

Google data: Word classes in Japanese

About the Google data on this handout:

- Collected on September 29, 2011 unless otherwise noted
- All search terms were enclosed in quotation marks (" ")
- Comparison structures and percentages are provided because raw numbers of Google hits are usually not useful; raw G-hit numbers vary widely for different lexical items

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

	raw G-hits	% of total
行ったが itta ga 'go-PAST' + ??	648,000,000	83.9%
行ったを itta o	43,700,000	5.7% (quotations!)
行ったに itta ni	80,900,000	10.5%
<i>total</i>	<i>772,600,000</i>	

	raw G-hits	% of total
自転車が ziten-sya ga 'bicycle NOM'	8,480,000	25.3%
自転車を ziten-sya o	14,700,000	43.8%
自転車に ziten-sya ni	10,400,000	31.0%
<i>total</i>	<i>33,580,000</i>	

- *-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

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

	raw G-hits	% of total
高いです takai desu 'high COPULA'	26,500,000	97.7%
その高い sono takai	634,000	2.3%
<i>total</i>	<i>27,134,000</i>	

	raw G-hits	% of total
自転車で ziten-sya desu 'bicycle COPULA'	238,000	78.7%
その自転車 sono ziten-sya	64,300	21.3%
<i>total</i>	<i>302,300</i>	

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

(3) Combining (1) and (2): *sono* ___ *o* frame

	raw G-hits	% of total
その行ったを sono itta o	0	0.0%
その高いを sono takai o	2,560 (quotations!)	0.4%
その自転車を sono ziten-sya o	615,000	99.6%
<i>total</i>	<i>617,560</i>	

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

- "の報告": Examples from the top 30 Google hits (報告 *hookoku* 'report') [February 2010]

バグの報告	バグ	<i>bagu</i>	'bug (programming)'
安全性情報の報告	安全性情報	<i>anzensei-zyoohoo</i>	'safety information'
大井昌和の報告	大井 昌和	<i>Ooi Masakazu</i>	personal name (a manga artist)
事故等の報告	事故	<i>ziko</i>	'accident'
	等	<i>nado</i>	'etc.; things like that'
ポーランドからの報告	ポーランド	<i>Poorando</i>	'Poland'
	から	<i>kara</i>	'from'

(5) 'i-adjectives' and 'na-adjectives' with forms of the copula: Are the patterns different?

	raw G-hits	% of total
高いです takai desu 'high COP-FML'	45,800,000	54.1%
高いだ takai da ...COP-INFML'	1,030,000	1.2%
高いでした takai desita ...COP-PAST-FML'	3,680,000	4.3%
高いだった takai datta ...COP-PAST'	70,600	0.1%
高かったです takakatta desu 'high-PAST COP-FML'	34,000,000	40.2%
高かっただ takakatta da ...COP-INFML'	82,500	0.1%
<i>total</i>	<i>84,663,100</i>	

	raw G-hits	% of total
静かです sizuka desu 'quiet COP-FML'	2,320,000	7.7%
静かだ sizuka da	2,590,000	8.6%
静かでした sizuka desita	21,600,000	72.0%
静かだった sizuka datta	3,500,000	11.7%
—		
—		
<i>total</i>	<i>30,010,000</i>	

Key to color codes:

very rare	0– 4.9%
rare	5– 14.9%
frequent	15– 64.9%
very frequent	65– 100%

(6) Nouns and adjectival nouns with case particles

'quiet'	静かです sizuka desu 'quiet COP-FML'	静かな sizuka na 'quiet NA'	静かの sizuka no 'quiet GEN'	静かが sizuka ga 'quiet NOM'	静かを sizuka o 'quiet ACC'	<i>total</i>
Raw G-hits	2,320,000	41,200,000	264,000	28,100	21,100	43,833,200
% of total	5.3%	94.0%	0.6%	0.1%	0.0%	
'beautiful'	きれいです kirei desu	きれいな kirei na	きれいの kirei no	きれいが kirei ga	きれいを kirei o	<i>total</i>
Raw G-hits	13,100,000	50,700,000	1,100,000	206,000	694,000	65,800,000
% of total	20.0%	77.1%	1.7%	0.3%	1.0%	
'bicycle'	自転車です zitensya desu	自転車な zitensya na	自転車の zitensya no	自転車が zitensya ga	自転車を zitensya o	<i>total</i>
Raw G-hits	1,060,000	27,800	16,900,000	8,480,000	14,700,000	41,167,800
% of total	2.6%	0.0%	41.1%	20.6%	35.7%	
'sick'	病気です byooki desu	病気な byooki na	病気の byooki no	病気が byooki ga	病気を byooki o	<i>total</i>
Raw G-hits	3,040,000	1,190,000	27,100,000	11,500,000	10,300,000	53,130,000
% of total	5.7%	2.2%	51.0%	21.6%	19.4%	
'healthy, lively'	元気です genki desu	元気な genki na	元気の genki no	元気が genki ga	元気を genki o	<i>total</i>
Raw G-hits	21,800,000	31,500,000	18,100,000	20,300,000	16,000,000	107,700,000
% of total	20.2%	29.2%	16.8%	18.8%	14.9%	

- Do any of these seem to behave as predicted for N?
- Do any of these seem to behave as predicted for AN?
- Are there any cases that don't fit our expectations?