

Guide to using the BCCWJ 少納言 (Syoonagon) corpus

- Balanced Corpus of Contemporary Written Japanese (BCCWJ) 少納言 (Syoonagon) corpus
[<https://shonagon.ninjal.ac.jp/>] | temporarily unavailable
→ use the Bonten interface [https://bonten.ninjal.ac.jp/bccwj/string_search] for now
- English-language overview [<https://link.springer.com/article/10.1007/s10579-013-9261-0>]
Maekawa, Kikuo et al. (2014). Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation* 48(2): 345-371.

The Balanced Corpus of Contemporary Written Japanese (BCCWJ) was created by The National Institute for Japanese Language and Linguistics (NINJAL; 国立国語研究所). The Syoonagon interface is publicly available and allows full-text string-match searches. (Access to the morphologically parsed version of the corpus requires a license for the Tyuunagon interface.)

Accessing and using the corpus

- (1) **Access the Syoonagon corpus (Bonten interface) here:**
[https://bonten.ninjal.ac.jp/bccwj/string_search]
- (2) Enter your search term in the search box (検索文字列 *kensaku-mozi-retu* ‘search string’)
- (3) Optional: Restrict your search by register (レジスター *rezisutaa*)
 - Default setting is すべて *subete* ‘all’
 - Click the button レジスター設定 *rezisutaa settei* ‘register settings’ to select register(s)
 - Click the button 設定する *settei-suru* to save settings and close box
 - (a) See a translation of the register selection box on the next page of this handout
 - (b) You can also see the list of registers in English here (Maekawa et al. 2014):
<https://link.springer.com/article/10.1007/s10579-013-9261-0/tables/1>
(and for a fuller explanation of what the registers are, see the Maekawa et al. paper)
- (4) Optional: Restrict your search by date range (年代 *nendai*)
 - Default setting is すべて *subete* ‘all’
 - Click the button 年代 設定 *nendai settei* ‘date range settings’ to select date range(s)
 - Click the button 設定する *settei-suru* to save settings and close box
- (5) Click the button 検索 *kensaku* ‘search’ at bottom
- (6) Reset the search options to default by clicking 検索条件を初期化する *kensaku-zyoken-o syokika-suru* ‘initialize search conditions’

(7) Translation of options for selecting register (terms as given in Maekawa et al. 2014)

レジスター設定 "Register" (genre) settings

		Core	Non-core
(Subcorpus Register)		☒ コア	☒ 非コア
出版・新聞	Publication Newspapers	☒	☒
出版・雑誌	Publication Magazines	☒	☒
出版・書籍	Publication Books	☒	☒
図書館・書籍	Library Books	☒	☒
特定目的・白書	Special-purpose White papers		☒
特定目的・ベストセラー	Special-purpose Best-selling books		☒
特定目的・知恵袋	SP Bulletin board 'Yahoo! Chiebukuro'	☒	☒
特定目的・ブログ	SP Blog ('Yahoo! Blog')	☒	☒
特定目的・法律	SP Laws		☒
特定目的・国会会議録	SP Minutes of the National Diet		☒
特定目的・広報誌	SP Local gov't publicity newsletters		☒
特定目的・教科書	SP School textbooks		☒
特定目的・韻文	SP Poetry verses		☒

設定する

Last updated March 2, 2021