

## 文構造の漸進的予測を可能にする日本語の諸特徴の分析

木田敦子<sup>1,2</sup> 高梨克也<sup>2</sup> 乾裕子<sup>2,3</sup> 井佐原均<sup>1,2</sup>

{kida, takanasi, hinui, isahara}@crl.go.jp

1 通信・放送機構 2 通信総合研究所 3 計量計画研究所

### 1. はじめに

聞き手は進行中の文のある時点において、これから話し手によって発話されるであろう事柄を予測できる場合がある。この現象から、文の文法的特徴の中にこうした予測を可能にする特徴が潜在していると考えられる[木田01]。本研究では「聞き手は進行中の文中のどの時点で、何を予測できるのか?」という問題を解明していくために有効な手順と、それによって解明された現象の持つ文法的意味について考察する。

機能をより重視した日本語文法理論[益岡91]においては、このような聞き手への伝達に関わる現象として、いわゆる呼応現象が重視されている。本研究では、呼応を、文の前半に生起する「呼」の要素から、これと共起する「応」の要素が後続部分に生起することが予測できる関係と捉える。本稿では以下に挙げる四つを作業の中核に据える(図1)。1) 呼応関係を文の予測可能性の解明の出発点として採用し、2) コーパス中の実例に呼応関係がどの程度生じているかを確認する。さらに、3) 個々の呼応関係のみから聞き手の予測を説明するだけでは不十分な点を発見するための分析手法を紹介するとともに、4) こうした手法によって明らかになった傾向について報告する。

### 2. 出発点となる作業データの作成

#### 2.1 モダリティの階層関係に基づく辞書と規則

呼応関係に関する従来の知見を整理するために、モダリティの階層関係[益岡91]に基づいて呼応関係テーブルを作成した。呼応関係テーブルは、「呼」カテゴリと「応」カテゴリのリストとそれぞれの呼応規則、そして各カテゴリと規則に該当する代表的語句を表にまとめたものである[表1]。

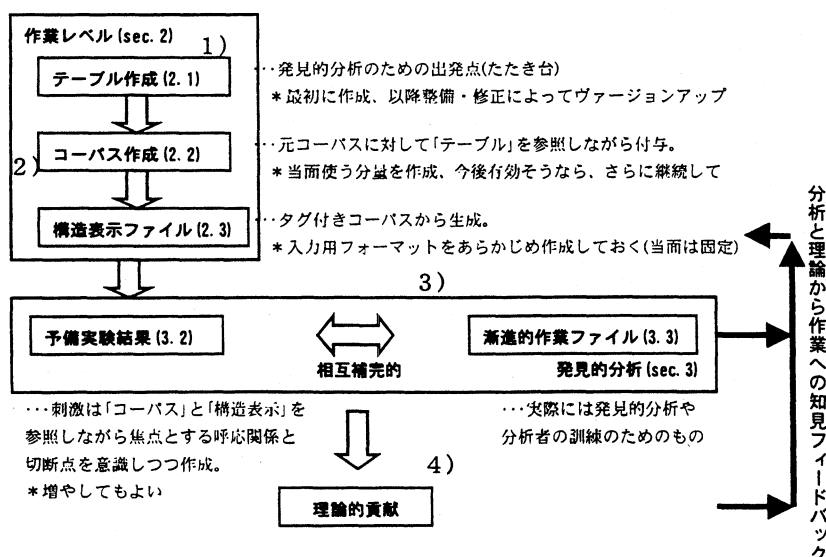
#### 2.2 辞書と規則を参照したコーパスタグ付与作業

##### 2.2.1 分析対象データ

『日本語話し言葉コーパス』モニター版(2001)を使って、呼応関係タグの付与のためのデータを作成する。本コーパスを採用した理由は、これが主に講演という、聴衆に向けた自然な談話を収録したものであり、しかも対話ではなく独話であるため、対話の場合に現れるような聞き手の実際に反応には影響されずに考察が可能となるためである。このコーパスには、学会講演158件(約43.5時間)、模擬講演233件(約43時間)が収録されている。今回は内容が専門的である学会講演は使用せず、テーマに一般性がある模擬講演のデータ10件を使用し、455文の作業データを準備した。

##### 2.2.2 呼応関係タグの付与

『日本語話し言葉コーパス』から作成した作業データに、2.1で作成した呼応関係タグを付与した。



[表 1] モダリティ階層関係に基づく呼応関係辞書

メタ	カテゴリ	辞書	規則		
呼	A:伝達態度	ねえ、おい	A-a		
	B:ていねいさ	-	B-b		
	C:表現類型	ぜひ、なんて	C-c		
	D:真偽判断	たぶん、どうも、いったい	D-d		
	E:価値判断	-	G-g		
	F:説明	-	H-h		
	G:みとめ方	決して、必ずしも	-		
	H:テンス	むかし、かつて、もうすぐ	-		
	I:取り立て	-	-		
	応	i:取り立て	は、も、ばかり、でも、くらい	その他	
		h:テンス	～た(有標形)、[述語の無標形]	} 気づいた点を記入	
g:みとめ方		ない			
f:説明		のだ、わけだ			
e:価値判断		ことだ、ものだ、べきだ			
d:真偽判断		だろう、らしい、ようだ、か			
c:表現類型		て下さい、なあ			
b:ていねいさ		です、ます			
a:伝達態度		ね、よ			
呼応規則		A-a	ねえーよ、おいーよ		}
		B-b	-		
	C-c	ぜひーて下さい			
	D-d	たぶんーだろう、どうもーらしい			
	E-e	-			
	F-f	-			
	G-g	決してーない、必ずしもーない			
	H-h	むかしーた、もうすぐー[無標形式]			
	I-i	-			
	呼応規則	= α			

[表 2] 呼応関係タグ付きコーパスからの構造表示

規則	要素 1	2	3	4	...	n-5	n-4	n-3	n-2	n-1
A-a										ね
B-b										です
C-c										
D-d	どうやら								よう	
G-g		必ずしも							ない	
H-h		昨H						て		
									太郎は	出かけ

### 3. 作業データを利用した分析手法の提案

本章では、第二章で作成したファイルを使用して、呼応関係を含む聞き手の予測に関わる現象を分析する方法を述べる。

#### 3.1 構造表示ファイルからの発見

モダリティの階層関係に基づく呼応規則から構造表示ファイル[表 2]を作成することによって、以下のような点を分析することができる。

##### 3.1.1 「呼」「応」のいずれかの要素がないもの

構造表示ファイルからは、「呼」「応」のいずれかの要素が生起していない例を見つけることは容易である。そのうち特に重要なのは、「呼」はないが「応」がある場合である。予測される側の「応」要素があることから、予測を生み出す要因である対応する「呼」要素があるはずではないかという仮定が生まれる。たとえば、[森本 94]では価値判断「べきだ」が副詞「当然」と共起することが述べられている。

##### 3.1.2 複数の呼応規則間の相関傾向

従来の呼応関係の理論においては、個々の呼応規則の妥当性は単独で検証されるものであり、複数の規則が相互に影響しあっている可能性については考察されていない。そこで、多数の文について構造表示ファイ

ルを作成、比較することによって、複数の呼応規則間の共起傾向を発見することが必要となる。第四章で述べるように、これらの共起傾向の中には、関連する呼応関係が相乗的に予測を高めている場合が含まれているのではないかと予想される。

### 3.2 実験による発見

ある文を途中で切断し、後続部分の内容を予測させる実験を行うことによって、構造表示ファイルに現れた呼応関係が進行中の文中のどの時点で予測可能になるかを特定することができる。

#### 3.2.1 一文内比較

その際、当該文中の異なる切断点を選択することによって、一文から複数の刺激を作成し、各々の切断点での予測結果を比較することができる。こうした結果の中で特に注目すべきものは、

① 規則上の呼応関係よりも早い時点で予測に成功している場合

(1a) ライオンは 彼らは 自分達が 人間より 強いことをよく知っているので 車で 近付いても < >  
【正解】決して 逃げたりは しませんでした

(1b) ライオンは 彼らは 自分達が 人間より 強いことをよく知っているので 車で 近付いても 決して < >  
【正解】逃げたりは しませんでした

みとめ方(否定)の「ない」は「決して」と呼応関係を形成するとされているが、(1a)の段階で、14人中5人が「逃げない」を予測できている。もちろん、この予測には「ライオン(動物)」と「人間」の関係と「車で近付く」という出来事からの知識レベルでの推論が関わっていることは否定できないが、逆接の接続助詞「ても」の影響に注意しなければならない。このように、ある特定の「応」要素が対応する「呼」要素の出現以前に予測可能になっている場合、予測が早い時点で可

能になった要因は「切断点」以前の要素にもあると考えなければならない。

- ② 時点 t で一旦形成されていた予測が後の時点 t+1 で変更される場合

t=それにしても あれだけの 至近距離で 野生の 象の 親子を 見れるなんていうことは <めったにない、など>

【正解】夢にも 思っていないでした

t+1=それにしても あれだけの 至近距離で 野生の 象の 親子を 見れるなんていうことは 夢にも <思わなかった、など>

【正解】思っていないでした

ここでは時点 t で生み出されていた予測が時点 t+1 での新出要素「夢にも」の影響を受けて変化していることから、聞き手の予測にとっては呼応規則は個々に独立に機能しているとは限らないことが示唆される。

### 3.2.2 被験者間比較

ひとつの刺激に対する複数の被験者の回答を比較することによって、a) 各被験者回答の質的な偏差を明らかにできるだけでなく、b) 正答とは異なっている場合にも、もし多くの被験者がともに類似の誤った予測をしているならば、当該の切断点以前の要素の中に誤った予測を生み出す要因が特定できるはずであり、こうした要素は後続して生起する要素によって生み出される予測と影響しあうものであることが予想される。

### 3.3 漸進的分析ファイルの作成を通じての発見

予備の実験では、ある切断点における複数の被験者の回答、すなわち「応」要素の予測を比較参照できることにより、1) 呼応関係がより明確に現れる現象や、2) それらの現象が従来の狭義の呼応関係から逸脱しながらも複数の被験者によって支持されるといった興味深い結果を得ることができている。

しかし、従来の単一カテゴリ同士の固定化された呼応関係を超えるような広義の呼応関係にはどのような要素が含まれるのか、これを明らかにするためには文の要素が対応しうる可能性をすべて調べなくてはならないことになる。こうした可能性を網羅していない場合、刺激の作り方、被験者の負荷などの点において不適切な実験デザインとなる危険性がある。

したがって、ここでは、分析者が文頭から一要素ごとを認識するたびに後続する要素をどのように予測できるか、また、認識する要素が増えることによってどのように予測が変化するかの可能性を調べる作業を行った。以下、作業手順について説明する。

#### 3.3.1 漸進的分析ファイル作成手順

ここでは、455 文の作業データのうち、任意の 50 文に対して下記の作業を行った。

- ① 文の先頭から観察する要素を増やし、後続する要素が予測できた地点で、分析者がその要素を記述する。

観察要素：<予測要素>

例) つい：

つい：<うっかり忘れてました>

- ② 予測できる要素が複数の種類に渡る場合、できるだけ多種の要素を記述する。

つい：<うっかり忘れてました/食べてしまいます>

- ③ 要素が増えることによって、a) 予測が変わればそれを記述し、b) 変わらなければ直前の要素から予測できる後続要素をそのまま継承する。

a) つい昨日：<のことです/知りました>

b) ついつい：<うっかり忘れてました>

cf. \* つい昨日うっかり忘れてました

この手続きの利点は、ある文に対して分析者の目視による判断から可能な数の切断点を選択することができ、かつ切断点ごとに予測が変わる可能性をすべて記述することができる点にある。手順③で見られるように、観察要素の違いによって後続要素が変わることが明らかである。以下、分析例を示す。

羽田から 千歳空港まで 約一時間三十分ぐらいで すく  
着く < {の | ん} です / φ > [ここで文が終わってもよい]

↓

羽田から 千歳空港まで 約一時間三十分ぐらいで すく  
着く 飛行機 < を使いました / に乗りました / に搭乗しました >

「～すく着く」までを見ると、この動詞に助詞助動詞などがついて文が終わる印象を受ける。だが、「飛行機」が出現すると、この文がまだ続くことが明らかになるので、予測が変わる。

#### 3.3.2 「予備実験結果」との比較

以上述べたように、実験による発見と漸進的分析ファイルによる発見では、得られる知見が異なっている。漸進的分析作業は潜在的な可能性の発見のための作業であるのに対し、実験結果の分析は仮説すなわち刺激文に対する検証である。以下、両者の違いをまとめる。

- 1) 実験結果の分析では回答と正答との比較が主目的だが、漸進的分析作業では作業結果と正答との比較を客観的に正当化することはできない。
- 2) 実験結果の分析では切断点の選択数 (= 文からの刺激作成数) をあまり増やせないが、漸進的分析作業では原理上はすべての形態素間を切断点とすることができる。
- 3) 実験結果の分析では収集できる回答のバリエーシ

ンは各被験者が最も思いつきやすかったものだけに限られる(他の回答を思いついても書かない、など)、漸進的分析作業では作業者の努力が続く限り潜在的に可能な回答ヴァリエーションを増やしていくことができる。

このように、実験結果の分析と漸進的分析は相互補完的な作業であり、この作業を通して得られる結果は性質の異なるものであるため、明らかにしたい問題に応じて適宜使い分ける必要がある。

## 4. 解明された傾向と文法体系の問題点

### 4.1 呼応関係の拡張

広義の呼応関係を形成する文法カテゴリには、「呼」/「応」カテゴリともに、既存の規則体系以外のものもあると思われる。こうした広義の呼応関係には、格関係、提題(取り立て)、接続(従属節)などを含めることができる。

即ち 宗教の 中身を 問わず 政治と宗教が 結び付く  
と 戦争が < >

【正解】起き易い という ことです

14人中12人が「起きる」または「起こる」と予測できている。ここで動詞の予測が可能になったのは、「戦争が一起こる」の格関係による。

友人なんかは 電車に 乗ってて すりの 経験が あった  
りとか そういう 人って いますけど < >

【正解】私は たまたま そういう 経験が 今まで 生きて  
きて 何にも なかったんで

14人中12人が「私は経験がない」という内容を予測できている。提題の「(友人なんか)は」と逆接続「けど」から、「私は」も「経験がない」が予測可能になっている。

### 4.2 立体的予測(一対多、多対一)

個々の呼応関係は単独で予測を生み出すだけでなく、他の関係との相関によって予測を強化したり、理論上新規な予測を生み出したり、一旦形成された予測が修正されたりすることがある。従って、進行中の文のある時点における予測は立体的であるため、個々の呼応関係を単独で見るとはならず、これらの相互関係の影響を解明しなければならない。

#### ① 一対多

共同作業を 一緒に して 大部屋で 寝泊まりと  
決して < >

【正解】綺麗な ところでは ないんですけども

「呼」要素が「-」に対して、「応」要素が「多」の予測関係の例である。14人中12人が「決して-ない」

の呼応関係から「ない」を予測している。また、「楽な生活ではなかった」など、「奇麗」「楽」などと同様の肯定的な修飾語が頻出(14人中9人)しているが、これは「決して」から「ない」が予測可能になったことによって新たな予測が生み出され、さらに「ない」によって否定されるはずの肯定的評価語が予測されるに至っている。

#### ② 多対一

僕はですね 高校三年 の時にですね ずっと バレーボ  
ールを 高校時代 中高 バレー やってたんですけども  
それを 引退してから あまり やることが なくなっ  
てしまっ  
てですね 受験勉強 も < >

【正解】ろくに せずにですね 独り旅を するようになり  
ました

14人中11人が「全くやる気がしなかった」など否定的な内容を予測している。取り立て「も」によって対比の存在が示され、対比される対象として先行要素中の「なく」が選ばれることによって、「ない」が予測可能となっており、多対一の予測であると言える。

## 5. まとめと今後の展望

本稿では、呼応関係タグ付きコーパスの構築方針とこれを用いた発見的分析のための手法の提案を行った。この手法によって、進行中の文中における漸進的予測を可能にする文法的諸特徴の解明が可能になる。今後は、1) タグ付きコーパス及び構造表示ファイルの継続的作成、2) 実験的手法による分析、3) 分析結果に基づくテーブルの修正、の3点を循環的に行うことによって、さらに分析を進める予定である。

### 参考文献

- 木田敦子・乾裕子・神崎享子・高梨克也・井佐原均 2001. 「構文論から見た対話 - 円滑な話者交替を可能にする構文構造 -」『人工知能学会研究会資料』 vol. SIG-SLUD-A102, pp. 33-38.  
益岡隆志 1991. 『モダリティの文法』くろしお出版。  
森本順子 1994. 『話し手の主観を表す副詞について』くろしお出版。

<sup>1</sup> 科学技術振興調整費開放的有効研究費の交付を受け、国立国語研究所、通信総合研究所、東京工業大学の三者が共同で開発を進めている日本語発音音声の音声言語コーパス。