

形式文法による擬似文の生成

新 田 義 彦

1 はじめに

形式文法 (formal grammar) により, 日本語および英語の擬似文 (pseudo sentence) を記述・生成する思考実験を行った結果について報告する。

ここでいう擬似文とは, 国語文法や英文法が規定するような完結な文, 主語 + 述語 + 修飾語句 + 終止符, Subject + Verb + Object + Period のような文として閉じた形式を必ずしも具備しない文を意味する。典型的には, 詩文にみられるような名詞句だけの言いかけで途切れてしまう文断片, 主語や補語のない宣伝文 (キャッチフレーズ) などである。擬似文であってもコミュニケーションや感懐・感興の伝達などの言語機能は十分に果たせる。

形式文法は句構造文法と言い換えてもよい。句構造文法という概念には, 正規文法, 文脈自由文法そして文脈依存文法, などのクラスが含まれる。処理が単純明快な正規文法を中核に据えて, 所々文脈自由文法を使う方針で思考実験に臨む。

2 文の形式的定義

まず単純素朴な文の生成から思考実験を始める。

句構造文法 (PSG, Phrase Structure Grammar) の一種である, 文脈自由文法 (CFG, Context Free Grammar) の文法規則は, 一般に $A \rightarrow X$ の形式を持つ。ただし, $A \in N$, $N =$ 非終端記号の集合, $X \in (N \cup T)^*$ = 終端記号または非終端記号からなる列。

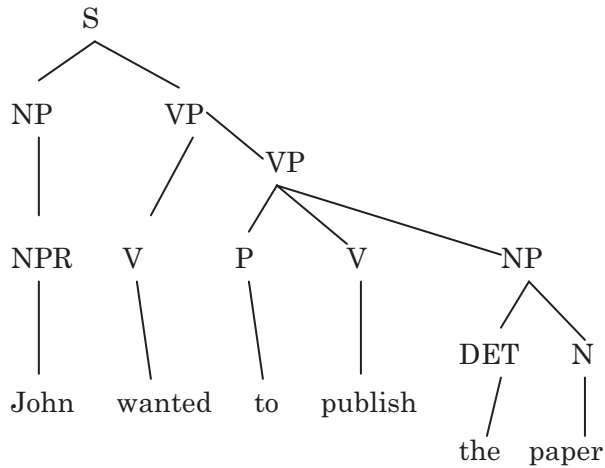
$(N \cup T)^*$ は空記号 (ε) を含む。

これに対して, 文脈依存文法 (CSG, Context Sensitive Grammar) の文法規則は, 一般に $ZAW \rightarrow ZXW$ の形式を持つ。つまり, 非終端記号 A は, 前後を Z と W で挟まれているという文脈 (環境) にあるときに限って, 記号列 X に書き換えられる。ただし, $A \in N$, $N =$ 非終端記号の集合, $X, Z, W \in (N \cup T)^*$ = 終端記号または非終端記号からなる列。 $(N \cup T)^*$ は空記号列 (ε) を含む。

文法規則が, $A \rightarrow a$ または $A \rightarrow aB$ ただし, $A, B \in N$, $a \in T$ の形式のものが, 正規文法。特に $S \rightarrow \varepsilon$, [$S (\in N)$ は初期記号] を追加することもある。 ε は空記号。

・英文とその構文木, およびそれを生成する文脈自由文法 + 文脈依存文法規則の例:

註: 直観的な理解のしやすさに注目して, ここでは, 奥津敬一郎の生成日本文法論 [18] が, 取り扱っている生成規則と超記号に準拠した記法を使った。



S → NP VP

NP → NPR

NP → DET N

VP → V VP

VP → P V NP

NPR → John, Mary, Bill

N → paper, man, cow

V → wanted, tried, publish, meet, eat, published, want

P → to

DET → the, a, an

しかし、このままだと、

John wanted to meet the paper.

や

John published to publish the paper

のような妙な文（非文）が生成されてしまう。これを防ぐには、下記のような文脈依存の規則の追加が必要となる。

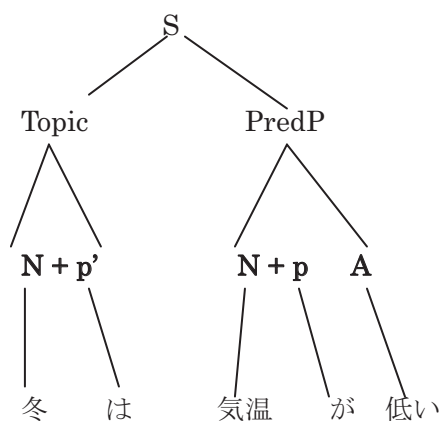
V VP → wanted to VP

V NP → publish NP

V NP → published NP

しかしながら、文脈依存の規則は形式的処理、つまり計算機処理がしにくいので、VPやNPを細分してその組み合わせに制限が導入されるような工夫をする必要がある。

変形生成文法に基礎を置く中島文雄 [24] の統語規則に準拠して、日本語文の生成をするもう一つの例を取り上げる。



$S \rightarrow \text{Topic PredP}$

$S \rightarrow \text{PredW}$

$S \rightarrow \text{PredP}$

$\text{Topic} \rightarrow \text{N} + \text{p}'$

$\text{PredW} \rightarrow (\text{Advl}) \text{PredHW} (+ \text{aux}) (+ \text{p})$

$\text{PredHW} \rightarrow \text{V}$

$\text{PredHW} \rightarrow \text{A}$

$\text{PredHW} \rightarrow \text{AN} + \text{v}$

$\text{PredHW} \rightarrow \text{N} + \text{v}$

$\text{PredP} \rightarrow (\text{Advl}) \text{N} + \text{p} (\text{N} + \text{p}) \text{PredW}$

$\text{N} \rightarrow (\text{Det} +) \text{N}$

$\text{N} \rightarrow \text{N} + \text{P} + \text{N}$

$\text{N} \rightarrow \text{AN} + \text{P} + \text{N}$

$\text{N} \rightarrow \text{MN} + \text{P} + \text{N}$

$\text{N} \rightarrow \text{S} + \text{N}$

$\text{N} \rightarrow \text{n}$

$\text{Aux} \rightarrow (\text{記述は省略})$

n : 形式名詞 文を名詞化する 「こと」「の」, 「の」を準体助詞と呼ぶ文法家もいる.

v : 形式動詞, 「だ」「である」「です」「であります」「でございます」 など.

註 : 形式動詞の定義にはこれと異なる考え方も多いが, ここでは単純にこのように考えることにする.

p : 助詞 +p+ : 連体助詞, +p : 格助詞, -p : 接続助詞, +p : 終助詞,
 および +p' : 副助詞

Det : 限定詞, 「この」, 「その」, 「あの」, 「どの」, 「我が」, 「或る」, 「いわゆる」, 「あらゆる」

註 : 限定詞は、英語文法に準じた用語である。上記の語は、伝統的な国文法では連体詞として定義することが多いが諸説あり、統一見解はない。

以上文献 [24] に準拠して、日本語文の単純な統語規則の素描を与えた。「格」の扱い「文節」, 「自立語」, 「付属語」の扱い, 副助詞「は」と格助詞「が」の区別, 文の依存構造, 入れ子構造などは論じていない。ほんの一部の統語規則しか与えていないことを注意しておく。

3 日本語文法研究の潮流概観

「日本語文法研究の潮流」は本論文のような小さ論文の中の1章で取り扱えるテーマではないが、そのさわりの部分だけをさらりと概観し、本論文で扱う形式文法の能力の限界（いわゆる身の程）を明らかにする一助としたい。

日本語の伝統文法、山田文法と時枝文法における「詞」と「辞」の概念、入れ子構造などの概念を統合して、近代的な数学的文法論への筋道をつけた、国文法学者、故水谷静夫氏の論文を、参照しながら潮流概観をする。水谷論文では、文法論を近代的な代数式に載せオートマトンによる、受理と生成に堪える構造に仕上げる提言と素描で終わっている。

現代の日本語文法研究では、日本語文の構文的多義性、助詞「は」「が」「の」の微妙な使い分けの問題、「格」の存在と不存在の問題などを精緻に掘り下げた研究がなされている。それらの詳細は、文献 [19] [20] [21] [22] [23] などに述べられているが、本論文では現代の文法論にまでは踏み込まない。水谷論文の示唆する範囲、奥津敬一郎 [17] [18] そして中島文雄 [24] などの論考レベル、つまり直観的な生成文法論の範囲で、形式文法を考え、擬似文の生成を試みている。現代の精緻な文法理論に足を踏み入れると、文の生成を頑健な処理として広範に実行することが困難になることを避けたためである。

「助詞」や「格」の精密な扱いをする形式文法とその文生成への適用は今後の課題としたい。

さて、伝統文法における「詞」と「辞」の対立的設定、分別的設定は、次のように約言できる。文構成において、陳述したい内容本体（いわば陳述命題、メッセージの本体）を「詞」といい、文の作成者、発話者の陳述態度や陳述の様相に関わる部分を「辞」として区別する。「辞」は英語文においては、助動詞がこの機能を持つことが多い。

以下、水谷論文「機械処理のための日本語文法」 [25] に沿って概観を進める。

- 日本語文の機械処理つまり形式的処理がしにくいのは、日本語文法の整備が未完のせいばかりでない。日本語側にも問題がある。つまり形式的文法性が弱く複雑な日本語運用の実態にも責任がある。
- 例
 - (1) 会長を私に引き受けろと言うのか。
 - (2) 私に会長を引き受けろと言うのか。

人は自然に [当然のこととして] (1) と (2) を同義と受けとめる。(1) は CFG (2型文法) で簡単に生成あるいは認識ができるが、(2) はそうは問屋が卸さない。陳述の統序関係 (註 : 語句の順

体言に吸収されてしまい、外部に文法的影響を与えないのが普通であるが、ロクナ、ナンノ、ナニ、などは範囲を超えて句末陳述辞を拘束する。

- こういった様な [今まで看過されてきた] 文法現象を扱える形式的文法はまだ存在しない、と水谷は指摘している。機械処理は特殊ルーチンにすべしと主張している。

%否定表現と呼応%

(7) 今まで記号論理を習わなかった。→否定の意味

(8) 今まで二記号論理を習った。

→ 新田の語感では少し不自然な表現か？

(9) 今まで記号論理を習っていた。

∴ 上記3例を観察すると、(7) のマデを陳述副詞にするワケには行かない。

註：ワケとナイの呼応の指摘は新田の追加。

Cf. 今まで (2年間) 記号論理を習った。→ OK

* 今まで記号論理を習った。→ NG

(7)' 今まで2年間記号論理を習わなかった。→否定の意味 →肯定の意味合いも出てくる 今習っている、という意味合いもある。

%論文 [25] のむすび%

- 日本語の単文ですらこんなに複雑な問題がある。
- 自然な感じで日本語文法を形式化するのは、予想外に困難である。ここで形式化とは、機械処理可能にすること、計算機プログラムで記述可能とすることである。生成文法の枠で考えることが重要である。
- 生成文法が強力に働くのは、語順が優先的な構文を持ち、語形変化しない言語に対してではないかと思われる。日本語はこれに該当しない。
- 論文 [25] の指摘範囲において、最難問は、陳述の統序機能が絡む問題であった。
- 「表現の実質的内容を写す部分」および「それに対する表現主体の陳述作用を写す部分」を、区別して論じたが、区別を設けるべきところに [無理矢理] 区別を立てたというそりを受けられるかもしれないという不安もある。
- しかしながら、この区別は意味論上大切な区別である。

%表現内容と陳述作用の区別は重要%

繰り返しになるが、まとめる。

(a) そうダと思わナイ

(b) そうデナイと思う

の区別は、先ほど述べた：「表現の実質的内容を写す部分」および「それに対する表現主体の陳述作用を写す部分」の、区別をしないと説明できない。

註：新田の言語感覚では、下記のように言うのが自然のように感じられる。もちろんニュアンスと「思いの強度」は変化するが。

- (a)′ そうダとハ思わナイ
 (b)′ そうデハナイと思う

%意味の扱いには：公理的集合論の記法，述語論理が有効，必須%

(8) 人は死ぬべき存在だ.

$$\forall x \{x \in H \supset x \in S\}$$

(9) 人は複数的な存在だ.

$$\forall x \exists y \{ (x \in H) \supset (y \in H) \wedge (y \neq x) \}$$

$$\therefore \#(H) \geq 2$$

註：() は 新田が見やすさのため補足した.

ここに，H = 人間の集合，

S = 死を免れぬものの集合

#(H) = 集合Hの濃度（カージナリティ）

注意：(8)と(9)は，構文がひどく似ているのに，意味はすこぶる異なる．この異なりは，記号論理や集合論の概念を使わぬと説明しにくい．

- 先ほどみた意味の相違する文の構文生成は，よほど ad hoc な 深層構造と変換規則でも持ち込まない限り実現しないだろう．（註：一部 新田が，論文 [25] を解釈して表現の変更した）
- 全体として「(日本語の) 機械処理文法，形式的文法」に対して，消極的（≡悲観的）見解を述べてしまったが，[まっとうな] 自然言語研究者としては，そうそう楽天的にはなれない．
 （註：一部 新田が，論文 [25] を解釈して表現変更した）

4 形式文法が扱う構文規則の要件

形式文法の構文規則は，下記の要件を満たす必要がある．この章で扱う構文規則は，文字列，単語列などを，文法の定める品詞タグ列，句構造標識タグ列，命題記号列，意味記号列などの超記号列へと変換するための規則である，これらは一種の数学的なパターン変換記号である．このような理由により，「規則」の代わりに「パターン」という呼び方を使うことにする．

さて，構文規則つまり構文パターンの満たすべき条件は，下記のようにまとめられる．

- 1) 表層的文字列や表層的単語列が記述されているレベルから，品詞タグ列，句構造標識タグ列，命題記号列，意味記号列，などの様々な汎化パターンのレベルまでを包括可能であること．
- 2) 任意の文型パターンは，新規入力文（つまり初見の表層的文字列や表層的単語列）と比較照合可能であること．
- 3) 任意の表層文あるいは他の文型パターンに対する，当該文型パターンの適用可否条件が明確に記述できること．
- 4) 上記比較照合の結果は，一致の程度を定量的に評価可能であること．
- 5) 日本語文，英語文，その他独語文，仏語文，スペイン語文，中国語文，ハンゲル語文，などに対し

て、言語中立的・汎言語的な文型パターン既述が可能であること。

6) 文法的制約, 意味的制約, 統語的制約, 語用論的制約, などの付加的条件や特殊処理(註: デモンプログラム処理など)の追加記述が可能であること。

7) パターン間の一致度や近似度計算が, 効率よくかつ曖昧性を持たずに実行可能であること。

8) 機械処理可能であると同時に, [パターン作成者にとって] 可読性が高いこと。

註: 異言語の文型パターン集合, たとえば日本語文型パターン集合 JPS と英語文型パターン集合 EPS の間において, パターン J (\in JPS) と意味的一致または意味的近似度の高いパターン E (\in EPS) を検索することが, 日英機械翻訳の原理的処理動作となる。たとえば, 英語文の意味的パターンと日本語文の意味的パターンの間の近似度が高いこと, つまり両者の概略一致が検出されると, 適当な語彙インスタンスの補完により, 英語文から日本語文への機械翻訳が実現することとなる。

5 構文パターン記述言語の概要

5.1 構文パターン記述作業の要件

まず構文パターン記述作業を行う際の前提条件について述べる。

- 1) パターン記述の原材料データは, 日本語文と英語文の対訳コーパス, あるいはテキストとその要約テキストの対応コーパスである。本論文では具体例を日英機械翻訳に設定したため, 日本語と英語に重点を置いた論述をするが, 言語系は, 日本語, 英語以外のもの(たとえば独語, 仏語, スペイン語, ポルトガル語, スウェーデン語, 中国語, 韓国語, スワヒリ語, など)であっても構わない。
- 2) 対訳コーパスあるいは要約コーパスを参考としつつ, 言語分析者あるいは自動分析ツールプログラムが, 文型パターンを抽出記述する。
- 3) 文型パターンの原理形態は, “P \rightarrow Q” と書ける。P と Q は一般に文字, 単語, あるいはパターン記号, などからなる有限列(シーケンス)である。
- 4) 入力シーケンスの全体または一部が左辺パターン P に一致したとき, または十分なスコアの近似度を持ったとき, この入力シーケンスは, パターン Q に変換される。
- 5) パターン変換 (Pattern Transduction) “P \rightarrow Q” が, 低レベルパターンから高レベルパターンに向かう変換である場合が, 汎化変換 (Generalization Transduction), 高レベルから低レベルに向かう変換である場合が特化変換 (Specialization Transduction), 同レベルの変換の場合が等価変換 (Equalization Transduction) である。
- 6) 変換は並列的かつ多段階に行われるため, 理論的には非決定性の一致計算過程である。実際上は, パターン一致判定デーモンプログラムおよび, 優先度判定デーモンプログラムにより, 逐次決定的に順序を決めつつ実行される。
- 7) 左辺パターン P と右辺パターン Q の対応は, 一般に多対多 (n 対 m) の関係である。絞込みはパターン一致判定デーモンプログラムがもたらす情報を参考としつつ行う。
- 8) コンピュータ内に知識データベースとして蓄積されるパターン記述において, 直接的に左辺パターン P と右辺パターン Q を対応させておく必要はない。適当なポイントまたはインデクス, あるいはアクセスキーにより対応関係情報を付与しておけばよい。

- 9) 入力文字列または単語列 S が、パターン変換を次々と施され (理論的には、同時並行的に複数のパターン変換を非決定的に施され) た結果、もはや変換できないような終端パターン (一般に文字列または単語列) T に到達したならば、文 S から文 T への翻訳もしくは言い換え・要約などの言語処理が終了したことになる。

Transduction $f: S \rightarrow T$, where f refers to Pattern Database PD

- 10) 入力シーケンスの全体がある単一のパターン P_0 のみに一致する場合は稀である。通常、入力シーケンス全体あるいはその一部分 (註:つまり部分シーケンス) が、何らかのパターン $P_i (i=1, 2, \dots)$ に一致もしくは近似する。また部分シーケンスの切り出し方も多様性や重複性を持つ。したがってパターン一致・近似判定は非決定性の並列処理となるが、これをデーモンが適当な評価情報とヒューリスティックスにより決定性処理として実行制御する。
- 11) 以下の記述においては、左辺パターン P と右辺パターン Q を併記せず、それぞれ別個に記述することとする。対応付けはインデクスもしくは、参照記号により行う。
- 12) パターンの一致判定処理を円滑化するために、パターン記述にいくつかの階層 (レベル) を設ける。一般にパターン一致・近似判定は、低レベルの一致判定から段階的に高レベルの一致・近似判定に上昇するが、低レベル一致と高レベル一致が混在してもよい。このようなレベル混在の制御はデーモンプログラムが行う。

5.2 構文パターン記述言語が定める基本構文

基本構文パターンが扱う構文 (syntax) は下記ようになる。構造定義を与える記述の枠組み、つまり“メタパターンの記述”が「構文パターン記述言語」の骨格である。

- 1) 構文パターンとは、パターン構成要素を1文字以上の空白を区切り記号として並べた有限列である。
- 2) パターン構成要素とは、文字、単語、語句、または記号である。
- 3) パターン構成要素において記号以外はすべて、表層言語の要素である。
- 4) 記号は、何らかの統語的汎化記号、意味的汎化記号、または関数呼び出し、他パターン参照記号、である。
- 5) 記号は、表層的言語要素と区別するため、すべて頭に1個以上の“#”を冠する。
- 6) 汎化のレベルは“#”の個数により表す。つまり“# a ”は第1レベルの汎化記号、“## a ”は第2レベルの汎化記号、“### a ”は第3レベルの汎化記号、 \dots である。“#”の個数が多いほど汎化の程度は大きい。ここで a は任意の文字列であるが、記号の意味が明確となるよう工夫する。
- 7) “#”の代わりに“#1”、“##”の代わりに“#2”、“###”の代わりに“#3”、 \dots などと記してもよい。
- 8) 記号のうち特にデーモン関数は、“# $f a$ ”という構文を持つ。ただし a は任意の文字列である。
- 9) 記号のうち特に意味記述用命題関数または述語関数は、“# $p a$ ”という構文を持つ。ただし a は任意の文字列である。述語関数はその引数 (項、ターム) を括弧の中に収めて表示できる。
- 10) 記号のうち特に他パターン参照記号は、“# $g a$ ”または“# $e a$ ”または“# $s a$ ”という構文を持つ。ただし a は任意の文字列である。“# $g a$ ”は汎化 (上位) パターンへの参照記号であり、“# $e a$ ”は等価 (等位) パターンへの参照記号であり、“# $s a$ ”は特化 (下位) パターンへの参照記号である。

註：パターン記述において，“#”，“##”等は誤解を生じない限り省略した。また他パターン参照記号の代わりに“⇒”により直接被参照パターンを明示・併記した場合もあることを注意しておく。

- 11) パターン記述は、すべて正規表現（註：Chomsky 言語階層の意味で3型言語）として扱われる。したがってパターン記述記号の上位に位置する記号（＝メタ記号）や正規表現構文が使える。
- 12) メタ記号やメタ構文を含むパターン記述は、原理的に有限状態トランスデューサまたは有限状態オートマトンにより、入力記号〔単語〕列（入力シーケンス）の受理、パターン変換、パターン照合、および出力記号〔単語〕列（出力シーケンス）の生成、が行われる。これらの処理の一部を適宜デーモンプログラムが代行する。有限状態オートマトンが受理（処理）できる正規言語は、一般に“A → a または A → aB”という書き換え規則により形式的に定義できる。ここで、A または B は、非終端記号を表し、a は終端記号を表す。

13) 正規言語表現のメタ記号またはメタ構文を、下記のように定める。

① 基本構文：

- ϵ : 空記号列
 c : 通常文字（表層文字）
 ab : 並び
 a* : 繰り返し
 a | b : 選択

② 省略表記：

- . : すべての文字を | で区切って並べたもの
 a+ : aa*
 a [3] : aaa
 a [3, 5] : aaa | aaaa | aaaaa
 a? : ϵ | a
 [a-z] : a ~ z を | で区切って並べたもの
 [^a-z] : a ~ z 以外の文字を | で区切って並べたもの

③ 一般構文：

- ϵ : 空列, 長さ 0 の正規表現
 c : 通常文字, 文字そのもの
 ab : 並び, a と b をこの順に並べたもの
 $\backslash c$: エスケープ, 直後の文字の意味を変更する, 通常文字以外の機能を付与
 . : 任意文字, 任意の 1 文字
 a* : 繰り返し, a のゼロ回以上の繰り返し
 a+ : 繰り返し, a の 1 回以上の繰り返し
 a {n} : 指定回数の繰り返し, a の n 回の繰り返し

$a \{n,\}$: 指定回数の繰り返し, a の n 回以上の繰り返し
$a \{n, m\}$: 指定回数の繰り返し, a の n 回 ~ m 回の繰り返し
$a?$: オプショナルシヨ, a があっても無くてもよい
$[]$: 文字クラス, 並びに含まれる文字
$[^]$: 文字クラス, 並びに含まれない文字
$a b$: 選択, a または b
$()$: グループ/部分式, グループ/部分式を作る
$\yen1 \sim \yen9$: 後方参照/キャプチャ, 部分式を参照する
\wedge	: アンカー, 行頭にマッチ
$\$$: アンカー, 行末にマッチ
$[: :]$: 文字クラス式, 名前に対応する文字クラス
$[..]$: 照合記号, 名前に対応する照合要素
$[= =]$: 等価クラス, 照合要素と等価な文字

- 14) パターン記述の正規言語処理, 特にパターンマッチング処理においては, 適宜, 完全網羅マッチング (greedy matching) と怠慢マッチング (lazy matching) を使い分けるが, その制御はデーモンプログラムが行う。

6 構文パターン記述要素の概略

構文パターンの記述は概略的に言うと, 表層文字, 表層単語, または汎化記号 (統語記号や意味記号), 他パターン参照記号, 特殊処理用デーモン関数呼び出し記号, などのシーケンスである。処理の原則はパターンマッチング処理, つまり有限状態オートマトンによる受理処理や生成処理であるから, パターン (シーケンス) は原則的には正規表現, つまり正規言語の要素となる。しかし部分的には正規言語を越えて文脈自由言語となる場合もある。

日本語構文パターンまたは英語構文型パターンは, 以下に示すような「表層文字」, 「表層単語または表層句」, 「記号」, 「変数」, 「関数」の5種類の要素からなるシーケンスとして定義される。ただし, “変数” および “関数” は〔広義の〕記号の一種とみなせる。広義記号には「他パターン参照記号」または「他パターン参照関数」が含まれる。“関数” は一般に, パターンの照合・変換・生成という機能の一部を分担するデーモンプログラム呼び出しを行う。このゆえに「デーモン関数」と呼ばれることもある。“デーモン” という呼称は “悪魔 (デーモン) のように巧妙に何でもできる” というニュアンスを表現する修辞である。

またパターン記述の一部となっている “関数” が, 自分の所属するパターンの変換・生成・変形などを行なうこともできるという点を強調して, 「マクロ関数」と呼ばれることもある。

またパターンの全体あるいは部分の意味を表現するための記号も, “関数の一種” (注: 命題関数または述語関数) である。

これらの記述要素の並び (つまりシーケンスとしてのパターン) は, その上位構文規則である「正規言語の構文規則」と「正規言語のメタ記号」の支配下に置かれる。

「変数」は汎化処理において、表層語句を適当なパターン記述記号と対応（バインド）させる役割りを担う。さらにまた変数は、“意味記述用の述語表現における項（ターム）を表現する手段”という役割りも担う。

変数 X が他の変数 Y に対して、関数のように振舞うとき、これを “ $X(Y)$ ” と記して、 X あるいは $X(\cdot)$ のことを、「変数関数」と呼ぶ（後述）。

6.1 変数要素

変数は、パターン記述（シーケンス）の部分を表す記号である。表される部分的表現の文法的属性に対応して、単語、句、節を表す変数に大別される。いずれも、アルファベットの大文字と数字（インデクス）で表現される。日英機械翻訳用パターン記述においては、数字（インデクス）は、日本語文型パターン内で出現する変数の順番を表す。

日本語入力文と日本語文型パターンの照合に成功したときは、変数には該当する日本語表現の部分がバインドされる。すなわち、変数にその値として日本語表現が代入（インスタンス）される。また英語文型パターンで使用される変数の場合は、その変数はバインドされている日本語表現の英語訳を意味する。従って、変数を媒介として、日本語文型パターンにおける日本語表現と英語文型パターンにおける英語訳が対応付けられることとなる。

動詞、形容詞、あるいは用言などの活用語に対しては、日本語・英語共通に変数にはその原形がバインドされる。また、英語名詞では、原則として変数は冠詞を含む名詞を意味するものとする。但し、これは、文型パターン記述を頭脳作業として推進する場合の理論的取り決めである。実際の実装計算機処理においては、日本語・英語間で変数による冠詞情報の対応付けは不可能であるので、英文生成の最終段階の形態素調整で冠詞の有無等を決定する。

例) 日本語側：私達→N1, 食べる→V2, 等

英語側：N1 → we, V2 → eat, 等

6.2 関数要素

文型パターン中で、表層文字列と共に使用される。日本語文型パターンで使用された場合は、該当する部分的表現が持つべき形式と条件を示し、入力日本語との照合に成功すれば、引数で使用された変数に該当する表現がバインドされる。また、英語文型パターンで使用されると、該当する部分において関数定義された英語表現が生成される。

関数は、デーモン関数（後述）を除き、いずれも多重合成使用（注：関数の引数として関数を使用すること）を可能とする。また、関数の引数には、文字列、変数、関数を含む表現が使用されるが、表現の属性は、関数毎に定められた属性（「単語」、「句」、「節」のいずれか）を持たなければならない。

日英文型パターンで使用される関数は、「変数関数」、「要素抽出関数」、「文字列関数」、「形式指定関数」、「構文合成関数」、「デーモン関数」の6種類である。それらの形式と機能の概要を以下に示す。関数名はその綴り字を“#f・・・”のように記述することが統語仕様上の取り決めであるが、以下の部分では見やすさのために、“#f”の頭2文字は省略した。表層語句の文字列との混同はないであろう。

1) 変数関数

変数名が関数名として使用される関数で、①動詞、形容詞、形容動詞を名詞に変換するもの、②名詞を動詞に変換するもの、③動詞句、節を名詞句に変換するもの、④名詞句を動詞句に変換するもの、等がある。いずれも英語文型パターンのみで使用され、引数で指定された表現を関数名で指定された文法的属性の表現に変換する。

例) N (V 1) : 動詞 V 1 を名詞化した表現を意味する。

註：英語においてはすべての動詞や形容詞に対して名詞が存在するわけではない。また、その逆の場合も存在するとは限らない。そのためバインドされている引数によっては、この関数は実行不能となる。そのような現象を減少させるため、近い将来、V、N 等の変数に文法的、意味的な制約条件を設け、入力文と日本語文型パターンを照合する段階で、必要な条件を満足しているかどうかを判定する方法を導入する予定である。

2) 文字列関数

関数名として字面を使用した関数で、日本語文型パターンでのみ使用される。引数には、変数が使用され、それによって表される表現が関数名で示される字面を持つことを意味する。但し、原文との照合の結果、引数とする変数には、その字面を除いた表現がバインドされる。

例) 「外は大変暑かったので」は、「past (大変 (CL1)) ので」との照合に成功し、CL 1 := 「外は暑い」がバインドされる。

3) 要素抽出関数

英語文型パターンのみで使用される関数で、引数にバインドされた表現から特定の文要素を取り出す。

例) CL1 = We eat anapple のとき、subj (CL1) = we

註：現時点の仕様では、要素抽出関数には、節の中から、①主語を取り出すものと、②目的語を取り出すものがあるが、②については廃止の方向で検討中である。その理由は以下の通りである。すなわち、節変数によって表される表現の一つは、該当する表現が主語を持っていること（パターン任意要素でも良い）であるため、①の動作は保証される。しかし、目的語のない節もあるので、②の動作は保証されないからである。

4) 形式指定関数

単一の引数を持つ関数で、動詞や動詞句、節を過去形、可能形、否定形、受身形などに変形するなど、引数で指定された表現に対して関数名で指定された構造の表現を示す。

実行結果として、日本語文型パターンで使用された場合、引数には、該当する日本語表現がバインドされる。ここで、該当する日本語表現とは、パターン内の関数に相当する表現に該当する関数の逆関数を適用した結果の表現である。例えば、

例) teiru (V1) を日本語表現「登っている」と照合した結果では、変数名 V1 に、「登っている」の原型「登る」がバインドされる。

英語文型パターンで使用される場合は、引数の変数に対して日本語側でバインドされた日本語表現を英訳した後、関数名で指定された表現を生成する。

注：この関数は、日本語側では、与えられた入力文の中の適用可能な表現要素にもれなく適用される必要がある。その理由は、以下の通りである。

英語パターンでこの関数が使用される場合は、引数に対して英語文型パターン内の関数で変形可能な表現がバインドされていないが、これは、1対1の対訳例文からのパターン化（デフォルトパターン）の場合、すべてのパターン化の条件として保証される。同じ条件が、意味類型化された後、対応関係にあるパターン同士で成り立つためには、日本語文型のパターン化において、形式指定関数の使用可能ところで、もれなく使用されている必要がある。

5) 構文合成関数

複数の引数によって指定された表現を組み合わせ、so that 構文、to for to 構文、so as to 構文など関数名で指定された表現を合成する関数で、英語文形パターンでのみ使用される。この関数で使用される引数には、それぞれの関数で定められた条件を満足する表現がバインドされていることが必要である。

注：英語文形パターンでこの関数を使用する場合は、使用する関数の種類に応じて、対応する日本語文型パターン側の節変数をマクロ関数によって定義し、節の内部構造を指定して使用する必要がある。

6) デーモン（マクロ）関数

変数関数の一種である。引数としての変数（注：つまりカッコの中にある本来の変数）にバインドされる表現の構造を、関数名となっている変数に従って、定義・変形しなおす関数である。この関数は、適用対象の存在するストリング上に、適用対象ストリングと対等の外見で存在するため、マクロ関数とかデーモン関数と呼ばれる。

デーモン（マクロ）関数は、日本語文型パターンでのみ使用される。（注：ただし本研究が日英機械翻訳に特化されているから、このような制約が付与されるのであって、一般のパターンベース自然言語処理では、もちろんこのような制約はない。）対象とする表現が引数で指定された構造を持つ場合、その表現が関数名として使用されている変数の値となる。対象とする表現が引数で指定された構造を持たない場合は、照合に失敗する。

7 形式的構文を抽出するためのコーパスの例

形式的文法を構築するために使用したバイリンガル・コーパス [S. Ikehara et al. 2002] の一部を以下に示す。構文要素記号が一部付与されている。

AN030150-00 1 110

この事を成就するには絶対秘密が肝心だ Absolute secrecy is vital to the success of the scheme.

LJ014982: この事を成就するには絶対秘密が肝心だ。 LE000569: Absolute secrecy is vital to the success of the scheme.

WJ011579: #1 [REN2] /N3 を / 成就するには / #4 [(絶対 | ぜったい)] /N5 が /AJV6.

WE009995: #4 [Absolute] N5 be AJ6 to the success of #1 [AJ2] N3.

PJ085990: VP1 には / #2 [ADV3] /N4 が /AJV5. PE005744: #2 [AJ (ADV3)] N4 be AJ5 to NP (VP1).

AN030154-00 1 110

この縁談は皆賛成だが肝心の本人が不承知だ Everybody approves of the match, all but the party concerned herself, who objects to it.

LJ013389: この縁談は皆賛成だが肝心の本人が不承知だ。 LE014578: Everybody approves of the match, all but the party concerned herself, who objects to it.

WJ008917: #1 [REN2] /N3 は /N4/ 賛成だが / #5 [肝心の] /N6 が / 不承知だ。 WE102726: N4 approve of #1 [AJ2] N3, all but N6 concerned N6. pron. reflex, who objects to N3. pron.

PJ070234: NP1 は /N2/N3. da が / #4 [肝心の] /N5 が / 不承知だ。 PE077015: N2 V (N3) of NP1, all but N5 concerned N5. pron. reflex, who objects to NP1. pron.

AN065115-00 1 110

彼の病気は一時少康を得たが危篤に陥った There was a lull in his illness, which has again taken a dangerous turn.

LJ100438: 彼の病気は一時少康を得たが危篤に陥った。 LE103819: There was a lull in his illness, which has again taken a dangerous turn.

WJ005275: #1 [N2 の] /N3 は / 一時 /N4 を / 得たが / 危篤に / 陥った。 WE115405: There was N4 in #1 [N2. poss] N3, which has again taken a dangerous turn.

PJ075591: NP1 は / 一時 /N2 を / 得たが /VP3. kako. PE104771: There was N2 in NP1, which VP3.

AN065163-00 2 110

好きで勉強するのではない, しょうことなしに勉強するのだ I do not love labour for labour's sake, but I am obliged to do so.

LJ061604: 好きで勉強するのではない, しょうことなしに勉強するのだ。 LE039809: I do not love labour for labour's sake, but I am obliged to do so.

WJ042562: < N1 は > / 好きで / 勉強するのではない, / < N1 は > / しょう / ことなしに / 勉強するのだ。

WE023750: (N1 | I) do not love labour for labour's sake, but (N1 | I). pron be obliged to do so.

PJ007412: < N1 は > /AJV2/ 勉強するのではない, / < N1 は > / しょう / ことなしに / 勉強するのだ。

PE018903: (N1 | I) do not V (AJV2) labour for labour's sake, but (N1 | I). pron be obliged to do so.

AN066148-00 3 110

不景気のために商売を止めた商人がある Some tradesmen have shut up shop owing to the badness of the times.

LJ117240 : 不景気のために商売を止めた商人がある. LE080728 : Some tradesmen have shut up shop owing to the badness of the times.

WJ123292 : 不景気の / ために /N1 を /V2. kako/N3 が / ある. WE109362 : Some N3 V2. pft N1 owing to the badness of the times.

PJ068510 : NP1 の / ために /VP2. kako/N3 が / ある. PE100096 : Some N3 VP2. pft owing to NP1.

AN066149-00 1 100

彼は息子に商売を譲って隠居した He has made over his business to his son, and retired from active life.

LJ107827 : 彼は息子に商売を譲って隠居した. LE021834 : He has made over his business to his son, and retired from active life.

WJ001637 : N1 は /N2 に /N3 を / 譲って /V4. kako. WE002251 : N1 have made over N1. poss N3 to N1. poss N2 , and V4. past.

AN066165-00 1 100

昼間寝て夜働く商売がある Some people sleep by day and work by night.

LJ088779 : 昼間寝て夜働く商売がある. LE000100 : Some people sleep by day and work by night.

WJ109980 : TIME1/V2 て /TIME3/V4/ 商売が / ある. WE109581 : Some people V2 by TIME1 and V4 by TIME3.

AN072197-00 3 110

それが即ち僕のいつも言うことだ That is precisely what I always say.

LJ029131 : それが即ち僕のいつも言うことだ. LE083012 : That is precisely what I always say.

WJ049530 : N1 が /#2 [即ち] /N3 の /#4 [いつも] /V5/ ことだ. WE073312 : N1 be #2 [precisely] what N3 #4 [always] V5.

PJ028480 : N1 が /#2 [即ち] /N3 の /#4 [TIME5] /V6/ ことだ. PE060262 : N1 be #2 [precisely] what N3 #4 [ADV (TIME5)] V6.

AN072203-00 1 110

彼を怒らせれば即ち僕の身の破滅だ His displeasure means my ruin.

LJ111986 : 彼を怒らせれば即ち僕の身の破滅だ. LE001325 : His displeasure means my ruin.

WJ103325 : N1 を / 怒らせれば /#2 [即ち] /N3 の / 身の /N4. da. WE002421 : N1. poss displeasure means N3. poss N4.

PJ059898 : N1 を /V2 せれば /#3 [即ち] /NP4. da. PE003862 : N1. poss N (V2) mean NP4.

AN082707-00 1 100

あれだけ欠点があっても偉いことは偉い With all his faults, he is a great man.

LJ007864: あれだけ欠点があっても偉いことは偉い。 LE000007: With all his faults, he is a great man.

WJ037637: < N1 は > / あれだけ / N2 が / あっても / AJ3 / ことは / 偉い。 WE123630: With all (N1 | I). poss N2 , N1. pron be AJ3 man.

AN082708-00 1 100

あれだけ勉強したのに失敗した He has failed after all his labours.

LJ007869: あれだけ勉強したのに失敗した。 LE000167: He has failed after all his labours.

WJ000959: < N1 は > / あれだけ / 勉強したのに / < N1 は > / V2. kako. WE000299: (N1 | I) V2. pft after all (N1 | I). pron. poss labours.

AN082713-00 1 110

これだけあれば当分間に合います This sum will do for the present.

LJ017951: これだけあれば当分間に合います。 LE000040: This sum will do for the present.

WJ113994: これだけ / あれば / #1 [当分] / 間に合います。 WE116493: This sum will do #1 [for the present].

PJ087780: これだけ / あれば / VP1. teinei. PE105662: This sum will VP1.

AN082714-00 2 110

あればあるだけ欲しくなる The more you have, the more you will want.

LJ008006: あればあるだけ欲しくなる。 LE093926: The more you have, the more you will want.

WJ112799: あれば / あるだけ / < N1 は > / 欲しくなる。 WE112396: The more you have , the more (N1 | you) will want.

PJ087101: あれば / あるだけ / < N1 は > / AJ2. joutaihenka. PE102512: The more you have , the more (N1 | you) will V (AJ2).

AP001632-00 1 110

おべっか者が集まれば悪魔は夕食に出かける。 When flatterers meet, the devil goes to dinner.

LJ009628: おべっか者が集まれば悪魔は夕食に出かける。 LE119487: When flatterers meet, the devil goes to dinner.

WJ054212: N1 が / V2 ば / N3 は / N4 に / V5. WE122123: When N1 V2 , N3 V5 N4.

PJ031573: N1 が / V2 ば / N3 は / VP4. PE004757: When N1 V2 , N3 VP4.

AP001635-00 3 110

危険性のある箇所については、各章で図示されています。 As applicable, potential danger is indicated on the figures shown in the various sections.

LJ054053: 危険性のある箇所については、各章で図示されています。 LE008582: As applicable, potential danger is indicated on the figures shown in the various sections.

WJ067683: N1 の / ある / 箇所については、 / N2 で / 図示されています。 WE044829: As applicable ,

N1 be indicated on the figures shown in N2.

PJ003164 : NP1/VP2. reru. teiru. teinei. PE034416 : As applicable , NP1 be VP2. past.

AP001644-00 1 110

日曜日だったので、そのモデルハウスは見物人でにぎわっていた。 As it was a Sunday, the show home was busy with people having a look.

LJ094511 : 日曜日だったので、そのモデルハウスは見物人でにぎわっていた。 LE008757 : As it was a Sunday, the show home was busy with people having a look.

WJ110283 : TIME1 だったので、/#2 [REN3] /N4 は /見物人で /にぎわっていた。 WE044920 : As it was TIME1 , #2 [AJ3] N4 be. past busy with people having a look.

PJ079272 : TIME1 だったので、/NP2 は /見物人で /にぎわっていた。 PE034504 : As it was TIME1 , NP2 be. past busy with people having a look.

AP001648-00 1 100

今日現在で10以上の文書を作ってデータベースに登録してある。 As of today, I have made over 10 postings of documents into this database.

LJ063840 : 今日現在で10以上の文書を作ってデータベースに登録してある。 LE008846 : As of today, I have made over 10 postings of documents into this database.

WJ110026 : TIME1/ 現在で / < N2 は > /NUM3 以上の /N4 を /V5 て /N6 に /登録してある。 WE044974 : As of N1 , (N2 | I) V5. pft over NUM3 postings of N4 into this N6.

AP001655-00 1 100

サーバから印刷作業を開始すると、緑ランプがすぐに点滅し始める。 As soon as a printing job is started from the server, the green light will start to flash.

LJ040460 : サーバから印刷作業を開始すると、緑ランプがすぐに点滅し始める。 LE008923 : As soon as a printing job is started from the server, the green light will start to flash.

WJ048632 : N1 から /N2 を /V3 と、/N4 が /すぐに /V5. kaishi. WE045045 : As soon as N2 be V3. ed from N1 , N4 will start to V5.

AQ007544-00 3 111

医者は何も心配する理由はありませんと断言してくれた。 The doctor assured us that there were no grounds for alarm.

LJ046732 : 医者は何も心配する理由はありませんと断言してくれた。 LE089315 : The doctor assured us that there were no grounds for alarm.

WJ091578 : N1 は /何も /心配する /N2 は /ありませんと /V3. tekureru. kako. WE067510 : N1 V3. past us that there were no N2 for alarm.

PJ055936 : N1 は /何も /V2/N3 は /ありませんと /V4. tekureru. kako. PE055117 : N1 V4. past us that there were no N3 for N (V2).

CJ001121 : N1 は /CL2. teinei. hitei と /V3. tekureru. kako. CE000832 : N1 V3. past us that CL2.

AQ056364-00 3 110

腹を立てることは、難しい交渉では逆効果を招きかねない。 Losing one's temper can be counterproductive in delicate negotiation.

LJ118620 : 腹を立てることは、難しい交渉では逆効果を招きかねない。 LE064820 : Losing one's temper can be counterproductive in delicate negotiation.

WJ123386 : 腹を / 立てる / ことは、 / #1 [難しい] / N2 では / 逆効果を / 招きかねない。

WE060707 : Losing one's temper can be counterproductive in #1 [delicate] N2.

PJ083205 : VP1 / ことは、 / NP2 では / VP3 かねない。 PE108835 : VP1 can VP3 in NP2.

AQ093621-00 2 110

彼の目はきよろきよろ部屋をながめ、何かないか物色していた。 His eyes restlessly roamed the room, searching for something.

LJ100624 : 彼の目はきよろきよろ部屋をながめ、何かないか物色していた。 LE034245 : His eyes restlessly roamed the room, searching for something.

WJ065510 : N1 の / N2 は / ADV3 / N4 を / ながめ、 / 何か / ないか / V5. teiru. kako. WE093910 : N1. poss N2 ADV3 roamed N4 , V5. ing for something.

PJ069436 : NP1 は / ADV2 / N3 を / ながめ、 / 何か / ないか / V4. teiru. kako. PE084924 : NP1 ADV2 roamed N3 , V4. ing for something.

AQ093631-00 3 110

彼らの視線が背中に突き刺してくるような気がした。 She felt their eyes stab into her back.

LJ110825 : 彼らの視線が背中に突き刺してくるような気がした。 LE075739 : She felt their eyes stab into her back.

WJ027345 : < N1 は > / N2 の / N3 が / N4 に / V5. tekuru. suitei / 気が / した。 WE024269 : (N1 | I) felt N2. poss N3 V5 into (N1 | I). pron. poss N4.

PJ013107 : < N1 は > / NP2 が / VP3. tekuru. suitei / VP4. kako. PE013609 : (N1 | I) VP4. past NP2 VP3.

AQ093640-00 1 110

煙が目にしみて涙が出た。 The smoke made my eyes water.

LJ048791 : 煙が目にしみて涙が出た。 LE098668 : The smoke made my eyes water.

WJ118318 : 煙が / 目に / しみて / < N1 は > / 涙が / 出た。 WE112972 : The smoke made (N1 | I). poss eyes water.

PJ034518 : N1 が / 目に / しみて / < N2 は > / 涙が / 出た。 PE101873 : The V (N1) made (N2 | I). poss eyes water.

AQ093691-00 1 100

彼らの服装が服装なのでイギリス人には給仕と見えた。 The costume gave them the appearance

of waiters to English eyes.

LJ110902 : 彼らの服装が服装なのでイギリス人には給仕と見えた. LE088453 : The costume gave them the appearance of waiters to English eyes.

WJ123158 : 彼らの /N1 が / 服装なので /N2 には /N3 と / 見えた. WE085050 : N1 gave N2. obj the appearance of N3 to N2 eye.

AQ093693-00 3 110

それをヨーロッパ人が見たのはそのときが初めてだった. This was the first time it had been seen by European eyes.

LJ030906 : それをヨーロッパ人が見たのはそのときが初めてだった. LE110688 : This was the first time it had been seen by European eyes.

WJ100248 : N1 を /N2 が /V3. kako/ のは / そのときが / 初めてだった. WE116501 : This was the first time N1 had been V3. ed by N2 eye.

PJ059680 : N1 を /N2 が /V3. kako/ のは / そのときが /NP4. da た. PE105665 : This was NP4 N1 had been V3. ed by N2 eye.

AQ145561-00 5 110

突然ガクンと揺れるのを感じ、やがてすべてのものがガタガタ揺れだした. I felt a sudden lurch and everything started shaking.

LJ093216 : 突然ガクンと揺れるのを感じ、やがてすべてのものがガタガタ揺れだした. LE041024 : I felt a sudden lurch and everything started shaking.

WJ020434 : < N1 は > /#2 [突然] /#3 [ガクンと] / 揺れる / のを /V4, / やがて / すべての / ものが /#5 [ガタガタ] /V6. kako. WE014201 : (N1 | I) V4. past (a | an) #2 [sudden] lurch and everything V6. past.

PJ020366 : < N1 は > /VP2/ のを /V3, / やがて /NP4 が /#5 [ガタガタ] /V6. kako. PE008250 : (N1 | I) V3. past NP (VP2) and NP4 V6. past.

AQ145564-00 1 110

突然ガクンと揺れて機は離陸した. With a sudden lurch the plane left the ground.

LJ093215 : 突然ガクンと揺れて機は離陸した. LE121033 : With a sudden lurch the plane left the ground.

WJ014925 : #1 [突然] /#2 [ガクンと] /V3 て /N4 は /V5. kako. WE123407 : With #1 [sudden] V3 N4 V5. past.

PJ003993 : #1 [ADV2] /#3 [ガクンと] /V4 て /N5 は /V6. kako. PE112939 : With V6.

AQ145569-00 2 110

地下鉄がカーブを曲がったとき彼女がよろめいて私にぶつかった. She lurched into me as the subway went around a bend.

LJ088275 : 地下鉄がカーブを曲がったとき彼女がよろめいて私にぶつかった. LE077260 : She

lurched into me as the subway went around a bend.

WJ051547 : N1 が /N2 を /V3. kako/ とき /N4 が / よろめいて /N5 に / ぶつかった. WE103210 :

N4 lurched into N5. obj as N1 V3. past N2.

PJ031814 : N1 が /VP2. kako/ とき /N3 が / よろめいて /N4 に / ぶつかった. PE082092 : N3

lurched into N4. obj as N1 VP2. past.

AQ145571-00 1 100

私たちのほうへよろめきながら近づいてきた。 He lurched toward us.

LJ068065 : 私たちのほうへよろめきながら近づいてきた。 LE025546 : He lurched toward us.

WJ028615 : < N1 は > /N2 の / ほうへ /V3 ながら / 近づいてきた。 WE013451 : (N1 | I) V3. past toward N2. obj.

AQ145575-00 3 110

富には確かに人を引きつけるものがある。 There is a certain lure in riches.

LJ117474 : 富には確かに人を引きつけるものがある。 LE102705 : There is a certain lure in riches.

WJ062432 : N1 には /#2 [確かに] / 人を / 引きつける / ものが / ある。 WE114107 : There is (a | an) #2 [certain] lure in N1.

PJ037266 : N1 には /#2[ADV3]/VP4/ ものが / ある。 PE103579 : There is (a | an) #2[AJ(ADV3)] NP (VP4) in N1.

AQ145583-00 3 110

太陽を浴びた輝く砂浜は世界中どこでも観光の目玉であるようだ。 Sun-drenched golden beaches seem to be a universal tourist lure.

LJ086489 : 太陽を浴びた輝く砂浜は世界中どこでも観光の目玉であるようだ。 LE081810 : Sun-drenched golden beaches seem to be a universal tourist lure.

WJ122138 : 太陽を / 浴びた / 輝く /N1 は / 世界中 / どこでも /N2 の /N3. da ある (ようだ | 様だ). WE109967 : Sun-drenched golden N1 seem to be a universal N2 N3.

PJ091570 : 太陽を / 浴びた /V1/N2 は /ADV3/ どこでも /N4 の /N5. da あるようだ。 PE100510 : Sun-drenched AJ (V1) N2 seem to be a AJ (ADV3) N4 N5.

AQ145590-00 1 110

誘惑にのってとうとう破産した。 He was lured on to bankruptcy.

LJ123714 : 誘惑にのってとうとう破産した。 LE030673 : He was lured on to bankruptcy.

WJ046201 : < N1 は > / 誘惑に / のって /#2 [とうとう] / 破産した。 WE021183 : (N1 | I) be. past lured on to bankruptcy.

PJ021060 : < N1 は > /VP2 て /#3 [(到頭 | とうとう)] /V4. kako. PE012316 : (N1 | I) VP2. past on to N (V4).

AQ145591-00 1 100

満月の魔力に誘われてつい庭に出てしまった。 The magic of a full moon lured him out into the garden.

LJ121750 : 満月の魔力に誘われてつい庭に出てしまった。 LE093112 : The magic of a full moon lured him out into the garden.

WJ064758 : N1 の /N2 に /V3. reru て /つい /N4 に / 出てしまった。 WE098034 : N2 of N1 V3. past him out into N4.

AQ145595-00 1 110

毎年春と秋には海岸の灯火に誘われて無数の小鳥が死ぬ。 Every spring and fall the lights along the coast lure countless birds to destruction.

LJ121629 : 毎年春と秋には海岸の灯火に誘われて無数の小鳥が死ぬ。 LE014533 : Every spring and fall the lights along the coast lure countless birds to destruction.

WJ123623 : 毎年 /TIME1 と /TIME2 には /N3 の /N4 に /V5. reru て /REN6/N7 が / 死ぬ。

WE047701 : Every N1 and N2 N4 along N3 V5 AJ6 N7 to destruction.

PJ092050 : 毎年 /TIME1 と /TIME2 には /NP3 に /V4. reru て /NP5 が /V6. PE036761 : Every N1 and N2 NP3 V4 NP5 to N (V6).

AQ145597-00 1 110

復讐欲が彼の意識の端に今にも押し寄せてくるかのように潜んでいた。 Desire for revenge lurked threateningly at the edge of his consciousness.

LJ118518 : 復讐欲が彼の意識の端に今にも押し寄せてくるかのように潜んでいた。 LE012830 : Desire for revenge lurked threateningly at the edge of his consciousness.

WJ049137 : N1 が /#2 [N3 の] /N4 の /N5 に /今にも /押し寄せてくるかのように /V6 でいた。

WE071744 : N1 V6. past threateningly at N5 of #2 [N3. poss] N4.

PJ029803 : N1 が /NP2 に /今にも /押し寄せてくるかのように /V3 ていた。 PE053423 : N1 V3. past threateningly at NP2.

AQ145601-00 3 111

あらゆる批評の背後に敵が潜んでいると信じているらしい。 He seems to believe that an enemy lurks behind every criticism.

LJ007691 : あらゆる批評の背後に敵が潜んでいると信じているらしい。 LE027802 : He seems to believe that an enemy lurks behind every criticism.

WJ112679 : あらゆる /N1 の /背後に /N2 が /V3 でいると / < N4 は > /V5. teiru. rashii.

WE036555 : (N4 | I) seem to V5 that N2 V3 behind every N1.

PJ068751 : NP1 の /背後に /N2 が /V3 でいると / < N4 は > /V5. teiru. rashii. PE029751 : (N4 | I) seem to V5 that N2 V3 behind NP1.

CJ000844 : CL1 と / < N2 は > /V3. teiru. rashii. CE000452 : (N2 | I) seem to V3 that CL1.

8 擬似文生成実験

まず、水谷静夫 [27] [28] による擬似俳句の自動生成の先駆的試みについて検討する。次に、前章で示したコーパスから構文パターンを作成し、これを出発点として擬似文を生成する思考実験を行った結果の一部を示す。思考実験では、MS-Excel 上で作動するマクロプログラムを部分的に利用した。

8.1 俳句形式の擬似文の生成

文献 [27] [28] では、Basic Program による擬似俳句の生成を試みた結果を報告している。1979 年ごろ、コンピュータの処理能力が十分高くはなく、日本語処理能力もまだ初期段階であった。カタカナ文からようやく漢字仮名交じり文の入力が可能になった頃である。擬似俳句の生成には、それなりの工夫と苦勞があったことが論文の論調から偲ばれるが、先駆的な擬似文生成の実験として評価できる。

取り上げられている擬似俳句の生成結果を列挙すると下記のようなになる。

文献 [27] から：

花明り人の行末つくづくと
古寺に斧こだまする寒さかな
春の月人をさがして行きつ来つ
汝が墓を訪ひ来て偲ぶ秋の月
わが恋は空の果てなる白百合か
チロチロと物煮るかまど蟬時雨
散る桜みそ汁にほふ村に入る
これやこの流れて果てはシャボン玉
風呂をたく姉さん被り夏の月
冬の月石の地蔵が道しるべ
メーデーや足音去りゆく石だなこみ
野分して森のはつれの水車小屋

文献 [28] から：

秋の月長き坂道空に消ゆ
冬の月異郷に住みてはや四年
春の月石の地蔵が道しるべ
夏の月岸打つ波や松の声
逝く年や俳句を作る蹴算機

けさの初春川遠くうねりけり
秋草や人の行く末つくづくと
鶯に西行庵を尋ね行く
夕桜よそに精だす賃仕事
Paris の秋我が影にさす酒の酔ひ

春雨や君が瞳に燃ゆる夢
 秋の日に昨日のごとく今日もまた
 夏過ぎてまた立ち出づる仮の宿
 初雁の森のはつれの水車小屋
 麦踏みて垣根も結はぬ住まひかな
 長雨を忘れも果つる日々にして
 山桜たつにたたれぬ旅衣
 雲の峯思ひ出もある昔かな
 梅の香にもと大名の隠者とか
 夏祭さればよ銭は欲しきもの

文献 [27] [28] が記述している擬似俳句生成プログラムの概略動作は、ユーザが入力した上句（5音）の季節や句末の文字（や）あるいは体言、用言などの品詞を判定し、データベース内に蓄積されている中句下句（7 + 5 = 12音）を適当に検索して、接続し全体を出力するものである。いわば文字列の選択結合処理である。

今日の自然言語処理の水準からすれば、文字列の選択結合処理は容易い仕事であるが、擬似文生成にオートマトンの考え方、プログラムによる文字列処理が有効であることを示した先駆的工作として高く評価すべきと筆者は思う。

上句（五）と中句下句（五七）の単純結合ではなく、上句（五）を出発点の核として、中句下句（五七）を新たに（本論文が取り上げているような）擬似文生成処理で作ろうとすると、相当に難しい課題である。俳句の満たすべき条件を取り込んだ俳句作成文法などを開発し実装する必要がある。また、歳時記の機能を持つ俳句レキシコン（あるいは俳句オントロジー）の開発実装も必要になる。この問題は、擬似俳句の形式的生成の問題として次期の研究論文で取り上げたい。

8.2 コーパス文の再生に接近する擬似文の生成

前章で示したコーパスから構文パターンを作成し、これを出発点として擬似文を生成する思考実験を行った結果の一部を以下に示す。

AN030150-00

この事を成就するにはには絶対秘密が肝心だ

S → Advl + S2

Advl → S-p → S1 には ; S1 = この事を成就する ; S2 = 絶対秘密が肝心だ

註:上記2つの構文パターンが蓄積されていれば、擬似文生成ができる。構文パターンは、原則的に“→”によるプロダクション（導出）により記述される。構文パターンの導出の中途を省略して、直接的に表層単語列を示す場合には、“=”を使った。

AN030154-00

この縁談は皆賛成だが肝心の本人が不承知だ

S → Topic + Advl + S2

Topic → N は = この縁談は

Advl → S-p → S1 だが; S1 = 皆賛成; S2 = 肝心の本人が不承知だ

AN065115-00

彼の病気は一時少康を得たが危篤に陥った

S → Topic + Advl + S2

Topic → N は = 彼の病気は

Advl → S-p → S1 が; S1 = 一時少康を得た; S2 = 危篤に陥った

AN065163-00

好きで勉強するのではない, しようことなしに勉強するのだ

S → S1 + “-p” + S2

“-p” → のではない, (註: 接続助詞の概念の拡大適用)

S1 → Advl1 + V1; Advl1 → V で = 好きで; V1 = 勉強する

S2 → Advl2 + V2; Advl2 → S-p → Sに; S = しようことなし;

V2 → V1 + p2; p2 → のだ

AN066148-00

不景気のために商売を止めた商人がある

S → N + p + V; p → が; V → ある

N → S1 + N2 ; N2 → 商人

S1 → Advl + S2

Advl → N3 + “のために”

S2 = 商売を止めた

AN066149-00

彼は息子に商売を譲って隠居した

S → Topic + Advl + S2

Topic → N は = 彼は

Advl → S-p → S1 て; S1 = 息子に商売を譲る; S2 = 隠居した

AN066165-00

昼間寝て夜働く商売がある

S → Advl + S2

Advl → S-p → S1 て; S1 = 昼間寝る; S2 = 商売がある

AN072197-00

それが即ち僕のいつも言うことだ

S → Topic + S1

Topic → N が即ち = それが即ち 註： 提題概念 (Topic) の拡大版である

S1 → N1 + “の” + N2 + “だ”

N1 → 僕

N2 → V + “こと”

V → Adv + V1; Adv → いつも; V1 → 言う

(以下 擬似文生成の例示は略す)

擬似文生成用の形式文法の構成にあたっては、名詞 (句)、動詞 (句) などの語彙情報の整備と同時に、副詞句 (Adv1) の概念を拡大して、多種多様な語句を構文パターンとして収集整備することが、肝要であることが判明した。この拡大概念は、伝統的国語国文法で行う「名詞句 + (種々の) 助詞類」という要素的分析法を放棄して、直截に副詞的機能を持つ語句を抽出する分析法を意味する。言語学的には、異論が多い短絡的な方法に見えるが実際の効果や効用は極めて大きいことが分かった。

9 おわりに

基本構文のパターンを形式的に記述する実験フェーズを終了し、その概要について述べた。基本構文パターンに、具体的な表層単語を代入することが形式的文生成の基本動作である。

現状の水準では妥当な文と同時に、形式は整っているが意味のない文 (非文) も大量に生成される。

たとえば、「岡本君は A 高校で数学を教えている」の生成と同時に、「A 高校は数学で岡本君を教えている」や「数学は岡本君で A 高校を教えている」などの生成を許してしまう。もちろんこの例の場合には、生成を抑止する規則の構築は容易であるが、このような例が広範な語彙分布空間で発生することを抑止できるような、簡潔かつ完備な形式的規則の構築は相当に面倒である。

非文の生成を強力に抑制するための形式的規則の体系を構築することを次の課題としたい。

参考文献

- [1] L. Bentivogli, and E. Pianta, “Exploiting Parallel Texts in the Creation of Multilingual Semantically Annotated Resources: the MultiSemiCor Corpus”, *Natural Language Engineering*, Vol.11, No.3 (2005) pp.247-261
- [2] K. Church, I. Dagan, W. Gale, P. Fung, B. Satish and J. HELFMAN, “Aligning Parallel Texts: Do Methods Developed for English French Generalize to Asian Languages?” *Proceedings of the Pacific Asia Conference on Formal and Computational Linguistics* (1993)
- [3] S. Ikehara et al. “Semantically Equivalent Language Transformation Method Based on Analogical Thinking Principle”, (in Japanese), *Journal of Artificial Intelligence Society of Japan* (2002)
Also in : 電子情報通信学会技術研究報告. TL, 思考と言語 102 (491), 7-12, 2002-11-29
- [4] A. Kinyon, “A Language-Independent Shallow-Parser Compiler”, *Proc. 39th ACL Ann. Meeting (European Chapter)* (2001) pp.322-329
- [5] ホップクロフト&ウルマン著 (野崎, 他訳, 言語理論とオートマトン, サイエンス社 (刊) (1971)
- [6] E. Macklovitch and H. Marie-Louise, “Line ‘em up: Advances in Alignment Tecnology and Their Impact on Translation Support Tools”, *AMTA* (1996) pp145-156
- [7] R. Mihalcea, and M. Simard, “Parallel Texts”. *Natural Language Engineering* Vo.11, No.3 (2005) pp.239-246
- [8] J. Munday, *Introducing Translation Studies*, Taylor & Francis Group (2009)
- [9] Y. Nitta, “Idiosyncratic Gap: A Tough Problem to Machine Translation”, *Proc. Comp. Linguistics, COLING’86* ACL (Assoc. Comp. Ling.) (1986)

- [10] Y. Nitta, "Problems of Machine Translation: From a Viewpoint of Logical Semantics", *Economic Review of Nihon University*. Vol.72, No.2, Nihon University, Tokyo: (2002) pp.23-42
- [11] Y. Nitta, "The Utility and Problem of Insufficient Machine Translation", *Economic Review of Nihon University*. Vol.80, No.4 (2001) pp.1-54
- [12] A. Pim, *Exploring Translation Theories*, Routledge, Taylor & Francis Group (2010)
- [13] M. Saraki and Y. Nitta, "The Semantic Classification of Verb Conjunction in the "Shite" Form", *Proceedings of Spring IECEI Conference*, IECEI Japan (2005)
- [14] M. Saraki, ed. and Y. Nitta, "Regular Expression and Text Mining (in Japanese) Second Printing", Akashi-Shoten (2008) 312p
- [15] A. G. William and K. W. Church, "A Program for Aligning Sentences in Bilingual Corpora", *Computational Linguistics*, Vol.19, No.3 (1993) pp.75-102
- [16] Hangeveld (Editor), Simon C. Dik: *The Theory of Functional Grammar 1&2*, FGS 20, Mouton de Gruyter (1997)
- [17] 奥津敬一郎, 「ボクハ ウナギダ」の文法, くろしお出版 (1978)
- [18] 奥津敬一郎, 生成日本文法論, 大修館書店 (1974)
- [19] 仁田義雄, 日本語文法研究序説, くろしお出版 (1997)
- [20] 石綿敏雄, 現代言語理論と格, ひつじ書房 (1999)
- [21] 三原健一, 日本語の統語構造, 松柏社 (1894)
- [22] 三原健一, 構造から見る日本語文法, 開拓社 (2008)
- [23] 森山卓郎, ここからはじまる日本語文法, ひつじ書房 (2000)
- [24] 中島文雄, 日本語の構造, 岩波新書 373, 岩波書店 (1987)
- [25] 水谷静夫, 機械処理のための日本語文法, 電気学会雑誌 V.93. N11 (1972-11)
- [26] 水谷静夫, 構文解析試論, NII-Elec. Lib. Service, 東京女子大学論集 (水谷静夫氏の東京女子大学助教授時代 発行年不知)
- [27] 水谷静夫, 俳句を作る計算機, *ibid.* 東京女子大学論集 (1979-5)
- [28] 水谷静夫, 再び俳句を作る計算機, *ibid.* 東京女子大学論集 (1979-5 以降 発行年月不知)