

XML タグ付け方法

1 XML タグ

1.1 方言箇所には<hg></hg>タグをつけ、その中に要素項目 target='' (方言を含む文節全て), bs='' (さらに細かい文節), kihon='' (基本形), type='' (方言の種類, 添付書類 1 を参照) を加える.

例: <hg target='' bs='' kihon='' type=''></hg>

1.2 タグやスペースはすべて半角。要素間 (「kihon=''」と「type=''」との間など) は半角スペースを入れる。

2 方言を抜き出す範囲

5 文節程度を上限とし、直感で句点を入れた方が良く感じる長さにする。つまり程良い長さで方言を抜き出す。方言を抜き出す範囲を以下のタグで区切る。フレーズ階層→方言の記述階層→基本形→方言分類の順でタグ付けをする。以下、例。

例: <hg target=' だけんね。短…長いときは全然きにならんかったわけ' bs=' だけんね。' kihon=' ダケン' type=' voc' bs1=' 気にならんかったわけ' kihon1=' ンカッタ' type1=' neg/jug'></hg>

3 タグつけの際の注意点

<hg>の中に複数の方言要素が現れた場合 (上の例参照)。方言要素を含む文節もしくは短いまとまりを<bs>タグでくくる。<bs>が複数ある場合、先頭は数値無し、以降 1、2…と通し番号を付ける。<bs>の範囲に複数の方言要素が現れる場合 (例えば bs1 と bs2 の中に、それぞれふたつの方言要素がある場合)、以下のように記述する。(XML タグ付けマニュアル 2014 年 2 月 23 日版の④の 6 参照)

3.1 例: bs1=' ◇◇' kihon1=' ○○' type1=' ▽▽' kihon1-1=' ☆☆' type1-1=' □□'

bs2=' ◆◆' kihon2=' ●●' type2=' ▼▼' kihon2-1=' ★★' type2-1=' ■■'

<添付>

添付書類①: XML タグ付けマニュアル 2013 年 03 月 04 日版

添付書類②: XML タグ付けマニュアル 2014 年 02 月 23 日版

添付書類③: タグ表

添付書類①：XML タグ付けマニュアル 2013 年 03 月 04 日版

- ・方言を抜き出す範囲:5 文節程度を上限とし、直感で句点を入れた方が良く感じる長さ。つまり程良い長さで。方言を抜き出す範囲を以下のタグで区切る。

<hg target='' kihon='' type=''></hg>

- ・タグやスペースはすべて半角。要素間（「kihon=''」と「type=''」との間など）は半角スペースを入れる。

- ・フレーズ階層→方言の記述階層→基本形→方言分類 の順でタグ付け。以下、例。

※3 日版から変更

```
<hg target=' ビニールば着てから ころころ寝とったろうが' bs=' ビニールば' kihon='  
バ' type=' case' bs1=' 着てから' kihon1=' テカラ' type1=' conj' bs2=' 寝とったろう  
が' kihon2=' ツ' type2=' son' kihon2-1=' トル' type2-1=' time' kihon2-2=' ガ'  
type2-2=' time'></hg>
```

※変更点→<p>タグが無くなる。実質<hg>が<p>の階層に。<hg>の中に複数の方言要素が現れた場合（上の例参照）。方言要素を含む文節もしくは短いまとまりを<ph>タグでくくる。<bs>が複数ある場合、先頭は数値無し、以降 1、2…と通し番号を付ける。<bs>の範囲に複数の方言要素が現れる場合（例えば bs1 と bs2 の中に、それぞれふたつの方言要素がある場合）、以下のように記述する。

```
bs1=' ◇◇' kihon1=' ○○' type1=' ▽▽' kihon1-1=' ☆☆' type1-1=' □□'  
bs2=' ◆◆' kihon2=' ●●' type2=' ▼▼' kihon2-1=' ★★' type2-1=' ■■'
```

※target の取り方：発話が短い場合、target と bs が一致することがある。この場合、bs は必要ない。

例：

A：昨日、下通りで牡蠣ば食べたたい。

<hg target=' 下通りで牡蠣ば食べた' bs=' 牡蠣ば' …>

A：食べたたい。

<hg target=' 食べたたい' kihon=' タイ' type=' end' …>

→<bs=' 食べたたい'>は必要ない。

・ 方言的な要素がひとつの単位に重なって見られる場合、それぞれの要素を別にタグ付ける。

※「遅かったたい」の「たっ」は「たと」の「と」が促音化したもの。従って「ト」と「促音化 (son)」についてタグ付けを行う。以下、例。

```
<hg target='だけん遅かったたい' kihon='ダケン' type='voc' kihon1='ト' type1='end'  
kihon2='ツ' type2='son' kihon3='タイ' type3='end'></hg></p>
```

添付書類②：XML タグ付けマニュアル 2014 年 02 月 23 日版

- ① タグ付け業者依頼用フォルダーの中の以下の pdf を見ながらそれに対応する word 書類にタグ付けをお願いします。

003_110325_0823A (タグ：しまもと) .pdf

009_003A_100202_0000 (タグ：よしと) .pdf

019__110406__2315A (タグ：わだ) .pdf

024_110322_1825② (タグ：おおば) .pdf

- ②pdf には、方言の書き起こしが書かれており、その文章中の方言の部分にタグが降られています。

例) 003_110325_0823A (タグ：しまもと) .pdf の一ページ目より

B: あ、そうそう。(だけんね、) (A: うん) 短…長いときは全然(気にならんかったわけ) (A: うん) その、(何ヶ月(ほっとい)とつても) ただ下が伸びてくるだけでしょ? (A: うん) 色もずっと黒かったし。
(A: うん) でもこんなやって(切ってしまったらね) (なんて言う) この、切った (A: あ、) 切り口とか気になるわけ。

- ③この青い (大きい) 開始かつこの部分に半角で<hg target=' ' …のタグを入れてください。赤の (小さい) 括弧の部分は、<hg の中のbs=' ' の中に書いてください。

例えば、この B の文はタグが入ると以下ようになります。

```
<hg target=' だけんね。短…長いときは全然きにならんかったわけ' bs=' だけんね。' kihon=' ダケン' type=' voc' bs1=' 気にならんかったわけ' kihon1=' ンカッタ' type1=' neg/jug' > 文</hg>
```

④タグについての注意事項

- i) タグは、基本的に以下の形式です。

```
<hg target=' ' bs=' ' kihon=' ' type=' ' >文</hg>
```

- ii) target=' ' の中には青い括弧でくくられている部分がすべて入ります。

- iii) bs=' ' の中には赤い括弧でくくられている部分がすべて入ります。青い括弧中に何回か出現します。その場合、bs から始めて、bs1、bs2、bs3 と番号を増やして行ってください。

- iv) kihon=' ' の中にはカタカナでかかかれている方言部分が入ります。kihon=' ダケン' kihon1=' ンカッタ' など。方言部分は、' ' でくくります。

v) type=' ' の中にはアルファベットでかかれていますコードが入ります。
type=' voc' type1=' neg/jug' などは、' 'でくくります。

vi) <bs>の範囲に複数の方言要素が現れる場合（例えば bs1 と bs2 の中に、それぞれふたつの方言要素がある場合）、以下のように記述する。（このケースはあまりないと思います。）

bs1='◇◇' kihon1='○○' type1='▽▽' kihon1-1='☆☆' type1-1='□□'
bs2='◆◆' kihon2='●●' type2='▼▼' kihon2-1='★★' type2-1='■■'

vii) 発話の途中の相槌→（ ）に入れて、発話者の発話終了後に記す。

★注）タグやスペースはすべて半角。要素間（「kihon=' '」と「type=' '」との間など）は半角スペースを入れる。