日本語和語動詞に関する語彙概念構造辞書構築の試み

加藤 恒昭　　畠山 真一　　坂本 浩　　伊藤 たかね
東京大学 大学院総合文化研究科言語情報科学専攻

1　はじめに

語彙概念構造（LCS）は，動詞の意味を意義素の組み合わせにより表現する体系であり [4]，統語処理と意味処理を橋渡しする言語知識として期待されている [7][1]．我々は言語学と言語処理の双方に有用な言語資源となることを目指して，日本語和語動詞の基本語約 1000 語について，LCS 辞書の網羅的な構築を試みている．1000 語という規模は，それが言語処理において有用なものとなるために必要と判断したためである．加えて，広範囲の動詞について検討することを通じて LCS の体系を整理することも目指している．和語動詞を選んだのは，漢語動詞については先行する取り組み [8] があることが大きな理由であるが，身近な動詞を取り上げることで，多義性等，LCS 辞書構築における問題を積極的に明らかにしていこうという意図がある．本稿ではこの LCS 辞書構築の方針と現状について報告する．

2　構築の方針

辞書構築の対象となる動詞は文献 [5] にあげられている基本語彙から助動詞と思われるものや漢語を除いた 971 語とした．

1000 語に関する LCS 構築という大規模な作業は，個人では困難であり，複数人への分割分担は内省判断の揺れが問題となるため，複数人による検証可能な調査に基づいて，実証的に作成していくことを試みている．このことを通じて，いわゆる動詞の典型性や従来より提案されている様々なテストの安定性についても考察できると期待している．検討の初期段階では言語学の背景を持たない多数の一般被験者に対してアンケート形式で文の適切性を判断させることを考えたが，設問に用いる文言の選択を含めて高い信頼性を得るための設問を構成することが困難であることから断念した．例えば「向かう」は「高速道路を北に向かう」とヲ格をとることは明らかであるが,「向かう」が「〜を」をとるかという質問に一般被験者が正しく回答できるかは疑問である．加えて「経路ヲ格」「起点ヲ格」等も一般的には馴染みのない言葉であろうがこれに代わる適当な表現も見あたらない．一方で個々に文を与えてその適切性を判断させれば信頼性は高いと思われるが，個々の文の判断に止まり一般性がないという批判に耐えられないと判断し，行わないこととした．結局，言語学の背景のある少数（4 名）の被験者に対するアンケート形式の文の適切性判断に基づく構築を行なっている．アンケートの設問はある形式の文が可能か，その解釈がどうなるかを問う選択式のものとなっており，回答に際しては判断の根拠とした文をあわせて記入させている．

LCS の体系についてもアンケートの設問をあまり多くしないこと，このような手法で再現性の高いレベルに留めること等の必要性から，細分化はおさえ，基本的なものを中心とした．結果としてアンケートではヲ格の振る舞いを中心として基本構文に着目し，主に有限性（telicity）と状態性（stativeness）について訊ねており，LCS もそれに応じたものとなっている．

3　LCS 体系

現在採用している LCS の体系は以下のようなものである．

働きかけと活動を表現するものとして，[?　ACT ON ?]，[?　ACT] をおいた．これらには意志性があるものとし，意志性のない働きかけと活動は [?　ACT- ON ?]，[?　ACT-] と分析する．

状態は，特徴もしくは位置を表現する [?　BE AT ?] と所有を表現する [?　BE WITH ?] のふたつと，それぞれの否定である [?　BE NOT AT ?] と [?　BE NOT WITH ?] とする．それぞれの 2 番目の変数については，項で指定される (z) か動詞の意味として定まっている (C) かが区別されるのが普通だが，後述のように今回はその区別の基準を作ることが困難であったため現状では分離せず，z/C と表現している．また，項として現れるものそのものではなくそれと関連した何かに関する場合，C(?) という表現を用いている．

状態変化は [BECOME **状態**] で表し，移動は [?　MOVE VIA ? TOWARD ?] とした．なお，着点や方向を表す TOWARD については項として現れない場合があるが，経路を表す VIA についてはすべて項として現れると考えて

いる．

因果性の表現には CAUSE を用い，[**働きかけ/活動** CAUSE **状態変化/ 移動**] の形で表現する（以下，A/B は A もしくは B を表す）．ただし，[[x ACT-] CAUSE **状態変化/移動**] の形式は，**状態変化/移動**そのものと区別されないため用いない．また，姿勢保持のように状態の変化でなく維持を引き起こす活動もありうるはずだが，今回は考えていない．

なお，これらの組み合わせによって得られるすべての LCS が現時点で得られている LCS 辞書に現れているわけではない．そのような欠落が本質的なものなのか否かについては今後の検討課題である．

4 アンケート設計

アンケートを構成する設問を図 1 に示す．このアンケートではある設問の回答に応じて次に回答する設問が指定される形となっており，すべての動詞についてすべての設問を回答するものでない．図中で選択肢の後に矢印で示されたものが次に回答すべき設問を示している．全 12 問のアンケートであるが，ある動詞についての設問数は最少で 3 問，最多で 7 問である．可能な回答パタンは 100 を超えるが，アンケート構成を簡潔にするために，不可能な（複数の設問の回答が矛盾する）パタンが含まれているので，実際に現れる分類はもっと少ない．図に示された設問はその内容を簡潔に示す表現であり，実際のアンケートにはもう少し丁寧な説明と例文が付与されている．

設問 1「テアル」は，動詞に意志性のあるの働きかけがあり，その結果として目的語項に変化があり，その結果状態が残存するかを訊ねている．つまり，LCS が [[x ACT ON ?] CAUSE [y MOVE ...]/[BECOME [y ...]]] の形となるかを調べている．この設問 1「テアル」は事前に行った予備検討で揺れが少ないと判断され，これらの判断を行ういわばショートカットして採用されている．

設問 2「ヲ格」，**設問 4「経路ヲ格」**，**設問 5「起点ヲ格」**は，ヲ格をとり得るかと，もしとる場合はその性質を調べている．経路ヲ格をとるものは [x MOVE VIA y ...] を，起点ヲ格をとるものは [BECOME [x NOT BE AT y]] を含む．

設問 3「タ後」は，結果状態の残存を訊ねているが，ここを通るのは設問 2「ヲ格」でヲ格をとらないと判断されたものであるので，この結果状態は主語項に関するものとなる．従って「はい」と回答されものには，[BECOME [x ...]] が含まれる．

設問 6「時間デ」は，有限性の有無，つまり変化や終了の時点があるかを訊ねている．ここを通るのは設問 1「テアル」で目的語項の結果状態が残存しないと判断されたものであるので,「完了解釈で可能」と回答されたものにはそれ以外の変化があるか移動の終了があることになる．

設問 7「スル」は設問 3「タ後」，設問 6「時間デ」で結果状態が残存しないとされたものについて，活動や働きかけの有無を訊ねている.「V する」が未来の事態と解

設問 1 「X が/を V してある」の形式が可能か？
(a) 可能 → 設問 9　(b) 不可能 → 設問 2

設問 2 「X を V する」の形式が可能か？
(a) 可能 → 設問 4　(b) 不可能 → 設問 3

設問 3 「V した後，今 V している」の形式の表現は当然のことを述べているか？
(a) はい → 設問 11　(b) いいえ → 設問 7
(c) その形式自体が不可能 → 設問 7

設問 4 移動の経路や起点を「X を」の形式で表現可能か？
(a) 可能 → 設問 5　(b) 不可能 → 設問 6

設問 5 「X を V する」の形式を「X から V する」の形式に変更可能か？
(a) はい → 設問 12　(b) いいえ → 設問 8

設問 6 「X 分で Y を V した」の形式が可能か？可能な場合，その解釈は動作や行為の完了かそれとも開始か？
(a) 完了解釈で可能 → 設問 11
(b) 開始解釈で可能→ 設問 7
(c) 不可能 → 設問 7

設問 7 「V する」と「V している」の形式について，
(a) 共に現在の事態を表す → 設問 12
(b) 前者は後者と異なり未来の事態を表す → 設問 12
(c) 「V する」が不可能 → 設問 12
(d) 「V している」が不可能 → 設問 12

設問 8 移動の着点を「X に」の形式で表現可能か？
(a) 可能 → 設問 12　(b) 不可能 → 設問 12
(c) そもそも移動がない → 設問 11

設問 9 V の表す行為や活動が対象の物理的移動，その所有権や情報の移動を含むか？
(a) 物理的移動を含む → 設問 10
(b) 所有権や情報の移動を含む → 設問 12
(c) 移動はない → 設問 11

設問 10 「X 分間 Y を V している」の形式が X 分間の間ずっと Y が物理的に移動し続けていると解釈可能か？
(a) 可能 → 設問 12　(b) 不可能 → 設問 12
(c) そもそも移動がない → 設問 11

設問 11 V の表す行為や活動による変化の結果状態はあらかじめ定められているか？
(a) はい → 設問 12　(b) いいえ → 設問 12
(c) そもそも変化がない → 設問 12

設問 12 「一生懸命」「いやいや」等の副詞を付与した文が可能か？
(a) 可能　(b) 不可能

図 1: **アンケートを構成する設問**

釈される場合，活動や働きかけや状態変化がある．

設問 8「着点ニ格」は，設問 5「起点ヲ格」で MOVE が含まれるとされたものについて，[MOVE x VIA y]，[MOVE x VIA y TOWARD z] の切り分けを行っている．

設問 9「物理的移動」，**設問 10「時間」**は，設問 1「テアル」で MOVE/BECOME が含まれるとされたものについて，MOVE/BECOME を切り分け，更に BECOME の場合，結果状態が特徴に関するものか物理的位置に関するものかを切り分けている．

設問 11「対象変化」は，BECOME を含むものについて，変化する状態である [x BE AT ?]/[x BE WITH ?] の? 部分が項で指定される (z) か動詞の意味として定まっている (C) かを区別する．「変化がない」とされたものは BECOME を含まない，もしくは対象以外の何かが変化すると考えられる．

設問 12「意志性」は，意志性の有無を訊ねている．ACT/ACT-，ACT ON/ACT- ON がそれぞれ区別されるが，[[x ACT-] CAUSE **状態変化/移動**] の形式は用いないため，[[x ACT] CAUSE **状態変化/移動**] と**状態変化/移動**とが区別される場合がある．

実際のアンケートにおいては，接触打撃動詞は容易に状態変化動詞へと振る舞いを変えるという観察 [2] から，設問 1「テアル」，設問 6「時間デ」では,「すべての目的語で」もしくは「特定の目的語で」可能であるかの判断をさせている．その効果については現在分析中である．

設問 1「テアル」，設問 3「タ後」，設問 6「時間デ」は一般には結果状態の残存あるいは有限性を訊ねるテストとして同じ振る舞いをするとされているものであるが，このアンケートでは，設問 1「テアル」は目的語項の状態変化のみを取り出せるらしいこと等に着目し，これら３つの質問を使い分けている [6]．これらのテストの振る舞いの差異については現在更に詳しい調査を行っている．また，この類のテストでよく用いられる「テイル」の解釈が結果状態と継続中のいずれとなるかは，予備検討で揺れが大きいことが明らかとなったため，用いていない．

5　構築の現状

前節で述べたアンケートを言語学の背景のある４名に実施した．結果の一致率を見ると，4 人もしくは 3 人の結果が一致する動詞が 971 語中 450 語であった．内容を見たところ，設問 11「対象変化」で訊ねている状態に関する部分が項で指定されるか動詞の意味として定まっているかの区別と，設問 9「物理的移動」，設問 10「時間」の BECOME の結果状態が特徴に関するものか物理的位置に関するものかの区別とに個人による一定した判断の違いがあることがわかった．また，これらの設問はその訊ね方が他と較べて抽象的，つまり文の適切性や解釈ではなく LCS の内容そのものを訊ねている感もあり，これらの特徴について分析していくことは今後の課題として，状態については [?　BE AT z/C] 等で表現し，z/C の区別やそれが特徴であるか場所であるかの区別はおかないこととした．このような処理の結果，4 人の判断が一致した動詞は 234 語，3 人の一致が 255 語（ここまでで 489 語），2 人が一致して残りの 2 人の判断が割れているものが 303 語，2 人ずつに割れたものが 87 語，その他が 98 語である．およそ半分の動詞について 3 人以上の一致が得られている．4 人もしくは 3 人の一致が得られた動詞の一部とその LCS を表 1 に示す．括弧内の数字はそこに分類された動詞の数を示す．ここで LCS は前節で述べた設問の位置付けから導いているが，それだけでは一意に定まらないため，実際に得られた動詞の意味等から内省的に判断して決定している．設問 6「時間デ」や設問 7「スル」の意味付け，設問 11「対象変化」で「(c) 変化がない」とされたものの意味付けが難しく，現時点で幾つかの動詞がそこに分類されている表 1 の 7,17,18,22 についても，LCS が決定できていない．また，15 については incremental theme として位置づけたが，これも自明ではない．

結果を考察したところ，以下についてはほぼ適切に分類され，問題なく LCS が割り当てられている．

- 位置変化，状態変化，作成の他動詞（14,16），ただし働きかけと思われるものが若干混ざっている．
- 位置変化，状態変化の自動詞（5,6,9,10,11），ただし一部（6）は，格パタンとのリンクが一定していない．
- 移動の自動詞（8,12,13）．
- 状態動詞（19,20,21），ただし 19 は「ている」との関係で分類されていない．

一方，以下の問題も明らかになった．

- 活動の自動詞 (1,2) については，ヲ格以外のテストを行わなかったため，働きかけがあると思われるもの（「取り組む」「つきあう」等）が含まれている．また，このクラスの典型的な動詞である「励む」「遊ぶ」等は意志性の判断（1,2）で割れている．
- 意志性のある働きかけの他動詞（3）については,「なぐる」「たたく」等が問題なく収まっているが,「なでる」「こする」は判断が分かれた，また,「間違える」「待つ」など目的語項への働きかけと捉えることに違和感のあるものもある．意志性のないもの（4）については，意志性の判断に疑問のあるものがある．

6　今後の課題

今後は，まず，判断が分かれた動詞，特に 2 人の判断が一致して残りの 2 人の結果が割れているものについて，

表 1: 得られた LCS 辞書の一部

No.	LCS	動詞の例
1	[x ACT]	取り組む，努める，つきあう，話しかける (11)
2	[x ACT-]	光る，流行る，栄える，ぶつかる，泣く (20)
3	[x ACT ON y]	鳴らす，はげます，禁じる，なぐる，間違える (55)
4	[x ACT- ON y]	つぶやく，犯す，知る，笑う，病む，恨む，喜ぶ (12)
5	[BECOME [x BE AT C/z]]	止まる，乾く，直る，こわれる，消える，濡れる (115)
6	[BECOME [x BE WITH z]]	気づく，わかる，見つかる，間に合う，掛かる (6)
7		似る (1)
8	[x MOVE VIA y]	漂う (1)
9	[x ACT] CAUSE [BECOME [x BE AT C/z]]	立ち止まる，座る，だまる，腰掛ける，しゃがむ (6)
10	[x ACT] CAUSE [BECOME [x BE WITH z]	応じる (1)
11	[x ACT] CAUSE [BECOME [x BE NOT AT y]]	去る，出る，抜ける，離れる，降りる，はいる (6)
12	[x ACT] CAUSE [x MOVE VIA y]	歩く，走る，越える，通る，泳ぐ，横切る，越す (10)
13	[x ACT] CAUSE [x MOVE VIA y TOWARD z]	戻る，向かう (2)
14	[x ACT ON y/C(y)] CAUSE [BECOME [y BE AT z/C]]	収める，取り出す，汲む，戻す，掛ける，盛る (173)
15	[x ACT ON y] CAUSE [x MOVE VIA y]	飲む，教わる，教える，歌う，読む，説く，見る (15)
16	[x ACT ON y] CAUSE [y MOVE TOWARD z]	動かす，巻く，運ぶ，流す (4)
17		脱ぐ，選ぶ，さえぎる，計る，飼う，用いる (8)
18		追い出す，高める，治す，履く (4)
19	[x BE AT z/C]	違う，似合う，まさる， 面す，沿う，そびえる (21)
20	[x BE WITH C(y)]	思う，きらう，好む，有する，恐れる，惜しむ (14)
21	[x BE WITH C]	におう，聞こえる，苦しむ (3)
22		望む (1)

多数派の判断が信頼できるかを調査する．これにより，現在 LCS が与えられていない動詞についての検討を進めていくと共に，同じ（タイプの）LCS を与えられる動詞の事例を増やし，その特徴，適切な LCS について考察していく．次に，今回行わなかったヲ格以外に関するテスト，今回は適切な結果が得られなかった BECOME の結果状態の分類について考えてゆき，LCS の詳細化を試みていく予定である．また，多義性の問題 [3] をどのような枠組みで扱うか，アンケート調査において多義性のある動詞をどう扱うかも検討していく必要がある．

7 おわりに

日本語和語動詞の基本語約 1000 語についての LCS 辞書構築の方法論と現状について述べた．規模の確保と実証性の重視のために複数の被験者によるアンケート結果を分析することによって構築を進めている．現在，約半数の動詞について揺れの少ない判断がなされており，それらについて LCS の割当を行っている．

本研究は 21 世紀 COE「心とことば－進化認知科学的展開」の一環として行われている．構築された LCS 辞書等，本研究の成果は本 COE ホームページ[1]等を通じて公開していく．

[1] http://ecs.c.u-tokyo.ac.jp/home/home.html

参考文献

[1] 降幡健太郎他.「語彙概念構造を用いた機能動詞結合の言い換え」言語処理学会第 10 回年次大会，pp. 504 - 507，2004.

[2] 畠山真一他.「日本語動詞の LCS 推定に関して－他動詞を中心に－」情報処理学会研究会報告，2005-NL-165，pp. 1 - 8，2005.

[3] 畠山真一他.「語彙概念構造と多義性」言語処理学会第 11 回年次大会，P1-12，2005.

[4] 影山太郎.「動詞意味論－言語と認知の接点」くろしお出版，1996.

[5] 国立国語研究所「日本語教育のための基本語彙調査」秀英出版，1984.

[6] 坂本浩他.「動詞識別テストとしてのテアル構文の有効性について」言語処理学会第 11 回年次大会，P6-5，2005.

[7] 竹内孔一他.「語彙概念構造を利用した複合名詞内の係り関係の解析」情報処理学会論文誌，Vol. 43 No. 5，pp. 1446 - 1455，2002.

[8] 竹内孔一他.「語彙概念構造による動詞辞書の作成」言語処理学会第 10 回年次大会，pp. 576 - 579，2004.