

コーパスとしてのインターネットと初級ドイツ語学習

岩 崎 克 己
外国語教育研究センター

0. はじめに

インターネットを利用した外国語学習の分野では、いまなお、ドリル型学習の配信が主流であり、それらの多くは、e-learningと呼ばれ、社会的認知度もこの分野が最も高い。それに対し、インターネットが持つ、1) 本物 (authentic) の情報リソースとしての側面、2) 協調学習 (collaborative learning) のプラットフォームとしての側面、3) あるいは自己表現や情報発信のための手段としての側面等を利用したプロジェクト学習や自己発見型学習なども、社会的認知度はそれほどではないが、徐々に増え始めている。本稿ではそのなかの1つとして、インターネットを言語コーパスとして利用した語彙学習の事例を取りあげたい。

今日、我々は、Googleに代表される検索エンジンを用いることで、インターネット上から様々な情報を引き出すことができ、インターネットは、人類共通の一大情報リソースと見なされている。しかし、この考え方をもう少し進めるならば、膨大なテキストデータを含むインターネットは、単なる情報リソースとしてだけでなくそれ自体を1つの言語コーパスと見なすこともできる。本稿では、巨大な言語コーパスとしてのインターネットを外国語学習に利用する試みの、初級ドイツ語授業への応用例について紹介する。

1. コーパスを利用した外国語学習 — 歴史的な回顧 —

外国語学習へのコーパスの利用は、欧米では1980年代のいわゆるコミュニカティブCALLの時代(岩崎 2009, 30-54)にさかのぼる。それは、PC(パーソナルコンピュータ)がまだ高価であり、教育機関におけるコンピュータの台数が1機関当たりせいぜい数台という時代である。もちろん、今日の日本におけるように学生が1人1台使って作業を行うことを前提としたCALL教室は存在していない。教育現場におけるコンピュータ利用は、島型に机が配置された教室の壁際にせいぜい1台から2台のPC置かれるか、別室に数台のPCが用意され、特定の作業のためにそこへ行って、複数の学習者が一緒に作業するという形が普通であった。こうした技術的な条件が、逆にコンピュータをパートナー学習やグループ学習の道具として使うという外国語学習におけるコンピュータの新たな役割をクローズアップし、大型コンピュータに繋がった個々のデータ端末に向かって学習者がひとりで行う、それまでのCAI型ドリル学習とは異なった、CALLのタイプを生み出した。

ただし、当時のPCは技術的な性能がまだ低く、特にハードディスク等の大規模記憶媒体を内蔵していなかったため、コーパスを利用した外国語学習は、実際には教案のようなアイデア(Hardisty/Windeatt 1988)の段階にとどまるか、実際の作業は、当時ようやく実用化されたワードプロセッサの検索機能やエディタプログラムのgrep検索機能(複数ファイル間での文字列検索と一覧表示機能)を利用するのがせいぜいであった(Higgins/Johns 1984, 85, 88-94)。したがって、コーパスの外国語教育への応用が実質的に始まるのは、1990年代以降、大容量外部記憶装置が安価になるとともに小型化し、大量のデータの保管や移動が比較的容易になってからであ

る。初期の時代の最も重要な事例は (Tribble/ Jones 1990) に紹介されているもので, KWIC (Key Word in Context) 形式のデータのワークシートを基に, コーパスを使った以下のような様々な作業のプロトタイプが紹介されている。

- 1) 文脈を利用した検索語の推理
- 2) 一連の関連した語のコーパス検索を通じた文法的な特徴の研究
- 3) 同音異義語や類義語の研究
- 4) コーパスで検索した結果を分割して, グループごとに分析させ, 後にクラス全体に報告させるようなグループ作業
- 5) 穴埋め問題の作成
- 6) 検索語の左の文脈と右の文脈を分割してシャッフルし, それを再び元に戻させるような並べ替え問題の作成
- 7) 学習者自身が書いたものを利用した補習課題の作成

(Tribble/ Jones 1990: 61)

このうち, KWIC 形式の検索結果を基に, その検索語を空欄に換えたり, 実際には存在しないナンセンス語 (phantasy word) に置き換えたり, 検索語の左右のデータ行をシャッフルし並べ替え問題を作ったりするなどの多くのアイデアとそれを使ったグループ作業のコンセプトは, その後のコーパスを利用した外国語学習に大きな影響を与えた。また, 検索語を空欄に換えてワークシートを作るなどのアイデアは, 後に Monoconc¹⁾ などのコーパス検索ソフトの機能の一部としても実装された。

1990年代半ば以降, インターネットが登場すると, いくつかのコーパスがネットワーク上で公開されたり, ネットワークを通じたコーパスデータのやり取りが容易になったりして, コーパスの利用は一層進んだ。さらには, ブラウザ上でも使えるコーパス検索ソフトを使ってインターネット経由で直接利用できるオンラインコーパスも登場してくる。現状では, BNC や Bank of English などの均衡コーパスの一部が有料で提供されている事例²⁾を除けば, オンライン上で自由に利用できる本格的なコーパスはまだ少ない。しかし, 今日では, ワープロソフトの使用の一般化などにより公的な性格をもったテキストのほとんどが最初から電子テキストとして作られているため, インターネット上でアクセス可能なコーパス, あるいは, 広い意味でコーパスと見なせるテキストの量は多い。たとえば, ドイツ語の例でも, IDS (Institut für Deutsche Sprache: ドイツ語研究所) が公開している COSMAS II (<https://cosmas2.ids-mannheim.de/cosmas2-web/>) は前者の典型であり, ヨーロッパ議会の多言語議事録1996-2009 (<http://www.statmt.org/europarl/>) や, おもに著作権の切れた文学テキストなどを中心にボランティアの手で電子化され, 公開されている Gutenberg Project (http://www.gutenberg.org/wiki/Main_Page) 中のドイツ語テキストなどは, 後者の例である。これらのテキストは, いずれも自由にアクセスできるため, それらをダウンロードし, コーパス検索ソフトを使って自分の PC 上で自由に利用することができる。上記の例の根底にあるのは, オンライン・オフラインの利用に限らず, インターネット上の特定の, しかも, ある程度信頼性のあるテキストのみをコーパスとして利用するという考え方である。しかし, 発想を転換し, たとえばインターネット上におけるドイツ語テキスト全体を1つの巨大な言語コーパスとして見なせば, たとえば, 検索エンジンをコーパス検索ソフトとし

て使うことで、インターネットそれ自体をコーパスとして利用することも可能になる。

英語の分野では、日本でも専修大学の佐藤弘明氏により、インターネット上の検索エンジンの機能を一部利用し、一定の検索語をもとに Web ページ上のテキストを検索し、その結果を KWIC 形式で提供する試みなどが数年前からすでに行われている。また、海外では、検索・表示の対象となる Web ページ数には制限があるものの、ドイツ語などの他の言語でも使える WebCorp LSE (<http://www.webcorp.org.uk/index.html>) のような、Web ページを対象としたコーパス検索ソフトも存在する。ただし、これらは、インターネットという巨大なデータを検索する必要上、いずれも検索の際に、Google などの既存の検索エンジンの機能に依拠せざるを得ない。そのため、プログラムの開発、公開、使用に対し制限が課せられていることが多い。また、オンライン上で動かせるこの種のプログラムを開発するだけの技術を誰もが持てるわけではない。それゆえ、インターネットを言語コーパスと見なしながらも、特殊な検索ソフトの独自開発は考えず、むしろ既存の検索エンジンの機能を単純に利用するだけで何ができるか、を考えるというアプローチも重要であろう。以下では、初級ドイツ語の学習者を対象にした、こうした CALL 事例について報告する。

2. 巨大コーパスとしてのインターネット

外国語学習におけるコーパスとしてのインターネットの利点は、大きく以下の3点にまとめられる。

- 1) インターネット上では、実際に使用されている本物の学習言語データが提供されている。
- 2) インターネット上には、言語使用レベル (register) や文体の多様性においてもその量においても、通常のコーパスとは比較できないほど膨大な言語データがある。そのため、特に出現頻度が少ない事例や、最新の表現であるため辞書などにはまだ記載されていないようなデータも調べられる。
- 3) インターネット上では、意味のわからない語であっても、具体的な概念を示す語であれば、たとえば検索エンジンの持つイメージ検索機能を利用することで、視覚的情報を利用することができる。

もちろん、検索用ソフトとして、既存の検索エンジンを使う場合、それほど高度な検索は期待できない。しかし、いくつかの技法を覚えれば、複数の単語からなる文字列検索やワイルドカード検索などの特殊な検索も、ある程度は可能である。たとえば、検索エンジン Google を使う際に便利な技法の代表例としては、以下の3つが挙げられる。

1) 2重引用符 (“ ”) の利用

2重引用符を使うことで特定の語連鎖を含む語を検索できる。たとえば Google 検索の場合、検索語として zum Frühstück esse ich という語連鎖をそのまま入れても、zum Frühstück esse ich という文字列を含むページを検索してくれるわけではない。Google 検索では、入力語が複数の場合それらの語の AND 検索をするだけである。すなわち zum と Frühstück と esse と ich という4つの単語を含むページであれば、それらがばらばらに散らばっていても、またどんな順番で出現してもヒットしてしまう。したがって、複数の単語をその連鎖の順番を

含む特定の文字列として検索する場合は、“zum Frühstück esse ich”のように、全体を2重引用符で囲む必要がある。

2) アスタリスク (*) の利用

アスタリスクは、方程式でいう変数 x のように、1つまたはそれ以上の任意の単語からなる文字列の代わり (=ワイルドカード) として使うことができる。また、2重引用符の中でも使うことができるので、たとえば“ich warte * der Post”などの形で使えば、Post (=郵便局) という女性名詞3格の単語と共起可能な前置詞をリストアップするための、文脈を指定した検索も可能になる。

3) 除外を意味するマイナス記号 (-) の利用

マイナス記号は、そのあとに続く単語が含まれてはならないという条件を表すのに使用できる。したがって、たとえば ich lerne と -kennen を組み合わせることで、kennen lernen (=知り合う)ではなく lernen(=知っている)の活用形を含む事例のみを検索することも可能である。

以上、外国語学習におけるコーパスとしてのインターネットの利点について述べたが、他方で、その欠点として以下の2つがよく指摘される。

- 1) 実際に使用されている本物の言語データは、多くの場合そのままでは初級の学習者にとっては難しすぎる。
- 2) インターネット上には、外国人の学習者が学習するのは好ましくないような言語使用レベルの表現や、非母語話者(時には一部の母語話者)によって書かれた誤った表現など、文法的にも文体的にも適切でない表現が存在する。

本物の言語データは初級の学習者にとって、難しすぎるという1つ目の主張は、インターネットをコーパスとして利用する場合のみならず、コーパスを利用した自己発見型学習全般に対してよく投げかけられる疑念である。一般に、外国語学習へのコーパス利用は、検索結果として得られる KWIC 形式などのコンコーダンスラインを読み解くための一定程度の言語能力を必要とするため、上級者向けの課題として考えられることが多い。実際に、過去の実践を見ても、中・上級の学習者を対象としたものが多く、初級の学習者を対象とした事例は少ない(岩崎 2005)。オフラインの場合は、検索対象とするコーパスとして、たとえば教科書本文やレベル別多読教材(easy readers)のように語彙レベルや文法レベルの限られたもののみ使うという方策も考えられる。しかし、インターネット上のテキストを対象とする場合、テキストのレベルそのものは原理的に言って操作できない。そのため、初心者でも遂行可能な課題となるよう、課題の内容とレベルを調整し、その難易度を下げることがとりわけ重要になる。2つ目の問題点としてインターネット上に「もともとある不適切なデータ」の存在が挙げられているが、確かにこれは深刻な問題である。これらは、検索エンジンを利用する場合、通常のコーパス検索ソフトを使う場合とは異なり、細かな条件を指定した検索がほとんどできないこと、コンコーダンスラインの表示やそのソート、コロケーションの表示や統計的な情報の利用がいずれもできないこと等に起因

する「検索手段の限界から出てくる不適切なデータ」の問題と相まって、言語学的にアプローチする場合はとりわけ致命的である。したがって、現状では、検索エンジンを使ったインターネット上の検索では、ある特定の用例があるということを確認する以外は、言語学的に基礎付けられた厳密な主張をすることは難しいであろう。それに対し、言語教育に利用する場合は、学習すべき代表的な用例を調べたり、2つの表現のどれがより適当かを判断したり、大まかに何が言えるかということの確認でも充分意味があるので、検索結果に表れる「不適切なデータの問題」は、それらの意味を判断するための学習方略を教えることである程度回避することができる。これらの問題点の回避のための方策については、次節で具体的に触れたい。

3. コーパスとしてのインターネットを用いた学習事例

事例1：計量表現と適切な名詞の結合

対象となる学習者：学習開始後1年以内の初級者

学習形態：2人から3人の小グループ

作業内容：図1のワークシートを使い、以下のようなドイツ語の数量表現と使われる典型的な名詞を、それぞれ2個から3個程度見つけ、出現件数と共にシート上に図示する。

学習すべき計量表現：eine Tasse / ein Glas / eine Flasche / eine Dose / eine Kanne / ein Kännchen / einen Becher / eine Tüte / eine Packung / ein Päckchen / ein Stück / eine Scheibe / ein Kilo / 100 Gramm / ein Liter

Google 検索でこれらの単位を表す単語がどんな名詞と結びつか調べ、例にならいヒット件数とともに書き込もう。

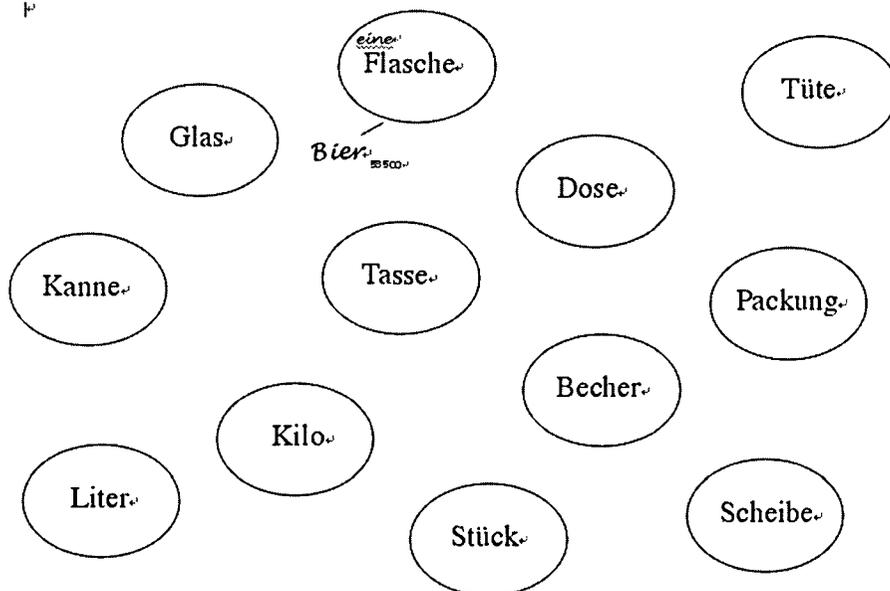


図1：使用するワークシート

初級者向けの授業でも可能な具体的な課題としては、「コップ一杯の水 (ein Glass Wasser)」や「一本のビール (eine Flasche Bier)」などの計量表現と名詞の結合を小グループで調べ、アソシオグラムなどの形で視覚的に表現し発表する等の課題が考えられる。一般に、教科書等では、スペース等の問題もあって、計量表現と名詞の組み合わせはごく少数の例が載っているに過ぎない。しかし、どの名詞の前にどの計量表現が使われるかは語感に関わる問題であって、コップや瓶などの一部の典型的な結合を除けば、ドイツ語学習者にとってもそれほど簡単ではない。この課題では、計量表現と名詞のコロケーションを調べることで、計量表現の自然な使い方を学ばせることを目的としている。具体的な手続きとしては、図1にある計量表現、たとえば Tüte を出発点に、“eine Tüte*”を検索語として Google³で調べ、(*)の部分によく出てくる単語を列挙する。すると、たとえば、Gummibärchen, Chips, Bonbons, Haribo, Backpulver, Luft, M&M's, Trockenhefe, Popcorn, Mandeln, Pfefferminztee, Erdnüsse, Ananas, Süßigkeiten, Rosinen, Suppen-Pulver, Lakritz, Reis, Puddingpulver, Butterkekse, Schokolade, Kirschen, Pommes, Waschmittel, Tropicfrutti, Maronen 等の名詞が得られる。しかし、インターネット上には、大量のデータがあり、中には間違っただけのものもある。したがって見つかった組み合わせが、たまたまヒットしただけの例外的なものか、一緒に使われる頻度の高いコロケーションなのかを確認するために、次のステップとして、アステリスク (*) の箇所をこれらの見つかった単語に置きかえて、再度検索する検証作業が必要である。たとえば、上の例では、Gummibärchen が見つかったら次のステップとして “eine Tüte Gummibärchen” で調べさせる。これで、かなりの数のヒットが再度得られたなら、これはコロケーションと見なせる。こうした手続きで、ワークシートに挙げた計量表現と共に使われる名詞を調べ、その名詞とヒット件数を含むアソシオグラムを完成させる。その際、単語の意味がわからなかった場合は、その単語を使って、Google のイメージ検索⁴ をさせる。一般に、計量表現と使われるのは、Gummibärchen や Pommes のような具体的な事物名詞や Haribo などの具体的な製品名であることが多い。そこで、イメージ検索のしかたさえ教えてやれば、初級の学習者でも、1) 単語の語形態、2) 計量表現の意味からの予測、3) 画像情報、等を手がかりに当該単語の意味を簡単に類推できる。

初級学習者の場合、見つけた単語の全てを理解したり覚えたりする必要はない。「1つの計量表現についてそれと一緒に使われる代表的な名詞を2個か3個ずつ調べ、その件数を書き込んだアソシオグラムを完成する」という具体的な課題があるので、それを実現するために必要な範囲内の情報を、Google 検索等を通じて、インターネット上から見つけられれば良い。多くの場合は、アステリスク (*) が指す単語を理解し、その組み合わせの件数の多寡からコロケーションであるかどうかを分析するだけでいいので、本物のドイツ語データの中で答えを捜すという作業をしながらも、難易度はそれほど高くなく、初級の学習者でも十分にこなすことができる。

事例2：文の断片を検索語に使い特定の語場に属する語をシステムティックに見つけ出す

対象となる学習者：学習開始後1年以内の初級者

学習形態：2人から3人の小グループ

作業内容：次ページ図2からなるワークシートを使い、以下のようなドイツ語の文の断片を用い、アステリスクの箇所に現れる(=同じ文脈で使われる)範列的(paradigmatic)な関係にある名詞を見つけ、シート上の表を完成させる。

検索に使う文の断片の例：

“Lieblingsfarbe ist */ “ich spiele gern */ “In meiner Freizeit * ich gern”/
 “ich esse gern */ “ich esse nicht gern */ “zum Frühstück esse ich */
 “ich trinke gern */ “ich trinke nicht gern */

Aufgabe:左の欄の表現をもとに、どんな単語がヒットするかで調べよう。わからない単語は、Bildsuche で見てみよう。

□□Google で調べる表現	□□□君の場合	□□□□□□□□□□Google で見つけた単語とヒット数	
		□□習ったことがあるもの	□□初めてみたもの
“Ich esse gern.”			
“Ich esse nicht gern.”			
“Mein Lieblingsessen ist.”			
“Ich trinke gern.”			
“Ich trinke nicht gern.”			
“Mein Lieblingsgetränk ist.”			
“zum Frühstück esse ich.”			
“ich spiele gern.”			
“In meiner Freizeit * ich gern”			
“ich bin * von Beruf”			
“ich bin von Beruf.”			
“Traumberuf ist.”			
“Ich schenke meinem Freund.”			
“Ich schenke meiner Freundin.”			
“Lieblingsfarbe ist.”			
“ich fahre mit dem.”			
“ich fahre mit der.”			

図2：使用するワークシート

たとえば，“ich esse gern */ “ich esse nicht gern */ “mein Lieblingsessen ist */ “zum Frühstück esse ich ”(私は*を食べるのが好きです / 私は*を食べるのが嫌いです / 私の好きな食べ物は*です / 朝食には、私は*を食べます) や “ich trinke gern */ “ich trinke nicht gern */ “mein Lieblingsgetränk ist ”(私は*を飲むのが好きです / 私は*を飲むのが嫌いです / 私の好きな飲み物は*です) のような文の断片を使えば、飲食物の範疇に属する単語だけを探し出すことができる。また，“Lieblingsfarbe ist ”(大好きな色は*です) は色を，“ich spiele gern */ “in meiner Freizeit * ich gern”(私は、*をするのが好きです / 暇なときには*するのが好きです) は、趣味やスポーツ・楽器名等をリストアップするのに使える。さらに，“ich bin * von Beruf”/ “ich bin von Beruf */ “ich möchte später * werden”(私の職業は*です / 私は将来に*になりたい) では現在の職業や将来つきたい職業を，“ich schenke meinem Freund */ “ich schenke meiner Freundin ”(私は彼に*をプレゼントする / 私は彼女に*をプレゼントする) では典型的なプレゼントのアイテムを，“ich fahre mit dem */ “ich fahre mit der ”(私は*に乗っていく) は交通手段を、それぞれ探し出すのに使える。このように、すでに習った表現を手がかりに、その表現とともに使う語群をシステムティックに見つけ出すという課題を出せば、語彙を増やす作業と既習単語のコロケーションを学習する作業を同時に行うことができる。その際、たとえば、

Luftgitarre (エアギター) などの、ドイツ語教員さえ知らない単語や新しい言葉でまだ辞書にすら載っていないような単語を、学習者自身が自ら見つけることもまれではない。このような作業では、たとえ Luftgitarre のような未知の単語に出会った場合も、学習者は、その語が使われる文脈(たとえば“ich spiele gern”)や語場(おそらく楽器かスポーツ)の情報は利用できる、ある程度はその単語の範囲を特定できる。また既習語である Gitarre との部分的な同一性などの語形態に関する(morphological)な情報も利用できる。そのうえ、すでに述べたように、Google イメージ検索を使うことで、視覚的な情報も利用できる。動詞や副詞などの場合は、そもそもドイツ語と日本語では、対象の捉え方自体が異なることが多く、語場のずれも大きいので、初級段階の学習者に1つの単語に対応する一義的な情報をひとりで見つけ出させるような使わせ方は危険である。また、特に動詞などの適切な語法は、ひとりでは見つけられないことも多い。しかし、Luftgitarre の例のように、視覚的な情報を含むいくつかの補助手段があり、語場も特定できる具体的な名詞を捜すのであれば、初級段階の学習者でも十分可能である。

なお、この種の語彙検索が単なる機械的な操作に終わらないようにするためには、こうした作業を、たとえば Landeskunde (ドイツ事情)に関する小作文などのより大きな課題の中に埋め込むことなども重要である。具体的な課題例としては、「ドイツ語圏の人達が日頃朝食にどんなものを食べているかを調べ、日本人の典型的な朝食と対比しながら報告しなさい」、「ドイツの若者がどんな趣味を持っているか調べ、皆さんの趣味と比較して見よう」、「ドイツの若者の理想の職業について調べ皆さんの将来の夢と対比させて簡単に報告してください」などの課題を出して、ドイツ語圏の人々と自分たちを比較する数個の文章と画像などからなる簡単なレポートを Word 等で作って提出させたり、WebCT や Moodle などの LMS (Learning Management System : 学習管理システム) のフォーラムなどに投稿させたりすることなどが考えられる。

最後に、インターネットを通じた自己発見型学習の際に必要な様々な学習方略についても触れておきたい。インターネット上には、既に触れたように、一定量の「不適切なデータ」が常に存在する。しかし、初級段階の学習者には検索によって得られたデータの適切性を判断するだけの言語能力はまだ無い。したがって、その適切性を間接的な形で判断したりするための学習方略を意識的に教える必要がある。そのいくつかを、簡単に列挙すると以下ようになる。

- 1) Google 検索や Google イメージ検索で得られた情報には、誤りや偶然的に一致しただけで必要とする情報ではないものも含まれていることが多い。したがって、原則として取捨選択や検証が必要である。
- 2) アスタリスク (*) を含む文の断片を使った検索で得られた語は、アスタリスクの部分その語に置き換えた完全な文を使って再度検索し、その件数を調べることで、たまたま出現した特殊な例か、頻繁に使われるコロケーションかを検証できる。
- 3) 検索に使用する文の断片の長さや含まれる語彙により、出現件数には差がある。したがって、ある用法の適切さは、絶対的な件数だけでなく、関連する類似表現の出現件数との相対的な比率で判断する。
- 4) 検索による出現件数が少ない場合や逆に多すぎる場合には、検索に使う文の断片を、目的に応じて工夫する必要がある。たとえば、“mein Lieblingsessen ist *”でヒット数が少なければ、mein を削って、“Lieblingsessen ist *”に変えたり、“zum Frühstück esse ich *”の esse ich の部分を isst man や essen wir に変えたりすることで、追加的な情報が得られる。

また、動詞の原形ではヒット数が多すぎるならば、適宜活用させたり、gern / heute などの一般的な副詞を追加する。

- 5) 元になる単語を結合させて新しい単語を作るドイツ語の造語法の特徴を考慮し、長めの単語の場合は、未知の単語でも、その中に形態素として知っている単語が含まれていないかを常に意識し、必要に応じて部分ごとに捜す。

この中で特に重要なのは、3つ目の学習方略である。同じような文脈で、A という語の出現回数と他のたとえば B や C という語の出現回数を調べ、それらの間に非常に大きな差が見られた場合、出現回数の少ないものについては、留保が必要である。検索に使う文脈としての文の断片の長さ、検索される語が一般的に多用される語か特殊な用語かという条件もあるので何件以下なら不適切という絶対数は一概に言えない。しかし、インターネット上のデータの規模からすれば出現件数が三桁以下の場合には注意が必要である。ただし重要なのは、絶対数でなく同種の表現と比較した場合の出現頻度の比率である。同一の文脈で、A という語を用いた表現と B という表現を用いた表現の差にたとえば 30 : 1 という開きがあった場合、たとえ B の表現が1,000件以上出現していても不適切な可能性が高い。というのも人間は実際には誤りを犯しやすく、30回に1回という比率は誤りの頻度としてもけして多いほうではないからである。学問的に厳密な結論を出すというのであれば、より厳密なアプローチが必要であるが、初級者を対象とした課題の遂行という観点からは、上記の学習方略を教え、頻度と比率の両方の基準で多いものから1つか2つを典型的な表現として覚えるという大まかな指針だけで十分実用になる。

4. おわりに

インターネットは、それ自体が巨大な言語コーパスでもあり、外国語学習のさまざまな側面にかかわる共通の土俵 (interface) でもある。また、インターネットは多言語メディアでもある。たとえば、Wikipedia の代表的な項目は、日本語だけでなく他の言語でも登録されていることが多い。以下は、言語コーパスとしての検索ではなく、情報リソースとしての検索型学習の例であるが、たとえば、Wikipedia の日本語ページで「村上春樹」を調べ、その作品リストを確認した上で左側の Deutsch の項目をクリックし、ドイツ語ページと日本語ページの対応を見ていけば、学生諸君が時々愛読書として挙げる村上春樹の『ノルウェイの森』が、ドイツ語ではその主人公の一人である直子という女性の名前を使って、Naokos Lächeln (直子の微笑) というタイトルであることを発見するのは、初級の学習者でもそれほど困難ではない。実際にインターネットを調べ、自己紹介の課題作文中で、Ich lese gern Murakami Haruki. Mein Lieblingsroman ist "Naokos Lächeln". と書いてきた学生がいたが、筆者はそれによって、『ノルウェイの森』の独訳名を初めて知った。また、同じような要領で日本のヒットチャートの調べ方とヨーロッパのヒットチャートのページを教えただけで、両者のデータ表示の並行性を手がかりにその週のドイツのチャートを調べ、気に入った曲のテキストを Songtext 等のサイト (<http://www.magistrix.de/>) で調べて印刷し、その曲のミュージックビデオを配信する YouTube の URL を付けて発表した学生もいた。

以上の例が示すように、言語コーパスとしてであれ、情報リソースとしてであれ、インターネット検索を外国語学習のための意識的な手段として教え、そのための学習方略を教えることは、外国語での自律学習の能力を高めることに繋がる。というのもインターネットを利用した自己発見型学習は、それを通じて学習者が新しい情報を効果的に学べるという意味では学習の手段であ

るが、インターネットを利用してドイツ語を使って何かを発見できるようになること自体は、学習方略の習得も含め、現代におけるドイツ語の運用能力の1つでもあり、したがってその獲得はドイツ語学習の目的でもあるからである。初級外国語授業におけるCALLの貢献分野の1つとして、こうした活動にももう少し重点が置かれるべきではないだろうか。

参考文献

- Aston, G. (ed.) (2001): Learning with Corpora. Bologna, Italy: CLUEB.
- Biber, D. / Conrad, S. / Reppen, R. (1998): Corpus Linguistics. Cambridge: Cambridge University Press. (邦訳: 齋藤俊雄他訳 (2003): 『コーパス言語学』南雲堂.)
- Bimmel, P. / Rampillon, U. (2000): Lernerautonomie und Lernstrategie. Fernstudienreihe 23. Berlin / München: Langenscheidt.
- Dodd, B. (1997): Exploiting a Corpus of Written German for Advanced Language Learning. In: Wichmann et al. (1997).
- Granger, S. ed. (1998): Learner English on Computer. London and New York: Longman.
- Hardisty, D. / Windeatt, S. (1988): CALL. Oxford: Oxford University Press.
- Higgins, J. / Johns, T. (1984): Computers in Language Learning. London: Collins Educational.
- Hunston, S. (2002): Corpora in Applied Linguistics. Cambridge: Cambridge University Press.
- Iwasaki, K. (2005): Einsatzmöglichkeiten eines deutsch-japanischen Parallelkorpus als ein Mittel für exploratives Lernen im DaF-Unterricht, Neue Beiträge zur Germanistik Band 4 /Heft 4, pp.166-181, JGG.
- Kennedy, G. (1998): An Introduction to Corpus Linguistics. London and New York: Longman.
- Knowles, G. (ed.) (1997): Teaching and Language Corpora. London and New York: Longman.
- Tribble, C. / Jones, G. (1990): Concordances in the Classroom. London: Longman.
- Wichmann, A. / Fligelstone, S. / McEnery, T. / Knowles, G. (ed.) (1997): Teaching and Language Corpora. London and New York: Longman.
- 岩崎克己 (2005) : オンライン型日独パラレルコーパスを利用した自己発見型ドイツ語学習の試み. 広島外国語教育研究 第8号, 広島大学外国語教育研究センター, pp.11-44.
- 岩崎克己 (2009) : 日本のドイツ語教育とCALL - その多様性と可能性 -. 博士論文 (広島大学総合科学研究科2009年10月提出 全283頁)
http://vu.flare.hiroshima-u.ac.jp/german/compling/iwasaki_call.pdf
- 齋藤俊雄 / 中村純作 / 赤野一郎 (編) (1998) : 『英語コーパス言語学』研究社.
- 鷹家秀史 / 須賀廣 (2002) : 『実践コーパス言語学 英語教師のインターネット活用』桐原ユニ.
- 中尾浩 / 赤瀬川史朗 / 宮川進吾 (2002) : 『コーパス言語学の技法 I. テキスト処理入門』夏目書房.

注

- 1) <http://www.athel.com/> 参照。
- 2) それぞれ <http://www.corpora.jp/~scn2/bnc.html>, <http://www.corpora.jp/~scn2/wordbanks.html> 参照。
- 3) 以下では、煩雑さを避けるために Google Deutschland を Google と略称するが、Google Deutschland は検索エンジン Google のドイツ語検索用インターフェースであり、<http://www>.

google.com/intl/de/ からアクセスできる。あるいは、日本語の Google で、入力欄の右の「言語ツール」から「ドイツ語」を選んでもよい。なお、通常の Google、すなわち日本語検索用インターフェースの Google をそのまま使っても、検索語がドイツ語なので、主としてドイツ圏のサイトにしかヒットしないため、検索結果に大きな違いはない。したがって、初級の学習者で、切り替えなどで、画面がドイツ語になる等の負荷に耐えられないような場合は、Google Japan のままでもいい。ただし、デフォルトでは、ドイツ語のドメインにあるサイトのみを捜してくれるので、一般に、Google Deutschland に切り替える方が、間違っただデータが出てくる比率が少ない。

- 4) Google イメージ検索についてもドイツ語インターフェースの Google Bilder を使わせたいのであれば、<http://www.google.com/imghp?hl=de> からアクセスできる。あるいは、Google Deutschland で、検索欄の上のリンク Bilder をクリックしてもよい。

ABSTRACT

Using the Internet as a Corpus in German Classes at the Beginner's Level

Katsumi IWASAKI

Institute for Foreign Language Research and Education
Hiroshima University

Nowadays, the Internet has become a huge common information resource for society. But if we shift our viewpoint a little further, the Internet itself could be considered not only as an information resource but also as a huge Corpus, including an enormous number of socially and stylistically varied texts. In this paper, we discuss the concept of learning activities in German classes for beginners, using the Internet as a huge German corpus, and the search engine Google as a kind of concordance software. In so-called mainstream CALL practices, which prefer to call themselves e-learning rather than CALL, the online distribution of drill-type learning materials still plays a main role. But here, in contrast, we want to show the possibility of the Internet for supporting explorative language learning.

First, we give a brief historical overview of corpus use in foreign language education from the early 1980s to today.

Second, we discuss the three main advantages of the Internet as a corpus for learners:

1. the Internet offers authentic data,
2. the Internet offers a wide variety of texts,
3. the Internet makes it possible to use so-called image search.

We also pay attention to the widely-alleged two main disadvantages of the Internet as a corpus for learners at the beginner's level:

1. authentic data are often too difficult,
2. the Internet often contains a certain portion of not only stylistically but also grammatically false data, which are produced by non-natives and also by natives.

Third, we illustrate the concept of concrete searching activities with the search engine Google, such as finding the examples of proper combinations of quantity indications and material nouns and presenting these results graphically, or systematic searches of the words which belong to the same word field with the help of expressions including asterisks as wild cards.

Finally we discuss some learning strategies which are important for explorative learning with the Internet as a corpus, in order to avoid the negative influence of the above-mentioned disadvantages.