

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 6月 7日現在

機関番号：12601

研究種目：挑戦的萌芽研究

研究期間：2010～2012

課題番号：22652038

研究課題名（和文） 大規模語彙知識を融合した語彙概念構造体系の構築

研究課題名（英文） On Constructing a LCS System integrated with Large Lexical Knowledge

研究代表者

加藤 恒昭 (KATO TSUNEAKI)

東京大学・大学院総合文化研究科・教授

研究者番号：60334299

研究成果の概要（和文）：

複合動詞に着目し、言語処理で利用されている言語知識の分析を行った。形態素解析用言語知識に対して、動詞間の形態的・意味的關係を記述すると共に、動詞の複合と派生に関する關係に基づいた整理を行った。計算機可読辞書の意味記述について、複合動詞の意味においてその構成要素のどんな意味がどのような關係で現れているかを分析した。これらの分析を通じて、言語処理に用いられている複数の語彙知識を関連づけ、語彙概念構造理論の視点で整理する足掛かりを得た。

研究成果の概要（英文）：

Focusing on complex verbs, language resources, dictionaries, used in natural language processing were analyzed. Verbs in a dictionary used for morphological analysis were annotated by their morpho-semantic relations and structural relations, such as composition and deliberation. Meaning descriptions of complex verbs in a machine readable dictionary were analyzed from the perspective of their relations to the meanings of their components. These analyses provide some clues to organizing multiple language resources and providing those with a background of lexical conceptual structure theory.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	1,100,000	0	1,100,000
2011年度	700,000	210,000	910,000
2012年度	700,000	210,000	910,000
年度			
年度			
総計	2,500,000	420,000	2,920,000

研究分野：人文学

科研費の分科・細目：言語学・言語学

キーワード：語彙概念構造，語彙的オントロジー，語彙意味論，複合動詞

1. 研究開始当初の背景

語彙概念構造(LCS)や、生成語彙論(GL)の提案を受けて、語彙意味論研究は大きく進展

していたが、その主たる関心は、興味深い個別の言語現象の説明にあり、網羅的な理論体系や大規模な言語知識の構築への努力は希薄であった。一方、現代社会のテキスト情報

の氾濫に対処すべくコンピュータによる言語処理が盛んに行われており、そこでは広い範囲の現実のテキストが扱われ、機械学習や人手によって大規模な言語知識が構築されていた。しかし、言語処理の内容は表層的なレベルに留まっており、その知識も形式性の不足、粒度の不均一、ノイズの多さ等の問題を含んでいた。広い範囲の現実のテキストの処理に役立つような大規模な言語的知識に語彙意味論研究からの知見を反映させて形式生を高めること、そのような規模出の扱いを通じて語彙概念構造理論を再検討し、その理論的なカバー率をあげることが期待されていた。

2. 研究の目的

特に複合動詞に着目し、言語処理で用いられている大規模な言語的知識を語彙概念構造と関係づける。これを通じて、(1) 意味関係に基づく語義記述である大規模語義関係記述（語彙的オントロジ）と語彙概念構造の関係性を明らかにし、(2) 動詞の複合と派生に関する語彙概念構造構成方式と項構造（下位範疇）変化規則の形式化を行う。そして、これらの実施の過程で必要となる語彙概念構造理論の枠組みの拡張を行う。最終的には言語処理で用いられている言語知識と整合し、実用規模の言語処理に耐え、意味解釈の基礎までを提供する語彙概念構造体系へと展開することを目指し、それを通じてより高度な言語処理のための資源構築の基礎を確立する事を目的とする。

3. 研究の方法

(1) 実際の言語処理に利用されている語彙知識を、大規模語義関係記述で用いられている意味関係や、動詞の複合と派生に関する関係性に基づいて整理し、分析を行う。これらの関係性について、複数の語彙知識の間の整合性を調査する。これを通じて、言語処理に用いられている複数の語彙知識を関連づけ、語彙概念構造の視点で整理する足掛かりを得る。
(2) 複合動詞とその構成要素である動詞との意味的关系を明らかにするために、計算機可読辞書の語釈文を対象とした分析を行う。パターンマッチに基づく計算機分析と人手による深い分析の両方を行い、複合動詞の語義の形式的な表現と語義自動獲得の可能性を検討する。

4. 研究成果

(1) 形態素解析で用いられる辞書（IPADIC version 2.70）に含まれる動詞について、その形態・意味関係の記述を行った。異表記関

係にある項目をまとめて語（動詞）を定義し、それらの間に、可能、使役、受動等の文法的態、自他対応という語彙的態に関する関係を記述した。加えて、動詞を語構成の観点から分類し、その構成要素と関連づけた。加えて、これらの語彙的態に関する関係と別の言語資源（基本語データベース、計算機用日本語基本辞書 IPAL）との対応関係を調査した。

形態・意味関係の記述は以下からなる。

- ① PADIC の辞書項目は、表記、読み、活用型で特徴づけられる。形態素解析用辞書の場合、すべての辞書項目はそれぞれひとつの表記、読み、活用型を持つ。それに対し、語（動詞）を関係記述の基本要素として定義し、ある語とその語の表記となりうる表記を持つ辞書項目とを関係づける。辞書項目どうしの異表記関係もこの関係を通じて得られる。
- ② 動詞間を、可能関係、使役関係、能動受動関係、自他関係という形態・意味関係で関係づける。
- ③ 動詞を語構成の観点から分類し、派生動詞、複合動詞である場合は、その構成要素と関係づける。

これらの記述を、上位下位関係を持つ関係データベースとしてまとめている。

②の関係は、形態的特徴、意味的特徴に基づいて定義され、以下のように分類・定義される。

可能関係: 動詞 $v1$ と $v1$ の語幹+((ra)r)eru の形態をもつ動詞 $v2$ との関係。このうち、 $v2$ が「 $v1$ することができる」という意味を持つ対を意味関係ありとする。

使役関係 1: 動詞 $v1$ と $v1$ 語幹+(s)asu の形態をもつ動詞 $v2$ との関係。このうち、「A が B に $v2$ する」と「B が $v1$ する」という含意を持つ対を意味関係ありとする。

使役関係 2: 動詞 $v1$ と $v1$ の語幹+(s)aseru の形態をもつ動詞 $v2$ との関係。意味関係の有無の判断は使役関係 1 と同様。

使役関係 3: 形態的に使役関係 1,2 にない共に他動詞の $v1, v2$ で、使役関係 1 と同じ意味関係が成り立つもの。

能動受動関係 1: 動詞 $v1$ と $v1$ の語幹+(r)areru の形態をもつ動詞 $v2$ との関係。「A が B を $v1$ する」と「B が A に $v2$ する」もしくは「A が B に $v1$ する」と「B が A に $v2$ する」という含意を持つものを意味関係ありとする。

能動受動関係 2: 形態的に能動受動関係 1 にない $v1, v2$ で、「A が B を $v1$ する」と「B が A に $v2$ する」もしくは「A が B に $v1$ する」と「B が A から $v2$ する」という含

意を持つ関係。
 自他関係：「AがBをv1する」と「Bがv2する」という含意を持つ他動詞v1と自動詞v2との関係。

③では、語基（活用語尾を除く意味的な単位である形態素）ひとつからなる単純動詞、複数の語基からなる複合動詞、語基に接辞が伴った派生動詞への分類を行う。以下に示す10分類（実際にはこれ以外に古語等、特殊な分類が4種類ある）を設けた。「品詞」がその品詞である語に由来する派生動詞、「品詞+」がその品詞である語と動詞からなる複合動詞である。

単純：[歩く][壊す]等、ひとつの語基からなる単純動詞。

動詞+：[打ち壊す(打つ, 壊す)][売れ残る(売れる, 残る)]等、動詞連用形+動詞の複合動詞。

名詞：[大人びる(大人)][大人ぶる(大人)][春めく(春)]等、名詞からの派生動詞。

名詞+：[色褪せる(色, 褪せる)][役立つ(役, 立つ)]等、名詞+動詞の複合動詞。

形容：[暖める(暖かい)][楽しむ(楽しい)]等、形容詞からの派生動詞。

形容+：[近付く(近い, 付く)][若返る(若い, 返る)]等、形容詞+動詞の複合動詞。

形動：[静まる(静か)][馬鹿げる(馬鹿)]等、形容動詞からの派生動詞。

形動+：[主立つ(主, 立つ)][真面目くさる(真面目, くさる)]等、形容動詞+動詞の複合動詞。

接辞：[ざわめく(ざわ, めく)][べとつく(べと, つく)]等、単独で語にならない接辞や擬態語の一部に「めく」「つく」等が付属した派生動詞。

接辞+：[すっ飛ぶ(すっ, 飛ぶ)][蹴落とす(蹴, 落とす)]等、単独で語にならないような接辞や擬態語の一部と動詞からなる、動詞由来の派生動詞。

記述された関係や分類の数を表1,表2に示す。

表1 付与された形態・意味関係

	関係数	意味関係数
可能	1,382	1,164
使役1	161	132
使役2	100	69
使役3	4	-
能動受動1	65	61
能動受動2	9	-
自他	612	-

表2 動詞の語構成分類

分類	語彙数
単純	3,112
動詞+	2,904
名詞	523
名詞+	341
形容	90
形容+	23
形動	33
形動+	3
接辞	51
接辞+	215
その他	68

得られた関係について、他の言語資源との対応づけの可能性や一貫性の程度について検討を行った。

まず、人間向けの国語辞書を基に作成された計算機可読辞書であり、一般の辞書に見られるような語釈文（語義の記述）を備えている基本語データベースとの語彙の重なりを調べた。IPADICに較べて基本語データベースに含まれる語の数が少ないこともあり、形態素解析で得られる語に対して語義をふるることができる割合は十分に多いわけではないことが観察された。複合・派生動詞は、単純動詞以上に基本語データベースに含まれないIPADICの語が多い。対応づけられるのは、楽観的にみても50%以下(1,798/3,653)であった。複合動詞の語義は、一部については構成要素から構成的に得られるとはいえ、一般にそれらを構成要素の意味から推測する方法は確立されていない。これら複合・派生動詞の意味処理もしくは意味記述の充実が必要であることが明らかとなった。

次に、辞書項目数は多くないが、動詞について、豊富な形態・意味情報を記述している計算機用日本語基本辞書IPALと作成した関係記述との一貫性を調査した。それぞれの関係で傾向は異なるが、一致していない関係が多数見られることがわかった。例えば、自他関係の意味判断の基準においては、自発性の有無をどう捉えるか、使役関係や能動受動関係との境界をどこに引くか等のズレが影響して、記述に違いが出ている。

(2) 人間向けの国語辞書を基に作成された計算機可読辞書である基本語データベースに

含まれる複合動詞について、その語釈文を調べ、複合動詞とその構成要素である動詞との意味関係、複合動詞の意味がその構成要素の意味からどのように構成されているかを調べた、パターンマッチを用いた計算機による分析と人手による分析の両方を行った。

第一の分析では、語釈文に対して形態素解析を行い、その結果に以下のパターンを適用することで、見出し語の構成要素である v1 (前方の動詞)、v2 (後方の動詞) がどのような形式で語釈文に現れているかを調査した。

終止形式

「v[基本形].」

「v[連用形]たり v[連用形]たりする」

連用形式

「v[連用形],」「v[連用形]て」

否定形式

「v[未然形]ず」「v[未然形]ない。」

「v[未然形]ないで」

この分析によって、以下の例のように、構成要素の動詞が複合動詞の語義にどのようにかかわっているかが明らかになる。

売れ残る：商品が売れないで残る。

(v1 否定形式, v2 終止形式)

あきれ返る：完全にあきれる。

(v1 終止形式)

言い表わす：思っていることを言葉で表す。

言葉で表現する。

(v2 終止形式)

撃ち殺す：鉄砲などで撃って殺す。

(v1 連用形式, v2 終止形式)

書き添える：文章や絵の側に添えて書く。

(v1 終止形式, v2 連用形式)

結果を表3に示す。

表3 語釈文における構成要素の出現

	否定	連用	終止	連用+終止
v1	10	359	327	662
v2	0	51	280	326
v1+v2	-	-	-	95

2,286 語義中の 893 語義、約 1/3 の語釈文に v1 もしくは v2 が含まれていることがわかった。終止形式での出現はその語 (の語義) がその複合語の意味的主辞 (語義の主たる部分) であることを示唆し、連用形式での出現は、その語が、付帯状況や様態を表わし意味的主辞を修飾する役割を担っていることを示唆する。v1 はその両方の役割を果たすこと

があること、v2 の出現は v1 に比べて少なく、その役割は意味的主辞が大半であることがわかった。

第二の分析では、複合動詞の意味の中で、その構成要素のどんな意味が現れているのか、その構成要素の意味がどのように組み合わせられて (もしくは、組み合わせられないで) 構成されているのかを分類した。語釈文において、v1, v2 やその意味がどのように現れているかを以下の分類に基づいて分析した。

A 構成要素である語そのものが語釈文に現れている。

B1 構成要素の動詞本来の意味と関連する内容が語釈文に現れており、その内容は構成要素の主たる語義となっている。

B2 同上であるが、その内容は構成要素の比較的特殊な語義、比喩的な語義となっている。

C 構成要素と関連するが、動詞単体では持ち得ない構成要素独特の内容が語釈文に現れている。

D A~C いずれでもないが、語釈文の一部が構成要素の本来の動詞として意味と関連すると判断される。

E 以上のいずれにも当てはまらない。

v1, v2 の語義 (もしくはそれに関連する語義) がどのような関係を持って組み合わせられ、複合動詞の意味を構成しているかを以下の分類に基づいて分析した。

並列 「v1 したり v2 したりする」

「v1 し、かつ v2 する」

様態 12 「v1 しながら、v2 する」

「v1 して、v2 する」

「v1 するように v2 する」

手段 12 「v1 することで v2 する」

「v1 することを通じて、v2 する」

手段 21 「v2 することで v1 する」「v2 することを通じて、v1 する」

原因 12 「v1 した結果、(意図しなかったのに) v2 する」

様態 21 「v2 ったように v1 する」「v2 する程に v1 する」

埋込 12 「v1 する何かに、v2 する」

補文 12 「v1 することを (に) v2 する」「v1 することを、し v2 する」

格 12 「v1 を v2 する」

副詞 12 主たる意味は「v2 する」で、v1 はそれを強める、もしくは v1 が v2 の仕方を副詞的に修飾する。

副詞 21 主たる意味は「v1 する」で、v2 はそれを強める、v2 が v1 の仕方を副詞的に修飾する。

抽象化 語義の抽象性や比喩性が高く、v1, v2 と直接の関係が読み取れない。

これらの分析を基本語データベース」より抽出した複合動詞 1,314 語の 2,184 語義（語釈文）より，無作為に選択した 1,065 語義に対して実施した。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔学会発表〕（計 3 件）

- ① 加藤恒昭，語釈文を用いた複合動詞の特徴分類，言語処理学会第 17 回年次大会，2011 年 3 月 9 日，豊橋技術科学大学
- ② 加藤恒昭，形態素解析用辞書における動詞の形態・意味関係の記述，言語処理学会第 18 回年次大会，2012 年 3 月 16 日，広島市立大学
- ③ 加藤恒昭，形態素解析用辞書における動詞の形態・意味関係，情報処理学会第 207 回自然言語処理研究会，2012 年 7 月 27 日，北海道大学

〔図書〕（計 1 件）

- ① 東京大学言語情報科学専攻（編），東京大学出版，言語科学の世界へ，2011 年，pp214-228

6. 研究組織

(1) 研究代表者

加藤 恒昭 (KATO TSUNEAKI)
東京大学・大学院総合文化研究科・教授
研究者番号：60334299

(2) 研究分担者

伊藤 たかね (ITO TAKANE)
東京大学・大学院総合文化研究科・教授
研究者番号：10168354
林 良彦 (HAYASHI YOSHIHIKO)
大阪大学・大学院言語文化研究科・教授
研究者番号：80379156

(3) 連携研究者

()

研究者番号：