

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（PubMed論文抄録からの総語数約1億語）における単語の使用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二十数万語に及んだが、そのうちの2,600語でコーパス全体の約85%が書かれていた。そこで、この頻出単語ランキングの上位3,000語の中から論文執筆の際に重要である約1,100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を上位5,000語の中から約300語を選んで追加した。総計1,430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。

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カバーするであろう。

✿ 本書の特徴

本書には次のような特徴がある。

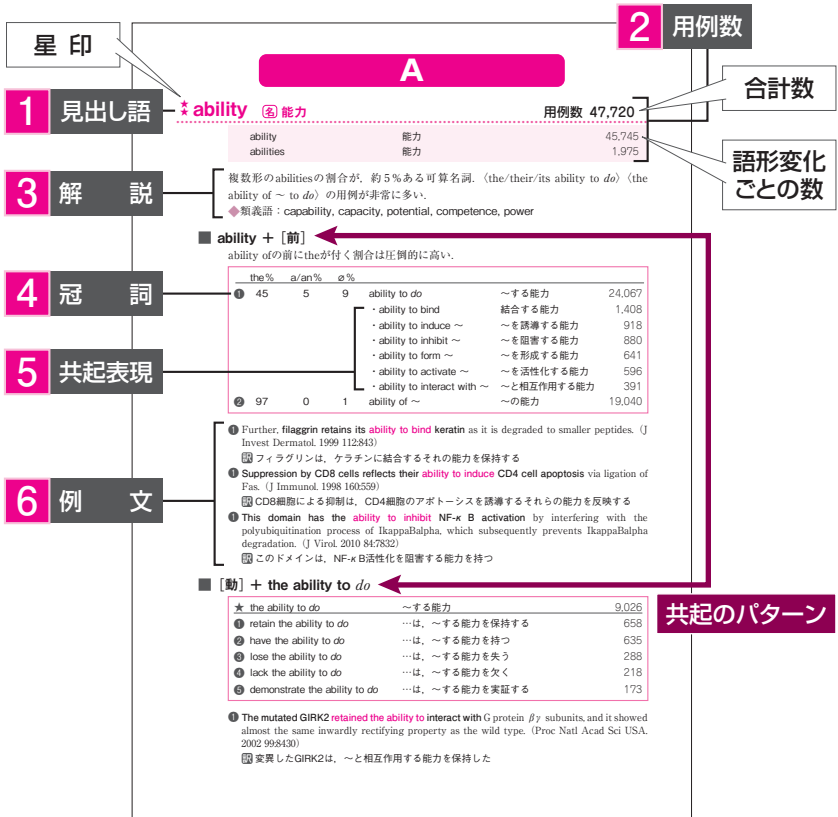
- ・ 収集したすべての単語にその語形変化ごとの用例数（1億語のコーパス中での使用回数）を示した。
- ・ 各単語の前後にどのような単語がよく用いられるかという共起表現を、その用例数と共に示した。
- ・ PubMed論文抄録から典型的な例文を引用し、それを日本語訳と共に示した。

✳ 本書の構成と活用法

本書は、以下のように構成されている。

1 見出し語

見出し語は、論文執筆に重要な1,430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（使用回数）に応じて、星印を付してある。



2 用例数

すべての見出し語に対してその語形変化ごとの用例数と合計数を示した。ただし動詞の ing 形については、名詞としての用例が多い場合は別項目として扱った。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く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を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ためである。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、よく使われる類義語・反意語もできるだけ収録した。

4 冠詞

名詞の見出し語について、「名詞＋前置詞（同格 that 節）」のパターンごとに定冠詞 the が付く割合、不定冠詞 a あるいは an が付く割合、無冠詞の割合を示した。この3つの合計に、代名詞・数字など冠詞と同時に使うことができない単語が付く割合を加えると100%になる。

5 共起表現

本書ではすべての見出し語に対して、代表的な共起表現を使用回数と共に収録した。それらをパターンによって分類し、使用回数順に示した。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」「形容詞＋名詞」などがある。

6 例文

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

✳ 本書の活用法（補足）

今回の改訂で新たに名詞＋前置詞の組み合わせごとの、冠詞の使い分けについて示したので補足しておく。一般に「名詞＋of」の前には、the が付くことが非常に多い。とは言っても、必ずthe が付くという訳でもない。たとえば、aim of の前に定冠詞 the が付く割合は99%である。また、ability of の前に the が付く割合は97%である。つまり、これらの前に通常は the を付けるはずである。もし、the ではなく、不定冠詞 an を付けると判断するなら、その場合には相当地慎重に吟味しなければならない。もちろん、複数の目的があつてそのうちの一つという意味なら、an aim of にするのが適切であろう。the aim of は、その目的が一つであることを意味するからだ。

これほど極端ではないが、abundance of の前に the が付く割合は71%、accuracy of の前に the が付く割合は77%と、これらも非常に高い。したがって、ここでも the を付けることを基本に考えるのがよいであろう。一方、ablation of の前に the が付く割合は7%しかなく、activation of の前に the が付く割合は18%しかない。これらの場合には、前の例とは逆に、the を付けないことを基本に考える方がよいだろう。the が付かない理由は、これらの意味が「～を切除すること」「～を活性化すること」で、何かの性質を表すわけではないからであろう。

では、of 以外の前置詞の場合はどうだろうか。名詞によって事情は異なるが、一般的には the が付く割合がぐっと低くなる。たとえば同じ単語でも、abundance in の前に the が付く割合は5%、accuracy in の前に the が付く割合は12%とかなり低い。したがって、これらの前には the は付けないことを基本に考えるべきである。加えて、不定冠詞が付く割合もそれぞれ0%と5%と低く、無冠詞の割合がそれぞれ65%と61%と高いので、どちらも無冠詞で使うことを基本に考えるべきである。このように、後ろに続く前置詞によっても、状況が大きく変わる点に注意しなければならない。

名詞の使い方を考える場合、冠詞の問題だけでなく、可算・不可

算および単数・複数についても合わせて考えることが必須である。可算・不可算の区別は、複数形の割合で概ね判定できる。複数形の割合が低ければ低いほど、自信を持って不可算名詞であると言える。可算名詞と不可算名詞の境目は、おそらく1~2%ぐらいであるが、それ以外にも様々な問題があるので、コメントを参照していただきたい。例えば、abundanceやaccuracyを無冠詞で使う場合には、文中でこれらを不可算名詞として扱うことになる。しかし、abundanceは、複数形の用例もかなりある(約10%)ので、可算名詞としても使われることがわかる。その場合には、単数形に不定冠詞が付くこともあり得るわけである。同じ単語でも、状況によっては可算・不可算が変わってくることに注意しよう。

最後に、複数形の割合が非常に高い名詞にも注目していただきたい。たとえばabnormalityの場合は、複数形のabnormalitiesの割合が85%と圧倒的に高い。このような単語は、複数形で使うことを基本に考えるのがよいであろう。

以上のように英語の名詞の使い方は、日本語の名詞と比べると非常に複雑である。適切に判断するのが難しいと感じるときには、よく使われるパターンを基本に考えてみるとよいだろう。本書をこのように活用することによって、英語の間違いを減らすことができるはずである。

LSD コーパスの1億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分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「英語共起表現」を利用すれば調べることができる。

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。すでに国内外の多くの英語辞書で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学術誌（約 100 誌）を選び、2000 年から 2012 年までの 13 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の抄録を収集して、1 億語からなるライフサイエンス英語コーパスを作成した。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1億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%をカバーしている。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/>) において英語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、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集を行った。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)