

語彙概念構造による意味役割の形式化と複数役割の割り当て

松林 優一郎

宮尾 祐介

相澤 彰子 §

国立情報学研究所コンテンツ科学研究系, § 東京大学情報理工学系研究科

{y-matsu, yusuke, aizawa}@nii.ac.jp

1 はじめに

近年登場した大規模な意味情報付き言語資源は、意味解析の実験主義的な研究手法を飛躍的に加速させてきた [7]。特に、PropBank[4] や FrameNet[9] といった意味役割付きのコーパスは、項構造解析や意味役割割り当てなど意味解析の基盤技術構築に欠かすことのできない資源となっている。しかし、これらの資源が研究の基礎的資源として広く受け入れられている一方で、その意味論の理論的側面には疑問が残る。例えば、意味役割割り当ての正解コーパスとして最も利用されている PropBank では、意味役割ラベルの意味的な不整合さが報告された [10]。VerbNet[5]、FrameNet の意味役割はより矛盾の少ないラベルセットであるものの、その定義は形式的な方法で与えられておらず、役割の意味や性質は不明瞭なままである。

もう一つの問題は、既存の資源が述語のそれぞれの項に対して一つの役割しか割り当てないという仮定を置くことである。例として、表1に動詞 throw を含む英文中の意味役割を示した。既存の資源では、主語に当たる John に対して、Arg0、Agent といった一つの役割のみを割り当てているが、実際には、我々はこの文から「John は、投げる動作の動作主であり、かつ、ボールの移動における始点である」という二つの情報を自然に読み取ることが出来る。このように、一般に、項は動詞が表現する事態において複数の役割を帯びている場合が稀ではない。しかし、既存の資源では、本来語彙が内包している項の役割のいくつかを、単純に無視しているか、一つのラベルに圧縮している。

この問題は既存の資源におけるラベル定義の不明瞭さの問題とも関連している。後に示すように、VerbNet では複数の異なるタイプの意味的性質が一つのラベルに詰め込まれている。これが意味する事は、既存資源の意味役割は必ずしも本質的に独立な性質を持ったラベルではないという事である。我々は、それぞれの項が持つ役割に明確な定義を与え、また、複数の意味役割を正確にタグ付けする事によって、項構造の理論的な枠組みだけでなく、これまで汲み取れなかった二次の役割による意味処理を用いる応用技術にも貢献できると期待している。

本論文では、語彙概念構造 (LCS) の理論を拡張することで、動詞に関する実用的な語彙意味論の枠組み

表 1: 既存の資源における単一役割の割り当ての例

Sentence	[John]	threw	[a ball]	[from the window].
Action	Agent		Patient	
Movement	Source		Theme	Source/Path
PropBank	Arg0		Arg1	Arg2
VerbNet	Agent		Theme	Source
FrameNet	Agent		Theme	Source

を示す。同時に、LCS の基本述語を通して意味役割を再定義することにより、意味ラベルの定義を形式化し、明確化する。また、拡張された LCS の枠組みにより、複数役割の割り当てが自然に実現できることを示す。論文の前半部では、我々の拡張した LCS についての理論的側面を説明し、後半部では、我々の定義する意味役割と既存資源の意味役割との比較を与える。

2 関連研究

言語学の領域では、項に複数の意味属性を割り当てる試みがいくつか成されてきた。Gruber[2] は具体的な例を示しながら、項が一つの役割しか持たない事に必然性がない事を主張した。Rozwadowska[8] は change, cause, sentient という三つの特徴によって意味役割の意味素性を分解する方法を提案し、従来の主題役割をこれら三つの特徴の組み合わせによって再定義した。このような特徴分解の方法は役割の定義を整理することを可能にしたが、一般的に、特徴の組み合わせの数は、言語における実質的な組み合わせの数よりも多く [6]、従って、特徴分解の方法は、特徴の可能な組み合わせを理論的に制限することが要求された。

Culicover ら [1] は意味役割を行為の役割と知覚の役割という二つのグループに分割し、複数の役割は必ずそれぞれのグループから一つずつ選ばれらると説明した。Jackendoff[3] は語彙の意味を複数の基本述語を用いて表現する LCS の理論を提案し、その意味構造の中で、行為層と主題層という二つの層を用いて Culicover らの二つのセットと類似するモデルを立てた。これらのモデルでは、行為と変化に関わる役割が区別されることで、複数の意味役割の割り当てが可能となった。

我々は、Jackendoff[3] の LCS 理論を基本として、同一の出来事における複数の活動や変化を記述出来るように拡張する。具体的な拡張内容は、次節に述べる。

$$\left[\text{cause}(\text{affect}(i,j), \text{go}(j, \left[\begin{array}{l} \text{away_from}(\text{locate}(\text{in}(i))) \\ \text{from}(\text{locate}(\text{at}(k))) \\ \text{toward}(\text{locate}(\text{at}(l))) \end{array} \right])) \right]$$

図 1: 動詞 *throw* の語彙概念構造

表 2: 主要な基本述語とその意味機能

述語名	意味機能
$\text{state}(x, y)$	第一項が、第二項で指定される状態にある。
$\text{cause}(x, y)$	第一項で指定される行為が、第二項で指定される変化を引き起こす。
$\text{act}(x)$	第一項が、自身に作用する活動をする。
$\text{affect}(x, y)$	第一項が、第二項に作用を与える。
$\text{react}(x, y)$	第一項が、第二項の影響を受けて、自身に作用する活動をする。
$\text{go}(x, y)$	第一項が、第二項で指定される経路に沿って変化する。
$\text{away_from}(x)$	変化の始点
$\text{from}(x)$	変化の始点の方向
$\text{via}(x)$	変化の経由点
$\text{toward}(x)$	変化の着点の方向
$\text{to}(x)$	変化の着点
$\text{along}(x)$	変化の経路となる連続的な対称

3 LCS による意味役割の形式化と複数役割の表現

語彙概念構造 (LCS) [3] は、語彙が表現する出来事や状態を、汎化された意味の構造として記述するための体系である。語彙の意味は、複数の基本述語によって分解され、その組み合わせ構造で表現される。例えば、図 1 は英語の動詞 *throw* の概念構造を表しているが、この中には *cause*, *affect*, *go*, *away_from*, *toward*, *locate*, *in*, *at* といった述語が含まれている。LCS 内の各項は、語彙が取る文中の項によって埋められる。表 1 の例文中の *throw* の場合、図 1 の *i, j, k* が、それぞれ *John, a ball, the window* によって埋められる。

図 1 と表 2 で示すように、LCS の基本述語は、その組み合わせによって、動詞が表現する、行為-変化-状態という連鎖の部分または全部を表現出来るように設計されている。従って、用いられる各基本述語は、この連鎖における形式的な意味的機能を表す。例えば、図 1 の *throw* の例では、*cause*, *affect*, *go*, *away_from*, *from*, *toward* の複合により、行為 (*i* が *j* に作用) - 変化-状態 (*j* が、*i* → *k* → *l* と移動) の全体が表現されている。

我々の拡張した LCS 体系を図 2 に与える。基本的に、我々の LCS は Jackendoff[3] の理論に基づく。主な相違点は、(1) *cause*, *act*, *affect*, *react*, *go* の可能な組み合わせを明示的に制限している点、(2) 述語 *comb* の導入によって、一つの出来事における複数の活動や変化を記述出来るようにした点、(3) Jackendoff の理論における *GO*, *STAY*, *INCH* の三つの変化に関する述語を一つの述語 *go* に統一した点、(4) 変化を伴う殆どの出来事が、場所の変化として比喩的に表現される点である。本論文の話題において特に鍵となる拡張は (2) であるが、これは Jackendoff の提案した述語 *EXCH*

$$\text{LCS} = \left[\text{EVENT} + \left[\text{comb EVENT} * \right] \right] \quad \text{STATE} = \left\{ \begin{array}{l} \text{locate}(\text{PLACE}) \\ \text{orient}(\text{PLACE}) \\ \text{extent}(\text{PLACE}) \\ \text{connect}(\text{arg}) \end{array} \right\}$$

$$\text{EVENT} = \left\{ \begin{array}{l} \text{state}(\text{arg}, \text{STATE}) \\ \text{go}(\text{arg}, \text{PATH}) \\ \text{cause}(\text{act}(\text{arg1}), \text{go}(\text{arg1}, \text{PATH})) \\ \text{cause}(\text{affect}(\text{arg1}, \text{arg2}), \text{go}(\text{arg2}, \text{PATH})) \\ \text{cause}(\text{react}(\text{arg1}, \text{arg2}), \text{go}(\text{arg1}, \text{PATH})) \\ \text{manner}(\text{constant})? \\ \text{mean}(\text{constant})? \\ \text{instrument}(\text{constant})? \\ \text{purpose}(\text{EVENT} | \text{arg}) * \end{array} \right\}$$

$$\text{PLACE} = \left\{ \begin{array}{l} \text{in}(\text{arg}) \\ \text{on}(\text{arg}) \\ \text{cover}(\text{arg}) \\ \text{t}(\text{arg}) \\ \text{inscribed}(\text{arg}) \\ \text{beside}(\text{arg}) \\ \text{around}(\text{arg}) \\ \text{near}(\text{arg}) \\ \text{inside}(\text{arg}) \\ \text{at}(\text{arg}) \end{array} \right\} \quad \text{PATH} = \left\{ \begin{array}{l} \text{away_from}(\text{STATE})? \\ \text{from}(\text{STATE})? \\ \text{via}(\text{STATE})? \\ \text{toward}(\text{STATE})? \\ \text{to}(\text{STATE})? \\ \text{along}(\text{arg})? \end{array} \right\}$$

図 2: 拡張された LCS の記述体系¹。演算子 +, *, ? は一般的な正規表現の文法に従う。|, {} は選択を表す。

表 3: 提案する LCS 体系における意味役割のリスト

役割	定義
Protagonist	事態の視点となる要素
Theme	状態や状態の変化が言及される要素
State	ある要素の現在の状態
Actor	自身の状態を変える/保つ行為を行う要素
Effector	他の要素の状態を変える/保つ行為を行う要素
Patient	他の要素によって自身の状態を変えられる/保たれる要素
Stimulus	他の要素の行為のきっかけとなる要素
Source	ある変化の始点
Source_dir	ある変化の始点の方向
Middle	ある変化における経由点
Goal	ある変化の着点
Goal_dir	ある変化の着点の方向
Route	ある変化についての線状の経路

の自然な拡張である。ただし、我々の場合、*comb* は *EXCH* のような反対向きの移動を表すだけでなく、主題となる出来事に併発して起こる任意の出来事を記述する述語として定義される。また、我々は動作主が複数の行為を行なう場合に、主題となる出来事を複数記述することも許容している(「交換する」など)。このような拡張により、複雑な出来事において、併発する動作や変化を取り零すことなく、各項の役割を正確に記述することが可能となる。

図 2 から見て取れるように、LCS の基本述語は入れ子の構造となっている。特に、表 2 に示す上位部分(図 2 の *EVENT*, *PATH* に相当)の述語は、行為-変化-状態の連鎖に直接関係する機能を持ち、従って、これらの述語のスロットを埋める項は、この連鎖の範疇において、形式的な意味機能が与えられる。この意味機能は、従来解析に用いられてきた主題役割と大まかに対応が取れる。そこで、比較のために、LCS が表現する意味機能に対して、表 3 のような独自の役割ラ

¹簡単のため、各述語の細かな属性値は省略してある。

表 4: 基本述語の項と定義した意味役割との対応

述語	第一項	第二項
state(x, y)	Theme	State
act(x)	Actor	-
affect(x, y)	Effector	Patient
react(x, y)	Actor	Stimulus
go(x, y)	Theme	PATH
away_from(x)	Source	-
from(x)	Source_dir	-
via(x)	Middle	-
toward(x)	Goal_dir	-
to(x)	Goal	-
along(x)	Route	-

表 5: 60 の語彙における各役割の出現数

役割	単一	複数	増加率 (%)
Theme	21	108	414
State	1	1	0
Actor	12	13	8.3
Effector	73	92	26
Patient	77	79	2.5
Stimulus	0	0	0
Source	11	44	300
Source_dir	4	4	0
Middle	1	8	700
Goal	42	81	93
Goal_dir	2	3	50
Route	2	2	0
w/o Theme	225	327	45
Total	246	435	77

ベルを定義する²³。表 4 には、基本述語の項と役割ラベルの対応表を記した。各ラベルは基本述語の機能に基づいて、明確な役割が定義され、また、述語の各スロットにおよそ一対一対応することが分かる。唯一述語の項に一致しない役割として、Protagonist があるが、これは出来事の主題となる式の最初の項に割り当てられるもので、主題の式とその他の付随的出来事を区別する意味素性である。定義される役割の数は 13 であり、この数は VerbNet の主題役割と比べて約半分だが、これについての詳しい考察は次節で述べる。

LCS のような述語分解の方法は、文中の一つの項が、概念構造中の複数のスロットを埋めることを可能にする。例えば、図 1 の例では、動詞 throw の項 i は affect の第一項と away_from の内側の項として、二度出現している。また、我々の拡張により、述語 comb で導入される副次的出来事における役割も記述できるようになる。我々は、60 個の日本語動詞（事態性名詞を含む）について LCS の辞書を作成し、複数役割を認めることによって、どのくらいの役割が新たに付与されたかを検証した（表 5）。60 の動詞は京都大学テキストコーパス [11] 中に最頻出（ただし、最初の 100

²³我々の定義するラベルは従来の意味役割ラベルと一部名前が似通うが、定義を異にする。純粋に表に書かれた定義に基づき、それ以外の意味的属性を仮定しない。

²⁴ただし、LCS の核心となる情報は既に構造的に直接含まれており、定義するラベルは比較の目的の為であることに注意されたい。

表 6: 各言語資源における意味役割の数

資源名	フレーム依存	役割数
LCS	no	13
VerbNet (v3.1)	no	30
FrameNet (r1.4)	yes	8884

語を除く)のものから選び、記述したフレームの数は 97 であった。単一役割の場合の数は、対象の構造中に最初に出てきた役割を数えることで算出した。結果として、複数役割を認めた場合、単一役割の場合に対して 1.77 倍の役割が付与されることが分かった。増加の主な理由としては、作用動詞における Patient と Theme が意味機能に基づいて明示的に分離される事で Theme が増加した事、移動系の動詞における Effector の一部が Source や Goal としてタグ付けされた事、また、comb 関数による副次的出来事の追加で、Effector 等が増えた事などが見受けられた。

4 既存リソースとの比較

表 6 に、我々の LCS、VerbNet、FrameNet で定義されている役割の数を示す。役割の数はそれぞれの資源が表現する意味属性の数や、その属性の一般性に関する。VerbNet は我々の LCS に対して約二倍の役割を持ち、FrameNet ではその数は数百倍にのぼる。

現在、意味役割の定義には二つの作法がある。一つは、語彙に対してそれぞれ固有の役割を定義するフレーム依存の方法であり、もう一方は、全ての動詞間で共通に使われる一般性を持った役割を定義する方法である。FrameNet の役割数が比較的莫大なのは、FrameNet が役割の一般性よりも、個々の語彙における詳細化された項の意味を尊重して意味役割を定義するからであり、したがって、FrameNet は語彙意味論や意味役割の形式的定義を目的とする我々の資源とは異なるベクトルを持った資源と言える。

一方、LCS と VerbNet はどちらもフレームに独立な意味役割の定義を取るが、VerbNet の方が役割数が多い。しかしこれは LCS の表現力が VerbNet に劣ることを直接意味するものではない。我々は人手による入念な分析により、VerbNet の意味役割と、それに対応する LCS における表現の対応表を作成した（表 7）。表 7 に示す対応表は、タグ付けコーパスに基づいた分析ではないため、完全なものではないが、ここで少なくとも言える事は、VerbNet の意味役割が、LCS では別の層へと分離される、三つの異なる情報を混合しているという事である。それらは、(1) 行為一変化一状態の連鎖における意味機能、(2) 選択制限、(3) 項同士の構造的な関係である。結果として、VerbNet には (1) における機能が重複する役割がいくつも見受けられる。例えば、Destination, Recipient, Beneficiary の三つは変化を伴う動作において終了時の状態 (LCS では Goal) という同じ属性を持っている。この三つの役割の違いは、より詳細化された状態の派生型 (被

表 7: VerbNet の意味役割と対応する LCS における表現⁴

VerbNet の役割 (使用数)	LCS における対応する表現
Actor (9), Actor1 (9), Actor2 (9)	構造中の対称形をした二つの式におけるそれぞれの Actor もしくは Effector
Agent (212)	(Actor \vee Effector) \wedge Protagonist
Asset (6)	Theme \wedge その対象に関する変化の Source 部分が (locate(in()) \wedge Protagonist) \wedge selPref(valuable entity)
Bene ciary (9)	Goal \wedge locate(in()) \wedge selPref(animate \vee organization) \wedge \neg (Actor \vee Effector) \wedge 変化した対象が selPref(bene cial) を満たす
Cause (21)	(Effector \wedge selPref(\neg animate \wedge \neg organization)) \vee Stimulus \vee 周辺的役割
Destination (32)	Goal
Experiencer (24)	react() に関する Actor
Instrument (25)	(Effector \wedge selPref(\neg animate \wedge \neg organization)) \vee 周辺的役割
Location (45)	(Theme \vee PATH 内の役割 \vee 周辺的役割) \wedge selPref(location)
Material (6)	Theme \vee (Source \wedge 変化の Goal が locate(t()) \wedge Goal が selPref(physical_object) を満たす)
Patient (59), Patient 1(11)	Patient \vee Theme
Patient2 (11)	(Source \vee Goal) \wedge connect()
Predicate (23)	Theme \vee (Goal \wedge locate(t())) \vee 周辺的役割
Product (7)	Theme \vee (Goal \wedge locate(t()) \wedge selPref(physical_object))
Proposition (11)	Theme
Recipient (33)	Goal \wedge locate(in()) \wedge selPref(animate \vee organization)
Source (34)	Source
Theme (162)	Theme
Theme1 (13), Theme2 (13)	二つの変化する対象 (Theme) がある \vee state 式の対象 (Theme) と状態 (State)
Topic (18)	Theme \wedge selPref(knowledge \vee information)

所有) を表現したり、選択制限や、項の間の構造的関係を表す部分による。これらの役割を分離する事は、統語論との関係を含めながら項の振る舞いを分析する点においては有意義だと考えられるが、意味表現の形式化という点では、述語分解の方法で与えられる明示的な意味属性の方がより有益である。LCS では、選択制限の情報は動詞の項 (図 1 の i, j, k, l) の属性として直接記述できる。また、LCS は、それ自体が語彙内の項の構造情報を表すため、複雑な項同士の構造的関係を意味ラベルに埋め込む必要がない。結果として、LCS の意味役割は、その一般性を高く保ちながら、役割の数を少数に保つことが出来る上、枠組み全体では、VerbNet の意味役割と同等の情報を表現可能である。

5 おわりに

本論文では、現在の意味役割付与の枠組みにおける二つの問題に取り組んだ。一つは意味役割の定義の形式化であり、もうひとつは複数の意味役割の割り当ての問題である。我々は、語彙意味論における述語分解の手法がこの問題の解決に適することに着眼し、Jackendoff による LCS の理論を、複数の併発する出来事を表現できるよう、自然な形で拡張した。60 の日本語動詞における LCS の辞書の統計では、単一の役割を割り当てられる場合に比べて、我々の方法で 77% の新たな役割が割り当てられることを示した。

VerbNet や FrameNet などの主要な資源と比較して、我々の LCS における意味役割の定義は、その意味機能が基本述語の示す行為-変化-状態の連鎖にお

る役割に限定されているため、より形式的で明瞭である。また、LCS 自体は、選択制限や項の構造的な関係を役割ラベルとは別の層の情報として記述出来ることも示した。一方で、LCS は語彙の意味に依存しない周辺的な意味役割については規定しないため、これらの役割の厳密な定義には、さらなる議論を必要とする。

我々は、このような語彙分解的な解析が、意味解析分野の基礎研究及び応用研究を加速させると期待しており、今後は、我々の LCS 辞書を用いた意味タグ付きコーパスの設計を予定している。

参考文献

- [1] P.W. Culicover and W.K. Wilkins. *Locality in linguistic theory*. Academic Press, 1984.
- [2] Je rey S Gruber. *Studies in lexical relations*. PhD thesis, MIT, 1965.
- [3] Ray Jackendo . *Semantic Structures*. The MIT Press, 1990.
- [4] Paul Kingsbury and Martha Palmer. From Treebank to Prop-Bank. In *Proceedings of LREC-2002*, pp. 1989–1993, 2002.
- [5] Karin Kipper, Hoa Trang Dang, and Martha Palmer. Class-based construction of a verb lexicon. In *Proceedings of AAAI-2000*, pp. 691–696, 2000.
- [6] Beth Levin and Malka Rappaport Hovav. *Argument realization*. Cambridge University Press, 2005.
- [7] Llus Màrquez, Xavier Carreras, Kenneth C. Litkowski, and Suzanne Stevenson. Semantic role labeling: an introduction to the special issue. *Computational linguistics*, Vol. 34, No. 2, pp. 145–159, 2008.
- [8] B. Rozwadowska. Thematic restrictions on derived nominals. In W Wilkins, editor, *Syntax and Semantics*, Vol. 21, pp. 147–165. Academic Press, 1988.
- [9] J. Ruppenhofer, M. Ellsworth, M.R.L. Petruck, C.R. Johnson, and J. Sche czyk. FrameNet II: Extended Theory and Practice. *Berkeley FrameNet Release*, Vol. 1, , 2006.
- [10] Szu-ting Yi, Edward Loper, and Martha Palmer. Can semantic roles generalize across genres? In *Proceedings of HLT-NAACL 2007*, pp. 548–555, 2007.
- [11] 黒橋貞夫, 長尾真. 京都大学テキストコーパス・プロジェクト. 人工知能学会全国大会論文集, pp. 58–61, 1997.

⁴フレームの定義に五回以上出現した VerbNet の役割を解析対象とした。ここに示す対応表は部分的で整合性の高い対応のみを抜き出して示している。完全な対応表現は、個々の語彙 (や動詞クラス) の意味に高く依存することに注意されたい。