

FSNLP 勉強会 3.2

小町守, mamoru-k

平成 17 年 9 月 27 日

3.2 句構造

語はただ古い順に並べられているわけではない。言語は語順(*word order*) に制約を課している。しかしながら、文中の語は、たとえばネックレスにビーズが連なっているように、品詞の羅列として単に並べられているというわけでもない。そうではなく、語は句(*phrase*)、つまりひとつの単位としてまとめ上げられた語のグループに編成されるのだ。統語論(*Syntax*) というのは語順と句構造の規則と制約についての学問である。

句構造の基本概念の一つに、語のある種のグループが構成素(*constituent*) としてふるまう、というものがある。その語のグループが構成素だと分かるのは、それがいろいろな位置に現われることと、統語的に常に一つの展開に決まるということから判断できる。(3.29) と (3.30) の例文は、名詞と修飾語のグループが、位置と句の展開によって構成素と判断できる証拠を示している。

1. I put *the bagels* in the freezer.
2. *The bagles*, I put in the freezer.
3. I put in the fridge *the bagles* (that John had given me)

これはソシユール言語学での並列関係(*paradigmatic relationship*) の概念である。ある特定の統語的な位置(たとえば上記の名詞句構成素) で互いに交換可能な要素は、全てあるパラダイム(*paradigm*) のメンバーである。対照的に、2 つの語が統語的關係(*syntagmatic relationship*) にあるというのは、それらの語が *sewed clothes* もしくは *sewed a dress* のような句(統語体(*syntagma*) ともいう) を形成するときである。このような関係にある語のクラスで重要なものとして、共起(*collocation*) 関係がある(第 5 章)。

この節では主要なタイプの句についていくつか簡単に触れ、言語学者が句構造をモデル化するのに使うテクニックを紹介する。要点としては、英語の文は典型的に全体で下図のような句構造を持っているのを示すことである。

文全体は S というカテゴリーを与えられる。文は通常主語となる名詞句と動詞句とに書き直される。

名詞句 (*noun phrase*) 名詞は名詞句(NP) に埋め込まれていることが多い。名詞句とは文の統語単位の一つであり、名詞についての情報が集約されている。名詞は名詞句の主辞(*head*) であり、名詞句の統語的な性質を決める中心的な構成素である。名詞句は動詞の項となり、動作の関係者や動作そのもの、もしくは動詞によって描写される状態を示すこともよくある。名詞句は通常限定詞(なくてもよい)、0 個以上の形容詞、名詞という構成素からなり、この順番で配置され、時に前置詞句や修飾節が後置修飾する。名詞の修飾節は関係節(*relative clause*) と呼ばれる。次に示すのは巨大な名詞句で、上に挙げたことのほとんどを示している。

前置詞句 (*prepositional phrase*) 前置詞句(PP) は前置詞を主辞とする句で、名詞句を補部を持つ。前置詞句は他の全てのタイプの句の中に現われることができる。前置詞句は特に名詞句と動詞句の中によく見られ、空間や時間的な位置、またはそれ以外の属性を示すことが多い。

動詞句 (*verb phrase*) 名詞が名詞句の主辞となるのと同様で、動詞が動詞句 (VP) の主辞となる。一般的に、動詞句が文内の統語的に動詞にかかる全要素をまとめ上げる (大部分の統語理論で動詞句は主語の名詞句を含まないのを除く)。動詞句の例のいくつかを (3.33) に挙げる。

形容詞句 (*adjective phrase*) 複雑な形容詞句(AP) はほとんど見かけないが、次のような文で太字に示されるような句が形容詞句の例である。

句構造文法

文の構文解析を行なうと語の意味から文の意味を決定する方法が分かる。たとえば、文に書かれているイベントで誰が何を誰に対してしたのか、といったことが分かる。

文 (3.34) と (3.35) は同じ語を使っているが、違った意味を持っている。最初の文では、Mary から Peter に本が移動しているが、2 番目の文では Peter から Mary に移動している。誰が何を誰に対して行なったのかは語順によって分かる。

ラテン語やロシア語といった言語のように、意味の変化を伴わずに文中の語の順番を入れ替えることができる言語もある。これらの言語は格を用いて誰が何を誰に対して行なったのかを示しているのだ。この種の言語は語順自由 (*free word order*) 言語と呼ばれ、語順が行為者を示すためには使われないことを意味する。そして、語順は重に談話構造を示すために使われることが多い。そういった言語とは違い、英語のような言語は文中での語の移動に関してもっと制約が強い。英語では、基本的な語順は主語 — 動詞 — 目的語である。

(3.36) *The children* (主語) *should* (助動詞) *eat spinach* (目的語).

一般的に、この語順は特定の「ムード」カテゴリーを表現する目的のみにおいて変更される。疑問文 (*interrogative*) (質問) においては、主語と最初の助動詞の位置が逆 (*invert*) になる。

(3.37) *Should* (助動詞) *the children* (主語) *eat spinach* (目的語)?

表現文中に助動詞がなかった場合、*do* の派生形の一つが文頭の位置に現われる (*Did he cry?*)。命令文 (*imperative*) (指図や要求) においては主語はない (言及されている人物であると推論される)。

(3.38) *Eat spinach!*

基本的な文は疑問文や命令文と対置されるときは平叙文 (*declarative*) と呼ばれる。

語順の規則性は書き換え規則 (*rewrite rule*) によって捉えられることが多い。書き換え規則というものは「カテゴリー → カテゴリー*」という形式をしており、左側の記号を右側の記号列に書き換えることができるということを言っている。言語の文を生成するには、開始記号 (*start symbol*) 「S」(文) から始める。単純な書き換え規則を下に示す。

(3.39)

右側の規則は前節で導入された統語カテゴリー (品詞の記号) を対応するカテゴリーの単語 1 つに書き換える規則である。文法のこの部分はよく語彙(*lexicon*) として切り離されている。これらの規則の性質として、ある統語カテゴリーが 1 つ以上の統語カテゴリーか語に書き換えられることがありうる。そのように書き換えられることができるかどうかは、純粹にカテゴリーに依存しており、全く周囲の文脈に依存していない。そのため、こういった句構造文法はよく文脈自由文法(*context-free grammar*) と呼ばれることもある。

これらの規則を用いて、文を導くことができる。(3.40) と (3.41) に簡単な導出 2 つを挙げる。

(3.40)

(3.41)

句構造を表現するのにもっと直観的な方法は、木構造を用いることである。木の葉ノードは終端ノード(*terminal node*) と言い、内部ノードは非終端ノード(*nonterminal node*) と言う。このような木構造においては各非終端ノードとその直下の娘、言い換えると部分木(*local tree*) は書き換え規則を適用することに対応する。娘の順番は文の語順に相当し、全体の木は文法の開始記号となる 1 つのルートノードを持つ。木 (3.42) と (3.43) は (3.40) と (3.41) の導出に対応する。木の各ノードは構成素として仮説を立てているなにかを表現している。

(3.42)

(3.43)

最後に、構成素を示す 3 番目の方法としては、(ラベルつき) 括弧(*bracketing*) を用いる方法がある。括弧の集合で構成素を区切り、ラベルをつけることによって非終端ノードのカテゴリーを示すことができる。(3.43) のラベルつき括弧表記を (3.44) に示す。

(3.44) [*S* [*NP* [*AT The*][*NNS children*]] [*VP* [*VPD ate*] [*NP* [*AT the*][*NN cake*]]]]

書き換え規則という観点での自然言語の文法の定式化が持っている特徴の一つに、再帰性(*recursivity*) がある。再帰性とは、書き換え規則を何回も適用することができる配置がある、という性質のことである。ここで例に挙げた文法では、PP が NP を含み、さらにその NP が別の PP を含むことができる。そのため、図 3.1 の例のような再帰的な展開を作ることができる。この例では、前置詞句の列は書き換え規則のサイクル「NP → NP PP; PP → IN NP」を複数回適用することによって生成される。この導出ではこのサイクルを 2 回適用しているが、3 回、4 回、もしくは 100 回適用することもできる。

再帰性のおかげで、たった 1 つの VP や NP のような非終端記号をいくつもの単語に展開することができる。(たとえば、図 3.1 では記号 VP は 9 単語に展開されている: *ate the cake of the children in the maintains*) このことの一つの帰結として、共通の書き換え規則によって生成された統語的に結びついている 2 つの単語は、文の導出が進行するに従って間に邪魔する語が入ってきて離れることも起こりうる。この種の現象は非局所的依存関係(*non-local dependency*) と呼ばれる。なぜなら、文内で離れたところに出現する 2 単語であっても統語的な依存関係にありうるからである。

局所的ではない依存関係の例として、主語と動詞の一致(*subject-verb agreement*) が挙げられる。主語と動詞の一致というのは、文の主語と動詞の数と人称が一致することである。*She walks, He walks, It walks* に対して *I walk, You walk, We walk, They walk* といった形がある。つまり、動詞が末尾に三人称単数を示す *-s* を持っているのは、主語が三人称単数であるときであり、またその

ときに限る，ということである．主語と動詞は次の例のように他の語や句が間に割り込んで一致する．

(3.45) The **women** who found the wallet **were** given a reward.

直後に隣接する単語だけを見ると *the wallet was* と言わないといけないかのように見える．この文に完全な統語的分析を加えることによってのみ，*The women* が主語であり *to be* の形が複数形である必要があることが分かるのである．

非局所依存の別の重要な例として，*wh*-移動(*wh-extraction*)のような長距離依存関係(*long-distance dependency*)として知られているクラスがある．この名前は(3.46b)の *which book* のような句が，深層的な位置((3.46a)に示されているように動詞のあと)から「表層」的な位置((3.46b)に示されているように文頭)に移動(抽出)するという理論に基づいている．

(3.46) a.

(3.46) b.

このような移動理論とは関係なくとも，*buy* と *which book* の間の長距離依存関係を認識する必要があることは明らかである．そうでなければ，*book* が *buy* の項であることが分からないことになる．

非局所現象は，局所依存をモデル化する *n*-グラムのような統計的自然言語処理的アプローチにとって課題の一つになっている．*n*-グラムモデルだと，(3.45)における *wallet* の後ろの語は *was* であって *were* ではないと予言する．これらの問題は11章の冒頭でより深く議論する．

いろいろな種類の句構造文法を見てきたが，その最後の特性として，空ノードがある．空ノード(*empty node*)は非終端記号が空に書き換えられるときに現れる．たとえば，主語のNPがかうても *Eat the cake!* と言うことができることを考えてみると， $NP \rightarrow \emptyset$ という規則を追加してもよいだろう．すると，このNP非終端記号規則はNPを空に書き換えることができる，という規則になる．これはしばしば木のノードの下に \emptyset が *e* を書くことによって表現される．この概念を使うことによって，(3.46b)の木は(3.47)の構造を持つことができる．

この節で見てきた句構造の単純なモデルは言語の文脈自由性(*context-free*)という観点を前提としている．たとえば，一度「VP」を「VBD NP」に，そして「*sewed* NP」に展開すれば，NPを自分が望むどんな名詞句にでも置き換えることができる．NPを展開する方法を決めた時点で，動詞 *sewed* によって与えられる文脈にはアクセスできなくなるのである．このように文脈へアクセスができないことは，文脈自由言語の重要な性質である．VPを *sewed clothes* のような自然な句に展開することはできるが，それと同様 *sewed wood blocks* のような無意味な展開を選択することも簡単にできるのである．

必要な依存関係を含めるといったことは確率的言語解析の中心トピックであり，12章で議論する．

3.2.2 依存関係: 項と付加語

もう一つ重要な概念として従属語(*dependent*)という概念がある．次のような文を考える．

(3.48) Sue watched the man at the next table.

Sue と *the man* は目撃するという出来事の従属語であり、これらが動詞 *watch* の 2 つの項である。PP *at the next table* は *man* の従属句であり、*man* を修飾している。

通常、名詞句は動詞の項となることが多い。動詞の項はさまざまなレベルで記述される。一つには、意味役割(*semantic role*) によって項を分類することもできる。動作の動作主(*agent*) はなにかをしている人や物であり、受動主(*patient*) はなにかがされている人や物であり、道具(*instrument*) や目的(*goal*) といった役割はまた別の意味関係のクラスを記述している。二つ目に、意味役割で記述する代わりに、文法関係という観点から項の統語的な可能性を記述することができる。英語の動詞は全て主語(*subject*) を取る。主語とは動詞の前に登場する名詞句である。また、英語の動詞には目的語(*object*) を取るものも多い。目的語とは通常動詞の直後に現われる名詞句である。代名詞は動詞の主語となるとき主格となり、動詞の目的語となるとき目的格となる。(3.49) で再び取り上げるが、以前の例では、*children* は *eat* の主語 (子どもは食べるという動作の動作主となっている) であり、*sweet candy* は *eat* の目的語 (甘い飴は動作が行なわれている対象であり、動作の受動主である)。

(3.49) Children eat sweet candy.

candy の形態素的な形は変化しないことに留意せよ。英語では、代名詞が目的格で使われるときに形態を変える唯一の名詞である。

動詞のあとに目的格の名詞を 2 つ取る動詞もある。

(3.50) She gave him the book.

この文では、*him* が間接目的語(*indirect object*)(受取人 (*recipient*), つまり間接的になにかを得る人) であり、*the book* が直接目的語(*direct object*)(受取人を記述する) である。他のこのような動詞は送る動詞かコミュニケーションする動詞である。

(3.51)

これらの動詞は受取人が前置詞句として出現するような項の代替表現を持つことが多い。

(3.52)

格付与がある言語は普通これらの NP を区別し、受動主を対格、受取人を奪格で表す。

意味役割と文法関係の間には、たとえば動作主はだいたい主語であるといったような体系的な関係があるが、同時に乖離もある。*Bill received a package from the mailman.* という文では郵便配達人が動作主である。意味役割と文法関係の間の関係性は態の交替でも変わる(前に説明しなかったが表 3.3 で示してある素性の一つ)。多くの言語では能動態(*active voice*) と受動態(*passive voice*) の間に区別を設けている。能動態は動詞の項を表現するデフォルトの方法、つまり動作主が主語で受動主が目的語に表現される方法に当たる。

(3.53) Children eat sweet candy.

受動態では、受動主が主語になり、動作主が間接的な役割になる。英語ではこれは 2 つの項の順番が逆になり、動作主が *by*-前置詞句を用いて表現されることを意味する。受動態は助動詞 *be* と過去完了を用いて構成される。

(3.54) Candy is eaten by children.

他の言語では、受動態による態の交替は単に格付与の変化と動詞の形態変化だけで行なわれることもある。

サブカテゴリー化

ここまで見たように、動詞は関係する実体(人、動物、物)の数によって異なっている。このような違いの一つに、他動詞(*transitive verb*)と自動詞(*intransitive verb*)の区別がある。他動詞は(直接)目的語を持つが、自動詞は目的語を持たない。

(3.55)

文(3.55a)では、*a bottle of whiskey* が *brought* の目的語となっている。動詞 *bring* は目的語抜きで使うことはできない。つまり、*She brought.* と言うことはできない。動詞 *walk* は自動詞の例である。文(3.55)においては目的語がない。だが、動作の場所を表現する前置詞句が存在する。

文法学者は動詞の従属語を分類しようとしている。文法学者が最初に区別するのは項と付加句である。主語、目的語、直接目的語は項である。一般的に、項(*argument*)は動詞の動作に中心的に関わっている実体を表わしている。多くの項はNPとして表現されるが、PP、VPまたは節として表現されることもある。

(3.56)

項は主語と主語以外の項とに分けられる。主語以外の項はまとめて補部(*complement*)と呼ばれる。付加句(*adjunct*)とは動詞にそれほど強く結びついていない句のことである。補部のほとんどは義務的(たとえば *bring* の目的語は義務的)だが、付加句は常にオプションである。付加句は補部より簡単に動かすこともできる。付加句の典型的な例としては時間、場所、動作の方法や状態を示す句である。たとえば、次の例文で動詞が記述しているような状態がそれである。

(3.57)

従属節(*subordinate clause*)(文の中の文)は付加句またはサブカテゴリー化された項のいずれかであり、動詞に関するさまざまな関係性を表現することができる。以前(3.28)で見た例では、(a)は項の節を示しており、残りは付加句である。

付加句と補部を区別することが難しいときもある。前置詞句 *on the table* は最初の文では補部である(*put* によってサブカテゴリー化されており省略することができない)が、2番目の文では付加句(オプション)である。

(3.58)

(3.59)

伝統的な項/付加句の区別は実のところ伝統的な言語学のカテゴリー観を反映したものである。多くの場合、次の例文のように、どちらとも言える。

(3.60)

イタリックで示されたPPが動詞が記述している事態に中心にかかっていると見なすことができるかどうかははっきりしない。統計的自然言語処理のアプローチでは、動詞と従属語の関係性の度合について議論することは意味のあることかもしれない。

動詞が許す補部のタイプに従った動詞の分類のことをサブカテゴリー化(*subcategorization*)と呼ぶ。動詞が特定の補部をサブカテゴライズすると言う。たとえば、*bring* は目的語をサブカテゴライズする。サブカテゴライズされた項のリストを例文つきで下に掲げる。

主語 The *children* eat candy.

目的語 The children eat *candy*.

前置詞句 She put the book *on the table*.

叙述形容詞 We made the man *angry*.

原形不定詞 She helped me *walk*.

to 不定詞 She likes to *walk*.

分詞句 She stopped *singing that tune* eventually.

that 節 She thinks *that it will rain tomorrow*. *that* は通常省略することができる。She thinks *it will rain tomorrow*.

疑問節 She is wondering *why it is raining in August*. She asked me *what book I was reading*.

これらの補部の大部分はこれまで見たような NP や AP のような句単位であるが、最後のエントリーは S より大きい単位であり、句単位ではない。節 *why it is raining in August* は完全文 *it is raining in August* と先頭の付加構成素によって構成されている。このような「大きい節」は S' (「S バー」と発音する) 構成素と呼ばれる。関係節と主節疑問文も S' 構成素として分析される。

動詞はいくつかの項パターンが可能であることが多い。動詞が現れることのできる特定の項セットはサブカテゴリー化フレーム (*subcategorization frame*) と呼ばれる。以下に英語でよくあるサブカテゴリー化フレームを挙げる。

自動詞 NP[主語] *The woman walked*.

他動詞 NP[主語], NP[目的語] *John loves Mary*.

二重目的語他動詞 NP[主語], NP[直接目的語], NP[間接目的語] *Mary gave Peter flowers*.

自動詞と PP NP[主語], PP *I rent in Paddington*.

他動詞と PP NP[主語], NP[目的語], PP *She put the book on the table*.

文補部 NP[主語], 節 *I know (that) she likes you*.

他動詞と文補部 NP[主語], NP[目的語], 節 *She told me that Gary is coming on Tuesday*.

サブカテゴリー化フレームは補部の統語的規則性を捉えることができる。選択制限 (*selectional restriction*) と呼ばれる意味論的規則性もある。たとえば、動詞 *bark* は主語として犬が選択される。動詞 *eat* は目的語として食べられるものが選択される。

(3.61)

(3.62)

選択制限に従わない文は変に聞こえる。

(3.63)

選択制限は 8.4 節で詳しく議論する。

3.2.3 X' 理論

上で示した句構造規則は、自然言語における句が生成される方法に構造的な性質がなく、節内でのいろいろな種類の依存関係を示すための規則性もないことを暗示している。しかしながら、最近の統語論はこのような規則がたくさんあることを強調している。重要な考え方の一つに、語が句の主辞(*head*)になるというものがある。名詞句と前置詞句について話す理由は、名詞句と前置詞句がそれぞれ名詞と前置詞からなる構成素であるからだ。名詞や前置詞が名詞句と前置詞句のヘッドになる。言語学者は句のヘッドの周りに依存するものを配置する方法に大局的な構造的な性質があることを詳細に議論している。ヘッドはヘッドの補部と小さな構成素を形成する。この構成素は修飾語によって修飾されより大きな構成素となり、最終的に指定詞(*specifier*)と結合して構成素を形成することができる。たとえば主語と限定詞のようなもので最大の句を形成することができる。一般的な見取り図の例が(3.64)に示されている。

(3.64)

中間の構成素は N' ノード(「 N バーノード」と発音する)として参照される。これは基本的に2レベルのバー理論(XP を X'' と考える)であるが、名詞が任意の数の形容詞句による修飾を受けられるように、 N' のレベルで再帰的に修飾語を付加(*adjunction*)することができる。これより多くのレベルのバーを使う理論やこれより少ないレベルのバーを使う理論を採用する人もいる。

この議論の最後のステップとしては、語順には差異があるかもしれないが、構成素の一般的なパターンは句タイプの間で反復されているということである。このアイデアは、 X は語彙カテゴリー間の変数として、 X' 理論と呼ばれる。

3.2.4 句構造の曖昧性

ここまで書き換え規則を文の生成(*generation*)に用いてきた。だが、書き換え規則は構文解析(*parsing*)に用いることのほうが多い。構文解析とは、特定の単語列を生成する導出や句構造木を再構築するプロセスのことである。文から構成された句構造木のことを解析木(*parse*)と言う。たとえば、(3.43)に示す木は文(3.41)の解析木である。

ほとんどの場合、特定の単語列を生成する句構造木はいくつもある。英語の包括的な文法に基づいた解析器は通常ある文に対する解析を何通りも与える。この現象は句構造の曖昧性、もしくは統語的曖昧性(*syntactic ambiguity*)と呼ばれている。イントロダクションの例(1.10)、*Our company is training workers.*は統語的に曖昧な文の例である。特に頻繁に出現する統語的曖昧性の一つには付加の曖昧性(*attachment ambiguity*)がある。

付加の曖昧性とは、2つの異なるノードによって生成される句があるときに起きる。たとえば、(3.39)の文法に従うと、(3.65)の前置詞句 *with a spoon* を生成するのに2つの方法がある。

(3.65) the children ate the cake with spoon.

この前置詞句は図3.2(a)に示す解析木のように動詞句の子として生成することもできるし、図3.2(b)に示す解析木のように片方の名詞句の子として生成することもできる。

図 3.2

異なる付着は異なる意味を持つ。動詞句への「高」位置での付着は子どもたちがケーキを食べるときに使った道具についての言明になるが、「低」位置での名詞句への付着はどちらのケーキ（スプーンがついたケーキであって、たとえば冷やされているケーキではない）が食べられたのかについての言明になる。そのため、付着の曖昧性を解消することは正しい意味解釈を与えるために重要である。

統語的曖昧性についてよく研究されているサブクラスとして袋小路文(*garden path*) といった現象がある。袋小路文とは、読んでいて突然使えないことが分かるようなパスに誘導されるような文のことである。たとえば、文中で突然一見そこには属さないように見える語が追加で見つかるような場合がある。

(3.66) The horse raced past the barn fell.

文(3.66)は(Bever 1970)から取った文で、袋小路文の恐らくもっとも有名な例である。語 *barn* に到達した時点でほとんどの人は大まかに言って「The house ran past the barn.」という意味に相当する解析を構築しているが、その次にこの解析に単純に付加できない追加語 *fell* に出くわす。*raced* までバックトラックし、*The horse fell after it had been raced past the barn.* という意味に相当する全く異なった解析を構築しなければならないのである。袋小路文は最初間違った解析にはめられ、そのあと正しい解析を構築するためにバックトラックするという現象である。

袋小路文は話し言葉ではほとんど問題にならない。というのは、通常、意味的な選好性とコミュニケーションの格率に従う話者の寛容性、そしてイントネーションのパターンの3つ全てにより、袋小路文は回避される(MacDonald et al. 1994; Tanenhaus and Trueswell 1995)。文(3.66)でそれが分かる。文(3.66)では *horse* と *raced* の間にイントネーションの小休止があり、聞き手は *raced* が関係節を導く可能性、つまり主節の動詞ではない可能性を取り除く。しかしながら、袋小路文は英語の書き言葉で複雑な文を読解するときには本当の問題になりうる。

この節では統語的曖昧性による1つ以上の解析がある文の例を見てきた。ほとんどの文はこのタイプである。しかし、文が解析を持たないこともありうる。その理由は、この文法でカバーされない文の生成にある規則が使われることかもしれない。他の可能性としては、文が非文法的(*ungrammatical*) つまり統語的に正しくないからかもしれない。下に非文法的な文の例を挙げる。

(3.67) *Slept children the.

非文法的であることと意味的に正しくないことを区別することは重要だ。次のような文はおかしいが、それが変に聞こえるのは意味的な解釈が首尾一貫していないからであり、(3.67)がいかなる解釈も持たないのとは異なる。

(3.68) a. Colorless green ideas sleep furiously.

(3.68) b. The cat barked.

統語的なおかしさを示すために以前導入した記号(*)と区別するために、意味的、語用論的、文化的なおかしさを示すためにはハッシュ記号(#)が使われることが多い。