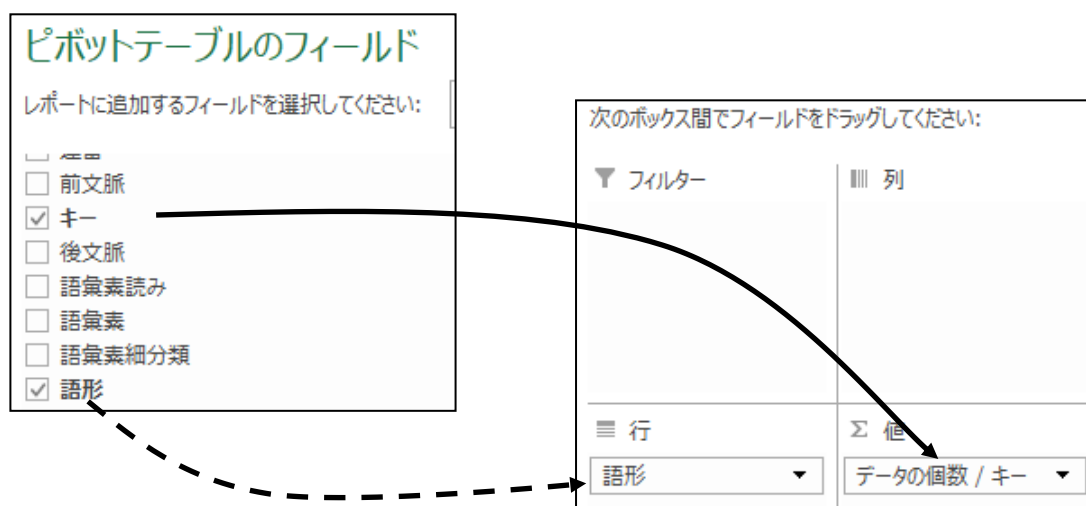


Web 資料 2

ピボットテーブルを使った集計

この資料では、「中納言」でダウンロードした検索結果ファイルを使って語や語形、表記の頻度を集計する際の基本的な手順を説明する。

- (1) 検索結果ファイル (csv ファイル) を開き、メニューから [ファイル] → [名前を付けて保存] を選択する。[ファイルの種類] を「Excel ブック」に指定し、任意の名前を付けて保存する。
- (2) メニューから [挿入] → [ピボットテーブル] を選択する。
- (3) 語形の頻度を集計する場合、以下のとおりフィルターと [行] [値] を指定する。
行:「語形」をドラッグ&ドロップ。
値:「キー」をドラッグ&ドロップ。



※1 第3章の例題1では、新たに付与した媒体という情報を使って、媒体別に頻度を集計する。この場合、[列] に「媒体」をドラッグ&ドロップする。



※2 語の頻度を集計する場合は「語彙素」を、表記の頻度を集計する場合は「書字形」を「行」にドラッグ&ドロップする。
レジスター別に集計する場合は「行」に「レジスター」をドラッグ&ドロップする。なお、表の見やすさを考えて、「行」と「列」とを入れ替えてもよい。

※3 「行」「列」には、複数の項目をドラッグ&ドロップすることができる。以下は、語別の表記頻度表（第6章・例題1）を作成する際の「行」「列」の指定である。

次のボックス間でフィールドをドラッグしてください:

▼ フィルター	■ 列
≡ 行	Σ 値
語彙素_語彙素細分類 ▼	データの個数 / キー ▼
書字形 ▼	

上記のように指定して表を作成した場合、以下の手順で表の体裁を整える（Web資料1の「3.語彙表の作成」参照）。

- 「ピボットテーブルツール」→「デザイン」から以下の設定をする。
 - ①「小計」→「小計を表示しない」を選択。
 - ②「レポートのレイアウト」→「表形式で表示」を選択。
 - ③「レポートのレイアウト」→「アイテムのラベルを全て繰り返す」を選択。

本書・付録1で解説したとおり、BCCWJの短単位解析に用いた形態素解析用辞書 UniDicは、見出しが階層的な構造になっている。この階層的な構造はBCCWJにも反映している。

検索結果のkwicファイルを見ると分かったとおり、語形・書字形は、以下のような情報となっている。

語形：「キー」の読みを表記し、活用語の場合は、それを終止形にしたもの。

書字形：「キー」に示された表記のそのまま、活用語の場合は、それを終止形にしたもの。活用しない語の場合、「キー」と「書字形」は全く同じ情報。

どのような語形が、又はどのような表記が使われているか調査する場合、活用形の違いまで見る必要はない。そのため、活用する語の語形・書字形は、終止形になっているのである。