

Augmented Dependency Grammar(ADG)の意味構造(II)

3R-3

村木 一至
NEC Corp.

1.はじめに

自然言語の意味解釈・理解のFrame-of-Referenceとして、依存文法を意味主導へと強化したAugmented Dependency Grammar(ADG)を提案している[1]。本稿では、ADG計算モデルの為に知識記述詞の内、物と属性・属性値間の内在的依存の記述に供する記述詞について報告する。

2.背景・動機

認知記号主義研究の代表選手であった人工知能や自然言語処理が壁に突きあたり、新たな方向を模索している。知識のような物を、書いても書いても、決して望むような判別、認識、推論、理解と言った人間知能の欠片も作れないという壁[6,7]である。事例・用例指向、コーパスベースの統計技法を取り入れて、限界を越えようとする一連の自然言語研究の流れにあって、それらの方法論とも、相性の良い言語の意味記述の方法が望まれる。単に思いつき（検証の手を欠く）の知識を書き貯めることによる、過度の抽象化や恣意性の混入は、元々知識を記述する方法を欠いていることによる。トップダウン的内包手法に、事実上“Grounded”された知識の記述を行うボトムアップ的外延手法[1]を重ねるためのモデル記述法が必要である。

文、発話のモデルは、以下のことを事実上“Ground”させて記述されなければならない。陳述、様、相、命題、命題文の構造、文を成り立たせる句の構造 etc.、陳述、様、相と命題を連結する関係、命題を成り立たせるその要素間の関係、句を成立させるその要素間の関係などである[2]。

	モデル	制約
	格ボタン	
構文レベル	(格シフト・格タイプ)	1文1格 語順 係り受け
意味レベル	ダニ格モデル 姉妹関係	スコープ
写像マップ(MAP)		
概念レベル	関係概念 深層格ボタン	1文1格 依存関係

図1 Augment Dependency Grammar計算モデル[2]

3. Augmented Dependency Grammar(ADG)

日本語の“係り受け”の構文論的内容は、簡潔で理解し易いが、文中の特定の項とその係り先項を選択的に決定する力は貧弱である。日本語処理システムの中の自動係り受け決定機能が実用的に十分に動作するほど豊かな依存規定の記述が望ましい。ADG[1]では、その豊かな記述はより意味に近い、より語彙的な知識として与える。

依存文法ADGは簡潔且つ少数の制約が、語意味モデルの依存構造知識の適用を制御するモデルである。左図1のように、ADG[2]の制約は依存文法の規定する構文内容のいわゆる連体(用)形表現と体(用)言の係り受け規定を付随してその左から右への係り受けの方向規定(その係り受け非交差の原理からなる。ADGは、それに加えて事象表現語(概念)の固有な下位範疇化構造である表層格ボタン、深層格ボタン及び、両ボタン間の個別対応知識(MAP)によって、文の選択的解釈を規定する。これによって各事象表現(用言、事象概念を含む体言)に係る項(格要素)のなんたるかを概念レベルから構文レベルまで規定(説明)できる。

ADGでは陳述、様、相と命題を連結する関係を13種の様相概念種、2種の時相概念、48種の陳述を含む補助概念と事象概念(命題)との関係として規定している。語句を構成する要素間の関係には、加えて、物と属性、事象とその様態(属性)という内在的強連結(依存)関係があり、自然言語の選択的解釈にはこの知識が欠かせない。特にこの属性と物(事象)との関係は、固有の強い連結であるが為に、自然言語の経済原則により、両者の文中での係り受けを明示的に示す構文標識が出現しない。その内在的関係は読み手にとってほぼ無意識の関わりであり、発話の解釈に特段に標識を必要としないのである。だから、この物とその属性に關する個別知識は読み手がコンピュータであるときにすべてなんらかの知識として与えておく必要がある。推論上推論において近年対象のOntology(事実から抽出した詳細モデル)[4]が重視されている。ここで言う内在的強連結関係は、物・事のOntologyそのものである。

具体属性	属性軸名	例	転用	抽象属性	軸名	例
形(立体)	#	球状の				
形(面)	#	四角い	+	真面目さ	真面目-不真面目	四角い 四角四面の
	原形-変形	いびつ	+	心の素直さ	素直-偏屈	いびつな 歪んだ
	直-曲	まっすぐ	+	心の素直さ	素直-偏屈	捻くれた 臍曲り
地形	#	山がちの				
起伏	平-凸凹	凸凹の	+	平和さ	平坦-波瀾	平らな 平和 平穩

表1. 属性(軸)名の抽出例

4.物(事象)とその属性(様態)の数え上げと同定

発話理解に必要な常識である、物(事象)とその固有の属性(様態)の構造知識記述のために、“物・事象=>属性(属性名=>属性値)”のスキーマに基づく属性リストアップと属性名の開発を行った。物(事象)の固有な属性の全数を数え上げるために、その属性(属性値)を具現すると考えられる、日本語の形容詞、形容動詞、副詞を中心に、約7300表現を収集した。主に、分類語彙表[8]とPIVOT翻訳システムの日本語辞書より抽出し、最終的には、“的、様”など造語成分によって造語派生的に得られる周辺形容動詞を除外するようにした。次に、各属性表現が記述する属性の名を逐一命名した。結果として、7300表現から1036種の属性(軸)名を抽出した。

属性名を数え上げるとき、以下のような分類と記述を行った。

- 1) 具体物の属性と抽象物(心含む)の属性の別
- 2) 属性値の極性(両極性、片側極性)と絶対値の別
- 3) 属性の転用(具体->抽象属性への転用)
- 4) 属性を持つ概念語の例

結果の一部を表1に示す。左半分が具体物の属性名と属性軸名、右半分が抽象物の属性の属性名と属性軸名で、中央カラムは、同一表現が具体から抽象へ転用可能(+)か否かを保持している。例えば、“まっすぐ”は具体物の形属性の一つである“直曲”の極軸を表現する属性表現(属性名+属性値)であり、心の素直さ/素直・偏屈極軸の表現として転用されることを示す[5]。属性軸名のカラムが“#”は、同一軸の程度差を表現する別の表現がない事を示す。例えば、“四角い”と言う表現の軸名は“#”であり、その度合いの異なりを表す独立した表現がないことを意味する。

5.類似先行研究との比較

日本語の属性概念の研究が岡田[3]によって精力的になされている。それによれば2300の属性表現(形容詞類)を分類語彙表[8]から抽出して、その内部構造を単純、複合の観点から分析し、属性表現の造語

を分析すると共に、属性概念の深層格パタン7種を提案している。これにより、極性を持つ属性(値)詞の文表現が整理され、属性に作用する動作の意味を表現する文生成を簡易に行うことができた。また、分類語彙表の相の類の意味分類や、自然言語理解システムの為の形容詞、形容動詞などの意味分類が行われ、ある程度の属性名(記述用の名標)が数え上げられている。ただし、具体的に物や事象とその性質を詳細に記述するには、不十分である。

本稿の属性軸名はこの物に固有な性質を記述する有力な道具と考える。また、言語的には、各属性表現の転用能力や、属性軸の極性有無の判別はそれによって造られる文構造を規定するので、文解析・生成の語用論的制約として有効に機能させる事ができると考える。

6.まとめ

物・事象とそれが持つ固有の性質を物・事象のモデル知識として記述するために、日本語属性詞を分析して、まず、属性(軸)名1036個を抽出した。今後は、ADGが定義する“語の意味”に関する知識記述の一端をこれによって記述し、評価する予定である。

[参考文献]

- [1]Muraki, et.al, "Augmented Dependency Grammar for Language Comprehension", Second European AI Conf.'86
- [2]村木, "概念表現と機械翻訳", 野村編「言語処理と機械翻訳」, 講談社, '91
- [3]岡田, 「語の概念の表現と蓄積」, 電子情報通信学会, '91
- [4]溝口, "知識の共有と再利用研究の現状と動向", 人工知能学会誌, Vol.9, No.1, '94
- [5]諏訪, 他, "比喩の計算モデル"等, 情報処理, Vol.34, No.5, '93
- [6]横井, "知識処理と自然言語処理の融合としての大規模知識ベース - 電子辞書から知識アーカイブへ", 人工知能学会誌, Vol.8, No.3, '93
- [7]Lenat, et.al, "Building Large Knowledge Based Systems", Addison-Wesley, '90
- [8]岡研, 「分類語彙表」, 秀英出版, 第5版