

文の理解における予測について

井 口 靖

要旨：私たちは、テキストを理解しようとする時、ある程度その内容を予測しながら読んだり、聞いたりしていると考えられる。それは文内に留まる場合もあるし、文を超えて次の文を予測する場合もある。文理解の予測に関する研究はこれまでも行われてきたが、文内の予測もテキスト全体のテーマや分野に大きく依存するという点が見過ごされてきている。予測の際、私たちはテキストのテーマや分野に応じて、心的辞書に格納された該当語彙を活性化させていると仮定できる。ここでは、文の理解における予測、コロケーションと予測の関係、テキスト全体のテーマや分野と予測メカニズムとの関連について、これまで行われてきた研究を概観し、それらを心的辞書の働きとして説明する可能性について展望を述べる。

0. はじめに

私たちが外国語を聞いて理解しようとするときには、母語の場合にはないさまざまな困難さがある。ひとつには、その速度に慣れないために音が拾えないという初歩的なこともあるだろう。たとえ音が拾えてもそれが結合した単語の意味を知らなければ理解できない。さらに、文全体の意味が理解できたとしても話し手の真意が理解できたわけではない。たとえば「今度ぜひ遊びに来てください」と言われて、本当に行っていいものかどうかは状況に依る。ここに文化の違いが出てくる可能性も高い。最後はコミュニケーションの場における総合的判断になる。このように外国語を理解するためにはいくつかの難関を越えなければならない。

以上は言語学的に見れば、音韻論、形態論、統語論、および形態論・統語論に対応する意味論、さらには、発話の場における語用論の問題となる。当然のことではあるが、ある発話を理解するということはこれらさまざまなレベルでの問題になる。ただ、母語にせよ、外国語にせよ、話されたものあるいは書かれたものの理解ということを考えると、見落とされている側面があるような気がする。

たとえ母語であっても、突然言われたこと、あるいは、初めて聞く話というのは理解するのに苦勞する。「えっ」と問い返すことも多いだろう。逆に外国語でも知っているニュースは比較的楽に理解できるというのは外国語を学んだ経験がある人には納得できることであろう。これは何を意味しているのであろうか。

私たちは文章を理解しようとする場合、ある程度その内容を予測しながら聞いたり、読んだりしているのではないか、というのがここでの仮定である。だから、予測しやすい内容については理解しやすいし、予測しにくい内容については理解が追い付いていかない場合がある。もちろん予測したにしてもその内容は次に続く語や文から検証されたり、修正されたりするのであろうが、次の語や文が現れるまで私たちは待っているわけにはいかない。母語の場合には言語内・言語外の知識から次の内容も予測しやすいが、外国語の場合には予測のための知識が不足しているということもあるのではなかろうか。

寺村（1987: 57）は外国人が苦勞するのは『あたまから』『流れに沿って』理解していく力が充分でないから』だとし、その力は「聴いた瞬間にその聴いた部分を理解するだけでなく、その後でどういう語の連なりが来るかをも瞬間に予測する能力を含んでいる」のだとしている。

児玉（2008: 89）は「われわれは、母語を習得する過程で、特定の動詞が特定の名詞と結合し、特定の名詞が特定の形容詞と結合することを知る。そのうち、母語を習熟するにつれて、語と語、文と文のつながりのパターンを知るようになり、特定の名詞から特定の動詞や形容詞が連想されたり、後続の文が予想されたりする。」と述べている。

さらに、スタッブズ（2006: 27）は「他人が言ったことや書いたことを解釈できるのは、1つには、これから生じることを前もって予測できるからである。人間のコミュニケーション能力には、起こりうること・しばしば起こること・典型的に起こることについての（たいていは無意識的な）知識が含まれる。」としている。

これらは言語学者が自らの経験的に則して述べているものと思われるが、脳科学者である酒井（2011: 96）も「人間の場合は、先読みの能力をうまく言語にも流用しているので、一つの言葉を眼や耳にただけで、それに付随する意味や次に来るべき言葉を予測できる。それは、ほとんど本能のようなものだ。」としている。

このように理解における予測というのはだれしもが感じていることであろう。ここでは文を理解するための予測ということを整理・考察し、今後の研究のあり方について検討することにする。

1. 文内の予測と文を超えた予測

文内の予測においては日本語における先駆的な研究がある。寺村（1987）は「その先生は私に国へ帰ったら父の生きているうちに早く財産を分けて貰えと勧める人であった」という夏目漱石の「こころ」の文を「その先生は」「私に」「国へ」のように頭から区切って提示し、被験者にその先を予測させている。「その先生は」だけでは予測文は非常に広範囲であるが、次に「私に」を聞いただけでかなり絞られ、先に進むうちに一致が進み、「ネイティブスピーカーというものは考えようによっては驚くほどの正確さで先を予測するものだ」としている。寺村は表題からもわかるように、主に文法的な観点からこの結果を考察している（詳しくは下記参照）。

寺村の研究を受けて、その後さまざまな調査・研究が行われている。内田ほか（1995）は文頭の名詞句に「が」がつくか「は」がつくかで、「が」は動作性述語を、「は」は状態性述語を予測させる傾向にあるとし、時制も「が」が過去時制、「は」が非過去時制を予測させるとする。市川（1993）、津留崎ほか（1997）はこのような予測調査を日本語学習者にまで広げている。また、菊池ほか（1996）は「友人に聞くと、国際感覚というのは」を提示して、母語者と学習者が「という」「らしい」「そうだ」「ようだ」などのモダリティ表現をどの程度予測できるかを調査している。このように主として予測の研究は日本語教育の中で進められてきた。

以上は、その文の内部における予測を中心に扱っているが、酒井・山内（2004）は「なかなか」「あげく」「確かに」に続く予測をとりあげ、「なかなか」は単文の中だけで説明でき、「あげく」は複文レベルまで及び、「確かに」は次の文まで影響を及ぼすとしている。松浦（1996）は母語者と学習者においてニュース文の「責任をとって」など「テ形接続」に続く文の内容を点数化して予測にどのような要素がかかわっているかを考察している。堀口（1989）は話し手

の発話に対する聞き手の予測を聞き手の反応から分析している。これも文を超えた実際のコミュニケーションの場における予測研究であるが、文を超えた予測、つまり次にどのような文が予測されるのかということに関するまとまった研究には石黒（2008）がある。石黒（2008: 63）は予測を「当該文を読んで感じられる情報の不全感を後続文脈で解消しようとする理解主体の意識の動き」とし、後続文の予測を「関係連続の予測」「接続関係の予測」「具体的内容の予測」に分類している¹。

母語でも外国語でも、その文の内容を理解し、ある程度次に来る文の内容を予測できるというのはきわめて重要なことであり、これができてこそ実際にコミュニケーションが成り立つと思われる。その意味で文を超えた理解の研究は重要である。しかし、ここでは基本的に文内にとどまってみたい。それは、文内の予測自体まだ大部分が解明されていないということがあり、また、調査方法としても形式面からより客観的にとらえやすく、特に語のレベルからもアプローチしやすいのではないかと考えるからである。

ただし、文の理解も実はそれが使われる場、つまり、その文が実際に使われているテキストのテーマや分野などにかなり依存するのではないかと考えられる。そもそもその文に現れる語彙自体がテキストの種類によって異なってくる。聞き手や読み手は特別な場合を除いてあらかじめそのテキストのテーマはある程度知っているはずであり、それがなければ本来予測は成り立たないのではないだろうか²。ここでは、文と文のつながりに関する予測は取り扱わないが、テキストのテーマや分野が文の予測に及ぼす影響についてはのちほど考察する。

2. 文法予測と内容予測

寺村（1987）はまず「その先生は」という語句を提示し、被験者にその先を予測させる。次に「その先生は私に」を提示し、同様に予測させる。ここで興味深いのは、「その先生に」まででは「名詞+ダ」、「形容動詞+ダ」、形容詞、動詞が現れ、しかも動詞では現在形が多いのに対し、「私に」が提示されると動詞のみ、しかも過去形がより予測されたと言う。これを寺村は「私に」が現れることにより、「その先生が」が「ガ格」つまり主格の読みを受け、この二つの格から動詞が連想され、それから「個別・具体的動的事象」が浮かび上がり、過去形になったと推測している。

寺村は文法的観点から予測をとらえており、「ガ格」と「ニ格」が現れることによって必然的に述語も予測されるということであるが、実際はそれだけにとどまらず、それが文の内容の予測にも影響を及ぼしていると思われる。

特定の語が内容の予測に影響を及ぼすということについては、たとえば、酒井・山内（2004: 2 ff.）の「なかなか」に続く予測に示されている³。

- (1 a) 彼がなかなか ()
 (1 b) 彼はなかなか ()

で予測をさせると、(1 a) では動詞の否定形（「来ない」など）が多く、(1 b) では形容詞（「優しい」など）が多いという。これは寺村（1987）や内田ほか（1995）の研究と一致する。しかし、

(2 a) バスがなかなか ()

(2 b) バスはなかなか ()

では両方とも動詞の否定形がほとんどだという。これについては単なる「バス」では特定のバスが思い浮かべられないため、それについての状態を述べるのが難しいのだろうとしている。しかし、たとえば次の例ではどうだろうか。

(2 c) 再就職はなかなか ()

(2 d) ドイツ語はなかなか ()

「できない」や「話せない」もあるだろうが、「難しい」「おもしろい」なども予測されるだろう。バスについてそもそもその性質や状態を述べる状況を思い浮かべるのが難しいということではないだろうか⁴。つまり、文法的なことがらだけでは片付かない問題ということである。

また、酒井・山内（2004: 5 ff.）では、「あげく」の分析の際に、「ようやく」「やっと」が現れると必ず望ましい内容が現れるとしているが、これは「あげく」の問題ではなく、「ようやく」「やっと」が望ましい内容を予測させるのだと思われる⁵。

(3 a) 夜遅くまで考え続けたあげく、やっと ()

(3 b) さんざん苦労したあげく、ようやく ()

「やっと」「ようやく」がなければ「答えはでなかった」とか「失敗した」などの望ましくない内容が予測されるが、「やっと」「ようやく」があれば「思いついた」「完成した」などが予測されるとする。

これはドイツ語の *erst* にも通じるものがある。

(4 a) *Es ist erst zehn Uhr.* (やっと 10 時だ)

(4 b) *Er ist erst Dozent.* (彼はやっと講師になった)

(4 c) *Der Kampf hat erst gerade begonnen.* (戦いはようやく今始まったばかりだ)

井口（2000: 74）は（4 a）（4 b）のような状態表現での *erst* の機能を「話し手は、文で表現された状態が終わって次の段階になっていると想定していたことを表す」としている。つまり、（4 a）では現時点が 10 時より後だと想定していたが、実際は 10 時であったことになり、（4 b）では Professor になっているかと思っていたが、実際は Dozent だったということになる。また、（4 c）のようなできごと表現においては *erst* は「話し手は、基準点において、文で表現されたできごとが起こった後の状態だと想定していたことを表す」（井口 2000: 77）としている。そうすると、（4 c）はすでに戦いの真っ最中だと思っていたことを示すことになる。*erst* 自体は直接的に望ましい内容、望ましくない内容を示すものではないが、状態にせよ、できごとにせよ、話し手の想定がさらに先にあったことになり、そのため待ち望んでいたことと受け取られることが多いのであろう。

このように特定の語が後の内容を予測させるということは大いにあることである。特に副詞

や接続詞は文内だけではなく、文を超えた内容の予測をさせることがある。たとえば、石黒(2009: 167)は「たしかに」「もちろん」「ほんとうに」という副詞が逆説を導いているかどうかをコーパスを用いて調査している。「たしかに」の逆説誘導率は70%であるのに対し、「ほんとうに」は0.9%であったとしている。おそらくは母語者の場合にはこれを経験的に知っており、予測に役立っていると思われる。

以上は、機能語、もしくはそれに類するもの⁶が、ある程度文内の次の内容、もしくはその文に続く内容を予測させるものであることを示しており、語彙の機能的な意味に関係するものであろう。一方で、寺村(1987: 63)は、「その先生は私に」で動詞が「言ウ類とクレル類に絞られているだけではなく、『～せよ／～するように』言うという形をとることが、かなり高い率で予測されている」としている。これについて寺村は「先生」と「私」の関係のあり方が予想されるのだろうが、よくわからないとしている。

寺村は文法的知識を主に考察しているが、そこに現れる語彙の意味、その現実世界での含意が文の予測に影響していることを示唆している。さらに「その先生は私に国に帰ったら」まで提示したところ「～ように言った」と結びつくという予測がほとんどで、しかもその内容は一定の範囲にあるとしている。それを寺村(1987: 66)は「日本人なら誰でもこの文脈で思いつきそうな行動だ。予測内容が言語外的知識、大げさにいえば文化に関わる部分も含むということの一例といえるだろう」としている。これについては「先生は」「私に」「国へ」という関係から想定できる場面はある程度限られ、それに沿って予測が行われているためだと言えるだろう。そして、その場面を思い浮かべることがこの文の予測であり、それがこの文の理解に決定的な役割を果たすものと思われる。

たとえば主体が「(その)学生は」「妻は」「(その)店員は」であればまったく異なる場面が想定され、まったく異なる動詞が予測されるかもしれない。そのように考えると、文法的な予測とともに語が持つ意味がもたらす予測というのも文理解に大きな役割を果たすものと考えられる。それは言いかえると一種の「コロケーション」と見ることができる。

次に予測という観点からコロケーションを整理しておく。

3. コロケーション

コロケーションは人によってさまざまな定義がある。

Mel'cuk (2009: 33 ff.) は、慣用連語素 (phraseme) を「語用論的慣用連語素」と「意味論的慣用連語素」に分け、後者をさらに「イディオム」「コロケーション」「擬似イディオム」に分ける。「慣用連語素」というのは「シニフィエとシニフィアンを制約なしにかつ文法規則に則って組み立てることができないような句」で、そのうち「語用論的慣用連語素」というのは特定の場面で特定の言い方しかしないというような場合(例: Caesar Salad: All you can eat シーザーサラダ、食べ放題)、決まり文句などである。もうひとつの「意味論的慣用連語素」は、AとBからできている表現で、その意味にAの意味もBの意味も含まない場合が「イディオム」(例: shoot the breeze 無駄話をする)、AとBにさらに別の意味が加わる場合が「擬似イディオム」(例: start a family 子供を産む)であり、これに加えて、「コロケーション」がある。これは、たとえば、land a JOB (職を見つける)、strong COFFEE (濃いコーヒー) などであるが、2つの構成素のうちの1つにはもとの意味 (JOB, COFFEE) が残っており、

JOB, COFFEE に「依存した形で選択されている」land, strong は別の意味を持っているとする。ここでは場合によってはイディオムともされるかなり限られた共起を「コロケーション」としている。

スタッブズ (2006: 32) はコロケーションを「連続するテキスト内において、互いに数語の範囲 (span) 内に共起する傾向を持つ 2 語ないしそれ以上の語彙的關係をいう。たとえば、PROVIDE (提供する) は help (援助) や assistance (助力), money (金銭), food (食料), shelter (保護), information (情報) など、人間が必要とする価値のあるものを指す語としばしば共起する。これらの語は、当該動詞の頻度の高い共起語 (collocate) の一例である。」としている。これが通常のコロケーションの理解であろう。

堀 (2009: 7) は「コロケーションとは、語と語の間における、語彙、意味、文法等に関する習慣的な共起関係を言う」とし、「語彙的コロケーション」「意味的コロケーション」「文法的コロケーション」に区分している。「語彙的コロケーション」とは、スタッブズの言うような動詞+名詞、形容詞+名詞、副詞+形容詞などに見られる「語と語の相性の問題」だとしている。「意味的コロケーション」は「語とある特定の意味領域との相性の問題」で、たとえば、It provided them with a bad environment. が不自然なのは provide が通常好ましいものについて使うからであるとしている。「文法的コロケーション」は step into, calm down, look into のような「前置詞や副詞といった特定の文法機能と習慣的に共起する関係をもつ動詞のコロケーション」だとしている。

井口 (2011: 117) は、ドイツ語研究所 IDS コーパスで調査したところ、unerträglich (耐えられないほど) + heiß (暑い) の結びつきが 209 例あったのに対し、unerträglich+kalt (寒い) は 6 例しかなかったことから、außergewöhnlich, außerordentlich, besonders, denkbar, höchst, recht, sehr, so, überaus, ungemain, ungewöhnlich, ungleich, weit など程度が高いことを表わす程度副詞がどのような形容詞や副詞を修飾しうるかを IDS のコーパスを用いて分析している。調査の結果、程度副詞によって修飾する形容詞や副詞が異なり、かつ、必ずしも通常頻度が高いとされる形容詞や副詞を修飾するとは限らず、程度副詞とそれが修飾する語との間には何らかの関係があることが判明したとしている。これは堀 (2009) の語彙的コロケーションに当たるものであろう。

また Quasthoff (2011: 95) ではたとえば Bier (ビール) には次のような共起語が挙げられている。これも語彙的コロケーションと言ってよいだろう⁷。

動詞と

1 格で： schäumen ・ zischen

4 格で： brauen ・ herstellen ・ produzieren ■ anbieten ・ verkaufen ■ besorgen ・ einkaufen ・ holen ・ kaufen ・ konsumieren ・ mitbringen ■ ausgeben ・ bestellen ・ verlangen ■ kaltstellen ・ kühlen ■ ausschenken ・ zapfen ■ einschenken ・ servieren ■ genießen ・ trinken ■ verschütten ■ einlagern ・ lagern

3 格で mit B. anstoßen ・ zuprosten

形容詞と

groß ・ klein ■ eiskalt ・ kalt ・ kühl ■ lauwarm ・ warm ■ frisch ・ gepflegt ・ zünftig ■ herb ・ leicht ・ milde ・ stark ・ süffig ■ gepanscht ・ sauer ・ schal ・ schlecht ■

ausländisch • importiert ■ einheimisch • heimisch • lokal • regional ■ alkoholfrei • naturtrüb • obergärig

ただ、予測ということ考えた場合、このコロケーションが常に有効であるとは言い切れない。もし Bier という語が最初に来たならば schäumen (泡立つ), zischen (しゅっと音を立てる) のような動詞を予測することができるかもしれないし、Bier が目的語でも先にあったなら、brauen (醸造する) などの動詞を予測することができるかもしれない。ドイツ語の場合、動詞と主語、目的語の順序が定まっているわけではない。動詞が主語や目的語の前に来ることもあるし、特に完了形、受動態、語法の助動詞の文では動詞が文末に出てくる。ここで挙げた例では双方向的にある程度の予測は可能に思える。

Bier→schäumen, brauen

- (5 a) Das Bier schäumt zu stark. (このビールは泡立ち過ぎる)
 (5 b) Wie wird Bier gebraut? (ビールはどうやって醸造されるのか)

schäumen, brauen→Bier

- (6 b) Warum schäumt das Bier? (なぜビールは泡立つのか)
 (6 a) Wir brauen Bier. (私たちはビールを醸造している)

これに対して、動詞が前に来た場合、brauen はまだよいとしても、herstellen (生産する), verkaufen (売る), holen (取ってくる), bestellen (注文する) から Bier を予測することは可能だろうか。これら動詞は必ずしも Bier と特別な関係があるわけではない。

Quasthoff (2011: IX) はコロケーションは unsymmetrische Struktur (非対称的構造) をしているとしている。上の例では Bier からは herstellen, verkaufen などに対して、ある程度強いコロケーションのための力が生じると言えるが、herstellen, verkaufen から Bier に対しては非常に弱い。つまり、コロケーションには方向性があるということになる。これは「予測の力」にも置き換えることができよう。

しかし、そのような場合であっても、たとえばレストランや酒場であったなら、holen, bestellen からかなり高い確率で予測される候補の中に Bier が含まれることであろう。つまり、コロケーションから見た予測にも場面が大きな役割を果たしていると考えられる⁸。

井口 (2011) の程度副詞+形容詞の共起の中には語彙的というよりも「意味的なコロケーション」ではないかと思わせる例もある。たとえば denkbar はネガティブな意味合いを持つ形容詞や副詞と共起しやすく、10%以上の確率で knapp (乏しい)、5%以上の確率で schlecht (悪い) が予測できる。

また、井口 (2012 a) では、程度が高いことを表わす程度副詞 bedeutend, beträchtlich, ungleich, weit, weitaus, wesentlich は形容詞、副詞の比較級と結びつくことから、どのような形容詞や副詞と結びつくかを調査しているが、比較級に関しては原級ほどは程度副詞との結びつきの偏りは見られなかったとしている。比較級を修飾するということは、それぞれの形容詞や副詞に含まれる意味の程度の高さを述べるのではなく、比較した場合の差の程度をさまざまに修飾しているのではないかと推察される。これらが「比較」という意味自体と共起することなら、堀 (2009) の「意味的なコロケーション」に含まれることになる。

恒川（2011）は *beginnen*（始める）が *mit+3* 格名詞と使われる場合と 4 格名詞と使われる場合の名詞の内容を比較して、「行為の明示性（行為性）の高い名詞」は *mit+3* 格名詞と結びつく傾向が見られ、「行為の明示性（行為性）の低い名詞」は 4 格名詞と結びつく傾向があるとしている。また、恒川（2011）、恒川（2012 a）は、これらはそれぞれ特定の副詞と結びつく傾向があるとしている。そうすると、文法形態や特定の語が文の内容を予測させている可能性があることになる。

以上のように、予測の観点からコロケーションを利用するとすると、特に「意味的コロケーション」が問題になってくると思われる。ただ、「語彙的コロケーション」はある程度機械的にコーパスから導き出すことが可能であるが、「意味的コロケーション」はコーパスから抜き出した例についてひとつひとつ内容の検討を行うことになる。それはある意味では予測の研究にとって避けることのできない宿命ではあるが、予測する内容をどのように分類、整理するのかという問題が大きい。

上で Bier の予測に関して場面の働きが大きいとしたが、次に、テキストのテーマや分野が文の予測に影響を及ぼすという観点からアプローチを試みる。

4. 心的辞書

スタップズ（2006: 8）は「場合によっては、ただ 1 つの語や句によって、テキスト・タイプがじゅうぶん特定できることもある」としている。ということは、逆にテキスト・タイプがわかっているならば、そこに生じる語や句もまた予測できるということになる。実際、スタップズ（2006: 28）は「異なったテキスト・タイプは、異なった予測のパターンを持つ。たとえば、大半の恋人募集広告では、限定的な動詞しか用いられず、そこに典型的な主語と目的語が付随する。意味のパターンは非常に単純であって、語彙や文法の大部分も予測可能である。ただし、語彙的多様性はかなり生じる余地がある。」とも述べている。

ここで語彙の大部分が予測可能としながらも「語彙的多様性」があるというのは矛盾しているようにも見える。恒川（2008）はドイツ語の語彙を「深さ（頻度）」と「広さ（当該語彙が他の素材テキストにも共通して出現するか）」の観点から調査・検討しているが、スタップズの「テキスト・タイプ」に関係する語彙というのは「狭い語彙」ということになる。そこに「語彙的多様性」があるとすると、「浅く狭い語彙」、つまり特定の分野で使われ、かつ頻度も低いもの、ということになるであろう。しかし、予測という観点からは「浅く狭い語彙」は予測しやすいとは言えない⁹。母語者ならば、頻度はさほど高くないがその分野の話なら必ず出てくる語彙をいうものを備えているのかもしれないが、外国語の学習においてはそのような語彙の習得は最も後の段階になる。それに対して、特定のテキスト・タイプにおける「深く狭い語彙」はある程度確定できる可能性もあり、それは予測に大いに役立つにちがいない。それはそのテーマや分野の話になれば必ず出てくる語彙を収集してリスト化するということになる。

これは聞き手の理解という観点からは、あらかじめテキストの内容から出現しそうな語彙を頭の中で活性化し、備えておくということである。実証されたことではないが、ごくふつうにありそうなことである¹⁰。本を読むときには書籍名、あるいは、その章の題目が頭にはいつているであろうし、新聞を読むときにはまず見出しから入る。その時点ですでに予測が働いている。講演を聞きに行くときにも演題からその内容を予測するであろうし、場合によっては、あ

の演者ならこのような話だろうと予測することさえある。これはその内容を予測することであるし、それは特定の語彙を予測することでもあるだろう。

語彙というのは私たちの頭の中に蓄えられているが、それは紙の辞書のようにアルファベット順に並んでいるのではないだろう。由本 (2011: 5) は「心的辞書とは、ただ個々に単語が羅列されてその意味が記載されているものではなく、さまざまなカテゴリーや連想のネットワークのもとに、膨大な情報が整理されているものであり、しかもそれらが、適切にしかも瞬時に利用できるように体系化されて蓄積されていると考えられる。」と述べている。

私たちが文を理解するときに、特定の語彙を活性化させているというのは現在のところ実証するのは難しい。しかし、それを推測させる現象はある。寺尾 (2002) は、言い間違いがどうして起こるのかを分析しているが、文脈的誤りと非文脈的誤りがあるとし、文脈的誤りは「『統合関係』、つまり前の語から後の語への順序関係」から発するものであり、非文脈的誤りは「『連合関係』、つまり実際に現れた要素とその位置に現れることが可能だったにもかかわらず現れなかった『隠れた』要素との関係」から発するものとしている。この「隠れた要素」というのが話し手の頭の中で活性化された要素と見ることができるであろう。活性化されているからこそその本来なら「隠れた要素」が表面に出てきて、言い間違いとなるのである¹¹。予測ということは話し手の頭の中を予測するということである。そうするならば、聞き手においてもできる限り話し手と同じ要素を活性化しておこうとする働きがあるに違いない。

それではあるテーマにおいて活性化される語彙はどのように調査できるだろうか。直接活性化されているものを調べることは難しいが、実際に出力されたものは調べることができる。特定の分野における語彙を調査することによってある程度その分野において現れやすい語彙というのは特定できるのではないだろうか。

恒川 (2010) は動物保護の分野でのキーワードを抽出する試みを行っている。黒田 (2012) は自動車試乗記の分野で 16 万語のコーパスを作成し、その頻度を一般のコーパスの頻度と比較している。その結果、特定の分野で頻度が高いものはほとんど名詞と形容詞で、分野別コーパスでは非常に頻度が低いものや特殊な意味を持ったものがしばしば上位にあがる (たとえば Diesel (燃料としてのディーゼル・ディーゼルエンジン・ディーゼル車)、Verbrauch (燃費; 通常は「消費」)、Kofferraum (トランク)、Lenkung (ステアリング)、serienmäßig (標準装備に含まれた)、Höchstgeschwindigkeit (最高速度) など) などの指摘をしている。

また、井口 (2012 b) では、ドイツの週刊誌 der Spiegel が Fukushima というテーマでまとめた記事をコーパスとし (Token 約 7.5 万語、Type 約 8 千語)、頻度分析を行なっている¹²。上位は主として冠詞、前置詞、接続詞であるが、11 位にはすでに Fukushima が挙がり、22 位 Tepco (東京電力)、26 位 Japan, 28 位 Reaktor (原子炉)、31 位 japanisch, 35 位 radioaktiv (放射能の)、43 位 Tsunami, 44 位 AKW (=Atomkraftwerk 原子力発電所)、46 位 Katastrophe (大災害)、55 位 Erdbeben (地震)、58 位 Tokio, 60 位 Atomkraftwerk, 82 位 havariert (破損した)、84 位 Daiichi, 87 位 Radioaktivität (放射能)、89 位 Strahlung (放射線)、106 位 Betreiber (事業主)、111 位 Atomkatastrophe (原子力災害)、114 位 Cäsium (セシウム)、127 位 Kernschmelze (メルトダウン)、157 位 Millisievert (ミリシーベルト)、159 位 Brennstab (核燃料棒)、183 位 Sperrzone (立入禁止区域)、184 位 Tschernobyl (チェルノブイリ) などが特徴的であるとしている。これらを含む 200 位までで文章全体の語数 (Token) の約 60% をカバーしており、90% をカバーするには約 2300 語の語彙が必要であるという¹³。単純に計算しただけで半分以上の

内容を理解するために上記のような「特有の」語彙を知っている必要があるということは言えるだろう。

上では任意に目立つ語彙を挙げたにすぎず、何をこの分野特有の語彙とするかは一般のコーパスと綿密に比較していかなければならない。ドイツ語研究所の DeReWo は頻度をランクで示しているために、単純に比較はできないが、Fukushima, Tepco, Tschernobyl などの固有名詞は当然含まれていないが、Tsunami, AKW, havariert, Radioaktivität, Atomkatastrophe, Cäsium, Millisievert, Brennstab, Sperrzone¹⁴などは1万語レベルを超えるか、そもそも25万語のリストにない。

これらは Fukushima に特有の語彙であるが、一般に使われる語であっても、ここではふつうの頻度よりもより多く使われる語というものも見受けられる。たとえば、41位 Mensch [104]¹⁵ (人), 42位 Regierung [416] (政府), 52位 März [1730] (3月), 79位 Wasser [299] (水), 73位 Arbeiter [2396] (作業員) などである。また、91位 Tag [108] (日), 97位 Stunde [264] (時間), 132位 Woche [209] (週) など基本語であるが、頻度がある程度高めになっているのは時間的緊迫が問題になっているためであろう。これらもここでは活性化されていなければならないものと言える。

これだけを見ても予測においてそのテキストのテーマや分野における特定の語彙の活性化が必要であることが明確である。

5. 終わりに

この小論では最初に文内において特定の文法事項や語彙が文の内容を予測させる可能性について検討した。そして、次に特定の分野の理解においてはその分野における語彙が活性化されており、それが理解に役立っているのではないかとの仮説を立てた。現れる語彙がある程度特定され、表現される内容が特定されるということは予測が非常にしやすくなる。これは文内のコロケーションにも影響を及ぼす可能性がある。

これまでも特に外国語学習の観点から「花」「鳥」などの分類や「買い物」「旅行」などの場面別で語彙集が作られてきた¹⁶。それらは今後コーパスで検証される必要があるが¹⁷、現実生活では利用価値の高いものであろう。しかし、そこには予測研究という観点からは根本的な問題が存在する。それは、そもそもどのようなテーマや分野が立てられるべきかということである。テーマや分野の立て方は無数であり、また仮に立てたとしてもその境界が明確に存在するわけでもない。

しかし、心的辞書において語彙がきちんとテーマ・分野別に分類されているとは到底思われない。もともとテーマや分野というのは人為的なものであるということもあるが、たとえテーマや分野があったとしても同じ語がいろいろなテーマや分野に関連付けられているはずである。さらには、そのような語同志が複雑な連想関係にあるということも想定できる。

心的辞書を完全に再現することは困難だとしても、将来、可能な辞書記述の可能性としては、ある語はその意味だけではなく、その語と syntagmatic に関係する語（語彙的・意味的コロケーション）、paradigmatic に関係する語（隠れた要素）が記述されるだけではなく、テーマや分野に関連付けられていなければならない。それは言語の知識を超えるものであり、さまざまなテーマに多元的に関連付けられたものであろう。しかもテーマは現実世界の動きにつれて変動

する。たとえば「原発」は「エネルギー」や「環境」だけではなく、今や「地震」や「津波」にも関連する。そうなるともはや紙の辞書では対処できず、必然的に電子的なものとなる。

それがもしも実現したならば、特定の分野の話を聞いたり、読んだりしようとするときにはあらかじめある一定程度の語彙を抽出できるし、それは話したり、書いたりする場合にも同様である。場合によっては、辞書を引く場合にも予測される可能性の高い語や意味を順序付けて提示することもできるかもしれない。

それが一部実現していると思われるところがある。携帯電話などで行われている漢字変換はそれまでの頻度から予測した変換候補を提示している。あるいは、コンピュータ上での漢字変換でもその前後の語句からある程度予測して漢字候補を提示する。たとえばこれにテーマや分野が考慮されるようになるとさらに変換効率が上がるのではないだろうか。

さらに、心的辞書の解明は教材開発にも寄与することであろう。読むテキストとしてはそのテーマで予測される語彙が適度に含まれるテキストが自然なテキストだということになるし、また、会話教材としてはあらかじめ予測される語彙をあげておくことにより、スムーズな会話が成立する。成田（2005）はドイツ語の質問にドイツ語で答えさせるためのテキストで次のような手法をとっている。学生は、まずテキストを自分で読み、自分でその内容についての質問に答える。授業では、教師がドイツ語で質問するが、その語彙はすでにテキストや自習用の質問に与えられており、テキストにも日本語付きでリスト化されている。そこには一部追加の語彙もある。つまり、あらかじめそのテキストのテーマで現れそうな語彙を提示することによって、教室でのドイツ語の質疑が可能となっているのである。

これらの意味で予測という研究分野は今後心的辞書という脳の働きの解明に関わるだけでなく、電子辞書や教材の今後のあり方にも関わる。脳内は直接に見ることができないため、実際に現れた言語資料からたどるしかないが、大規模コーパスの分析という方法論がその研究を可能にしてくれるものと思われる。

註

これは井口を代表とする平成 21～24 年度科学研究費補助金基盤研究 (C)「ドイツ語テキスト及び文における語彙出現予測分析とその和独辞典・教材への応用」(課題番号: 21520436; 分担者: 恒川元行九州大学教授、成田克史名古屋大学教授、黒田廉富山大学准教授)の成果と今後の展望をまとめたものである。

- 1 石黒（2008）は予測関係の文献情報も豊富でここでもそれを参考にした。
- 2 そのテキストのテーマや分野も知らずに聞き始めたり、読み始めたりすることがないとは言えないが、その場合でも最初のいくつかの文からまず何を取り扱ったものかを知ろうとするのではないだろうか。
- 3 酒井ほか（2004: 2f）は「なかなか」の意味を日本語教育誤用例研究会『類似表現の使い分けと指導法』（1997）に従って、「もどかしさを表わす用法」「手強さを評価する用法」「良さを評価する用法」としている。動詞の否定形と結びつく場合「もどかしさ」、形容詞と結びつく場合「手強さ」「良さ」を表わす。
- 4 これはインフォーマントテストの限界と同種のことを示していると思われる。特に意味的整合性を調べるテストにおいてはインフォーマントが状況を思い浮かべることができるかどうかで容認可能となったり不可能となったりするのはよく経験することである。
- 5 中村（2010）は「やっと」は「かろうじて、長い時間待っていたの意」で、「ようやく」は「いろいろあったその後でといった意味合い」とし、「実現時に意識のある『やっと』」に比べ、この語は実現に至る過程に焦点を当てた表現とされる」としている。

- 6 通常副詞は機能語に入れられないが、たとえば「ようやく」と「今日」「ゆっくりと」などでは具体的意味内容（命題内容）をもつかどうかという点において異なると思われる。井口（2000）ではドイツ語の副詞を命題内機能と命題外機能に分けて記述している。ここでは命題外機能を持つものを「機能語に類するもの」と呼んでおく。
- 7 Quasthoff（2011: X）は Leipzig 大学 Herder/BYU コーパスに基づき、名詞 2346、動詞 617、形容詞 290 の見出し語で約 19 万通のコロケーションを収録している。語の間の・と■の区切りは原典のまま。
なお、日本語については近年、姫野（2004）、金田一（2006）、小内（2010）などのコロケーション辞典が出ている。英語については、コーパスに基づく 2500 語の見出し語と 2 万弱のコロケーションを収録したとする塚本（2012）がある。
- 8 ドイツ語において動詞が文末に来るような場合、聞き手は主語、目的語、状況語などを記憶にとどめておき、動詞が聞こえてからそれらを一括して処理するのか、それとも、ある程度の要素が聞こえたときに動詞を予測しているのかということも興味深い問題である。寺村（1987）の実験はこの後者を想定したものとと言えるが、ただ、寺村の場合、予測するように指示したために反応があったということにすぎず、必ずしも常にそのような予測をしながら聞いているという裏付けにはならない。なお、寺尾（2002: 181）は「成人の記憶や注意が一度に処理できるまとまりの数は、七を基準に、少なくとも五、多くて九の範囲に入るという G・ミラーの古典的仮説」を紹介しているが、人間の脳がどの程度まで保留にしながら一括処理できるのかということに関わってくるのかもしれない。
- 9 樺島（1981: 109）は「延べ語数の九割を占める見出し語を習得していても、読もうとする個々の文章の内容がどれほど理解できるかということになれば疑問である。どんな文章にも共通して使われる語彙は、どんな文章にも共通する内容は理解させるだろう。しかし、今読んだり聞いたりしようとしている文章の、独自の内容を理解させる力は弱いだろう。」としている。そして「意味内容の伝達において重要な語は（中略）使用範囲が限られていて、その文章の内容に密接な関係がある、特殊な語」（樺島 1981: 110）で、「その意味の範囲が狭く、はっきりしている語」（樺島 1981: 111）という。これは「浅くて狭い語」ということになる。そうすると予測との関係から言うと「その位置にどのような語が来るかを、文脈から予測しにくいほど、文や文章の意味内容を構成する上で重要な語である」（樺島 1981: 119）ということになる。
確かに情報量という点から考えると、予測できるということは情報量が比較的少なく、予測できないということは情報量が比較的多い、ということであるから、予測できない語が内容面で重要だと言うことはできよう。樺島（1981）の指摘は予測と文章理解という点から示唆に富むものであり、今後の検討課題である。なお、ドイツ語におけるこの問題に関しては恒川（2012 b）を参照。
- 10 脳科学者酒井（2011: 95）は友人の葬儀の後、いつも歩く道ではじめて葬儀社の看板を見つけた経験について「これは、今気にかかっている問題のリストを脳があらかじめ用意していて、それに該当する情報が入ってきたときには、瞬時にそれを意識に上らせているかのようである。」としている。
- 11 寺尾（2002: 65）は、ある女性が汗をかいたので家に帰って着替えをしようと言ったときに下着については触れなかったが、nach Hause（家へ）を nach Hose（下着へ）と言い間違った「フロイトの言い間違い」の例をあげている。Hose が活性化されていたために起こったと言えるのである。
- 12 TreeTagger によりレマ化し、AntConc を用いて頻度表を作成したもの。
- 13 Tschirner（2008: 3）は、あるテキストを理解するためには 95-97% の語彙を知っていなければならない、2000 語では 90%、4000 語で 95% をカバーするとしており、基本語彙として 4000 語を抽出している。ただし、註 9 で述べたように、残りの 10% あるいは 5% に情報量の高いその分野特有の語が含まれている可能性があり、それなしにテキストを本当に理解できるかどうかという問題がある。
- 14 ここで複合語をどのようにとらえるかという問題が出てくる。複合語がその構成語の意味の総和だけでできている場合には、構成語の意味を知っていれば複合語の意味も導き出せるが、そこになんらかの別の意味が加わる場合も少なくない。これも母語者としての予測能力に関わるのかもしれない問題であるが、ここでは考慮しない。
- 15 [] は Tschirner（2008）による順位。

- 16 たとえばドイツ語では Vig (2002)、荻野 (2003) など。特定分野のみの辞書としては鈴木ほか (2012) がある。
- 17 Tschirner (2008) はコーパスを用いた頻度調査に基づいて基本語彙約 4000 語を選びだし、テーマ別に配列している。

参考文献

- 石黒圭 (2008) 『日本語の文章理解過程における予測の型と機能』ひつじ書房
- 石黒圭 (2009) 「文章理解における予測研究方法と可能性」『第二言語としての日本語の習得研究』12号、159-173
- 市川保子 (1993) 「外国人日本語学習者の予測の能力と文法的知識」『筑波大学留学生センター日本語教育論集』8号、1-18
- 井口靖 (2000) 『ドイツ語文法シリーズ5 副詞』大学書林
- 井口靖 (2011) 「ドイツ語程度副詞の共起語の頻度に関する調査 (1)」三重大学人文学部文化学科『人文論叢』28号、117-129
- 井口靖 (2012 a) 「ドイツ語程度副詞の共起語の頻度に関する調査 (2)」三重大学人文学部文化学科『人文論叢』29号、163-174
- 井口靖 (2012 b) 「Fukushima コーパスにおける特定語彙調査」第7回予測研究会口頭発表
- 内田安伊子・池上摩希子・大野早苗・大島弥生・長友和彦 (1995) 「予測文法研究 (1) : 『が』と『は』の予測機能について」お茶の水女子大学言語文化学会『言語文化と日本語教育』9号、134-159
- 小内一 (2010) 『てにをは辞典』三省堂
- 荻野蔵平 (2003) 『日本語から引く知っておきたいドイツ語』小学館
- 樺島忠夫 (1981) 『日本語はどう変わるか 一語彙と文字一』(岩波新書) 岩波書店
- 菊地民子・天野千春・猪狩美保・嶽肩志江・大島弥生・長友和彦 (1996) 「予測文法研究 : モダリティ表現の予測能力とその習得について」お茶の水女子大学言語文化学会『言語文化と日本語教育』11号、12-25
- 金田一秀穂監修／学研編集部編集 (2006) 『知っておきたい日本語コロケーション辞典』学習研究社
- 黒田廉 (2012) 「特定分野のテキストの出現する語彙についての試行的調査 一自動車試乗記を例として一」『富山大学人文学部紀要』第57号、139-148
- 児玉徳美 (2008) 『ことばと論理 一このままでいいのか言語分析』開拓社
- 酒井邦嘉 (2011) 『脳を創る読書 一なぜ「紙の本」が人にとって必要なのか』実業之日本社
- 酒井智香子・山内博之 (2004) 「『なかなか』『あげく』『確かに』の用法に関する予測文法的研究」実践国文学会『実践国文学』66号、1-14
- 鈴木敦典・保阪靖人・成田克史・福元圭太 (2012) 『IT 基本用語辞典 (独・日・英)』同学社
- スタップス、マイケル著／南出康世・石川慎一郎監訳 (2006) 『コーパス語彙意味論 一語から句へ』研究社
- 塚本倫久 (2012) 『プログレッシブ英語コロケーション辞典』小学館
- 恒川元行 (2008) 「小規模な語彙調査に基づく基本語彙抽出の試み」九州大学大学院言語文化研究院言語研究会『言語科学』第43号、107-117
- 恒川元行 (2010) 「特定分野語彙抽出のための予備的調査—“Tierschutz (動物保護)” の分野を例として」九州大学大学院言語文化研究院言語研究会『言語科学』第45号、97-115
- 恒川元行 (2011) 「『～を始める』の意の beginnen の競合用法と文脈の予測」九州大学大学院言語文化研究院言語研究会『言語科学』第46号、137-145
- 恒川元行 (2012 a) 「研究報告 : beginnen と副詞の共起」九州大学大学院言語文化研究院『言語文化論究』第28号、217-222
- 恒川元行 (2012 b) 「テキストにおける名詞語彙の働きについて」九州大学大学院言語文化研究院『言語

科学』第47号、83-90

津留崎由紀子・中嶋敦子・金志宣・近藤彩・齋美智子・広田妙子（1997）「予測文法研究：後続文完成課題におけるJSLとJFLの予測能力について」お茶の水女子大学言語文化学会『言語文化と日本語研究』13号、182-209

寺尾康（2002）『言い間違いはどうして起こる?』岩波書店

寺村秀夫（1987）「聴き取りにおける予測能力と文法的知識」『日本語学』6巻3号、56-68

中村明（2010）『日本語語感の辞典』岩波書店

成田克史（2005）『答えはドイツ語』同学社

姫野昌子監修（2004）『日本語表現活用辞典』研究社

堀正広（2009）『英語コロケーション研究入門』研究社

堀口純子（1989）「コミュニケーションにおける聞き手による予測の型」筑波大学文芸・言語学系『文藝言語研究言語篇』7号、1-18

松浦恵津子（1996）「ニュース文聴解における予測能力 —テ形接続を中心とした日本語母語話者と日本語学習者との比較—」お茶の水女子大学言語文化学会『言語文化と日本語教育』12号、46-57

由本陽子（2011）『レキシコンに潜む文法とダイナミズム』開拓社

Meřćuk, Igor（2009）「コロケーションと語彙関数」Cowie, A.P.編／南出康世・石川慎一郎監訳『慣用連語とコロケーション—コーパス・辞書・言語教育への応用—』くろしお出版、27-68

Tschirner, Erwin（2008）Grund- und Aufbauwortschatz. Deutsch als Fremdsprache nach Themen - Lernwörterbuch. Cornelsen.

Quasthoff, Uwe（2011）Wörterbuch der Kollokationen im Deutschen. De Gruyter.

Vig, Sylvia 著／渡部重美監修（2002）『ドイツ語生活単語ノート』三修社