

付録

日本語のフルコンテキストラベルの形式

HTS バージョン 2.3 の日本語用のデモ^{*1}に基づく、フルコンテキストラベルの形式は以下のとおりです。

$$p_1 \hat{p}_2 - p_3 + p_4 = p_5 / A: a_1 + a_2 + a_3$$

$$/ B: b_1 - b_2 - b_3 / C: c_1 - c_2 + c_3 / D: d_1 + d_2 - d_3$$

$$/ E: e_1 - e_2 ! e_3 - e_4 - e_5 / F: f_1 - f_2 \# f_3 - f_4 @ f_5 - f_6 | f_7 - f_8 / G: g_1 - g_2 \% g_3 - g_4 - g_5$$

$$/ H: h_1 - h_2 / I: i_1 - i_2 @ i_3 + i_4 \& i_5 - i_6 | i_7 + i_8 / J: j_1 - j_2$$

$$/ K: k_1 + k_2 - k_3$$

表 1 は各コンテキストの意味をまとめています。POS は part-of-speech (品詞) の略語です。

表 1: 日本語のフルコンテキストラベルの形式

p_1	当該音素から見て 2 つ前の音素	
p_2	先行音素	
p_3	当該音素	
p_4	後続音素	
p_5	当該音素から見て 2 つ後の音素	
a_1	アクセント核と当該モーラの位置の差	-49 ~ 49
a_2	当該アクセント句中の当該モーラの位置 (前方向)	1 ~ 49
a_3	当該アクセント句中の当該モーラの位置 (後方向)	1 ~ 49
b_1	先行単語の POS タグ	
b_2	先行単語の活用 of 型 (五段活用、サ行変格活用等)	
b_3	先行単語の動詞の活用形 (未然形・連用形・終止形・連体形・假定形・命令形)	
c_1	当該単語の POS タグ	
c_2	当該単語の活用 of 型 (五段活用、サ行変格活用等)	
c_3	当該単語の動詞の活用形 (未然形・連用形・終止形・連体形・假定形・命令形)	

^{*1} http://hts.sp.nitech.ac.jp/archives/2.3/HTS-demo_NIT-ATR503-M001.tar.bz2

d_1	後続単語の POS タグ	
d_2	後続単語の活用の型 (五段活用、サ行変格活用等)	
d_3	後続単語の動詞の活用形 (未然形・連用形・終止形・連体形・假定形・命令形)	
e_1	先行アクセント句に含まれるモーラ数	1 ~ 49
e_2	先行アクセント句のアクセント核	1 ~ 49
e_3	先行アクセント句が疑問形かどうか (疑問形でない: 0、疑問形: 1)	
e_4	undefined context	
e_5	先行アクセント句と現在のアクセント句の間にポーズが挿入されるかどうか	
f_1	当該アクセント句に含まれるモーラ数	1 ~ 49
f_2	当該アクセント句のアクセント核	1 ~ 49
f_3	当該アクセント句が疑問形かどうか (疑問形でない: 0、疑問形: 1)	
f_4	undefined context	
f_5	当該呼気段落中の当該アクセント句の位置 (アクセント句単位、前方向)	1 ~ 49
f_6	当該呼気段落中の当該アクセント句の位置 (アクセント句単位、後方向)	1 ~ 49
f_7	当該呼気段落中の当該アクセント句の位置 (モーラ単位、前方向)	1 ~ 99
f_8	当該呼気段落中の当該アクセント句の位置 (モーラ単位、後方向)	1 ~ 99
g_1	後続アクセント句に含まれるモーラ数	1 ~ 49
g_2	後続アクセント句のアクセント核	1 ~ 49
g_3	後続アクセント句が疑問形かどうか (疑問形でない: 0、疑問形: 1)	
g_4	undefined context	
g_5	後続アクセント句と現在のアクセント句の間にポーズが挿入されるかどうか	
h_1	先行呼気段落中のアクセント句数	1 ~ 49
h_2	先行呼気段落中のモーラ数	1 ~ 99
i_1	当該呼気段落に含まれるアクセント句数	1 ~ 49
i_2	当該呼気段落中のモーラ数	1 ~ 99
i_3	文中の当該呼気段落中の位置 (呼気段落単位、前方向)	1 ~ 19
i_4	文中の当該呼気段落中の位置 (呼気段落単位、後方向)	1 ~ 19
i_5	文中の当該呼気段落中の位置 (アクセント句単位、前方向)	1 ~ 49
i_6	文中の当該呼気段落中の位置 (アクセント句単位、後方向)	1 ~ 49
i_7	文中の当該呼気段落中の位置 (モーラ単位、前方向)	1 ~ 199
i_8	文中の当該呼気段落中の位置 (モーラ単位、後方向)	1 ~ 199
j_1	後続呼気段落に含まれるアクセント句数	1 ~ 49
j_2	後続呼気段落中のモーラ数	1 ~ 99
k_1	文中の呼気段落数	1 ~ 19
k_2	文中のアクセント句数	1 ~ 49

k_3	文中のモーラ数	1 ~ 199
-------	---------	---------