

ウェブインテリジェンス最終レポート

平成 31 年 8 月 9 日

学籍番号 29114154

氏名 PHAM DUY

1 課題 1

1.1 手法

cabocha の解析結果を用いて、一つの形態素に対して、品詞が「名詞, 固有名詞, 人名」になっている形態素は人名の形態素と判断する。この規則にしたがって、以下の処理を行う。

- 全文章の中に人名の形態素を全て抽出して、出現頻度を表すリストを作成する
- 人名の出現頻度のリストより人名の出現回数を計算する

1.2 実装の説明

課題 1 は 2 つのメソッドで実装する。

- `getPersonName` メソッドパラメータ：
 - － `sentences` 解析した全ての文のリスト

戻り値：

- － 人名の出現頻度を表すリスト

解析結果から人名を表す形態素を検索し、人名の出現頻度を表すリストを返す。このメソッドの中に、全ての形態素を回して、その形態素の品詞が「名詞, 固有名詞, 人名」になっていると、その形態素の表層を文章に含まれている全ての人物のリスト（人名が重複不可）を表す `personNameList` セットに追加する。すると、上記の手法で取得できる全ての人物が `personNameList` で表される。しかし、上記の手法で取得できなかった人物が存在するので、取得できなかった人物を指定しないといけない。`exceptionPersonName` リストは取得できなかった人物を表すリストである。`exceptionPersonName` リストの全ての要素を `personNameList` に追加する。追加した後、もう一度解析した全ての形態素を回して、その形態素の表層格が `personNameList` に含まれるなら、人名の出現頻度を表す `personNameFreq` リスト（人名の重複は可能）に追加する。この `personNameFreq` リストは `getPersonName` メソッドの戻り値である。

- `countPersonFreq` メソッド
パラメータ：
 - － `sentences` 解析した全ての文のリスト
 - － `countPersonFreq` 上位の名前を返す数
 - － `exceptNameList` 考慮しない人物（例外の名前など）のリスト

戻り値：

- 人物の人名とその人名の出現回数を表すハッシュマップ

getPersonName メソッドを用いて、人名の出現頻度を表すリストを取得する。このリストの中に、考慮しない人物の人名を全て削除する。その後、人名の出現回数の計算を行う。計算した結果は result ハッシュマップである。このハッシュマップのキーは人物の人名であり、値はその人名の出現回数である。

1.3 実行結果

実行に使用した文章：まだらのひも

リンク先：https://www.aozora.gr.jp/cards/000009/files/50717_36994.html

課題 1 の実行結果：

上位 3 人の名前：

[ワトソン, ロイロット, ホームズ]

上位 3 人の名前とその名前が現れる回数：

{ワトソン=11, ロイロット=15, ホームズ=83}

1.4 考察

品詞が「名詞, 固有名詞, 人名」になっている形態素は人名を表す形態素であることが確認できたが、この手法で取得できない人名がある。この点により、使っていた cabocha 形態素解析器は文書の全ての形態素を正しく解析することができると言えない。これで、全ての人名を抽出するには、形態素解析器で取得できない人名をあらかじめ指定することが必要になった。

2 課題 2

2.1 手法

一つの文の中に、かかり先が -1 になっている文節は述語と判断し、文章の中の全ての述語を検索する。それぞれの述語に対して、述語項構造を表す PredArgStruct クラスのインスタンスを作成し、格の種類と格要素をセットする。文の中の全ての述語項構造を表す PredArgs クラスのインスタンスを作成する。その上で、係り受け解析結果を含む Sentence オブジェクトから PredArgs オブジェクトへ変換する getPredArgs メソッドを実装する。

2.2 実装の説明

手法で説明したように、課題 2 を実装するために、2 つのクラス :PredArgStruct クラスと PregArgs クラス、係り受け解析結果を含む Sentence オブジェクトから PredArgs オブジェクトへ変換する getPredArgs メソッドを実装する必要がある。

- PredArgStruct クラス

このクラスは述語項構造を表す。クラス変数として、述語項構造を表す `HashMap(String,String)` `args` を宣言し、述語を表す変数 `Pred`（形態素）を宣言する。述語項構造 `args` のキーは格の種類を表し、`args` の値は格要素を表す。

- PredArgs クラス

このクラスは `ArrayList(PredArgStruct)` を継承して、`PredArgStruct` オブジェクトの集合を表す。このクラスの `getSurface` メソッドは全ての述語項構造を表す文字列を返す。

- getPredArgs メソッド

このメソッドは `Sentence` クラス内に宣言し、係り受け解析結果を含む `Sentence` オブジェクトから `PredArgs` オブジェクトへ変換する役割を担う。このメソッドのパラメータは指定する表層格のリストである。戻り値は述語項構造の集合を表す `PredArgs` オブジェクトである。指定した表層格のリストにしたがって、文章内の全ての述語項構造を検索する。検索した結果は `PredArgStruct` オブジェクトとして保存し、一文ずつの全ての `PredArgStruct` オブジェクトを `PredArgs` オブジェクトに追加する。

- 全ての述語項構造を抽出するためには、全ての表層格を指定する必要がある。日本語の全ての表層格は以下になる。

が, に, を, は, から, まで, で, より, と, へ

2.3 実行結果

文章の全ての述語項構造：

ある{ガ格=事件}

考える{ヲ格=手際}

する{ト格=ひとつ, ハ格=もの}

ため

一蹴{ト格=かかわりたくない}

事件

ない{ニ格=他, ハ格=もの}

話

共用{デ格=ベイカー街, ハ格=我々, ヲ格=下宿}

ため

こと{ハ格=こと}

ある{ガ格=理由, ハ格=それに}

知る{ハ格=私, ヲ格=こと}

ねじ曲げる{ガ格=いずれも, ヲ格=真実}

立つ{ニ格=枕元}

回る{ヲ格=一五分}

向く{ハ格=私, ヲ格=方}

悪い

こう

ワトソン

声

具合{ハ格=「今朝」}

やってくる{ヘ格=ところ}

ところ{ハ格=僕}

まさか

依頼

面会{ニ格=僕}

待つ{デ格=居間}

考える{ト格=ある}

かかわる{カラ格=はじめ，ハ格=君}

決める{ニ格=君}

もの

ない{ハ格=もの}

よい

解き明かす{ヲ格=問題}

入る{ヘ格=居間，ニ格=一緒，ハ格=のちに}

立ち上がる{カラ格=椅子，ハ格=顔に}

挨拶{ニ格=朗らか，ハ格=ホームズ}

シャーロック{ガ格=「僕」}

ワトソン{デ格=友人，ハ格=こちら}

結構

つける{ニ格=暖炉，ヲ格=火}

方

いかが{ハ格=コーヒー}

ふるえる

ふるえる

言う{ト格=「」，デ格=声，ハ格=女性}

ホームズ

おそろしい

動揺{ニ格=ほど}

獲物

とれる

わかる{ガ格=すべて}

置く{ニ格=腕，ヲ格=手}

です

お{デ格=汽車}

見える{ガ格=切符，ニ格=あいだ}

いらっしゃる{マデ格=駅}

見つめる{デ格=目，ヲ格=友人}

こと

ほほえむ{ト格=「」，ハ格=ホームズ}

ござる{ガ格=跳ね，ニ格=左袖}

つく

限る{ニ格=場合}

おっしゃる

乗る{ニ格=始発，ハ格=わたくし}

耐える{ハ格=わたくし}

おかしい{ハ格=このままで}

誰

なれる{ニ格=力}

思い出す{ヲ格=こと}

ある{ガ格=こと}

教える{ヲ格=住所}

お{ヲ格=わたくし}

光

思う{ト格=できるか}

めくる{ハ格=ホームズ，ヲ格=事件簿}

思い出す

ワトソン

致す{ヲ格=助力}

報酬{ガ格=事件そのもの}

とき

お話{ヲ格=すべて}

役立つ{ニ格=考えるの}

話す{ト格=」，ガ格=依頼人}

なる{ガ格=こと，ニ格=気}

こと

くらい

わかる

見抜く{ハ格=あなた}

申す{ト格=ヘレン・ストーナ}

住む{ニ格=一緒}

ござる{ニ格=ストーク・モラン}

うなずく{ハ格=ホームズ}

存じる{ハ格=名}

裕福

広がる{ハ格=西，マデ格=ハンブシア}

崩す{デ格=賭博，ハ格=一族，ヲ格=身}

始末

甘んじる{ニ格=清貧，ハ格=先代}

取る{ヲ格=学位}

開業{ヲ格=医院}

殴る{ヲ格=執事}

免れる{デ格=ところ，ハ格=死刑}

なる{ニ格=人}

未亡人

赤ん坊{ハ格=とき}

譲る{デ格=遺言，ニ格=ロイロット博士，ハ格=金}

なる{ニ格=決まり}

亡くなる{ハ格=母}

遭う{デ格=近く，ニ格=鉄道事故}

なる{ニ格=こと}

こと{ハ格=わたくしども}

変わる{ガ格=様子}

閉じこもる{ニ格=中}

する{ヲ格=喧嘩}

ひどい

三

逃げる

つける{ト格=なる，ガ格=手}

突き落とす{ヘ格=中，カラ格=欄干，ハ格=父，ヲ格=鍛冶屋さん}

ほど

言う{ト格=野宿してよい}

ござる{ト格=よる，ハ格=父}

買い入れる{カラ格=現地，ヲ格=動物}

なる{ハ格=屋敷に}

存じる{ト格=なった}

見る{デ格=手，ハ格=こと}

同じ{ト格=わたくし，ハ格=髪}

お{ハ格=「お姉さま」}

二

つく{ニ格=死}

難しい{ガ格=付き合い，ハ格=人と}

許す{ハ格=ことだけ}

なる{ト格=運び，ハ格=ジュリア}

失う{ハ格=わたくし，ヲ格=姉}

見る{ヲ格=方}

お話

お話し{マデ格=「隅々」}

焼き付く{ト格=はっきり，ニ格=記憶}

一{ハ格=使っておりますの}

ござる{ガ格=居間，ニ格=中程，ハ格=寝室}

なる{ガ格=一番目，ニ格=わたくしの}

並ぶ{ニ格=廊下，ハ格=入り口}

よろしい{デ格=説明}

ある

見える{ガ格=芝地，ハ格=窓から}

思う{ト格=なったのではない}

悩ます{ニ格=臭い}
参る{ヘ格=部屋}
する{ヲ格=話}
申す{ヲ格=こと}
ヘレン
聞く{ヲ格=吹いているの}
何
吹く{ヲ格=口笛}
聞こえる{ガ格=口笛}
覚める{ト格=聞こえる, ガ格=目}
思う{ト格=聞かなくちゃ}
仕業
もの
安眠
こと
閉める{ヲ格=戸}
聞こえる{ガ格=音}
答える{ト格=, ガ格=ホームズ}
かける{ニ格=部屋, ヲ格=鍵}
いつも
眠れる
お話{ヲ格=先}
眠れる{ハ格=夜}
する{ガ格=胸騒ぎ}
双子{ハ格=わたくしと}
感じる{ニ格=敏感, ヲ格=こと}
鳴る{ハ格=窓}
聞こえる{ガ格=声}
声
飛び出す{ヘ格=廊下, ハ格=わたくし}
する{ガ格=気}
する{ガ格=音}
開く{ガ格=扉}
呑む{ハ格=わたくし, ヲ格=息}
進み出る{ト格=ふらふら, デ格=千鳥足}
抱きとめる{デ格=両腕, ハ格=わたくし, ヲ格=身体}
倒れる{ニ格=場}
震わせる{ト格=ぶるぶる, ヲ格=手足}
絞り出す{ヲ格=声}
助ける
ひも
言える{ガ格=物}
出会う{ト格=父}

失う{ハ格=姉，ヲ格=氣}
引き取る{ハ格=姉，ヲ格=息}
最期
ひとつ
挟む{ト格=」，ガ格=ホームズ，ヲ格=口}
確か{ハ格=金属音というの}
間違い
思う
聞く
服
まま
箱{ハ格=左手に}
なる{ニ格=ご覧，ヲ格=辺り}
なるほど
なさる{ヲ格=取り調べ}
もの
分かる{ハ格=こと}
閉ざす{ハ格=夜に}
分かる
こと
通す{ガ格=釘，ニ格=横}
ござる{ハ格=ことに}
見つかる{ハ格=身体に}
何{ハ格=結果}
亡くなる{デ格=原因，ハ格=お姉さま}
思う{ト格=死んだのだ}
いる{ニ格=中}
おりる{ニ格=どこか}
ある{カラ格=それ，ハ格=お心当たり}
ひも
ちょっと
振る{ニ格=横，ヲ格=首}
海
お話{ヲ格=先}
暮らす{カラ格=「事件，マデ格=先頃}
現れる{ガ格=方}
お{ニ格=クレイン・ウォータ，ハ格=お名前}
予定
なる{ニ格=こと}
こと
聞こえる{ガ格=口笛}
見あたる{ハ格=部屋に}
参る{ヘ格=こちら，ニ格=お願い，ヲ格=お助け}

行動

すべて{デ格=それ，ハ格=お話}

ある

お{ヲ格=こと}

こと

折り返す{デ格=つまん，ハ格=ホームズ}

残す{ニ格=手首}

受ける{ヲ格=扱い}

言う{ガ格=ホームズ}

隠す{ニ格=袖の下，ヲ格=手首}

わかる{ガ格=加減}

続く{ガ格=沈黙}

ある{ガ格=事情}

開く{ト格=」，ヲ格=口}

ある

見せる

申す{ト格=出る}

大丈夫

訳{ハ格=出てもらうの}

ワトソン

参る{デ格=ふたり}

なさる{ハ格=あなた}

存じる{ト格=済ませたい}

存じる{ト格=間に合うか}

伺う{デ格=ふたり}

ある{ガ格=こと}

いかが

ない

軽い{ガ格=心}

お待ち{ヲ格=お目にかかるの}

退室{ニ格=静か，ハ格=」依頼人}

ワトソン

訊ねる

思う{ト格=話だ}

そう

なる{ニ格=こと}

なる

付く

こと

存在

事実

謎

音{ハ格=（これ）}

考える{ハ格=解き明かせそうだと}
する{ガ格=ロマたち，ヲ格=何}
ある{ガ格=穴}
思う{ト格=ある}
確かめる
何ごと
立ちほだかる{ガ格=大男，ニ格=入り口}
風
振り回す{デ格=手，ニ格=トップ・ハット，ヲ格=狩猟鞭}
ぎりぎり
猛禽{ハ格=顔}
ホームズ{ガ格=「どっち」}
名前
名乗る{ハ格=まず}
問い返す{ニ格=静か，ハ格=「友人」}
*{ハ格=「わし」}
先生
切り返す{ト格=「」，ニ格=おだやか，ハ格=ホームズ}
おかけ
来る{ヘ格=ここ，ガ格=娘}
見る{ヲ格=入るの}
する{ニ格=お前，ヲ格=何}
悪い{ガ格=寒さはたち}
言う{ハ格=ホームズ}
聞く{ト格=しゃべりおった}
怒る{ハ格=老医師}
いい{ガ格=出来，ハ格=方}
動じる{ト格=「」，ハ格=友人}
はぐらかす
ふるわす{ハ格=「闖入者」，ヲ格=鞭}
ガキ
知る
お節介{ハ格=職}
笑う{ト格=にこり，ハ格=「友人」}
浮かべる{ヲ格=笑み}
笑う{ト格=くすくす，ハ格=ホームズ}
傑作
よろしく{ハ格=際，ヲ格=戸締まり}
寒い{ガ格=すきま風}
帰る{カラ格=こっち}
つつこむ{ニ格=こと，ヲ格=首}
つける
後悔{ト格=する}

見る

折り曲げる{デ格=手}

こる

出る{カラ格=部屋，ハ格=老医師}

人物

披露{ヲ格=こと}

伸ばす{ニ格=まっすぐ}

暴挙

なる{ニ格=薬味}

なる{ニ格=こと}

行く{ト格=朝食}

行く{ヘ格=博士会館，デ格=あと，ハ格=僕}

思う{ト格=ある}

一時

書き込む{ガ格=抜き書き，ニ格=いっぱい}

見る{ヲ格=遺言状}

なる{ハ格=計算しなくて}

ところ

受け取れる

なる{ガ格=うま味，ニ格=少しだけ}

損なう{ニ格=ほど}

無駄{ハ格=仕事}

証明{ト格=ある}

場合

分かる{ト格=邪魔になる，ハ格=老人}

急ぐ{ヘ格=ウォータルー駅，デ格=呼ん}

ありがたい{ト格=忍ばせてくれる}

対抗

足りる{デ格=それだけ}

揺る{ヲ格=田舎道}

雲{ハ格=空に}

含む{ハ格=空気，ヲ格=香り}

感じる{ハ格=私，ヲ格=対比}

沈む{ニ格=考え，ハ格=友人}

指さす{ヲ格=彼方}

ま

なる{ニ格=森，ハ格=頂上}

のぞく{ガ格=屋根}

ごぜ{デ格=「そう」}

がする{デ格=お屋敷}

答える{ガ格=御者}

やる{ヲ格=普請}

ホームズ{ト格=「」}

行く{ガ格=「そこ」}

村{ガ格=「あそこ」}

指さす{ハ格=「御者」，ヲ格=集落}

*

歩く

*

眺める{ハ格=ホームズ}

賢明{ガ格=まま}

引き返す{ヘ格=方，ト格=がたごと，ハ格=馬車}

いい{ガ格=方}

言う{ガ格=ホームズ}

相違{ニ格=思う}

立つ

*

参る

挨拶{ハ格=依頼人}

にじみ出る{ガ格=うれしさ，ニ格=顔}

お待ち

する{ニ格=熱烈，ヲ格=握手}

帰る{ニ格=めった，マデ格=夕方}

いただく{ヲ格=光栄}

説明{ニ格=手短，ヲ格=事}

なる{ニ格=真っ青，ハ格=ストーナ嬢，マデ格=唇}

こと

つける{ヲ格=あと}

鳥

言う{ト格=何}

方{ハ格=用心すべき}

追う{ガ格=人間，ヲ格=自分}

こと

差し上げる

使う{ニ格=有効，ヲ格=時間}

お願い{ヘ格=方，ヲ格=ご案内}

のびる{ヘ格=左右，ガ格=棟，ニ格=翼状}

体{ハ格=一方}

手入れ

見て取れる{ガ格=こと}

見える{ハ格=姿}

調べる{カラ格=外側，ハ格=ホームズ，ヲ格=窓}

部屋{ガ格=近いの}

休む{デ格=真ん中，ハ格=わたくし}

です

見える{ハ格=ものに}

ない

口実

いわく

面す{ガ格=寝室}

ある{ハ格=窓}

もの

できる{ハ格=通り抜け}

忍び込める{ハ格=そちらから}

かける{ヲ格=（かんぬき）}

従う{ニ格=指示，ハ格=ストーナ嬢}

尽くす{ト格=こじ開けよう，ハ格=ホームズ，ヲ格=手}

いく

ない{ガ格=隙間}

埋め込む{ニ格=石壁}

ある{ガ格=ところ，ハ格=理論に}

通れる{ヲ格=鎧戸}

調べる

出る{ニ格=廊下}

向かう{ニ格=部屋，ハ格=ホームズ}

田舎

ある{ガ格=化粧台，ハ格=左側に}

すべて{ガ格=加えたの}

敷く{ガ格=ウィルトン絨毯，ニ格=中央}

違い{ニ格=もの}

見回す{ハ格=ホームズ，マデ格=隅々}

通じる{ヘ格=どこ}

指さす{ハ格=ホームズ，ヲ格=綱}

乗っかる{ガ格=先，ニ格=上}

通じる{ヘ格=部屋}

真新しい{ヨリ格=もの}

つける{ニ格=二年前}

つける{ガ格=お姉さま}

話

致す{デ格=自分，ハ格=こと}

不

お待ち

近づける{ニ格=床，ハ格=「ホームズ，ヲ格=頭}

調べる{ハ格=床に，マデ格=隙間}

調べる{ト格=すむ，ニ格=同じよう，ヲ格=木材}

回す{ヘ格=下，カラ格=上，ヲ格=壁}

引っ張る

飾り

つながる{ニ格=針金}

面白い

動く

わかる{ガ格=「訳」}

気づく{マデ格=今}

妙

引っ張る{ヲ格=綱}

ある{ガ格=点}

あける{へ格=部屋，ハ格=大工，ヲ格=穴}

入る{ガ格=外気}

もの

頃

ござる{ガ格=ところ}

通風

参る{ニ格=調べ，ヲ格=部屋}

もの{ハ格=家具}

もの{ガ格=肘掛椅子，ニ格=わき}

調べる{ハ格=ホームズ}

何{ハ格=「中に」}

叩く{ハ格=ホームズ，ヲ格=金庫}

書類

ごらん{ヲ格=中}

前

ぎっしり{ガ格=書類，ハ格=中に}

いる{ガ格=猫，ニ格=「中」}

ない

なさる

下ろす{ハ格=ホームズ，ヲ格=小皿}

おりる{ニ格=うち，ハ格=猫}

存じる

ない

増える{ガ格=こと}

調べる{カラ格=それ，ニ格=念入り，ハ格=ホームズ，ヲ格=部分}

はっきり

面妖{ハ格=これ}

犬

結ぶ

ワトソン

なる{ニ格=輪}

まったく

末

未恐ろしい

*

出る{へ格=方}

寄せる{ハ格=友人，ヲ格=眉根}

とき{ハ格=それ}

待つ{マデ格=終わる}

*

言う{ガ格=ホームズ}

こと

いたす

切迫{ハ格=「事態」}

できる{ハ格=保証}

お

なる

見つめる{デ格=目，ヲ格=ホームズ}

絶対{ハ格=「これ」}

説明

宿屋

申す{ト格=クラウン}

見える{カラ格=あそこ，ガ格=窓}

下がる{ヘ格=部屋}

合図{カラ格=そこ，デ格=ランプ，ニ格=僕ら，ハ格=あなた}

部屋

いただける

大丈夫

お{ニ格=僕ら}

なさる

突き止める{ヲ格=正体}

付く{ガ格=見当}

かける{ニ格=袖，ヲ格=手}

お願い

何{ハ格=死因}

先{ガ格=つかむの}

合う{ガ格=考え}

亡くなる{ハ格=姉}

違う

ある{ガ格=原因}

いたす

台無し{ハ格=知れて}

お{ヲ格=気}

安心

手配{ヲ格=寝室}

できる{ガ格=こと，ハ格=窓から}

見える{ガ格=帰宅するの}

覆う{ハ格=巨体}

見える

漏れる{ガ格=光}
ワトソン
言う{ガ格=ホームズ}
腰掛ける{ニ格=椅子，ハ格=我々}
ためらう
見える{ガ格=危険，ニ格=目}
お{ハ格=「私」}
助かる{ト格=「来てくれる」}
決まり{デ格=行く}
ありがたい
見逃す{デ格=部屋，ハ格=私}
する{ハ格=僕，ヲ格=推理}
思う{ト格=変わらない}
引き
付く
見る
思える
通れる
踏む{ト格=あるだろう}
ホームズ
本気
匂う{ガ格=煙草}
分かる{ト格=はず、}
違い{ニ格=もの}
気づく
導く{ガ格=答え}
なる{ガ格=あれ，ニ格=害}
一致{ガ格=時期，ニ格=奇妙}
亡くなる{ガ格=女性}
気が付く{ニ格=何か}
ある{ガ格=つながり}
ない{ハ格=ところ}
固定{デ格=金具，ニ格=「床」}
ある{ニ格=寝台どこ}
動く{ハ格=「ご婦人，ヲ格=寝台」}
呼ぶ{ト格=綱}
鳴る
張る{ト格=「」，ハ格=私，ヲ格=声}
分かる{ガ格=こと，ニ格=おぼろげ}
食い止める{デ格=瀬戸際，ヲ格=犯罪}
そう
なる{ニ格=犯罪者}
筆頭{ハ格=プリチャド}

行く{ハ格=僕ら，ヲ格=上}
抜ける{ハ格=気，マデ格=終わる}
する
包む{ニ格=闇，ハ格=方向}
灯す{ガ格=明かり}
立ち上がる{ガ格=ホームズ}
窓
告げる{ト格=泊まるかもしれない、}
なでる{ガ格=風，ヲ格=頬}
導く{ト格=使命へ，ヲ格=我々}
入れる{ヘ格=中}
出る{ト格=かいくぐる，ニ格=芝地}
駆け込む{ヘ格=中}
呑む{ハ格=私，ヲ格=息}
耳打ち{ト格=すぐ，ニ格=私，ヲ格=忍び笑い}
家族
つぶやく{ト格=」，ハ格=ホームズ}
ど{ヲ格=こと}
いる
かみつく{ガ格=そいつ，ニ格=肩，ハ格=瞬間に}
正直
見渡す{ハ格=友人，ヲ格=中}
ない{ト格=とき，ハ格=変わり}
ささやく{デ格=声，ハ格=ホームズ}
台無し{ガ格=物音，ヲ格=計画}
うなづく{ト格=分かった，ハ格=私}
じっと
お
うなづく
ない
関わる{ニ格=命}
準備
椅子{ハ格=君}
置く{ニ格=隅，ハ格=私}
立てかける{ニ格=寝台，ハ格=ホームズ，ヲ格=杖}
並べる{ニ格=そば，ヲ格=蠟燭}
包む{ニ格=闇，ハ格=我々}
ない{ハ格=こと}
ない
座る{ハ格=友人}
張りつめる{ニ格=同じよう，ヲ格=気}
待つ{ニ格=闇，ハ格=我々}
一{カラ格=外，ガ格=鳴き声}

相違{ニ格=放たれた}
響く{デ格=空，ガ格=鐘，ニ格=おごそか}
強い{ニ格=十五分ごと}
長い{ガ格=十五分間}
起こる{ガ格=何か}
消える
匂い
つける{デ格=部屋，ニ格=（ランタン），ヲ格=火}
する{ガ格=気配，ニ格=静か}
残る{ガ格=匂いだけ，ハ格=あと}
澄む{ヲ格=耳}
聞く{ガ格=物音，ニ格=唐突}
音
食らわせる{デ格=杖，ニ格=引き綱，ヲ格=一閃}
ワトソン
叫ぶ{ガ格=ホームズ}
見える{ハ格=私に}
する{ニ格=耳，ハ格=口笛}
できる{ガ格=こと}
顔
見上げる{ハ格=ホームズ，ヲ格=方}
聞こえる{ガ格=悲鳴}
なる{ト格=叫び}
いう{ト格=覚ました}
立ちすくむ{ハ格=ホームズ}
戻る{ニ格=静けさ，ハ格=あたり}
今
吞む{ハ格=私，ヲ格=息}
知らせ
答える{ト格=」，ガ格=ホームズ}
最善{ガ格=結末}
たまえ
行く{ヘ格=部屋}
進む{ハ格=ホームズ，ヲ格=廊下}
ない{ハ格=中から}
入る{ヘ格=中，ハ格=ホームズ}
かける{ニ格=引き金，ヲ格=指}
奪う{ニ格=光景，ハ格=私，ヲ格=目}
見える{ガ格=金庫}
かける{ヲ格=腰}
ある{ガ格=鞭}
もの
にらみつける{ヲ格=隅}

しめつける{ヲ格=頭}
ない
ひも
ささやく{ハ格=ホームズ}
出る{ニ格=一步前, ハ格=私}
もたげる{ヲ格=鎌首}
張る{ガ格=ホームズ, ヲ格=声}
蛇
十
必然{ガ格=返る}
はまる{ニ格=穴, ハ格=策士}
追い込む{ヘ格=中, ヲ格=毒蛇}
報告{ヘ格=警察, ヲ格=次第}
投げ入れる{ヘ格=中, ハ格=ホームズ}
真相{ガ格= 以上}
話す{ニ格=簡単, ハ格=こと}
送り届ける{ヘ格=元, ハ格=我々}
結論づける{ト格=事故死}
教える{デ格=車中, ハ格=ホームズ}
下す{ハ格=僕, ヲ格=推定}
一{ト格=行う}
いる{ガ格=ロマ}
口走る{ガ格=ご婦人, ヲ格=言葉}
表す{ヲ格=こと}
追いやる{ヘ格=方向, ガ格=それ}
考える
はっきり{ト格=ものではない, ハ格=それ}
向ける{ニ格=綱と, ハ格=注意}
浮かぶ{ガ格=考え}
よぎる{ガ格=姿, ニ格=頭}
出る{ガ格=信憑性}
ある{ガ格=もの}
言う{ト格=悪知恵}
いい{ガ格=都合}
当然
こと
呼び戻す{ハ格=博士, ヲ格=蛇}
する{ヲ格=訓練}
戻る
下りる{ヘ格=寝台, ハ格=蛇}
なる{ト格=犠牲}
たどり着く{ニ格=前}
分かる{ガ格=こと}

必要{二格=ため}
ない
相違{二格=音}
移る{ト格=握ろう, 二格=行動}
食らわせる
逃げ失せる{へ格=通風口, ハ格=蛇}
なる{ト格=結果}
かみつく{二格=人間}
感じる{ヲ格=呵責}

2.4 考察

表層格だけで述語項構造を作成したが、深層格まで推定するために機械学習を用いる方法が流行っている。機械学習を用いて二の深層格を推定する参考資料：https://www.anlp.jp/proceedings/annual_meeting/2014/pdf_dir/D6-2.pdf

3 課題3

3.1 手法

最初に、探索役（使用した文章の中にはホームズとワトソン）の道場人物を取り除き、登場頻度が上位3名の登場人物を抽出する。この作業は課題1で作成した countPersonFreq メソッドを用いることで、実装できる。次に、一文ずつを考え、その文の中に取得した人物が現れるかどうかを検知する。一人でも検知できれば、課題2の getPredArgs メソッドでその文の述語項構造を抽出し、表としてまとめる。文章その順番で検索するため、抽出した述語項構造も時系列順に並べられる。結果は標準出力に表示する表ではなく、JTable に表示する表である。

3.2 実装の説明

課題3を実装するのは PersonRelAnalyzer クラス内の getKnowledgeGraph メソッドである。パラメータ：sentences 解析した全ての文のリスト最初に、countPersonFreq メソッドで探索役（使用した文章の中にはホームズとワトソン）の道場人物を取り除き、登場頻度が上位3名の登場人物を抽出する。考慮しない人物をホームズとワトソンと指定する。そして、一文ずつを考え、その文の中に取得した人物が現れるかどうかを表す personWithFlag という HashMap を作成する。この HashMap のキーは人物の名前、値は false と初期化する（現れていないという意味）。文の全ての形態素を検知し、人物の名前と一致する形態素があれば、対応する人物の personWithFlag の要素の値を true とセットする。そして、personWithFlag の全ての値の中に、true になっている値があれば、その文の述語項構造を getPredArgs メソッドで取得し、対応する出力データを作成する。出力データはそれぞれの登場人物が現れる述語項構造を表す。一文は表の一行として処理を行う。

3.3 実行結果

ヘレン	ロイロット	ジュリア
	話	
申す{ト格=ヘレン・ストーナ}	知る{ハ格=私, ヲ格=こと}	
	住む{ニ格=一緒}	
	ござる{ニ格=ストーク・モラン}	
	未亡人	
		赤ん坊{ハ格=とき}
	譲る{デ格=遺言, ニ格=ロイロット博士, ...}	
	閉じこもる{ニ格=中}	
	ひどい	
	言う{ト格=野宿してよい}	
		存じる{ト格=なった}
		なる{ト格=運び, ハ格=ジュリア}
	なる{ガ格=一番目, ニ格=わたくしの}	
	思う{ト格=なったのではない}	
ヘレン		
ひも		
音{ハ格= (これ)}		
	*{ハ格= 「わし}	
	がする{デ格=お屋敷}	
	帰る{ニ格=めった, マデ格=夕方}	
	部屋{ガ格=近いの}	
	もの{ハ格=家具}	
	台無し{ハ格=知れて}	
	見える{ガ格=帰宅するの}	
	匂う{ガ格=煙草}	
	行く{ヘ格=部屋}	
	かける{ヲ格=腰}	
	真相{ガ格= 以上}	

3.4 考察

課題3には課題1と課題2で作成したメソッドに基づいて道場人物を除き、登場頻度が上位3名の登場人物が現れる述語項構造を時系列順に表にまとめるので、課題3の結果は課題1と課題2の結果に依存する。しかし、解析した述語項構造は表層格だけで作られたものだったから、各述語項構造の対象と関連がわからないので、登場頻度が上位3名の登場人物が現れる述語項構造の表が十分な情報を持たない。

4 課題4

4.1 手法

登場頻度が上位3名の登場人物間の関係をわかりやすくするような可視化手法: 課題3の結果から、各人物が現れる述語項構造の述語はその述語項構造の前後または同じ時系（課題3の表の同じ行）に、他人が現れる述語項構造の述語と同じであれば、その二人の登場人物の間に関係があると考えられる。

4.2 考察

課題3の結果から、ヘレンとロイロットと一緒に住んでいることが推定できる。例えば、以下の図の赤い枠の中には推定することを表す。

深層格まで述語項構造を作成できるなら、登場人物の間の関係を取り出すことができると考えられる。

ヘレン	ロイロット	ジュリア
	話	
	知る{ハ格=私, ヲ格=こと}	
申す{ト格=ヘレン・ストーナ}		
	住む{ニ格=一緒}	
	ござる{ニ格=ストーク・モラン}	
	未亡人	
		赤ん坊{ハ格=とき}
	譲る{デ格=遺言, ニ格=ロイロット博士, ...}	
	閉じこもる{ニ格=中}	
	ひどい	
	言う{ト格=野宿してよい}	
		存じる{ト格=なった}
		なる{ト格=運び, ハ格=ジュリア}
	なる{ガ格=一番目, ニ格=わたくしの}	
	思う{ト格=なったのではない}	
ヘレン		
ひも		
音{ハ格= (これ)}		
	*{ハ格= 「わし}	
	がする{デ格=お屋敷}	
	帰る{ニ格=めった, マデ格=夕方}	
	部屋{ガ格=近いの}	
	もの{ハ格=家具}	
	台無し{ハ格=知れて}	
	見える{ガ格=帰宅するの}	
	匂う{ガ格=煙草}	
	行く{ヘ格=部屋}	
	かける{ヲ格=腰}	
	真相{ガ格= 以上}	

深層格まで推定できるなら、
ここにも[住む]という述語項構
造が入ると予想した

図 2: 課題 4 の考えた図