「少納言」「中納言」検索結果活用ツール

田野村忠温（大阪大学大学院文学研究科）

Corpus Tools for *Shonagon* and *Chunagon*

Tadaharu Tanomura (Osaka University)

１．はじめに

「現代日本語書き言葉均衡コーパス(BCCWJ)」検索サイト「少納言」「中納言」の検索結果を利用するための小ツールbccwj2excelとsortKWICを作成した。いずれも検索結果をソート（並べ替え）してExcelに収めるものである。日本語版Windows上で作動する。

bccwj2excelは、少納言または中納言で画面上に表示された検索結果を処理の対象とする。sortKWICは、中納言で［検索結果をダウンロード］のボタンを押して取得した検索結果およびその他一般の日本語のKWIC索引を対象とする。

両ツールはそれぞれ次のWebページで公開している。もっともこれらのURLをタイプするより、Googleなどで「bccwj excel」（または「bccwj エクセル」）「sortkwic」などと指定して検索したほうが手っ取り早い。

bccwj2excel: http://www.tanomura.com/research/bccwj2excel/
sortKWIC: http://www.tanomura.com/research/sortKWIC/

２ bccwj2excel

２.１ インストール方法

上記の bccwj2excel のページを開いて「bccwj2excel のインストール」のリンクをクリックし、表示される「ファイルのダウンロード」ダイアログで[実行]ボタンを押す（セキュリティの警告には「実行する」や「はい」で応じる）。続いて表示される「bccwj2excel のインストール」のダイアログで[OK]ボタンを押すと、デスクトップに次のような２つのアイコンが作られる。それぞれを通常版、フルデータ版と呼ぶ。

通常版は検索された用例に著者名と書名だけを添えて出力する。フルデータ版は少納言・中納言の提供するすべてのデータ項目を出力する。

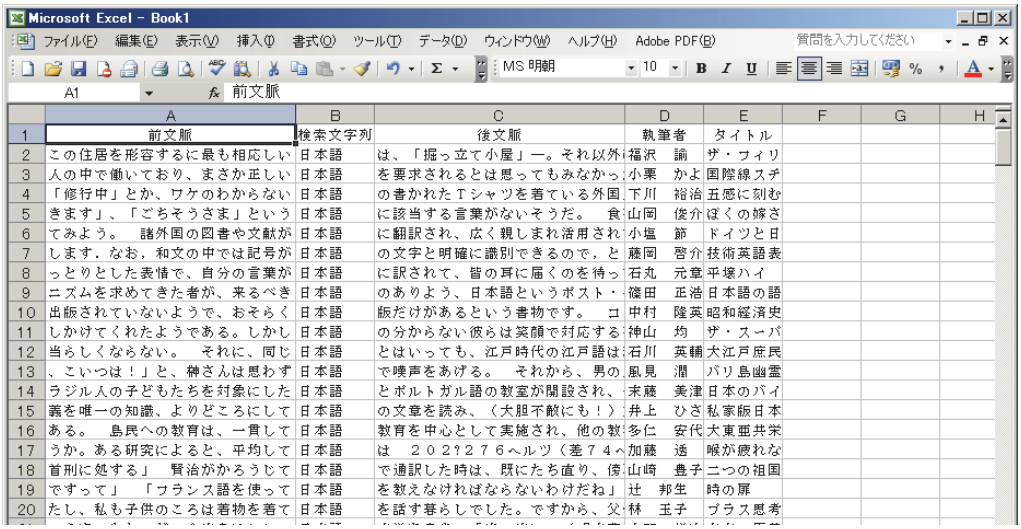
一方だけ使う場合は、使用時に迷わずにすむよう他方を消去するのが便利であろう。アンインストールするにはアイコンをごみ箱に移すだけでよい。

２.２ 用法

少納言または中納言で語句を検索して画面上に表示された検索結果をソートしてエクセルに格納するには次のようにする。

1) Internet Explorer を使って少納言または中納言のサイトで語句を検索
2) 検索結果の画面上で右クリックして「ソースの表示」を選び、表示されたソースを Ctrl+A で全選択して Ctrl+C でコピー
3) bccwj2excel のアイコンをダブルクリック

これによりエクセルの新しいブックが開かれ、各シートにソート済みの検索結果が入力される。必要に応じて列の幅を適宜調整して利用する。前後の文脈は最初各十数文字だけ表示されるが、幅を広げればより広い文脈を見ることができる。



	A	B	C	D	E	F	G	H
1	前文脈	検索文字列	後文脈	執筆者	タイトル			
2	この住居を形容するに最も相応しい	日本語	は、「掘っ立て小屋」一。それ以外	福沢　諭	ザ・フィリ			
3	人の中で働いており、まさか正しい	日本語	を要求されるとは思ってもみなかっ	小栗　かよ	国際線スチ			
4	「修行中」とか、ワケのわからない	日本語	の書かれたＴシャツを着ている外国	下川　裕治	五感に刻む			
5	きます」、「ごちそうさま」という	日本語	に該当する言葉がないそうだ。　食	山岡　俊介	ぼくの嫁さ			
6	てみよう。　諸外国の図書や文献が	日本語	に翻訳され、広く親しまれ活用され	小塩　節	ドイツと日			
7	します．なお，和文の中では記号が	日本語	の文字と明確に識別できるので，と	藤岡　啓介	技術英語表			
8	っとりとした表情で、自分の言葉が	日本語	に訳されて、皆の耳に届くのを待っ	石丸　元章	平壌ハイ			
9	ニズムを求めてきた者が、来るべき	日本語	のありよう、日本語というポスト・	篠田　正浩	日本語の語			
10	出版されていないようで、おそらく	日本語	版だけがあるという書物です。　コ	中村　隆英	昭和経済史			
11	しかけてくれたようである。しかし	日本語	の分からない彼らは笑顔で対応する	神山　均	ザ・スーパ			
12	当らしくならない。　それに、同じ	日本語	とはいっても、江戸時代の江戸語は	石川　英輔	大江戸庶民			
13	、こいつは！」と、神さんは思わず	日本語	で喚声をあげる。　それから、男の	風見　潤	バリ島幽霊			
14	ラジル人の子どもたちを対象にした	日本語	とポルトガル語の教室が開設され、	末藤　美津	日本のバイ			
15	義を唯一の知識、よりどころにして	日本語	の文章を読み、〈大胆不敵にも！〉	井上　ひさ	私家版日本			
16	ある。　島民への教育は、一貫して	日本語	教育を中心として実施され、他の教	多仁　安代	大東亜共栄			
17	うか。ある研究によると、平均して	日本語	は　２０２１２７６ヘルツ（差７４ヘ	加藤　透	喉が疲れな			
18	首刑に処する」　賢治がかろうじて	日本語	で通訳した時は、既にたち直り、傍	山崎　豊子	二つの祖国			
19	ですって」　「フランス語を使って	日本語	を教えなければならないわけだね」	辻　邦生	時の扉			
20	たし、私も子供のころは着物を着て	日本語	を話す暮らしでした。ですから、父	林　玉子	プラス思考			

検索結果は 3 つのモードでソートされ、各シートに収められる。

・モード１：　先行文脈に基づくソート（上図）
・モード２：　検索文字列＋後続文脈に基づくソート
・モード３：　後続文脈に基づくソート

検索文字列が単一の文字列の場合はモード 2 とモード 3 のソートの結果は同一になるので、モード 3 のシートは作成されない。
その他、細かい点を補足すれば以下の通りである。

- 先行文脈は自動的にセルの中で右寄せに配置されるので、列幅の大小にかかわらず途切れなく読むことができる。
- 通常版では書名に副題と巻号を添えて出力する。
- bccwj2excel は処理結果のディスクへの保存は行わない。必要に応じて手動で保存する。
- 実行終了時に表示されるポップアップメッセージは数秒後に自動的に消える。
- コピー後のソースは不要なので、開かれたメモ帳・エディタは閉じる。
- Internet Explorer 以外のブラウザには対応していない（bccwj2excel は動作するがデータを正しく変換できない）。

３　sortKWIC

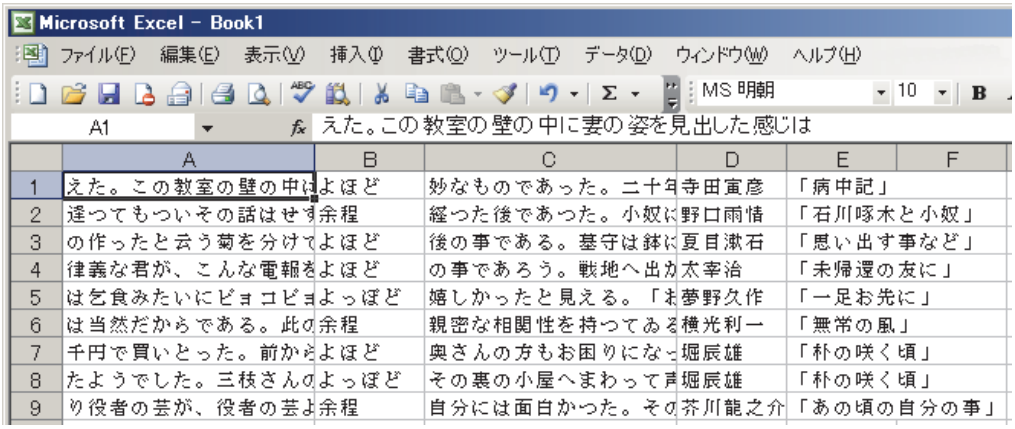
３.１　インストール方法

sortKWIC のインストール方法は 2.1 で説明した bccwj2excel の場合と共通である。sortKWIC のページを開いてインストールする。インストールが終わると、デスクトップに次のような２つのアイコンが作られる。それぞれを通常版、フルデータ版と呼ぶ。

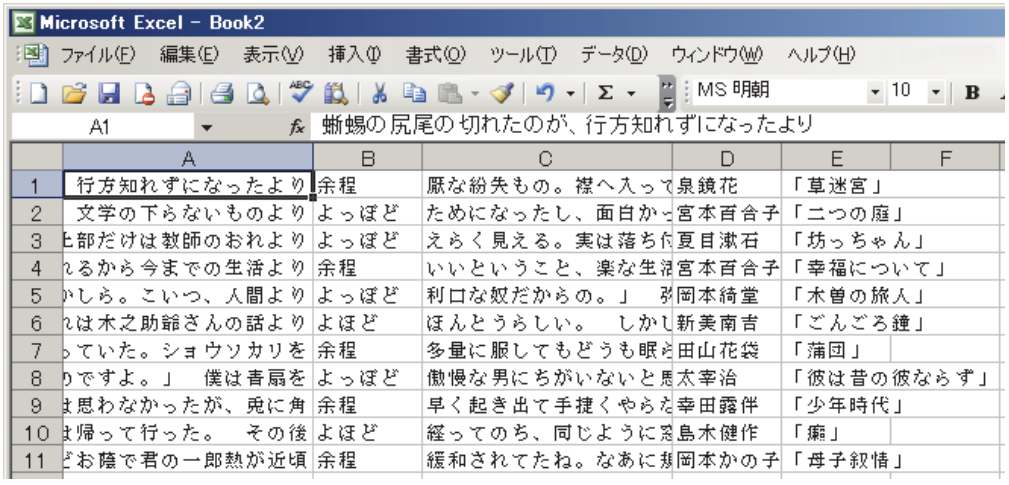
通常版は、中納言の検索結果を処理するとき、用例に著者名と書名だけを添えて出力する。フルデータ版は中納言の提供するすべてのデータ項目を出力する。その他の KWIC 索引を処理するときはどちらを使っても処理結果は同じである。

３.２　用法１──エクセルから

エクセルに次のように KWIC 索引が入っているものとする。



Ctrl+A で全選択して Ctrl+C でコピーし、sortKWIC のアイコンをダブルクリックする。これによりエクセルの新しいブックが開かれ、各シートにソート済みの検索結果が入力される。必要に応じて列の幅を適宜調整して利用する。



bccwj2excel の場合と同じく、検索結果は 3 つ（ないし 2 つ）のモードでソートされ、各シートに収められる。

その他、細かい点を補足すれば以下の通りである。

- 中納言で取得した検索結果を通常版で処理するときは、書名に副題と巻号を添えて出力する。
- sortKWIC は処理結果のディスクへの保存は行わない。必要に応じて手動で保存する。
- 中納言でダウンロードした 3 万件以上の KWIC 索引を処理できることを確認している。ただし、件数の上限はデータや環境に依存する。
- 上の例では先行文脈、検索文字列、後続文脈が A～C 列に入っているが、連続する 3 列ならばどこでもかまわない。

３.３　用法２──タブ区切り形式データ

sortKWIC でソートする KWIC 索引はタブ区切り形式データでありさえすればよく、エクセルに入っている必要はない。

例えば、中納言の［検索結果をダウンロード］を使って検索結果を取得してエクセルに格納するには次のようにする。

1) ［検索結果をダウンロード］ボタンを押し、「ファイルのダウンロード」ダイアログで［開く］を選ぶ
2) メモ帳などで開かれた検索結果全体を Ctrl+A、Ctrl+C によってコピー
3) sortKWIC のアイコンをダブルクリック

コピー後の検索結果は不要なので、開かれたメモ帳類は閉じる。

青空文庫所収の文学作品 3,410 件から語句を用例を検索する「日本語用例検索サイト」（http://www.tokuteicorpus.jp/team/jpling/kwic/）での検索の場合は、「検索結果をダウンロード」にチェックを入れてから［検索］ボタンを押す。あとの手順は上の 1)の後半以下と同じである。

付記

・両ソフトウェアの内容は無保証です。ご自身の責任においてご利用ください。
・両ソフトウェアは日本語版 Windows 上で動作します。動作確認は Windows XP＋Internet Explorer 6＋Excel 2003 で行っています。新しい環境でもおそらく動くと思いますが未確認です。
・両ソフトウェアの作成には Ruby 1.8.7（http://www.ruby-lang.org/）と Exerb 5.3.0（http://exerb.sourceforge.jp/）を使用させていただいています。