社会科学論集　第 149・150 合併号　2017.3

《論文》

発話ジャンルと文節係り受け木の形態的特徴

高　松　　亮

1　背景と目的

1.1　発話行為の社会的文脈と発話のスタイル(1)

一般に我々が言葉を用いるときには，表現したい意味内容が同一であっても，必要に応じてさまざまなスタイル（文体）を採用する。話し言葉で「あしたは晴れるらしい。」という常体での表現が用いられることは少なく，例えば「あすは晴れるそうです。」「あしたは晴れるんだって。」「あした晴れらしいぜ。」といったように，様々な表現が用いられる。書き言葉においても，新聞の文体，論文の文体，文芸作品の分野固有の文体など，さまざまなスタイルが存在する。

ある話者が発話を行なう際には，その話者が既に習得しているスタイルの中から，その発話の社会的な文脈に良く整合しているものを選択する可能性が高い。ここでの社会的な文脈とは例えば，発話場面の種類とその場面における話者の役割，やりとりをする話者同士の社会的関係性，発話行為の話者による社会的な位置付け，発話を伝達する文化・社会的さらには物理的な媒体（メディア）の種類等々，様々なものを想定できる。本論文ではそのような社会的文脈を発話のジャンルであると考え，「発話ジャンル」と呼ぶことにする。

発話ジャンルを規定する代表的な要因とその具体例を以下に示す(2)。

1）発話場面の種類と話者の役割：レストランの客，相撲中継のアナウンサー，世間話をする職場の同僚，など
2）話し手同士，あるいは話し手と聴き手の社会的な関係性：親疎の度合，年齢の差異，ジェンダー，上司 / 部下，先生 / 学生，など
3）話者による発話行為の社会的な位置付け：公的 / 私的，発話行為が話者自身の社会的評価・役割に与える影響，など
4）メディア（媒体）：やりとりのトポロジカルな構造（多対多か，一対多か，など），やりとりに参加する人数，やりとりの即時性に対する要求の程度（会話か，手紙か，など），物理的実体の揮発性の程度（話し言葉のように音としての発話がただちに消滅するか，書き言葉のように長期にわたって消滅しないとみなせるか），やりとりが主として片方向か双方向か，など

1.2　スタイルと文節係り受けの構造

発話とは，話者が言語的単位を組み合わせた（統べた）統語的な構造物としての文を産出する行為であるから，言語的単位を組み合わせた構造（すなわち統語的構造）の傾向がジャンルによってどのように異なるか，という観点からスタイルを捉えることができる可能性がある［Biber and Vasquez2008］［Iwasaki2005］［Chafe1994］［Halliday and Hasan1989］［Chafe1982］。

言い換えると，どのように言語的単位を組み合わせるかというルールには，発話された文が文（文法的に許容される文）か非文（文法的に許容されない文）かという大きな区別にかかわるもの以外に，「そのジャンルによりふさわしい統語的構造はどのようなものか」という範疇が存在し，我々はそれをジャンルに応じて適用している，ということになる。

したがって，ある発話が持つ統語的な構造を何

らかの特徴量で表すことができれば，ある発話ジャンルのスタイルの統語的な特徴を定量的に記述し，異なる発話ジャンル間の統語的な特徴の差異を明らかにすることが可能になる。

日本語の文節と文節との間に修飾―被修飾の関係があることを文節同士の係り受けと呼ぶ。修飾する側の文節を係り元，修飾される側の文節を係り先という。文節係り受けは文の構成要素同士の統語的関係の一種であるから，係り受けの様態を定量化すれば，それを用いて文のある種の統語的特徴を表すことができる。

そのような手法を用いた既存研究としては，係り受け関係を有する 2 つの文節間に存在する文節の個数，すなわち係り受けの距離が，拡張された Zipf の法則に従うことを指摘した研究［丸山・荻野 1992］，小説の書き手が変わっても係受けの距離の分布はほとんど変化がないことを示した研究［金 1996］，ある文節の係り先になっている文節が，さらに他の文節へと係る，ということが繰り返される際の繰り返しの回数に注目し，ある文における最大の繰り返し回数を「係り受けの次数」と定義して，発話ジャンルによって係り受けの次数の平均値が異なることを示した研究［国立国語研究所 1955］などがある。

これらの研究はいずれも，係り受け関係にある 2 つの文節がどの程度離れているか，という 1 次元的な捉え方から構成された特徴量に基づくものである。ところが，以下に述べるように，文節同士の係り受け関係には 2 次元的な広がりを持つ木構造をなすという特徴がある。そのような係り受け木の形状，すなわち形態的特徴に注目したスタイルについての研究は存在しなかった。

以下では，文節係り受けの構造を木構造として表現したものを文節係り受け木，あるいは単に係り受け木と呼ぶ。図 1 に係り受け木の例を示す。図に示すように，各文節を「ノード」，係り元がなく，係り先がある文節を「葉」，係り元はあるが係り先のない文節を「根」と呼ぶ。係り受け関係のあるノード間を結ぶ線を「エッジ」，あるノード P から根 R に向かってエッジをたどる最短経路を考えるとき，経路上のノード Q に到達するまでに経たエッジの数を「P と Q の距離」，P から根 R までの距離を「ノード P の高さ」，ある木の葉から根までの高さの最大値を「木の高さ」という。

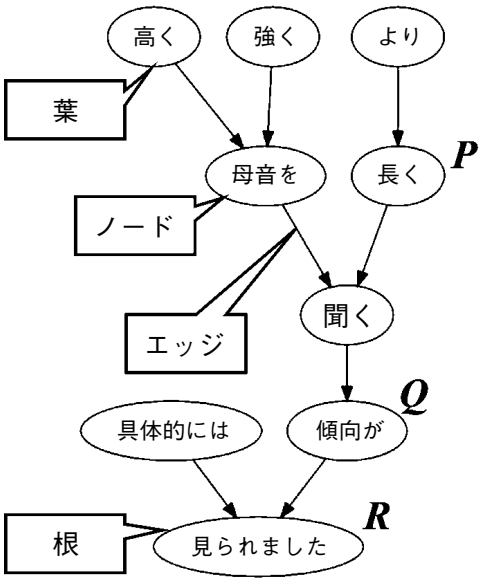

図 1　文節係り受け木の例

本論文では，係り受け関係を木構造として捉え，その特徴量に基いて発話を分析することで，発話ジャンルに依存した統語的なスタイルの傾向を明らかにすることを試みる。木構造はノードとエッジの繋り方のみで一意に定まるのであるから，ノードとエッジの繋り方の特徴を表す適切な量を定義できれば，ある文節係り受け木によって表現された，統語的な構造の持つ特徴を定量化できる。

2　分析対象と分析の方針

本論文において分析の対象とした言語資料は，『日本語話し言葉コーパス』（Corpus of Spoken Japanese, 以下 CSJ）に含まれる「学会講演」と「模擬講演」の 2 種類の講演データである(3)［国立国語研究所 2006］。

「学会講演」は，理工学，人文，社会の 3 領域におよぶ種々の学会における研究発表を収録した

表 1：分析対象の種類と規模（括弧内は共通話者 6 名についての値）

	話者数	節単位総数	木の総本数
学会講演	70（6）	8516（790）	8723（794）
模擬講演	107（6）	9675（613）	10046（640）

ものである。収録音声の多くを占める理工学系の学会では，発表者が男性の大学院生であることが多いため，話者に年齢と性別の偏りがある。また，発話は概してあらたまり度が高い。

一方，「模擬講演」は，できるだけ年齢と性別のバランスをとった一般話者による，日常的話題についての講演である。話者があらかじめ数種類の中から指定されたテーマに基づいて，具体的な講演内容を決めてタイトルをつけ，数名の聞き手の前でスピーチをおこなった模様が収録されている。また，発話は概して学会講演よりもくだけたものとなっている。

両者を社会的文脈という観点から見ると，いずれも複数の，必ずしも個人的に面識や親密さがあるわけではない聴衆の前で，単独の話者が特定の話題についての詳細な説明を筋道をたてて話し言葉で即時的に行なうという点では共通しているが，学会講演は学術的な研究についての発表という専門性の高い場面であって，発話には高い客観性や論理性が要求されるのに対し，模擬講演は日常的な専門性の低い事柄についての発表であって，発話には多少の主観性や論理的な流れの揺らぎがあっても問題にはならないという点が異なる。したがって，学会講演と模擬講演は学術的専門性の有無，客観性・論理性に対する要求の程度という要素が異なる発話ジャンルとみなすことができる。

以上のように発話ジャンルを構成する社会的文脈のうち特定の要素のみが異なり，かつ文節係り受けについての精度の高いラベル付けが施されているデータ同士を比較すれば，発話ジャンルに依存して統語的なスタイルがどのように変化し得るかの一端を係り受け木の特徴量を用いて明らかにできると考えられる。

なお，学会講演と模擬講演の話者の中には双方に収録されている話者が若干名いる。これらの話者を以下「共通話者」と呼ぶことにする。共通話者はいずれも学術的なリテラシーを十分に習得した研究者であり，学会で発表するスタイルと日常的な話題について講演するスタイルの双方を習得していると考えられる。共通話者は学会講演と模擬講演で異なるスタイルで発話するものと考えられるが，そのスタイルの差異が，全話者についてのスタイルの差異と一致するならば，その差異は主として学会講演と模擬講演という発話ジャンルの違いから来るものであって，学会講演と模擬講演それぞれに属する話者の平均的属性の差異から来るものではない，ということを明らかにできる。以下の分析ではそのような観点から共通話者と話者全体についての分析結果を比較し，双方の結果に共通するスタイル上の差異を，学会講演と模擬講演のスタイルの差異と考える。

一般に話し言葉では文の終わりが書き言葉に見られるような文末表現の形をとらず，明示的に示されないことが多い。そこで CSJ では文を記述の単位としては用いず，その代わりに節単位という，節に基いた概念を用いている［国立国語研究所 2006］。係り受け先の記述は個々の節単位を超えない範囲としているため，ほとんどの場合 1 本の係り受け木は 1 個の節単位に対応する。ただし，係り元があって，係り先のない文節が節単位中に複数存在する場合もあり，その場合にはそれぞれの文節を根に持つ複数の木を考えることにする。

表 1 に学会講演と模擬講演の話者ならびに節単位の数，および係り受け木の本数を示す。

3　分析と考察

3.1　はじめに

学会講演と模擬講演のデータは，年齢や性別と

いった話者の属性の分布が同一ではないため，両者の統計的な性質を単純に比較するべきではないが，前述したように両者に共通の話者（共通話者）が 6 名おり，共通話者の場合と全話者の場合それぞれについて比較することで，特徴量の異同の原因がジャンルなのか母集団の違いなのかをある程度判断できる。以下では，係り受け木の形態的特徴を表現する特徴量として，木の高さのような大域的な特徴と，ある文節に対して係る文節の個数やその平均値のような局所的な特徴について分析する。なお，係り受け木のうち，係り元，係り先の両方が存在しない 1 個の文節のみからなる木は，その多くがフィラーなどであるため，分析対象から除外している。

3.2　大域的特徴

3.2.1　木の高さの頻度

係り受け木の高さの相対頻度の分布を図 2 および図 3 に示す。いずれの図においても学会講演の方が模擬講演よりも分布の幅が狭く，相対的に高い山の度数が多い。両者に共通する特徴としては，学会講演および模擬講演とも木の高さが 2 で最大値の頻度となり，それよりも木の高さが高くなるにしたがって頻度が単調に減少することがあげられる。学会講演の木の高さの平均値は 3.45（全話者）および 3.27（共通話者），模擬講演の木の高さの平均値は 2.98（全話者）および 2.88（共通話者）であり，学会講演が模擬講演よりも木の高さの平均値が大きい。

また，全話者と共通話者の双方で同様の傾向を示すことから，学会講演と模擬講演による分布形の差異は，母集団の属性の偏りというよりは，ジャンルに起因する違いであることが推察される。

［国立国語研究所（1955）］においては「係り受けの次数」という，係り受け木の高さと同等のパラメータを用いて文の構造を分析しており，ニュース音声と日常的な場面における対話音声それぞれに表れる文の次数を比較した結果，ニュース音声（平均値 3.76）が対話音声（平均値 1.77）よりも次数の高い文が頻出すると指摘している（平均値は筆者による再計算）。参考のために本論文における値も含め，木の高さの値の順に並べると，

ニュース音声 > 学会講演 > 模擬講演 > 日常対話

となる。

ニュース音声は独話で，かつ改まり度が高く，本論文における学会講演に近い性質を持っている。また，本論文における模擬講演は比較的くだけた状況における独話であり，日常の対話とニュースや学会講演の中間的な性質を有すると考えられ，このことが木の高さの平均値の大小にも表れているものと考えられる。

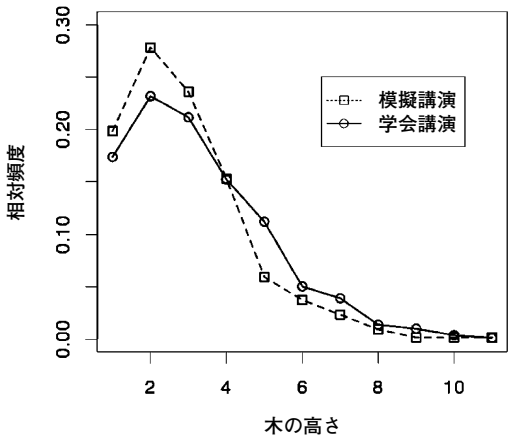

図 2　木の高さの頻度（全話者）

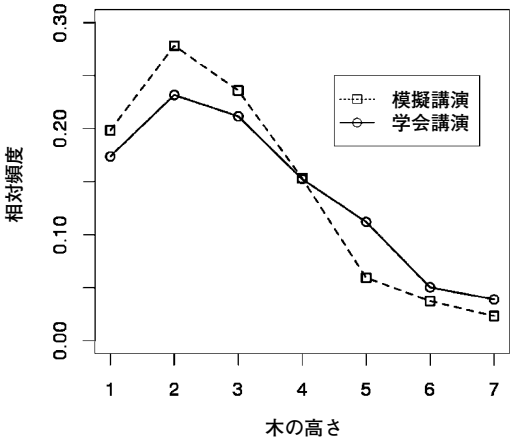

図 3　木の高さの頻度（共通話者）

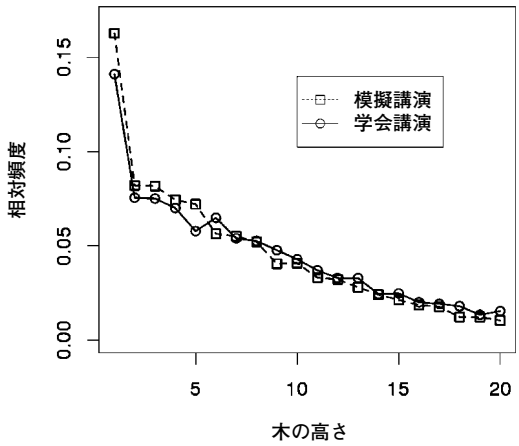

図 4　木に含まれる文節数の頻度（全話者）

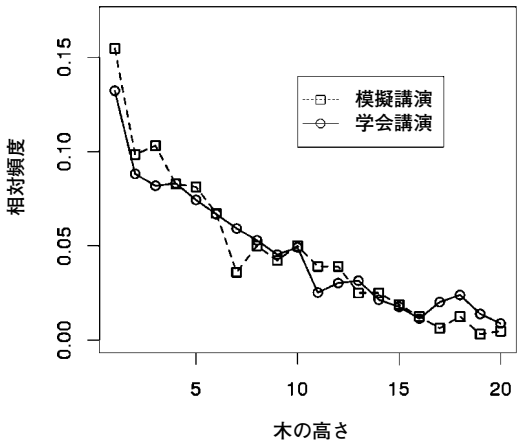

図 5　木に含まれる文節数の頻度（共通話者）

3.2.2 文節数の頻度

1 本の係り受け木に含まれる文節の数は，木の規模の大小を表現するパラメータの一つである。図 4 および図 5 に文節数の相対頻度の分布を示す。図より，木の高さの頻度の場合と同様に，共通話者の場合も，話者全体の場合もかなり類似した傾向があることがわかる。すなわち，いずれの場合も文節数 2（図の最も左側のプロット）の頻度が例外的に高く以降単調に減少すること，文節数が 2 においては模擬講演の頻度が高く，3 から 5 程度の範囲ではその差はわずかになり，それよりも文節数が多い領域においては，逆に学会講演の方がわずかに頻度が高いことがわかる。

これらの特徴が図 4 と図 5 に共通して見られることから，文節数の頻度分布の傾向も，学会講演と模擬講演というジャンルの違いから生じていることが推察される。

3.3　局所的特徴

3.3.1　係り元の文節数

係り受け木の局所的な特徴のうちもっとも基本的なものとして，ある文節に注目した場合に，その文節に係る文節（係り元）の数が n 個である場合の頻度を考える。図 6 および図 7 に係り元の数の相対頻度の分布を示す。なお，縦軸は相対頻度の常用対数である。

いずれのグラフもプロットの傾きがほぼ負の直線上にのっていること，係り元の数が 0，すなわち文節が葉である場合の相対頻度が学会講演と模擬講演とで一致すること，係り元の数が 0 ～ 3 ないし 4 個の領域では学会講演が，それ以上の領域では模擬講演が，それぞれわずかずつ頻度が高いことがわかる。共通話者と全体話者で傾向が一致することから，学会講演と模擬講演の間に見られたわずかな差異が，スタイルの差異から生じたものである可能性がある。

3.3.2　根の文節に係る文節数

係り受け木の根に相当する文節に，n 個の文節が係る場合の相対頻度を求めたものを図 8 および図 9 に示す。

学会講演は文節数 2 において最大値を，模擬講演は文節数 1 において最大値を取る。また，学会講演の方が分布の幅が相対的に狭い。これらの傾向が両方の図において見られることから，以上の差異が学会講演と模擬講演のスタイルの違いから生じている可能性がある。

3.3.3　葉の高さと葉の累計係り元数

ある葉の高さが n であるとき，葉から根まで辿って行く際に通過する各文節 N_i（i=1,2,…,n）が係り元を d_i 個ずつ持っているなら，d_i の合計数をその葉の累計係り元数と呼ぶことにする。葉の高さと累計係り元数の平均値の関係を図 10 お

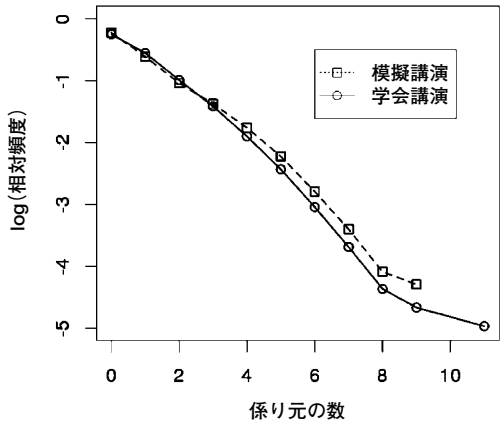

図 6　係り元の文節数の頻度（全話者）

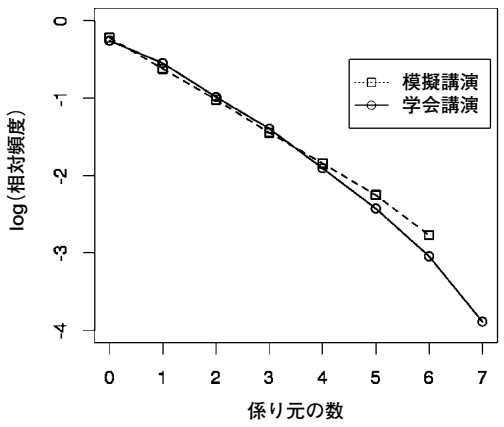

図 7　係り元の文節数の頻度（共通話者）

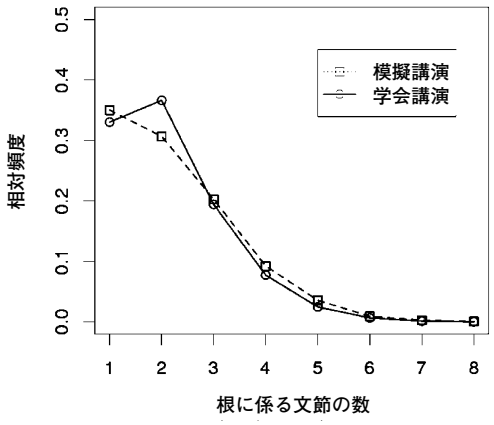

図 8　根の文節に係る文節数の頻度（全話者）

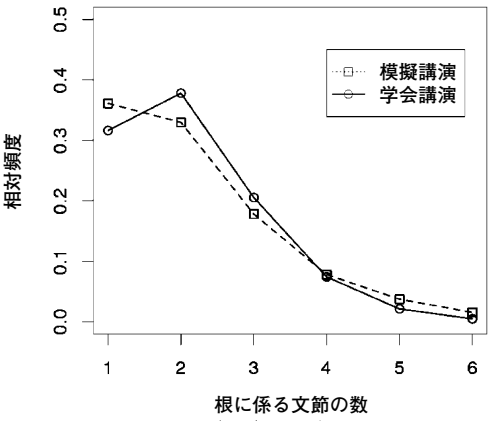

図 9　根の文節に係る文節数の頻度（共通話者）

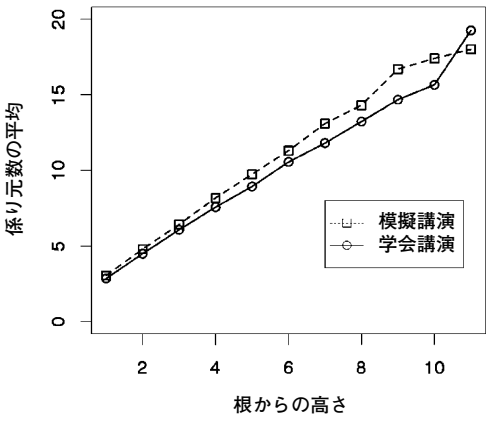

図 10　累計係り元数の平均値（全話者）

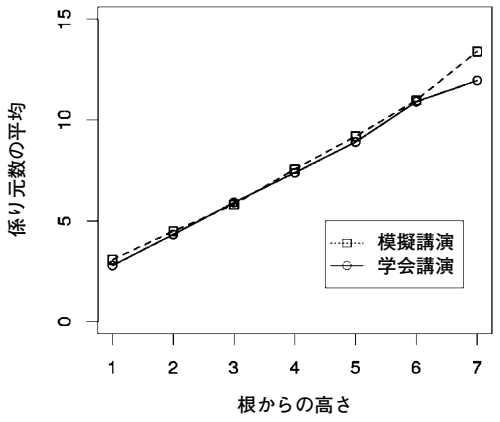

図 11　累計係り元数の平均値（共通話者）

よび図 11 に示す。

全てに共通する特徴として，葉の高さが 1 から 6 ないし 7 程度までの範囲においては，プロットの傾きが正の直線上に良くのっていることが挙げられる。全話者においてはこの直線の傾きが模擬講演と学会講演とで異なり，学会講演の方が傾きが若干小さく，葉の高さが高くなった場合の係り元数の増加が少ない。一方，共通話者においては，学会講演の方が傾きが小さい点は全話者と同じではあるが，その差はごくわずかである。したがって，傾きの差異が発話ジャンルに起因している可能性はあるが，話者によってはそれほど明確な差が生じないことがあることがわかる。

4　まとめ

本論文で得られた知見のうち，ジャンルによる量的な差異に関するものをまとめると次のようになる（A は学会講演，S は模擬講演を指す）。

- 大域的特徴
  - 木の高さの分布の平均：A > S
  - 木の高さの分布の幅： A < S
  - 文節数 2 の木の相対頻度： S > A
- 局所的特徴
  - 葉の相対頻度：A=S
  - 根に係る文節数の分布の最頻値：A= 2 / S= 1
  - 根に係る文節数の分布の幅：A < S
  - 高さ n の葉から根までの累積係り元数：n に比例して増加（比例定数は A < S）

以上から，学会講演は木の高さが高く，高さの分布の散らばりも小さいこと，模擬講演は文節数が 2（すなわち高さで言えば 1）の木の相対頻度が相対的に多いことがわかる。また，高さ n の葉から根までの累積係り元数は n にほぼ比例するが，学会講演の方が比例定数が小さいことから，葉が高い位置にあっても，根からその葉までの経路での枝分れがより少ない。これらの特徴を一言でまとめれば，学会講演は木の高さが高いが，枝分れの少ない構造を持つ傾向がある，ということになる。そのような構造の実例を図 12 に示す。

この図の構造を今回得られた知見に沿って考えると，次のように説明できる。テ形節『僕のとってもかわいい〜書いて』から文末『求めています』に至る経路は 10 回分もの多くの係り受けを経る。また，係り元の個数が 1 個の文節が 9 箇所，2 個の場合が 6 箇所，3 個の場合が 2 箇所であり，それ以上の係り元がある文節は存在しない。学会講演のような場面においてこのような形態の文を用いることは，ある経路の係り受けの回数を増やすことで論理的に複雑な意味内容を表現する一方，文節に係る文節数を減らし単純な修飾構造を用いることで，話し手が文を生成する際と，聞き手が文を理解する際それぞれの認知的負荷を減少させ，意味内容を正確かつ確実に伝えることに寄与している可能性がある。

一方，図 6 および図 7 に示されるように，文節の出現頻度は係り元の数が多いほど急速に減少するが，模擬講演の方がよりロングテールな傾向を持つことから，模擬講演には 1 つの文節に多数の

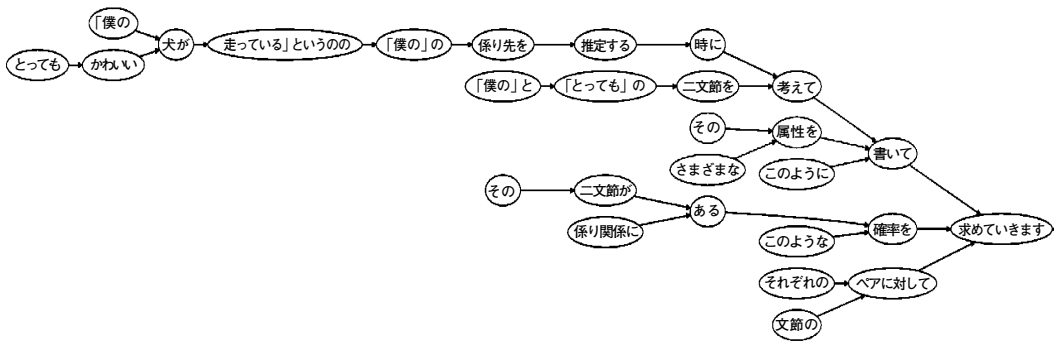

図 12　高さが高く枝分かれの少ない係り受け木の例（学会講演）

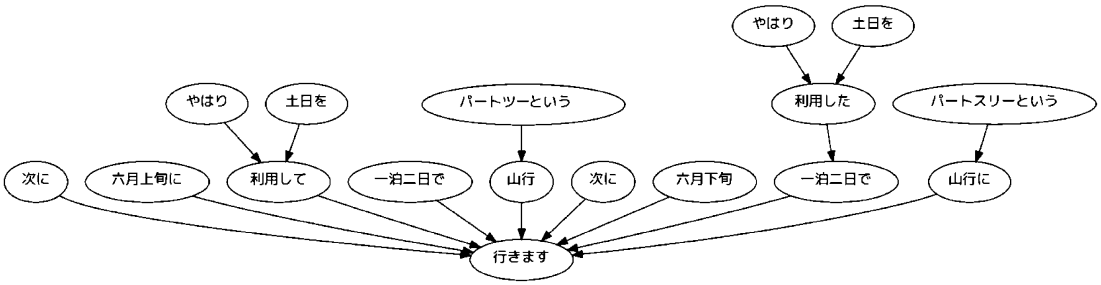

図 13　1 つの文節に多数の文節が係る構造の例（模擬講演）

文節が係る表現が相対的に多いことがわかる。そのような構造の実例を図 13 に示す。述語『行きます』に多くの係り元からの係り受けがあること，高さ 1 の葉が 5 個，高さ 2 のものが 4 個，高さ 3 のものが 2 個と，葉が低いことがわかる。学会講演に比較して模擬講演においては，論理的に複雑な意味内容を説明する必要性がないので，係り受けの回数を増やす必要がないこと，それゆえに述語に多くの係り受けを集中させても聞き手が文を理解することは容易であることが，このような構造を生んでいる可能性がある。

5　まとめと今後の課題

本論文では係り受け木の形態を表す特徴量として，大域的なものと局所的なものとを用い，学会講演ならびに模擬講演という社会的文脈の異なる 2 つの発話ジャンルの統語的なスタイルが，それぞれどのような傾向を持つのかを調査した。その結果，学会講演は木の高さが高く枝分れが少ない傾向があるのに対し，模擬講演は木の高さが低く枝分れが多い傾向が見られることを示した。

今後の課題としては次のようなものを挙げることができる。

学会発表には理工学，人文，社会の異なる領域の学会が含まれている。領域それぞれの学術的リテラシーには異なる部分があり，例えば人文系の学会では原稿を読み上げるような発表スタイルが存在するのに対して，理工系の学会はそのようなことが許されない。このようないわば「文化的差異」が発話の統語的スタイルに与える影響を明らかにする必要がある。

また，学会講演と模擬講演のデータに含まれる話者の性別や年齢の分布には差があり，それが両者の統語的スタイルにどの程度影響しているかをより精密に調べることも重要な課題である。

より長期的な課題としては，まず，節単位に含まれる接続節の構造を考慮した分析をすることで，発話ジャンルによるスタイルの違いについてより明確な知見を得る必要がある。また，本論文で得られた学会講演と模擬講演の係り受け木の特徴が，発話生成過程ならびに文理解における認知的負荷と実際に関連しているか否かを明らかにすることも重要であろう。

謝辞

本論文は筆者が国立国語研究所に外来研究員として 2012 年 4 月から約 1 年間にわたって滞在した折に行った研究が元になっている。言語学に関していわば初学者であった筆者にさまざまな御示唆を下さった国立国語研究所の前川喜久雄氏ならびに小磯花絵氏，このような貴重な機会を得るきっかけを下さった早稲田大学人間科学学術院の菊池英明氏，そして通称「モニター室」で日々ご一緒させて頂いた方々に謝意を表する。

《注》

(1) 本論文では話し言葉・書き言葉にかかわらず人が文章を表出することを発話，話し手ないし書き手のことを話者と呼ぶ。

(2) これらが全ての要因でないこと，個々の要因と他の要因が分離できることを示したものではないことは言うまでもない。

(3) 正確には，これら講演データのうち，文節係り受けについて手作業による綿密なラベル付けが施されている，「コア部分」を分析対象とした。

参考文献

［Biber and Vasquez 2008］ Biber, Douglas and Cammila Vasquez “Writing and Speaking”, in Handbook of research on writing, ed. C. Bazerman, pp.535-548, Routledge, Oxford, 2008

［Iwasaki 2005］ Iwasaki, Shoichi “Multiple-grammar hypothesis: a case study of Japanese passive constructions”, Phylogeny and Ontogeny of Written Language, Kyoto University, August 17, 2005.

［Chafe1994］ Chafe, W. “Discourse, consciousness, and time.”, Chicago and London, The University of Chicago Press, 1994.

［Halliday and Hasan 1989］ Halliday, M.A.K. and R. Hasan “Language, context, and text: aspects of language in a social-semiotic perspective”, Oxford, Oxford Univ. Press, 1989.

［Chafe1982］ Chafe, W. “Integration and involvement in speaking, writing, and oral literature.” in D. Tannen (Ed.), Spoken and written language: Exploring orality and literacy, pp. 35-54, Ablex, 1982

［丸山・萩野 1992］ 丸山 宏，荻野 紫穂，“日本語における文節間係り受け関係の統計的性質”，情報処理学会全国大会講演論文集，vol.45, no.3, pp.173-174, 1992 （http://ci.nii.ac.jp/naid/110002889591 よりダウンロード可能）

［金 1996］ 金 明哲 “文節の係り受け距離の統計分析”，社会情報：札幌学院大学社会情報学部紀要 , vol.5, no.2, pp.1-11, 1996.（http://hdl.handle.net/10742/754 よりダウンロード可能）

［国立国語研究所 1955］ 国立国語研究所，“談話語の実態”，国立国語研究所研究報告 8, 1955
（http://db3.ninjal.ac.jp/publication_db/item.php?id=100170008 よりダウンロード可能）

［国立国語研究所 2006］ 国立国語研究所，“日本語話し言葉コーパスの構築法”，国立国語研究所研究報告 124，2006（http://www.ninjal.ac.jp/csj/k-report-f/CSJ_rep.pdf よりダウンロード可能）

《Summary》

The Relation Between Genres and the Morphological Tendencies of Dependency Trees of Spoken Japanese

Ryo Takamatsu

Abstract

In this paper, we focus on a quantitative analysis of stylistic differences using features which represent local and global forms of dependency trees. We investigate differences between two styles of spoken Japanese (“Simulated Public Speaking (SPS)” and “Academic Presentation Speech (APS),” which are included in Corpus of Spoken Japanese). The results show that the forms of dependency trees are “tall and narrow” in APS and “low and wide” in SPS, and this demonstrates the fact that the proposed features, while effectively depicting distinct tendencies, arise from different styles.

Keywords: style analysis, Japanese linguistics, dependency tree, Japanese bunsetsu unit dependency, Corpus of Spoke Japanese