

## TreeTagger タグセット変更記録

- TreeTagger では、前置詞と従属接続詞に IN タグをふるため、NICE-NNS では、TreeTagger の IN タグを前置詞 IN と従属接続詞 CJ(新設)に分割。

※注：NICE-NS は未処理

表1：タグセット変更前 NICE-NNS に IN が振られていた項目と件数

項目	頻度	品詞	項目	頻度	品詞	項目	頻度	品詞
although	61	接続詞	about	541	前置詞	after	143	前置詞 or 接続詞
because	484	接続詞	above	18	前置詞	as	466	前置詞 or 接続詞
if	573	接続詞	across	6	前置詞	before	76	前置詞 or 接続詞
lest	1	接続詞	against	37	前置詞	except	14	前置詞 or 接続詞
once	3	接続詞	along	8	前置詞	for	1373	前置詞 or 接続詞
so	86	接続詞	among	32	前置詞	like	191	前置詞 or 接続詞
that	1195	接続詞	around	57	前置詞	since	56	前置詞 or 接続詞
though	59	接続詞	at	594	前置詞	than	210	前置詞 or 接続詞
whether	55	接続詞	behind	15	前置詞	till	3	前置詞 or 接続詞
while	55	接続詞	besides	1	前置詞	unless	3	前置詞 or 接続詞
			between	68	前置詞	unlike	1	前置詞 or 接続詞
			beyond	3	前置詞	until	49	前置詞 or 接続詞
			by	505	前置詞			
			despite	8	前置詞			
			during	34	前置詞			
			from	427	前置詞			
			in	2522	前置詞			
			inside	3	前置詞			
			into	50	前置詞			
			near	15	前置詞			
			of	2739	前置詞			
			off	2	前置詞			
			on	719	前置詞			
			onto	1	前置詞			
			out	6	前置詞			
			outside	16	前置詞			
			over	66	前置詞			
			per	11	前置詞			

through	102	前置詞
throughout	10	前置詞
toward	13	前置詞
towards	3	前置詞
under	23	前置詞
up	1	前置詞
upon	2	前置詞
via	6	前置詞
with	580	前置詞
within	3	前置詞
without	88	前置詞

- 接続詞のみの機能を持つものは、機械的に IN から CJ に変更。
- 前置詞のみの機能を持つものは、変更なし。
- 接続詞と前置詞の両方の機能を持つ 12 項目約 2,500 件を、1 件 1 件文脈を確認し、IN (前置詞)または CJ (従属接続詞)に振り分けた。

表 2: 振分け後の件数

	CJ	IN
after	43	99
as	92	356
before	33	39
except	1	13
for	5	1368
like	15	147
since	43	13
than	190	20
till	0	3
unless	3	0
unlike	0	1
until	21	28