

基本語彙と頻度 実践と課題

大藪正彦

1. はじめに

本稿では、日本におけるドイツ語学習またはドイツ語教育・研究において参照できるような、頻度に基づくドイツ語基本語彙リスト(5000語)の作成について報告する。

基本語彙の選定についてはこれまで様々な提案や批判がなされており、日本独文学会でも、「基礎語彙策定の試み」(1993年春)、「ドイツ語基本語彙」(2011年秋)などのシンポジウムが企画されてきた。そのような

場で筆者が耳にする印象的な言葉は、「既存の語彙リストに対して屋上屋を架すようなことを行ってはならない」というものである。しかし、「屋上」と呼べるような語彙リストがそもそも存在しているであろうか。つまり、広く一般の使用に供されており、ごく一般レベルの日本人学習者・教師が気軽に、簡単に、参照できるような基本語彙リストはないのではないか¹。

現在では各種コーパスに基づいた複数の語彙リストが参照可能な状況にある。それらを用いれば、自らコーパスを構築することなく、「高い頻度」を示し、「幅広い使用域」に出現する語彙(図1で言えばA部分)を取り出すことが、個人のレベルでも十分に可能である。本稿ではその実践例を示しながら、同時にそのようにして出来上がった基本語彙リスト(以下では「本リスト」と呼ぶ)の有用性や課題について検討する²。

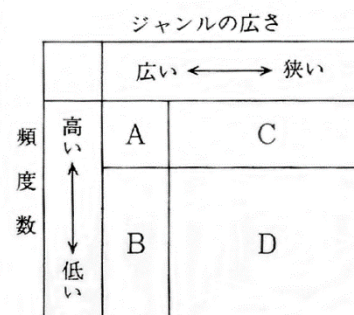


図1 語彙のあり方(千野 1986: 59)

2. 参照した語彙リスト

今回参照した語彙リストは次のものである。なお、それぞれのリストの特徴については山田(2009)に詳しい。また、リストのもとになっているコーパスについては在間(2009)を参照。

DeReWo

- 作成：Institut für Deutsche Sprache (IDS)
 - 参照可能な語彙リスト：derewo-v-30000g-2007-12-31-0.1 (Grundformliste) ほか
 - コーパス：DeReKo (総語数 22 億、新聞を中心とした書き言葉コーパス)
 - 文献：IDS (2007) ほか
 - ウェブサイト：<http://www1.ids-mannheim.de/kl/projekte/methoden/derewo.html>
- 上位 1 万語を参照。以下 **DRW** と略記。

¹ 教育的観点から選定された単語集はもちろん多数存在する。岩崎(2012)に簡単な紹介がある。中には頻度順を採用しているものもあるが、実際は教材などを調査対象とした、「頻出順」とでも言うべきものである。このような状況で Jones/Tschirner(2006)は現代ドイツ語の頻度に基づいたほとんど唯一と言っていい本格的頻度辞典である。本稿ではこの辞典の検証も行いつつ、日本人学習者を念頭に置いた基本語彙リストの作成を目指す。

² 本研究は科学研究費助成事業(基盤研究(C)25370478)の助成を受けている。

deWaC

- 作成：Trento 大学 Baroni 氏ほかによる WaCky プロジェクト
- 参照可能な語彙リスト：deWaC unigrams (lemmas) ほか
- コーパス：deWaC（総語数 17 億，de ドメインを中心としたウェブコーパス）
- 文献：Baroni (2009)
- ウェブサイト：<http://wacky.sslmit.unibo.it/>
- 上位 1 万語を参照。以下 **DWC** と略記。

Jones/Tschirner (2006)

- 頻度辞典：*A Frequency Dictionary of German*. 約 4000 語
- コーパス：Herder/BYU-Korpus（総語数 420 万，均衡コーパス）
- 文献：Tschirner (2005)

Tschirner (2008)

- 上記 Jones/Tschirner (2006) に対する追加語彙リスト。約 1000 語
- 両者を合わせた約 5000 語に加え，Jones/Tschirner (2006) の巻末に挙げられている固有名詞，略語も追加し，約 5140 語を参照。以下両者を含め **J&T** と略記。

Deutscher Wortschatz

- 作成：Leipzig 大学情報学研究所のプロジェクト
- 参照可能な語彙リスト（語形リスト）：100, 1000, 10000
- コーパス：新聞中心の書き言葉コーパス（総語数 3 億）
- 文献：Quasthoff/Wolff (1999)
- ウェブサイト：<http://wortschatz.uni-leipzig.de/html/wliste.html>
- 手作業でレマ化を行った結果の約 6000 語を参照。以下 **DtW** と略記。

このうち，DRW と DWC の二つは 10 億語単位のコーパスに基づいており，文字通り大規模コーパスに基づく語彙リストである。山田 (2009) によれば，四つのリストのうち，DRW と DtW は書き言葉的傾向を示すのに対し，DWC と J&T は話し言葉的傾向を示す。また J&T (2006) は，出版された頻度辞典として唯一厳密なレマ化（見出し語化）がなされている。一方で DtW はレマ化を行っていない語形リストである（残りの二つは機械処理による不完全なレマリスト）。基本語彙選定の際にはもちろんレマ化されているリストの方が有用であるが，後述の通り，見出し語の選定には様々な問題が含まれるため，場合によって語形の頻度を参照できるのは都合がよい³。細かく見ていけば様々な問題があるにせよ，基本語彙の選定を行うにあたり，都合のよい条件がそろったと見なすべきであろう⁴。

³ レマ化の方法は各リストでそれぞれ異なっており，場合によっては求める見出し語の頻度が不明なこともある。例えば DRW では略語などは取り上げられていない。

⁴ コーパスの評価のポイントとして，現代性，代表性，均衡性などが挙げられるが，このうち現代性については大きな問題はないであろう。また，代表性の問題については大規模コーパス二つの存在によって，均衡性については複数のコーパスを参照することによって——もちろん程度問題ではあるが——解消されるものと考えられる。

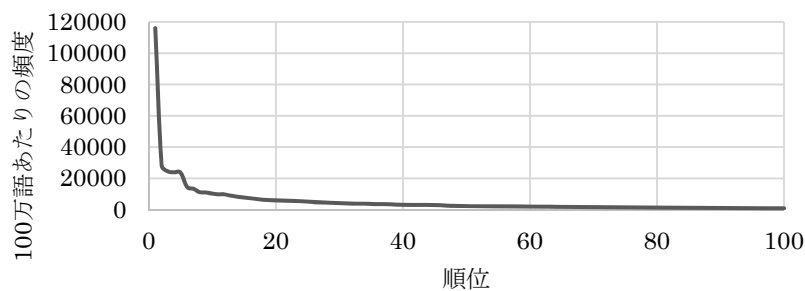
3. 語数・レベルの設定

J&T(2006) は, Nation(2001) らの英語に関する研究をもとに, 4000~5000 語程度を単語学習目標の目安として設定している (Tschirner 2005 も参照)。この数字はテキストの種類に応じて 80~90%程度のカバー率を示すもので, 未知の単語をコンテキストから予測できたり, 間接的に学習できるようになる目安でもあるとされる。本リストでも, 過去の研究を踏まえ 5000 語を選定することとした。その際, 次に示すように, 1000 語ごとに 5 段階のレベルを設定し, 最初のレベルのみさらに下位区分する (上位語は頻度の上で大きな差があるため。グラフ 1 を参照)。

表 1 レベルの設定

レベル	語数	語彙レベル	頻度順位
レベル 1A	500 語	500 語レベル	1~ 500 位
レベル 1B	500 語	1000 語レベル	501~1000 位
レベル 2	1000 語	2000 語レベル	1001~2000 位
レベル 3	1000 語	3000 語レベル	2001~3000 位
レベル 4	1000 語	4000 語レベル	3001~4000 位
レベル 5	1000 語	5000 語レベル	4001~5000 位

グラフ 1 J&T(2006) に基づく上位 100 語の頻度



4. 見出し語の立て方

基本語彙の選定においては, そもそもどの単位を語と見なすべきか, どの語形を基本形と見なすべきかといった問題が常についてまわる。具体的には, 語形 (同音異義語, 大文字・小文字, 単数形・複数形), 変化形・派生形 (現在分詞, 過去分詞, 名詞化, 女性形), 外来語, 固有名詞などの扱いが問題となる。実際のところ, これは何が正しいのかア priori に論じ得る問題ではなく, 個々のケースごとに判断していくしかないという性質の問題である (在間 2012)⁵。

ここではまず, 唯一書籍の形で出版されている J&T(2006) を例に, 見出し語の立て方の難しさを見てみたい。次の表 2 は, Tschirner (2005) に挙げられている Herder/BYU コーパスの語形リスト, レマリスト, および最終的に出版された J&T(2006) の上位 10 語を並べたものである。それぞれ異なっているのが分かる。また, J&T(2006) では派生形・変化形などが準見出し語として扱われている場合があるので, 下の表ではそれらの語を括弧に入れて示している。

⁵ IDS (2007) では DRW のレマ化の方針が紹介されているが, 機械的なレマ化の問題点や限界などを知ることができる。山田 (2009) は DRW や DWC のレマ化の傾向について分析しており興味深い。なお J&T (2006) のレマ化についても, その導入の中で簡単に紹介されている。

表 2 語形リスト・レマリスト・頻度辞典（書籍）

ランク	Herder/BYU		J&T
	語形リスト	レマリスト	頻度辞典（書籍）
1.	die	die	der, die, das
2.	und	der	und
3.	der	und	sein <verb, aux>
4.	in	sein	in (im)
5.	ich	in	ein
6.	den	ein	zu
7.	ist	dem	haben
8.	sie	das	ich
9.	nicht	zu	werden
10.	mit	haben	sie (Sie)

次に、J&T (2006) の見出し語の立て方として問題になりそうな事例を挙げておく。客観的な基準を設けることの困難さが見て取れる。

- (a) 一つの見出し語とされているケース（括弧内の語は見出し語とはされていない）
- 大文字・小文字：deutsch (<Deutsch), Mal (<mal), Morgen (<morgen), sie (<Sie), verhalten (<Verhalten), Wert (<wert), wiedersehen (<Wiedersehen) など
 - 変化形・派生形：Abend (<abends), deutsch (<Deutsche), lang (<lange <adv>), möglich (<möglichst), spät (<später), vorübergehen (<vorübergehend) など
- (b) 一つの見出し語のもとに複数の語形が併記されているケース
- 異なる語形を併記：allein/alleine, bisschen/bissel, daran/dran, gern/gerne, gucken/kucken, herein/rein, nein/nee/nö, nichts/nix, Norden/Nord, selbst/selber など
 - 略語を併記：Abbildung/Abb., Aktiengesellschaft/AG, beziehungsweise/bzw., Kilogramm/kg, Universität/Uni, zirka/ca. など
- (c) 二つの見出し語とされているケース
- 同綴：laut <prep> – laut <adj>, sein <verb> – sein <pron>, verlegen <verb> – verlegen <adj>, weitere(r, s) <adj> – weiter <adv> など
 - 大文字・小文字：bedenken – Bedenken, dunkel – Dunkel⁶, essen – Essen, glauben – Glauben, leben – Leben, vorgehen – Vorgehen, vorhaben – Vorhaben など

本リストの見出し語の選定においては、体系性という学問的な観点よりも、日本人学習者にとっての有用性という実用的・実地的な観点が優先する。長くなるが、原則とした基本方針を以下に列挙する。

見出し語の基本方針

- **大原則**：日本語を母語とする者が、ドイツ語学習またはドイツ語教育・研究において参照できる頻度リストの作成を目指す⁷。

⁶ ただし添えられている例文では *im Dunkeln* という形が用いられている。

⁷ ドイツで刊行されている単語集・辞書などではしばしば普通名詞と固有名詞を区分して扱い、固有名詞が除外されることが多い。しかしヨーロッパの地理に馴染みのない日本人学習者にとって、ドイツ語の地名は普通名詞と同様に学習していく必要がある。例えばそのような地名もリストに含めることをこの原則は意味している。

- 本リストでは文法と語彙の区分を想定する。文法の領域に委ねることができるものとは異なる限り文法に委ねるものとし、語彙項目としては取り上げない（例えば *ich, mir, mich* は語形が大きく異なるが、*ich* のみを見出し語として取り上げる）。動詞であれば不定詞、名詞であれば単数 1 格の形を基本形として見出し語にする⁸。
- 本リストは語形（綴り）を見出し語の規準とする。従って、同音（同綴）異義語も、綴りが同じであれば同一の見出し語とする。ただし大文字・小文字は区分する。
- アラビア数字、ローマ数字、物理単位記号・化学記号・数学記号などの記号、および単純なアルファベットは見出し語としない（ただし略語については下記参照）。
- 略語も見出し語の対象とする。既存の独独辞典を参照する（例えば z.B.）。
- 変化形・派生形（現在分詞・過去分詞・名詞化など）を独立した見出し語と見なすかどうかは、既存の独独辞典を参照して決定する（例えば *bekannt, Bekannte*）。
- 名詞の女性形については、男性形のもとにまとめるものとする。
- 外来語を見出し語と認めるかどうかは、既存の独独辞典を参照して決定する（例えば *Team*）。
- 固有名詞のうち、人名（姓・名）は本リストには含めず、別途リストをまとめるものとする。ただし、《神話》《聖書》関連の人名を見出し語と見なすかどうかは、既存の独独辞典を参照して決定する（例えば *Jesus Christus*）。外国人名は本リストからも別リストからも除外する。
- 固有名詞のうち、国名・都市名・地名・河川名などは、ドイツ語圏（ドイツ・オーストリア・スイス）とドイツ語圏以外を区分して扱う。ドイツ語圏のものについては、原則としてすべて取り上げる。ドイツ語圏以外のものについては、国名をのぞき、原則としてドイツ語特有の特徴（綴り、発音、名詞の姓）を持つものに限り取り上げる。機関名・企業名・商標名など、その他の固有名詞もこれに準じる。
- 地域的バリエーションは、原則として《ドイツ》および《標準語》に限定する。
- 多義語のうち、本リストの方針に従うと本来見出し語にならない語義を含んでいるものであっても、他の語義での使用が十分に認められる場合は、通常通り見出し語扱いとする（例えば普通名詞・固有名詞としての *König*）⁹。
- 正書法は、新正書法（Duden 2006 以降）によるものとし、表記上のバリエーションがあるもの（例えば *Fotografie* と *Photographie*）は Duden の推奨形・非推奨形の記述に従い、推奨形の方を基本形（主見出し語）とする。
- ふつう語尾の付いた形で用いられる語は、一般的な語尾を付した形で見出し語とし、語尾部分はイタリックとする（例えば *einige*）。

⁸ 人称代名詞は性・数ごとに 1 格の形を見出し語とする（*ich, du, er, sie, es* など）。冠詞・冠詞類は性・数・格にかかわらず一つの見出し語のもとにまとめる。このうち定冠詞・不定冠詞はそれぞれ *der/die/das, ein/eine* と表記し、冠詞類は男性 1 格の形で代表させる（例えば *dieser, mein*）。なお、指示代名詞 *der/die/das* は冠詞と同一見出しになる。

⁹ 品詞情報の付与された DWDS の Kernkorpus で調べると、*König* については 90%以上が普通名詞としての使用。一方 *Ernst* では、普通名詞としての使用が約 30%、固有名詞としての使用が約 70%になる。しかしながら、概算上、普通名詞だけをカウントしたとしても十分に 5000 位以内に入ると考えられるので、このような場合は、固有名詞も含めた頻度で見出し語に取り上げる。なお、DWDS については在間 (2009) を参照。

5. 見出し語の処理例

上の基本方針において、語形、変化形、外来語については既存の独独辞典を参照することとした。実際に参照した辞書は次の三点である(それぞれ作業時における最新版を利用)。

- Duden Universalwörterbuch. (以下 **DUW** と略記)
- Langenscheidt Großwörterbuch Deutsch als Fremdsprache. (以下 **LGT** と略記)
- PONS Großwörterbuch Deutsch als Fremdsprache. (以下 **PON** と略記)

本リストでは、これらの辞書すべてに見出し語として挙がっている場合のみ¹⁰、見出し語として採用することにする。以下、具体例をいくつか挙げておく。

次は *folgend* の各リストおよび辞書における記載状況である(数字は順位を示す。DtW は語形リストのためここでは取り上げない)。

表 3 語彙リストの記載状況

見出し語	DRW	DWC	J&T
folgend	—	254	—

表 4 辞書の記載状況

DUW	LGT	PON
×	○	○

語形 *folgend* は DUW において見出し語として扱われていないため、本リストでも見出し語とはせず、動詞 *folgen* のもとにまとめる。なお、DWC では *folgend* が 254 位、*folgen* が 345 位で、*folgend* の方が上位に来ている。このような場合、次のように *folgen* を 254 位に修正して処理する。

表 5 見出し語の処理

見出し語	DRW	DWC	J&T
folgen	274	254	202

次は名詞化の例である。もとの形容詞の順位もあわせて示す。

表 6 語彙リストの記載状況

見出し語	DRW	DWC	J&T
Jugendliche (jugendlich)	— (404)	489 (4808)	932 (4897)

表 7 辞書の記載状況

DUW	LGT	PON
○	○	○

このように並べてみると、DRW では、*der/die Jugendliche* も形容詞 *jugendlich* のもとにまとめてレマ化されたのだと推測できる(複数の語彙リストを参照することによってこのような推測が可能になる)。*Jugendliche* は三つの辞書すべてに記載があるため、本リストにおいて見出し語の候補となる。なお、*Jugendliche* のレベルを確定する際は(本稿 9 節)、DRW をのぞいた他のリストの順位を参照して決定することになる。

次は外来語の例である。

表 8 語彙リストの記載状況

見出し語	DRW	DWC	J&T
Champion	2219	—	3526

表 9 辞書の記載状況

DUW	LGT	PON
○	○	○

¹⁰ ただし、基本形を参照するよう指示があるだけの「変化形見出し語」は考慮しない。

この結果、Champion も見出し語の候補になるわけであるが、実は DtW で最も上位に（そして唯一）登場する Champion の語形は、連語の Champions League である¹¹。IDS のコーパス DeReKo で調査してみたところ、語形 Champion が約 2 万件ヒットするのに対し、Champions は約 8 万件のヒット、さらにこのうち約 6 万件が Champions League としての使用であることが分かった。つまり Champion という語が使用される場合、その約 60% が Champions League としての使用であるということである。次のように処理しておく。

表 10 見出し語の処理

見出し語	DRW	DWC	J&T
Champion (Champions League)	2219	—	3526

正書法も問題となる。まず次のような場合がある。

表 11 語彙リストの記載状況

見出し語	DRW	DWC	J&T
naja	—	3863	384

一語書きの naja という形は、実は旧正書法（Duden 1991）でも、新正書法（Duden 2006 以降）でも認められておらず、na ja と分かち書きすることになっている¹²。本リストでは、たとえ実際に使用が確認できる場合であっても、正書法上認められていない語形は採用しない（正書法の問題については大藪 2005 も参照）。

次は複数の綴りが存在する事例である。

表 12 語彙リストの記載状況

見出し語	DRW	DWC	J&T
potentiell	3811	2426	—
potenziell	6850	—	2633

表 13 見出し語の処理

見出し語	DRW	DWC	J&T
potenziell (potentiell)	3811	2426	2633

旧正書法では potentiell という綴りのみが認められていたが、新正書法で potentiell, potenziell 両方の綴りが認められることになった。Duden は potenziell を推奨しているので、本リストでは上の表 13 のように処理する。

困難なケースとして分かち書き・一語書きの問題がある。

表 14 語彙リストの記載状況

見出し語	DRW	DWC	J&T
mithilfe	—	—	1282

旧正書法では mit Hilfe という分かち書きのみが認められていたが、新正書法では mithilfe (推奨形)、mit Hilfe (非推奨形) の二通りの綴りが認められることになった。DWDS の Kernkorpus 20 (20 世紀の言語資料に基づいたコーパス) で調べてみると、mit Hilfe という語結合は 4662 件ヒットする。他の語と比較してみると、これは本リストのレベル 2 相当であ

¹¹ DtW には、2 語以上から成る連語が見出し語として挙げられている事例が 100 例以上ある。例えば vor allem, nicht nur, zum Beispiel など。

¹² J&T (2006: 3) では新正書法に従ったとされている。

り、J&T のレベルとも合致する。このような場合は、他のリストに挙がっていないからと言って候補から除外するのではなく、見出し語「mithilfe (mit Hilfe)」を立て、最終的に当該のレベルで処理するものとする。慎重な手作業が必要とされる部分である。

6. 選定作業のための準備

DWC は、記号・数字なども含むため、最初にそれらは除外した上で (表 15) 上位 1 万語を参照する。DtW は語形リストであるため、手作業でレマ化を行っておく。ただし頻度は分からないので、上記方針に従い基本形にまとめ、重複を削除していただくだけの作業である (表 16)。その結果 6022 語の見出し語数となった。

表 15 DWC の処理

ランク	オリジナル	処理後
1.	d	d
2.	,	und
3.	.	sein
4.	und	ein
5.	sein	in

表 16 DtW の処理

ランク	オリジナル	処理後
1.	der	der/die/das
2.	die	und
3.	und	in
4.	in	von
5.	den	zu

J&T についても、上記方針を適用すると、最終的に計 5140 語が見出し語として残る。なお、順位はそれぞれのリストに掲載されているオリジナルの順位に従う¹³。また正確な順位の分からない巻末記載の語については、頻度から導き出せる該当ランク・レベルを参照するものとする。参考までに、各リストの上位には次のような語が来る。

表 17 各語彙リストの上位語

ランク	DRW	DWC	J&T	DtW
1.	d-	d	der, die, das	der/die/das
2.	in	und	und	und
3.	und	sein	sein	in
4.	sein	ein	in	von
5.	ein	in	ein	zu
6.	zu	werden	zu	mit
7.	werden	zu	haben	sich
8.	von	von	ich	auf
9.	haben	mit	werden	für
10.	mit	haben	sie	sein

7. 選定作業 (1) ——見出し語候補の選定

本リストでは、大規模コーパスに基づく DRW および DWC の二つのリストを優先的に参照する。まず次の手順で見出し語の候補となる語を選定しておく。

- (1) DRW, DWC から交互に計 7000 語 (各 3500 語) まで見出し語を取り出す。1~7000 位の順位が付いたこのリストを以下「仮リスト (1)」と呼ぶ。

¹³ Tschirner (2008) では追加された 4000 位以降の語の情報しか記載されていないが、部分的に J&T (2006) のリスト (1~4000 位部分) にも変更が加えられているため、Tschirner の選定した 5000 語の全容は結局のところ不明である。

(2) J&T (5140 語), DtW (6022 語) との重複状況を調べ, 上の 7000 語に含まれない語は新たに上げる。その際, 当該の語が DRW, DWC の 1 万位以内に含まれている場合は, その順位も調査する。

この結果, 8223 語が見出し語の候補として選定された。次の表 18 に各リスト間の重複状況を示す。なお, 表 19 に示すように, 最終的に DRW から 7055 語, DWC から 7114 語が取り上げられた。

表 18 各リストの重複状況

DRW	DWC	J&T	DtW	語数	
○	○	○	○	3892	(4 リスト重複: 3892 語)
○	○	○	×	719	
○	○	×	○	1129	(3 リスト重複: 1992 語)
○	×	○	○	45	
×	○	○	○	99	
○	○	×	×	543	
○	×	○	×	100	
○	×	×	○	369	(2 リスト重複: 1448 語)
×	○	○	×	142	
×	○	×	○	285	
×	×	○	○	9	
○	×	×	×	258	
×	○	×	×	305	(重複なし: 891 語)
×	×	○	×	134	
×	×	×	○	194	
計				8223	

表 19 各リストの語数

DRW	7055 語
DWC	7114 語
J&T	5140 語
DtW	6022 語

この表から分かる通り, 3892 語が 4 リストすべてに, 1992 語が 3 リストに含まれる。これらを合わせると 5884 語である。この時点で, この 5884 語を中心に基本語彙選定を行えばよいという見通しが立った。

8. 各リストの比較

この段階で各リストの語の順位を比較することが可能になる。J&T の検証も兼ね, リスト間でずれの見られる例を J&T を中心にいくつか紹介しておく (表の数字は順位を示す)。

表 20 J&T にのみある語の例

見出し語	DRW	DWC	J&T	DtW
aha	—	—	747	—
mithilfe	—	—	1282	—
vorübergehen	—	—	2446	—
vorhin	—	—	2490	—
murmeln	—	—	2548	—
weggehen	—	—	2725	—
Gleichung	—	—	3005	—

表 21 J&T にのみない語の例

見出し語	DRW	DWC	J&T	DtW
anstatt	3753	2950	—	3572
Broschüre	3941	3037	—	3812
Umweltschutz	3587	3403	—	2250
nützen	2457	3997	—	5405
Jubiläum	2285	5141	—	3691
drucken	4893	4171	—	3713
Ankunft	4864	4445	—	3487

表 22 J&T でのみ上位の語の例

見出し語	DRW	DWC	J&T	DtW
halt	2483	1482	219	1902
Paar	1806	2683	284	2435
Morgen	1079	1702	311	1138
irgendwie	2332	1339	366	1982
anfangen	1380	1385	373	3374
wohnen	1126	1209	380	2556
ach	3566	2148	404	3589

表 23 J&T でのみ下位の語の例

見出し語	DRW	DWC	J&T	DtW
Januar	490	479	1301	225
hören	570	413	1557	725
Werbung	1763	1753	2965	1229
kostenlos	1956	1462	3323	2072
bundesweit	1870	1892	4074	1665
anlässlich	2237	2013	4550	2591
Verteidigung	2897	2889	5014	1878

これまでの議論から、ずれの原因が推測できる語も一部含まれるが（本稿 4 節, 5 節を参照）、ここではこれ以上立ち入らない。

9. 選定作業（2）——レベルの確定

以上の作業を踏まえ、4 リストに基づいた基本語彙 5000 語を確定する。原則として DRW, DWC の 2 リストに基づく順位（仮リスト（1））を優先的に参照し、これに J&T の状況を加味しながら調整を行う。DtW は語形リストであることを考慮し、順位は参考程度にとどめる。選定方法の概略は次の通り¹⁴。ただしレベル 1A については、極めて高頻度でかつ極めて広い使用域に現れる語を取り出すという目的から、他のレベルより基準を厳しく設定する（後述）。

選定および順位づけの基本方針

- 本リストは、基本語彙リストとして使用域の広さを重視する¹⁵。そのため、原則として 3 リスト以上に掲載されている語を選定対象とする。ただし、レマ化の事情により特定の語形がリストから漏れている可能性が考えられる場合は、個々の事情に応じて検討し、判断するものとする。
- 少なくとも 2 リストにおいて同じレベルにあるかどうかを順位づけの目安とする¹⁶。ただしリスト間のばらつきが大きい場合は個別に検討する¹⁷。
- 1 リストのみが上位に位置している場合（例えば 4 リストのレベルが 1A-2-2-3 のような場合）、上位の順位は採用しない。下位のリストが近接している場合（2-2-3 あるいは 2-3-3 のような場合）は、その中のより上位のレベルの順位を採用する。
- すべてのリストのレベルが異なる場合（例えば 1A-2-3-5 のような場合）、特に使用域・時代などの問題が考えられるため、DWDS などの調査結果も参照しながら、個々に判断するものとする¹⁸。特に問題なければ、上位のリストを除外した残りのリストの中間レベルの順位を採用する（2-3-5-5 のような場合はレベル 4）。

¹⁴ ここに示す選定方法は、実際には試行錯誤を繰り返しながら決定したものである（例えば注 17 など）。その点において選定には選定者の主観が入らざるを得ない。

¹⁵ 本稿冒頭の図 1 で言うと A~B の部分。A の次は C ではなく B の部分が優先される。

¹⁶ 平均を取らないのは、各リストにおけるレマ化の状況が不明瞭なことがあるため。つまり、不当にレベルが下がることを避けるため。ただし、レベル 1A はのぞく。

¹⁷ 特に DRW, DtW で、新聞特有の語彙が上位に来る場合があるため、DRW, DtW のみ高順位の語については DWC, J&T を優先的に参照することとした。

¹⁸ 特に都市名（とりわけ欧米圏以外の都市名）については、新聞記事を多く含む DRW, DtW で高頻度になっている可能性が高いため、DWDS のデータも活用する。

- 多義語などについては、適宜コーパス調査を実施する。

レベル 1A

- 原則として 4 リストすべてに掲載されており、DRW, DWC, J&T の 3 リストすべてにおいて 1000 位以内に入る語を選定の対象とする。677 語が対象となる。3 リストの順位の平均値を出して順位を決定する。1～500 位までをレベル 1A とする。

レベルの確定について一つだけ例を挙げておく。見出し語 Fußball は次の表に見る通り、4 リストすべてに含まれているが、最も上位の DRW は他のリストと順位が離れている¹⁹。一方で J&T と DtW ではともに同じレベルに位置する。上記の方針に従い、Fußball はレベル 2 (1001～2000 位) の候補になる。

表 24 仮リスト (1) の順位

見出し語	DRW	DWC	J&T	DtW
Fußball (レベル)	447 (1A)	2872 (3)	1497 (2)	1304 (2)

続いて、作業の過程を一部だけ紹介しておく。

レベル 1B

- (1) レベル 1A 候補となった 677 語のうち、レベル 1A から漏れた 177 語を 501～677 位で確定。レベル 1A の処理段階でレベル 1B に回された 1 語を 678 位で処理。
- (2) 679 位以下を「仮リスト (1)」に基づき原則通りに 1000 位まで処理。「仮リスト (1)」の 1044 位まで (DRW 882 位, DWC 849 位まで) がこれに収まる (仮の 679～1000 位)。J&T の 882 位までのところでレベル 1B の候補になり得る語を調査すると 101 語あった。この 101 語と、処理済みの下位 101 語 (先の仮の 679～1000 位のうち、900～1000 位の語) と 3 リストの順位の平均値を比較し、その結果上位になった 101 語を最終的に 900～1000 位として組み入れる。その結果入れ替わったのは 49 語。
- (3) ここに含まれなかった 49 語については、あらためて原則通り処理して行くものとする (1001 位とはしない)。

以下、同様の処理を進めて 5000 語を選定する。次がその作業結果である。

¹⁹ DRW では、ハイフンでつながれた複合語 (例えば Fußball-Bundesliga や Fußball-WM) を各構成語に分解して処理している。そのため Fußball が高頻度になった可能性がある。

表 25 作業結果（一部）

レベル 1A	レベル 1B	レベル 2	レベル 3	レベル 4	レベル 5
ab	Abend	abgeben	Abbau	abbrechen	Aachen
aber	ablehnen	Abgeordnete	abbauen	Abenteuer	Abb.
ähnlich	abschließen	abhängig	Abbildung	Abfall	abbilden
alle	absolut	Ablauf	abends	Abgabe	abdecken
allein	acht	Abschluss	abhalten	abholen	Abfahrt
allerdings	AG	Abschnitt	abhängen	Abitur	abheben
allgemein	Aktion	absehen	Abhängigkeit	Abkommen	ableiten
als	aktiv	Abstand	ablaufen	ablösen	Abrechnung
also	aktuell	Abteilung	ablegen	abschaffen	abreißen
ab	Alter	ach	Ablehnung	Abschaffung	absagen
...
zweit	Zweck	zweimal	zustande	zweitens	Zweig
zwischen	zwölf	zwingen	zustimmen	zwingend	Zyklus

10. 精緻化に向けて——日本人にとっての頻度

以上の作業を通じ、頻度に基づく基本語彙リストがいちおう完成した。ただし、ここでの頻度とは母語話者による使用頻度のことである。日本におけるドイツ語教育への応用を考えた場合、日本人のドイツ語使用という観点を加味する必要がある（大藪 2009）。一例を挙げると、**Japan** という語は日本人がドイツ語を使用する場合は高頻度になると予想されるが、今回のリストではレベル 2（1001～2000 位）に位置している。このような問題を考慮し、あらためて基本語彙リストを調整しようとするならば、独自の調査が必要となってくる。

今回一つの試みとして独自に小規模な語彙調査を行ってみたので、以下に簡単に紹介しておきたい。対象としたのは次の二つのテキストである。

- 小島泰（監修）『ドイツ人が日本人によく聞く 100 の質問（全面改訂版）』三修社. 2009.（**100F** と略記）
- 福本義憲『ドイツ語会話 110 番』旺文社. 1989.（**K110** と略記）

前者は日本の社会・文化などを説明した文章（書き言葉）、後者は日本人を登場人物とした日常会話集（話し言葉）である。また、前者は 100 の、後者は 110 のテーマから成り、両者ともに文脈のある、ある程度まとまったテキストを構成しているため、調査対象として都合がよいと考えた。

各テキストの総語数、基本形の数、および両者を総合して上位に来る語は次の通りである²⁰。なお、二つのテキストは総語数が異なるため、比較に際しては 1000 語あたりの頻度を基準にしている。

²⁰ 単語リストの作成には Tecely（山田善久氏作成 <http://staff.gifu-keizai.ac.jp/~yamada/>）を利用した。ただし、レマ辞書は本リストの基準に従って独自に作成した。分離動詞、同形の変化形などは手作業で処理してある。また、『ドイツ語会話 110 番』は旧正書法で書かれているため、新正書法に直して調査を行った。

表 26 テクストの語数

	総語数	基本形
100F	17077	3365
K110	4139	705
計	21216	3656

表 27 調査結果（上位語）

ランク	見出し語	100F	K110	平均
1.	der/die/das	86.432	49.770	68.101
2.	sein	30.509	36.241	33.375
3.	ich	0.527	60.401	30.464
4.	Sie	0.176	50.012	25.094
5.	in	36.892	7.490	22.191
6.	es	15.752	21.261	18.507
7.	ein/eine	19.968	16.912	18.440
8.	und	26.820	7.490	17.155
9.	ja	3.279	23.677	13.478
10.	zu	13.293	10.872	12.082

(数字は 1000 語あたりの頻度)

基本形の数で計 3656 語のこの語彙リストと先の頻度に基づく 5000 語リストの上位 500 語を比較してみると、重複は 273 語であった。つまり、約半数の 227 語が重ならないことになる。このうち、今回の独自調査のリストにのみ出てくる語を 3 タイプに分けて以下に挙げてみる。

(a) 日本的な語彙

asiatisch, Buddhismus, buddhistisch, Erdbeben, *geisha*, Kaiser, Kimono, *ryokan*, Sake, Japan, Japaner, japanisch, Japanisch, Präfektur, Reis, Teezeremonie, Tempel, Yen ...

(b) 日常会話の語彙

Appetit, Augenblick, bitte, Dank, danke, entschuldigen, Entschuldigung, gern, gerne, hallo, leidtun, lieber, Moment, nein, schade, wann, Wiederhören, Wiedersehen ...

(c) 旅行会話の語彙

Bahnhof, Bier, Café, Deutsch, Englisch, essen, Essen, Fisch, Fleisch, Hauptbahnhof, Kaffee, kaufen, Post, Restaurant, Supermarkt, trinken, verkaufen, Wein, Zug ...

まず (a) は日本社会・文化に関わる語彙である。調査したテキストのうち『ドイツ人が日本人によく聞く 100 の質問』に頻出する語であると想像される。次の (b), (c) のタイプについては厳密な線引きは難しいが、どちらも『ドイツ語会話 110 番』に頻出する語ではないかと想像される。これら (a)~(c) に挙げた語の中には、日本人がドイツ語を使用する上において欠かせない基本的な語も含まれていると思われるが、そのような語が一般的な頻度調査では上位に入っていないということが分かる。

11. おわりに

以上本稿では、既存の四つの語彙リストを利用した、頻度に基づく基本語彙リスト作成の試みについて報告した。明らかになった問題点をまとめておく。

(a) コーパスの代表性と均衡性の問題

既存の頻度辞典である Jones/Tschirner (2006) も、他の語彙リストと比較するとかなりの程度ずれが認められ、絶対的な基準とはなり得ないことが分かった。Jones/Tschirner は手作業を前提とし、比較的小規模の 420 万語のコーパスを利用しているが、それが一因であるのかもしれない。ではコーパスの規模を大きくすれば良いのかということ、もちろんそういう問題でもない。「代表性・均衡性」の問題は常について回る。今回筆者が作成した語彙リストももちろん例外ではない。だからと言って客観的手法を放棄することも一方でナンセンスであろう。頻度分析はより良い語彙学習のための基盤を提供するものである。筆者としては、研究者や教員が、その目的や関心に応じて、独自の語彙リストを作成するためのノウハウを蓄積していくことが重要であると考えている。

(b) 特殊事情への配慮とその比重の問題

日本人にとっての頻度ということ考えた場合、ドイツ語圏のコーパスに基づいた単純な語彙リストだけでは不十分であることも分かった。どのような基準で、どの程度調整するかという点については様々な考え方があろう²¹。この点については今後の課題としたい。

また、高頻度の語と学習上優先すべき語が常に一致するとは限らないということも容易に想像できることである。教育的観点から作成されたドイツ語の語彙集は数多く存在する。また多くの学習独和辞典でも重要語のラベル付けが行われている（今道 2009 を参照）。それらで重要語と見なされている語と今回作成した頻度に基づく語彙リストとの間にどの程度の食い違いがあるのかという点も今後検証し、実用面に応用していく必要があるだろう²²。

最後にこの関連で、基本語彙に含まれない語の重要性についても言及しておきたい。冒頭の千野 (1986) の図で言うと C の部分である。個々の事情に応じて変動のある部分であるが、ジャンルごとの高頻度語の選定も一方で取り組んでいかなければならない課題である²³。

参考文献

- Baroni, Marco et al. (2009): 'The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora'. *Language Resources and Evaluation* 43 (3): 209-226.
- 千野栄一 (1986): 『外国語上達法』岩波書店。
- Duden (1991): *Rechtschreibung der deutschen Sprache*. 20. Aufl. Mannheim: Dudenverlag.
- Duden (2006): *Die deutsche Rechtschreibung*. 24. Aufl. Mannheim: Dudenverlag.
- IDS (2007): 'Benutzerdokumentation zum Produkt Korpusbasierte Wortgrundformenliste DeReWo v-30000g-2007-12-31-0.1. Technical Report IDS-KL-2007-02'. IDS Mannheim.
- 今道晴彦 (2009): 「ドイツ語教育における重要語彙選定の現状と課題」『ドイツ語教育』14, 40-47.
- 岩崎克己 (2012): 「日本の DaF におけるドイツ語基本語彙へのアプローチ」岡村三郎ほか (編)『ドイツ語基本語彙』(日本独文学会研究叢書 88), 45-66.

²¹ 大学英語教育学会が選定した英語基本語リスト、いわゆる「JACET8000」では、BNC (British National Corpus) の語彙リストに依拠しつつも、独自調査を踏まえ、日本人学習者を想定した順位の補正が行われている (JACET 2003 を参照)。

²² ただし、具体的なドイツ語学習の状況・目的・方法は実際には様々である。例えば岩崎 (2012) は、日本の大学の教養教育という枠組みの中での基本語彙について論じている。

²³ 一つの試みとして恒川 (2008) がある。本論集所収の黒田論文も参照。

- JACET (2003): 『大学英語教育学会基本語リスト』 大学英語教育学会.
- Jones, Randall L./Tschirner, Erwin (2006): *A Frequency Dictionary of German*. London: Routledge.
- Nation, I. S. P. (2001): *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- 大藪正彦 (2005): 「新正書法についての批判的コメント」『ドイツ語教育』10, 67-72.
- 大藪正彦 (2009): 「コーパスとドイツ語教育——日本人のための「自然なドイツ語」の提供を目指して」
田中愼 (編) 『コーパスをめぐって』(日本独文学会研究叢書 67), 41-56.
- Quasthoff, Uwe/Wolff, Christian (1999): 'Korpuslinguistik und große einsprachige Wörterbücher'. *Linguistik online* 3, 2/99.
- Tschirner, Erwin (2005): 'Korpora, Häufigkeiten, Wortschatzerwerb'. In: Heine, Antje u.a. (Hg.): *Deutsch als Fremdsprache*. München: Iudicium, 133-149.
- Tschirner, Erwin (2008): 'Das professionelle Wortschatzminimum im Deutschen als Fremdsprache'. *Deutsch als Fremdsprache* 4/2008, 195-208.
- 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」『言語科学』(九州大学言語文化研究院) 43, 107-117.
- 山田善久 (2009): 「語彙使用頻度調査は今どこまで可能になっているか」在間進/時田伊津子 (編) 『大規模コーパスを用いたドイツ語研究』(日本独文学会研究叢書 61), 23-35.
- 在間進 (2009): 「ドイツ語コーパスの現状」『国文学——解釈と鑑賞』1月号, 150-157.
- 在間進 (2012): 「頻度に基づく「基本語彙」リスト作成の問題点と展望」岡村三郎ほか (編) 『ドイツ語基本語彙』(日本独文学会研究叢書 88), 3-12.

今後ホームページ上で情報を発信していく予定である。

<http://homepage3.nifty.com/moz/wortschatz/>