

局所化した単一化文法とその表現

佐野 洋 福本文代

(財) 新世代コンピュータ技術開発機構

1 はじめに

我々は、自然言語の分析に中核を占める文法の形式化を行い、機械処理に資する要請のため、文法の記述を進めている。そのアプローチは論理文法を基本に置いており、単一化の枠組の中で文法の規則を組み立てている。他方、工学的な有用性を裏付けるため、上述の形式化と文法の実証や評価を行う自然言語実験支援環境の開発も行っている。後者のシステムについては既に報告した[8]。本稿では、論理文法に基礎を置く文法の記述枠組とその形式について述べる。

2 文法枠組

一般的に、単一化文法は文脈自由文法の一般化とみることができる。但し、その統語上の範疇は一つの表象に限定されないという特徴をもつ。特定表象に限定されない統語範疇は素性とその値の対の束により一般に表現される。素性と値の束は、統語上の特徴のみならず、形態特徴や意味解釈の枠組を単一の表現で記述できる。したがって、構成規則、いわゆる文法規則を簡略化するだけでなく、適切な構造を作るために素性値の一致を単一化により行えるのである。

単一化は二つの素性値構造の矛盾のない併合である。多くの句構造ベースの文法理論により提唱される各々の原則は素性の集合の一致を指定する。その方向性が、非明示的になるものの、単一化は二つの統語範疇の依存関係を表現する方法でもある。

素性値の束による範疇の表現は、統語上の特徴だけでなく形態や意味の情報を含む複合構造である。よって、単一化枠組は意味解釈や語用論的解釈までも融合できる形で統合化して扱える潜在能力を持つ。

3 局所化した単一化文法

単一化に基づく文法記述では、宣言的な規則の記述が可能だが、文構造の構成上の振るまいが基本的に語彙に記載の情報に強く依存する。単一化という操作のみで、統語対象を扱う際に、語彙依存の文法規則はしばしば記述に制限が加わる。例えば、連体句の規則を考えた場合、空格、あるいは特殊名詞句を被修飾句に持つ場合などは、比較的明らかになるものの、意味的な関係を構

成する連体句は統語構造を確定できない。「大根を切る音」に対して統語上の構成を規定することは難しい。また、係助辞でその機能が特定される句については、多くの場合、格役割を仮定する。しかしながら「三番線は発車します」という駅構内の極めて身近なアナウンスの分析に労を要する。

上述の問題点の解消のため統語分析による単一化を制限し構造上の機能的適用を取り入れた。機能的適用は、直観的には、局所的に単一化を一時棚上げした形で構造を汲み上げ、意味情報や語用論的情報が得られた時点で、再び単一化を行おうとするものである。

3.1 表現形式

文法規則は比較的基本的な句構造規則を用いる。

$$C_0 \Rightarrow C_1, C_2$$

文法の記述形式は次の要素から成る。以下には、要素と働き、並びにデータ表現を示す。

表1 構成要素

範疇名	統語範疇名	定項
SYN	発話に依存しない素性値	素性値の束
X	量化される対象	変項
REL	発話に依存する素性値	素性値の束
F	機能適用の対象となる要素	差分リスト
PRO	伝搬素性値	差分リスト

構文上の形式的な記述には、一般性への考慮から DCG の記述を仮定している。この表現形式は、範疇名が終端記号あるいは、非終端記号に対応し、他の要素は引数位置に展開される。基本構成要素はこれらのうち、SYN, REL, F の三つ組により表される。これにより、語彙の記述も行えるほか、句の記述はもとより文の記述も三つ組により表現される。

SYN と REL は共に素性とその値の束で表現され、当該範疇を特徴づけている。その違いは量化対象を含むかどうかによる。例えば、「話(す)」が用言であって、音韻変化の特徴として強変化の値を持ち、その読みが「はな(す)」であるといった情報は SYN 部分にある。

他方、REL 部分には「話す」の関数機能として、引数を二つとり、第一引数が主体であるとか第二引数が内容であるといった情報が含まれる。この情報は、発話の状態によって特徴づけられるものである。F には、一

般には下位範疇化の情報が含まれ「が」格と「を」格でマークされる構成要素を下位範疇化の対象とすることが記述される。

日本語は統括要素の範囲内では、統括される幾つかの要素は語順が比較的自由である。これらを扱うためにFの内容は要素の集合として扱われる。基本的に格割当てに使用され、機能適用の対象となっている。

この要素が差分リストのデータ表現を持つ背景は、いわゆるゼロ代名詞を明示的に表現することである。「汗を流すことはよい」のように主格が義務的にゼロ化されるような場合を考慮している。

PROは、束縛素性などの値を伝搬させるもので、再帰代名詞などは必要に応じて文内で束縛される。また、いわゆる代名詞などは文外束縛対象として分析結果に明示的に現れ、構文の分析以降にその束縛が委ねられる。

3.2 機能的適用の構造

図1は連体句の構成の規則を示す。構造にも依存するものの連体句と主名詞の間に意味的關係が成立すれば、統語上は適格とされる。本文法枠組では、空格や主名詞の性質に依存する統語上の曖昧性とせずに機能的適用として扱う。

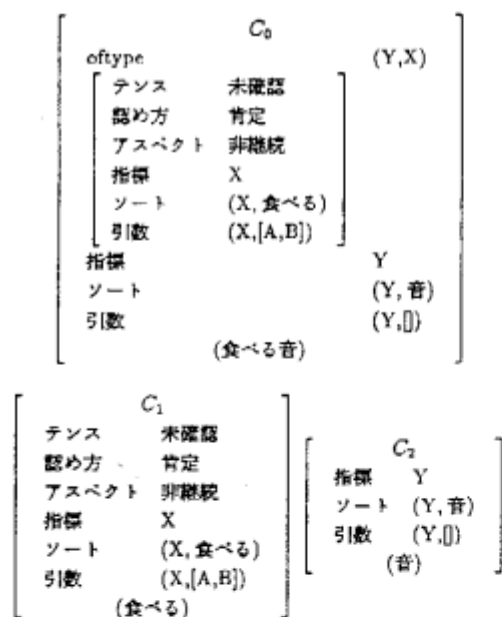


図1. 連体句での機能適用の例

oftypeは機能的適用が行われたことを示す表象である。他に数種存在するがいずれも統語上の現象に対応している[9]。基本的に修飾句については機能的適用を施す。統語的曖昧性を保持した構造をもって統語分析を行っており、この構造は統語解析後、意味処理あるいは語用論的处理の際に解消されるべきものである。形態素解析あるいは統語解析といった意味での区分はないが、単一化という枠組の中でモジュール性を維持している。

3.3 文法記述の実際

我々は、この枠組によって日本語文法を記述している。現在の版の文法規模は分析対象が単文の範囲に限られている。この形式化によれば、文の表現形式も語彙の表現形式も特定の句の形式にも区分がない。従って、いわゆる形態素解析と構文解析を同時に行うことである。語の構成も句の構成も同様に扱うことが可能で、必要に応じて単一化が成される。統語上その構成を保留する際には、局所的に単一化が制限される。一種の制約が課された形で処理が進む。

基本的に単一化に基礎を置く文法は、処理の方向性を仮定しない。確かに解析と生成の非対称性はあるものの、扱う範囲を限ることで文の解析と生成を同一の文法で行える。

4 まとめ

単一化の機能を局所的に制限することで、宣言的記述能力を損なうことなく機能的適用を実現した。構文の解析の後、構造上の非確定部分は意味情報や語用論的情報を使い、同じ単一化の枠組により解消されることが期待できる。逆に論理表現に近い枠組から、統語上の曖昧性を内在する構造を含む記述に変換することで文の生成が可能である。

参考文献

- [1] Fernando C.N. Pereira, David H.D. Warren. (1980), *Definite Clause Grammars for Language Analysis—A Survey of Formalism and a Comparison with Augmented Transition Networks*, Artificial Intelligence 13 (1980), 231-278.
- [2] 郡司隆男 (1986), 句構造文法 情報処理, Vol.27, 1986, No.8 pp868-pp875
- [3] Kevin Knight. (1989), *Unification: A Multidisciplinary Survey* ACM Computing Surveys, Vol.21, No.1, March 1989.
- [4] Marc Moens, Jo Calder, Ewan Klein, Mike Reape, Henk Zeevat. (1989), *Expressing generalizations in unification-based grammar formalisms*, Fourth Conference of the European Chapter of the ACL Proceedings, 1989.
- [5] 松本裕治, 清野正樹, 田中徳穂. (1983), BUPトランスレータ, 電総研彙報, Vol.47, No.8 (1983).
- [6] Michael Barlow. (1988), *Unification and Agreement*, Report No.CSLI-88-120, February 1988.
- [7] Robert C.Moor. (1989), *Unification-Based Semantic Interpretation* Proc. 27th ACL, Canada, 1989.
- [8] 佐野洋, 田中裕一, 佐々木博司, 小野寺浩, 木下聡. (1990), 論理文法の枠組における言語分析支援環境, 自然言語処理研究会報告, 90-NL-77, 情報処理学会研究報告, 1990.
- [9] 佐野洋, 堀本文化, 田中裕一. (1990), 説明的記述に基礎を置く文法 -SFTB-, ICOTテクニカルレポート, TM-885, 1990.
- [10] 首藤公昭, 吉村賢治 (1986), 日本語の構造とその解析, 情報処理, Vol.27, 1986, No.8 pp947-pp954