

平成 29 年度 東京外国語大学オープンアカデミー
東京外国語大学語学研究所 企画
『コーパスから見えることば・文化・社会』
2017 年 10 月 10 日 (火) 第 1 回
「コーパスから見えるドイツのことば・文化・社会」
東京外国語大学教授
成田 節

講座の開始に当たって

はじめに、このオープンアカデミーの講座を企画いたしました語学研究所所長として一言ご挨拶いたします。実は去年の秋も、このようなテーマで講座を企画したのですが、その時は「コーパス」というテーマがよくわからないという印象を与えてしまったようで、受講者の集まりが足りず開講することができませんでした。今年も「コーパス」をキーワードとしてはいるのですが、言葉だけではなく、「コーパスから見える『言葉・文化・社会』」という、世の中のことにも目を向ける企画にしたところ、おかげさまで開講することができました。

講座は全 6 回です。今日は、私、成田がドイツ語についてお話します。来週以降はフランス語、朝鮮語、英語、アラビア語、最後は日本語を取り上げます。アラビア語の回はコーパスを使わず「アラブ映画のシナリオから見えたこと」というテーマでお話します。

本講座の講師は基本的に全員言語研究を中心にしているので、看板のわりに「文化」と「社会」の影が薄いと感じるかも知れませんが、なるべく「文化・社会」を入れようとしていますので、大目に見ていただければと思います。

今日の内容

それでは早速始めようと思います。パワーポイントを用意してきましたのですが、後半は実際にドイツのサイトにアクセスして、いろいろと実際に検索してお見せしようと思います。

まず今日の内容をざっと紹介します。この講座の柱の一つとして「コーパス」という言葉がありますが、まずはじめに、「コーパス」とはどういうものなのか

イメージを持っていただくために大雑把な説明をしてから、どんなコーパスがあるのか、実際にサイトにアクセスしながら見ていこうと思います。その後、今から 27、8 年前のコーパスに関する状況について、体験に基づいてお話ししたいと思います。それから、ドイツで大きなコーパスを公開しているところが 2 つあるのですが、そのうちの一つ、マンハイムという町にありますドイツ語研究所 — Institut für Deutsche Sprache (IDS) が公開している COSMAS II という検索システムについて、このコーパスを使ってどんなことができるかということと、もう一つの大きなコーパスであるベルリンのベルリン・ブランデンブルク科学アカデミー (BBAW) が公開している DWDS というコーパスについて、このコーパスを使うとどんなことができるかということをお話ししたいと思います。

コーパスとは

まず、「コーパス」がどういうものなのかについてですが、コーパスとは簡単に言うと、ものすごく沢山の量の文章を集めたものです。ただ集めただけではなく、言語研究のために検索ができるように加工してあります。「加工」と言っても、新聞記事の内容を変えるというようなことではなく、ある記事について「何年の新聞のどのあたりの記事」であるとか、読み方によっては検索できない部分があるとしたら、その読み方をつけるであるとか、文の「はじめ」・「終わり」・「改行」といった情報を加え、検索した結果が、どこから取ってきた文章なのかがわかるようになっていきます。あるいは、コーパスによっては品詞をつけてしまって、品詞で検索できるようにするといった加工がしてあります。コーパスとは、このような加工をした言語テキストの集合体です。

コーパスの中にはどんなものが収録されているかということ、例えば新聞記事、小説、実用書、メール、公開書簡などの手紙、日記といった文章、あるいは文章だけではなくて、講演の録音、インタビューの録音、ラジオ放送などが、文字に起こしたものとして、あるいは音声データとして入っているコーパスもあるようです。もともとは必ずしもそうではなかったのですが、現在「コーパス」というと、コンピューターで検索可能な電子データ化された大量の言語テキストの集合体をさすことが普通です。「コーパス」という語自体は必ずしもコンピューターで検索できるものに限らないのですが、今は普通コンピューターで検索できる大量のものを指します。コーパスという言葉は初めて聞いた方も、これでイメージは持っていただけたでしょうか。

各コーパスの概要

次に、コーパスにはどんなものがあるかですが、日本語のコーパスですと、国立国語研究所の「現代日本語書き言葉均衡コーパス」というものが公開されています。均衡コーパスというのは、例えば一つの小説からたくさんテキストを持ってくるでも偏りがあるので、新聞や小説といったいろんなジャンルから、時代的にもいろいろと取り混ぜて、ある一定期間の偏りのない代表的な姿が見えるように組み立てたコーパスのことです。それから、英語だと、かなり昔のものだともおもいますが、BNC (British National Corpus) というものが公開されています。ドイツ語では、例えば、ドイツ語代表コーパス (Das Deutsche Referenzkorpus - DeReKo) — この Referenz というのは、「参照」というような意味ですが、Referenzkorpus 「代表コーパス」というと、「均衡コーパス」と同じで、その時代を代表するような偏りのない、その言語の代表的な姿がわかるようなコーパスのことです。他に、ベルリン・ブランデンブルク科学アカデミーの各種コーパスなどがあります。

それでは、それぞれのコーパスのはじめのページを見てみようと思います。

(現代日本語書き言葉均衡コーパスのウェブサイトを表示)¹ 「書き言葉均衡コーパス」の規模はどのくらいのものだろうと想像されますか？(ウェブページを読み上げ) この『現代書き言葉均衡コーパス』は現代日本語の全体像を把握するために構築したコーパスであり、現在、日本語について入手可能な唯一の均衡コーパスで、書籍全般、雑誌全般、新聞、白書、ブログ、ネット掲示板、教科書、法律などのジャンルにまたがって1億430万語のデータを格納しており、各ジャンルについて無作為にサンプルを抽出しています。」本ならその本を一冊全部ではなく、例えばある本から200ページ分、別の本から200ページ分というように、いろいろなところからサンプルをとってきて組み立てるという発想です。1億語というと、どのくらいでしょうか。あまりピンときませんが、30年前ごろ初めてコーパスを見たころは220万語や100万語のドイツ語コーパスを「すごいな」と思ったものです。後で見るドイツの大きなコーパスは、今では100億語が収録されています。昔とは比べ物にならないほど膨大な量の言語データが検索できるよ

¹ https://pj.ninjal.ac.jp/corpus_center/bccwj/

うになっています。

(BNC のウェブサイトを表示)² BNC のウェブサイトの What ist the BNC? を見えます。「ブリティッシュ・ナショナル・コーパスは、1 億語のコレクションで、書き言葉と話し言葉、いろいろな分野から取ってきた、イギリス英語の広いセクションにまたがるような大規模なものになっている」とあります。20 世紀後半から検索ができます。このコーパスは以前からあるものですが、ここにある XML という形式で出ているものが 2007 年の最新版です。一番大きいものは、イギリス英語の全体像が見られるようになっているようです。

(COSMAS II のウェブサイトを表示)³ そしてドイツ語の COSMAS II です。ここにあるように、316 億 8000 万のデータが収録されています、⁴ ただし、ほとんどが新聞などです。今、新聞はインターネットでも見るできるので、権利の問題を解決するだけで、データを打ち込む手間はなく、どんどん規模を増やすことができ、数字が膨らんでいます。

(DWDS のウェブサイトを表示)⁵ これは 20 世紀のドイツ語コーパスという 1 億 2000 万語のコーパスで、時代ごと、新聞など分野ごとにも検索結果を出すことができるようになっています。21 世紀のドイツ語というプロジェクトも始まっていますが、まだ 1500 万語程度です。15 世紀から 20 世紀前半までのコーパスという古いものもあり、これも 2 億語を超えています。15 世紀から 20 世紀前半までのコーパスは、ちょうど作っているときにベルリンに行って話を聞いたことがあるのですが、古い本をスキャンして、OCR で文字化して組み立ててあります。昔のドイツ語の本なので、いわゆる「亀の子文字」で書かれているのが普通の時代なので、正確な OCR は難しく、最終的には人間が読んでチェックすることになります。それから新聞コーパスなどがあります。ベルリンのサイトで使えるものを全部合わせると、55 億 801 万語の規模のデータが入っていることになりま

² <http://www.natcorp.ox.ac.uk/>

³ <http://www.ids-mannheim.de/cosmas2/uebersicht.html>

⁴ 2019 年 2 月 4 日の時点では 547 億語。

⁵ <https://www.dwds.de/d/k-referenz#kern>

す。⁶

コーパス検索システム

コーパスは検索システムがないと使えません。それぞれ検索システムがあり、現代日本語書き言葉均衡コーパスには「中納言」や「少納言」といった、いかにも和風の検索システムがあります。私は国語研究所からアカウントをもらったまま、ちょっと複雑そうでまだ使ったことがないのですが、いろいろな検索ができます。もう一つの検索システムが NINJAL-LWP といって、これはすぐに使うことができます。⁷ ページを開くとすぐに検索画面で、ここ（検索窓）に検索項目を入れて検索します。今選んでいるオプションは、単語と単語の結び付きを調べる検索です。例えば、「上る」がどのような語と結びついているかを見てみると、「に上る」というのがたくさんあります。これがどんな風に使われているかを見ると、「人に上る」——これは「何人に上る」、「50 人に上る」という使われ方ですね——のように、「上る」の前にどんな表現があるかということも簡単に検索できます。あとでドイツ語と日本語の比較をやってみようと思います。

BNC にも、**Phrase in English** という検索システムがあります。これは例えば中央大学などでは図書館で契約しているようです。試しに、**beer** を検索してみましよう。3570 あります。こうして、どんな形容詞がつくかといった単語の使い方を見ることができます。

ドイツ語の COSMAS II と DWDS についてはあとで説明します。

初期のドイツ語コーパス

コーパスが登場した初期のドイツ語コーパスはどうだったかについて話します。当時西ドイツで、1970 年前後から、大量の文書を集め、機械で検索して研究に役立てるということが始まったそうです。最初のコーパスはマンハイムのコーパス **Mannheimer Korpus 1** で、延べ語数で約 220 万語です。異なり語数としてはもっと少なくなります。これは先ほどのマンハイムの国立ドイツ語研究所で開発しています。

もうひとつ、私自身が初めに使ったコーパスで、**Limas-Korpus** というコーパス

⁶ 2019 年 2 月 4 日の時点では 135 億語。

⁷ <http://nlb.ninjal.ac.jp/search/>

を紹介します。これは約 100 万語の均衡コーパスで、500 の様々のジャンルの様々な文書から任意の 2000 語を抜き出して作った、かなりバランスのとれたコーパスです。マイクロフィッシュのフィルムで流通していました。今はウェブでも見られますし、先ほどのマンハイムのコーパスの中にも入っています。

Mannheimer Korpus 1 を利用した研究として、Klaus Brinker という人の、「現代ドイツ語の受動態」と言う研究があります。先ほどの Manheimer Kopus 1 を使い、現代ドイツ語の受動態の全体像を記述したものです。受動態の問題点の一つとして、動作主が表示されているかどうかということがよく話題になります。例えば

1. Das Auto wurde von einem Mechaniker repariert. その自動車は修理工によって修理された。

2. Das Auto wurde repariert. その自動車は修理された。

1. は von einem Mechaniker 「修理工によって」という動作主が表示されている受動文です。2. のように、「誰によって」という動作主を表さないこともあります。文法書には、「ドイツ語の受動文は動作主が表示されないことが多い」と書かれているのですが、実際にはどうなっているのかというのを調査したわけです。

	文芸作品	娯楽小説	教養書	新聞	実用書
+動作主	13.4%	20.4%	16.3%	14.6%	2.2%
-動作主	86.6%	79.6%	83.7%	85.4%	97.8%

全体として動作主の表示がない受動文が圧倒的に多いということがわかりました。それから、文書のジャンルによって違いがあるということもわかりました。あるいは、受動文で使われやすい動詞や使われにくい動詞があるといったこともわかりました。この論文は受動文の動作主に関していつも引用されますが、そのもとになったのはコーパスのデータなのです。

日本のドイツ語学では 1990 年頃から

1990 年ごろから日本のドイツ語学では「コーパス」という言葉が使われるようになりました。先駆けは九州大学の樋口忠治というトーマス・マンの研究者でした。樋口氏はトーマス・マンのある語の用法を調べるのにコンピューターを利用

することを考えました。サイトの説明をみると、「このテキストデータベースに収められたファイルは、樋口忠治先生が独力で、しかもキーボードでの手入力という、まさに気の遠くなる地道な作業によって完成された」と本当だとは思いますが、疑いたくなるようなことが書かれています。⁸ 今なら、インターネット上にある電子ファイルを取ってくるのが（技術的には）簡単にできますし、本をスキャナで読み込んで OCR で文字化してスペルチェックをするといったこともできるので、簡単になっています。「これはこの当時手作業で九州大学大型計算機センターの協力で作った」ということです。トーマス・マンの全集をすべて、その後はゲーテも全部収録しています。マンハイムの研究所のコーパスには、日本の樋口教授が作ったものとして入っています。ただ、当時はまだコーパスを研究に使う時代ではなかったので、作ったものをどのくらい研究に使ったのかは良く知られてはいません。

その後、マンハイムのコーパスがフロッピーディスクで買えるようになりました。私も大阪大学の細谷行輝氏が買ったものを使わせてもらいました。それからさっきの Linas-Korpus といったものを使って、90 年ごろからコーパスを使うようになりました。

それまではどのように研究していたかと言うと、例えばある単語の使い方を知りたい、辞典には例がすくないと思ったときは、2 つ方法がありました。一つは例文を作って母語話者にチェックをしてもらい、一方は言えて他方は言えないことから、この単語はこのような意味で用法はこうだと論じる方法です。例えば——今思うとかなり人工的に思えるのですが——「部屋に入る」という時に、*treten* という動詞と前置詞句を使って *in sein Zimmer treten* という言い方と、*treten* に *be-* という前綴りがついた *betreten* という動詞と対格の目的語を使って *sein Zimmer betreten* という言い方があり、この 2 つの言い方の違いとして、*in sein Zimmer treten* の方には「2、3 歩」という過程の副詞句をつけることができるのに対し、*sein Zimmer betreten* という方は「2、3 歩」という副詞句をつけて過程に目を向けることはできないということ、例文をつくって、ドイツ語母語話者に聞いて示していました。

3. Er trat ein paar Schritte in sein Zimmer. 彼は 2、3 歩部屋に入った。

⁸ http://flc.kyushu-u.ac.jp/cms4/index.php/j-higuchi_corpus/

4.*Er betrat ein paar Schritte sein Zimmer. 彼は2、3歩部屋に入った。

ときとして誘導するようなやり取りをすることもありました。

日本語でも、「この例どうですか」と言われても、どういう状況を想定するかによって答えはかわります。でも、それを言い出してしまうとはじまらないので、ある程度割り切って「よいか、ダメか」ということでやっていました。正直に言う、かなり怪しいところもありました。

分析方法の転換(個人的なきっかけ)

ある時、論文を読んでいて、「こういうやり方の方がいいな」と思ったことがありました。⁹ 日本語の例文で「心臓をナイフで刺した」と「心臓にナイフを刺した」というのは、どちらも日本語として成り立ちます。

5. ジョンハ彼女ノ心臓ヲナイフデ刺シタ

6. ジョンハ彼女ノ心臓ニナイフヲ刺シタ

この論文の編者である国広哲也と言う人は、「刺す」について手元に集まった用例を「～に～を」型と「～を～で」型に分けてみると、用法の違いがおぼろげに見えてくると言っています。「～に～を」型の方は刺した結果何か有用なものが生産されたり、求める状態が生じたりするのに対し、「～を～で」型の方は傷害の例であったと言うのです。実際に国広哲也が挙げている例を見てみると、「～に～を」型の方は、

- a. (...) 串に肉片とわけぎを刺し続ける (...) —深田祐介『革命商人』下巻
- b. (...) くろもじの木の子にだんごをさして (...) —井上靖『夏草冬濤』
- c. (...) あの葉をちぎって土にさしておく (...) —立原正秋『はましぎ』
- d. (...) それが目を出している楸(ひさぎ)のすぐ傍に、そのがじゅまるの枝を挿した (...) —團伊玖磨『又パイプのけむり』
- e. (...) 鼈甲の簪(かんざし)をタボの右寄りに刺すと (...) —萩原葉子『蕁麻の家』

⁹ 国広哲也編集 (1980)「日英語比較講座 第2巻 文法」編者補説 (p. 269ff.)

これはどれも傷害の例ではありません。それに対して、「～を～で」型の方は、傷害の例です。

- f. 知夫は短刀で腹を刺されて、路上でうめいているところを (...) —黒岩重吾『人間の鎖』
- g. 殺したのは、酔いどれた外国の船員で、ジャックナイフで彼を刺したのだった。—同上
- h. (...) 注意された事に立腹して、注意した中川さんと奥さんを果物ナイフで刺して重傷を負わせている。—團伊玖磨『又パイプのけむり』朝日新聞社

こうして、無理に1対1の例文を作って母語話者に判断してもらうよりも、実際に例を集めてみた方が、違いが見えてくるのではないかと思うようになりました。

1990年頃、コーパス利用を始める

コーパスの存在を知る前はどのようにしていたかと言うと、実際に小説などを読んで、出てくるか出てこないかわからないものを慎重に探していました。この方法は時間がかかるし、時間をかけたら集まるという保証もありませんでした。90年ごろにコーパスの存在を知って、先ほど挙げた阪大の細谷先生のところで使わせてもらったり、Limas-Korpusをつかって例文を集めたりするようになり、例文を集める時間は非常に短くなりました。これでなにをやったかという、似たような意味の表現の違いについて、例をたくさん集めて調べました。

raten も beraten も「助言をする」という動詞です。raten の方は目的語（助言の相手）を「～に」という与格で表し、beraten は「～を」という対格で表します。

- 7. ich rate ihm 私は彼に助言する。
- 8. ich berate ihn 私は彼に助言する。

先行研究では、「特に違いはない」だとか、beraten の方は「強く助言する」と

いう意味だとか——どの程度なら「強い助言」と言えるのかなど分からないと思うのですが——いったことが書かれています。実際に例をたくさん見てみると、与格目的語の方は、文の中だったり文の外だったりしますが、「具体的な助言の内容」が必ず書いてあるということがわかりました。反対に、**beraten**の方は「具体的な助言の内容」が一切出てきません。つまり、与格の **raten** は「～に...することを助言する」という意味で、**beraten**+対格目的語は、個別的な助言内容を問題とするのではなくて、「～に助言者として対応する」という意味だと考えられるわけです。

私のコーパス利用

「コーパスから見る言葉と社会」という話をしているので、私がコーパス言語学の専門家だと思われるかもしれませんが、私は「コーパス言語学」の専門家としてではなく、ドイツ語の実証的研究や、あるいはドイツ語の授業の準備のために、とくに独作文の添削の補助手段としてコーパスを利用しています。添削の際は、何か変だけどこが変かわからない文や、良いのか悪いかわからない文について、似たような文があるかないか、どのくらいあるかを調べるといった使い方を行っています。コーパスの専門家には数学の知識や統計の知識が必要ですが、私は統計処理には既成のプログラムを利用して、あつまった用例は、ほぼ手作業で分類しています。

IDS (Mannheim)の COSMAS II

マンハイムのドイツ語研究所 (IDS) の 300 億語を超える大きなコーパス (COSMASII) を使ってどんなことができるかを紹介します。これは私がやったことではなくて、この研究所の方が行ったことです。ドイツ語のイメージとしては、文法がかっちりしていて、その点とつきにくくて、変な例外があまりないといったイメージがあるのではないかと思います。語形変化などはとくに初級文法のはじめにやっていて、例外はないというイメージがあるわけですが、実はその通りでないということが、コーパスをみるとわかってきます。

dieses Jahr「今年」という表現の語形変化において、属格の形は **dieses Jahres**「今年の」という形が、本来正しい形です。ところが、1980 年ぐらいから、新聞のようなきちんと校正された文章にも **diesen Jahres** という形が出てくるということが話題になっています。2009/11/9 の段階で、**dieses Jahres** という形が 76927 件に対

して *diesen Jahres* という形が 8863 件と、*dieses Jahres* という正しい形の方が圧倒的に多いことは多いのですが、*diesen Jahres* という形もかなり見られます。

これと同じように *dieses* を *diesen* とする間違いが、*dieses Monats* 「今月の」という別の表現では、*dieses Monats* という正しい形が 2404 件に対して *diesen Monats* という誤った形は 180 件とずっと少なく、*dieses Wochenendes* 「今週末」という表現では正しい *dieses* という形が 475 件に対して、誤った *diesen* は 0 件、*dieses Tages* 「今日の」という表現では正しい *dieses* 1861 件に対して、誤った *diesen* は 9 件、*dieses Morgens* 「今朝の」という表現でも正しい *dieses* 46 件に対して誤った *diesen* は 0 件とやはり少なく、なぜか「今年の」という *dieses Jahres* だけがダントツで誤った形が多いわけです。

(2009 から 2013 年の誤用例の推移表を見る)

2010 年には *diesen Jahres* という形がさらに増えています。他の名詞の誤った形はあまり増えていません。2013 年も *diesen Jahres* という形がさらに増えています。*dieses Ns* という形がみんな *diesen Ns* に変わってしまっているわけではなく、ある名詞の場合にだけ、誤用が増えているということがわかります。

これに対し、ドイツ語研究所では、「2008 年 12 月のデータで、*dieses Jahres* という正しい語形が 65290 例に対し、*diesen Jahres* という誤った語形が一割近くの 5712 例あり、*dies* は冠詞なので中性属格の語尾は *-es* が正しいものの、本来形容詞の語尾である *-en* がつけられる例が意外と多く、ここ 20 年で急増している」と述べています。

1. *dieses Jahres* 65290 例

2. *diesen Jahres* 5712 例

他の名詞についても調査してみると、「誤用は *diesen Jahres* が圧倒的に多く、他には *diesen Monats* (今月の) が少し見られる程度である」とわかったとのことです。

ところで、「今年の」に対して「昨年の」は何と何かというと、*vergangenen Jahres* といいます。*vergangen* は形容詞です。「翌年の」は *nächsten Jahres*、*nächst* も形容詞です。

3. vergangenen Jahres 「去年の」 25290 例

4. nächsten Jahres 「翌年の」 14281 例

ドイツ語研究所は、「去年」、「今年」、「来年」のうち、「過ぎ去った」の *vergangen* が形容詞で語尾は *-en*、「次の」 *nächst* も形容詞で語尾が *-en* なので、*dies* の語尾も *-en* だと考える人が出て来ているのではないかという結論を出しています。「前年の」も *letzten Jahres* で形容詞、「来年の」 *kommenden Jahres* も、もとは動詞ですが形容詞の語尾になっています。

5. letzten Jahres 「前年の」 12682 例

6. kommenden Jahres 「来年の」 6502 例

このように、規定語が形容詞の場合を調査して、この「誤用」が高頻度である理由を推測したわけです。「去年・昨年」と「翌年・来年」の規定語がどれも形容詞なので、「今年」の「今」(*dies-*)も形容詞の語尾にする「誤用」が増えているということです。これも、実際に使われた大量のデータを見てみて初めて分かることです。

BBAW (Berlin)の DWDS

マンハイムのコーパスは規模が膨大なのですが、少し使いにくいところがあり、最近では遠ざかっていました。ベルリン・ブランデンブルク科学アカデミー (BBAW) のコーパス (DWDS) はとっつきやすく、最近新しくなり、なかなか魅力的なサイトになっています。DWDS というのは、*Digitales Wörterbuch der Deutschen Sprache* 「ドイツ語電子辞典」の頭文字をとったものです。もともと辞典のサイトで、1961年から1977年に東ドイツで刊行された *Wörterbuch der deutschen Gegenwartssprache* (WDG 現代ドイツ語大辞典) という6巻本の辞典を基礎にして、この紙の辞典を電子データ化して、そこに100,000,000語からなる20世紀のドイツ語均衡コーパスを利用して現代化・デジタル化するという Berlin-Brandenburgische Akademie der Wissenschaften 「ベルリン・ブランデンブルク科学アカデミー (=BBAW)」のプロジェクトでした。その後コーパスの分量も増え、プロジェクトで使用可能な辞書も数が増え、いろんな辞典を取り込み、いろ

いろな観点から検索できるようになりました。もともとサイトの名前もドイツ語電子辞典ですので、辞典としても使えますが、コーパスの言語データから用例を探したり、コロケーション、その語とともに使われる語や頻度、あるいは結びつきの強さをしめす他の指数を調べたりすることができます。

例えば「おいしいビール」という組み合わせが、仮に 1 億語のうち「ビール」が 100 回、「おいしい」が 200 回出てきたとします。200 回のうち 100 回が「ビール」と一緒に出てきたとしたら、頻度は 100 回でも、これはかなり結びつきが強いといえます。一方で「おいしいウイスキー」が 500 回出てきたとしても、「ウイスキー」が 2 万回、「おいしい」が 3 万回あるとしたら、かなり結びつきは弱いということになります。こうした結びつきを示す指数があるのですが、こうしたものも検索できるようになっています。

ある語と別の語が近い位置に現れることを語と語の共起と言います。近い位置に現れるからと言って結びついているとは限らないですが、例えば「食べる」と「食べ物」が近い位置に現れていれば「何かを食べる」という結びつきであると推測することができます。正確さを期する場合は元の文を見て、きちんと結びついているのか、位置は近いけれども結びついていないのかを調べて確認してみる必要があります。

DWDS の Wortprofil「語の特徴」という検索システムで、ドイツ語の「食べる」という動詞 *essen* の目的語、それから「飲む」という動詞 *trinken* の目的語を調査してみました。これはドイツ人が何を食べることが好きかということではなくて、ドイツ語として「食べる」や「飲む」という表現が何を目的語として用いられるかということです。そこからひよっとしたら、実際にドイツ人が何を食べるのが好きかという推測もできるかも知れませんが、直接関係しているかどうかはわかりません。この結果を、日本語の NINJAL-LWP で「食べる」と「飲む」の目的語を調べてみた結果と比べてみて、何かが見えてくるかを調べてみましょう。

実際に DWDS のサイトを使って実演してみます。DWDS のサイトにはメニューが 3 つあって、辞典を調べたい人は *Wörterbücher*「辞典」、コーパスで用例を調べたい人は *Textkorpora*「テキストコーパス」、統計的なデータが欲しい人は *Statistische Auswertungen*「統計的評価」のメニューから調べます。今回は *Statistische Auswertungen*「統計的評価」を使い、その中の *Wortprofil*「語の特徴」という検索

システムを使い、essen「食べる」と入力してみます。essen という動詞と結びつく主語や副詞などの語を、安定していると言われる指数である logDice によって、上位 20 語まで検索します。logDice のほかに、単純な頻度でも検索できます。単純な頻度で検索すると、だいぶ結果が変わります。

いくつかの項目に分かれて検索結果がでます。Überblick「概観」という項目は全体的な結果です。hat Adverbialbestimmung「は副詞規定を持っている」という項目は、検索した語がどのような副詞と結びつくかを表しています。「満腹に食べる」や「好んで食べる」、「お昼に食べる」などです。hat Akk./Dativ-Objekt「は対格・与格目的語を持っている」という項目が、目的語になる名詞を表しています。各語の右側に logDice の数値が表示されます。共起の確率が高いものほど logDice の数値が大きくなります。その右隣りに表示されているのが単純な頻度です。2 語を比べたときに、一方の方がもう一方よりも、単純な頻度では多く、logDice では少ないということもあります。他に、検索した語がどんな前置詞句と結びつくか、どんな語と並列的な関係になるか、どんな語を主語にとるかを表す項目があります。今回は目的語のところを見ます。

ドイツ語の essen は上位から順に、Fleisch「肉」、Gemüse「野菜」、Brot「パン」、Fisch「魚」、Eis「アイス」、Kuchen「ケーキ」、Apfel「リンゴ」、Pizza「ピザ」、Schokolade「チョコレート」、Obst「果物」、Schweinefleisch「豚肉」、Suppe「スープ」、Ei「卵」、Käse「チーズ」、Salat「サラダ」、Kartoffel「ジャガイモ」、Wurst「ソーセージ」、Nudel「パスタ」、Steak「ステーキ」、Banane「バナナ」、Bratwurst「焼きソーセージ」、Kleinigkeit「軽いもの」、Brötchen「プレートヒェン：外が硬く中が柔らかい小さく丸いパン」、Abendbrot「アーベントブロート「ドイツで昔一般的だった温かくない夕食」、Schnitzel「カツレツ」、Rindfleisch「牛肉」、Currywurst「カレーソーセージ」、Auster「牡蠣」といった目的語をとります。牡蠣は、実は例を見てみるとすべて外国の話がドイツ語で書かれたもので、ドイツ人が牡蠣を食べるということではないようです。「ストラヴィンスキーがどこそこで牡蠣を食べた」というような例でした。

日本語の方を見てみると、ご飯、料理、ラーメン、弁当、夕食、肉、給食、餌、野菜、カレー、ケーキ、食べ物、魚、そば、パン、昼食、おやつ、お昼、草、うどん、夕食、お菓子、ご馳走、ランチ、寿司、飯、葉、食品、おにぎり、ステー

キといった目的語をとります。確かに（ドイツ語と）違いますが、この違いに何か感じるものはあるでしょうか。色々あると思いますが、まず言えることは、文化以前に言葉の問題として、日本語の「食べる」は動物も主語になります。「餌を食べる」や「草を食べる」は動物が主語です。ドイツ語の *essen* は、主語は人間で、動物の場合は *fressen* という別の動詞を使います。ただしこれは文化の問題なのか言葉の問題なのかわかりません。2つ目として、日本語の方は具体的に何を食べたのかが分からないことが多いのではないのでしょうか。「ご飯を食べる」という表現には「ライスを食べる」という意味と「食事をする」という意味があると思いますが、用例を実際に見てみると、「食事をする」という用例の方が多いようです。料理、弁当、夕食、朝食、食べ物、昼食、おやつ、お昼、夕飯、ご馳走、ランチ、飯、これらは具体的な料理ではなくて、「食事」ということです。ドイツ語はどうかというと、こういう例はほとんどありません。ドイツ語で「お昼を食べる」と言わないかということ全然そんなことはないのですが、実は、そういう時には対格目的語を使わないのです。その点で、今回の調査は同一条件ではないかも知れません。日本語では「お昼をたべる」というときも「を」を使いますし、「肉を食べる」でも「を」を使いますが、ドイツ語では「お昼を食べる」という時には、目的語なしで *zu Mittag essen* 「昼に食べる」というので、もしかしたら、こういう事情によるのかも知れません。ただ、このデータだけで考えてみると、ドイツ語では具体的に何を食べたのかを言おうとする傾向があるような気がします。

それから、これは遊びのようなものですが、第6位にアイスが来ています。これもドイツらしい気がします。ケーキ、リンゴ、チョコレートと甘いものが上位にきています。これにはひょっとすると食文化が関係しているかも知れません。

それからもう一つ、文化の問題なのか言葉の問題なのか悩んだのですが、ドイツ語の方には、豚肉、ソーセージ、ステーキ、焼きソーセージ、カツレツ、牛肉、カレーソーセージといったものがあります。食べ物関係には段階があります。食べ物の下に肉、肉の下に牛肉、豚肉、鶏肉、そのさらに下にソーセージ、さらにその下に焼きソーセージ、煮たソーセージ...というように、より基本的なレベルとより具体的なレベルがあります。ドイツ語は、より具体的なレベルの語が上

位 30 位に多いような気がします。それに対して、日本語は（具体的なレベルの語が）あまりありません。ただこれも、結果の表示の仕方をどうプログラムしてあるかという事情に原因があるかも知れません。例えば「たぬきうどん」や「きつねうどん」はどうなっているでしょう。もしかすると、「うどん」をみたら、「たぬきうどん」や「きつねうどん」が一緒になってしまっているかも知れません。確認してみましょう。「うどん」、「讃岐うどん」、「うどん」、「伊勢うどん」、「うどん」...そんなに「何々うどん」は多くないようです。半分以上はただの「うどん」です。本当は「うどん」と「何々うどん」を別々に検索するように設定して検索しなくてはわからないんですが、とりあえず手元にある検索結果からやや強引に違いを導き出そうとするなら、ドイツ語では具体的な料理や下位区分にあたるものがよく表現に現われるのに対して、日本語では、具体的なものではなく、「食事を」のようなものが良く表現に現われるということが言えるのではないかと思います。

同じように **trinken** と「飲む」を見てみると、やはりドイツ語の方は下位区分が多いようです。お酒の種類や、エスプレッソやカプチーノのような下位区分されたコーヒーの種類が表現されています。対する日本語の方には、大雑把に目的語を表示する傾向があるのかも知れません。

DiaCollo を利用して

もう一つ、是非お見せしたいと思っているのが、**DiaCollo** というサービスです。これを使って、ドイツの週刊新聞 **Die Zeit** の 70 年分の記事を検索し、**Atomenergie** 「核エネルギー」という語がそれぞれの時代にどういう語と出てきているかを調べました。はじめは **Atomausstieg** 「脱原発」を調べようかと思ったのですが、これだと方向性が決まってしまうから、肯定的な話題も否定的な話題も出てくるだろうということで **Atomenergie** という語を調べました。これはドイツのことに限らず、イランのことなんかも出てきます。**Atomenergie** という単語の前後 5 語以内にどんな表現が出ているかということ時代ごとに見てみます。

2 年ごとにベスト 10 を表示させます。1946 年には **Kontorolle** 「制御」という語が出て来ます。配布資料 (**DiaCollo** „**Kernenergie**“ 参考語彙) を見ていただけます。1. に挙げた語彙がどちらかというと肯定的な文脈で出てきたキーワードです。

Kontrolle「制御」、friedlich「平和的」、wirtschaftlich「経済的」、neu「新しい」、industriell「産業の」、Anwendung「応用」、Nutzung「利用」、Gebiet「(医療、産業などの)領域」、Entwicklung「発展」などです。例としては ein internationales Amt für die Kontrolle der Atomenergie zu schaffen「核エネルギー制御国際事務局(を設立する)」だとか、Möglichkeiten einer friedlichen Verwendung der Atomenergie「核エネルギーの平和的使用の可能性」、in der Anwendung der Atomenergie auf medizinischem、biologischem und technischem Gebiet「医療、生物学、工学領域での核エネルギーの応用」などです。

反対に 8. に挙げた語彙は否定的な文脈にあったキーワードです。Frage「(これでいいのかという)問い」、Ausstieg「脱却」、rot-grün「赤と緑の(社会民主党と緑の党)」——これは(ドイツの政党にはシンボルカラーがあり)赤がドイツ社会民主党、緑が緑の党で、(この2党が)1998年に政権をとったことで、(ドイツの政策が)反原発に変わりました——それから Veranstalterangabe「(デモの)主催者発表」、Brückentechnologie「(新技術ができるまでの)繋ぎの技術」といったキーワードです。例としては、einen allmählichen Ausstieg aus der Atomenergie「核エネルギーからの漸進的脱却」、... rot-grünen Regierung in Deutschland das letzte Stündlein der Atomenergie geschlagen?「ドイツの赤緑政権..核エネルギーの最後の時が来る?」、Das Märchen von der Atomenergie als Brückentechnologie「核エネルギーはつなぎの技術だというおとぎ話」などです。

5. はどちらともいえないキーワードで、Erdgas「天然ガス」、Kohle「石炭」、international「国際的」、Ministerium「省」、Iran「イラン」、Recht「権利」、Deckmantel「隠れ蓑」、Nuklearwaffe「核兵器」といった語です。具体例は im russischen Ministerium für Atomenergie「ロシア核エネルギー庁」、unter dem Deckenmantel der zivilen Nutzung der Atomenergie an Nuklearwaffen zu arbeiten「平和的核エネルギー利用の隠れ蓑の下で核兵器を開発する」などです。

では、(1946年から2010年までどんなキーワードが出てくるのかを年代を追って可視化した映像を)流してみようと思います。(用例数に沿った大きさの字で)時代ごとにキーワードが出たり消えたりします。みなさんご想像のように、ある

時代に（用例数が増えて）字が大きくなります。1946年から1979年までは Anwendung「応用」や Nutzung「利用」などが多いですが、1986年に初めて Ausstieg「脱却」が出て来ます。1996年にまた Ausstieg「脱却」が出てきて、2006年にアンゲラ・メルケルの CDU・CSU が政権を取りましたが、メルケル首相は原子力推進派でしたので、一度は Nutzung「利用」が増えます。そして、ある時に一気に文字が大きくなります。もちろん2011年です。

一つ一つ見ていきますと、戦後すぐ、1946年には Kontrolle「制御」、すこし飛んで1954年は Anwendung「応用」や Gebiet「領域」——何の領域かという、具体的には生物学、医学、農業などの領域での核エネルギーの実用的な応用などですね——それから friedlich「平和的」——兵器ではなくエネルギーとしてということですね——それから wirtschaftlich「経済的」など、プラス指向のキーワードが出てきます。1956年には Anwendung「応用」、neu「新しい」、wirtschaftlich「経済的」、industriell「産業の」、Entwicklung「発展」、friedlich「平和的」などが、1962年には Nutzung「利用」、friedlich「平和的」、1964年は Erdgas「天然ガス」、friedlich「平和的」などが出てきます。時間が押しているの少し飛ばします。1986年に初めて Ausstieg「脱却」が出て来ます。1994年に Ministerium「省」、これはロシア核エネルギー省のことです。チェルノブイリは1983年だったでしょうか？（受講生「1986年です」）1986年ですか？そうすると、チェルノブイリがあったから1986年に Ausstieg「脱却」が出てきたわけですね。1996年に rot-grün「赤（ドイツ社会民主党）と緑（緑の党）」が政権を取ります。2004年に Ausstieg「脱却」が一度現れなくなります。メルケルが政権を取った頃です。そして2010年、2年ごとで表示しているの2011年もここに入っています。急に増えていますが、あの（原発事故の）影響で、メルケル首相が政策を転換した後のものもここに含まれています。Brückentechnologie「繋ぎの技術」、Deckmantel「隠れ蓑」、aussteigen「降りる、やめる」などが出てきています。このような感じで、こじつけもあるかも知れませんが、70年分の新聞のデータを処理して見てみると、キーワードだけではなく、核エネルギーに関しての社会の動きも読み取れるのではないかと思います。（完）

コラム †

1984年夏・東ドイツ・ドレスデン



1984年7月東ベルリン。東側から見たブランデンブルク門。
門の直ぐ後ろは「ベルリンの壁」。フェンスから先は立ち入り禁止だった。

…・* ◀──▶ *・…

ドイツ語教員夏期講習の間宿泊した学生寮ではユーゴスラビアとブルガリアの同世代と3人部屋。ある時ユーゴの超イケメン君から「タカシ、知り合いが西の100マルク（約一万円）を東の800マルクと交換しないかって言うんだけど。8倍だよ。」正規ではもちろん1対1。「だいじょうぶかな。」「平気だよ。」「じゃあ、はい100マルク。」（一時間後）「タカシごめん。値切られて600にしかならなかった。」「十分だよ。ダンケ。」（早速街の古本屋へ。本は安いから600あれば山ほど買える。）選んでから店員に「持ち帰れないので、日本に送ってくれますか?」「もちろん。」「じゃあ、送料も今払います。」「いやいや、支払は日本に本が届いてからです。ドルでスイスの口座に振り込んで下さい。」（えっ、じゃあこの本の山は幾らになるの?!）本に使えなかった600マルクは講習の仲間との飲

食に消えた。

当時インターショップという、西側の通貨でしか買えない店があった。西側の通貨をもっているだけで優越感を感じながら、コーヒー（ネスカフェですが）とかブランデーなどを購入。ある時、洋酒を買いに行ったときのこと。店員に一万円札を渡すと、確認してくると事務室にさがる。しばらくして戻ると「このお札は使えません。」「なんで？この前はこのお札で買えたよ。」「カタログと違うんです。」「おかしいなあ。」「(カタログを持ってきて見せながら)「ここのこの文字がこのお札にはありません」「えっ？あつ、これは『見本』という意味。本物のお札には『見本』と印刷されていないよ。」「いや、しかし、カタログと違うお札は受け取れません。」「これは『見本』。『見本』なんて書いてあったら、それこそ偽物だよ。」「とにかく私としてはこのお札は受け取れません。」(あほか・・・)

35年前の夏、ドイツ民主共和国のドレスデンでのことでした。

ドイツ語を知るための3冊

・*・...‡...・*・*・*...‡...・*・*...‡...・*・*



清野智昭／山田敏弘「日本語から考える！ドイツ語の表現」
白水社 2011年

日本語学の専門家が日本語の例文を挙げてポイントを解説し、各言語の専門家が対応する外国語の文を提示して解説を加え、表現方法や発想の違いを考えるという語学好き向けのシリーズの1冊です。ドイツ語初級を一通り学んだ後、ドイツ語を深く理解したい方に最適な本です。

...・*・* ————— *・*・*・*

山口裕之「映画に学ぶドイツ語 — 台詞のある風景 —」
東洋書店 2012年

「嘆きの天使」「ブリキの太鼓」「ベルリン天使の詩」「ラン・ローラ・ラン」「グッバイ・レーニン」...
どれもドイツ映画です。1920年台から現代までの名作30本を厳選し、それぞれの印象的な台詞を中心に据えて作品を読み解きます。もちろん生きたドイツ語の勉強にもなります。



...・*・* ————— *・*・*・*



大園正彦「サッカーを楽しむドイツ語」三修社 2018年

サッカーと言えばドイツ、ドイツと言えばサッカー。ドイツ語よりサッカーが好きと言うドイツ語研究者が長年にわたって収集・整理したサッカー用語集。2014年ブラジルワールドカップ決勝戦などのLive-Ticker（ライブニュース）の対訳なども楽しめます。（サッカーに関心のない方、すみません。）

...・*・* ————— *・*・*・*