

概念依存図式と英語文生成

村木一至

日本電気(株) C & C システム研究所

1. はじめに

現在、意味表現から表面文を生成するための課題は以下のプロポジションを言い表わせる。

「伝達したい意味内容と読者の意図から如何にして単語・文体の選択を可能にするか。」

本稿は、ADG (Augmented Dependency Grammar) という枠組を提案し、プログラマティック機能 (P-func) による文体決定のモデルについて述べる。

文生成の研究は、(変形) 生成文法的能力 (competence)、その文法集合のコンパクトネス検証を目的としたランダムテキスト生成の研究に始まった。特定の構造 (ないし意味内容) から文生成を試みたのは S. Klein (1965)^[1] が最初であった。彼は、句構造解析木と同一文に対する依存木に生成用依存文法をランダムに適用し、生成された文は文法的には正しくとも意味内容を正しく伝えられとは限らなかった。

R. Quillian (1968)^[2] は、文法的健全性よりも意味妥当文生成を主目的とした文生成研究を行った。入力に彼の提案する意味網とし、意味網中の板上の論理関係を鍵として単語ないし文の選択を行った。彼はこの研究を通じ、S. Klein が出発点とした構造的依存関係よりも、意味依存関係が文生成により大きな機能を果たすことを示した。

これ以後意味内容を如何に文体として実現するかという問題は、AI 的言語研究の 1 話題となった。Simmons & Slowom (1972)^[3] の深層格意味網からの ATNE 用いた英文生成、N. Goldman

(1975)^[4] による CD 構造からの英文生成 (semantic primitive 構造に別網による逆語彙分割規則を適用して複合動詞を選択することに特徴があった。) の研究があった。

他方言語学からも、計算処理に向いた文法モデルの提案がなされた。LFG (Lexical Functional Grammar)^[4] はその代表例であり、AI 及び計算機科学が曖昧にしていた各種機能を、構成構造、機能構造、語彙規則に分割し、3 者間の写像関係を定式化しようとしている。

GPSG (Generalized Phrase Structure Grammar) は、^[6] 句構造規則、線形化規則、語彙規則の 3 つにより文法をモデル化しようとする。

AI 的、言語的観点からの研究の内、ADG^{[1][2]} は次のような機能を文生成、文解析を統一的に抱えるものがある。

- | | |
|----------------|--------|
| 1) 概念 (論理) 機能 | C-func |
| 2) プログラマティック機能 | P-func |
| 3) 語彙機能 | L-func |
| 4) 文法機能 | G-func |
| 5) 構文・形態属性 | S-feat |
| 6) 順序属性 | O-feat |

文生成に於いて、入力構造は概念依存図式 (Cnet) と呼ばれ上記 C-func, P-func が付与された深層格意味構造である。同構造に対し、辞書、依存文法規則、順序規則の適用により表面語、文、文体選択属性の付与が行われる。

この文法モデルによれば、変形操作の代わりに順序属性を導入すること

表層構造を規定し、入力 Cnet 内概念依存関係と、アラブマティック機能を諸概念機能に適用することにより、構文構造を規定する。

以下の節では、ADG 文法モデルの背景及び、ADG 文法モデルによる計算手順 LARGO の基本動作について述べる。最後に検証システム LARGO を通して文法モデルの能力について議論する。

2. ADG 文法の機能的背景

ADG 文法は、入力 Cnet 上に定式化される。Cnet は以下のように定義されるシステムである。

$$D1. CP \triangleq \{SYMBOL_i\}$$

$$D2. Cr \triangleq \{Relation_j\} \subset CP$$

$$CARD(Cr) = 31$$

$$D3. Cp \triangleq CP - Cr$$

$$D4. Cnet \subset Cp \times Cp \times Cr$$

$$D5. Ptbl \subset Cp \times Cp \times Cr \times func$$

SYMBOL_i は、語が持つ概念記号であり、近似的にその数は辞書単語項目に一致する。この集合 CP の部分集合 Cr の要素は現在 31 個の概念関係である。Cr_k は、AGENT, OBJECT, SUPER-SUB, PART OF, REASON 等が含まれ各々は 2 項述語と見ることを要する。

Cnet は、Cp × Cp × Cr の直積の部分集合であり、知識ベース Ptbl に入力文を写像したものの func 軸へのプロジェクションである。また、Cnet 中の Cr は有方向であり、論理的ハッド、デバインタントを指定する。

つまり、Cnet は、概念機能 C-func が与えられている。R. Quillian (1968)

が議論した、論理関係同定メカニズムは、ADG 文法による日本語解析モデル LEGATO (参考文献 [12] 参照) により実現されている。

また、日本語解析時の日本語文を決定した意図は、TOPIC, FOCUS として抽出され、Cnet 上に付与されている (P-func 付与)。実際には、提題助詞「は」、述部目的語、語順によって導き出される。

文の成り立ちを規定する為にくくりの機能は指摘されている。

Simmons & Slowm が取り上げた機能は、意味構造中の深層格機能であり、Fillmore が提案した普遍文法を踏襲している。ADG モデルでは、31 個の Cr 要素に含まれている。Cr_i は、この要素によって 2 つの意味 SYMBOL_{i, j} が連結され、その機能を C-func と呼ぶ。GPSG, LFG は、述語論理表記を採用している。Kowalski によれば、^[8] 例えは次の命題 1 は、命題 2 のように分解される。Cnet_i は、深層格を用い

命題 1. GIVE (TOM, MARY, BOOK)

命題 2. $X(GIVE, TOM, 1) \wedge$
 $X(GIVE, MARY, 2) \wedge$
 $X(GIVE, BOOK, 3)$

2 命題 3 のように記述する。とれる。

命題 3. $(GIVE, TOM, +AGT) \wedge$
 $(GIVE, MARY, +TAR) \wedge$
 $(GIVE, BOOK, +OBJ)$

or

$(TOM, GIVE, -AGT) \wedge$
 $(MARY, GIVE, -TAR) \wedge$
 $(GIVE, BOOK, +OBJ)$

or else

AGT, TAR, OBJ を概念機能 C-func と

考える。同表記中の関係者に付加される「-」,「+」は概念依存の方向を示す(以後「+」は省略する)。

このような表記を採用する理由は、少くとも表現力が同等であること、実際の文に於いて、AGT, TAR, OBJ等の概念記号の表層に於いて自語としても頻繁に出現することを直接的に表現できるからである。それに対し、命題1.の表記では、例えばGIVEとMARYの依存機能は別手段を設けない限り、明示化されない。

諸量機能L-funcは、文というもののその要素である単語から成り立つという意味を重要とする。L-funcを導入することにより、従来の変形文法と同構造規則なし変形規則を導入した文の基本構造を、諸量規則の同題として扱うことができる。例えば、概念記号「NEED」に対し、それを概念依存ヘッドとする表層構造を得るために、表層形 require, necessary, need等の単語選択により、同辞書項目中のL-funcにより基本構造が決定する。

LFG, GPSG等は、この機能をベースに非変形文生成を可能にする。

ADGモデルでは、概念ヘッド(例えば命題3.中のGIVE)に対する諸辞書内のL-ruleに次のような機能を与える。

与えられる概念ヘッドHに、と直接の概念ダイアコンティンゲンツにH_iに対し、H_iに対する諸語の各々のL-ruleは、

L-rule_l: {D_j} → {SUB, DOB, IOB, COM}

文法機能
G-func { SUBJECT
 Direct Object
 Indirect Object
 COMplement }

上記の字像を定義する。

基本的なL-ruleが、G-funcを付与する機能を持ち、同時に特定単語選択されることにより、構文構造の形式等が、各D_jに付与される。

以上説明したC-func, P-func, L-func, G-func, とその他S-featに関する演算は、依存規則D-ruleにより行われる。この機能は、C-funcを鍵として、P-funcの制約下でHi概念ヘッドに対応するL-ruleを選択的に適用して、文法機能G-func, 構文・形態属性S-feat付与を行う。

表層構造(諸順を含む)は、C-func, P-func, G-func, S-featを鍵とするO-ruleにより決定される。O-ruleは、出力として各量状単語のO-featureとして相対順序を付与し、構造の書き換えは行われぬ。同規則適用条件S-featには、構造バウンス属性等も含まれる。

図1. は、GPSG, LFG, ADGの概念的差違をまとめたものである。

GPSGでは、意味表現に内包論理式を採用し、構成(constituent)決定に同構造規則を用いる。構成決定の際には、同構造木を直接書き換える代りに、メタ規則を用いて、同構造規則を書き換える。これにより、同構造変形操作が回避できるが、メタ規則適用のタイミング制御について解決されていない。

構成記述を同構造規則、順序規則に分割しているが、順序規則は同構造規則集合のコンパクトネスを改良するだけの意義を付与されていると過ぎない。

LFGでは、論理構造、H-structure, C-structureの3つを文生成規定を行

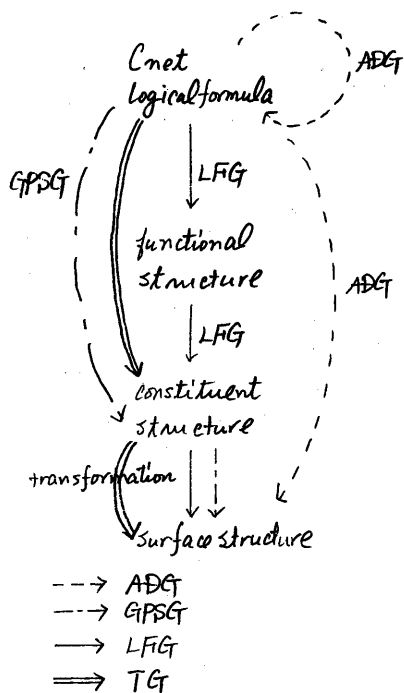


図1 モデルの比較

う。論理構造に対し諸彙挿入規則を適用し F-structure を得た後、態変形等の要求があれば、同 F-structure に単語辞書中の文法関係再付則を適用しながら C-structure, 表層文の生成を行う。つまり、構成自身は書き換えないで、F-structure の書き換えを回避する。

両者の変形回避法に対し、ADGは、全ての機能・属性は直接 Cnet に変形なしで行うされる。依存規則 D-rule, 順序規則 O-rule が分離されていることによりこれが可能となる。

3.ADG 機能・属性の関連

ADGは、Cnet が与えられたとき、そこから更語文を生成するために、まず何が主題として表現されるべきかを決

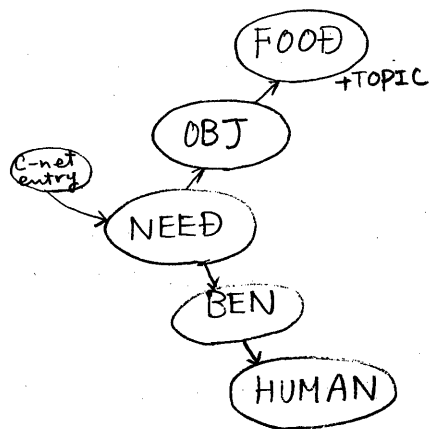
定する。この主題化によって、主節となるべき部分 Cnet の文法機能が決定される。

表1. は、機能・属性の主題化によってどう関係付けられるかを示すものである。なお、図2に、対応する Cnet を掲げる。

表1. 機能・属性関連

	FOOD	NEED	HUMAN
C-func	OBJ	C-head.	BEN
P-func	TOPIC	Relme.	
G-func	SUB	PRED	COMP
S-feat	+nom		+nom, +PREP
O-feat	1 < 2 < 3		

Food is {needed
necessary} for human



C-head —> C-dependent

図2. 対応する Cnet

図2の入力 Cnet に対し, C-head 「NEED」の単語辞書中の L-rule によつて, C-func, P-func により主語化が行はれる。

L-rule は, 図3の構造を持ち, MAP_i によつて, Cr を G-func を写像する。

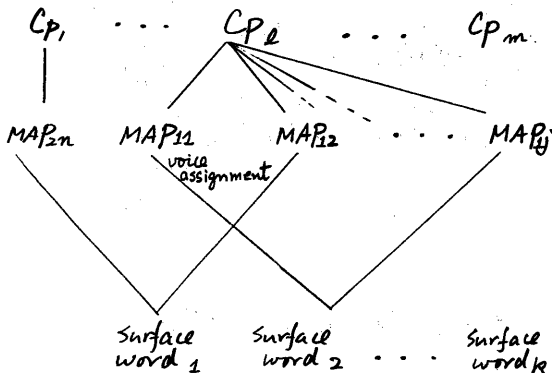


図3. Cpと単語辞書対応

MAP_i は, オ2節で定義した L-rule と同一であり, その適用は, C-head 「NEED」に対する P-func = TOPIC の付与と最大デフォント (FOOD) の SUBJECT を付与するように制御される。むしろ, こうした L-rule が存在しないときは, 単に C-head が持つデフォント (Df) を最大限に文法項へ写像する Maximal の L-rule が選択される。その結果, SUBJECT 化を伴う TOPIC は, 順序規則の適用を制御する。

G-func 付与後, 依存文法 D-rule の制御下で S-feat 付与が行はれる。

O-feat は, 上記 C-func, P-func, G-func, S-feat を条件とする2項対称順序付与規則 O-rule によつて付与される。^[5]

$$O\text{-rule:} \begin{bmatrix} \text{left} \\ C\text{-func} \\ P\text{-func} \\ G\text{-func} \\ S\text{-feat} \end{bmatrix} \begin{bmatrix} \text{right} \\ C\text{-func} \\ P\text{-func} \\ G\text{-func} \\ S\text{-feat} \end{bmatrix}$$

同規則は直接 Left-right の順序規定を行ない, その適用は G-func, S-feat によつて付与される1つの構文ヘッドにのみ適用される。

4. ADG の実現: LARGO-E-morph

ADG を言語翻訳系の英文生成モジュールとしてインポートした LARGO-E-morph を付録図1に示す。図中, LARGO は D-rule, O-rule と辞書内の L-rule を入力 Cnet に適用し, 文の構造, 語彙挿入, 構文・形態属性を付与する。その後 E-morph は, Cnet 上の構文・形態属性に形態系生成規則 M-rule を適用し, 辞書内容から表尺形を得るとともに, 順序属性に従った語順の整列を行う。

LARGO モジュールに於ける規則, 辞書引きの制御は, Cnet エントリーを指定される Cp から順次 depth-first に構造を探索する。辞書引きによる MAP_i の選択, それに伴う文法機能, 構文属性は, 各 Cp に対し2設定される FEAT と呼ぶ領域に保持される。又, 同時に順序規則が適用された結果の順序情報は, ORDER に保持される。

前掲図2. のサンプル Cnet に対する規則適用の結果は図4に示すものとなる。

規則適用の際, 予定 Cp 上での失敗が起ると, 直前に遡った Cp へ戻って代替規則, 辞書の適用を行う (バックトラック)。また depth-first 探索中, 端節 Cp に至り成功裏に戻る場合も, その途上出る Cp に対し規則適用が行はれる。こうした規則適用は, 既に確定している機能・属性の書き換えは行なわれない。

E-morph は LARGO から出力された C-net から, FEAT, ORDER 属性

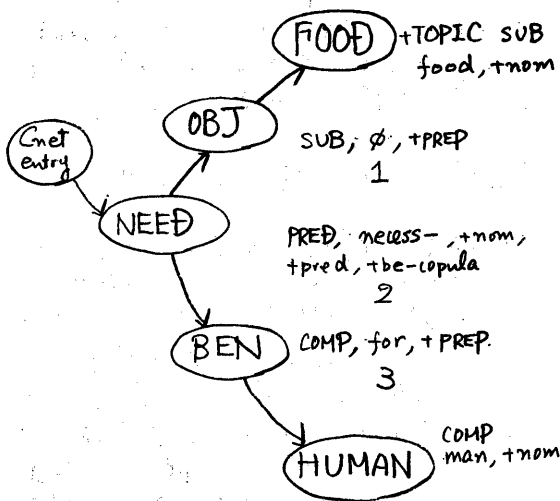


図4. サンプル(図2)に対する属性付与.

を鍵として表現文を生成する。m-ruleは、基本的なFEATを条件に辞書検索して形態を選択する。その後ORDERに従って再整理を実行する。例えば、図4中「NEED」の「be-copula」属性や、一般の助動詞の選択及び順序付けは、m-ruleが担う。

5.1 デバクッション

形態変形等は、LFGやADGの枠組では、それぞれ、F-structure上の文法関係再付規則、L-ruleを容易に解決できる。その他、通常raising, tough-movementなどと言われる構造変形に対してはどう扱うかGPSG, LFGでは述べられていない。

ここでは、raisingを例に採って、ADGの機能について述べる。図5は、日本語文S1, S2に対応するCnet1, 2である。

S1. トムは話し易い。

S2. トムに話すのは容易だ。

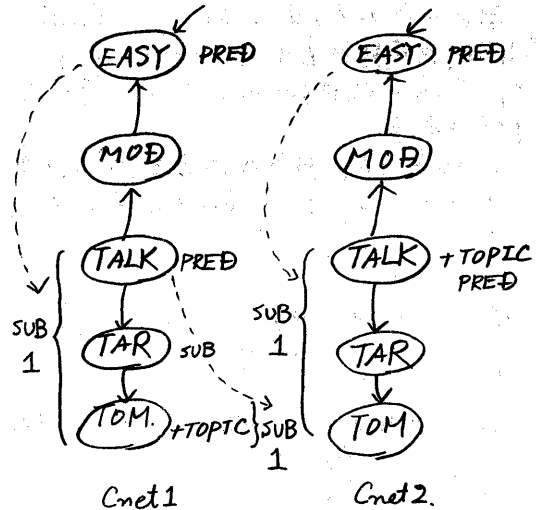


図5. Raisingに対する規則化.

Cnet1, Cnet2の差は「+TOPIC」の位置である。この差違と、Cnet1中の概念依存関係から簡単付規則適用を希望の文法を選択できる。

1. L-ruleにより、「EASY」の「MOD」に「SUB」が与えられる。この「SUB」を選択は、「TALK」の直接デイクメントに「+TOPIC」があることにより可能となる。

これよりCnet2は、「TALK」以下を「EASY」の表現形 easy が指定する補文構造を満足するように規則・辞書が適用される。

2. Cnet1では、補文構造「TALK」以下を生成する際、「TOM」の主観化が行われる。ここでは、「EASY」に関して行うためTALKに対する順序属性を「TOM」に遺伝させ、「TALK」自身は、自身の補文ヘッド easy の補文デイクメントの最後の順序が付けられる(文献[10]参照)。

上述の演算により, 入力文 S_i に対し, E_i が各々得られる。

E1. Tom is easy to talk to.

E2. To talk to Tom is easy.

この現象が扱われるのは, 順序規則を分離し, 橋文属性を直接 Cnet 節の属性としたことによる。 S_i の文が, Cnet のような「TALK」のみのディペンダントとして「EASY」, 「TOM」を表現することも重要である。これによつて, 「EASY」と「TOM」が, 元と「TALK」以下が橋文構造として埋め込まれるも, 概念的には「TALK」のシスターとして容易に情報伝達し合える。

ADG を英文生成の観点から説明した。まずオーキ, 日本語・英語の解析と生成が定式化出来ることを目標に提案した。このモデルに基づいた日本語解析システムと英語文生成システムにより日英翻訳実験を行っている。

翻訳実験が行える基盤モデルは, じめに解析・生成の全2つのレベルを網羅的に規定する必要がある。これに対し, 現在までに提案された文法は, 言語の特定の機能レベルについて述べるだけである。

故に, ADG は, 数多くの文法モデルをひき継いでおり, 各々の文法が個別的に扱えた C-func, P-func, L-func, G-func, S-feat, O-feat を明示的に統合した。

しかし, P-func について, 翻訳系以外では, 予期する TOPIC, FOCUS などが容易に取り出せるかの吟味が必要であるし, ADG に与えられる Cnet エンタの指定も翻訳系以外では文法モデルが設定しなければならぬことである。ADG はこれに対しては今後明確な答を用意したい。これについては,

長入力文を, 部延意図と伝達内容を変更しなくても分割し読み易い英語文生成を行うためにも今後一層検討を要するところである。

参考文献

- [1] S. Klein, 1965 "Automatic paraphrasing in essay format", Mechanical Translation
- [2] R. Quillian, 1968 "Semantic Memory" in Semantic information processing by Minsky, MIT Press
- [3] R. Simmons & J. Slocum, 1972, "Generating English Discourse from semantic networks, CACM 15.
- [4] J. Bresnan, 1982, "The mental representation of grammatical relations", MIT Press
- [5] H. Gaifman, 1965, "Dependency Systems and phrase structure systems", Information and Control, 8.
- [6] G. Gazdar and G. Pullman, 1982, "Generalized phrase structure grammar: a theoretical synopsis", Mimeo, Indiana Univ, linguistic club.
- [7] N. Goldman, 1975, "Conceptual Generation", in Conceptual information processing by R. Schank, North-Holland.
- [8] R. Kowalski, 1974, "Predicate logic as a programming language", IFIP 74, North-Holland
- [9] C. Fillmore, 1968, "The case for case", in Universals in linguistic theory by Bach and Harms, Holt, Rinehart, and Winston
- [10] 柴谷, 1979 "関係文法とは何か", 国語学
- [11] 村下, 石川, 1984, "VENUS: 概念構造からの文生成文法", 情報 28 大会
- [12] 村下, 市山, 1983 "日英機械翻訳システム VENUS の解析 x b = ズム", 情報研報 NL-40-7

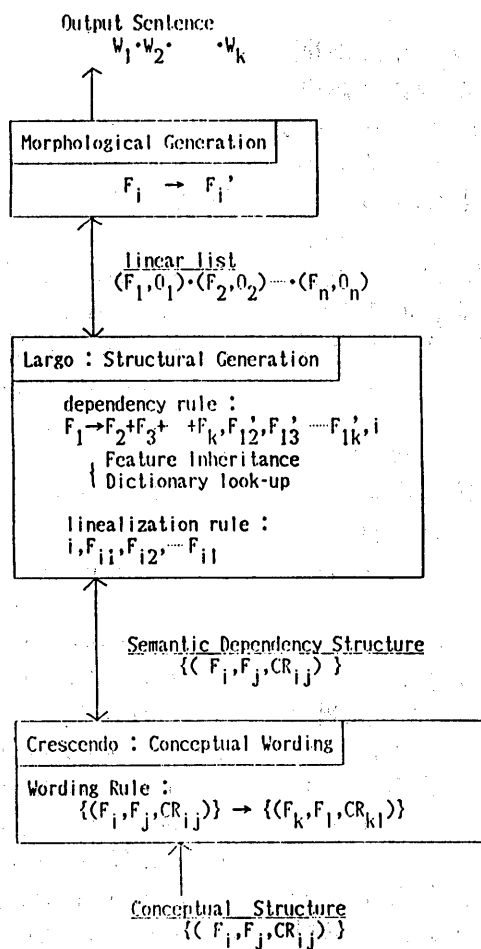


Fig 1. A.D.G. Framework : Synthesis