

関係子の体系の間のアラインメント

荻野孝野¹ 橋田浩一² 飯島満³ 関口知子⁴

¹日本電子化辞書研究所 ogino@edr.co.jp

²電子技術総合研究所 hasida@etl.go.jp

³明治薬科大学 ijima@horae.dti.ne.jp

⁴意味解析技術応用研究所 seki@sigmatics.co.jp

1. はじめに

文書の意味的な構造を記述するために、言語表現の間の意味的な関係を表現する関係子が用いられている。本稿では5つの関係子の体系を比較検討する。比較検討の過程で明らかになった問題点を整理し、タグつきデータの共有を進めるための基礎資料としたい。

2. 関係子

文書の情報内容を明示するため、文書構造や文内の単語の間の統語的依存関係、意味的關係などを示す XML によるタグ付けが行われている。下記は、「人が歩く」の「が」が行為者を意味することを示す GDA のタグ付けである。

人<ad sem="agt">が</ad>歩く

橋田は、タグによる意味的なアノテーションの社会的な普及を目指す GDA (Global Document Annotation) を提唱している[1]。EDR は、中間言語による意味の把握をめざし、その検討の材料として関係子付きコーパスを作成してきた[2]。UNU/IAS (国連大学高等研究所)は、やはり言語間の意味的情報の共通化をめざして、UNL (Universal Networking Language)を開発してきた[3]。また、国立国語研究所(NLRI: The National Language Research Institute)では、文構造解析のための「深層格と表層格の対応関係」に関する研究をまとめている[4]。NLRI の研究は、前3者とは異なり、情報の共有化を目指すものではないが、文の意味構造をとらえるという目的においては同様である。こういった基礎的な提案とは別に、実際にこれらのアノテーションをデータベース管理に使うために検討している GPM (General Plant Model) なども、事象を統一的な表現で管理する方法の一種である[5]。

異なるタグセットに基づいて開発されるデ

ータベースについて、タグを介して情報を共有化すること、および標準的なタグセットの提唱が可能かどうかについて検討材料を提供するため、これらのタグセットの部分である関係子について比較検討を行なった。

3. タグの諸相

タグによる記述の対象には下記のようにさまざまな側面がある。

- 1) 文書の特徴
文書の種類 (新聞記事、雑誌、論文、etc.)、作成日時、作成者、著作権者など
- 2) 文書の構造
表題、章、段落など
- 3) 品詞
名詞、動詞、形容詞など
- 4) 統語構造
依存関係、等位構造、同格など
- 5) 単語の意味
人名、地名、製品名など
- 6) 文法機能
主語、目的語、連用修飾語、述語
- 7) 事象の間の意味関係
原因、理由、条件、結果など
- 8) 主題役割等の意味的關係
動作主体(agent)、道具(implement)、対象や結果物 (object, patient, result)、上位関係 (superordinate) など
- 9) 後向きコミュニケーション機能
了解、否定、拒否、受諾など
- 10) 前向きコミュニケーション機能
命令、要請、感嘆、呼びかけなど

また、ひとつひとつのタグセットが扱う範囲もさまざまである。表1に関係子の個数を、表2に関係子の分布を示す。GDA では agt.rcp のような関係子の合成が可能なので論理的には無限個の関係子がありうる。表1の97というのはそのうちの代表的なものの

個数である。

GDA タグは上記の 1) ~ 10) をほぼ含むが、NLRI [4] の深層格は上記の 7)、8) に該当する。EDR も 3) ~ 10) に該当する情報を持っているが、そのうち 9)、10) は関係子でなく属性子や特別概念として扱っ

ている。また GPM も 7)、8) に該当し、関係子の体系は EDR と国語研を統合したものに近い。GDA では 6) ~ 10) を「関係子」として扱っているのが、本稿での比較も 6) から 10) を中心に行なう。

表 1：関係子の個数

関係子の体系	GDA	GPM	NLRI	EDR	UNL
関係子の個数	97	41	35	51	94

表 2：意味関係の種類ごとの関係子の分布

各欄の数値は当該の種類に属する関係子の個数。

対応する関係子を同じ行に示す。#は対応する関係子がないことを意味する。

種類	GDA	GPM	NLRI	EDR	UNL
文法機能	7 arg	3 MDF	1 #	3 modifier	3 mod
参加者	14 pat (patient)	14 OBJ	13 OBJ	13 object	12 obj
補足同格	4 cnt eg (example)	1 CNT #	1 COT #	1 modifier #	1 cnt #
因果関係	5 cau cntrst (contrast)	3 CAU #	3 CAU #	3 cause #	3 rsn #
包含関係と次元	19 tmx sup (superordinate)	15 DUR #	11 DUR #	14 duration #	15 dur #
論理的関係	4 and eq (equivalence)	0 # #	0 # #	2 and #	2 and #
その他の意味関係	23 ccm (circumstance) met (metonymy)	8 COC #	8 CON #	7 modifier #	6 con #
後向きコミュニケーション機能	9 rpy (reply YES) und (understand)	0 # #	0 # #	2 consent #	1 affirmative #
前向きコミュニケーション機能	12 ord (order) stt (statement)	0 # #	0 # #	6 imperative #	4 imperative #

4. 関係子の間の比較

関係子の間の対応には、以下に示すように定義がぴったり合う場合、ずれる場合、関係子名は同じだが定義にずれがある場合など、一致のパターンに次のような多様性がある。

(1) 対応する関係子がある

(1-1) 意味の範囲が等しい

(1-2) GDA の関係子の方が意味の範囲が狭い

(1-3) GDA の関係子の方が意味の範囲が広い

(1-4) 意味の範囲が重なるが包含関係にない

(2) 対応する関係子がない

以下、GDA、GPM、NLRI、EDR、UNL の

対応する 5 個の関係を/で区切って並べて示す。例の中の下線は、当該の関係を意味する助詞または助詞相当語句を示す。そのような助詞または助詞相当語句がない場合、当該の関係子で係る句を太字で示す。

(1)対応する関係子がある

(1-1) 意味の範囲が等しい

例：1958年に生まれた。

tim (時間的位置)/TIME/TIME/time/tim

例：川を下る。

via (経由)/place-through

/space-through/location/via

(1-2) GDA の関係子の方が意味の範囲が狭い

例：茶を沸かす

res (結果物)/ object/object/object/object

GDA では「茶を沸かす」、「家を建てる」などの「を」が「結果としての事物」かどうかで区別した関係子を与えているが、ここで比較している他の関係子の体系ではすべて object である。GDA 以外では、より広い範囲の関係子を割り当てることになる。

同様に、GDA の pat (被動作対象)も他の体系では object に該当する。

例：りんごを食べる

pat/object/object/object/ object

(1-3) GDA の関係子の方が意味の範囲が広い

例：海外から戻る

src/place-from/space-from/source

/initial place

例：子供からもらった服

src/origin/origin/source/origin

GDA の src および EDR の source は最初の状態も位置も移動元(発信人、与え手)も含んだ広いものであるが、GPM と NLRI および UNL の source は「位置の始まり」のみを指す source で「移動元になる人」は source に含まない。

(1-4) 意味の範囲が重なるが包含関係にない

例：学生3人, two spoonfuls of sugar

mob/object/object/object/object

GDA では mob (計量の対象; measured object)が一般の対象と区別されているが、他の関係子の体系にはこの区別がなく、すべて object である。

(2) 対応する関係子がない

表2で#が GDA 以外の体系において対応する関係子がないことを示す。

例：その家のドア

sup/###/###

例：雨にもかかわらず行く

cnc/###/###

以上、対応する関係子に対応関係の有無や関係子どうしの包含関係でみてみた。以下は、そのうち、(1.2)と(1.3)の範囲で、差異をより詳しく取り上げてみる(表3, 4)。

例：母が子供を心配する

agt/agent/exp/ agent/agent

agent は有意志行為の主体であるが、ここには表3に示すように「物理的行為、判断思考」などの行為のうち、「考える、決める、驚く、心配する」など、精神や思考に関わる行為や事象を含めるかどうか、含めるとして、その境界をどう定義するかが難しい。NLRI では「考える」など多少意図的なものでも心理的な行為は「感情の主体」と同じく「experiencer (経験主)」である。

有意志体としての「人間」が主体になった場合の関係子の違いを表3に示す。なお、ついでながら、「生まれる」、「病気になる」など、人の意志のコントロール外の変化の主体は、「対象」(a-object あるいは object)である。

表4に示す「相手」も様々な場合がある事例である。「相手」をおおまかに(1)与え手、(2)相手を必要とする行為(eg.戦う)の相手、(3)任意に人を伴う行為に分けると、対応は表4のようになる。

関係子の体系の対応に関する詳細についてはMMA2001の資料[7]を参照されたい。

表 3：行為主体を表わす関係子

		GDA	GPM	NLRI	EDR	UNL	例
有意志行為の主体	物理的行為、 身体的、抽象的 な意図的行為	agt	agent	agent	agent	agent	健がたたく 健が決める 健が解説する
	判断思考など 精神的行為	agt	agent	exp	agent	agent	健が思い出す、健が 信じる、健が心配す る、健が記憶する
感情の主体		exp	exp	exp	agent	agent	健が泣く 健が驚く
相互行為の主体		agt.rcp	agent	agent	agent	agent	二人が一緒になる

表 4：「相手」に該当する関係子の多様性

	GDA	GPM	NLRI	EDR	UNL	例
避ける相手	opp (opposit direction)	opponent place-from	opponent space-from	source source	source initial place	敵から身を守る 家を出る
与え手	src (source)	origin	origin	source	origin	健から貰う
受け手	gol (goal)	recipient	recipient	beneficiary	beneficiary	太郎に話す
相互性行為の同格主体	agt.rcp (agentive reciprocal partner)	agent	agent	agent	agent	二人が結婚する
相互性行為の相手	agt.rpt	partner	partner	agent	partner	花子と戦う
随伴相手	jnt (joint participant)	accompany	accompany	agent modifier	co-agent	健と一緒に歩く

5. 終わりに

言語やメディアを越えてデータの意味内容を何らかの形で共有することができれば、知識管理等の実際的应用と解析プログラムのトレーニングなどの基礎的な研究開発の両方に有用である。そのための関係子の体系のアラインメントについて報告した。以上で述べたようにさまざまな課題があるが、今後それらを克服する方法を探っていきたい。

参考文献

- [1] <http://i-content.org/GDA/tagset.html>
- [2] <http://www.ijent.or.jp/edr>
- [3] <http://www.unl.ias.unu.edu/publications/gm/>
- [4] 国立国語研究所：日本語における深層格と表層格の対応関係，国立国語研究所研究報告 113，国立国語研究所，1997 年 3 月，三省堂。
- [5] 芝尾絃一：発電プラント/プロダクト/モデル/共有システム開発報告書、(財)日本情報処理開発協会ステップ情報センター、1997。
- [6] 荻野孝野、小林正博：EDR 概念記述の抽象化による動詞パターンについて、2000 年 7 月，電子情報通信学会誌 *NLC2000-10~15*。
- [7] K. Hasida, T. Ogino, M. Iijima, T. Sekiguti, T. Inui: Alignment of Function-Word Semantics, Jan. 2001, *MMA2001*.