

# ドイツ語におけるコロケーション分析と その辞典記述の問題点

井口 靖<sup>1</sup>・恒川 元行<sup>2</sup>・黒田 廉<sup>3</sup>・成田 克史<sup>4</sup>・カン ミンギョン<sup>5</sup>

## 要 旨

母語話者は語を単独ではなく、語と語の慣用的な結びつきであるコロケーションで記憶し、活用していると考えられる。それを独和辞典に記述できれば、ドイツ語を習得するのに効率的になる。本稿では、コーパスを用いたコロケーション分析の問題点を指摘した後、実際にコーパスを分析してコロケーションを抽出し、その中に意味的な関連を見出すことを試みる。そして、コーパスの分析結果と独和辞典、独独辞典、コロケーション辞典の用例を照合して、その問題点を探り、コロケーションと関連して独和辞典における用例のあり方について検討する。

## 1. はじめに

語を単に組み合わせても自然な文が作られるわけではない。母語話者は語を単独ではなく、語と語の慣用的な結びつきであるコロケーションで記憶し、活用していると考えられる。それを第二言語として習得するために効率的なことは辞典に記述することであろう。現在ある独和辞典、独独辞典、あるいは、コロケーション辞典はそれに十分応えているのだろうか。私たちのプロジェクトでは、ドイツ語の大規模コーパスを用いてコロケーションを調査し、その結果を各種辞典と照合し、今後独和辞典上で活用していくことを目指している。本稿はその試みの一端を示すものである。

まず2節で、コーパスを用いたコロケーション分析について概観し、問題点を指摘する。3節では、実際にコーパスを分析し、コロケーションを抽出して、それを「意味的関連性」の観点から整理することを検討する。4節、5節では、コーパスの分析結果と独和辞典、独独辞典、コロケーション辞典の用例を照合して、現在の各種辞典の記述の問題点を探る。6節では、独和辞典における用例の記述のあり方について検討する。

## 2. コーパスを用いたコロケーション分析の問題点

本節では、コロケーション分析における問題点のいくつかを簡単に整理しておく。

まず、コロケーションとその関連概念について述べる。コロケーションは語と語の共起関係を指すが、N-gram や Cluster のように必ずしも語連鎖である必要はなく、文中で離れて生起する語との関係もコロケーションに含まれる。<sup>6</sup> その範囲は中心語の前後4語とされることが多いが、最近ではより広い範囲を想定する場合もある。共起語については、辞書記述への応用も考慮し<sup>7</sup>、品詞や主語・目的語といった文成分別に分類して捉えることが増えている。本稿の分析に利用した DWDS-Wortprofil<sup>8</sup> でも、品詞や文成分ごとの共起語が自動抽出される（これについては後述する）。また、共起語を個々の語で

はなく、意味タイプに分類して捉えることもある。たとえば、動詞「読む」の場合、共起語として本、新聞、雑誌、心、空気、表情といった語が考えられるが、前の3語はたとえば「掲載物」、後ろの3語は「言語以外」としてまとめることができる(井口2017)。<sup>9</sup>さらに、意味タイプにまとめた文成分がどのように組み合わさって文を形成するかを分析することもある。たとえば Zaima et al. (2011) では、動詞 *verbringen* の分析において、目的語に現れる *Freizeit, Urlaub, Wochenende* などを「Zeitdauer: was」、副詞 *hier, dort* や前置詞句 *im Ausland, in der Firma, zu Hause* などを「Ort: wo」、様態副詞 *aktiv, gemeinsam, sinnvoll* などを「Art und Weise: wie」にそれぞれ分類し、(語順も含めた)それらの組み合わせを構成素結合(Konstituentenverbindungen)の形で示している。これらも広い意味でのコロケーション分析に含まれるものと考えられる。このようにコロケーションという概念は広い意味で用いられ、ほかの関連概念と重なることも多いため、それぞれの研究においてその定義を明確にしておく必要がある。

次に、コーパスを用いたコロケーション分析の問題点について、主に本稿の分析に用いた DWDS-Wortprofil を例に述べる。DWDS-Wortprofil は、主語・目的語・付加語・前置詞句など、文成分ごとのコロケーションデータを自動抽出できる便利なツールであるが、使用上の問題点もある。一点目は、検索結果にエラーが含まれる可能性があるという点である。エラー率は検索条件によって異なるため一概に言えないが、分離動詞(正確には前つづりの処理)や主語または目的語として抽出される名詞にしばしばエラーが見られる。二点目として、データの自動処理に用いられる統計指標の問題が挙げられる。堀(2012: 16f.)は、次の Davies/Gardner (2010: 6) を引き合いに、統計処理に用いられる MI スコア<sup>10</sup>の問題点を指摘している。<sup>11</sup>

Using MI is sometimes more an art than a science. If the MI is set too low, then high frequency “noise words” show up as collocates, whereas if it is set too high, then only highly idiomatic collocates are found. As an example … (以下省略)

すなわち、実際の使用実態を反映した、正しいコロケーションデータを得るためには適宜人の判断による「調整(しきい値の変更)」が必要であるということである。DWDS-Wortprofil ではログダイス(logDice)<sup>12</sup>順または出現頻度(Frequency)順に検索結果を表示できるが、使用実態とやや異なる結果が得られる場合もあり、その意味では上記と類似した問題を含んでいる可能性が排除できない。たとえば、DWDS-Wortprofil で動詞 *verbringen* を検索してみると、結合する前置詞句の上位20位には、*im Gefängnis, hinter Gittern, im Exil, in Untersuchungshaft, in den Gefängnissen, in der Zelle, in Haft* といった逮捕・処罰関連の語句が7つも含まれている。この結果は、「結びつきの強さ」と「実際の使用頻度」は必ずしも比例しないことを示していると言える。この事例については、統計指標による影響の可能性のほかにも、さらに2つの要因が考えられる。一つは、動詞 *verbringen* と結びつく場所語句には、*in Paris* のような地名を含む場合や *hier, dort* のような副詞が多く、特定の場所名詞を含む前置詞句が比較的少ないため、*Gefängnis* のような場所名詞と *verbringen* との結びつきが統計的に目立つと考えられることである。<sup>13</sup> もう一つとしては、テキストの種類の問題が考えられる。現在インターネットで利用可能な DWDS-Wortprofil (Wortprofil 2012) は約18億語からなるコーパスに基づいているが、DWDS-Kernkorpus (Referenzkorpora: 約1億語)とは違って、テキストの種類がバランスよく含まれているとは言えない。新聞記事が大半を占める可能性が高く、それも上記結果の一因と考えられる。このように、一見便利な DWDS-Wortprofil のようなツールには落とし穴があることを認めざるを得ない。とはいえ、大規模コーパスを用いる場合、手作業の分析にも限界がある。正確で有効なコロケーション分析を行うためには、自動処理や関連ツールを適宜活用しつつ、手作業による確認・調整を加味していく必要があると考える。

### 3. コロケーションにおける意味的関連性の分析 ー副詞を中心語としてー

上で述べたように、コロケーションのとらえ方にはさまざまな考え方があり、単なる語と語の偶然的結びつきではなく、共起可能な語群の意味分析から特定の意味領域との相性の問題として捉えることもできる。スタップズ (2006:129) は「実際に使用されている言語の大部分は、拡張語彙意味単位 (extended lexico-semantic unit) で成り立って」おり、この拡張語彙意味単位は固定慣用句のようなものではなく、「意味スキーマ」として機能し、「語彙 (中心語と共起語)」、「文法 (連辞的結合)」、「意味 (特定語彙領域に属する語を選択する傾向性)」、「語用 (暗示的意味や談話的韻律)」の4つにモデル化できるとする。このうち「意味」はたとえば *commit* が「社会的に非難される犯罪・行為・犯罪行為」で特徴づけられる語句と共起する (同: 84) ことなどを言うとし、「語用」の「談話的韻律」というのは「話者の態度」に関わるもので、たとえば、英語の動詞 *cause* は「不快な出来事を表わす語と圧倒的に多く共起」し、*provide* は *help, money, food* など「望ましいものや必要なものを意味する語と共起する」 (同: 86) ことなどを指している。本節ではこれらコロケーションの明示的・暗示的・談話的意味上の関連をまとめて「意味的関連性」と呼ぶことにし、ドイツ語の副詞とその共起語との間の意味的関連性について試行的に調査する。

仁田 (2002) は日本語の副詞を網羅的に分類したものであるが、「起動への時間量」を表す副詞を「僅少所要型」「中期所要型」「長期所要型」に分けている (同 246ff.)。このうち「僅少所要型」に相当すると思われるドイツ語の副詞を意味に基づいて便宜的に3つのグループに分ける。

「突然」系	<i>abrupt, jäh, plötzlich</i>
「思いがけず」系	<i>überraschend, unerwartet, unvermittelt, unvermutet</i>
「すぐに」系	<i>bald, dringend, eilig, gleich, schnell, sofort, unverzüglich</i>

以下の3.1. と3.2. は、DWDS-Wortprofilの *ist Adverbialbestimmung von* のデータに基づき、これらドイツ語の副詞と共起する logDice 上位20位以内の動詞および形容詞とのコロケーションを調べ、意味タイプ<sup>14</sup>に基づいて整理したものである。紙幅の都合で「突然」系と「すぐに」系について報告する。なお、*gleich* は「同じ」という意味で使われた事例がほとんどなので分析から除外する。

#### 3.1. 「突然」系

##### 3.1.1. abrupt

何らかの動作や状態の「開始」「終了・中断」「転換」を表すものがある。

終了・中断: *abbrechen, abbremsen, abreißen, aufhören, beenden, bremsen, enden*<sup>15</sup>, *innehalten, stoppen, unterbrechen, verstummen*

開始: *einsetzen*

転換: *abwenden, aufstehen, umdrehen, wechseln*

##### 3.1.2. jäh

ここでも「終了」を表すもの、「転換」を表すものが見られる。*erlöschen* などは「消滅」と分類したが、「終了」と密接な関係がある。その他に「落下」を表すものがある。

終了: *abbrechen, beenden, enden, unterbrechen, verstummen*

消滅: *erlöschen, zerplatzen, zerreißen*

転換: *aufsteigen, erwachen, hereinbrechen, umschlagen*

落下: *abfallen, abstürzen*

### 3.1.3. plötzlich

「出現」「消滅」を表すもののほか、(2-1)のように一見 *plötzlich* とは意味的に相いれないと思われる *stehen*, *dastehen* など「状態」を表す動詞もある。これは、*entdecken*, *hören* など「知覚」を表すものがあることから、*plötzlich* に（突然に）「状態」が「知覚」されたことを表すと考えられる。また、*ziehen* は「ナイフなどを取り出す」という場面が多く、全体として「出現」である。また *bekommen* も *die Masern bekommen* のような例では「出現」と言える。このようにコロケーションを3語以上で考える必要がある場合もある。

(2-1) Und plötzlich *stehen* da Häuser, das Dorf nimmt Gestalt an. (Die Zeit, 26.12.2011 (online))

出現・開始：anfangen, auftauchen, erscheinen, kommen, scheinen, springen

消滅：verschwinden

転換：geraten

落下：fallen

状態：dastehen, stehen,

知覚：entdecken, fühlen, hören, merken, sehen, spüren, wiederfinden

その他：bekommen, ziehen

「突然」系では、「転換」が共通し、「終了」「開始」「消滅」「落下」が2語に共通する。*plötzlich* は「出現」「状態」「知覚」などが特徴的で、客観的事態というより、その事態を認識者側から描写していることが想定される。

## 3.2. 「すぐに」系

### 3.2.1. bald

「知覚」を表す動詞が目につくが、「判明」を表すものもこれに近い。「その他」に入れてあるが、*angehören* は *der Vergangenheit* と結びつく例が多く、目的語とともに「過去のものとなる」という意味の「転換」になる。*folgen* は(2-2)のように時間的なものが多く、「開始」とも言える。*verlassen* はある場所を去ることから「消滅」に関係するかもしれない。

(2-2) Es wäre wünschenswert, dass die zwei weiteren Bände der Trilogie dem ersten Band bald *folgen* können.

(Die Zeit, 27.08.2012, Nr. 34)

知覚：erkennen, feststellen, finden, merken,

判明：erweisen, herausstellen

消滅：sterben, verlieren, verschwinden

到着：bekommen, kommen, zurückkehren

開始：beginnen

転換：ändern, geraten

その他：angehören, folgen, heiraten, verlassen

### 3.2.2. dringend

最も目立つのは「必要」を表す動詞や形容詞<sup>16</sup>である。これらと近い「要求」や「助言・警告など」も多い。ただし、厳密に見ると「要求」「助言・警告など」はそれら行為までの時間が僅少であることを表すのではなく、それら動詞が求める行為（(2-3)では「努力すること」で、*appellieren* は発話と同

時に遂行されている)までの「時間量」が問題となっている。コロケーションがかなり高度な意味レベルで働いていると言えるのかもしれない。

(2-3) „Ich *appelle* dringend an alle zuständigen Stellen in der Türkei, sich um Deeskalation zu bemühen und mit den Demonstranten das Gespräch zu suchen“, erklärte der SPD-Politiker.

(Die Zeit, 02.06.2013 (online))

必要：(動詞)      bedürfen, benötigen, brauchen, gebrauchen, suchen

(形容詞)      notwendig, erforderlich, nötig

要求：appellieren, auffordern, bitten, gebieten, wünschen

助言・警告など：abraten, anraten, empfehlen, raten, warnen

### 3.2.3. eilig

会議などの「招集」、建物やプロジェクトの「設立」、「移動」を表す動詞が多い。

招集：anberaumen, ansetzen, einberufen, herbeirufen, herbeischaffen, zusammenrufen, zusammentrommeln

作成・設立：aufstellen, errichten, hochziehen, verfassen, zusammenschustern, zusammenstellen,  
zusammenzimmern

移動：hasten, nachschieben, zurückrudern, zustreben

### 3.2.4. sofort

(2-4) のように状態を表す *tot* が最も多い。「知覚」も多いので、*plötzlich* と同様、そのような状態を知覚したことを表すのだろう。「思考」「反応」「通報」も「知覚」と関連しているのかもしれない。*fallen* は実際には *einfallen*, *auffallen* など「落下」以外の意味である。

(2-4) Er stürzte 30 Meter in die Tiefe, war sofort *tot*.

(Bild, 15.04.2005)

知覚：auffallen, erkennen, merken, spüren

開始：aufnehmen, beginnen, einleiten

終了：aufhören, beenden, einstellen, stoppen

通報：alarmieren, anrufen, melden, rufen

思考：denken

反応：reagieren

状態：tot

その他：fallen, verlassen

### 3.2.5. schnell

*schnell* においては動作様態を表すものが多く<sup>17</sup>、それらを除いて分類すると次のようにまとめられる。*lernen* は言語などの習得が多い。

到着：bringen, kommen

消滅：vergessen, verlieren, verschwinden

転換：ändern, drehen, geraten

知覚：merken, finden

反応：reagieren

その他：lernen

### 3.2.6. unverzüglich

政治的、官庁的な文体で使用されることが多い。なお、「終了」には、abziehen（軍隊の撤退）や freilassen（解放）を含めた。ergreifen は実際には Maßnahme とともに用いられ、「開始」を表す。

開始：aufnehmen, einleiten, umsetzen

終了：abziehen, aufheben, beenden, einstellen, freilassen, zurücknehmen

通報：anzeigen, benachrichtigen, einberufen, informieren, melden, unterrichten, veröffentlichen, vorlegen, weiterleiten

移動：begeben

その他：ergreifen

「すぐに」系は「開始」「知覚」が3語、「消滅」「到着」「転換」「移動」「終了」「通報」「反応」が2語で認められた。dringend, eilig は特有の意味的関連性を示し、unverzüglich は文体的に特徴がある。その他のものについては共通の特徴も多く、意味の違いについてはさらなる考察が必要である。

### 3.3. コロケーションの辞典記述

以上の分析により、類似した意味の副詞もそれぞれある一定の限られた意味を持つ動詞や形容詞と共起することが多いことが明らかになった。辞典においてコロケーションを表示する場合、単に語を列挙するのではなく、意味的なまとまりで表示することにより学習の効率化が期待でき、類義語の意味の違いも浮き彫りになる。

## 4. コーパスとコロケーション辞典におけるコロケーション分析 一名詞を中心語として

### 4.1. 比較調査の目的・内容

本節は、若干の名詞語彙について、以下の2種のドイツ語コロケーション辞典の記述、および DWDS-Wortprofil リストを比較することで、以下の3つの仮説を検証することを目的とする。

#### ■仮説

- 1) ある語には、コロケーションとして一定数の共起語群が予想される。したがって、異なるコロケーション辞典でも、同一の見出し語には一定数の共通する共起語が、コロケーションとして採録されていることが期待される。
- 2) コロケーション辞典の採録する共起語は、頻度がひとつの重要な選択基準となっている（Buhofer et al. 2014:XIX, Quasthoff 2010:XII）。したがって、コロケーション辞典の採録語は、DWDS-Wortprofil の高頻度共起語リストとも一定の共通性が期待される。
- 3) 基礎語彙は、多種多様な語と共起する可能性をもつがゆえに幅広く用いられ、使用頻度も高くなる。この性質のゆえに、高頻度の基礎語彙には特徴的なコロケーションは認めにくく、むしろ低頻度語の方に特徴的な共起語が見られる（Möhrling 2011:37, Reder 2006:62）。

#### ■比較の対象とした辞書・コーパス

- ・ Buhofer et al. (2014) のコロケーション情報のうち、見出し語名詞を1格主語または4格目的語とする動詞（「VERBEN」の項）、および付加語形容詞（「ADJEKTIVE/ADVERBIEN」の項）



- Quasthoff (2010) のコロケーション情報のうち、見出し語名詞を 1 格主語または 4 格目的語とする動詞（「V:NOM」、「V:AKK」の項）、および付加語形容詞（「A」の項）
- DWDS-Wortprofil で抽出・表示される「共起語頻度順上位 20 位リスト」のうち、見出し語名詞を 1 格主語または 4 格目的語とする動詞（「ist Subjekt von」、「ist Akk./Dativ-Objekt von」の項）、および付加語形容詞（「hat Adjektivattribut」の項）

#### ■比較の対象とした語（見出し語）

- Buhofer et al. (2014), Quasthoff (2010) に見出しの共通する、A の冒頭の 6 つの単義名詞 Abend, Abfall, Abgeordnete, Abitur, Ablehnung, Absicht（したがって、Quasthoff で多義扱いの Abfahrt, Abschnitt は除外）
- 両辞典に共通する単義名詞のうち、Jones/Tschirner (2006) の頻度順位 4000 番以内の任意の 7 名詞 Brot, Computer, Entscheidung, Hund, Hut, Katze, Vogel
- 両辞典に共通する単義名詞のうち、Tschirner (2008) の頻度順位 4000 番台の任意の 6 語 Brunnen, Duft, Kälte, Müll, Obst, Suppe

## 4.2. 集計

次ページの集計は、比較調査の結果を示したものである。Buhofer（b と略記）、Quasthoff（同 q）が挙げている共起語、および DWDS-Wortprofil（同 d）で示される頻度順 20 位までの共起語を比較し、d, b, q の 3 つすべて、またはいずれか 2 つでの重なりを数え（「重なり数」）、「共起語数」（当該見出し語の共起語の異なり数）に占める割合（「重なり率」）を計算した。

なお、紙幅の都合で、「見出し語＝共起する動詞の 1 格主語（das Brot verschimmelt など）」の集計は全体を、また表 4-1「見出し語＝共起する動詞の 4 格目的語（Brot essen など）」では一部の見出し語を省略した。

表 4-1：見出し語＝共起する動詞の 4 格目的語

頻度順位	見出し語	d	b	q	共起語数	重なり数	重なり率
313	Abend	20	2	14	25	5	20.0
739	Computer	20	15	52	60	20	33.3
1046	Hund	18	21	13	37	11	29.7
1757	Brot	20	11	19	33	12	36.4
2689	Hut	19	12	5	19	12	63.2
3948	Abfall	20	15	26	41	14	34.1
4073	Brunnen	17	2	0	17	2	11.8
4088	Kälte	13	4	4	14	4	28.6
4319	Suppe	20	6	11	25	8	32.0
4596	Duft	19	8	4	22	8	36.4
4927	Müll	19	8	42	47	17	36.2
4935	Obst	20	5	11	25	9	36.0

表 4-2：見出し語と共起する付加語形容詞

頻度順位	見出し語	d	b	q	共起語数	重なり数	重なり率	平均
313	Abend	20	16	91	98	17	17.3	
542	Entscheidung	20	46	204	219	39	17.8	
739	Computer	20	7	41	52	14	26.9	20.7
1046	Hund	20	17	32	50	13	26.0	
1500	Katze	20	13	24	44	12	27.3	
1757	Brot	20	13	20	34	12	35.3	
1878	Vogel	20	6	41	52	11	21.2	
1917	Abgeordnete	20	8	44	58	12	20.7	26.1
2018	Abitur	19	3	10	23	7	30.4	
2494	Absicht	20	31	81	94	26	27.7	
2689	Hut	20	7	10	28	9	32.1	30.1
3384	Ablehnung	20	13	85	89	22	24.7	
3948	Abfall	20	6	34	42	14	33.3	29.0
4073	Brunnen	18	1	6	22	2	7.4	
4088	Kälte	20	15	33	36	19	52.8	
4319	Suppe	20	13	23	38	14	36.8	
4596	Duft	20	18	42	49	18	36.7	
4927	Müll	20	4	23	35	9	25.7	
4935	Obst	20	12	9	31	7	22.6	30.3

(注)「頻度順位」は Jones/Tschirner (2006), Tschirner (2008) による。

### 4.3. 若干の比較結果

#### 4.3.1. 重なりの低さ

上の集計が示すように、辞書・コーパス間の重なりは 2~3 割程度と概して低い。それぞれの「凡例」において、Quasthoff は「5000 万文」(2011:XII) のコーパスに、また Buhofer は「スイス・テキストコーパス (2000 万語)、DWDS-Kernkorpus (1 億語)、独自インターネットテキスト・コーパス (7.75 億語)」(2014:XIX) に基づくことが述べられているが、重なりがこの低さであることは、それぞれのコーパスの代表性、コロケーション抽出基準や抽出手順等に疑問を抱かせる。すなわち、両辞書間の採録共起語の共通性に関する上記の仮説 1、また DWDS データとの共通性に関する仮説 2 は、いずれも明確には確認されなかった。結果として、そもそも何を「コロケーション」と見なすのか、を再検討する必要性が示唆されていると考えられる。

#### 4.3.2. 少なすぎる／多すぎる共起語の採録

Quasthoff も Buhofer も、一般的に「見出し語 = 1 格主語」の場合の共起語 (動詞) がほとんど挙げられていない。たとえば、Buhofer の Abend, Brot, Müll, Obst では 0 例、Hut, Brunnen, Suppe では 1 例の



みである。Quasthoff でも、Abitur, Brunnen 0 例、Absicht, Obst 1 例、などである。4.2. で「見出し語＝1 格主語」の場合の集計表を省略したのは、このように母数の少ない例が多く、集計の意味が乏しいと考えられたためでもある。逆に、付加語形容詞（たとえば、belegte Brote など）の場合、Quasthoff は一般的に非常に多くの共起語を示している（表 4-2 参照。Abend 91 例、Entscheidung 204 例など）。当然ながら、名詞によっては 1 格主語になる用法が少ないもの、逆に多くの形容詞と共起するものもありえようが、現在の共起関係記述はコロケーション辞典としてまだ十分に熟していないと思われる。

また、本節で行った比較では DWDS の共起語表示数を機械的に 20 語に設定しており、そのため DWDS の自動データ処理に問題がなければ、表 4-1、4-2 とも d の母数は 20 となっている。この数値はしたがって、前者「見出し語＝1 格主語」の場合には多すぎ、逆に後者「付加語形容詞」の場合には少なすぎることであり、重なり率の計算にはゆがみが生じている。

### 4.3.3. 頻度と重なり率の反比例

他方、仮説 3「むしろ低頻度語の方に特徴的な共起語が見られる」に関しては、表 4-2 に、その傾向を多少見て取ることができるかもしれない。すなわち、19 名詞と共起する付加語形容詞の辞書・コーパス間の重なりを、頻度順 1000 番単位の平均（表 4-2 右端の数値）で見ると、1000 番以内の 20.7% から比較的頻度の低い 4000 番位の 30.3% へと、おおよそ段階的な上昇が認められ、辞書・コーパスが、頻度順が下がるほどより多く特徴的な共起語を示している、との解釈も可能と思われるからである。しかし、表 4-1「見出し語＝4 格目的語」の場合と同様、個々の数値にはばらつきがあり、仮説 3 の明確な確認とはならなかった。

## 5. コーパスと独和辞典・独独辞典におけるコロケーション分析 一動詞を中心語として一

### 5.1. 調査の目的・内容

本節では、再帰動詞 *sich ereignen* を例に、学習独和辞典および学習独独辞典で共起語句の記載状況を確認し、コーパスの分析結果と比較、検証した結果について述べる。コーパスの分析には主に DWDS-Wortprofil を使用するが、2 節で指摘されているような問題があることから、IDS の DeReKo のデータを手作業で分析した結果も用いる。

### 5.2. 学習独和辞典と学習独独辞典の用例

外国語の習得には、何と言っても文法を学習し、多数の語を記憶することが肝要であるが、それだけでは自然な表現を作ることができない。学習者がその言語らしい表現を書いたり話したりするには、語の結合パターンを知ることが不可欠である。コーパスの発達により、そのような語の共起関係を知ることも容易になってきたが、独和辞典は学習目的のものといえども、この点について記述が十分でないように思われる。表 5-1 は、学習独独辞典、学習独和辞典各 5 点、および（学習者用ではないが参考として）DS の用例から、再帰動詞 *sich ereignen* の主語となっている語句をまとめたものである。採用が多いものを上から順に並べてある。（略号は参考文献において各文献の後に（ ）で示す。）

独独辞典では共起語句を複数挙げているものが多く、採用している語句も全体として多様である。表現用の辞書である DS は当然としても、LG, GR は DS 以上に共起語句を記載し、重視していることがうかがえる。<sup>18</sup> 独和辞典は CA を除き、記載する共起語句は 1 つのみである。全体としても Unfall,

表 5-1

主語となっている語句	独独辞典						独和辞典				
	LG	GR	PG	DB	WG	DS	AC	AP	CA	CR	PR
Unfall	O	O	O		O	O		O	O		O
nichts Besonderes	O	O				O					
Zwischenfall	O	O				O					
Explosion				O			O		O		
etwas Seltsames						O				O	
nichts Außergewöhnliches	O	O									
Vorfall	O										
Unglück	O										
Zugunglück	O										
Erdbeben		O									
viel		O									
etwas Besonderes				O							
etwas Merkwürdiges					O						
etwas Schreckliches					O						
nichts						O					

Explosion, etwas Seltsames の 3 種類に限られ、いずれも独独辞典に存在する。おそらく、これらは独独辞典からの借用で、とくに Explosion, etwas Seltsames は、独独辞典での採用状況から、それぞれ DB, DS が出典と推測される。例文もしばしば独独辞典に酷似しており、コーパス利用を謳う AC も必ずしも例外ではない。

(5-1) Die Explosion ereignete sich am frühen Morgen. (AC)

参照：Das Zugunglück ereignete sich am frühen Morgen. (LG)

### 5.3. DWDS-Wortprofil による共起語句の検証

Wortprofil では、sich ereignen の「主語」(「hat Subjekt」の項)として 89 語<sup>19</sup>が出力され、そのうち 2 語 Uhr と Jahr は 5.5. の理由により排除した。表 5-2 は、出力結果から logDice 順および頻度順上位 20 位までの語句を取り出したリストである。

表 5-1 の語句のうち、名詞はほとんど (Unfall, Explosion, Zwischenfall, Explosion, Vorfall, Unglück, Erdbeben) が表 5-2 に認められる。辞書で採用のもっとも多い Unfall は Wortprofil でも logDice 順、頻度順ともに 1 位である。とくに頻度では、2 位の Vorfall と大差をつけている。<sup>20</sup> 表 5-2 にはないが、名詞 Zugunglück も Wortprofil の出力結果には現れる。<sup>21</sup> 表 5-1 にあり、Worprofil で出力されないのは、名詞的用法の viel、不定代名詞 nichts および nichts Besonderes, etwas Seltsames のような不定代名詞 nichts, etwas + 名詞化した形容詞>の組み合わせである。

表 5-2

logDice 順			頻度順		
順位	主語	logDice	順位	主語	頻度
1	Unfall	11.8	1	Unfall	2414
2	Vorfall	11.2	2	Vorfall	1506
3	Unglück	11.2	3	Unglück	1503
4	Explosion	10.7	4	Explosion	965
5	Zwischenfall	10.6	5	Zwischenfall	792
6	Anschlag	9.6	6	Anschlag	566
7	Tat	9.5	7	Tat	535
8	Verkehrsunfälle	9.2	8	sie	487
9	Überfall	8.9	9	es	468
10	Beben	8.9	10	Fall	401
11	Katastrophe	8.8	11	Katastrophe	287
12	Zusammenstöße	8.4	12	Verkehrsunfälle	285
13	Detonation	8.3	13	Überfall	254
14	Erdbeben	8.2	14	Beben	241
15	Fall	8.0	15	Angriff	200
16	Wunder	7.9	16	Dinge	184
17	Tragödie	7.9	17	Prozent	183
18	Angriff	7.9	18	Zusammenstöße	163
19	Bluttat	7.7	18	Erdbeben	163
20	Attentat	7.6	20	Szene	154

#### 5.4. DeReKo のデータの手作業による分析結果

手作業による分析は、DeReKo の「書き言葉コーパス」(Korpora geschriebener Gegenwartssprache) から 500 例を無作為抽出し、主語となっている語句を取り出し頻度順にリスト化するという形で行った。表 5-3 はその一部、頻度上位 1～15 位、異なり語数 161 語中 29 語を示したものである。

小規模な調査結果であるが、表 5-2 の頻度順リストと比較した場合、上位を構成する語はあまり変わらない。順序の違いはあるものの、12 位までは 10 語が共通している。<sup>22</sup> 1 位はやはり Unfall で、2 位以下との差は大きい。<sup>23</sup>

Wortprofil では出力されなかった viel, nichts は、手作業のリストでは、それぞれ頻度 1、3 ではあるが、存在する。Wortprofil では＜不定代名詞 etwas, nichts + 名詞化した形容詞＞の組み合わせはまったくなかったが、辞書と同一例ではないものの、nichts Bewegenderes, etwas Schlimmes が見い出された。不定代名詞の有無を別にすれば、名詞化した形容詞は異なり語数 10、累計頻度 13 で、用法としては一定数存在する。

表 5-3

順位	主語	頻度	順位	主語	頻度
1	Unfall	107	15	Beben	3
2	Verkehrsunfälle	49	15	Zusammenstöße	3
3	Vorfall	41	15	Erdbeben	3
4	Unglück	18	15	Dinge	3
4	Tat	18	15	Brand	3
6	Zwischenfall	14	15	Auffahrunfälle	3
7	Explosion	13	15	Geschichte	3
7	was	13	15	Prozent	3
9	Fall	9	15	es	3
10	Arbeitsunfälle	7	15	Busunglück	3
11	Anschlag	6	15	das	3
12	Katastrophe	5	15	eine	3
12	sie	5	15	nichts	3
14	etwas	4	15	Schlägereien	3
15	Überfall	3			

### 5.5. コーパスによる用例の検証と学習独和辞典

Wortprofil, DeReKo の分析結果と比較する限り、学習独和辞典（および学習独辞典）で採用されている再帰動詞 *sich ereignen* の共起語句は、概ね実際に使用されているようである。この結果が、少なくとも独和辞典がコーパスの成果を活用しているためか、<sup>24</sup> あるいは（印刷物からの手作業による収集、直観など）従来の手法で採用した例がたまたまコーパスでの分析結果と一致しているだけなのかはわからない。

しかしながら、このような検証により、これまで独和辞典の用例を記述する際、独和辞典の用例を信用し、直観的・経験的にそこから選ぶしかなかったのが、より一般的なものはどれかが判断できる。独和辞典にない用例を多数載せることもできる。独自の、一定の客観的データに裏付けられた独和辞典をつくることが可能となったのである。今後は、独和辞典の翻訳段階を脱した、日本人学習者により有用な学習独和辞典をつくることを目指したい。

本調査で判明した Wortprofil の問題点について述べておく。1 つは誤解析である。出力結果の中に *Uhr* があったが、使用例が次のようにすべて主語以外であったため、*Uhr* はリストから排除した。

(5-2) Die Einbrüche ereigneten sich zwischen Montag 18.30 *Uhr* und Dienstag 9.15 *Uhr*.

(Der Tagesspiegel, 01.07.2004)

*Uhr* の場合は、語義からも例文からも、まず主語でないと判断できるが、主語としても主語以外でも使われる語は厄介である。たとえば、*Prozent* は、20 例中 17 例では (5-3) のように主語であったが、(5-4) のように主語でない例も 3 あった。不適切な例が混じっている場合、共起語として判断していいのかどうか難しい。<sup>25</sup>

(5-3) Dort ereigneten sich 78 *Prozent* aller tödlichen Unfälle. (Die Zeit, 14.03.2007 (online))

(5-4) Mit 11 316 Unfällen ereigneten sich im Monat Mai 1,1 *Prozent* weniger Unfälle als im Vorjahresmonat.  
(Bild, 15.07.2003)

2つめは、抽出しない語句があるらしいことである。手作業によるリストと比較すると、名詞的に使われた他品詞の語および不定代名詞などは出力されないようである。プログラムが本来の品詞で解釈する、あるいは用法の違いと区別できないためかもしれない。

## 6. 独和辞典における用例分析

本節では、独和辞典における例文、中でも現実に使用された文を用例として用いることの意義について、コーパスやコロケーションと関連して考えてみたい。用例には、見出し語の語法を示すとともに、見出し語がどのような事態を描写するのに用いられるかを具体的に示すことで、意味の説明だけでは記述しきれない情報を与える機能がある。また、どのような話し手がどのような場面でそれを用いるかまでを示唆することもあり、辞書の中で非常に重要な役割を果たしている。

各種独和辞典がどのようにして用例を集め、選択しているかは定かではないが、類似あるいは同一の用例が多数見られることから、既存の独和辞典、独和辞典を参照していると推測される。その中で、『アクセス独和辞典 第3版』（以下、『アクセス』と略す）は、見返しで「③大規模コーパスを利用した初の独和辞典」を謳い、「■コロケーション（連語）調査を含む、用例の見直しを行ない、生活の息吹がより感じられる実用的な用例記述にしました」と記し、次の例を挙げている。

(6-1) Kein Anschluss unter dieser Nummer. この電話番号は現在使われておりません

(6-2) Gebührenpflichtige Parkplätze stehen zur Verfügung. 有料駐車場あり（施設の案内などで）

用例収集の情報源として電子化された大規模コーパスを活用することは、用例選択の幅を飛躍的に広げ、自立した選択を可能にするという意味で画期的であり、実際の使用例であるという点でこれに勝る確かさはない。

上記2例のうち最初のは Nummer の例文であり、文字どおり、使用されていない番号に電話をかけたときに自動で流れるメッセージと思われるが、Nummer（電話番号）と結びつく前置詞として unter（…のもとに）が示されていることから、例えば anrufen（電話する）を学んだ者が Rufen Sie mich unter 012345 an! (012345 [の番号] で私にお電話ください) のような文を組み立てる際に応用が利き、同時に「電話」と関連の深い重要語 Anschluss（接続）も知ることができる、好ましい例文といえる。他方、二つ目は Parkplatz（駐車場）の例文であるが、これに関連付けて覚えておいてもよいと思われるのは gebührenpflichtig（有料の）という比較的重要度の低い語であり、Parkplatz の例文というよりは zur Verfügung stehen（使用に供される）の例文、つまり、「使用に供される」ものにどんなものがあるかを示す例文のように思える。『アクセス』本体の見出し語 Parkplatz には、einen Parkplatz suchen <finden>（駐車場を探す〈見つける〉）、den Wagen auf dem Parkplatz abstellen（車を駐車場に置く）という例句も挙げてあるので、よく使う表現を見逃すことはないが、加えてこの例文を挙げることの意義については議論の余地がある。

ところで、今日のような電子的な大規模コーパスが実用化される前の独和辞典では、文学作品から用例を引くことがあったが、1960年代末から1980年代にかけて出版された、ともに6巻からなる Wörterbuch der deutschen Gegenwartssprache と Brockhaus-Wahrig, Deutsches Wörterbuch とでは考え方が正反対であることが興味深い。前者は、1) 広く知られた名言と受け取れるもの、2) 当該の見出し語

を使用するにあたり典型的ではない言い回し、3) 古くなったり使用がまれな言い回しの場合における文学作品からの引用の効用を述べている(第1巻、23ページ)。これに対して後者は、用例の選択が偶然であるという印象を与えないためには、採用される使用例は6巻本の一般的なドイツ語辞典の枠を大きく超えるとして、文学作品から引用はしないとしている(第1巻、15ページ)。

*Duden. Das große Wörterbuch der deutschen Sprache* は文学作品から引用することの是非についてその立場を表明していないが、上記、*Wb. d. dt. Gegenwartsspr.* と同様、文学作品から実例を引いている。その Parkplatz には次の例文がある: Ein P. für mindestens 100 Wagen, zurzeit leer (Frisch, Montauk 7) (少なくとも100台は入る駐車場、今はがらんとしている)。この *Duden* の見出し語 Parkplatz には二つの意味項目がある。第一に何台かの車を停めることができる比較的広い場所、第二に車1台を停めることができる駐車スペースであるが、上の例文は最初の意味項目に添えられている。つまり、Parkplatz には、車1台分の駐車場の意味だけではなく、何台もの車を収容できる駐車場の意味があることを証拠立てているのである。しかも駐車場とは縁の深い leer という形容詞まで付いた、なかなかよい例文であるといえよう。さらに、この例文の良さはその味わい深さにある。読めば、空っぽの大駐車場のわびしい情景が目につかび、それを見つめる人の気持ちを共体験し、100台入ることを見込んで造られた駐車場になぜ今は1台も停まっていないかという背景にまで思いを巡らせることができる。無論、*Brockhaus-Wahrig* が言うように、これも偶然選ばれたひとつの実例に過ぎず、他にも適切な文はいくらでもあるはずだが、たまたま選ばれたある特殊な状況における Parkplatz の描写ではあっても、その状況が与えられたときに、単なる Parkplatz という語がどのような表情の変化を見せるかを知ること、その語を深く理解することにつながり、実例が持つ力には侮りがたいものがあるといえる。

コーパスから選ばれた例文にも似たようなことが言えるのではなかろうか。文学作品からの引用の持つ強みが味わい深さだとすれば、コーパスからの引用の強みは生活の香り、『アクセス』の言葉を借りれば「生活の息吹」だろう。先に同書の Parkplatz の例文についてやや批判的に書いたが、これとてドイツで暮らし、運動施設や娯楽施設に車で行こうとするとときに駐車可能かどうかを調べて初めて遭遇する表現であり、日本ではまず目にすることはない生のドイツ語表現である。「味」や「香り」が辞書に奥行きを与えることは間違いない。『アクセス』は辞書をひと回り大きくすることにより、このような実際の言語使用に基づく例文やその他の企画のためのスペースを稼ぎ出したと思われる。同書と対比されることの多い『アポロン独和辞典 第3版』や『クラウン独和辞典 第5版』はそこまで踏み切っていない。それはそれぞれの編集方針によるのであろう。しかし、紙幅の多寡にかかわらず、「味も香りもある」独自性の高い例文をどう提示していくかは、いずれの辞書にとっても重要な課題であると思われる。各辞書の編者のこれからの「さじ加減」に大きな期待が寄せられる。

## 【註】

本稿は科学研究費補助事業基盤研究(C)(一般)(平成28年～平成31年)(課題名:「ドイツ語基礎語彙のコロケーションに基づく意味分析とその独和辞典記述方法の検討」 課題番号:16K02667 研究代表者:井口靖)及び基盤研究(C)(一般)「名詞を核とするコロケーションの収集と整理―日独対照表現データベースの作成―」(平成27年～平成29年度)(課題番号:15K02521 研究代表者:恒川元行)の助成を受けたものである。



- 1 三重大学教養教育機構教授 1, 3 節担当
- 2 九州大学言語文化研究院教授 4 節担当
- 3 富山大学人文学部教授 5 節担当
- 4 名古屋大学人文学研究科教授 6 節担当
- 5 東北大学高度教養教育・学生支援機構准教授 2 節担当
- 6 N-gram は、コロケーションより広い意味で用いられる場合もあれば、その一部しか捉えない場合もある (田畑 2012: 135).
- 7 辞書におけるコロケーションの扱いについては赤野 (2012) で概観することができる.
- 8 <https://www.dwds.de>
- 9 そもそもこのような自由結合はコロケーションと見なさない立場もある.
- 10 Mutual information (相互情報量) の略で、コロケーションの結びつきの強さを示す指標として用いられる総計量の 1 つ. 2 語のそれぞれの度数と共起度数, コーパスの総語数から得られる. 3 以上の値で統計的に有意な結びつきだと言われる (赤野他 2015: 403).
- 11 田畑 (2012: 119) にも同様の指摘が見られる.
- 12 コロケーションの結びつきの強さを示す指標として用いられる総計量の 1 つ. 2 つの要素の類似度を示し, 0 (2 語が全く共起しない場合) ~1 (2 語が必ず共起する場合) の値となるダイス係数に基づいて計算される (赤野他 2015: 401).
- 13 DWDS-Wortprofil には, 「人名」「地名」というように固有名詞を意味グループにまとめて表示する機能はない. なお, DWDS-Wortprofil で Gefängnis を検索語にして調べてみると, 「前置詞+ Gefängnis」との結びつきが高い動詞は sitzen (im Gefängnis), entlassen (aus dem Gefängnis), verbringen (im Gefängnis) などであった.
- 14 どのような意味タイプを設定するかが問題であるが, 本稿の範囲内であるべく共通のものを設定するように努めた.
- 15 たとえば, enden は endet, endenden と変化形で出てくるために, 全体で 20 にならないこともある.
- 16 logDice20 位に neu があがっているが, brauchte dringend neue Daten のように直接には dringend に関係していないと思われる.
- 17 wachsen, gehen, fahren, laufen, steigen, entwickeln など.
- 18 LG はコロケーション欄も設けている. 表 5-1 にはコロケーション欄の語句も含めた.
- 19 通例 100 語まで抽出できるが, sich ereignen の場合 89 語までであった.
- 20 87 語の累計頻度 14,710 中では, 16.41% を占める.
- 21 logDice 順で 64 位, 頻度順 73 位.
- 22 共通しない 2 語は Arbeitsunfälle と was.
- 23 累計頻度 500 中, 21.4% になる.
- 24 DS はコーパスについて言及しているが, どの程度使用したかは明らかにしていない. 他の独辞典はそもそもコーパスについて書いていない.
- 25 使用例の多数が該当しない場合, 排除した. Prozent は採用したが, Jahr は 3 例中 2 例が該当外だったため, 採用しなかった.

## 【参考文献】

- Möhrling, J. (2011) Kollokationen im Lernerwörterbuch - Anspruch und Wirklichkeit. *Linguistik online* 47, 3/2011.
- Reder, A. (2006) *Kollokationen in der Wortschatzarbeit*. Wien.
- Tschirner, E. (2008) Das professionelle Wortschatzminimum im Deutschen als Fremdsprache. *Deutsch als Fremdsprache* 4/2008.
- Zaima, S./ Imamichi, H./Tokita, I./Kang, M. (2011) Korpusbasierte Analyse der syntakto-semantischen Konstituenten-



- verbindungen des Verbs *verbringen*. Japanische Gesellschaft für Germanistik (Hrsg.) *Mapping zwischen Syntax, Prosodie und Informationsstruktur*. München. 103-120.
- 赤野一郎 (2012) 「辞書とコロケーション」 In: 堀正広 (編) (2012) 『これからのコロケーション研究』 ひつじ書房. 23-60.
- 赤野一郎他 (編) (2015) 『最新英語学・言語学用語辞典』 開拓社.
- 井口 靖 (2017) 「コーパスに基づく多義語の分析 ―日本語「読む」とドイツ語 *lesen* を例として―」 In: 『三重大学教養教育機構研究紀要』 第 2 号. 53-62.
- スタッフズ、M. (南出康世・石川慎一郎監訳) (2006) 『コーパス語彙意味論 ―語から句へ』 研究社.
- 田畑智司 (2012) 「文体とコロケーション」 In: 堀正広 (編) (2012) 『これからのコロケーション研究』 ひつじ書房. 107-152.
- 仁田義雄 (2002) 『副詞的表現の諸相』 (新日本文法選書 3) くろしお出版.
- 堀 正広 (編) (2012) 『これからのコロケーション研究』 ひつじ書房.

## 【辞典】

- Brockhaus-Wahrig (1980-1984) *Deutsches Wörterbuch*. Wiesbaden/Stuttgart.
- Buhofer, A. H. et al. (2014) *Feste Wortverbindungen des Deutschen: Kollokationenwörterbuch für den Alltag*. Tübingen.
- Davis, M./Gardner, D. (2010) *A Frequency Dictionary of Contemporary American English*. London / New York.
- Duden (1993-1995) *Das große Wörterbuch der deutschen Sprache*. 2. Aufl. Mannheim.
- Duden (2010) *Stilwörterbuch*. 9. Aufl. Mannheim. (DS)
- Duden (2013) *Basiswörterbuch Deutsch als Fremdsprache*. Mannheim. (DB)
- Jones, R.L. /Tschirner, E. (2006) *A Frequency Dictionary of German*. London.
- Kempcke, G. (2000) *Wörterbuch Deutsch als Fremdsprache*. Berlin. (GR)
- Klappenbach, R./Steinitz, W. (1969-1977) *Wörterbuch der deutschen Gegenwartssprache*. Berlin.
- Langenscheidt (2015) *Großwörterbuch Deutsch als Fremdsprache*. Berlin. (LG)
- PONS (2015) *Großwörterbuch Deutsch als Fremdsprache*. Stuttgart. (PG)
- Quasthoff, U. (2010) *Wörterbuch der Kollokationen im Deutschen*. Berlin.
- Wahrig (2008) *Großwörterbuch Deutsch als Fremdsprache*. Gütersloh/München. (WG)
- 小野寺和夫他 (2005) 『プログレッシブ独和辞典』 第 2 版 小学館. (PR)
- 在間 進他 (2010) 『アクセス独和辞典』 第 3 版 三修社. (AC)
- 在間 進他 (2011) 『新キャンパス独和辞典』 郁文堂. (CA)
- 根本道也他 (2010) 『アポロン独和辞典』 第 3 版 同学社. (AP)
- 濱川祥枝他 (2014) 『クラウン独和辞典』 第 5 版 三省堂. (CR)

## **Analyse der Kollokationen im Deutschen und Probleme bei ihrer Beschreibung im Wörterbuch**

Yasushi INOKUCHI, Motoyuki TSUNEKAWA, Kastufumi NARITA  
Kiyoshi KURODA und Minkyong KANG

### **Zusammenfassung**

MuttersprachlerInnen behalten keine einzelnen Wörter, sondern zusammenhängende Einheiten als Kollokationen im Gedächtnis und verwenden sie im Sprachgebrauch intuitiv. Wenn diese Kollokationen im (deutsch-japanischen) Wörterbuch beschrieben werden, ist das für Deutschlernende sehr nützlich. Die vorliegende Arbeit zeigt zunächst methodische Probleme bei der Analyse von Kollokationen auf. Anschließend beschäftigt sie sich damit, mit Hilfe von Korpora Kollokationen zu extrahieren und die semantischen Beziehungen zwischen den Kollokatoren zu analysieren. Darüber hinaus werden die Ergebnisse der Korpusanalyse mit den Beispielen in den Kollokations- sowie deutsch-japanischen Wörterbüchern verglichen, um die Beispiele im Wörterbuch zu überprüfen und zu verbessern.