

第3回

音楽と言語の構造認知

東条 敏 (北陸先端科学技術大学院大学)

ことばと音楽

我々は小鳥は「歌う」といいます。しかしそれは紛れもなく彼らのことばです。そのことばの意味するところは求愛であり、それ以上複雑なメッセージを仲間に伝えることはありません。しかしとにかく彼らは鳴管（咽喉に相当する器官）を使って発声し、それによってコミュニケーションを行う事実には変わりありません。実際、小鳥の歌には文法があります。岡ノ谷らの研究^{☆1}によればジュウシマツの歌はチョムスキー階層¹⁾における正規文法であり、有限状態オートマトンで表現されることが知られています。

小鳥の言語を歌というのはメタファにすぎませんが、では逆に、言語というのは歌と独立に存在し得たでしょうか。今日の我々のように舌や口の位置・形でさまざまな母音・子音を作り出すためには、その舌や口唇を動かす筋肉の発達が必要になります。すると舌や口唇が発達する以前において、たとえば鳥のように変形しない^{くちばし}嘴を持っていたころ、最も容易に発声にバラエティを持たせるための手段は音の強弱と高低であったと考えられます。すると今日我々がいうところの歌が、進化の過程において日常言語と同じルーツを持つものとするのも無理なことではありません²⁾。

ここからは生物学的に人間の言語が言語として独立したはるか後、文明の時代になってからの話です。音楽の出自として考えられるのは古代ギリシアのムシケー（*μουσική*）という概念であり、それが分化して片やミュージックに、片や詩文の朗読になったと言われていきます^{☆2}。ピタゴラスは弦をはじいて弦の長さがきれいな整数比のときに、すなわち周波数の波の一致がよくとれ

るときに音が調和するとして、2:3の比（ドとソの音程）で音を作っては1:2の比（1オクターヴの音程）で折り返すという操作でピタゴラス音階を作りました^{☆3}。その後音楽を発展させたのは宗教的儀式における祈りであり、ミサの唱和です。このためにうまく調和する音程を規定すべく、独自の音階＝教会旋法が作り出されました。以後、バロックよりクラシックの時代にかけて音楽の歴史とはこの和音とその進行に関する理論の整備と発展・拡大の歴史でもありました。しかしクラシック転じて芸術音楽は20世紀が明けるころは和音の崩壊をめざし、一般的な人気からは乖離していきます。

現在でも、ことばと音楽には不可分の関係があるとは巷間よく言及されることです^{☆4}。フランス語とシャンソン、イタリア語とカンツォーネはいずれもその言語でないと雰囲気が出ません。ベートーヴェン（Beethoven）はMuss es sein? Es muss sein!というドイツ語の会話の語調をそのまま弦楽四重奏曲第16番の第4楽章の主題にしました。バルトークは東欧・中欧の民謡を収集し、伝統的なドイツ・オーストリア音楽と異なるこれらの地域の音楽が、この地域特有の音韻体系とかかわる音律を含むことを明らかにし、20世紀の音楽の新しい地平を開くのに寄与しました。自然言語は言語によってアクセントを作り出す方法が異なり、音の強弱による方法、高低による方法、長短による方法があります。そして、これら強弱、高低、長短とはまさに音楽の概念です。するとある言語がその地域で歌われる歌・音楽と親密な関係にあるという主張もこれまた自然なことと思われます。

さて音楽の起源と言語の起源がこのように関係ありそうであれば、分析手法にも共通に通用するものがあるはずで、本稿では、これまで言語の研究に対して行われてきた方法論を音楽に適用することを試みます。

自然言語解析の方法論

自然言語処理の伝統的な解析においては大まかな流れとして、(i) 音韻認識、(ii) 形態素解析、(iii) 構文解析、(iv) 意味表現というシーケンスが考えられます¹⁾。(i)の音

☆1 岡ノ谷一夫：小鳥の歌からヘビットの言葉へ、岩波科学ライブラリー。

☆2 ゲオルギアードス(木村 敏 訳)：音楽と言語、講談社学術文庫。

☆3 芥川也寸志：音楽の基礎、岩波新書 E57。

☆4 小倉 朗：現代音楽を語る、岩波新書 E56 より IV 「言葉と音楽」。



韻認識とは、音すなわち空気振動の波形から、話者の発音を特定する操作です。荒っぽく言えばマイクロフォンで音を捉え、その結果文字列が出てくるまでのプロセスを指します。(ii)の形態素解析とはいくつかの文字を連結させて、単語（もしくは接頭辞・接尾辞）など最小限の意味をなす単位を認識するプロセスを指します。(iii)の構文解析とは、一言でいえば単語間の係り受けの関係を見出すことです。(iv)の意味表現とは、文の内容を計算機上で形式的な記法に表現し直したものです。

● プッシュダウン・スタックと CFG

まず自然言語での (iii) 構文解析を考えてみましょう。人間の言語が他の生物の言語と決定的に異なるのは、それが階層的な構文を持つということです。この階層的という概念を説明します。たとえば「渋谷に CD を買いに行く」という文を考えてみましょう。「渋谷に」は「行く」わけですし、「CD を」「買う」わけですからこれらの名詞句と述語の関係には係り受け関係があります。日本語はこの係り受け関係が交差しないこと、すなわち「係り」から「受け」に矢印を引くと矢印が交わらないことが原則です。よって「渋谷に行って CD を買う」は OK でも「渋谷に CD を行って買う」は不自然だというわけです。この係り受け関係を敷衍して、

(a) 単語間の依存関係は交差しない

ということは、我々の言語の重要な特徴の 1 つです。係り受けの矢印の始点・終点を括弧の開く・閉じるに対応させると、係り受け非交差は

(b) 括弧の開く・閉じるが適切な位置関係にあることと同じことになります。今度はプッシュダウン・スタックを考えてみましょう。(b)の括弧を開くという操作は、いつかそれが閉じられることを予測してスタックに積む (push) ことと考え、括弧を閉じるという操作はスタックから対応する開いた括弧を取りだす (pop) ことと考えられます。そして、プッシュダウン・スタックを持つオートマトン、すなわち

(c) プッシュダウン・オートマトン

で受理可能な言語はチョムスキー階層で

(d) 文脈自由文法

(Context Free Grammar ; 以下 CFG)¹⁾ と呼ばれます。これは英語を例にとれば、以下のように矢印の左辺がただ 1 つの項からなる生成規則による文法です^{☆5}。

〈文〉→〈名詞句〉〈動詞句〉

〈名詞句〉→〈冠詞〉〈名詞〉

〈動詞句〉→〈他動詞〉〈名詞句〉

〈動詞句〉→〈自動詞〉

(a)～(d) はすべて同等な概念であることを念を押しておきます。

さてプッシュダウン・スタックがあるとなぜ階層的になるのでしょうか。それは 1 つの係り受け関係が 1 つの文の中で語句としてまとまった意味を形成し、部分木を作り、文全体はこうした部分木を接合して、1 つの大きな木構造を作るからです。ここに部分木から全体木への階層が生まれます。プッシュダウン・スタックは人間の脳の中の一時記憶装置であり、発話されたある単語が一時的に蓄えられいつかそれと呼応する単語が来ることを予測するしくみです。逆に言えば生物進化のうちに人間だけがこのような記憶装置を具備することによって階層的な文を理解するようになったとも言えます。

さて一度こんな便利な記憶装置を身につけてしまったら、それが音楽を聴くときに活用されないということは逆に考えにくいのではないのでしょうか。すなわち人間は音楽を聴くときも、ある楽句の記憶をもとにそれと関連する楽句を予測するような聞き方をしていないのでしょうか。ここで言語から音楽へのアナロジーを考えます。

● 音楽へのアナロジー

さて自然言語解析の (i)～(iv) のプロセスで、どこまで音楽へのアナロジーが通じるのでしょうか。(i)の音韻認識は言語・音楽ともまったく共通であり、(ii)の形態素解析は近隣の音符を集めた楽句認識に相当すると考えられます。しかし音楽にも (iii) の構文解析の方法論が成り立つのでしょうか。すなわち楽句と楽句の間に何か有意義な関係を見出す操作が定義できるのでしょうか。これが本稿の主題です。というわけで「音楽の構文解析」は次節にスペースを割いて論じることにしてしましましょう。

この次の (iv) 意味解析というのは音楽にも通じるものがあるのでしょうか。自然言語にとっての意味表現は、伝統的には一階述語論理と呼ばれる論理を用いて一文に相当する情報を述語と引数の構造によって表現したものです。もう少し洗練された意味表現、たとえば DRT^{☆6} などにおいても意味の単位はこの述語構造です。とにかく自然言語解析にとっての意味とは、人工的な形式言語すなわち論理への翻訳と考えてください。したがって日常的な意味での「意味」というものとは異なります。さて、音楽においてはこれは何に相当するのでしょうか。少なくとも筆者には「音楽の意味」というのが自明な概念であるようには思えません。

☆5 自然言語には例外的に文脈自由でないような文もあります。たとえばオランダ語の従属節における動詞の後置については係り受け交差することが知られています。

☆6 Kamp, H. and Reyle, U. : From Discourse to Logic, Kluwer (1993).

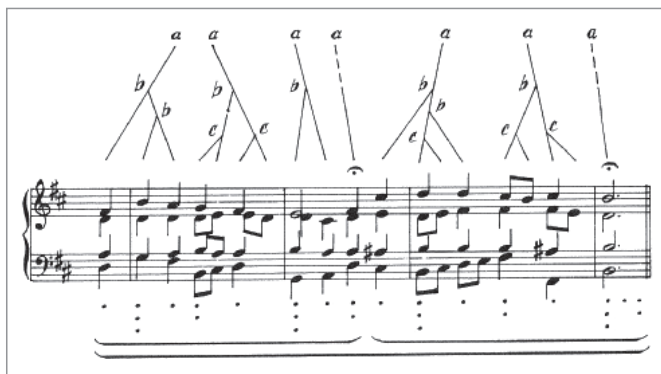


図-1 タイムスパン木の生成 (文献3), p.132)

ところで自然言語処理では (iii) 構文構造と (iv) 意味表現は合成性 (compositionality) ^{☆7} という考え方により、部分的な構文構造がその部分の意味をなし、部分と部分の結合が全体の意味を形成するという立場をとります。したがって、部分の意味の集積が必ずしも全体の意味にならないような熟語・慣用句は例外と考えます。先に述べたように構文が階層的ならば、意味も階層的に構成されるというわけです。よって (iii) と (iv) は一体の不可分のプロセスと考えます。音楽もある観点から構造解析をすれば、少なくともその観点は表現したものができるでしょう。本稿でも構文解析の結果はある種の意味と考えます。

自然言語解析にはさらにその先に (v) 語用論 (pragmatics) というものがあります。これは言語表現が実世界の環境に埋め込まれた中でどういう機能を持つかということまで考えるものです。すなわちまったく同じことばでも状況によってはその含意するもの (implicature) が変わると考えるわけです。音楽でも演奏される場所、時間、周囲の環境によって訴えるものが違うというのは OK ですが、これを形式的に表現するのは困難であり、本稿では考えないことにします。

音楽の構文解析

さて我々の音楽認識においても、隣り合う音符どうしの近接作用的な楽句認識に対して、楽句の記憶を用いたより大局的・遠隔作用的な認識があり、これをマクロ構

造の認識と呼ぶことにします。この分かりやすい例は繰り返しであり、メロディーの音列が記憶されれば、我々はそれを再び聞いたときそれと認識することができます。同じメロディーが音程を変えて現れたりまた拡大^{☆8}するくらいなら誰の耳にもそうと認識できるでしょうが、メロディーを反行・逆行^{☆9}させたりするとこれを認識するにはちょっとした耳の訓練が必要です。

それから楽曲の開始の記憶があれば、区切りとなる個所への期待感があります。4拍子を例にとれば、人間は4小節・8小節でそれぞれリズム的に区切りの感じを持ちます^{☆10}。

さてそれ以外にもマクロな依存関係、特に係り受けのような構造はあるでしょうか。答えの1つがカデンツ (Kadenz (独), cadence (英)) です。カデンツとは我々が音楽の終了時に感じる「終わった感じ」を創出する制約であり、これが満たされないと曲が中途半端、尻切れトンボで終わってしまったという感じを持ちます。いま大文字のローマ数字 I, II, III, …でそれぞれ音階の第1音 (ド)、第2音 (レ)、第3音 (ミ)、…上の三和音を表すことにしましょう。すると、I, V, IV, すなわちドミソ、ソシレ、ファラドの和音は、それぞれトニック (tonic)、ドミナント (dominant)、サブドミナント (subdominant) と呼ばれる機能を帯び、

{ Tonic-Dominant-Tonic
 Tonic-Subdominant-Dominant-Tonic
 Tonic-Subdominant-Tonic

という進行があると終止感を作り出すとされます。

さてカデンツがなぜマクロな係り受けなのでしょう。今回の「道しるべ」シリーズでは、第2回までに大まかに音楽理論 GTTM (Generative Theory of Tonal Music) ³⁾ の概略を述べました。そして、その中の大事な理論の1つとしてタイムスパン木の生成というのがありました。図-1はJ. S. バッハ (J. S. Bach) のマタイ受難曲 BWV244 からコラール 'O Haupt, voll Blut und Wunden' の冒頭部分ですが、タイムスパン木は互いに隣接する音どうしを比較して拍節構造上重要な音の枝を上へ延ばすという操作から成立します。すると木は葉のほうからボトムアップに形成されることになります。

こうして形成された木は、マクロな係り受け関係を適切に表現できるでしょうか。一般的には答えはNoです。すなわち楽句のある始まりのところの記憶があって、それに呼応した結末を予測するには、木をある高さまで辿って、そこから下に向かって制約を加えるようなトップダウンな操作が必要とされます。GTTMにおいては、このようなトップダウン操作をキャッチボールになぞらえ (文献3), p.133), 投げられたボールが遠隔でキャッチされる必要があることを述べています。そしてその例

☆7 特に形式意味論においては Frege の原理と言います。

☆8 メロディーの各音の音価を均等に引き延ばすこと。

☆9 音列を逆順に辿ったり、上下を反転させること。セリー (série (仏); 12音音列) の音楽には必ず現れる。

☆10 ところがビートルズの Yesterday は7小節区切り! 何にでも例外はあります。

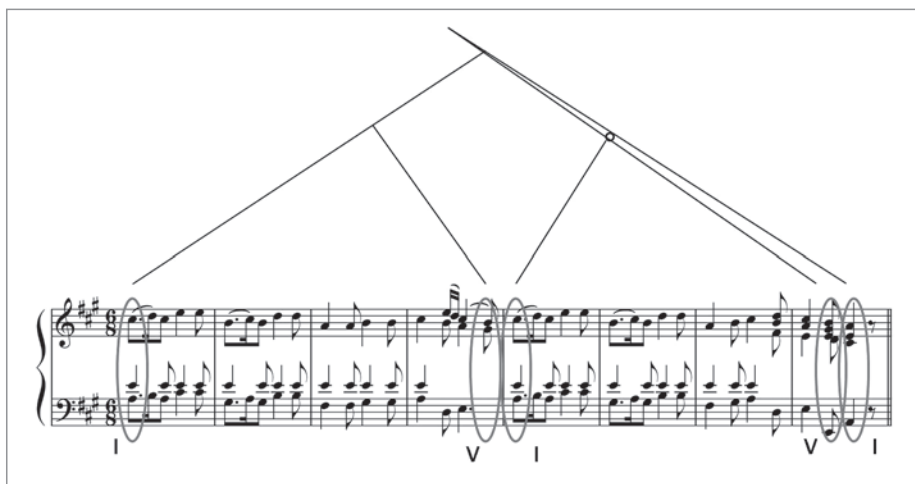


図-2 モーツァルト K.331 のカデンツ構成

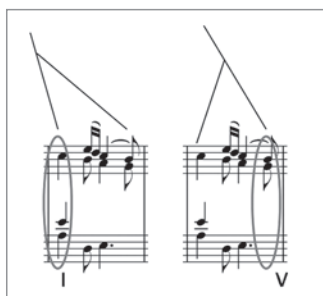


図-3 モーツァルト K.331 の第4小節

$K \rightarrow T, D, T$	$T \rightarrow T, T$
$K \rightarrow T, S, D, T$	$S \rightarrow IV$
$K \rightarrow T, S, T$	$S \rightarrow II$
$T \rightarrow I$	$S \rightarrow S, S$
$T \rightarrow VI$	$D \rightarrow V$

図-4 CFGによるカデンツ規則

としてカデンツについて説明がなされています。

図-2はモーツァルト（Mozart）のピアノ・ソナタイ長調 K.331^{☆11}の第一楽章冒頭の変奏曲の主題前半部です。この図-2は正しいタイムスパン木を示しており、最後のV-Iから延びる枝上の小さな丸はV-Iがカデンツとしてまとめられた処理の跡を示しています。

ところがこの第4小節（図-3）に注目すると、1拍目のA-E-C#はイ長調のIの和音、6拍目のE-G#-BはVの和音であり、比較すればIの和音のほうが拍節的に重要となるはずです（図-3左側）。ところが第4小節の最後のVは次の第5小節のIにつなぐための半終止（half-cadence）であり、この半終止を見せるためにはこのVからの枝が伸びていかなければなりません（図-3右側）。このことはボトムアップな操作だけでタイムスパン木ができないことを示しています。

とにかく我々は音楽を聞いて、楽譜を見て、このようなマクロな構造を認識しています。楽曲の開始の仕方はその終止（カデンツ）の形態に明らかな影響を与えます。このような構造は最初から作曲家がそのように企図して

設計したものであり、その構造が現実楽譜に書かれてあるにもかかわらず、音符の近接的な結合だけからその構造を見出すのは困難です。

ここで自然言語のパースが係り受け構造を適切に見出すしくみで動いていることを鑑み、いままで聞いた語句から将来現れる語句をトップダウン的に予測するメカニズムを用いて、楽曲のマクロな構造の認識に応用してみようと考えます。

カデンツの解析

● CFGの問題

それではさっそく文脈自由文法（CFG）によって和音列の生成規則を記述してみましょう^{☆12}。図-4ではTはトニック、Dはドミナント、Sはサブドミナントを意味し、最初の3つの規則のKはカデンツを表します。

しかしながら、このCFGによる表記では実際の和声進行に必要な制約を表現し切れていません。たとえば $S \rightarrow S, S$ という規則からIVとIIを連結して再びSになることが分かります。しかしIV-IIという連鎖はT-S-D-Tの中で用いることはできても、T-S-Tの中においては用いることができません。したがってこの2つのSを区別するために S_1, S_2 のような付記が必要です。同様に、カデンツの最初のTと最後にあるTはいつも同じもの

☆11 第三楽章が有名なトルコ行進曲。

☆12 文法理論を用いての和音の連結とカデンツにおける機能分析については、文献6)という先駆がありました。

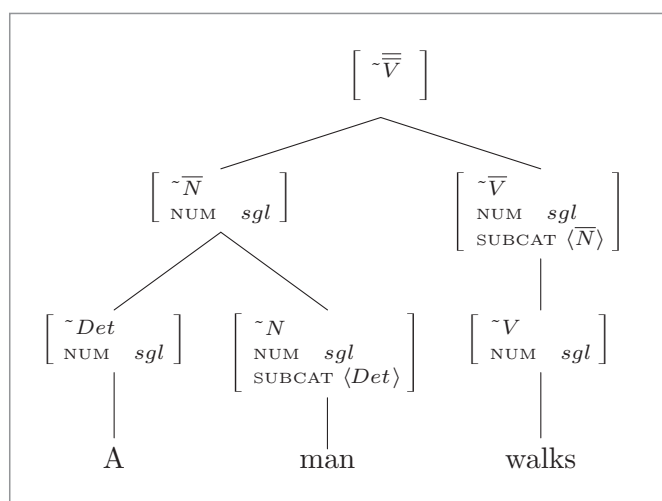


図-5 自然言語文の木構造例

ではありません。最後の T では I, VI を $T \rightarrow T, T$ によって連鎖させた $I-VI$ が現れることは稀です。よってカデンツ開始時の T と区別するために T_1, T_2 などとシンボルを区別する必要があります。これは同じトニックとしての処理の共通性を考えると無駄が多い記述となります。加えて七の和音や転回形、バス音の位置などの導入に伴うルールの追記には多くのシンボルの定義を伴い、知識表現としても非常に煩雑なルール体系となります。

以上の事態は、次の2点に着目することにより改善の可能性があります。

1. カデンツ生成規則の各シンボルは、それぞれ K, S, T, D として処理の共通性を保ったまま、内部には素性があると仮定し、その値によって規則の適用条件を変更できること。
2. その際、下位のシンボル列の中には、ある中心的な役割を果たすヘッド（Head；主辞）が存在し、それが主に上位シンボルの性格を決定できること。

このように内部素性とヘッドの概念によって CFG を拡張した文法としてヘッド駆動型の句構造文法（Head-driven Phrase Structure Grammar；以下 HPSG）⁴⁾ が有望です。この文法では、CFG がわずかに異なる規則でも別規則としていちいち列挙しなければならないのに対して、同類の規則をひとまとめでした規則を書くことができます。

● HPSG 入門

内部素性（属性）とその値のペアを縦に複数並べて角括弧でくくった構造を素性構造と呼びます。図-5は自然言語文“A man walks.”の構文木を生成した例です。木の各ノードは素性構造になっていて、各ノードでは素性 NUM（number）が値 sgl（単数；singular）であるという情報を持っており、この値が異なるものとは木構

造を構成できません。

1つの素性構造は型を持ち、その名前は先頭行に配置され‘~’に先行されて表されます。図中、ヘッドは型名の上にバーを付加して表現されています。すなわち‘man’の型は名詞（ N ）であることにより、‘a man’の型は名詞句（ \bar{N} ）となっています。同様に、動詞句（ \bar{V} ）を形成するヘッドは（ V ）であるとし、文全体（ \bar{V} ）は動詞句をヘッドとして形成されます。

またヘッドである子がそうでない子と結びついて親を形成する際、どのような型と結びつくかという情報が SUBCAT（subcategorization）という素性に書かれています。‘man’をヘッドとして冠詞（determiner）‘A’を結合する際、manの素性構造の中には SUBCAT〈 \bar{Det} 〉と指定されています。‘walks’が‘A man’（ \bar{N} ）を結合する際も walksには SUBCAT〈 \bar{N} 〉と指定されています。

また、HPSGでは図-6のように記された四角囲みのインデックスによるポインタを素性構造に付加する（素性構造の角括弧の前に先行させて書く）ことによって、以降そのポインタがその素性構造全体を指すことができます。ポインタを用いて異なる位置で同一の型や素性構造を共有することを構造共有と呼びます。

● カデンツ生成の文法とそのプロセス

それではさっそくカデンツの規則を HPSG で書いてみましょう⁵⁾。楽曲の和音列から自然言語文同様の木構造を生成するために、本稿ではカデンツの最後に現れる T の和音がカデンツ全体を支配するヘッドであるという指針を立てます。これは T がカデンツの最初にも最後にも共通に現れる中心的な和音であり、特に最後の T の和音は、 D や S に誘発されて終止を決定づける性格を持つことによるものです。本稿においては、 T, D, S の各項の和音の機能と、3つのカデンツ規則を素性構造を用いて定義します。具体的には、先に CFG で書いた各規則において、 T, S, D が単体のシンボルであったのを属性構造に表現します。特にカデンツ生成規則においては、トニック T に対して、 S, D を含んで最後の T に至る連鎖を \bar{T} 、カデンツ全体を \bar{T} と表記します。

図-6にカデンツ規則の主要な部分を示します。図-4の $S \rightarrow S, S$ に対しては $\bar{S} \rightarrow S, S$ というように HPSG 化しておきます。T-S-Tのカデンツでは \bar{S} の代わりに S を用いていることから IV-IIが自然に排除されていますが、 S と \bar{S} は型内部の情報が共有されているために両者には共通な処理が可能です。一方 $\bar{T} \rightarrow \bar{D}, T$ の中の \bar{D} を見ると、 $\bar{D} \rightarrow \bar{S}, D$ のように \bar{S} が用いられています。よって T-S-D-T の中の S は \bar{S} でもかまいません。いずれのカデンツも最後の T をヘッドとし、 S, D などを含めて \bar{T} を形成することでカデンツ \bar{T} を形成します。

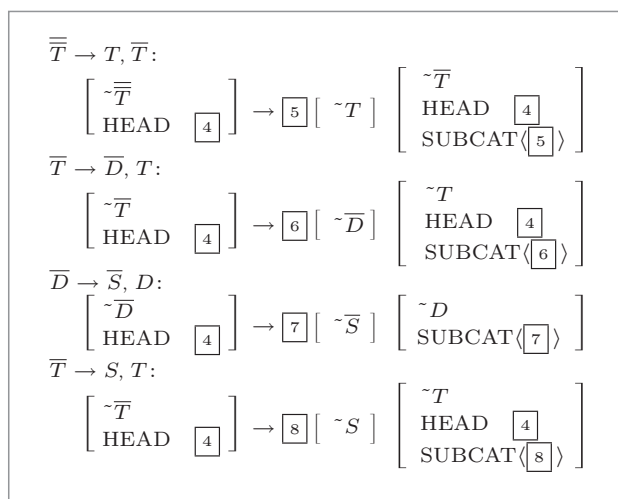


図-6 HPSGによるカデンツ規則

和音列に対して和音進行解析を行う場合、カデンツの規則ではII-Vの進行は許されますがV-IIの進行は許されていません。このような連結の方向性を担保するために、L (left), R (right) 素性を付加しておきます。また HEAD-DTR (head daughter) はヘッドとなる子、

COMP-DTRS (complement daughters) はその他の子を表します。

カデンツ規則を用いて、和音列からカデンツ構造を解析するプロセスについて、図-7で和音列

C - Am - F - Dm - G7 - C

を例に解説します。この和音列はハ長調のT-S-D-Tのカデンツであり、音度表記を用いてI-VI-IV-II-V-Iと表記できます。さてCFGの説明で挙げた規則(図-4)も内部素性によってHPSG化されていると考えてください。図-7の最初の2つの和音 I 、 VI に注目すると、まず $T \rightarrow T$ 、 T が適用でき、2つまとめた I の木構造が生成されます。次にこの T は図-6のカデンツ規則 $\bar{T} \rightarrow T, \bar{T}$ によって \bar{T} を予測します。そしてもし折りよく I が構成できれば、 VI が構成され、図-7の全体を作ることができます。

おわりに

楽譜は音符と休符が読む方向に(時間軸に沿って)並べられているという意味で、自然言語のテキストと同様

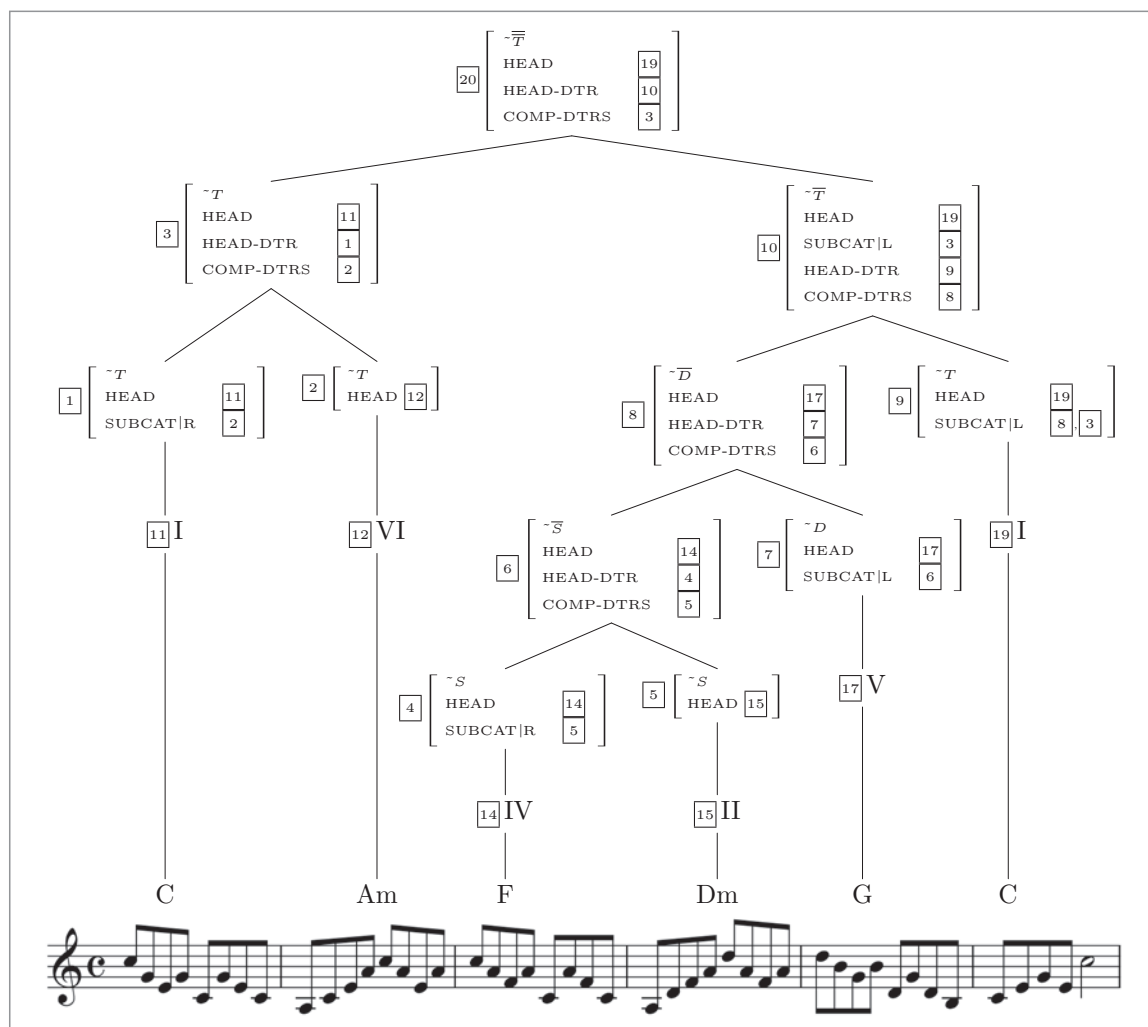


図-7
カデンツの解析例

なシンボル列と考えられます。もしそれが規則によって並べられたシンボル列ならば自然言語のパースを援用してみようという発想が生まれます^{☆13}。

本稿では音楽について、言語と音楽のアナロジーから出発し、自然言語の構文解析技術を使って楽譜の構造解析ができるのか、という試みについて述べました。まず GTTM においては、タイムスパン木を生成する際、完全にボトムアップな方法だけでは人間が認知できる音楽のマクロな構造が捉えきれないことがあるという視点を提示しました。そこで次の HPSG によるカデンツ解析となるわけですが、カデンツというのはたとえば I-V-I という連続を強制しているのではなく、I からスタートしたら、途中経過句があって、最後に V-I という結びがあればいいことになっています。そして経過句の部分の和声進行はある程度自由です。するとトップダウンな生成規則のみではなく途中経過を柔軟に扱える必要があります。さらに他の拍節構造認識のプロセスがボトムアップであることより、その操作と組み合わせたパースが要請されます。HPSG は文法「規則」を書くフォーマリズムであって、それがどのようなアルゴリズムで動くか、

☆13 たとえば遺伝子の中の DNA の配列も同じようにシンボル列です。DNA の解析にも自然言語のパースが試みられたことがありました。

すなわちトップダウン／ボトムアップをどのように切り替える戦略をとるかまでは規制しません。そこで、この切り替えの適切なアルゴリズムが必要となります。

参考文献

- 1) 東条 敏：自然言語処理入門，近代科学社 (1988)。
- 2) Wallin, N., Merker, B. and Brown, S. (ed.) : The Origins of Music, The MIT Press (2000)。
- 3) Lerdahl, F. and Jackendoff, R. : *A Generative Theory of Tonal Music*, The MIT Press (1983)。
- 4) Sag, I. A. and Wasow, T. : *Syntactic Theory - A Formal Introduction*, CSLI Publications (1999)。
- 5) Tojo, S., Oka, Y. and Nishida, M. : Analysis of Chord Progression by HPSG, in *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications* (2006)。
- 6) Winograd, T. : Linguistics and the Computer Analysis of Tonal Harmony, In *Journal of Music Theory*, Vol.12, No.1 (1968)。

(平成 20 年 7 月 1 日受付)

東条 敏 (正会員)

tojo@jaist.ac.jp

1981 年東京大学工学部計数工学科卒業，1983 年同大学院工学系研究科修了。1995 年同大学院博士（工学）。1983～95 年三菱総合研究所，1995 年北陸先端科学技術大学院大学情報科学研究科助教授，2000 年同教授。自然言語の形式意味論および人工知能の論理の研究に従事。人工知能学会，ソフトウェア科学会，言語処理学会，認知科学会各会員。

