

Bunsirou wa tamerawazuni sono yubi wo kuti ni hukumu to, kizuguti wo tuyoku sutta.
文四郎はためらわずにその指を口に含むと、傷口を強く吸った。

Example sentence from BCCWJ sample OT02_00007 (Miyaji et al. (2005) Kokugo 3. Mitsumura Toshio Publishing Company.):

Surface

Morphological analysis with
MeCab 0.994 and UniDic 1.3.12

Dependency structure analysis with
CaboCha 0.64

tokens = 19

morphemes (short unit words) as defined in
UniDic 1.3.12/MeCab 0.994

1 2
文四郎は

3 4 5
ためらわずに

6
その

7 8
指を

9 10
口に

11 12 13
含むと、

14 15
傷口を

16
強く

17 18 19
吸った。

characters = 29

one character per Unicode codepoint

chunks = 9

phrasal unit consisting of content word
and functional unit, as defined in
CaboCha; 'bunsetsu'

chunk depth = 2.78

average 'depth' of chunk; links of
distance greater than one increase
depth; higher values indicate greater
nesting complexity

depth

0
1
2
2
3
3
4
5
5

Syntactic

link distance = 2.44

average number of chunks between
source and linked chunks; higher values
indicate greater syntactic complexity

文四郎は
ためらわずに

その
指を

口に
含むと、

傷口を
強く

吸った。

predicates = 2

following definition in
(Shibasaki and Hara, 2010)

Vocabulary

JLPT word level = 0.45

average (pre-2010) Japanese Language
Proficiency Test word level; lower scores
correspond with higher difficulty

BCCWJ-LB word level = 4.56

average log-frequency of words in the
BCCWJ Library sub-corpus (LB); lower
scores correspond with higher difficulty

Writing system

writing system percentages

hiragana = 0.59

katakana = 0

kanji = 0.35

romaji = 0

symbols = 0.03

commas = 1