英語論文執筆のための
コーパスの使用

Using corpus to publish your research in English

奥村 泰之
(国立精神・神経センター)

於: 日本大学文理学部 (2009/09/19 15:30~)

コーパスとは

- 定義
  - ある目的のために集められたテキスト・データ
  - 機械可読テキストの集積
  - ある言語データを大量に収集し、電子化したもの
  - 言語資料を集めたデータベース

http://www2.kct.ne.jp/~suga/jfirst.htm

http://www.u-gakugei.ac.jp/~tefldpt/extra/corpus/WS10Mar.htm

http://www.med.osaka-u.ac.jp/pub/kid/clinicaljournalclub7.html

英語論文を執筆した感想

- 執筆道具としてのコーパスの重要性
  - 中程度 (執筆に必須ではないが，あると便利)
- 重要な執筆道具 (使用頻度順)
  - STROBE/CONSORT Statements
  - 類似の先行研究
  - 英語論文の執筆関連本
  - コーパス
  - 辞書 (英辞郎，Collins Cobuild-resource Pack on CD-Rom, Merriam-Webster Dictionary)

STROBE Statement

- 観察研究 (コホート研究，症例対照研究，横断研究) の報告のためのガイドライン
- 論文で書くべき「情報」が明確化されている
- 例: 目的の書き方
  - Objectives are the detailed aims of the study. Well crafted objectives specify populations, exposures and outcomes, and parameters that will be estimated.
  - They may be formulated as specific hypotheses or as questions that the study was designed to address.

PLoS Med 4: e296, 2007
PLoS Med 4: e297, 2007

"Our primary objectives were to 1) determine the prevalence of domestic violence among female patients presenting to four community-based, primary care, adult medicine practices that serve patients of diverse socioeconomic background and 2) identify demographic and clinical differences between currently abused patients and patients not currently being abused"

1) populations

2) exposures and outcomes

3) parameters that will be estimated

PLoS Med 4: e296, 2007
PLoS Med 4: e297, 2007

英語論文の執筆関連本

◎ Publication Manual of the American Psychological Association

◎ American Medical Association Manual of Style

◎ How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors, and Reviewers

◎ Displaying your findings: A Practical Guide for Presenting Figures, Posters, and Presentations

◎ Presenting Your Findings: A Practical Guide for Creating Tables

◎ アクセプトされる英語医学論文を書こう!—ワークショップ方式による英語の弱点克服法

○ 科学論文の英語用法百科〈第1編〉よく誤用される単語と表現

○ 英語論文によく使う表現

△ 英語論文表現事典

△ 初めての心理学英語論文

△ Publishing Your Psychology Research: A Guide to Writing for Journals in Psychology and Related Fields

△ 理科系のためのかならず書ける英語論文

△ 理科系のための入門英語論文ライティング

△ 英語論文に使う表現文例集

コーパス使用例

- コーパス使用の目的
  - 表現 (語法，文法，含める情報) を真似するため
- 使用例
  - 信頼性係数についての「結果」の書き方を調べる

検索: coefficient*

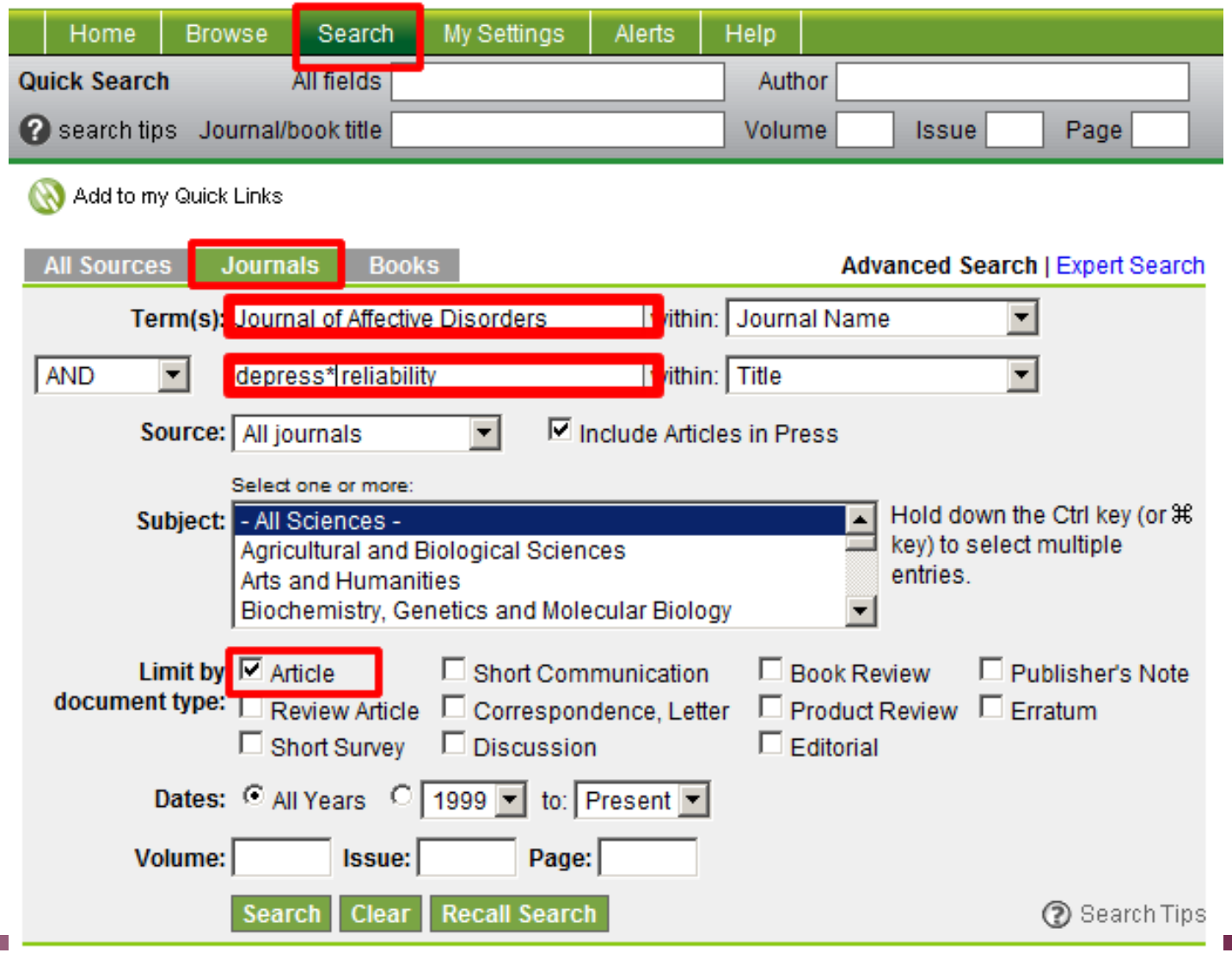
File Edit View Compute Settings Windows Help

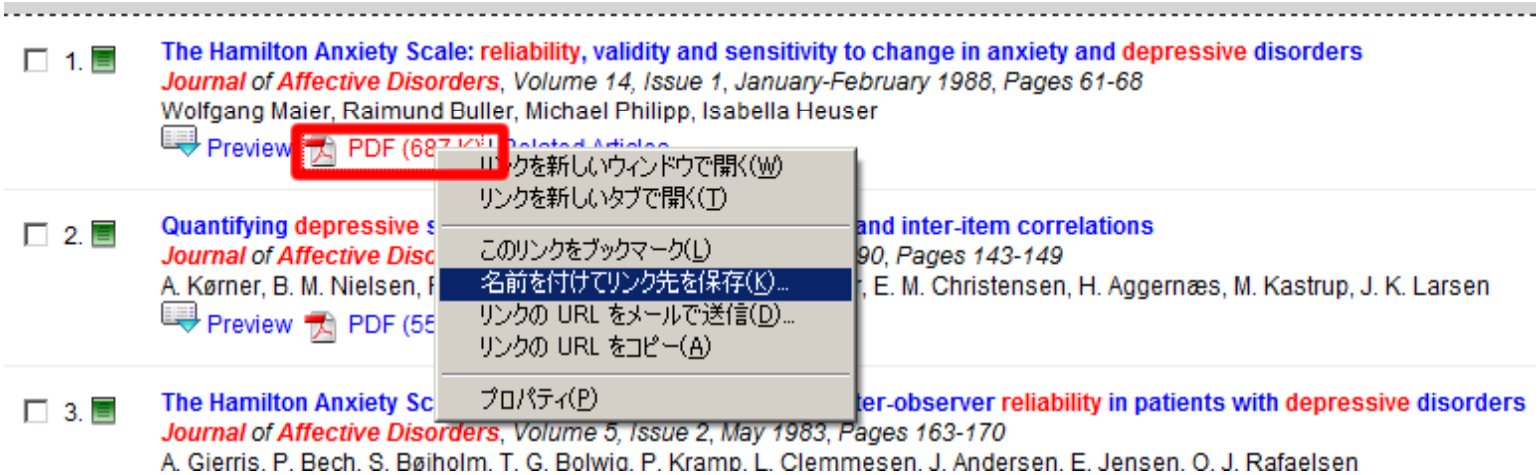
N	Concordance	Set	Ta
1	ritngc Intraclass correlation Maximum coefficient Protocol SIGH-SAD VAS 1		
2	none of the three methods TABLE 2 coefficients SD for measures of diurnal		
3	and calculate the Intraclass Correlation Coefficient (ICC) measure of reliability on		
4	z scores and Intraclass Correlation Coefficients (ICC). Next, we asked which		
5	Z scores and Intraclass Correlation Coefficients demonstrated a poor level of		
6	(DUAG, 1990) the intra-class coefficient of HAM-D,, and MES was		
7	this pattern. Thus, the Loevinger coefficient of homogeneity was above		
8	The HAM-D, produced an intraclass coefficient of 0.79 (P S O.OOI>. Table 7		
9	the HAM-D, produced an intraclass coefficient of 0.83 (P I 0.001). The		
10	of treatment an acceptable Loevinger coefficient (2 0.40) was obtained for the		
11	of treatment. At baseline the Loevinger coefficients were all very low. After 4		
12	scales had also the highest Cronbach coefficients after 4 weeks of treatment.		
13	MDS obtained an acceptable Cronbach coefficient. These scales had also the		
14	the 418 patients available or 88%). The coefficients were lower at baseline where		
15	coefficient alpha and the Loevinger coefficient of homogeneity for the two		
16	validity Table 4 shows the Cronbach coefficient alpha and the Loevinger		
17	the DUAG (1993) study the intra-class coefficients from the joint rating sessions		
18	analysis and Loevinger or Mokken coefficients) are those from the fourth		
19	al. (1995) 51 33 58 55 be measured. A coefficient of 0.30 is just acceptable and		
20	coefficient gives, like the Cronbach coefficient, an overall test for the items of		
21	and Bernstein, 1994). The Loevinger coefficient gives, like the Cronbach		
22	is the Loevinger and Mokken coefficients of homogeneity (Mokken,		
23	unidimensionality of a scale in terms of coefficients is the Loevinger and Mokken		
24	more items in a scale produce higher coefficients. The most adequate		
25	Another disadvantage of using coefficient alpha is that more items in a		
26	scale. However, if all items have a high coefficient they might be redundant.		
27	between the items of a scale. A coefficient of 0.70 or higher is considered		
28	this study we have used the Cronbach coefficient alpha to express the average		
29	validity (Cronbach and Loevinger coefficients as well as factor analysis) all		
30	A note on the concordance correlation coefficient. Biometrics 56, 324-325.		
31	1989. A concordance correlation coefficient to evaluate reproducibility.		

- the use of the Kendall coefficient concordance (W) was also high, W = 0.90, N = 9, P < 0,01.
- Regarding test-retest reliability, Spearman's coefficients for depressive, cyclothymic, hyperthymic and irritable temperaments were 0.79, 0.87..., respectively.
- At baseline only MES and to some extent MDS obtained an acceptable Cronbach coefficient. These scales had also the highest Cronbach coefficients after 4 weeks of treatment.

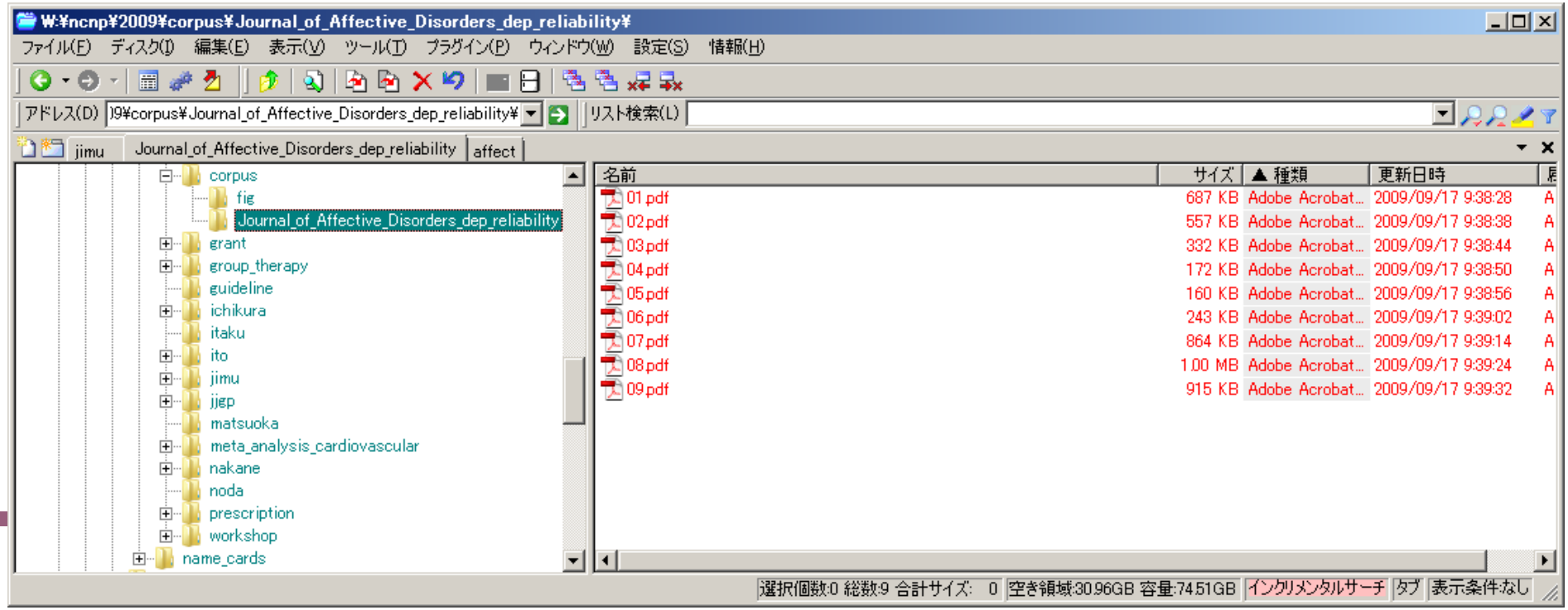
コーパスを作る

- PDFを収集
  - 全文が入手可能な文献データベース: PsycArticle, Elsevier Scienceなど
  - ドキュメント・スキャナの使用
- PDFをTEXTに変換
  - Zilla PDF to TXT Converter/Adobe Acrobat Pro
  - OCR ソフト (e.g., CapturePerfect 3.0)

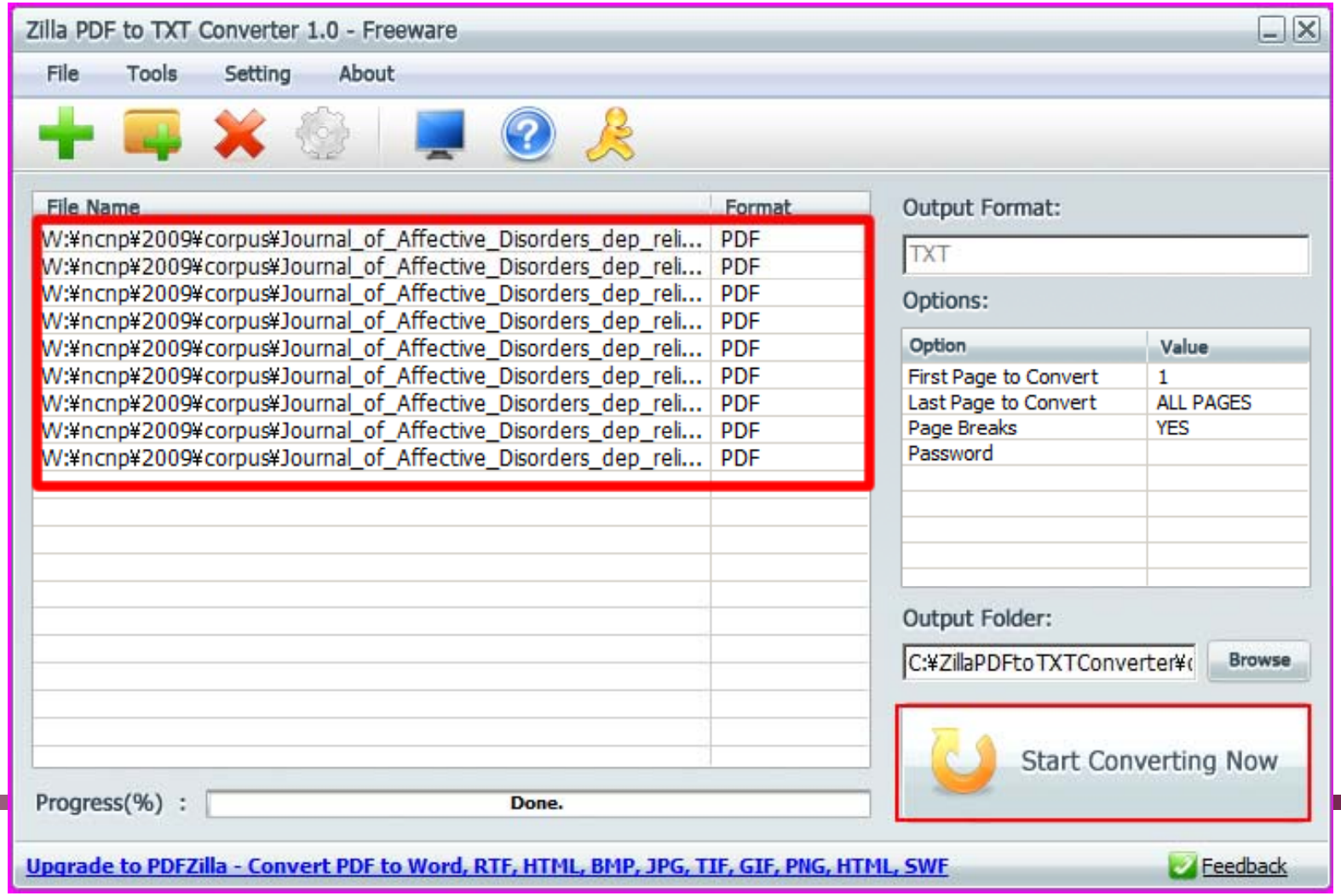

Home
Browse
Search
My Settings
Alerts
Help
Quick Search
All fields
Author
search tips
Journal/book title
Volume
Issue
Page
Add to my Quick Links
All Sources
Journals
Books
Advanced Search | Expert Search
Term(s):
Journal of Affective Disorders
within:
Journal Name
AND
depress*|reliability
within:
Title
Source:
All journals
Include Articles in Press
Select one or more:
Subject:
- All Sciences -
Agricultural and Biological Sciences
Arts and Humanities
Biochemistry, Genetics and Molecular Biology
Hold down the Ctrl key (or ⌘ key) to select multiple entries.
Limit by document type:
Article
Short Communication
Book Review
Publisher's Note
Review Article
Correspondence, Letter
Product Review
Erratum
Short Survey
Discussion
Editorial
Dates:
All Years
1999
to:
Present
Volume:
Issue:
Page:
Search
Clear
Recall Search
Search Tips


1.
The Hamilton Anxiety Scale: reliability, validity and sensitivity to change in anxiety and depressive disorders
Journal of Affective Disorders, Volume 14, Issue 1, January-February 1988, Pages 61-68
Wolfgang Maier, Raimund Buller, Michael Philipp, Isabella Heuser
Preview
PDF (68
リンクを新しいウィンドウで開く(W)
リンクを新しいタブで開く(T)
このリンクをブックマーク(L)
名前を付けてリンク先を保存(K)...
リンクの URL をメールで送信(D)...
リンクの URL をコピー(A)
プロパティ(P)
2.
Quantifying depressive
and inter-item correlations
Journal of Affective Disc
90, Pages 143-149
A. Kørner, B. M. Nielsen,
E. M. Christensen, H. Aggernæs, M. Kastrup, J. K. Larsen
Preview
PDF (55
3.
The Hamilton Anxiety Sc
ter-observer reliability in patients with depressive disorders
Journal of Affective Disorders, Volume 5, Issue 2, May 1983, Pages 163-170
A. Gierris, P. Bech, S. Bøiholm, T. G. Bolwig, P. Kramp, L. Clemmesen, J. Andersen, E. Jensen, O. J. Rafaelsen

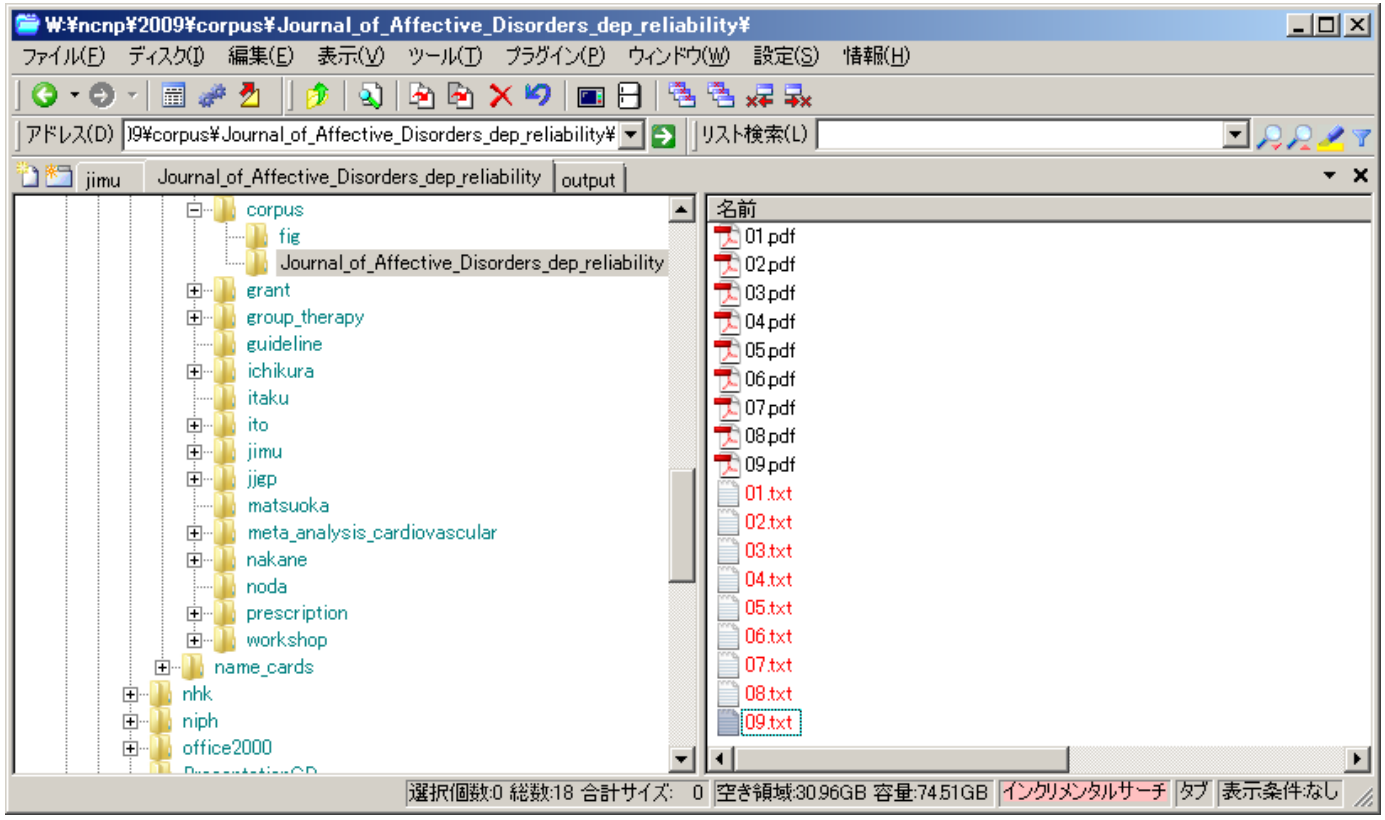
- フォルダ名は，どのような文献を集めたか，一目でわかるように付ける
  Journal_of_Affective_Disorders_dep_reliability



■PDFをTEXTに変換する



■TEXTはPDFファイルと同じ場所に保存



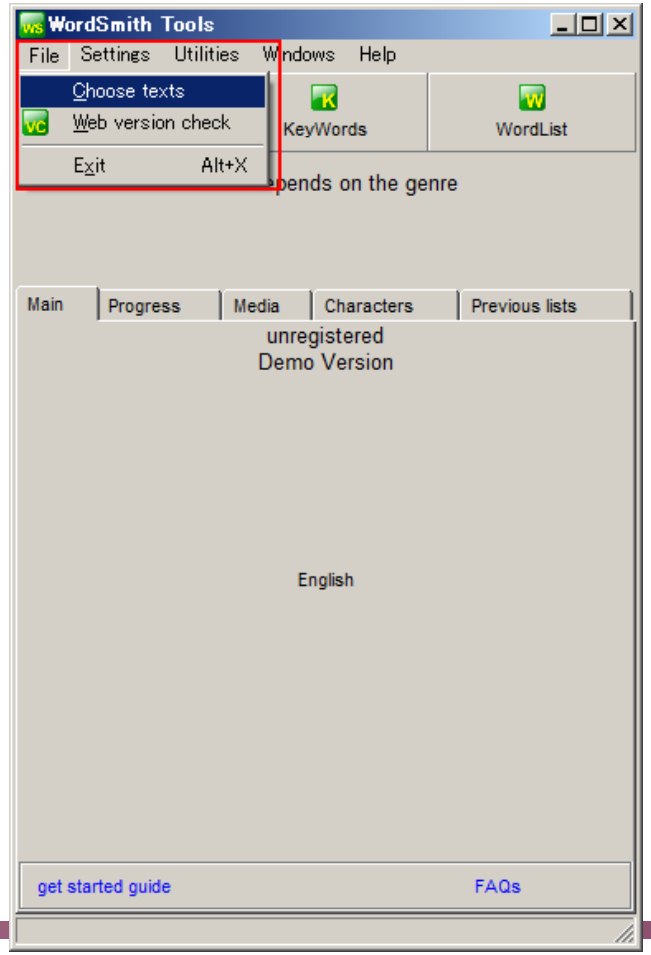
コーパス作成の方向性

論文を書きながら用途に応じて作る

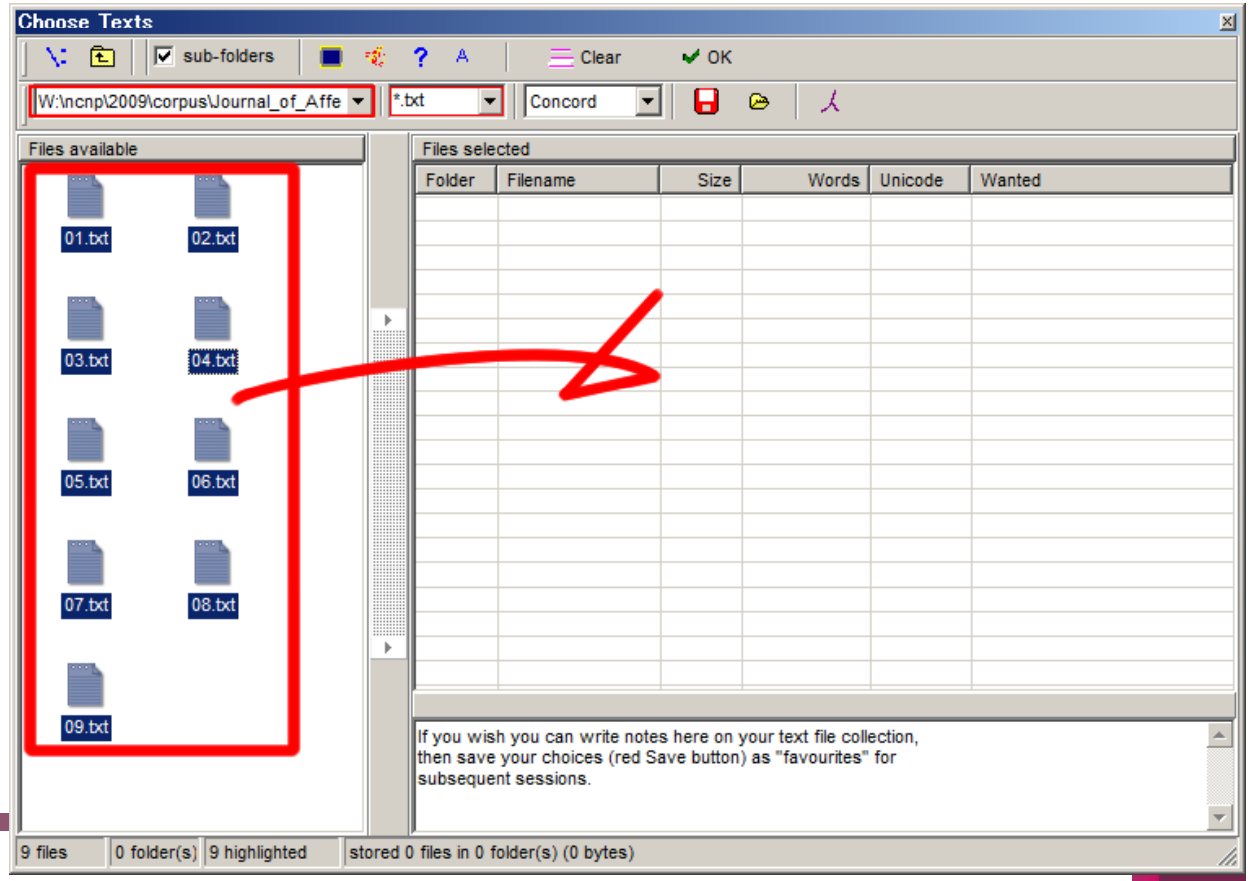
- 研究領域別
  - 抑うつの尺度構成
  - 抑うつの受診行動
- データ解析別
  - 因子分析の使用論文
  - 信頼性係数を求めている論文
- 使用尺度別
  - Beck Depression Inventoryの使用論文
  - HAM-Dの使用論文

WordSmithのセッティング

- WordSmithを起動し，Choose Textsを選択

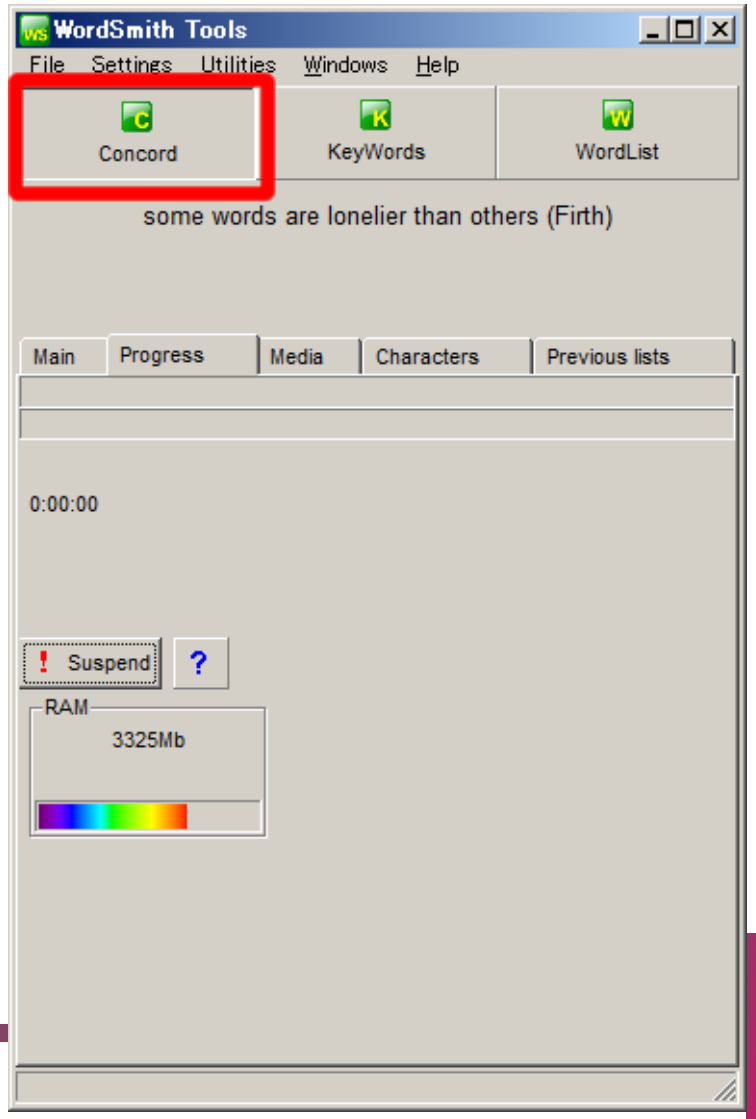


- フォルダ + 拡張子は.txtと指定
- TEXTをドラッグ＆ドロップ。
- 最後に「OK」

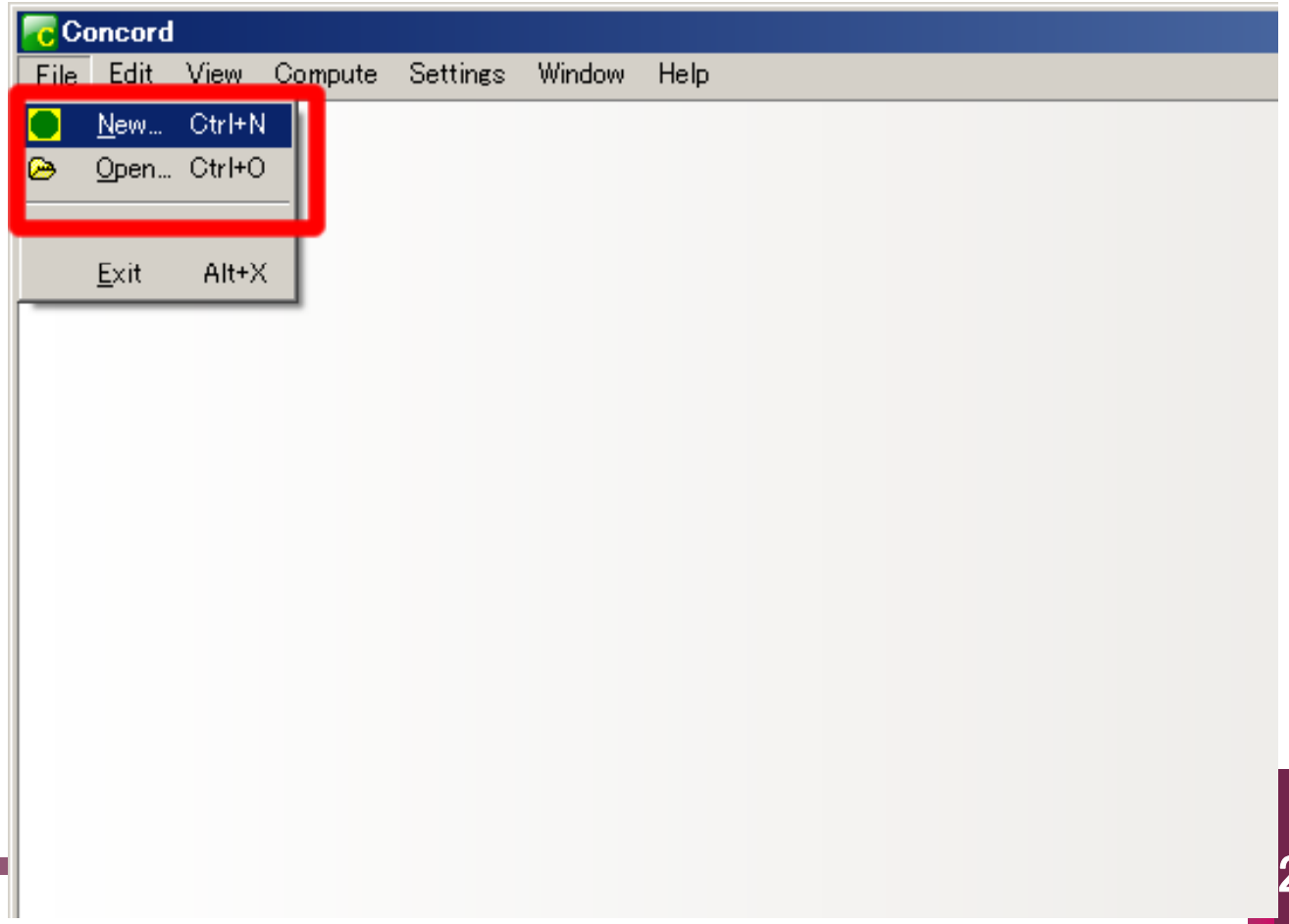


コンコーダンスを使う

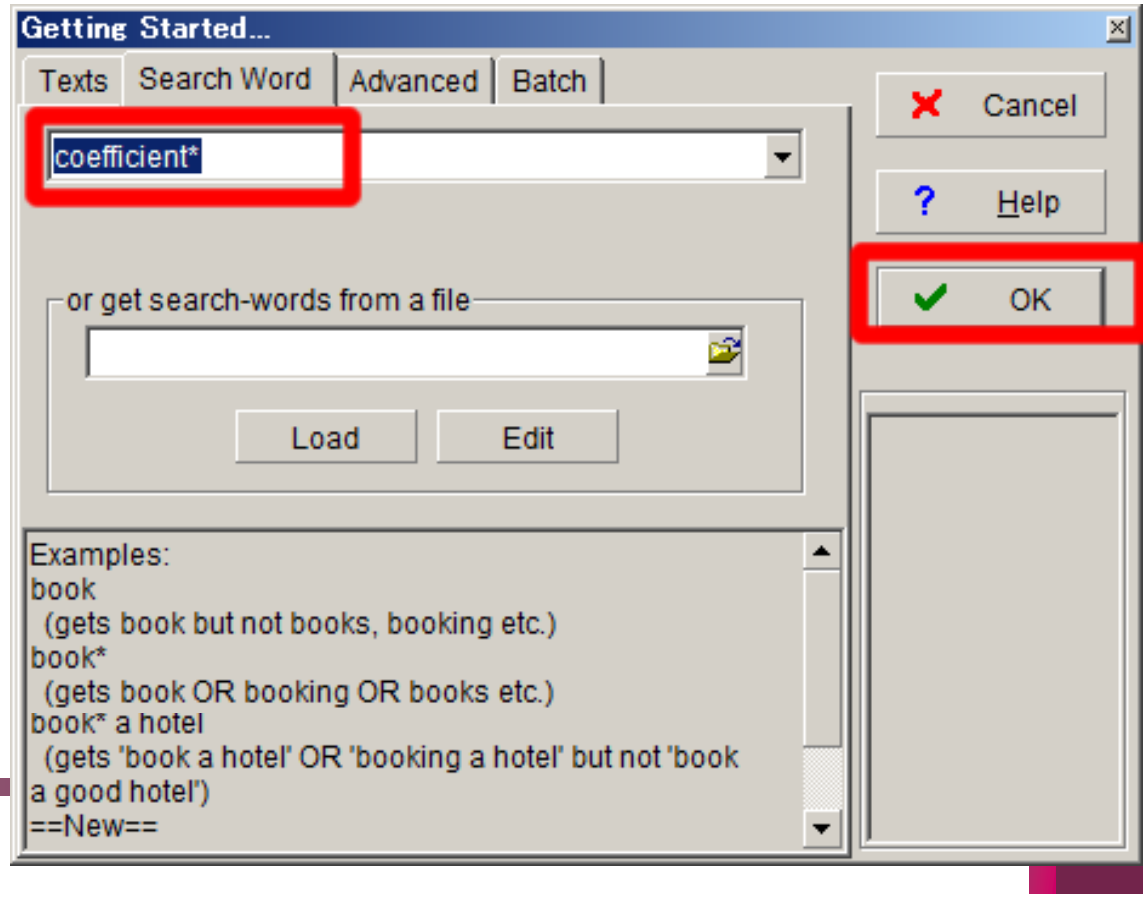
- Concordをクリック



■ File → Newをクリック (Ctrl + N)



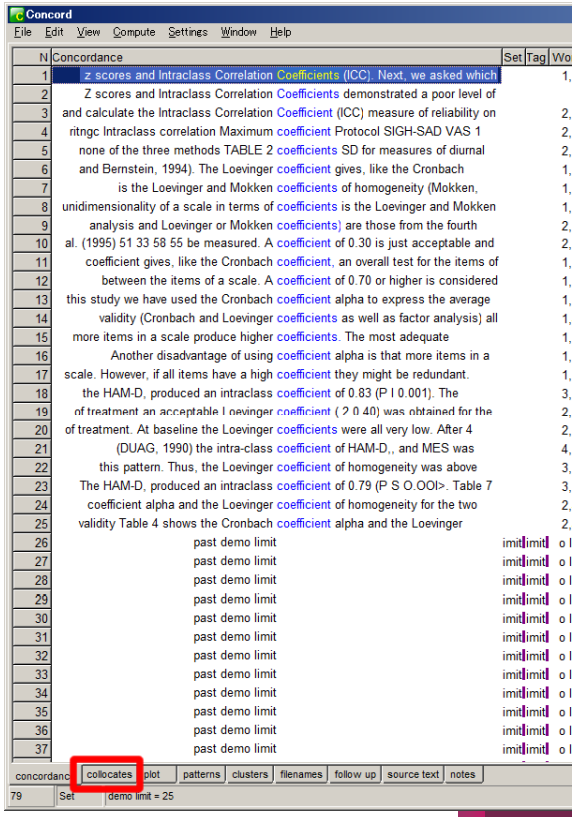
- 「coefficient*」と入力→OK
- coefficient* = (coefficient OR coefficients)



検索式の使い方

検索語	結果
1. book	book, BoOk, ...
2. *book	textbook, book, ...
3. bo* in	book in, books in, booking in, ...
4. book * hotel	book a hotel, book the hotel, ...
5. bo* in*	book in, book into, ...
6. book?	book, books, book., ...
7. ==Book==	Book
8. book/paper	book OR paper

- coefficient*が現れる文例が表示される
- F8を押すと，文例の前後の文字が伸びる
- Ctrl + F8を押すと，元に戻る
- 「collocate」タブをクリックする



- coefficient*と共起する文字が集計される
- キーワードを限定する時などに便利
- 「N:1」「Word: The」「L2:14」
  - 「The * coefficient*」が14件 (e.g., the intra-class coefficients)
- 「N:4」「Word: 0」「R2:10」
  - 「coefficient* * 0」が10件 (e.g., coefficient of 0.30)

Concord

File Edit View Compute Settings Window Help

N	Word	With	elation	Texts	Total	tal Left	al Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	THE	coefficient*	0.000	6	52	32	20	5	2	8	14	3	0	0	7	5	3	5
2	COEFFICIENT	coefficient*	0.000	7	48	4	4	2	0	1	0	1	40	1	0	1	0	2
3	COEFFICIENTS	coefficient*	0.000	8	41	1	1	0	0	1	0	0	39	0	0	1	0	0
4	0	coefficient*	0.000	7	27	1	26	0	0	1	0	0	0	1	10	5	4	6
5	OF	coefficient*	0.000	6	24	5	19	1	0	2	1	1	0	13	0	1	4	1
6	AND	coefficient*	0.000	7	23	12	11	1	2	4	5	0	0	2	2	0	5	2