

語意依存文法における語意項目の記述について

3G-8

河田 浩二

牧野 武則

東邦大学理学部情報科学科

1 はじめに

依存文法 [1] は、導きだす依存構造が、格文法で表現した意味構造に類似していること、非終端記号のように人工の記号を用いないで語と語の依存関係を表すなど利点が多い。実用のシステムでも採用され [2]、その有用性が検証されつつある。GPSG など単一化文法 [3] では語彙項目に語彙に係わる構文情報を集め、単一化操作で簡素な解析を実現している。

提案する語彙依存文法 [4] は、言語に共通の係り受けのモデルを設定し、語彙に依存する依存情報を語彙項目に集中し、さらに依存構造をネットワークで表現することで、簡素で柔軟な解析、生成を実現することを目的としている。ここでは、語彙項目の記述と日本語文の解析について紹介する。

2 係り受けモデル

依存文法は、語と語の依存関係から文の構造が決める。この時、依存関係にある2つの語の一方を主要素、もう一方を従要素と呼び、例えば、2つの語が動詞と名詞であれば、主要素が動詞になり、従要素が名詞となる。この主要素と従要素の関係は係り受け情報として与えられる (図1)。例えば、形容詞の場合、通常主要素として名詞をとり、その係り受け情報は連体修飾となる。この係り受け情報は各語の情報、つまり語彙項目によって決定される。

3 語意項目

語から得られる情報を語彙項目といい、大きく2つに分けられる。一つはその語が主要素である場

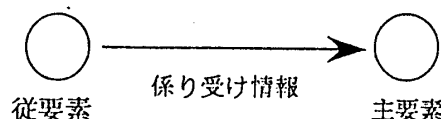


図1: 依存関係と係り受け情報

合、どのような従要素が係り受け可能であるかの情報 (主要素情報)、もう一方は従要素である場合、どのような主要素と係り受け可能であるかの情報 (従要素情報) である。主要素情報には、品詞と格情報があり、それらに対応する従要素情報として、修飾と格がある。

3.1 品詞

すべての語が持っている情報である。今回の日本語の実験では、名詞、動詞、形容詞、副詞、指示詞、連体詞の6つに分類した。

3.2 格情報

格情報とは、その語に従要素として係り受け可能な格を示すものである。格情報を示すのに次の記述法を用いる。ここで、係り受け可能な格を S_i とし、 $@$ は主要素の位置とする。

- (1) 記号 $[S_1, \dots, S_n]$ は、 $S_1 \sim S_n$ がそれぞれ格を持つ語で、 $[\]$ 内は順序を表し、文を成立させるにはそれらは不可欠な格である。
- (2) 記号 (S_1, \dots, S_n) は、その中の格の順序は望ましい順序であるが、順序を変えても省略しても文は成り立つ。

一般的な例でいうと、 $[S_1, @, S_2(S_3, S_4)]$ は主要素の直前と直後に S_1 と S_2 を必須とし、その後に S_3 または S_4 が、あるいは S_3 と S_4 、 S_4 と S_3 が続くことを意味する。

格情報は、動詞と形容詞が持っている。動詞はその語の意味的な分類と態によって幾つかの種類があり、例えば日本語では、「増加する」、「広がる」などは[(GA)@]、「行う」、「処理する」などは能動態の時[(GA,WO)@]、受動態の時[(GA,NI)@]などを格情報に持つ。形容詞の場合、日本語では格情報として[(GA)@]を持つ。

3.3 修飾

修飾は、語の品詞、活用、格マーカーから得られ、名詞、代名詞に係り受ける連体修飾と動詞、形容詞、副詞に係り受ける連用修飾がある。また、連用修飾にはその語の性質上、修飾先が限定されるものもあり、それらは動詞だけに係り受け可能なものを連用[動]、動詞と形容詞に係り受け可能なものを連用[動形]の様に示す。

3.4 格

格は、修飾の情報を意味的により限定するものである。日本語の場合、主に格マーカーによって得られ、連体修飾ではNO格、連用修飾ではGA格、WO格、KARA格、HE格、TO格、NI格、DE格、時格、提題格などがある。また、各格を持つ語はそれぞれの主要素に対し次のような意味を持つ。GA格を持つ語は、主要素が能動態の時、その行動の主体となり、受動態の時その対象となる。また、WO格、TO格を持つ語も行動の対象であり、KARA格やHE格を持つ語はその行動の方向を示す。DE格を持つ語は、提題、場所、時、手段などに意味が取れるがこれを特定するには、より深い意味解析ならび文脈解析を行う必要がある。

4 構文解析

係り受けの決定は極めて単純である。まず、従要素情報である修飾で係り受け可能な品詞を持つ語を検索し、その語を仮の主要素とする。そして、もし格が存在するならば仮の主要素の格情報を調べ、係り受け可能なならばその格に係り受け情報とし係り受けを決定し、格情報と矛盾があれば係り受け可能な別の語を探す。また格が存在しない時、その修飾に係り受け情報とし係り受けを決定する。そして構文解析では、文の総ての語に対しその係り受けを決定する。

NO	文節	品詞	修飾	格	態	格情報
1	開発者側が	名詞	連用[動形]	GA	*	*
2	用意した	動詞	連体	*	能動	[(GA,WO)@]
3	デモを	名詞	連用[動]	WO	*	*
4	見る	動詞	*	*	能動	[(GA,WO)@]

表 1: 要素リスト

では、例文「開発者側が用意したデモを見る」の解析の様子を示す。

日本語の場合、語の単位は文節となる。そこでまず、この文の形態素解析の結果から文節に切りそれぞれの語情報を記述した要素リストを作成する。例文の要素リストが表 1である。

そして、上記の手順で要素リストを参照しながら構文解析をする。その結果はネットワーク構造になり、例文場合図 2になる。

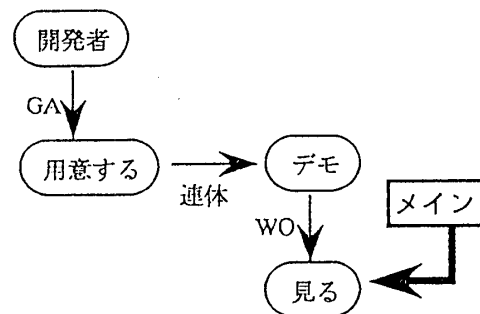


図 2: 構文解析結果を示すネットワーク

5 おわりに

本稿では、構文解析法の手法としての語彙依存文法を提案し、語彙項目の記述法について述べた。その結果、意味構造の一端を含んだ簡潔で柔軟な解析を実現できた。今後、より複雑な文および日本語以外の文に対して適用範囲を広げていく予定である。

参考文献

- [1] 児玉「依存文法の研究」研究社出版(1986)
- [2] Muraki, K., "Augmented Dependency Grammar for Language comprehension," Conference of European Chapter of ACL 85(1985)
- [3] 黒川「自然言語処理入門」近代科学社(1988)
- [4] 牧野、納富「語彙依存文法について」情報処理学会自然言語処理研究会、82-14(1991)