

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 4 月 25 日現在

機関番号： 14101
 研究種目： 基盤研究 (C)
 研究期間： 2009 ～ 2012
 課題番号： 21520436
 研究課題名 (和文) 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
 研究課題名 (英文) A Study on the occurrence of predicted words in German texts and sentences with applications to Japanese-German dictionaries and German textbooks for Japanese students
 研究代表者
 井口 靖 (INOKUCHI YASUSHI)
 三重大学・人文学部・教授
 研究者番号： 90151638

研究成果の概要 (和文) : 文を理解するために聞き手は出現する語彙の予測を行っている想定する。そのひとつはある語からそれ以降の語を予測する場合で、これについてはコーパスを用いてさまざまなコロケーションを調査した。もうひとつは語彙の出現そのものの予測で、それはテキストの種類、分野などから予測できるのではないかと仮定した。そこで、分野別コーパスを構築し、その語彙頻度を大規模コーパスと比較することによりいくつかの実例を提示した。これらの結果を独和、和独辞典や教材に反映することを試みた。

研究成果の概要 (英文) : This study is based on the assumption that a listener can predict words a speaker will say. This assumption was confirmed by researching various collocations via several corpuses. We also investigated the appearance of words in texts. We hypothesized that a reader could predict words that followed from types and branches of a text. We built a categorized corpus and showed some examples by comparing the frequency of words with that of larger corpuses. Applications of this research with respect to the making of German-Japanese and Japanese-German dictionaries and textbooks are discussed.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009 年度	1,700,000	510,000	2,210,000
2010 年度	500,000	150,000	650,000
2011 年度	500,000	150,000	650,000
2012 年度	900,000	270,000	1,170,000
年度			
総計	3,600,000	1,080,000	4,680,000

研究分野：独語学

科研費の分科・細目：言語学・言語学

キーワード：コロケーション・語彙・予測・コーパス・和独辞典・心的辞書

1. 研究開始当初の背景

従来から「コロケーション」という概念は一般にもよく知られ、辞典編集などで注目をされてきた。これについて見方を広げてみると、文の中に現れる語彙は先行する語彙に制限を受けるため、次に現れる語彙はある程度

予測がつき、それが文の理解に重要な働きをするのではないかと考えることができる。寺村(1987)はネイティブの聞き手には(文の一部を)「聴いた瞬間にその聴いた部分を理解するだけでなく、その後どのような語の連なりが来るかをも瞬間に予測する能力」がある

と仮定している。

その際、テキストの分野も重要な働きをされると考えられる。ある分野のテキストにおいてはそのテキストに現れる語彙の出現には一定程度の傾向があり、それも予測に大きな働きをするはずである。スタッブズ(2006)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」としている。

これらのことを背景に、文を理解するためにはどのような予測メカニズムがあるのかを明らかにしたいと考えた。

2. 研究の目的

この研究では、語彙の出現は周辺の語や文のテーマ・構造、テキストの分野の影響を受けるものであるという仮説を出発点とし、次のようなことを明らかにすることを目指した。

- ・特定の分野のテキストにおける語彙の現れは分野によりどの程度特性があり、その現れはどの程度予測できるのか
- ・文における語彙の現れはどのような情報からどの程度予測できるのか
- ・以上の成果は、言語教育にどのように応用可能か。

3. 研究の方法

- (1) 特定分野のテキストにおける語彙の出現について調査を行うために、いくつかの分野別コーパスを作成する。
- (2) 分野別コーパスにおける語彙の出現頻度を調べる。それを大規模コーパスと比較する。
- (3) 特徴的な分野別語彙についてそれが用いられた文での共起関係を分析する。
- (4) それらを基に予測メカニズムをモデル化する。
- (5) その成果を教材、和独辞典へ応用する。

4. 研究成果

(1) 分野別コーパスの作成と分析

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、予測という観点からは「浅く狭い語彙」は予測が難しく、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とすることができる。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。

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ドイツ語のキーワードを抽出する試みを行っている。黒田(2012)はドイツの雑誌の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通常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(例: 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))などが判明した。

また、井口(2013)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), 頻度分析を行なった概要が報告されている。上位は主として冠詞, 前置詞, 接続詞であるが, 22 位 Tepco (東京電力), 28 位 Reaktor (原子炉), 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), 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。また, 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。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ことになる。

このような調査により, 出現語彙を予測する場合, そのテキストのテーマや分野において特定の語彙を頭の中で活性化させていると想定できる。

(2) 文における予測分析

井口(2011)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などドイツ語で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」(語と語の相性)に当たるものであろう。

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堀(2009)の「意味的コロケーション」(語とある特定の意味領域との相性)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)では,程度が高いことを表わすドイツ語の程度副詞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という事なら,堀(2009)の「意味的コロケーション」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)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今後の課題として残った。

(3)応用

井口,恒川,黒田は『アクセス和独辞典』の編纂では中心的な役割を果たしたが,その際,コーパス分析で培ったコロケーションを特に重視した。しかしながらまだそれを体系化できるにまでは至っていないので,今後コロケーションを中心とした辞典の編纂を考えている。

予測研究は教材開発にも寄与する。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。井口(口頭報告)は原発事故を扱った教科書の語彙を井口(2013)で報告されている原発事故の雑誌記事の頻度と比較し,検討した。

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,教室での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。今後より多くの教員が使うことのできる教科書開発が課題である。

(4)心的辞書

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)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は,心的辞書の特定の語彙を活性化させていると考えられる。

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。そうであるならば,聞き手において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。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向かうべきであ

ろう。それは、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、今後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。

【参考文献】(5. にあげたもの以外)

- スタップズ, 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
- 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
- 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
- 成田克史(2005)『答えはドイツ語』同学社
- 堀正広(2009)『英語コロケーション研究入門』研究社
- 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

5. 主な発表論文等

(研究代表者, 研究分担者及び連携研究者には下線)

〔雑誌論文〕(計11件)

- 1) 「文の理解における予測について」, 井口靖, 三重大学人文学部文化学科『人文論叢』30, 2013, 123-136 (査読無)
<http://miuse.mie-u.ac.jp/bitstream/10076/12256/1/10C16168.pdf>
- 2) 「母語と第二言語における連語表現の処理の相違について —ドイツ語の連語表現に関する視線計測研究—」, 杉浦正利・成田克史・藤村逸子・山下淳子・梁志鋭, LET 中部支部研究紀要, 2013, 47-56 (査読有)
- 3) 「ドイツ語程度副詞の共起語の頻度に関する調査(2)」, 井口靖, 三重大学人文学部文化学科『人文論叢』29, 163-174, 2012 (査読無)
<http://miuse.mie-u.ac.jp/bitstream/10076/12224/1/10C16137.pdf>
- 4) 「テキストにおける名詞語彙の働きについて」, 恒川元行, 九州大学言語文化研究院『言語科学』No. 47, 83-90, 2012 (査読無)
- 5) 「研究報告: beginnen と副詞の共起」, 恒川元行, 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, 2012 (査読無)
<https://qir.kyushu-u.ac.jp/dspace/bitstream/2324/21810/1/p217.pdf>
- 6) 「特定分野のテキストに出現する語彙についての試行的調査—自動車試乗記を例として—」, 黒田廉, 『富山大学人文学部紀要』57, 139-148, 2012 (査読無)

- 7) 「ドイツ語程度副詞の共起語の頻度に関する調査(1)」, 井口靖, 三重大学人文学部文化学科『人文論叢』28, 117-129, 2011 (査読無)
- 8) 「『～を始める』の意の beginnen の競合用法と文脈の予測」, 恒川元行, 九州大学言語文化研究院『言語科学』No. 47, 83-90, 2011 (査読無)
- 9) 「学習独和辞典の現状と新たな試みについて」, 黒田廉, 『富山大学人文学部紀要』54, 165-178, 2011 (査読無)
- 10) 「ドイツ語話法詞と心態詞の多機能性について—*Vielleicht* haben Sie *ja* Recht. の例に関して—」, 井口靖, 三重大学人文学部文化学科『人文論叢』27, 47-58, 2010 (査読無)
<http://miuse.mie-u.ac.jp/bitstream/10076/11327/1/10C13920.pdf>
- 11) 「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—」, 恒川元行, 九州大学言語文化研究院『言語科学』No. 45, 97-115, 2010 (査読無)

〔学会発表〕(計3件)

- 1) 成田克史, 「ドイツ語正書法の再改変と独和辞典の対応—分離動詞を中心に—」, 日本独文学会東海支部, 2010年12月4日, 中京大学
- 2) 黒田廉, 「辞書作成の現場から」, 日本独文学会北陸支部, 2010年11月13日, 金沢大学サテライト・プラザ
- 3) 恒川元行, 「『～を始める』の意の beginnen の競合用法と『文脈の予測』」, 九州大学独文学会, 2010年5月1日, 九州大学文学部

〔図書〕(計5件)

- 1) 『アクセス独和辞典』三修社, (編集責任) 在間進・(編集執筆) 井口靖・恒川元行・黒田廉ほか32名, 総2072ページ, 2012
- 2) 『IT基本用語辞典(独-日-英)』同学社, 鈴木敦典・保阪靖人・成田克史・福元圭太, 総210ページ, 2012
- 3) 『言語研究の技法—データの収集と分析』ひつじ書房, 藤村逸子・滝沢直宏・成田克史 (第13章「音声の見方」執筆227-256)ほか8名, 2011
- 4) 『アクセス独和辞典第3版』三修社, (編集責任) 在間進・(編集委員) 井口靖・黒田廉・(編集執筆) 恒川元行ほか45名, 総2160ページ, 2010
- 5) 『アポロン独和辞典第3版』同学社, 成田克史 (責任編集執筆)ほか9名, 総1836ページ, 2010

〔その他〕

ホームページ

ドイツ語の各種辞書調査：
<http://www.hmt.u-toyama.ac.jp/Deutsch/jisho.html>

6. 研究組織

(1) 研究代表者

井口 靖 (INOKUCHI YASUSHI)

三重大学・人文学部・教授

研究者番号： 90151638

(2) 研究分担者

恒川 元行 (TSUNEKAWA MOTOYUKI)

九州大学・言語文化研究院・教授
研究者番号：70197747

成田 克史 (NARITA KATSUFUMI)
名古屋大学・国際開発研究科・教授
研究者番号：40128202

黒田 廉 (KURODA KIYOSHI)
富山大学・人文学部・准教授
研究者番号：00313578