辞書編集者から見た言語資源

―現代日本語書き言葉均衡コーパスについて―

辞書編集者　神永　暁

- 国語辞典における、コーパス活用の可能性
- 現行の辞書のコーパスの活用事例
- コーパスへの要望

国語辞典におけるコーパス活用の可能性
（辞書編集において、コーパスはどのように活用できるか）

1，見出し語選定の参考になる。

2，記述する意味の説明の参考になる。

3，例文の参考になる。

4，その語が使われる場面や文体、あるいは位相についての情報が得られる。

5，表記に揺れのある語の表記をどのように考えたらよいのか、その根拠が得られる。

6，ある語と連接しやすい語の特定が可能になる。

1，見出し語選定の参考になる

- 立項すべきかどうか検討中の語の使用頻度がわかる。
- 一定以上（といってもその基準は編集部で決めるしかないが）、頻度数の高いものは、使用範囲が広く重要語だと考えられるので、見出し語として立項する候補にできる。
- 従来、この項目選定の作業は一部編集者の勘に頼っていた部分だが、コーパスによって客観的な判断材料を得ることができる。

２，記述する意味の説明の参考にできる

- 意味の記述は、従来カードなどを使って集めた用例をもとに、帰納的に記述してきた。
- それだけだと、その語の多様な意味をじゅうぶんにカバーしきれないこともあった。
- コーパスによって、そうした意味用法の広がりを確実に押さえることができる。

２-2，記述する意味の説明の参考にできる

- 電子データであるコーパスは、その語の用例を並べ替えたり、分類したりすることができる。
- これによって多義語の意味記述の順序（ブランチ分けをした際の順序）も使用頻度順に変更することができ、より使いやすい記述内容にすることができる。

３，例文の参考になる

- 国語辞典には、例文と呼ばれるものが２種類ある。
- 典拠を示して引用した実際の使用例（『日本国語大辞典』など）。
- 辞書編集者が、典型的な例文を考えて作り載せたもの（中型・小型の国語辞典）。

３−2，例文の参考になる

- 現時点で、日本の文献を網羅的に収めたコーパスは残念ながら存在しない。
- したがって、現行のコーパスから、用例主義である『日本国語大辞典』のような辞書で引用する用例を探し出すことはほとんど不可能。
- 今後の課題。

３−3，例文の参考になる

- 編集者が作る例文は、コーパスによって適切、かつ典型的な例文を作成することができる。
- コロケーションに関しても、ある語がどういった語と共起しているのか、その実態を読み取ることができ、これも典型的な例文として辞書に載せることができる。

４　その語が使われる場面や文体、あるいは位相についての情報が得られる

- これに関しては、コーパスへの**要望**でもある。
- 俗語、雅語、文章語、口語、隠語、幼児語というきめ細かな情報が欲しい。
- 日本十進分類法（NDC）の情報はもとより、NDCに配された書籍をまんべんなく対象としたコーパスが望まれる。

５　表記に揺れのある語の表記をどのように考えたらよいか、判断のための根拠が得られる

- 「悪事にカタンする」の「カタン」は、「荷担」か「加担」か？
- 外来語の表記、化粧の意の「メイク」「メーク」、どちらが優勢か？
- コーパスの結果はいかに？→後述

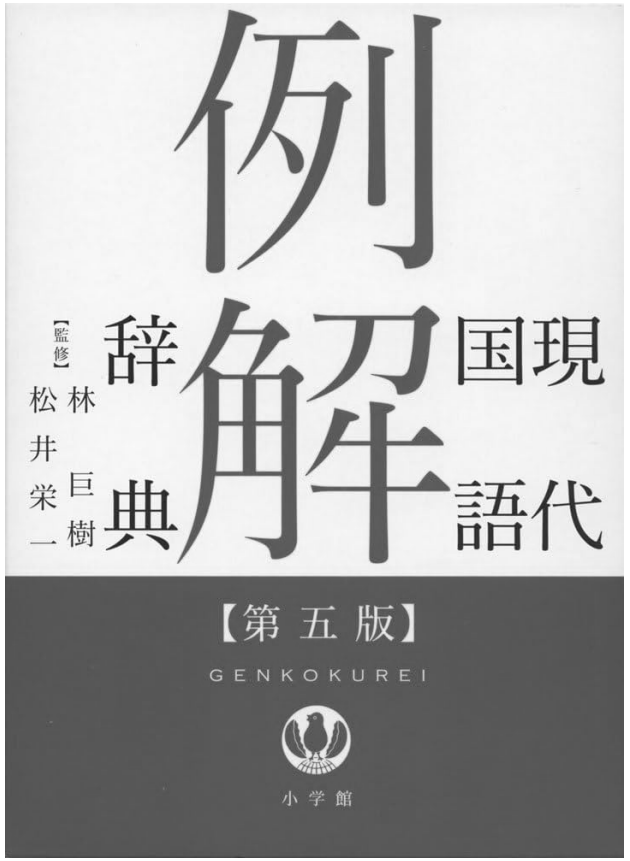
６ ある語と連接しやすい語の特定が可能になる

- たとえばある動詞について、それがどのような名詞や助詞を伴って使われるか、客観的な証拠をえられる。
- 【他動詞「流す」と自動詞「流れる」の場合】

それぞれ助詞を伴った形で、「ヲ流す」の場合どのような語と使われることが多いか、「ガ流れる」の場合ではどうかなど。

- このような情報は、従来の国語辞典ではあまり詳しく記述してこなかった。。

現行の辞典のコーパス活用事例



『現代国語例解辞典』第５版

（小学館）

1985年初版　2016年第5版

『現代国語例解辞典』第5版ではどのようにコーパスを活用したか

『**現代国語例解辞典**』凡例より

- 「見出し語形」「表記欄」「異形」「補注」などにおいて『日本語書き言葉均衡コーパス』を参考。
- コーパスをもとにしたコラムを250ほど掲載。

『現代国語例解辞典』における具体的な活用事例【見出し語形】

「アイデア」か「アイディア」か？

「メイク」か「メーク」か？

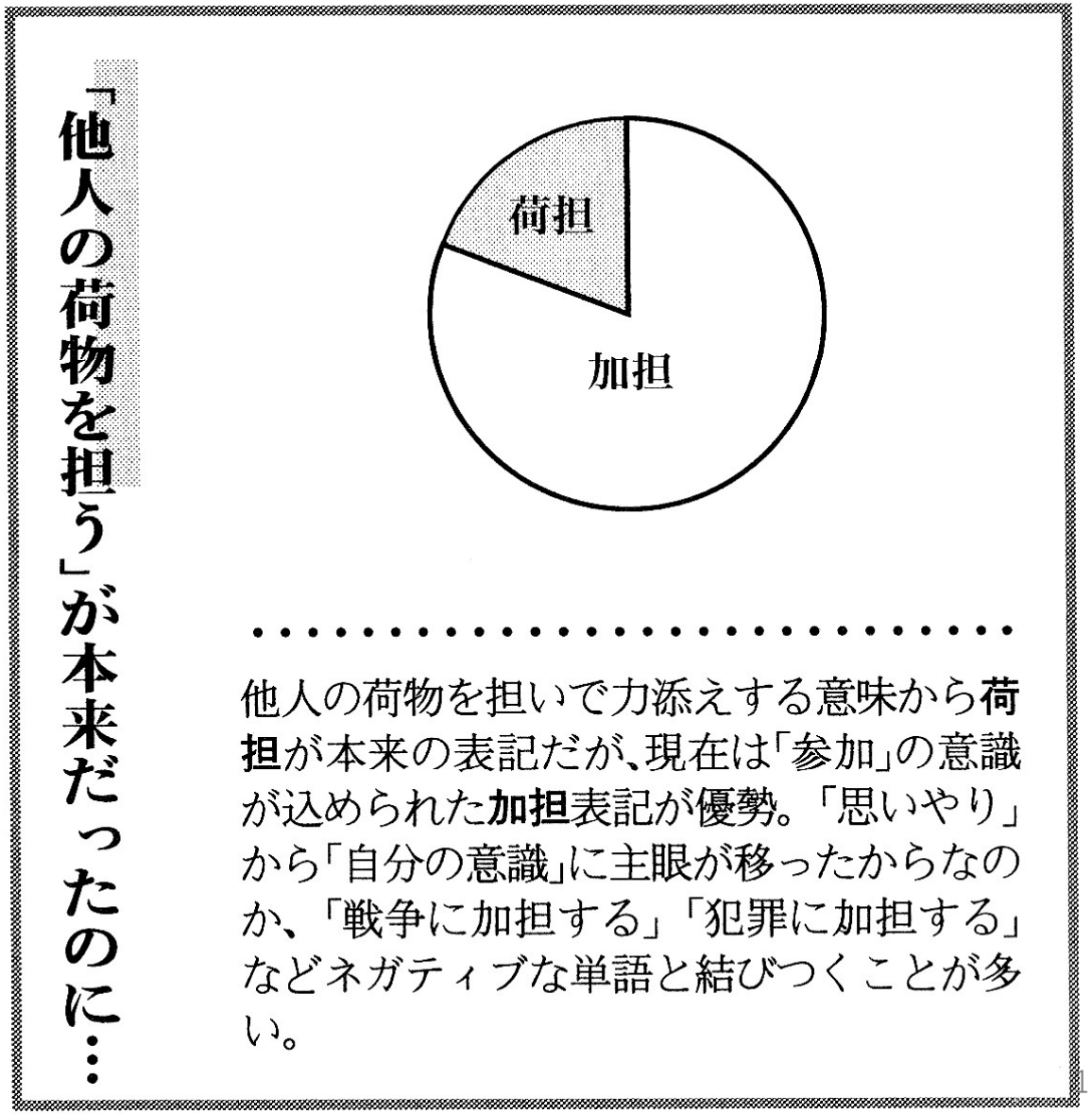
- 『現代国語例解辞典』では前の版（第４版）までは、「アイディア」「メーク」が本項。
- 『日本語書き言葉均衡コーパス』では、「アイディア」「メーク」は少数派。
- 第5版では「アイデア」「メイク」を本項にするように改めた。

『現代国語例解辞典』における具体的な活用事例【漢字表記】

**【「カタン」は
「荷担」か「加担」か？】**

〈他人の荷物を担う意から、「荷担」が本来の表記だが、現在は参加の意識から「加担」を用いることが多い〉

（『現代国語例解辞典』補注、右は同書コラム）



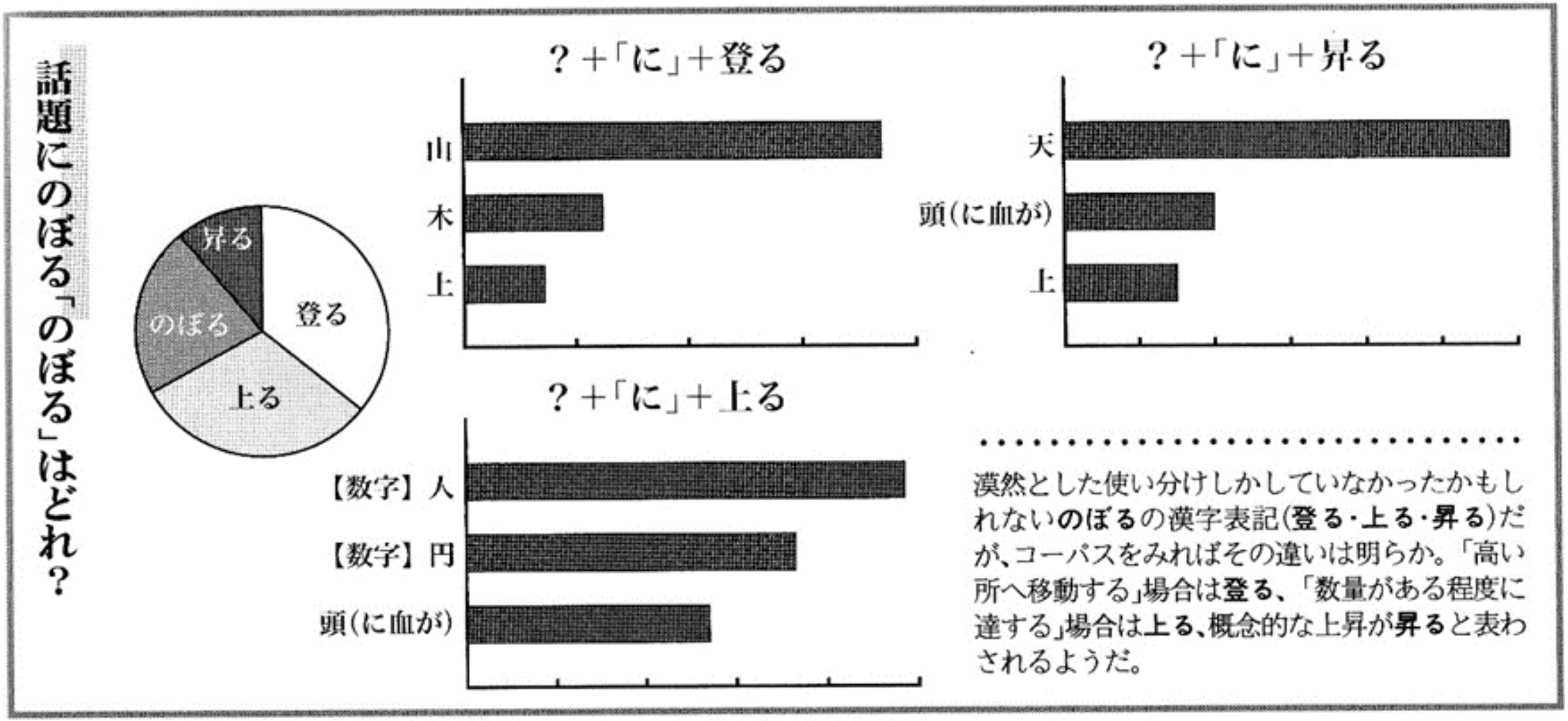
〔余談〕　漢字の書き分けに関しては、和語の異字同訓に関する情報を得る際に威力を発揮しそう！

【「のぼる」の漢字表記】

- 「のぼる」と書く漢字は、

  「上る」「登る」「昇る」の3つ。

- 2014年に発表された文化審議会国語分科会による「『異字同訓』の漢字の使い分け例（報告）」にもこの3つの表記の書き分けについて触れられている。


話題にのぼる「のぼる」はどれ？
昇る
登る
のぼる
上る
？＋「に」＋登る
山
木
上
？＋「に」＋昇る
天
頭（に血が）
上
？＋「に」＋上る
【数字】人
【数字】円
頭（に血が）
漠然とした使い分けしかしていなかったかもしれないのぼるの漢字表記（登る・上る・昇る）だが、コーパスをみればその違いは明らか。「高い所へ移動する」場合は登る、「数量がある程度に達する」場合は上る、概念的な上昇が昇ると表わされるようだ。

（『現代国語例解辞典』より）

今後のコーパスへの要望（辞書編集者の立場から）

(1)データとした格納したサンプルは、長めの文章が望ましい。

(2)さまざまなジャンルの書籍のものをまんべんなく収録する。

(3)書きことばに偏らずに、話しことば、またマンガの吹き出しのせりふなども対象にする。

(4)明治以降の書籍、文献を対象にする。

(5)より充実した検索ソフトの開発。

(6)その他、ほしい情報として

いわゆる誤用表現のデータ／分類語彙表の番号とのリンク／固有名詞の区別／詳細な読みの情報／絵文字／外字、画像で表示されたものも含めて