

Distributional evidence for word class: Corpus data

少納言 (Syoonaon) interface to the Balanced Corpus of Contemporary Written Japanese (BCCWJ) [<https://shonagon.ninjal.ac.jp/>] [[info \(English\)](#)]

Goals:

- Use corpus to explore distribution patterns that might distinguish word classes
- Identify some complications in using corpus data!

About the corpus counts reported here:

- Collected on September 24, 2026 using BCCWJ(1) +BCCWJ(2), all years/genres
- The Syoonaon tool searches for exact-string matches only
- “Control cases” and/or percentages are provided below because raw numbers of hits are usually not useful; overall frequencies vary widely for different lexical items

(1) If a word is immediately followed by case markers *-ga/-o/-ni*, is it definitely a N?

	raw hits	% of total
行った itta 'go-PAST'	25089	
行ったが itta ga	421	1.68%
行ったを itta o	0	0%
行ったに itta ni	35	0.14%

	raw hits	% of total
自転車 ziten-sya 'bicycle'	5608	
自転車が ziten-sya ga	324	5.78%
自転車を ziten-sya o	695	12.39%
自転車に ziten-sya ni	659	11.75%

- *-ga*: nominative marker for N, **or** 'but, although' (which can follow V and A)
- *-ni*: multiple uses, not all of which require it to follow N
- *-o*: accusative marker for N

(2) If a word can be preceded by a demonstrative such as *sono* 'that', is it definitely a N?

	raw hits	% of total
高い takai 'high-NPST'	30715	
高いです takai desu	679	2.21%
その高い sono takai	43	0.14%

	raw hits	% of total
自転車 ziten-sya 'bicycle'	5608	
自転車です ziten-sya desu	19	0.34%
その自転車 sono ziten-sya	18	0.32%

- An adjective or other modifier can intervene between a demonstrative and its N

(3) X no N: What word class is X?

- "の報告": Examples from search results (報告 *hookoku* 'report')

故障の報告	故障	<i>kosyoo</i>	'breakdown, failure'
松村さんの報告	松村さん	<i>Matumura-san</i>	personal.name-HONORIFIC
次長からの報告	次長	<i>zityoo</i>	'vice-chief; assist. mgr.'
	から	<i>kara</i>	'from'
大坂への報告	大阪	<i>Oosaka</i>	'Osaka' (place name)
	へ	<i>e</i>	'to, toward'
協会などの報告	協会	<i>kyookai</i>	'association'
	など	<i>nado</i>	'etc.; things like that'

(4) Adjectives and adjectival nouns with forms of the copula: Different patterns?

	raw hits	% of total
高い takai 'high-NPST'	30715	
高いです takai desu 'high COP-FML'	679	2.21%
高いだ takai da '...COP-INFML'	114 (but check!)	0.37%
高いでした takai desita '...COP-PAST-FML'	0	0%
高いだった takai datta '...COP-PAST'	0	0%

	raw hits	% of total
静か sizuka 'quiet'	6981	
静かです sizuka desu	64	0.92%
静かだ sizuka da	270 (but check!)	3.87%
静かでした sizuka desita	24	0.34%
静かだった sizuka datta	125	1.79%