岐阜経済大学論集 33 巻 4 号（2000 年）

〔研究ノート〕

ドイツ語の機械解析に向けて（3・完）[1]

山 田 善 久

0. 序
1. 自然言語処理における文解析のプロセス
2. ドイツ語の形態素解析
  2. 1. 人称変化詞（動詞類）の形態素規則
  2. 2. 格変化詞（名詞類）の形態素規則
3. Kongruenz の問題 ……（以上，第 22 巻第 2・3 号）
4. ドイツ語の構文解析
  4. 1. 基底句構造の構想
  4. 2. 不連続要素の処理
  4. 3. 自由語順の処理と構文規則 ……（以上，第 28 巻第 4 号）
5. 素性構造理論と辞書
  5. 1. 文脈自由文法と単一化文法
  5. 2. 単一化と素性に基づく文法理論
  5. 3. 言語情報としての記号
  5. 4. HPSG による辞書記述
  5. 5. 制約としての文法規則
6. 結　　語 ……（以上，本号）

5. 素性構造理論と辞書

自然言語を処理する際に用いる必須の道具立ては，一般に，文法と辞書である。これは機械処理の場合に限らず，我々が外国語を学習する際にも，同じような手立てを考える。前回までは，この文法と辞書のうち，もっぱら文

法の処理の問題を扱った。今回は主に辞書の問題を取り上げることにする。ただし単一化（unification）という手法を駆使する最近の文法理論では，文法と辞書の境界が，かなりあいまいになって来ているという側面がある。従来，文法（統語）規則として導入されていたものを，辞書の語彙項目の記述に移し替え，文法をできるだけシンプルにしよう，というのが理論の一般的な流れだからである。とくに今回ここで取り上げるHPSG（Head-driven Phrase Structure Grammar＝主辞駆動句構造文法）は，こうした傾向の中での最先端の理論と言うことができる。

そこで今回は，前回までに導入した構文規則をそのまま使用することは止め，あらためて，文法と辞書の関係を理論的に整理した上で，新しく機械へのインプリメントの問題を考えてみる。

5.1. 文脈自由文法と単一化文法

前回までに扱ってきた構文解析の規則は，原理的には，CFG（Context Free Grammar＝文脈自由文法）という形式文法システムに基づくものである。例えば，

(1)　Der Musiker spielt Geige.（その音楽家はヴァイオリンを弾く）

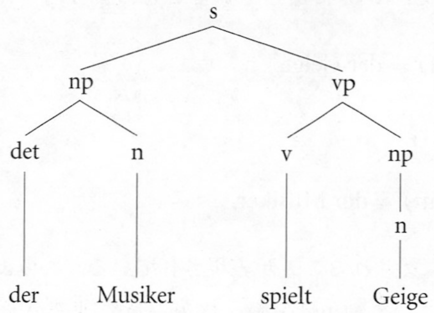
という文を解析するには，次の(2)のようなCFGが必要である。

(2)
```
s-->np,vp.
np-->det,noun.
np-->noun.
vp-->v,np.
det-->[der].
noun-->['Musiker'].
noun-->['Geige'].
v-->[spielt].
```

ちなみに(2)は，プログラミング言語Prologの文法記述形式であるDCG (Definite Clause Grammar＝限定節文法）の書式[2]にしたがった書き方であり，これもCFGの一種である。

(2)の各ルールの意味するところは，「矢印の左辺の記号を右辺の記号（列）に書き換えよ」ということで，始発のs（文）から規則が順次適用されて，最終的には(3)に見られるようなよく知られた木構造が生成される。

(3)



ところで，文法(2)は，規則の適用の仕方によっては，次の(4)のような非文法的な文が生成される可能性がある。

(4)　*Der Geige spielt Musiker.

(4)は「ヴァイオリンが音楽家を演奏する」という意味的な容認不可能性もあるが，もっと致命的なのは，女性名詞1格主語のGeigeが定冠詞derと結びついている非文法性である。DCGでは，以下の文法(5)に見られるように，

(5)
```
np-->det(GENUS),noun(GENUS).
np-->noun(GENUS).
det(m)-->[der].
noun(m)-->['Musiker'].
noun(f)-->['Geige'].
```

引数を追加することによって性の一致のような現象を明示的に表現でき，(4)のような非文法的文を排除することができるようになっている。

文法(5)は，文法(2)にあるカテゴリー det および noun に，文法的性を指定する引数を1つ追加したものである。1行目と2行目の規則の右辺に見られる GENUS がそれである。この GENUS は変項（variable）で[3]，ここでは m（男性）または f（女性）を値に取り得る。1行目の det および noun の引数が，同一名（つまり GENUS）になっていることにより，

det(m),noun(f) = der Geige

の可能性が排除され，

det(m),noun(m) = der Musiker

だけがフィルタリングされることが表現されている。このように DCG では，引数の追加によって，名詞の性・数・格の一致，動詞の人称・数の一致などを扱えるようになっている。またこれにより，規則の中の文法カテゴリーを大幅に削減することができるという効果もあり，かなり強力な文法となっていると言える。

しかしながら，こうした DCG による拡張にもかかわらず，一般に CFG に基づく構文解析には，克服できない難点があるように思われる。それは，文法規則の数の問題である。例えば，先の規則(2)において，np（名詞句）を展開する規則に，np-->det,noun と np-->noun の2つがあるが，ドイツ語名詞句の内部構造の可能性を考慮すると，もちろんこれだけで済むわけではない。付加語形容詞を含んだ構造（これには冠詞が先行する場合と無冠詞の場合とがある），2格や前置詞句による後置修飾語をともなう構造，関係文や同格をともなう構造，さらにこれらを複合的に含む構造，名詞句と名詞句の等位構造，付加語形容詞や2格および前置詞付加語などの繰り返しの可能性など，頭に浮かぶものを列挙していっても，名詞句の内部構造には，多種多様の可

能性があり，これらをすべてカバーするためには，かなりの数のルールが必要となるだろう。同じことは他の句構造についても言え，とくにvp（動詞句）は，文法(2)では他動詞についてのルール1つだけをあげているが，文形成の中核となる構造であり，他の句構造と比べて格段に複雑な規則が必要になることは，容易に推測できる。実際，実用的な自然言語処理システムを組むには，数百から数千オーダーの構文規則が必要であると言われている。

CFGは，計算機科学の分野では，いわゆるチョムスキー階層[4)]と呼ばれる形式文法システムの階層のうち，統語解析に適した文法として，プログラミング言語，とくにコンパイラの設計において，標準的に使われているということであるが，複雑多岐な自然言語にはそもそも適していないのではないか。人工言語と自然言語の間には，多少誇張めいて聞こえるかも知れないが，水槽と海ほどのスケールの違いがあるのではないか。そういう疑念が湧いてくる。

数の多少よりももっと本質的なことは，ある言語の文法において，構文規則の数がいったいいくつあるのか，その上限を決定できないことである。これを究めた理論的アプローチも実際的システムもいまだ存在しない。規則体系を閉じた集合として規定できない，もしくは規定することが困難ということは，理論的には相当深刻な弱点のように思われる。CFGをベースにして理論を組み立ててきた生成文法の命題の1つは，「有限の規則から無限の文を生み出す規則体系」の提示ということであるが，文法を有限の規則として明示的に規定できない以上，この命題は空理空論に過ぎなくなってしまう。

冒頭でふれたように，自然言語処理においての基本的な道具立ては文法と辞書であるが，このうち辞書は，有限として規定することができないのは明らかなので，上記の命題を実現するためには，どうしても文法の方を，有限として規定する必要がある。

この問題の解決策として，1960年代の生成文法初期の段階で，チョムスキーによって提案されたのは，変形という統語レベルの設定である[5)]。変形

規則によって，CFG の規則で生成される文の集合を制限することができ，有限の規則体系として文法を規定することが可能であると期待された。ところが研究が進展していくうちに，個々の文法現象ごとにさまざまな変形規則が提案され，中にはその場限りの場当たり的なものも現れるに及び，理論的妥当性を保持することができないほどの状況に立ち至った。

そんな中で出てきたのが，制約（constraints）という考え方である。1970 年代に入ると，この線に沿って，変形に対する制約についての研究が盛んに行われるようになった。1980 年代初頭の統率・束縛理論（theory of government and binding＝GB 理論）[6]は，その集大成と位置づけることができる。しかしこの GB 理論は，変形規則が移動変形（Move-α）1 つに制限されてはいるものの，依然として変形という記述のレベルを認める変形文法理論であった。

1980 年代に入ると，制約という考え方をさらに進めて，すべての変形的操作を制約として記述しようという理論，つまり変形のない生成文法が登場してくる。語彙機能文法（Lexical Functional Grammar＝LFG）[7]，一般化句構造文法（Generalized Phrase Structure Grammar＝GPSG）[8]，そしてここで取り上げる主辞駆動句構造文法（Head-driven Phrase Structure Grammar＝HPSG）[9]などがそれである。

これらの諸文法は，総称して単一化文法（unification grammar）と呼ばれることがあるように，単一化の操作を文法の基本原理としている点で一致している。この単一化については次節で詳しくふれることになるが，概略，素性（そせい）(feature）を使って文法機能を記述し，文の構成要素間の関係をこの素性の単一化として表す文法のことを言う。このような文法では，各辞書項目ごとの素性の記述，つまり辞書（lexicon）の記述に重点が移行することになり，文法規則は，限られた数の抽象的な句構造式型が，適格な句構造をチェックする制約として作動するかたちになる。ここに至って，ようやく文法の有限性への道筋が見えてきたように思われる。

5.2. 単一化と素性に基づく文法理論

単一化（unification）という言葉は，本来は数学の定理証明（theorem proving）分野の用語だそうであるが[10)]，本稿の読者の多くにとっては，プログラミング言語 Prolog の基本動作原理としての用語の方が，よりなじみ深いものと思われる。Prolog における単一化とは，2つのパターン（構造）を，出現する変数に適当な値を代入することによって，1つのパターンにすることである。例えば，

(6)	human('Klaudia', female).	%Klaudia は女性(female)である.
	human('Michael', male).	%Michael は男性(male)である.

という Prolog の節で表現された事実の集合があるとする。これに対して，

?-human('Klaudia', X).　　　%Klaudia は何(X)とマッチするか？

と質問して，プログラムを実行させると，Prolog インタープリタは次のような結果を返す。

X = female　　　%女性(female)とマッチする
yes　　　%単一化成功

つまりこの場合，質問の述語中の変項 X に female を代入すると，与えられた事実の節と同一のパターンをつくることができるので，単一化が成功したという。一方，

?-human('Iris', X).　　　%Iris は何(X)とマッチするか？

という質問に対しては，結果は次のようになる。

no　　　%単一化失敗

つまり，与えられた(6)の事実中に，述語 human の左辺が 'Iris' である節が見

つからないので，同一のパターンをつくることができず，単一化は失敗する。

さてLFG, GPSG, HPSGなど一連の単一化文法理論は，多かれ少なかれ計算機科学の影響下に開発されてきた経緯があるところから，そこでの単一化という操作も，以上見たPrologの単一化の動作とアナロジー的な関係があることは当然予想される。ただしそこには微妙な違いもある。Prologの単一化は，パターンの照合，要するにパターンマッチングのことであるのに対して，単一化文法における単一化は，後で見るように，むしろパターンの融合（マージ）という性格を持っている。「一群の，互いに矛盾しない構造から，各々の構造に含まれる情報をすべて含み他には何も含まないような構造を作り出す」[11]というわけである。そしてそのパターンとは，さまざまな言語的特性を素性の行列で表現した素性構造（feature structure）と言われるものである。以下では，HPSGの素性構造，および単一化プロセスの概略を，ごく簡単なドイツ語文(7)を例に，スケッチしてみよう。

(7)　Iris lächelt.（イーリスは微笑む）

(7)の構成要素Irisおよびlächeltは，HPSGでは，それぞれ(8)および(9)のような素性構造として表現される。この構造は記号（sign）と呼ばれる。

(8)(9)の素性構造の読み方を概略しておこう。まず各記号の左辺上にあるPHONは，音韻（phonology）のことで，ここで記号の音形が指定される。ここではアルファベット文字（Iris, laechelt）で簡略な表記がされているが，厳密には音韻論で用いられているような弁別的素性（distinctive feature）[12]の束を連ねた構造になる。

(8)

$$\begin{bmatrix} \text{PHON} & \text{Iris} \\ \text{SYN|LOC} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{MAJ} & \text{N} \\ \text{AGR} & \text{3S} \end{bmatrix} \\ \text{SUBCAT} & \langle\, \rangle \end{bmatrix} \\ \text{SEM|CONT} & \text{IRIS} \end{bmatrix}$$

(9)

$$
\begin{bmatrix}
\text{PHON} & \text{laechelt} \\
\text{SYN|LOC} & \begin{bmatrix}
\text{HEAD} & \begin{bmatrix} \text{MAJ} & \text{V} \\ \text{VFORM} & \text{FIN} \end{bmatrix} \\
\text{SUBCAT} & \left\langle \begin{bmatrix}
\text{SYN|LOC} & \begin{bmatrix}
\text{HEAD} & \begin{bmatrix} \text{MAJ} & \text{N} \\ \text{CASE} & \text{NOM} \\ \text{AGR} & \text{3S} \end{bmatrix} \\
\text{SUBCAT} & \langle\,\rangle
\end{bmatrix} \\
\text{SEM|CONT} & \boxed{1}
\end{bmatrix} \right\rangle
\end{bmatrix} \\
\text{SEM|CONT} & \begin{bmatrix} \text{RELN} & \text{LAECHELN} \\ \text{LAECHLER} & \boxed{1} \end{bmatrix}
\end{bmatrix}
$$

PHONの下にあるSYNは，統語（syntax）素性で，記号と記号との結合可能性が指定される。これは右辺に向かってさらに下位区分される。SYNの右隣のLOCは局所（local）素性であり，品詞，屈折，格，結合価など記号の内在的な統語性質を指定する。(8)(9)には現れないが，これと同じレベルにあるのが束縛（binding）素性で，これは関係代名詞や疑問表現などの非局所的な長距離依存関係を指定する。LOCはさらにその右隣にあるHEAD（主辞素性）とSUBCAT（下位範疇化素性）に区分される。この他，ここでは煩雑になるため省略しているが，第3の素性として，語彙的範疇と非語彙的範疇を区別するのに用いられるLEXがある。HEAD（主辞素性）にはさまざまなものがあるが，上の構造では，MAJ, CASE, AGR, VFORMが使われている。MAJは主要（major）素性のことで，V（動詞），N（名詞）などの品詞を指定する。CASEは文字通りNOM（主格），ACC（対格）などの格を指定する。AGRは動詞の人称・数など，一致（agreement）に関する情報である。VFORMは，動詞の定形（FIN），不定形（INF），過去分詞形（PSP）など動詞の屈折を指定する。SUBCAT（下位範疇化素性）は，動詞がどのような要素と結びつくか，その数と種類を指定する。例えば，lächelnのような要素1つ（つまり主語）だけで文を充足する自動詞は(10)，主語と4格目的語の2つの

要素が必要な他動詞は(11)のように表記される。

(10)　SUBCAT<NP[NOM]>

(11)　SUBCAT<NP[ACC],NP[NOM]>

この(10)(11)は，簡略化した表記法で，煩雑さを避けるためによく使われる書き方であるが，正式には(9)の SUBCAT の < > の中の記述（=次の(12)）のように素性構造を使って表現する必要がある。これは先の音韻の表記において，素性構造の代わりにアルファベット文字を使ったのと同じ事情である。つまり素性構造(12)は，(10)の NP[NOM] と同値であり，いずれも1格名詞句を表している。

(12)
$$\text{SUBCAT}\left\langle \begin{bmatrix} \text{SYN|LOC} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix}\text{MAJ} & \text{N}\\ \text{CASE} & \text{NOM}\\ \text{AGR} & \text{3S}\end{bmatrix}\\ \text{SUBCAT} & \langle\,\rangle \end{bmatrix} \\ \text{SEM|CONT} & \boxed{1}\end{bmatrix}\right\rangle$$

このように，下位範疇化素性により，主辞（head）が要求する構成要素の数と種類が記号ごとに明示されているので，HPSG では，従来の句構造規則のような文生成のための表示レベルを別個に持つ必要がなくなった。それに代わって文法（統語）規則としては，数個の抽象的な規則だけを設け，これが記号間の構成関係に対する制約（constraints）として働く仕組みになっている。

さて記号(8)(9)のそれぞれの最下行左端は，SEM|CONT であるが，これは意味・内容の属性を示す。本稿では，意味論の分野にまで立ち入ることはほとんどないが，概略すると，HPSG では，意味の表記に状況意味論（situation semantics）[13]を援用し，世界を個体（individuals）と関係（relations）とに即して分析する。大雑把に言うと，名詞的記号の意味は「個体」であり，動詞的記号の意味は「関係」となる。例えば記号(9)の場合，SEM|CONT の価から，その意味は LAECHELN という関係であり，タグ $\boxed{1}$ がその関与者を

示している。ちなみにこのタグ $\boxed{1}$ は，その上の SUBCAT 素性中の要素（つまり主語になる名詞）の SEM|CONT 素性の価（$\boxed{1}$）と同一の指標が付けられているので，結果として「文の主語になる個体が微笑む」という関係を意味していることになる。

素性構造についての説明が，少し長くなってしまったが，ここで HPSG における単一化の働き方を見ておこう。まず，先にふれた制約として働く文法規則の 1 つに，「句は，非語彙的主辞と 1 個の補助部（complement）からなる」という規則がある。形式化すると，概略(13)のようになる。

(13)　「SUBCAT < >」 → H[LEX −],C

これは，伝統的な句構造規則の S → NP VP や NP → Det Nom などの規則を包括し，これらにとって代わるものであるが，記号(8)と(9)の結合も文(S)を作るものなので，この規則の制約を受ける。まず記号(9)が規則(13)の H[LEX −]（主辞＝動詞句）と対応付けられ，次に(9)の SUBCAT 中の要素を取り出し，この統語素性が(8)の統語素性と比較照合される。これらの素性構造はそれぞれ(14)(15)の形をしている。この場合，互いに矛盾する素性の値を含んでいないので，単一化が可能となる。したがって(8)と規則(13)の C（補助部＝文の主語）を対応付けることができる。

(14)　(8)の統語素性

$$\text{SYN|LOC}\begin{bmatrix}\text{HEAD} & \begin{bmatrix}\text{MAJ} & \text{N}\\ \text{AGR} & \text{3S}\end{bmatrix}\\ \text{SUBCAT} & \langle\,\rangle\end{bmatrix}$$

(15)　(9)の SUBCAT 中の要素（NP）の統語素性

$$\text{SYN|LOC}\begin{bmatrix}\text{HEAD} & \begin{bmatrix}\text{MAJ} & \text{N}\\ \text{CASE} & \text{NOM}\\ \text{AGR} & \text{3S}\end{bmatrix}\\ \text{SUBCAT} & \langle\,\rangle\end{bmatrix}$$

(16)

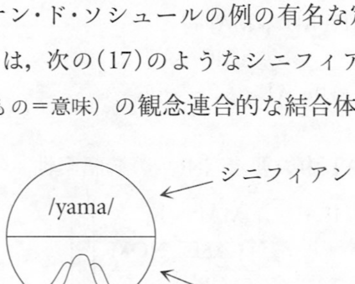
$$\begin{bmatrix} \text{PHON} & \text{Iris laechelt} \\ \text{SYN|LOC} & \begin{bmatrix} \text{HEAD} & \boxed{1} \\ \text{SUBCAT} & \langle\, \rangle \\ \text{LEX} & - \end{bmatrix} \\ \text{DTRS} & \begin{bmatrix} \text{HEAD-DTR} & \begin{bmatrix} \text{PHON} & \text{laechelt} \\ \text{SYN|LOC} & \begin{bmatrix} \text{HEAD} & \boxed{1}\begin{bmatrix}\text{MAJ} & \text{V} \\ \text{VFORM} & \text{FIN}\end{bmatrix} \\ \text{SUBCAT} & \langle \boxed{2} \rangle \end{bmatrix} \end{bmatrix} \\ \text{COMP-DTR} & \left\langle \boxed{2} \begin{bmatrix} \text{PHON} & \text{Iris} \\ \text{SYN|LOC} & \begin{bmatrix}\text{HEAD|MAJ} & \text{N} \\ \text{SUBCAT} & \langle\, \rangle\end{bmatrix}\end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

単一化の結果，(8)の統語素性は(15)と同一になり，格（CASE）の指定が付け加わる。値はNOMであり，これで主語1格であることが示される。(8)と(9)を結合した結果は概略(16)のようになり，これで文(7)の素性構造が示される。

5.3. 言語情報としての記号

すでに指摘したように，以上概観してきたHPSGの素性構造は，記号（sign）という名前が付いている。この「記号」という言葉から，すぐに思い浮かぶのはフェルディナン・ド・ソシュールの例の有名な定義であろう。彼によれば，記号（シーニュ）は，次の(17)のようなシニフィアン（示すもの＝音声）とシニフィエ（示されるもの＝意味）の観念連合的な結合体であるとされる。

(17)



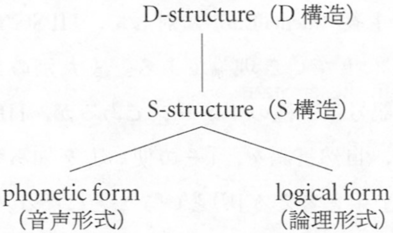
よく観察すると，実はHPSGの記号も，音声（PHON）と統辞（SYN）と意味（SEM）が1つの構造の中で統合的に記述されている。統辞は，この場合，音声と意味を媒介する機能を果たしている。このようにHPSGの記号もまた，言語における音声と意味の結びつきを明示的に構造化したものとみなされ，そこにソシュールの記号の発展した姿の1つを見ることができる。ただしそこには次のような原理的な差異もある。ソシュールにおいては，記号とは，言語共同体のメンバーの間で共有されている心理的な連合のシステムである。一方HPSGでは，「記号」は状況意味論的に再構成され，「記号とは，心理的な連合と解釈されるのではなく，シニフィアン（発話状況のタイプ＝言語表現）とシニフィエ（発話が記述する現実世界の中の物事のタイプ＝言語意味）との間に成り立つ関係（制約）を指す」[14]ということになる。つまりソシュールの心理主義（mentalism）に対して，HPSGは，ある種の現実主義（realism）を立脚点にしている理論である。また別の言い方をすると，ソシュールにとって，記号は一種のイメージであるが，HPSGにとっては，記号は言語情報であり，自然言語を，「その使い方を知っている人たちの間で交わされる情報の担い手と考える」[15]ということになる。

このような現実主義の立場はまた，HPSGの直接の源流であるチョムスキー理論とも，一線を画すものである。周知のようにチョムスキー理論は，心理学と深い関連を持ち，人間の言語能力の解明という認知科学的課題を持ったメンタリズムそのものの理論だからである。HPSGにとって，こういう側面の解明は当面の課題ではない。むしろ人間がコミュニケーションする，情報をやりとりするという事実に対して，十分に形式化された明示的な理論を提示することがHPSGにとっての問題なのである。

こうしたコミュニケーション過程，つまり話し手と聞き手という双方向性のモデルを構築するためには，実はチョムスキーの理論的枠組み（GB理論）は，そう最適なものとは言い難い。(18)がこのモデルの概略であるが，そこではモデルの中心に統語部門があり，その内部でD構造とS構造という

二層の構造レベルが想定される。それらを派生するために，前者では句構造規則，後者では変形規則といった規則体系が設定される。さらにこの統語部門の出力は，一方で音形部門の入力，他方で意味解釈部門の入力となり，ここに至ってようやく音声と意味との間の媒介が行われる。音声と意味がリニアに媒介されるのではなく，かなり屈折した複雑な外形をしている。これは，この理論が，当初からさまざまな改訂がなされているものの，一貫して統語部門をモデルの中核に置き，文の生成メカニズムを，派生規則の体系として記述する，いわば一方通行の話し手側の文法，という本質から起因するものと言えよう。

（18） GB 理論の表示のレベル



これに対して HPSG の場合，表示のレベルは，すでに見てきた「記号」だけであり，ここに音声，統辞，意味の情報が統合的に記述される。この記号は，話し手側と聞き手側の両方からアクセス可能な，可逆的（reversible）かつ双方向的（interactive）な性格を持っており，情報のやり取り，つまりコミュニケーション過程が，より自然なかたちでモデル化されていると言える。実際，コミュニケーション過程においては，言語情報（統語，語彙，意味，音韻）や文脈情報は絡み合い，相互に影響し合っていると考えられ，この関係は，さまざまなレベルを個別にモジュール化した，いわばノンリニアなモデルより，さまざまな情報を統合的に表示できるリニアでシンプルなモデルにおいて，より適切に記述できるように思われる。

HPSGは，しばしばGB, LFG, GPSGなどと並置され，生成文法の諸理論（つまりチョムスキー理論）の１つとして紹介されることが多いが，以上見てきたように，他の生成文法とは一線を画した独自の理論的射程を持っている。それは記号学の始祖ソシュールにまで先祖返りし，その伝統を批判的・発展的に受け継いだ現代記号理論の再構築であり，一方で計算機処理を強く指向した現実主義の言語観である。

5. 4. HPSGによる辞書記述

この節では，具体的にProlog言語による，HPSG辞書と文法規則のコーディングの問題を考えてみる[16)]。記述のサンプルとして次の(19)のドイツ文を取り上げる。

(19)　Ich wohne in dieser Stadt.

前節で概観したように，HPSGの文法規則は，要素を結合する際に，その構造の適格性をチェックする抽象的な制約というかたちをとり，その数も極めて少ない。必要な情報は個々の語彙項目に記述される。したがって，HPSGによるコーディングとは，その大部分が辞書の記述という作業に他ならない。

(19)には5個の構成要素（つまり単語）があるが，このうちwohneは動詞の定形，dieserは，冠詞の格変化形なので，通常の辞書と同様，この形をそのまま辞書項目としないで，その原形のみを辞書に登録する。つまりwohneの場合は不定形wohnen，dieserの場合は語尾を取ったdies[17)]を辞書項目とする。変化形から原形情報を得て辞書引きさせるためには，この連載の第一回に扱った形態素処理規則が必要である。今回のコーディングにあたって，この規則も前回のものから大幅に改訂している。変化の激しいドイツ語のような言語では必須の規則であるが，今は直接の話題ではないので，コードの

提示は省略する。(19)の他の3つの構成要素のうち，ich は不規則変化する代名詞，in は不変化の前置詞，Stadt は基本形（単数1格形）と同形の名詞であり，これらは，すべてこの形を辞書項目として登録する。

次にこれら5個の要素について，Prolog のコードで表現した HPSG 辞書記述の実際を見てみよう。(20)から(24)までがそれである。引数は3個あるが，2番目と3番目の引数（各項目末尾の N,N となっている部分）は，差分リストによって形態に対する制約を表した部分であり，辞書記述の本体は，1番目の引数である。ここでは HPSG の素性構造が，Prolog の節として表現されている。

(20)には，各行に Prolog のコメントを付けているので，その概略の意味は理解できることと思われる。つまり，ich は名詞（n）で，1人称単数形（s1），1格（nom）であることが，第3要素まで（最初の3行）で示されている。ちなみに，第1要素（MAJ）におけるカテゴリーと品詞の対応関係は次の通りである。

n	名詞
v	動詞
a	形容詞，副詞
det	冠詞(類)
p	前置詞

第2要素は，s（単数）と p（複数）および 1, 2, 3（人称）を組み合わせて述語が作られている。名詞の場合(24)に見られるように，m（男性），f（女性），n（中性）のいずれかを値とする引数が1つ追加されている。動詞の場合，不定形（inf），定形（fin）などの情報がここに記述される。第3要素で，名詞の場合，nom（1格），gen（2格），dat（3格），acc（4格）という格に関する情報が示される。

さて第4要素から第7要素（インデントされた4行分）までが SUBCAT（下位範疇化素性）であり，ここに前後の語句の接続条件が示されている。(20)の

ichの場合，4行がいずれも空リストであり，これにより他の要素との結合可能性がないことが示される。つまりこのichだけで句（この場合は名詞句）が飽和・完結してしまうわけである（これで代名詞であることが暗示的に表現されている）。この4行の意味は，次の通りである。

第4要素　自身が主辞である場合の前方（補助部）への接続条件
第5要素　自身が主辞である場合の後方（補助部）への接続条件
第6要素　自身が補助部である場合の前方（主辞）への接続条件
第7要素　自身が補助部である場合の後方（主辞）への接続条件

HPSGの下位範疇化素性は，表層の語順までは指定していない。語順は線形順序（linear precedence – LP）制約という別の一般的な原理で述べられることになっているからである。したがって，ここでの主辞・補助部，前方・後方

(20)
```
ich(
      [
            n,                  %MAJ  = 品詞
            s1,                 %AGR  = 人称・数（性）
            nom,                %CASE = 格
                  [],           %  ┐
                  [],           %  │ SUBCAT（下位範疇化素性）
                  [],           %  │
                  [],           %  ┘
            []                  %BIND（束縛素性）
      ]
      :
      私,                        %SEM（意味素性）
      N,N
).
```

```
(21)
wohnen(
        [
                v,
                inf
                _,
                        [[n,_,nom,[],[],[],[],[]]:X],
                        [],
                        [],
                        [],
                []
        ]
        :
        [X,は,住む],
        N,N
).
```

```
(22)
in(
        [
                p,
                _,
                _,
                        [],
                        [[n,_,dat,[],[],[],[],[]]:Y],
                        [[v,AGR,CASE,[],[],[],[],[]]:X],
                        [],
                []
        ]
        :
        [Y,に,X],
        N,N
).
```

```
(23)
dies(
        [
                det,
                base,
                _,
                        [],
                        [],
                        [],
                        [[n,s3(GENUS),CASE,[],[],[],[],[]]:X],
                []
        ]
        :
        この,
        N,N
).
```

```
(24)
'Stadt'(
        [
                n,
                s3(f),
                _,
                        [[det,_,_,[].[],[],_,[]]:X],
                        [],
                        [],
                        [],
                []
        ]
        :
        [X,町],
        N,N
).
```

といった場合分けのコーディングは，解析の効率という実用面を考慮した処置である。

次の第 8 要素は，関係詞や疑問形など非局所的な長距離依存関係を扱う束縛素性 (BIND) である。最後に，その下行のコロン“:”の後ろに，意味素性 (SEM) が置かれている。ここでは状況意味論ではなく，単なる日本語訳を対応させているが，これにより，簡単な日本語訳の出力が可能となっている。

5. 5. 制約としての文法規則

次に文法規則について見ていこう。(19)の文を解析するためには，ボトムアップ的に図式化すると，次の(25)a.～d. の(部分)構造をすべて解析できなければならない (下線部は主辞)。

(25) a. dieser Stadt (名詞句 = NP)
b. in dieser Stadt (前置詞句 = PP)
c. wohne in dieser Stadt (動詞句 = VP)
d. ich wohne in dieser Stadt. (文 = S)

これらの構造を，主辞の位置で分類すると，a と d が後置型，b と c が前置型となる。また，動詞句における人称変化，名詞句における格変化，前置詞句における格支配のような，何らかの意味で素性の一致によって語形が決定される構造を支配型と呼ぶことにすると，支配型は a と b と d，非支配型は，c と分類することができる。以上のことは，次の(26)のマトリックス表にまとめることができる。

(26)

	NP	PP	VP	S
主辞前置型	−	+	+	−
支 配 型	+	+	−	+

つまり値の組み合わせとしては，次の3つの型が抽出される。

Ⅰ ［－主辞前置，＋支配］型＝NP, S
Ⅱ ［＋主辞前置，＋支配］型＝PP
Ⅲ ［＋主辞前置，－支配］型＝VP

可能性としてはもう1つ，［－主辞前置，－支配］型が考えられる。実際，この構造もドイツ語に存在する。sehr gutのような形容詞句（AP）である。

Ⅳ ［－主辞前置，－支配］型＝AP

したがってここでは，このⅣを含め，以上4つの文法規則を定式化してみることにする。次の(27)～(30)がそれである。それぞれの規則に付けたコメントで，およその働きは理解していただけると思う。それぞれ引数は3個あり，前要素（第1引数）と後要素（第2引数）が結合して，その結合結果が第3引数で示される構造となっている。

(27)の文法規則Ⅰは，例えば(25a)の名詞句や(25d)の文に関わる規則である。(25a)なら，主辞である後要素Stadtが，前要素に女性3格形のdieserを要求して新たな名詞カテゴリーを生成することを表している。また(25d)なら，主辞である後要素wohne in dieser Stadtが，前要素に1人称単数形のichを要求して新たな動詞カテゴリー（この場合は文）を生成することを表している。

(28)の文法規則Ⅱは，(25b)のような前置詞句に関わる規則である。(25b)を例にとると，主辞である前要素inが，後要素に3格形のdieser Stadtを要求して新たな前置詞カテゴリーを生成することを表している。

(29)の文法規則Ⅲは，文の後ろの要素が前の要素を修飾する関係を表しており，(25c)のような動詞句に関わる規則である。(25c)を例にとると，後要素の前置詞句（カテゴリーは前置詞）in dieser Stadtが，主辞である前要素の動詞wohneに付いて，新たな動詞カテゴリー（動詞句）を生成することを表

している。

(30)の文法規則Ⅳは，(29)とは逆に，文の前の要素が後ろの要素を修飾する関係を表しており，すでに述べたように，例えばsehr gutのような形容詞句が，この規則で生成される。

(27)　文法規則Ⅰ(NP，S)

```
psr(
        [MAJ1|Spec]:Sem1,                      % 前要素（補足語）
        [
                MAJ,
                AGR,
                CASE,
                        [[MAJ1|Spec]:Sem1],
                        [],
                        Pre,                   後要素（主辞）
                        [],
                [],
        ]:Sem,
        [
                MAJ,
                AGR,
                CASE,
                        [],
                        [],                    結合結果
                        Pre,
                        [],
                [],
        ]:Sem
).
```

（28） 文法規則Ⅱ（PP）

```
psr(
        [                                         ┐
            MAJ,                                  │
            AGR,                                  │
            CASE,                                 │
                Init,                             │
                [[MAJ1|Spec]:Sem1:Tail],          │ 前要素（主辞）
                Pre,                              │
                Post,                             │
            []                                    │
        ]:Sem,                                    ┘
        [MAJ1|Spec]:Sem1,                         % 後要素（補足語）
        [                                         ┐
            MAJ,                                  │
            AGR,                                  │
            CASE,                                 │
                Init,                             │
                Tail,                             │ 結合結果
                Pre,                              │
                Post,                             │
            []                                    │
        ]:Sem                                     ┘
).
```

（29） 文法規則Ⅲ（VP）

```
psr(
        [MAJ1|Spec]:Sem1,                      % 前要素（修飾語）
        [                                      ┐
              MAJ,                             │
              _,                               │
              _,                               │
                   [],                         │ 後要素（主辞）
                   [],                         │
                   [[MAJ1|Spec]:Sem1],         │
                   []                          │
        ]:Sem,                                 ┘
        [MAJ1|Spec]:Sem                        % 結合結果
).
```

（30） 文法規則Ⅳ（AP）

```
psr(
        [                                      ┐
              MAJ,                             │
              _,                               │
              _,                               │
                   [],                         │ 前要素（主辞）
                   [],                         │
                   [],                         │
                   [[MAJ1|Spec]:Sem1],         │
              []                               │
        ]:Sem,                                 ┘
        [MAJ1|Spec]:Sem1,                      % 後要素（修飾語）
        [MAJ1|Spec]:Sem                        % 結合結果
).
```

6. 結　　語

以上提示したHPSGによるコーディングの試みは，もちろんまだほんの端緒に過ぎない。理論的進展と合わせて，解析できる文の数を増やし，辞書を拡充していくことが今後の課題である。

この連載も，その最初の試みからずいぶん時間が経過した。その間，理論面でも，実際面（コンピュータをめぐる環境）でも大きな変化があった。今回新しく，最先端の文法理論である主辞駆動句構造文法（HPSG）を取り上げたが，この理論は，1つのパラダイム転換とも言えるような新しいアイディアに満ちたものであった。私のドイツ語機械解析のプロジェクトも，現在，この方向に沿って新しい構想で始められており，1つの区切りをつける意味で，この連載も今回で終了することにしたい。新しい成果は，別の場所で公表していくつもりである。

〔注〕

1）本稿は，平成9年―12年度科学研究費補助金基盤研究(C)(2)課題番号09610522「ドイツ語テキストデータベース解析辞書構築に向けての理論的・実際的検討」の研究成果の一部である。

2）Prologシステムの内部では，(2)のDCG表現は，DCGトランスレーターによって，次のようなPrologプログラムに変換される。

```
s(S0,S):-np(S0,S1),vp(S1,S).
np(S0,S):-det(S0,S1),noun(S1,S).
np(S0,S):-noun(S0,S).
vp(S0,S):-v(S0,S1),np(S1,S).
det([der|S],S).
noun(['Musiker'|S],S).
noun(['Geige'|S],S).
v([spielt|S],S).
```

3) Prologでは大文字で始まる文字アトムは変項となる。

4) 文法規則に課せられる制限を強くしていくと，生成される言語のクラスが小さくなるという階層。制限のない0型文法，1型文法（文脈依存文法 Context Sensitive Grammar＝CSG），2型文法（文脈自由文法 CFG），3型文法（正規文法 Regular Grammar＝RG）があり，この順に制限が強くなる。入門的な解説としては，例えば，赤間世紀（1998）: 自然言語・意味論・論理，共立出版，などを参照。

5) Chomsky, N. (1965): Aspects of the Theory of Syntax. Cambridge/London: The MIT Press. 安井稔(訳)(1970): 文法理論の諸相，研究社，参照。

6) Chomsky, N. (1981): Lectures on Government and Binding. Dordrecht: Foris. 安井稔・原口庄輔(訳): 統率・束縛理論，研究社，参照。

7) Bresnan, J. (Hrsg.) (1982): The Mental Representation of Grammatical Relations. Cambridge/London: The MIT Press. 参照。

8) Gazdar, G./E. Klein/G.K. Pullum/I.A. Sag (1985): Generalized Phrase Structure Grammar. London: Blackwell. 参照。

9) Pollard, C./I.A. Sag (1987): Information-Based Syntax and Semantics. Vol. I: Fundamentals. CSLI Lecture Notes. Chicago: University of Chicago Press. 郡司隆男(訳)(1994): 制約にもとづく統語論と意味論 ── HPSG入門 ──，産業図書，および Pollard, C./I.A. Sag (1994): Head-Driven Phrase Structure Grammar. Chicago: University of Chicago Press. 参照。ドイツ語への適用の問題については Kiss, T (1995): Merkmale und Repräsentationen. Opladen: Westdeutscher Verlag. 参照。

10) 後藤茂樹（1984）: PROLOG入門，サイエンス社，p.121.

11) Pollard, C./I.A. Sag (1987)，前掲書，郡司隆男(訳)，p.9.

12) 例えば，ドイツ語の音素 /d/ は，[＋stimmhaft]，[＋alveolar]，[＋plosiv] などの弁別的素性を持つとされる。

13) Barwise, J./J. Perry (1983): Situations and Attitudes. Cambridge/London: The MIT Press. 土屋俊他(訳)(1992): 状況と態度，産業図書，参照。

14) Pollard, C./I.A. Sag (1987)，前掲書，郡司隆男(訳)，p.6. なお傍点箇所は筆者による補足。

15) Pollard, C./I.A. Sag (1987)，前掲書，郡司隆男(訳)，p.1.

16) この5.4.節および次の5.5.節でのコーディングにあたっては，多くの技術解説書を参考にしたが，とくに，高野真:「Prologによる機械翻訳システム」1〜6，PC WORLD 1987年10月号〜1988年6月号所収，からは多くの示唆を得た。

17) ここでは語尾を取った語幹を基本形としており，男性1格形を基本形とする伝統的なドイツ文法の方法とは異なる。