局は、仕事を続ける資格を永久に失ってしまうだろうという脅しによって、新鮮な空気と風景のもとで静養することに同意した。そういうわけであの年の早春、我々はコーンワル半島の先端、ポルドュー湾に近い小さなコテージで過ごしていたのだった。

SVO分解 & 短文化（例）

ID	Property	原文
例XXX	対象文	1897年早春、ホームズとワトソンはポル デュー湾のコテージにいた
例XXX	Who	ホームズ
例XXX	hasProp erty	いた
例XXX	Where	ポルデュー湾のコテージ
例XXX	When	1897年早春

推理小説ナレッジグラフの例



研究内容

KGの構築は主に下記の作業に分かれる。

1. KG化する部分の抽出
2. SVO分解 & 短文化

本研究では、

- KG構築手順のうち「SVO分解」の部分の自動化を行う。

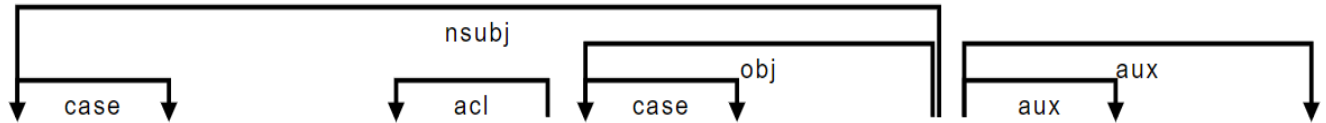
SVO分解の自動化プログラム

- 自然言語処理を行うためのツール「GINZA」を利用して、形態素解析及び係り受け解析を行う
- 二つの解析結果を利用して対象文から「主語」「述語」「目的語」を抜き出す。

係り受け解析と形態素解析の例

対象文：私は赤いリンゴを食べました

係り受け解析の結果



形態素解析の結果



私	は	赤い	リンゴ	を	食べ	まし	た
PRON	ADP	ADJ	NOUN	ADP	VERB	AUX	AUX

中核となる構文従属関係

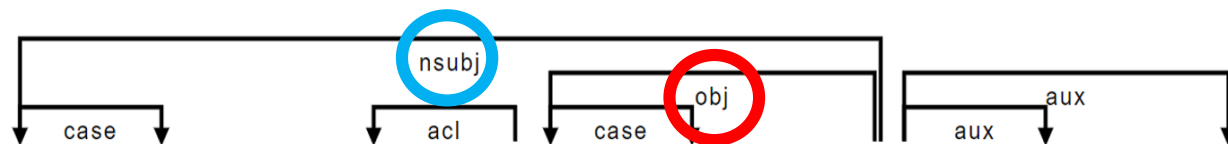
分類	構文従属関係	意味
主語	nsubj	主語名詞
	csbj	主部
	dislocated	転置
目的語	obj	目的語
	obl	斜格名詞
述語	無し(構文従属関係の始点)	
助詞	case	格表示

プログラム手法1

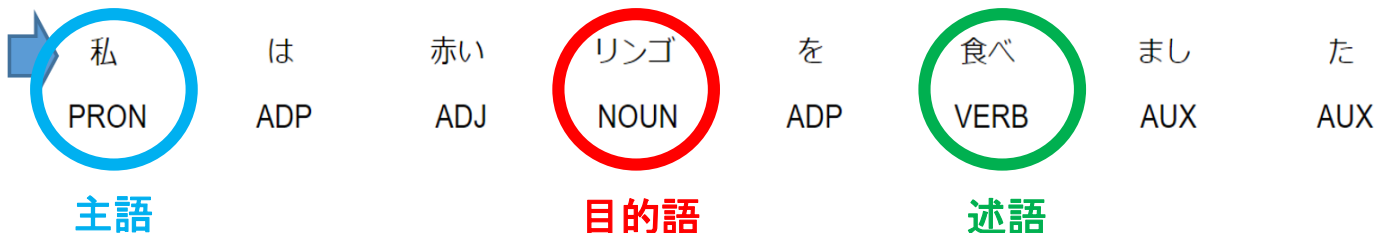
- 係り受け解析で得た構文従属関係を利用して、主語・述語・目的語となる単語を取得。

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

係り受け解析の結果



形態素解析の結果

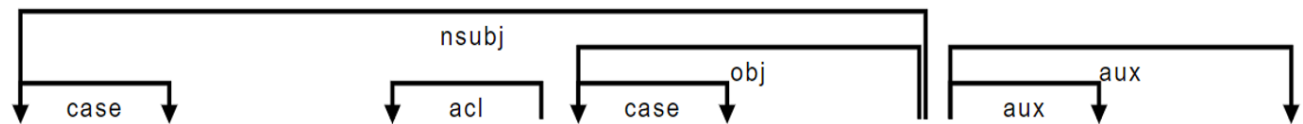


プログラム手法2

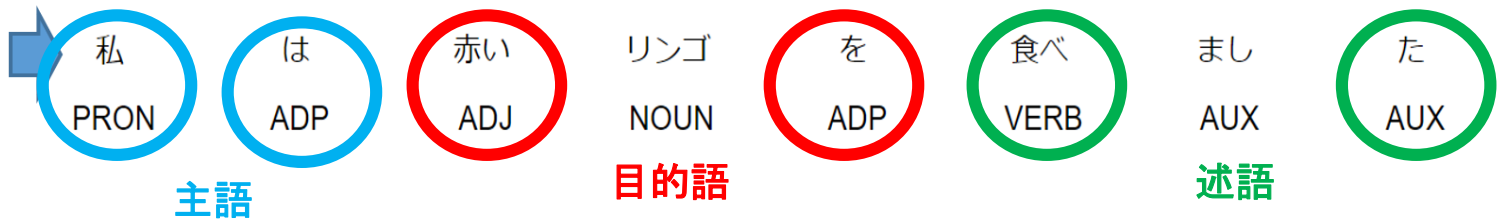
- 主語・述語・目的語となる単語の修飾語を取得するために各単語の係り受け先を辿り、それぞれ係り受け関係の中で先頭の単語と末尾の単語を取得。

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

係り受け解析の結果



形態素解析の結果

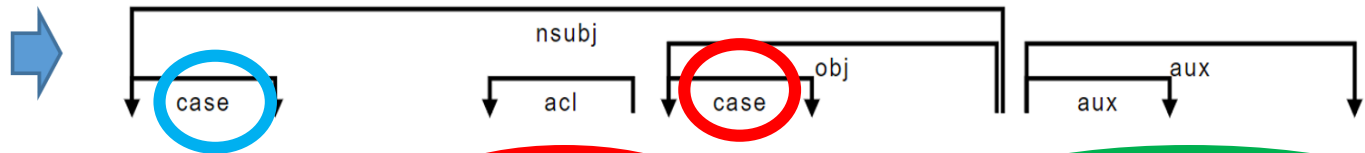


プログラム手法3

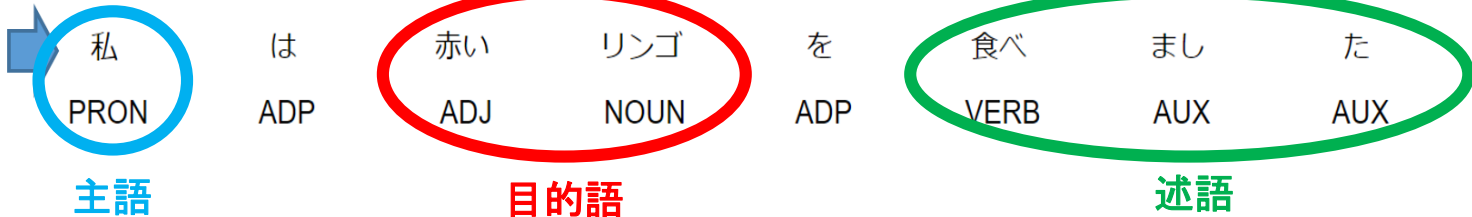
- 先頭の単語から末尾の単語まで順番に単語を結合していく。末尾の単語の構文従属関係がcase(助詞)の場合は末尾の単語は結合しない。

分類	構文従属関係	意味
主語	nsubj	主語名詞
	csubj	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

係り受け解析の結果



形態素解析の結果



SVO分解手法の評価

- ・ シャーロックホームズの推理小説6作品において評価を行う。
- ・ 正誤判定は正解データとの比較による完全一致で行う。
- ・ 評価する分類は主語・述語・目的語・総正解率・場面正解率の5項目で、正答率を表示する。

* 正解データはナレッジグラフ推論チャレンジで提供されているものを用いています。

* 場面正解率は場面(一つの文)の主語・述語・目的語が全て正解だった場合の正解率です。

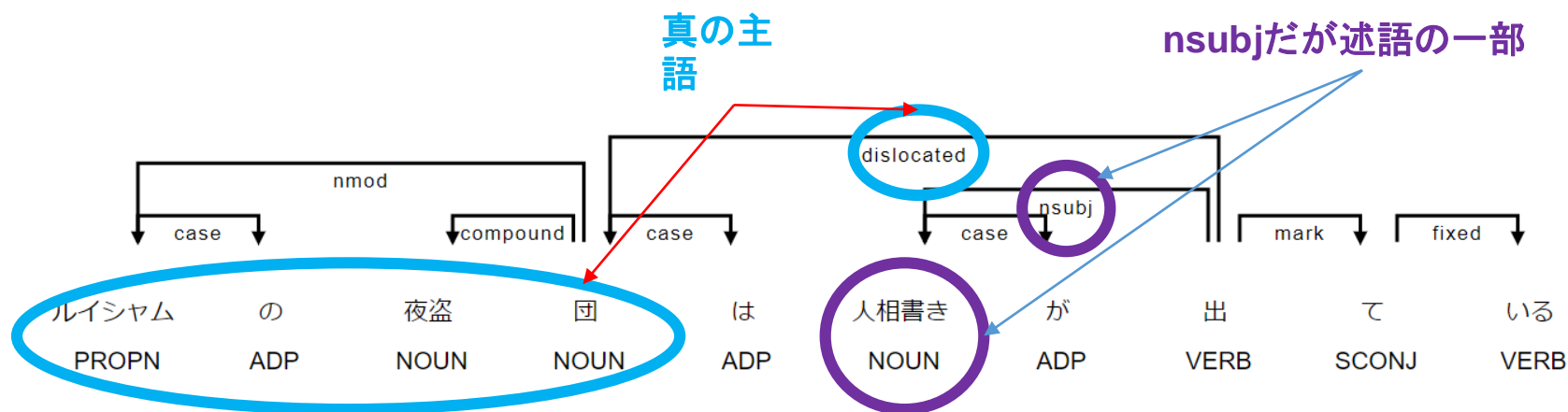
検証結果

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	0.6790	0.6773	0.3207
僧坊荘園	0.6875	0.5313	0.3644	0.5408	0.2201
踊る人形	0.9000	0.7591	0.6768	0.7946	0.4561
入院患者	0.9625	0.7051	0.6738	0.8294	0.6080
白銀号事件	0.9410	0.7028	0.6712	0.7871	0.4861
悪魔の足	0.9194	0.7043	0.6803	0.8123	0.5276
平均	0.8692	0.6557	0.6243	0.7403	0.4364

正答例

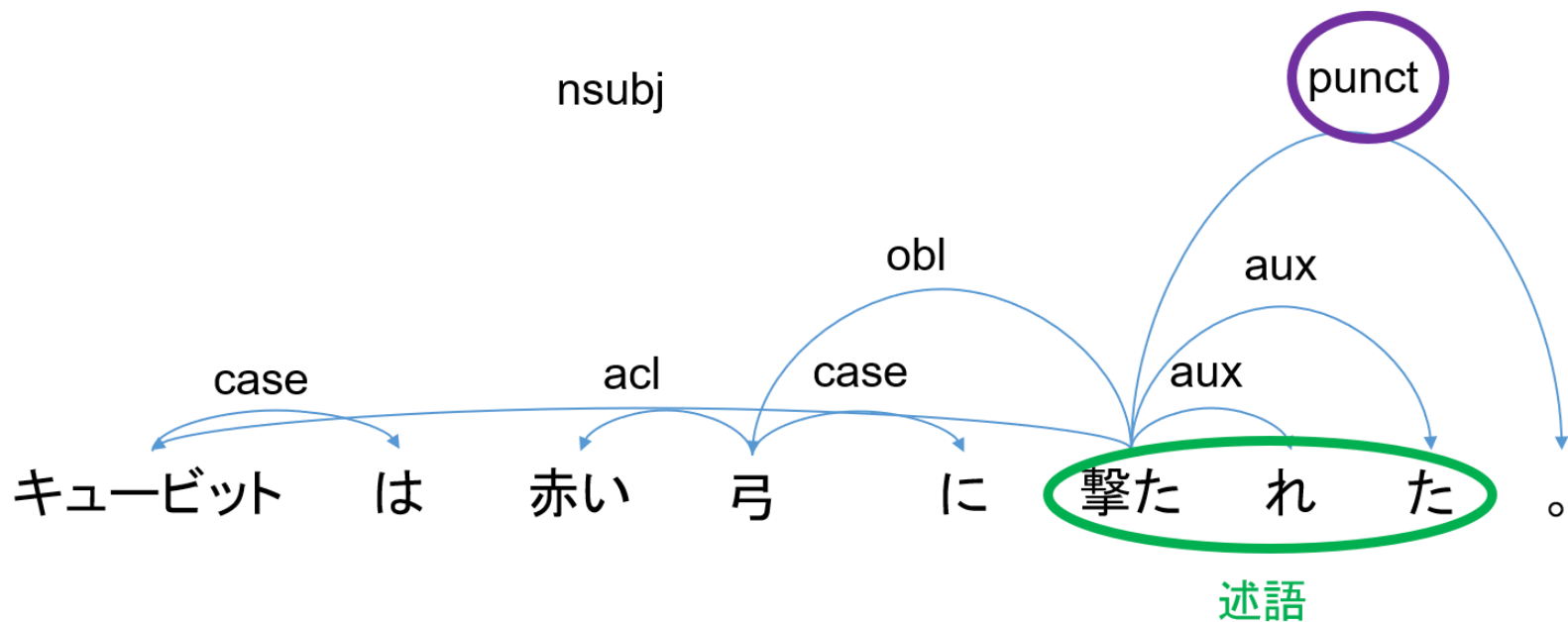
小説	分類	出力結果
花婿失踪事件	対象文	サザランドの父の友人は舞踏会に参加している
	主語	サザランドの父の友人
	述語	参加している
	目的語	舞踏会
僧坊荘園	対象文	ルイシャムの夜盗団は人相書きが出ている
	主語	ルイシャムの夜盗団
	述語	人相書きが出ている
踊る人形	対象文	キュービットとエルシィは撃たれた
	主語 1	キュービット
	主語 2	エルシィ
	述語	撃たれた
入院患者	対象文	パーシィ・トリヴェリアンの服装は地味である
	主語	パーシィ・トリヴェリアンの服装
	述語	地味である
白銀号事件	対象文	白銀号は競馬界で人気を保持し続ける
	主語	白銀号
	述語	保持し続ける
	目的語 1	競馬界
	目的語 2	人気

特別な例 (dislocated:転置)



* 基本のアルゴリズムとは別に、転置に関しては特殊な処理をしています。

特別な例 (punct:句読点)



* 基本のアルゴリズムとは別に、句読点に関しては特殊な処理をしています。

誤答例

小説	分類	出力結果	正解
踊る人形 場面15	対象文	キューピットは以下を言った	SV ×
	主語	キューピット	○
	述語	言った	○
	目的語	以下	16,17,19(場面ID)
悪魔の足 場面22	対象文	オーウェンとジョージとブレンダはモーティマーの家族である	S × ×
	主語1	オーウェン	○
	主語2	ジョージ	○
	主語3	ブレンダ	○
	述語	モーティマーの家族である	である
	目的語	(無し)	モーティマーの家族

まとめ

- 単語全体は7割前後、場面単位では4割前後、正解することが出来た。
- 主語に比べて述語や目的語の正答率が低いので、場面正解率を上げるためには述語や目的語の正答率を上げる必要がある。
- 他の場面を指す指示語があると上手くいかない
ので、まずは指示語とそれ以外を判別できるようにする。
- ズレをなくすために推理小説ナレッジグラフを構築する上での統一した基準を設ける。