

# 照応・時制の underspecified な談話表示構造

金山 博<sup>†</sup> 緒方 典裕<sup>‡</sup> 辻井 潤一<sup>†</sup>  
 {kanayama, ogata, tsujii}@is.s.u-tokyo.ac.jp

## 1.はじめに

本稿では、英文テキストの意味表示の手段として、UDRS(Underspecified DRS)[2][3]を HPSG 文法[4]により構文解析と同時に構成すること(第3節)及び、照応・時制の曖昧性を表示できるように「未決定子」を導入して UDRS を拡張すること(第4節)の2点について論じる。

談話表示理論(DRT)[1]では、モデルとの対応がとれる論理表現である談話表示構造(DRS)を用いて、文章の意味を表現する。しかし、談話中の文が量化子のスコープや複数名詞の読み方(collective/distributive)に起因する曖昧性を持つ場合、全ての解釈に相当する構造の論理和を用いて DRS を表現せざるを得ないという問題点がある。 $n$  文からなる談話の各文が  $m$  通りの解釈を持つ場合、談話全体の解釈は  $n^m$  通りとなり、事実上表現が不可能となる。

UDRS は DRS の拡張であり、談話における各文が上に挙げたような曖昧性を持つ場合にも、一つの論理構造で一つの文を表せるようにしたものである。UDRS により、曖昧性を持つ文であっても、各々の解釈の論理和を用いることなく真偽判定が可能となり、曖昧な文を用いての推論も可能となった。また、制約の追加により、文脈から曖昧性を単調に減少させることができる。

UDRS は、語により様々な形になり得るため、多数の語彙を扱うために、UDRS の構成要素を辞書項目に記述(語彙化)し、HPSG による構文解析と同時にボトムアップに構成する手法を提案する。これは、Frank & Reyle による UDRS の構成法[5]を改良したものである。

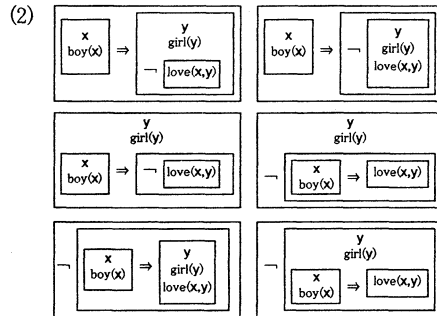
次に、UDRS を拡張し、照応解析・時制解析の際に生じる曖昧性を表現し、その表示系に意味論と推論規則を与えるを試みる。本稿では特に、曖昧性を含む表示に対して、制約を単調に加えて曖昧解消ができる例を示す。

## 2.UDRS について

DRS を構成する際には、量化子のスコープや複数名詞の読み方による曖昧性が問題となる。例として文(1)に対する論理表現を考えると、「every」「a」「否定」のスコープの順序により、(2)

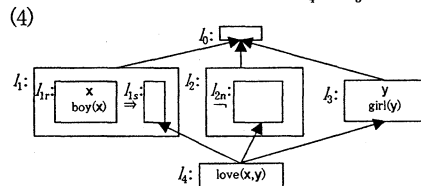
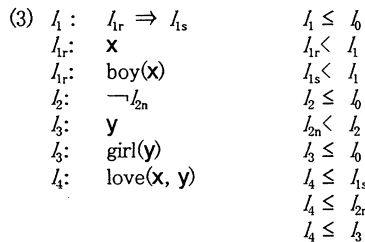
に示すような6通りの DRS が考えられてしまう。

(1) Every boy doesn't love a girl.



(1)の持つ6通りの解釈のいずれも無視せずに扱うには、(1)の文を(2)の6つの DRS を論理和で結んだ DRS として表現しなくてはならない。しかし、談話全体の DRS を構築する際には、論理和の数が増加し、扱うことが現実的でなくなる。

そこで、本稿では UDRS を用いて曖昧な文の持つ意味を表現することにする。UDRS は、曖昧性を一つの形にまとめた論理構造であり、DRS の断片と、それらの間の包含関係を表す半順序より成っている。(1)に対応する UDRS は(3)、それを視覚的に表した構造が(4)である。



(3)の左列は、指示対象と DRS 条件(condition)を、それが含まれる DRS の断片を指し示すラベ

<sup>†</sup> 東京大学理学部情報科学科

<sup>‡</sup> 日本学術振興会特別研究員

ル( $l_6 \sim l_7$ )と共に列挙したものである。なお、 $l_6$  は top-DRS(最外のスコープ)に相当する。

右列は、ラベル間の従属(包含)関係を示す。 $\leq$  は弱従属、 $<$  は強従属と呼ぶ。強従属は、条件文(every などの量化子)・否定などにより、必ず従属される DRS を含むことを表す。弱従属の場合は、二つのラベル  $l_1, l_2$  が同じ場所に位置する( $l_1=l_2$  となる)場合がある。

$l_4$  と  $l_5 \cdot l_6 \cdot l_7$  との間に定められている弱従属の制約により、述語 love が否定のスコープの中に入ること、及び love の引数 x,y が自由変数にならないことが保証される。

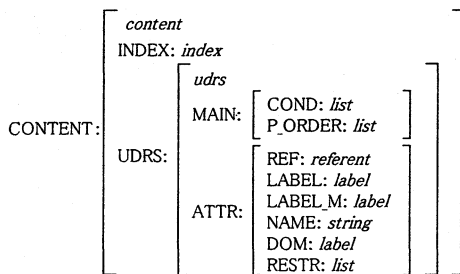
### 3. HPSG と UDRS の統合

前章で述べた UDRS を英文から構成する方法として、HPSG[4]による構文解析と同時にボトムアップに構成するという手法をとった。HPSG の CONTENT 素性(意味素性に相当する部分)の中に UDRS 素性を設け、各語が持つ UDRS の形を辞書項目中に記述した。

Frank&Reyle[5]においても、同様に HPSG 文法を用いた UDRS の構成について述べているが、本稿での方法は、より多くの情報を語彙化し、構成規則をより少なくしてある。3.3 節の最後に文献[5]との相違点を述べる。

#### 3.1 素性のデザイン

CONTENT 素性以下の構成を以下に示す。



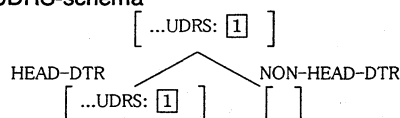
UDRS 素性のうち、MAIN 素性以下に、語または節が持つ UDRS-condition 及び半順序関係のリストがある。また、ATTR 素性以下には、他ノードと構造共有させるための素性を記述した。REF 素性は、導入された指示対象を表し、他のノードの述語の引数を埋めるために使われる。LABEL・LABEL\_M 素性はラベルを表し、半順序関係を構築する際に用いられる。NAME 素性には主に名詞の値、DOM 素性はラベルの上界([2]参照)、RESTR 素性は限定的に修飾されているラベルのリストである。

#### 3.2 UDRS-Principle

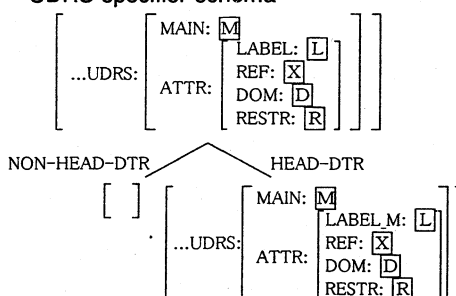
UDRS 素性の構成原理として、UDRS-Principle を追加した。その内容は、下表に示すように HPSG における6つの ID スキーマの種類により3通りに場合分けし、それぞれに応じた UDRS のスキーマが適用する、というものである。

head-subject-schema	} UDRS-schema
head-complement-schema	
head-marker-schema	
head-filler-schema	
head-specifier-schema	→UDRS-specifier-schema
head-adjunct-schema	→UDRS-adjunct-schema

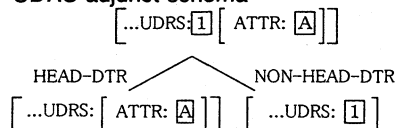
##### ・UDRS-schema



##### ・UDRS-specifier-schema



##### ・UDRS-adjunct-schema



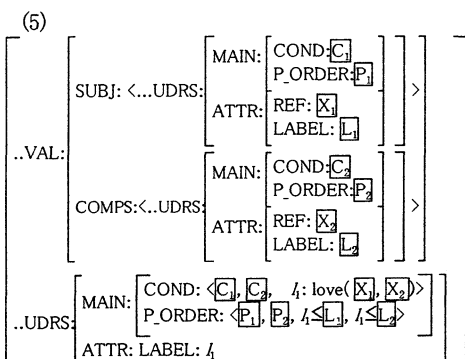
UDRS-schema は、主辞の UDRS 素性をそのまま親の UDRS 素性として継承させる。UDRS-specifier-schema は、名詞が限定詞を下位範疇化する際に用いられ、主辞の LABEL\_M 素性が親の LABEL 素性となるという点だけが UDRS-schema と異なる。また、UDRS-adjunct-schema の場合は、付加語(adjunct)が意味的主辞として働き、UDRS 素性が親に伝えられる。但し、他の属性(ATTR 素性)は主辞のものと単一化されることとした。

#### 3.3 Frank&Reyle の手法との比較

Frank&Reyle の手法では、動詞が名詞句を下位範疇化する際に生じる半順序関係が、UDRS

の構成規則である Semantics Principle が適用された段階で加えられるため、その他の文法構造については別の規則が必要となる。更に、親ノードの UDRS の値は、二つの子ノードの UDRS に関する制約の和となっている。

比較のため、本稿での手法における語彙項目中の UDRS 素性の例として、他動詞 love の例を(5)に示す。



我々の手法においては、(5)の P\_ORDER 素性のように、動詞の辞書項目の UDRS 素性自体が半順序関係を持っており、どのラベルとの半順序関係を規定すべきかが、構文解析の際に決定される。また、(5)の COND, P\_ORDER 素性は、下位範疇化される名詞句の COND, P\_ORDER 素性  $\langle C_1, C_2, P_1, P_2 \rangle$  をリスト中に含むため、主辞である love の側のノードが、もう一方のノードの値を含めて、親ノードの値となる。

関係代名詞や接続詞なども、辞書項目中に各語の振る舞いを記述することによって、3種のスキーマの枠組みで扱うことができた。

## 4. 照応と時制

### 4.1 照応の underspecification

通常の DRT では、照応詞(代名詞・定冠詞など)が出現した場合、その指示対象  $x$  に対して、 $x$  の置かれる位置への到達可能性(accessibility)を手がかりとして、適切な先行詞  $y$  を探し、 $x = y$  という条件を加える。一方、UDRSの枠組みでは、DRS の従属関係が確定していないため、到達可能性が定義できず、照応先の同定のためには新たな概念が必要となる。

本稿では、UDRS での照応詞の扱いとして、照応先となり得る指示対象の候補のみを規定し、照応先を直接一意に決定することなく論理構造を構成する。この表示における談話の意味は、既に規定されている制約のなかで可能な全ての

解釈とする。

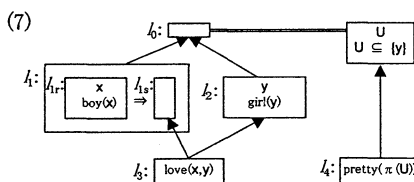
まず、先行詞の候補となる指示対象の集合  $U$ 、及び集合から一つを選択する関数  $\pi$  を導入し、照応詞の指示対象を  $\pi(U)$  なる形で表す。この  $\pi(U)$  を未決定子(underspecified discourse referent)と呼ぶこととする。 $U$  は、代名詞が出現した文に対する UDRS の top-DRS( $I_0$ )に置かれる指示対象で、照応先の候補となる指示対象の集合である。初期状態として、「 $U$  は照応詞が含まれる文及びそれ以前の文の指示対象のうち、性・数が一致し、かつ top-DRS から潜在到達可能性(以下に定義する)を持つものの集合である」という制約が加えられる。

**定義: 潜在到達可能性(potential accessibility)**

ラベル  $I$  から、ラベル  $I_1$  の位置にある指示対象  $x$  に潜在到達可能性があるとは、ラベル  $I$  からラベル  $I_1$  に至る弱従属のみからなる半順序の path があり、かつ強従属を含む path がないこと。

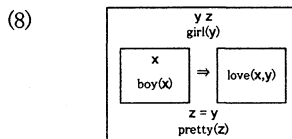
照応詞の指示対象を未決定子  $\pi(U)$  としてにより、表示として照応先を明示的に規定しなくても、次に示す例(6)(対応する UDRS は(7))のように、スコープの曖昧性との相互作用により照応先を絞れる場合がある。

(6) Every boy loves a girl. She is pretty.



1文目はスコープの曖昧性を持ち、2文目には代名詞  $\pi(U)$  がある。 $I_0$  から  $y$  へは潜在到達可能性があるのに対し、強従属の関係  $I_1 < I_1$  により  $I_0$  から  $x$  へは潜在到達可能でない。従って  $U = \{y\}$  となるため、 $\pi(U) = y$  が定まる。

この時、 $y$  は  $I_0$  から実際に到達可能でなくてはならない。従って、制約  $I_0 = I_0$  が自動的に付加される(これが行えることが潜在到達可能性である)ため、1文目のスコープの曖昧性も除去される。故に、UDRS(7)は、実質的に DRS(8)と等価であるといえ、スコープと照応の曖昧性が同時に解消されていることになる。



(9)の例もまた、文の解釈の時点では照応先が定められないが、文脈からの制約により曖昧性が解消される例である。(9)における照応に関する制約は UDRS で(10)のように記述される(スコープの関係は省略してある)。

- (9) John fought with Bill. He was wounded.  
And he was taken to a hospital by Bill.

(10)

j b	U <sub>1</sub>	U <sub>2</sub> h
John(j) Bill(b) fight(j,b)	U <sub>1</sub> ⊆ {j,b} be wounded(π(U <sub>1</sub> ))	U <sub>2</sub> ⊆ {j,b} hospital(h) take(b, π(U <sub>2</sub> ), h) b ≠ U <sub>2</sub>

2文目の he は、未決定子 π(U<sub>1</sub>)で表され、U<sub>1</sub> は Bill または John を含み得る。3文目は、目的語が再帰代名詞でないため、he と John とが同一の指示対象を指さないという制約が加えられる。U<sub>2</sub> ⊆ {j,b} と b ∉ U<sub>2</sub> の二つの制約より、U<sub>2</sub> ⊆ {b} となり、未決定子 π(U<sub>2</sub>) は b に定まる。

仮に、「二文で代名詞の主語が連続する場合にはそれらは同一のものを指す」という制約を加えると、2文目・3文目の he が Bill を指すことになり、2文目の照応先も確定させることができる。

## 4.2 時制の underspecification

DRT では、談話の時制解釈もまた一種の照応現象として扱われる。具体的には、時制・アスペクトを、location time  $t$ 、発話時間  $n$ 、reference point  $Rpt$ 、イベント  $e$ (または状態  $s$ )などの時間的オブジェクト相互の関係で記述する。

$Rpt$  は談話中に既に導入されたいずれかの時間的オブジェクトの指示対象と同定される。従って、談話途中の  $Rpt$  の同定は一種の照応と考えることができ、場合によっては上に述べた代名詞の照応と同様の照応の曖昧性が生じることになる。通常は  $Rpt$  は直前の文のイベントに同定されるが、次の談話では最後の文の  $Rpt$  の同定先が曖昧である。

- (11) John pushed Bill. He fell. Mary also fell.

この談話の時制の部分は、従来の DRT では(12)のように表示する(各行が各文に対応)。3文目では、reference point が  $e_2$  であるとし、 $e_1$  と異なるケースが無視されている。

(12)

	$t_1 < n$	$e_1 \subseteq t_1$	
Rpt := e <sub>1</sub>	$t_2 < n$	$e_1 < e_2$	$e_2 \subseteq t_2$
Rpt := e <sub>2</sub>	$t_3 < n$	$e_2 < e_3$	$e_3 \subseteq t_3$

これを変更し、我々は(13)のように書くことにする。ここでは、談話の1文目以外のイベントに関して reference point を代名詞の時と同様に関数  $\pi$  を用いた未決定子として定め、reference point の

後にイベントが行われたという条件を加える。照応先の候補を表す指示対象  $T_1, T_2$  を導入し、既出のイベント(または状態)の部分集合とする。

(13)

$t_1 e_1$	$t_2 e_2 T_1$	$t_3 e_3 T_2$
$t_1 < n$	$T_1 \subseteq \{e_1\}$	$T_2 \subseteq \{e_1, e_2\}$
$e_1 \subseteq t_1$	$t_2 < n$	$t_3 < n$
	$\pi(T_1) < e_2$	$\pi(T_2) < e_3$
	$e_2 \subseteq t_2$	$e_3 \subseteq t_3$

(13)の表記では、3文目の reference point の候補が  $e_1$  または  $e_2$  となっており、underspecified な状態となっている。これにより、Mary fell の因果関係が John pushed または He fell のいずれかに結びつけられることが表示できている。

## 5. おわりに

本稿で述べた第一のテーマである UDRS の語彙化であるが、僅か3つの構成規則で、構文解析の段階で得られる制約を構成することができた。更に、時制に関する制約、特に瞬間動詞と状態動詞の差などを語彙項目に導入することにより、時制を含む UDRS の構成が可能である。

照応と時制の underspecification については、文脈からの曖昧解消をする前段階としての役割について述べた。更に、Reyle らの枠組み[2][6]で提唱されている、UDRS の推論規則を導入すれば、実世界のテキストの解釈の際に照応やスコープの決定が困難である場合においても、推論により情報を得ることが可能になる。今後、本稿の後半部分についても実装したい。

## 参考文献

- [1] Kamp Hans and Uwe Reyle. *From Discourse to Logic*. The Netherlands: Kluwer Academic Publishers, 1993.
- [2] Uwe Reyle. Dealing with ambiguities by underspecification: Construction, representation and deduction. *Journal of Semantics*, 10(2):123-79, 1993.
- [3] Uwe Reyle. Co-indexing labeled DRs to represent and reason with ambiguities. *Semantic Ambiguity and Underspecification*, 1995.
- [4] Carl Pollard and Ivan A.Sag. *Head-Driven Phrase Structure Grammar*. University of Chicago Press and CSLI Publications, 1994.
- [5] Frank Anette and Uwe Reyle. Principle based semantics for HPSG. In *Proceedings of the 7<sup>th</sup> Conference of the European Chapter of the Association for Computer Linguistics*, Dublin, 1995.
- [6] On Reasoning with Ambiguities. In *Proceedings of EACL 95, Dublin*.