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。

BCCWJ sample
OT02_00007

Morphological analysis with
MeCab 0.994 and UniDic 1.3.12

tokens = 19

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

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
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	。
1		2			3	4		5		6			7		8	9		

characters = 29

one character per Unicode codepoint

chunks = 9

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