

Dynamic Lexicon 構築を目指した<Nスル>の考察 — 言語処理と認知言語学の接点

古牧 久典†

田中 穂積†

† 東京工業大学大学院 情報理工学研究科 計算工学専攻
{furumaki, tanaka}@cl.cs.titech.ac.jp

Building a Dynamic Lexicon:

Japanese Verbal Nouns Viewed from NLP and Cognitive Linguistics

This paper analyzes *N-suru* forms in Japanese, i.e. verbal nouns based on a Dynamic Lexicon. To understand language, NLP needs knowledge like corpora, but the data presented in the research conducted so far is hardly sufficient. However we cannot describe all the entries because the *N-suru* form is very productive. This often makes it virtually impossible to decide on a principled basis whether or not to include a given instance in the database. Therefore we propose a Dynamic Lexicon from a cognitive semantic point of view, based on Langacker[14,15,16,17,18]'s Dynamic Usage Based Model.

(keywords: *N-suru*, *internal control*, *a dynamic usage-based model*, *analyzability*, *construction*, *construal*)

1. はじめに

本稿は、Dynamic Lexical Resources 構築を目指し、日本語の<Nスル>型を例に、認知言語学的な立場から考察を行ったものである。

自然言語処理の分野では、例えば形態素解析として、単語あるいは形態素による分割を行うが、いかなるアルゴリズムを利用した場合にも、単語ないし形態素の単位が何であるかは常に問題とされる。仮に、形態素を「意味を担う最少の言語要素」と定義したときにも、「最少」が何であるかは、自然言語の意味を捉える上で重要である。例えば「処理する」は、「処理する」全体で動詞の形態素とも、部分の名詞「処理」も形態素である可能性も考えられる。「処理する」を{処理する}のように分割しない場合と{処理}{する}のように分割する場合の決定は、どう辞書に収録されているかで決まるか、あるいは、その統計的な手法で決定される。しかし、そうした手法では、「処理する」あるいは「処理」のどちらが「最少」の意味を持つかは、はっきりしない。こうした問題に対して、従来の辞書(例えば、分類語彙表、日本語語彙体系、IPAL電子辞書、あるいは紙の辞書)は、「処理する」「処理」の個別の単語を対象としており、「最少」の単位に関する記述や、動詞と名詞におけるそれぞれの意味の違いは記述されていない。

「語」の単位が何であるかを問う試みは、理論言語学でも盛んであり、例えば松本[21]の研究などに代表される。また、影山[8,9,10,11,12]などの形態論の研究においても、言語の最少の意味から、意味を計算によって導き出すことを目指しているが、解決には至っていない。<Nスル>という形式は、「処理する」、「研究する」など、様々な名詞に「する」をつけて動詞を作り出すことが可能であることから、非常に生産的であり、これらの表現を人間の認知能力では、状況に依存し、動的にその意味を解釈している。そこで、本稿では、<Nスル>における、名詞Nと「する」、

それぞれの意味を考えながら、<Nスル>全体の意味との関係性を明らかにすることで、動的な辞書構築を目指すためには何が必要かを考える。

自然言語の動的な辞書構築を目指すことは、自然言語処理にとっても有益であるばかりか、不可欠である。

2. 日本語<Nスル>

日本語の<Nスル>形式は、名詞(N)に「する」という形式を付加することで動詞として機能する用法である。伝統的には「サ変名詞」(「サ変動詞」)、あるいはMartin[20]などによれば、「動名詞」(Verbal Noun) (「動名詞による動詞」)とも呼ばれる。例えば、「処理する」、「研究する」など、様々な名詞に「する」を付けて様々な動詞の意味を作り出すことが出来るので、非常に生産性が高い。こうした用法を、国語学では、森田は9つに分類している。これらをさらに、山口[38]では、「和語名詞+スル」、「漢語名詞+スル」、「外来語+スル」、「副詞+スル」という形式に単純化している。

本稿の立場では、意味・機能による分類を試みる。意味による分類としては、影山[8,9,10,11,12]、杉岡[25]、伊藤[6]らによる形態論の研究だけでなく、廣濱[5]や宮島[22]らによる機能的な観点からの研究も注目すべきである。廣濱や宮島は、<Nスル>の動名詞は、「人間活動のわくに関すること」や「人間活動の中の精神および行為」がそうした名詞になりやすいとしている。こうした性質は、多くの動名詞に該当するものであるが、どこまで一般性があるかは、厳密に検討する必要がある。形態論の研究では、影山[9,10,11]は、<Nスル>の名詞Nの意味が<Nスル>の意味を決めるとし、そのNは「内在的コントロール」¹を

¹ 影山[9]では、例えば「爆発する」では、「爆発」に「内在的コントロール」があるとするが、「内在的コントロール」

行い、自他の交替といった文法現象にまで影響しているという指摘をしている。田窪[33]の主張では、「化」に接続する名詞は中身によって、「意味論的、語用論的要因」で自他の決定を行っているという指摘は影山の主張と類似点がある。〈Nスル〉に含まれる名詞Nの「内在的コントロール」、あるいは「意味論的、語用論的要因」とは何であるかを探ることが、本論の〈Nスル〉の分析となる。

3. 従来の辞書における〈Nスル〉

これまで〈Nスル〉という形式の扱いは、紙の辞書であれ、電子化されたものであれ、単語を網羅的に収録することを目指した辞書。基本と思われる単語のみ選択して収録した辞書に二分される。以下にその収録語数を示す。

辞書 ⁱⁱ	〈Nスル〉形式
「広辞苑 (第4版)」1992.	299 語
「大辞林 (第1版)」1992.	370 語
Oxford Basic Japanese English Dictionary. 1986.	284 語
「計算機用日本語基本動詞辞書 I P A L」1990.	66 語
「現代雑誌九十種の用語用字」 国立国語研究所 1998.	121 語
「日本語基本動詞用法辞典」 大修館書店. 1989.	171 語
「教育基本語彙の基本的研究」 2001.	171 語
「日本語語彙体系 NTT コミュニケーション科学研究所」1997.	3080 語
「分類語彙表 増補版」 国立国語研究所 1996.	7288 語
「EDR 電子化辞書 (ver. 2.0)」 1996.	19190 語

収録語数が最大の EDR (電子化辞書) は網羅的な辞書を目指している。一方、200~300 語前後のものは基本的な語の収録を目指している。しかし、例えば、EDR に未収録であり、日本語語彙体系に収録されているものをカウントしてみると 127 語ある。127 語が新語である可能性もあるが、例えば「実況する」、「新発売する」、「熟する」のように、必ずしも新語のみというわけではないことがわかる。また、基本的な単語を心理学的な実験により選択して辞書に収録しようという試み[13]は興味深い、収録の判断は非常に難しい。例えば、同じ収録語数の「基本動詞用法辞典」と「教育基本語彙の基本的研究」を比較してみると、共通に含まれる〈Nスル〉は 2 語しかみつからない。以上のことから、従来の辞書では、収録するかは恣意的に決められないと言わざるを得ない。

4. 認知言語学の立場

収録標準の恣意性に対して、認知言語学の立場 (特に認知文法の立場) [14,15,16,17,18,39,40] では、A Dynamic Usage-Based Model (動的使用依拠モデル) ⁱⁱⁱ という代案を提供[18]している。この節ではその考え方を示す。認知文法では、自然言語は「慣習的な言語的単位の構造化された目録」(a structured inventory of conventional linguistic units) と捉える。人間の認知能力や認知過程についての前提は最小限のものとし、様々な言語現象に対して、統合的な扱いを可能にする特徴がある。具体的には、認知能力・過程は認知的な構造体(cognitive assembly)によるものであり、「定着」(entrenchment)、「抽象化」(abstraction)、「比較」(compare)、「合成」(composition)、「連合」(associate)によって、様々な現象を説明する。こうした合成という考え方には、「構文」(construction)と呼ばれるような複合単位の意味が個別単位の意味に還元できない言語のゲシュタルト(gestalt)性を認める認知言語学の立場では強調する[3,14,18]。Langacker[18]はこうした構文スキーマ(constructional schema)を Figure 1^{iv}のよう示し、音形による[A]と[B]の合成[C]では、意味による(A')と(B')の合成(C')の意味から、(C'')の意味に発展することを強調する。

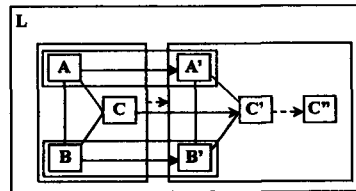


Figure1: Langacker [18] (pp. 24)

例えば「研究する」という表現は、「研究」の音形[A]、「する」の音形[B]による、合成的な音形「研究する」という[C]があるとする、と「研究」の意味<研究>の(A)、「する」の意味<する>の(B)による、合成(C)の<研究を行う>の行為の意味を経由し、(比喩的な意味など)発展した意味(C'')の<研究活動を定期的に行う>などの意味となる。

Langacker は、Figure2 の図式で、部分概念[A]と部分概念[B]の分析可能性の程度を示しており、複合単位(C)の意味は、突然新たに作り上げられるものではなく、操作によって作られ、その後、慣習的に定着していくと考える。

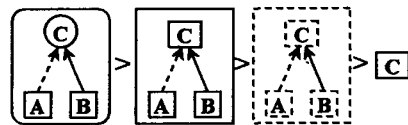


Figure2: Langacker [18] (pp. 38)

このことは、分析可能性 (analyzability) の問題に大きく

ⁱ ル」が何であるかについては定かではない。

ⁱⁱ ここにリストした辞書は、作成目的は様々であり、その規準は一致しない。

ⁱⁱⁱ 認知文法の考え方は、コネクショニズムの考え方とある意味ではつながり得るものだが、その連続性と非連続性については、[18,39,40]などを参照。

^{iv} 詳細は Langacker [18] (pp. 24)を参照。

関係する。分析可能性とは、より大きな記号構造の中に記号的構成要素の存在を認識して、全体の意味に貢献するかどうかの可能性である。認知文法では、この分析可能性に程度が存在し、部分で構成されるものと全体で構成されるもの間に連続性を認めることに主張の基盤がある。

最終的に、日本語<Nスル>における、名詞の意味と「する」の意味が、合成されたものは単に部分の意味だけでは還元できない意味があることを認め、動的な辞書を構築することによって解決しようとする本論の立場は、認知言語学の主張に基づくものであるといえる。

5. <Nスル>の動的辞書

認知文法の観点からすれば、分析可能性には程度が存在し、それが連続的であることを前節では述べた。ここでは、日本語<Nスル>の動的辞書構築にあたり、具体的に、どのように適用可能かをみる。

まず、複合単位が部分に還元不可能な語彙を扱う。この複合単位が部分に還元不可能なものを構文的用法^vと呼ぶ。例えば、「欲する」が構文的用法である。「欲する」の「欲」は個体として名詞一語としては存在せず、分割できない。ただし、この「欲」の個体に意味がないわけではなく、全体の意味を作り出すための重要な要素の一部である。こうした構文的な用法の例を挙げれば、「応ずる」、「関する」、「屈する」、「察する」、「準ずる」、「接する」、「属する」、「対する」などがある。

以下では、合成的な場合で、部分の意味が残るものを扱う。ただし、部分の意味が合成されるものとはいっても、実際には、単純な合成ではなく、複雑な操作を経て合成的な意味（構文的な意味）になっていることを強調したい。例えば、「研究する」全体の意味は、「研究」の中でも、その「一部」の意味が関与する。その関与の仕方が、影山[9,10,11]の「内在的なコントロール」と呼ぶものである。「内在的コントロール」は、「研究」と「する」が合成されたときに働くものである。そこで、まず「する」という形式の意味を解明する。そして、「研究する」の全体の意味は、「研究」における認知文法の「捉え方」(construal)と関係しており、その「捉え方」を解明することが、「内在的コントロール」の中身を解明することになる。その「捉え方」こそが「動的」となることを述べたい。

「する」のプロトタイプ的な用法は、影山[8](pp.256)でも述べられているように、「活動・動作をあらわす」と言える。「する」は、本動詞として単独でも、迂言的表現でも、共通する概念があると考えられる。そのスキーマ図式はFigure3^{vi}のように描くことができる。日本語の「する」は、活動、動作が行われる中で、2つの参与者によって、働きかける側(○)、働きかけられる側(□)の両者が存在する。

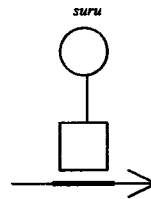


Figure3: [suru] in Japanese

ここで紹介したいのが「捉え方」(construal)である。認知言語学では、一般認知能力と連続的な言語の「捉え方」の認知能力が存在すると考える。例えば、「上り坂」と「下り坂」は、現実世界の同じ「坂」の焦点の当て方、つまり「捉え方」によって言語化(カテゴリー化)の違いが生じている。認知文法では、同じ事が「文法」にも生じていると考える。後に述べるように、「する」という表現では、働きかける側が焦点化される場合と、働きかけられる側が焦点化される場合がある。

影山[9,10,11]では、<Nスル>形式のNには、「内在的コントロール」が働いているとし、その「内在的コントロール」が何であるかは説明されていないが、認知文法では、prominence(顕著化)の能力が働いていると考える。prominenceとは文法における焦点化の仕方である。このprominenceこそが、本稿の「動的」と呼ぶものである。つまり、<Nスル>形式の名詞の焦点化は、以下の2つに分かれることになる。1つは、働きかける側と、もう1つは、働きかけられる側、それぞれが存在し、その仕方は「動的」なものである。

宮島[22]^{vii}は、日本語<Nスル>形式を構文的な機能による「転換」と呼ぶことが出来る例を挙げている。ここではPinker[23]やLevin[19]などに従い、「交替」(alternation)と呼ぶことにする。ここでは2つの交替を例に挙げる。一つ目は、<AガBヲNスル>⇄<BガNスル>の交替である。例えば「完成」の名詞による「太郎がプラモデルを完成する」という表現は、「プラモデルが完成する」と交替可能である。つまり、働きかけられる側のBは、ガ格で表現することが可能であり、名詞Nは働きかけられる側に焦点をあたえる意味である。宮島が挙げる例から抽出すれば、「移転」、「汚染」、「回復」、「確定」、「合併」、「継続」、「結合」などがそれに相当する。

一方、もう一つの交替は、働きかける側が焦点となるものであり、<AガBヲNスル>⇄<AガNスル>の交替である。例えば、「愛」、「挨拶」、「案内」、「調査」、「注意」、「注目」、「代表」などである。

さらには、働きかける側、働きかけられる側、名詞自体は両方可能なものも存在する。両者の交替を許すものとしては、「注射」や「影響」などがあり、どちらも可能であるため、文脈などによって、動的な辞書構築が行われなくては、解決できない。

以上、認知言語学的な視点から、動的な辞書構築に必要な

^v 「構文」の定義は、Goldberg[3]の「構文」とLangacker[14,15,16,17,18]の「構文」とでは定義は異なる。ここでは、元々の「構文」の発想の出発点であるFillmoreの基盤に沿ったLangackerの定義に基づく。

^{vi} 図の詳細はLangacker[14](vol.1.9.1.4),[16](pp.99,129)や山梨[40](pp.37)などを参照。

^{vii} ただし、宮島[22]の列挙した例は疑問が多い。

となるような、意味による分類を紹介した。

6. 結語

動的な辞書を目指し、理論言語学の研究成果を、言語処理に利用しようと試みる動向は、竹内[29,30,31,32]や高橋[26,27,28]においてもみられる。竹内では、TLCS と呼ばれる LCS(Lexical Conceptual Structure)により複合名詞の分類分けがされており、本研究との関係は興味深い。竹内は、LCS を利用し、複合語の係り関係を解析し、99.3%の精度が得られたと報告している。高橋は、GL(Generative Lexicon)[24]を利用し、「動的な語彙モデル」を目指した辞書構築を提案している。このモデルはまさに本論と並行しており、興味深い。GL のモデルで言えば、本稿で主張している「働きかける側」と「働きかけられる側」は、Qualia Structure のそれぞれ、Agentive Role と Telic Role と密接に関係する。今後、こうした研究も含め、今回提案した Dynamic Lexicon がより普及することを期待する。

参考文献

- [1] Barlow, Michael. 2000. "Usage, Blends and Grammar". In Michael Barlow and Suzanne Kemmer (eds.) *Usage Based Models of Language*. 315-346. CSLI.
- [2] Fong, Sandiway. 2002. "What Does It Mean To Be A Shelf?: Semantic Bleaching and WordNet". *Proceeding of the 1st International Global WordNet Conference*. India.
- [3] Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.
- [4] Grimshaw, Jane. 1990. *Argument Structure*. MIT Press.
- [5] 廣濱文雄. 1965. 複合してサ変動詞を作る漢字語の意味分類(2). 『山辺道』 12. 62-69.
- [6] 伊藤たかね・杉岡洋子. 2002. 『語の仕組みと語形成』(英語学モノグラフシリーズ 16) 研究社.
- [7] Jackendoff, Ray. 1990. *Semantic Structure*. MIT Press.
- [8] 影山太郎. 1993. 『文法と語形成』 ひつじ書房.
- [9] 影山太郎. 1996. 『動詞意味論』 ころしお出版.
- [10] 影山太郎ほか. 1997a. 『語形成と概念構造』 研究社.
- [11] Kagayama, Taro. 1997b. "Denominal Verbs and Relative Saliency in Lexical Conceptual Structure." In Kagayama ed., *Verb Semantics and Syntactic Structure*, 45-96. Kuroshio.
- [12] 影山太郎. 1999. 『形態論と意味』 ころしお出版.
- [13] 金杉友子ほか. 2002. 単語新密度に基づく基本的語彙の選定. 『情報処理学会研究報告』2002-NL-150, 119-124.
- [14] Langacker, Ronald. 1987, 1991. *Foundations of Cognitive Grammar Vol. 1, 2*. Stanford Press.
- [15] Langacker, Ronald. 1988. "A Usage-Based Model" In Brygida Rudzka-Ostyn(ed.), *Topics in Cognitive Linguistics*. Stanford University Press.
- [16] Langacker, Ronald. 1990. *Concept, Image, and Symbol: The Basis of Grammar*. Mouton de Gruyter.
- [17] Langacker, Ronald. 1999. *Grammar and Conceptualization*. Mouton de Gruyter.
- [18] Langacker, Ronald. 2000. "A Dynamic Usage-Based Model". In Michael Barlow and Suzanne Kemmer (ed.) *Usage Based Models of Language*. 1-65 CSLI.
- [19] Levin, Beth. 1993. *English Verb classes and alternations*. The University of Chicago Press.
- [20] Martin, Samuel. 1975. *A Reference Grammar of Japanese*. Yale University.
- [21] Matsumoto, Yo. 1996. *Complex Predicatives in Japanese: A Syntactic and Semantic Study of the Notion 'Word'*. CSLI.
- [22] 宮島達夫. 1994. 『語彙論研究』 ひつじ書房.
- [23] Pinker, Steven. 1989. *Learnability and Cognition*. MIT Press.
- [24] Pustejovsky, James. 1995. *The Generative Lexicon*. MIT Press.
- [25] 杉岡洋子・小林秀樹. 2001. 「名詞+動詞型の複合語」. 影山太郎編『日英対照 動詞の意味と構文』242-268 大修館書店.
- [26] 高橋幸. 2002a. 構成要素の辞書情報を基にした複合名詞の意味形成制約. 『言語処理学会第 8 回年次大会』, pp. 160-163.
- [27] 高橋幸. 2002b. 生成語彙論の枠組みに基づいた動的語彙モデルの構築. 『情報処理学会研究報告』, 2002-NL-150, pp125-131.
- [28] 高橋幸. 2002c. 言語処理における生成語彙論の辞書設定. 日本英語学会第 20 回大会発表資料.
- [29] Takeuchi, Koichi et al. 2001. "Categorising Deverbal Nouns Based on Lexical Conceptual Structure for Analysing Japanese Compounds". *Proceedings of the 2001 IEEE Systems, Man, and Cybernetics Conference*, pp. 904-909.
- [30] 竹内孔一ほか. 2002a. 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, 1446-1455.
- [31] 竹内孔一ほか. 2002b. 語彙概念構造を利用した助詞「に」に関する複合名詞の分析. 『情報処理学会研究報告』, 2002-NL-150, pp133-140.
- [32] 竹内孔一. 2002c. 語彙概念構造に基づく複合名詞の解析. 日本英語学会第 20 回大会発表資料.
- [33] 田窪行則. 1986. 一化. 『日本語学』, vol.5. no.3.
- [34] Talmy, Leonard. 2000. *Toward A Cognitive Semantics*. 2 vols. MIT Press.
- [35] Tanaka, Hideki. 1994. "Verbal Case Frame Acquisition from a Bilingual Corpus". *COLING 94*, 2:727-731.
- [36] 田中英輝. 1995. 『英日機械翻訳システムにおける基本動詞の曖昧性解消のに関する研究』, Doctoral Dissertation, 九州大学.
- [37] Tuggy, David. 1993. "Ambiguity, Polysemy, and Vagueness". *Cognitive Linguistics*. 4: 273-290.
- [38] 山口明徳. 2001. 「サ行変格活用」『日本語文法大辞典』明示書院.
- [39] 山梨正明. 1995. 『認知文法論』 ひつじ書房.
- [40] 山梨正明. 2000. 『認知言語学原理』 ころしお出版.