

本書の特徴および使い方

本書に収録した用語の収集は、ライフサイエンス辞書プロジェクトのLSDコーパス（15万件のPubMed論文抄録：3,000万語）における単語の出現回数に基づいて行った。このコーパスに含まれる単語の種類は合計約20万語であったが、3,000万語からなるテキストのうちの90%は約4,700語、85%は約2,600語の範囲で書かれていた。そこで、この頻出単語ランキングの上位3,000語のなかから論文執筆の際に重要である約1,000語を選択して本書の見出し語とした。さらにそれらの関連語などで使い方に注意が必要な単語や、重要な副詞、関係詞などを上位の3,000～5,000語のなかから約300語を選んで追加した。合計約1,300語の見出し語からなる本書は、個々の単語の使い方を示す活用辞典である。使い方にあまり特徴のない単語は使用頻度が高くても割愛してあるのでご了解いただきたい。論文抄録の文章の約86%は上位3,000語の範囲で書かれているので、本書に収録してある単語の使い方の情報を利用し、さらに個々の専門分野のキーワード、冠詞、代名詞、前置詞などを加えれば論文のほとんどの部分を書くことができるであろう。

✿ 本書の特徴

本書には次のような特徴がある。

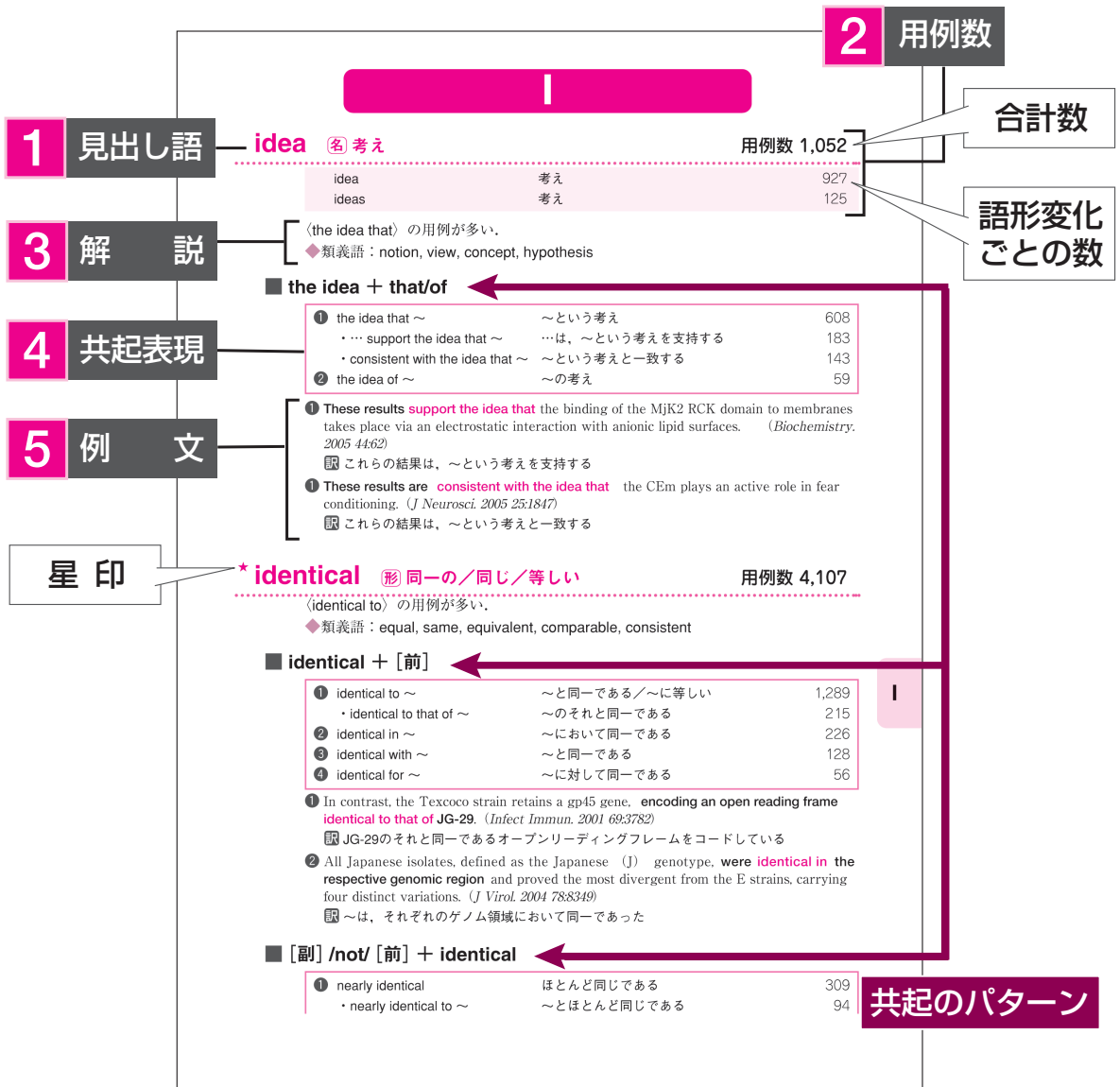
- ①収集したすべての単語とその語形変化ごとの用例数（3,000万語のコーパス中での出現回数）を示した
- ②ある単語の前後にどのような単語がよく用いられるかという共起表現について、その用例数と共に示した
- ③PubMed論文抄録から典型的な例文を引用し、それを日本語訳と共に示した

❖ 本書の構成と活用方法

本書は、以下のように構成されている（下記図参照）。

1 見出し語

見出し語は、論文執筆に重要な約 1,300 語からなる。その品詞と意味とは、論文でよく用いられるものに絞って示してある（一般的な用法での意味は必ずしも考慮されていない）。また、重要度（出現頻度）に応じて、★（星印）を 0 から 2 個付してある。



2 用例数

すべての見出し語に対して用例数の合計を示した。語形変化をもつ単語については、語形ごとの用例数も示している。これによって、生命科学分野の論文でそれぞれの単語がどれぐらいの頻度で用いられるかを知ることができる。頻度の高い単語を優先的に使い、さらに同じ表現の繰り返しを避けるためにやや頻度の低い単語も織りまぜながら論文を執筆するとよいであろう。語形変化ごとの用例数も示してあるので、名詞であればその単語が可算名詞であるか不可算名詞であるか、あるいは複数形で用いられることが多いかどうかなどの情報を得ることができる（1099ページコラム3参照）。また、他動詞であれば、能動態で用いられることが多いのか受動態が多いのか、現在分詞が多いのかなどを知ることができる。

3 解説

見出し語の使い方について、特によく見られる用法や注意すべき点などを示した。また、関連が深くてよく使われる類義語・反意語もできるだけ収録した。

4 共起表現

本書ではすべての見出し語に対して、代表的な共起表現（連語表現）を出現回数と共に収録した。それらを共起のパターンによって分類し、出現回数順に示した。共起のパターンとしては、「過去分詞＋前置詞」「自動詞＋前置詞」「過去分詞＋to不定詞」「副詞＋過去分詞」「名詞＋前置詞」「形容詞＋前置詞」などがある。共起表現の用例数は、示してある連語そのものの出現回数であり、それに含まれる単語の1つでも語形変化した場合の回数は含まれていない。

5 例文

代表的な例文と日本語訳が収録してあるので、実際の論文においてどのように使われるのかを知ることができる。ゴシック体の部分は、下記の日本語訳と対応している。

LSD コーパスの3,000万語の範囲に限っても、共起の組み合わせは非常にたくさんあり、残念ながらそのすべてを限られた紙面に収録することはできない。また、あまり収録する分量が多くても、かえって単語の用法のエッセンスを掴むことができずに逆効果であろう。そこで本書では、論文でよく用いられる個々の単語の使い方をわかりやすくコンパクトにまとめるように心がけた。さらに詳しいことを知りたい場合は、ライフサイエンス辞書プロジェクトのホームページ (<http://lsd-project.jp/>) で公開されている WebLSD の「共起表現検索」を利用すれば調べることができる。

(河本 健)

《本書中の記号について》

○見出し語、および「共起のパターン」中の品詞は、それぞれ以下の略号を用いて示した。

名	：名詞	動	：動詞	形	：形容詞	副	：副詞
前	：前置詞	接	：接続詞	動名	：動名詞	代名	：代名詞
助	：助動詞	現分	：現在分詞	過分	：過去分詞	冠	：冠詞
関	：関係詞	関代	：関係代名詞	関副	：関係副詞	疑副	：疑問副詞
接頭	：接頭辞	接尾	：接尾辞	名句	：名詞句	副句	：副詞句
過去	：(動詞の) 過去形	be	：be 動詞	do	：動詞 (の原形)		

ライフサイエンス英語コーパスについて

✿ コーパスとは

本書のもとになっているのは、ライフサイエンス辞書プロジェクトが独自に構築したライフサイエンス分野の専門英語のコーパスである。コーパスとは言語研究などのために一定の基準に従って収集された言語データのことをいうが、今日では「コンピュータで扱えるように体系化された大量の言語テキスト」すなわち「言葉のデータベース」の意味で使われることが多い。規模の大きな汎用の英語コーパスとしては、British National Corpus (BNC) や Bank of English (Cobuild corpus) がある。これら大規模コーパスのコンピュータ分析をもとにした数量的な視点を取り入れることによって、辞書の編纂方法（見出し語の選択、意味の記載順、例文の選択など）が大きく様変わりしたと言われている。英語辞書の分野では、すでに10年以上前から、COBUILD, LDOCE, OALDなどの主要な学習者用辞書はいずれもコーパスを全面的に取り入れた編纂を行っている。国内でも、『ウイズダム英和辞典』、『ユースプログレッシブ英和辞典』をはじめ最新の英和辞書はいずれもコーパスの活用を前面に打ち出している。

✿ ライフサイエンス英語コーパスについて

ライフサイエンス分野ではPubMedと呼ばれる無料の文献データベースが利用できることから、われわれはPubMedに収録されている学術論文の抄録を主な言語資料として、ライフサイエンス分野に特化した英語コーパスを構築した。生化学、分子生物学などの基礎的な分野から臨床医学などの応用分野に至るまで、ライフサイエンスのさまざまな分野を網羅する主要な89の学術雑誌を選び、2000年から2004年までの5年間にアメリカまたはイギリスの研究機関から出された論文約15万6,000報の抄録を収集した。そこから抽出した約123万件の文章を言語資料とした。すべての文章には、付帯情報としてPubMedの

登録番号をもたせており、分析結果から容易に元の論文抄録を参照できるように工夫されている。

❖ ライフサイエンス辞書とライフサイエンス英語コーパス

本書で利用したライフサイエンス英語コーパスには総語数にして約3,000万語の情報が含まれている^{*}。名詞や動詞の語尾活用を考慮すると、ユニークな単語の数は約20万語と見積もられるが、そのうち出現頻度の高い7万語でコーパス全体の99.2%をカバーしている。7万語という数は、現在のライフサイエンス辞書のサイズにほぼ相当する。こうして構築されたライフサイエンス英語コーパスは、ライフサイエンス辞書のオンライン検索サービス WebLSD (<http://lsd.pharm.kyoto-u.ac.jp/>) において共起検索の形で実装されている。任意の検索語に対してその場でライフサイエンス英語コーパスを検索し、語句の出現頻度を調べたり、前後の隣接語を数量的に捉えることができるようになっている。それだけにとどまらず、英和辞書における見出し語の選択、複合語の抽出、用法・用例の抽出などにもコーパスが活用されており、ライフサイエンス辞書のすべてがこのコーパスをベースにしていると言っても過言ではない。

❖ コーパスをもとに活きた英語を提示

科学論文の執筆を目的とした例文集や活用辞典はこれまでも多数出版されているが、本書がそれらと決定的に異なるのは、コーパスのコンピュータ分析によって得られた数量的なデータを基礎に置いている点である。すなわち、見出し語の選択のみならず、解説する用法や用例の選択にあたっては、頻度情報を最大限に考慮して編纂を行った。これによって、実際の学術論文で好んで使用される「活きた英語」を提示できているものと思う。WebLSDとあわせて、ぜひ論文執筆等に活用していただきたい。

(藤田信之)

^{*}ライフサイエンス英語コーパスはその後にも拡充を続けており、現在オンラインでは総語数6,000万語のコーパスを利用できるようになっている。