

# TSUKUBA WEB CORPUS

NINJAL-LWP for TWC (NLT)



Let's learn the characteristics of words

using a large-scale corpus with over 1.1 billion words!



**Over  
1.1 billion words**

Data book of  
over 1.1 billion words  
collected from websites  
in Japanese



**Comparative search  
between two words**

You can fully understand  
the co-occurrence  
relationships of vocabulary  
by comparing two words  
simultaneously



**Multilingual  
support**

The menus are available in  
Japanese, English,  
and Chinese  
(traditional and simplified)

Please give it a try!



WEB

<https://tsukubawebcorpus.jp/en/>

INQUIRY

[jp-kyoten@un.tsukuba.ac.jp](mailto:jp-kyoten@un.tsukuba.ac.jp)

Joint Development



筑波大学  
University of Tsukuba



日本語・日本事情遠隔教育拠点  
The Center for Distance Learning of Japanese and Japanese Studies

# What is the Tsukuba Web Corpus?

WEB <https://tsukubawebcorpus.jp/en/>

The Tsukuba Web Corpus provided by the University of Tsukuba is a large-scale Japanese corpus that is built with data comprising approximately 1.138 billion words collected from websites in Japanese. You can comprehensively investigate and compare the co-occurrence relationships, frequency of use, and grammatical behavior of vocabulary within the corpus by using **NINJAL-LWP for TWC(NLT)** as your dedicated search tool.

Let's give it a try now!



Easy operation allows you to investigate

without getting your thoughts interrupted.

**優しい** Freq=90,825 \*A search page for illustrative purposes

By Group | Pattern freq order | Basic

» Noun

Pattern	Freq	Ratio
優しい+名詞	37,064	
優しき+名詞	229	
優しい+接尾名詞	12,275	

Noun+particle «

Pattern	Freq	Ratio
…が優しい	3,700	
…は優しい	2,797	
…も優しい	1,455	

**優しい+名詞** 3,738 Types

Collocation	Freq	MI	LD
優しい人	1,674	4.02	4.10
優しい言葉	1,231	6.53	6.50
優しい気持ち	1,181	7.11	7.01
優しい心	882	5.87	5.86
優しいの	710	1.65	1.73
優しいほう	598	3.78	3.83
優しい声	534	6.24	6.11
優しいもの	526	2.69	2.76
優しい笑顔	522	8.21	7.55

**優しい人** 1,674 results

You can also see a dialogue that displays the previous and following sentences!

- やさしい人です。  
(今月の?スポットライト Bitter with the Sweets(ビターウィズスイーツ))
- 優しい人だった。  
(第23回 鳴門教育大学国語教育学芸発表資料 (平成20年))
- 優しい人なので。  
(真面目な年下男性へのアプローチの仕方 | OKWave)
- 優しい人である。  
(裁判傍聴記)

Filtering features allow you to search specific collocation, too!

Search results can be downloaded in XLS and CSV formats!

## Comparative search between two words



**優しい+名詞** 3,733 | **親切な+名詞** 712 \*A search page for illustrative purposes

優しい				親切な				LD Diff
Collocation	Freq	MI	LD	Collocation	Freq	MI	LD	
優しい人	1,674	4.02	4.10	親切な人	390	5.33	2.01	2.09
優しい言葉	1,231	6.53	6.50	親切な言葉	24	4.25	0.93	5.57
優しい気持ち	1,181	7.11	7.01	親切な気持ち	15	4.22	0.88	6.13
優しい心	882	5.87	5.86	親切な心	22	3.95	0.63	5.23

Refer to the user guide to learn in detail how to use the corpus!

The characteristics of words are easily sorted, and the collocations of both words are analyzed in detail!



「優しい人」 appears 1,674 times, while 「親切な人」 appears only 390 times

General trends can be grasped easily.



## Multilingual support

The menus in Japanese, English, and Chinese (traditional and simplified) are available at **NINJAL-LWP for TWC(NLT)** the search tool dedicated to the Tsukuba Web Corpus. Please use the multilingual support function if you find Japanese difficult to search in.

## User feedback



I was able to use the examples of natural expressions and grammar that are found in actual conversations and documents!

I managed to progress in my research on the usage pattern, syntax analysis, and semantics of contemporary Japanese!

I could check the frequency of words and expressions, as well as the context they are used in. This enabled me to deepen my understanding of their meaning and usage.