



# Appearance frequency of idiomatic phrase in fixed-length tag set of BCCWJ corpus

Tatsuya Kitamura, Kaede Tanijiri, Saya Kanazawa (Konan University, Japan)  
Yoshiko Kawamura (Tokyo International University, Japan)  
E-mail: t-kitamu@konan-u.ac.jp

## Research questions

How much time and effort Japanese language learners should allot to learning idiomatic phrases?  
Which idiomatic phrase should be learned preferentially?

→ We measured the frequency of use of idiomatic phrases for a Japanese text corpus.

## Methods

- ❖ Corpus: Fixed-length (1,000 characters) sample sets of the Balanced Corpus of Contemporary Written Japanese
  - 1,473 newspaper articles
  - 1,996 magazine articles
  - 10,551 books
  - 1,250 white papers
- ❖ Idiomatic phrase list: Over 6,000 phrases collected from Weblio web dictionary (<http://www.weblio.jp/>)
  - Notational variants of the idiomatic phrases were added manually.
- ❖ Find idiomatic phrases and calculate **the appearance frequency**
  - the average number of idiomatic phrases per 1,000 characters

## Results

Table 1: The appearance frequency of idiomatic phrases

Newspaper articles	0.65 phrases
Magazine articles	0.77 phrases
Books	1.12 phrases
White papers	0.09 phrases

Even in books, a mere 1 idiomatic phrase appeared in a 1,000-character sample.

**Learners should learn only highly frequently used idiomatic phrases.**

Table 2: The top 5 idiomatic phrases

Newspaper articles		Magazine articles	
1	気がする (39)	1	気がする (85)
2	手にする (24)	2	身に付ける (44)
3	身に付ける (20)	3	手にする (40)
4	姿を消す (18)	4	目にする (25)
5	汗を流す (16)	4	手に入る (25)

  

Books		White papers	
1	気がする (719)	1	身に付ける (19)
2	身に付ける (241)	2	一つとして (18)
3	手にする (202)	3	頭から (5)
4	自分と (130)	4	言うまでもない (4)
5	役に立つ (128)	5	日が浅い (3)

## Try Idiom finder !

[http://basil.is.konan-u.ac.jp/chuta/idiom\\_finder/](http://basil.is.konan-u.ac.jp/chuta/idiom_finder/)

