

分散表現型物語機能辞書の提案と試作

村井源 (はこだて未来大学 複雑系知能学科)

物語テキストの機械的な意味処理と自動生成を行う場合を想定し、手動分散表現辞書の構築手法を提案した。また、主体的語彙と犯罪・災害・死などのオチに関係の深い語彙を中心に、37種類の属性を持つ約8000語の辞書を試作した。手動分散表現辞書は構築の難易度は高いが、人間からも機械からも可読で、拡張性と応用性の高さを兼ね備えた辞書形式である。

Proposition and trial production of distributed representation dictionary for narratological functions

Hajime Murai (Department of Complex and Intelligent Systems, Future University Hakodate)

In order to carry out automatic semantic processing and text generation for story texts, construction method for manual distributed representation dictionary was proposed. Moreover, prototype manual distributed representation dictionary of about 8000 words based on 37 attributes was developed. Included vocabulary is focused on agent words and punch-line words such as criminal, disastrous situations and death. Though manual distributed representation dictionary is hard to construct, it is readable from both human and machine, and also it has both high scalability and applicability.

1. まえがき

文学研究には文章表現、作者や創作された時代背景の研究などとともに作品を構成する物語の構造自体を対象とする物語論、ナラトロジーなどと呼ばれる分野がある。ただし物語論自体は文学作品に限らず昔話や神話、映画や演劇など物語的な構造を持つすべての作品に対して適用可能なアプローチである。

伝統的な物語論は主として登場人物の機能や役割[1]、物語の意味的な構造[2]、語りの順序や文体など[3]の側面を切り口として、物語とは何かを問うものであった。用いられている手法は人間の解釈者が個々の作品を読み、その要素や構造を抽出するという人文学的でケーススタディ的な分析が主流であった。そのため、物語の構造としてどのようなものがあるか、どのような要素が影響しているかに関しては種々の知見が得られているものの、それらが総体としてどのように人間の感情を動かし、面白さを生み出しているのかという全体像や、物語の一般的性質は未だ明らかになっていない。

このような状況に対して近年はコンピュータを用いた物語論の研究が徐々に増えてきており、客観性を担保しつつ、大規模な作品の分析を行うことで物語の構造を科学的に明らかにしようという試みがなされている。しかしながら、現状としてはテキストマイニング等を用いた表面的な単語レベルの分類が主流であり[4]、伝統的な物語論が対象としてきたような解釈の結果として得られる高次の物語構造の抽出や分類には至っていない。この理由としては様々な原因が考えられ

るが、その一つとして、コンピュータに物語テキストの意味理解をさせるために必要となる適切な辞書群が存在しないことがあるのではないかと考えている。

そこで筆者は星新一作品などを対象に物語上の機能を示す言動的な語彙を分類するための動詞や動詞的機能を持つ複合的なフレーズ用の辞書[5]や、登場人物などの物語中で主体的な機能を持つ語彙を抽出するための辞書[6]などを作成してきた。これらの辞書を用いることでテキスト中に明示されている場合には、どの主体がどのような物語機能を担っているかをある程度特定することは可能である[7]。

しかし実際の物語テキストにおいては常識的に自明なことがらや、前後から類推可能なことがらは明示的には記述されずに省略されることが多い。例えば、「店」が物語の場面であることが語られると、他に特に説明で明示されなくとも「店員」がその場にいることは当然のこととして扱われる。そのため、人間が読む場合には明示されていない動作主体として「店員」の可能性があることを考慮に入れて解釈をすることができるが、コンピュータに分析させる場合には動作主体が明示されないためエラーとなってしまう。

また物語テキスト中での物語の登場人物等の主体を表す語彙は頻繁に言及されることになるが、それゆえに様々に言い替えたり省略されたりすることが少なくない。例えば「花子」について語っていた次の文で「その女」という言葉が記されている場合、人間の場合には無意識的に「花子」は女性の名前であり、かつ直前に出現するため「花子」と「その女」が同一人物である可能性を

考慮する必要があることは判断できるが、そこでは両単語が示す性別が等価であるというような単語に付随する情報が暗黙的に用いられている。

このように末端の単語レベルにも含有された常識的知識やそれに基づく省略や言い替え表現が多様多様に存在するため、コンピュータが物語テキストを適切に処理できるようになるためには主体と言動の辞書だけではなく、一般的な名詞や、様相を表現する形容詞・形容動詞なども含めた上で、さらに言外の意味として含まれる種々の常識的な知識をも内包した機械可読辞書の構築は不可欠であると考えられる。

ただし、人間がテキストを解釈する場合にどのような背景的知識が用いられるかということの全体像は未だ明らかになっていない。そのため、物語テキストの意味解釈に用いる機械可読辞書をどのような構成にすべきかということも残念ながら決定不能である。そこで、機械可読辞書の設計においては内容の変更や追加更新が容易であることが望ましいと考えられる。

また、物語テキストの処理のもう一つの側面として人工知能などを用いた物語テキストの生成の分野がある。テキストの生成においてもやはり、人間の場合は暗黙的常識的知識に基づいて文脈に合わせた単語や表現の選定が行われていると推測される。このような意味処理を人工知能に実現させる手法もまた未確立であるため、テキスト生成にも利用可能な機械可読辞書の構成としては、現段階では多くのアルゴリズムから利用可能な汎用的で単純な形式での実装が望ましいと考えられる。また、行われている意味処理がなんであるかを人間が確認可能であるという意味でも可能な限り単純な形式での構築が有用であると考える。

2. 機械可読辞書の構築法とその課題

機械可読の意味辞書を構築する方法論としては意味ネットワークやオントロジーなどの、概念を人間にも理解可能な記号や関係性の構造として記述する手法が一般的であった。これらの記述生成を自動化する試みも種々あるが自動化生成した場合には概して精度が低く、実用レベルで構築するためには最終的に人間による確認が必要である。例えば世界的に用いられる WordNet[8] や日本語で頻繁に利用される分類語彙表[9]などは基本的に人手によって構築されている。

このため、意味ネットワークやオントロジーなどの記号とその関係性からなる機械可読辞書は、辞書編纂者によって作成された通常の類語辞書等に近い精度を持たせることが可能である反面、その構築に多大なコストがかかるという性質を持つ。また、人手による構築であるため客観性の検証は困難である。語彙間の関係性や全体的構造も辞書構築者の解釈に依存するため、類語辞典と同様に構築者によってカテゴリの種類や構成が

異なってしまう。そのため、辞書構築者が想定した意味分類や推論には利用可能であるが、それ以外の様々に異なった観点から単語の意味を柔軟に比較するような処理の実現は難しくなる。

一方で近年 Word2vec などのニューラルネットワークや機械学習を用いた意味表現の自動獲得の手法が盛んになっている[10]。Word2vec などの自動的に意味表現を獲得する手法においては分散表現と呼ばれるような意味をベクトルとして表現する形式が主流である。Word2vec で得られたベクトルは意味の加法や減法の計算が可能であることから、大きな注目を集め現在も急速に関連研究が進められている。

Word2vec で得られた分散表現での加法や減法の例としてよく用いられるのは、

“King” - “Man” + “Woman” = “Queen”

であるが、ここでの “King”, “Queen”, “Man”, “Woman” はそれぞれ、訓練用のテキストデータ中で各単語の周辺に共起する語彙から学習された 300 次元程度のベクトルを示している。

“King”-“Man” + “Woman” というベクトル演算をした結果に最も近いベクトルを抽出すると実際に “Queen” が出てくるため人間がもつ単語の概念間の意味処理に近い作業をコンピュータ上で実現したということによって大きな話題となった。分散表現はベクトルで構成されるため加法や減法以外にも様々な数学的な変形や分析が可能であり、多様な類似度の計算が実現できる。

しかし、言語の性質上単に共起すれば意味が同じになるわけではなく、品詞や機能語、文法上の役割や能動態・受動態など様々な要因でテキスト中での単語が示す概念は異なりうる。実際に名詞では Word2vec から得られた分散表現は比較的高い精度を得られるものの、動詞での精度は低くなってしまっている[11]。おそらく、動詞の方が能動態・受動態などの態による違いや慣用句的な複合語による意味の変化の影響を強く受けるためと推測される。

また、分散表現で得られる類似度は多様な種類が内包されていると考えられるが[12]、ベクトルの各次元が何を意味するかに関しては個別的にいくつかの特徴量を取り上げた考察はあるが、全体的にその構造を理解する方法は発見されていない。そのため、ある類似度の高い語彙の関係が同義語なのか上位・下位語なのかあるいは全体語・部分語なのか判別することは困難である。また Word2vec によって得られる分散表現のベクトルは元となるコーパスの内容はもとよりや種々のランダムなパラメータによっても構造が変化するため、既存の分散表現に基づいた統合や、詳細化、高度化などによって発展させるようなことも困難である。構造が単純で人工知能等の情報処理分野との親和性は高いが、ブラックボックス性

が高く、追加や更新などの処理には不向きであると言えよう。

3. 多属性記述を拡張した手動分散表現

現状ではオントロジック的記述と分散表現的記述の双方にメリットデメリットがあるが、理想的には両方のメリットを合わせたような表記法が存在することが望ましいと考えられる。これらの統合を図る既存研究としては WordNet 等のカテゴリを用いて、分散表現の自動生成結果の精度を上げる研究などがいくつか行われてきている[12]。また逆に分散表現の結果から語彙の上位・下位関係を抽出する試みも始められている[13]。

本研究では、これらに対して人間の側からも機械の側からも理解可能な可読辞書形式として、分散表現自体をオントロジック的知識に基づき人手で構築する方法を提案したい。具体的には、人間が理解可能な形で各次元の特徴量(例えば、物質性、抽象性、生物か否か、知能を持つか否か等)を規定した上で、各特徴量に数値をいれたベクトルを作成するという手順である。例として鳥と飛行機の場合のオントロジック例とその分散表現への変換例をそれぞれ図1と表1に示す。

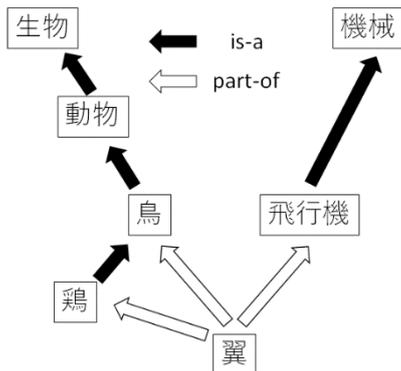


図1 オントロジック例

表1 分散表現例

	生物	動物	鳥	鶏	機械	飛行機
生命	1	1	1	1	0	0
能動	0	1	1	1	1	1
翼	0	0	1	1	0	1
飛行	0	0	1	0	0	1

図1ではごく単純なオントロジック型概念構造の例として is-a 関係と part-of 関係のみを使っている。Is-a 関係とは、概念の包含関係を示すもので、例えば「動物」は「生物」の下位概念であるため is-a 関係で接続される。Part-of 関係とは、ある二つの概念間に全体一部分の関係があること

を意味しており、例えば「翼」は「鳥」の一部であると同時に「飛行機」の一部でもある。

このような関係性を持つ語彙からベクトルでの分散表現を構築する方法は無数に考えられるが、単純化のために語彙の意味的な差異の最低限の弁別ができる次元を設定する方針を採用。図1の場合「生物」「動物」「鳥」「鶏」と「機械」「飛行機」の群を弁別するために、「生命」のあるなしを示す次元をまず付与する。次に「生物」と「動物」「鳥」「鶏」を弁別するために「能動」性の有無を示す次元を追加する。さらに、「動物」と「鳥」「鶏」の弁別と、「機械」と「飛行機」の弁別を可能とする「翼」の有無を示す次元を追加する。最後に、「鳥」と「鶏」を弁別するために「飛行」能力を示す次元を追加すると、全語彙を異なったベクトルで分散的に表現することが可能となる(表1)。なお、「機械」と「動物」に共通して「能動」性があり、「鳥」と「飛行機」に共通して「飛行」能力があることは図1の関係性には含まれず、表1で追加されたものである。

表1のように分散表現を記述すると、先ほどの Word2vec の場合と同じく、

$$\text{「鳥」} - \text{「生物」} + \text{「機械」} = \text{「飛行機」}$$

のような意味の演算処理が可能となる。

このように明示的に意味を定義された属性を次元として持つ分散表現するベクトルでは、特定の次元を複数組み合わせ(あるいは単独で)用いて距離の演算を行うことで、例えば生物全体の中での類似度や、特定の地域に存在する物の中での類似度、特定の機能の有無に合致する範囲の中での類似度など多種多様な類似度計算を実行可能になると期待される。

また各次元の意味が明示的であるため、各語彙が保持する概念の内容と、それらを用いた演算の過程及び結果は、人間の分析者からみて理解可能である。そのため、概念と属性の定義や演算結果を人間が確認することで、時間をかければ誤りの発見や訂正による改善での精度の向上が可能である。

表1のような記述方式はベクトル形式であるため、利用者側の必要に応じて適宜新規の次元(属性)を追加して柔軟に変更することも容易である。他の辞書構築者による次元を後からいくらでも併記的に追加可能であるため、従来のオントロジック構築の場合に問題となる辞書構築者の視点への依存も、利用者が異なる構築者によって記述された属性を複数用いる形でアルゴリズムを記述すれば回避することが可能となる。そのため利用目的に応じたカスタマイズや様々に異なる視点からの分類や類似度の計算に適した構造であると考えられる。

これらの次元で示される概念の属性は従来のオントロジック等でも記述されてきたものである

が、本研究ではさらに、各概念に付随する一般的な常識的知識を命題的な表現形式を含めた属性を追加することで記述する方式を提案する。

命題とは分野によって若干異なるニュアンスを持つ語ではあるが、本稿では言語データの分析手法である内容分析やプロトコル分析で用いられている用例に基づき、少なくとも一つの主体と言動的単語を含めた文の意味の最小単位を指すこととする[14]。

例えば、「毒物」を概念として定義し記述する場合を考える。「毒物」の人間が道具として利用する機能は「殺す」ことであるが、これは動詞一つで記述可能である。これに合わせて、一般的な物語テキスト中で頻出する毒物を食品や飲み物に混ぜて用いるという使用方法に関する常識的知識も辞書中に含めることを考える。

命題的な表現形式を例えば、「主体:言動:対象 1:対象 2」の形式で記述するとすれば、「人が毒物を食品に入れる」は、「人:入れる:毒物:食品」のように表現できる。命題的表現は形式化された文であるため、このように1つの単語で表現できないような複合的な意味構造も表現することが可能である。

辞書中にこのような常識も含めることができれば、「食品を食べてから苦しんで死ぬ」という描写があった場合に、それが毒による殺人を示唆する描写である可能性を機械的に推論できるようになる。物語中では死、中でも特に殺人に関する描写はばかした表現で非明示的に提示される場合が多いが、登場人物の生死、特に殺人による死は物語展開上重要であり、その解釈への可能性が開かれることの意義は大きいと考えられる。

また逆に物語を機械的に生成する場合にも、殺害の手法として毒物を用いる場合に、「食物に入れる」という物語的なプロットを作成するためのデータとして用いることが可能である。

従来の Word2vec による分散表現においても、周辺単語との共起頻度に基づく各概念の一般的な用いられ方、すなわち各単語に付随する常識的な用法が暗黙的にパラメータとして内包されていると推測される。ただし、提案手法の場合には、「死」など忌避される概念であるがゆえに直接的には言及されにくいことがらや、あまりに自明であるがゆえにほとんど言及されずに省略されるような知識に関しても明示的に記述することが可能である。

一方で、本提案手法の最大の問題点として構築には莫大な時間と人手を要する点が挙げられる。「鳥」と「飛行機」の例で挙げたような概念弁別のための共通の属性は、個別に分析者が検討を行う必要がある。また、異なった領域に類似の属性が存在するか(例:「鳥」と「飛行機」の「飛行」は類似している)を判断していく必要もある。このため、近縁の概念だけでなく、一見離れた概念との間にも類似関係がないか一通り検討を行

う必要性が生じる。また人手によるすべての辞書類に共通する欠点ではあるが、編集者の主観的判断や知識が最終的な構築物に反映されるという問題も内包される。

ただしこれらの欠点に対し属性の追加と修正の容易性というメリットを活かすことで軽減できる可能性はある。単純なベクトル記述の構造上複数分析者による属性記述は容易である。そのため、他の分析者による属性記述を将来にわたって漸進的に追記していくことで、語彙数を増やし、概念記述の精度や網羅性を改善していくことは可能と考えられる。また、前述のように複数分析者による異なった属性記述を利用アルゴリズムの側で合わせて用いるよう工夫することで、分析者個人の視点の偏りは改善できると考えられる。

そこで本提案においては、頻出語彙の最低限の弁別に必要となる属性を検討し、物語テキストの処理に有用と考えられる常識的な知識の記述の可能性を中心に設計を行っていく。

4. 属性記述の設計

属性記述の設計にあたっては従来のオントロジー工学での比較的固定的な属性 (Type 属性) と状況依存的に変化する役割的属性 (Role 属性) の二種類を中心として他に文体的な属性 (丁寧語、粗野語など) や各概念カテゴリに固有の属性などを必要に応じて追加して記述した。現在 37 種類の属性を規定している(表 2)。Type 属性は上位下位概念や全体・部分関係、形状や色・模様指定などは他の概念を指定する形で記述した。上位概念や全体・部分関係などは複数の概念が該当する場合があるがそれらはリストとして列挙する形で記述している。これら他の概念を参照する記述は従来のオントロジー表現に対応させることが可能である。それ以外は数値等に置き換え可能な定形表現として記述している。

Role 属性としては、どのような役割を果たすかを該当単語が言動として用いられた場合の意味や、該当単語を含む頻出の命題の形式で、言動辞書[5]の分類に基づいて記述している。頻出命題の検討においては国立国語研究所によるコーパス中の係り受け関係の分析結果を公開している NINJAL-LWP for BCCWJ [15]を参照している。他に必ずしも用例としては頻出ではないが知識として一般的であると判断した命題も含めている。命題表現に関しても、複数が該当する場合は列挙する形で記述している。

対象単語は分類語彙表[8]と日本語 WordNet[7]所収の名詞中で、物語の登場人物を示す主体的な概念に合わせて、物語のオチで頻出する犯罪や事件、死と病気、災害や事故等、武器・薬品・毒物と身体の部位の領域を中心に分散表現化を約 1000 種類 8000 語の語彙に対して行った。

表2 現段階の属性リスト

形式	属性名	内容
他の概念の指定	上位概念	is-aに相当する概念
	被含有概念	part-ofに相当する概念
	含有概念	part-ofの逆に相当する概念
	構成材料	構成する物質のリスト
	世界設定	特定の場所時間の指定
	形状	形を種類で分類
	色	色の指定
	模様	模様を種類で分類
数値等	全体・部分	全体か部分かの指定
	内部・外部	内部,境界,外部等の指定
	順序	前後等の指定
	時間前後	過去,現在,未来の指定
	時間指定	絶対時間か相対時間か
	時間量	秒,分,時等時間の単位
	空間次元	1,2,3次元のいずれか
	空間量	平均的なサイズの指定
	三相状態	気体,液体,固体
	透明	透明の場合に指定
	発光・光沢	発光する場合に指定
	位置	位置関係を示す場合に指定
	向き	方位等向きを示す場合に指定
	抽象	抽象的に用いやすい場合に指定
	自然	自然物の場合に指定
	人工	人工物の場合に指定
	情報聴覚	音声を示す場合に指定
	情報視覚	視覚を示す場合に指定
	情報記号	記号を示す場合に指定
	情報その他	電気や熱等を示す場合に指定
	性	生物等の場合に性別の指定
	数	複数である場合の指定
	丁寧粗野	丁寧・粗野表現の場合に指定
立場・価値	敬意や価値を含む場合に指定	
人称	代名詞等での人称の指定	
命題	言動・現象	言動として用いうる場合に指定
	主体言動	対象が主体の場合の頻出言動
	客体言動	対象が客体の場合の頻出言動
	その他言動	対象がそれ以外での頻出言動

現段階では属性の値は他概念の指定であったり命題であったりと非数値的な内容も含んでいるが、これらを数値に変換して一般的なベクトルとする方式はいくつか考えられる。

まず他の語彙が属性として複数個記述されている個所に関しては、当該属性に記述されうる語彙の種類だけの次元を用意し、それぞれの語彙が属性として含まれるかどうかを数値で表すという方法が考えられる。この場合、例えば上位概念に「生物」が含まれるかのみを示す次元や、部分概念として「翼」が含まれるかのみを示す次元を作成することになる。最終的なベクトルの次元数は多くなるが、概念の構造は数値的に扱いやすくなると考えられる。

他にも、各概念にIDを振り、それを数値として属性に入れる方式や、特定の次元に関しては数値型以外にもテキスト型を許容するなど、さまざまなベクトルとしての実装方が想定されうる。

命題に関しても、命題の種類ごとに次元を割り当てる方式や、命題の箇所はテキスト形式にすることを許容する方式など種々考えられる。

例として表1と同じ概念を記述した場合のベクトルの主要部分を表3に示す。表3中での属性で他の概念記述は個々の次元にし、命題形式での表現はテキストで記述する場合を示す。また命題中でのAは当該単語を示す。

これらの属性を用いた物語テキストの処理としては例えば下記のようなものが考えられる。

1) 上位概念や含有概念を用いた対象の言い替えの特定/表現の多様化:「鉄砲」を「武器」への言い替えや、「引き金」で「鉄砲」の操作を示すなどの物語中で一般的に用いられる換喩・提喩表現に対応ができるようになる。

2) 被含有概念、世界設定等を用いた場所に付随する常識の補完:地の文での明示的な説明がない場合も「学校」に主体として「先生」や「学生」が存在することを前提として状況の推論が可能。生成においても、「会社」には「課長」や「部長」を登場させるなど場に適した登場人物の配置に活用が可能。

3) 物語生成での状況に応じた概念の抽出や入れ替え:例えば主人公の死の場面で、その場に登場しても不自然ではない凶器から別の物を選んで話を改変することや、死因を別の形に代える場合に適切な現象を選択することなどができるようになると考えられる。出来事を示す名詞等にその不幸度や幸福度などの指標も付加できれば、物語中での展開に合わせたハッピーエンドやバッドエンドを選択したり入れ替えたりすることなども可能になるであろう。

5. 今後の課題と展望

オントロジー工学の知見に基づき人手による分散表現辞書の設計を提案し、語彙の種類約 1000 種類、語数約 8000 語の試作辞書を作成した。今後は物語のパターンデータ中の各要素の必要条件と照らし合わせることで物語中のオチや凶器、

登場人物の役割などを交換した物語の自動生成を行う予定である。ただし、このためには試作辞書の拡張や、物語の分析及び生成での記法の有用性の確認が必要である。また、複数分析者による記述一致度の統計学的な検定も必要であると考えている。

表3 「鳥」と「飛行機」の関係の概念記述例

		生物	動物	鳥	鶏	機械	飛行機
上位概念 (is-a)	主体	1	1	1	1	0	1
	生物	1	1	1	1	0	0
	動物	0	1	1	1	0	0
	鳥類	0	0	1	1	0	0
	道具	0	0	0	0	1	1
	乗り物	0	0	0	0	0	1
含有概念 (part-of)	翼	0	0	1	1	0	1
	足	0	0	1	1	0	0
	胴体	0	1	1	1	0	1
	頭部	0	1	1	1	0	0
	ドア	0	0	0	0	0	1
座席	0	0	0	0	0	1	
主体言動	A:生きる, A:死ぬ, A:増殖, A:絶滅	A:動作, A:呼吸, A:食べる, A:飲む	A:飛ぶ, A:生む:卵	A:走る, A:not飛ぶ	A:壊れる, A:動作	A:行く, A:連れる, A:運ぶ, A:飛ぶ	
客体言動	親:生む:A			人:食べる:A, 人:貰う:A:卵	人:使用:A, 人:作る:A, 人:買う:A		

参考文献

- [1] Propp, V.: Morphology of the Folk Tale, University of Texas Press, (1968).
 [2] Greimas, A. J.: “Sémanique structurale: recherche de méthode, Larousse, (1966). 田島宏(訳), 鳥居正文(訳): 構造意味論—方法の探求, 紀伊國屋書店, (1988).
 [3] Genette, G.: Discours du recit in Figures III, Seuil, (1972). 花輪光(訳), 和泉涼一(訳), 物語のディスクール:方法論の試み, (1985).
 [4] Moretti, F.: Distant Reading, Verso (2013).
 [5] 村井 源: 言動分類による物語機能辞書の汎用化に向けて, 人文科学とコンピュータシンポジウム論文集, Vol. 2017, No. 2, pp. 225-230, (2017).
 [6] 村井 源: 主体語彙辞書を用いた物語テキスト中の主体推定システムに向けて, 人文科学とコンピュータシンポジウム論文集, Vol. 2017, No. 2, pp. 209-214, (2016).
 [7] Murai, H.: Prototype algorithm for estimating agents and behaviors in plot structures, Int. J. Computational Linguistics Research, Vol. 8, No. 3, pp. 132-143 (2017).
 [8] Miller, G. A.: WordNet: a lexical database for

- English, Communications of the ACM, Vol. 38, No. 11, pp. 39-41, (1995).
 [9] 国立国語研究所: 分類語彙表-増補改訂版, 大日本図書, (2004).
 [10] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., Dean, J.: Distributed representations of words and phrases and their compositionality, In Advances in neural information processing systems, pp. 3111-3119, (2013).
 [11] Schwartz, R., Reichart, R., Rappoport, A.: Symmetric Patterns and Coordinations: Fast and Enhanced Representations of Verbs and Adjectives, Proceedings of NAACL-HLT 2016, pp. 499-505, (2016).
 [12] Vindula, J., et al.: Deriving a representative vector for ontology classes with instance word vector embeddings, Innovative Computing Technology (INTECH), 2017 Seventh International Conference on. IEEE, pp. 79-84, (2017).
 [13] 市瀬龍太郎, 荒川直哉: 分散表象とオントロジーの関係, 第 29 回人工知能学会全国大会予稿集, 2I4-OS-17a-5, PDF, (2015).
 [14] 海保 博之, 原田 悦子: プロトコル分析入門—発話データから何を讀むか, 新曜社, (1993).
 [15] 国立国語研究所: NINJAL-LWP for BCCWJ, <http://nlb.ninjal.ac.jp/search/>, 2018/10/2 参照.