

KNPで付与される feature 一覧

河原 大輔

京都大学 大学院情報学研究科

2013年1月23日

形態素に付与される feature

カテゴリ	カテゴリの情報 (JUMAN の意味情報)
ドメイン	ドメインの情報 (JUMAN の意味情報)
漢字読み	漢字の読みが音読みか訓読みか (JUMAN の意味情報)
連語	連語の一部であることを示す (JUMAN の意味情報)
自他動詞	(JUMAN の意味情報)
使役動詞	(JUMAN の意味情報)
代表表記	(大半は JUMAN の意味情報)
弱時相名詞	「夏休み」「途中」など特定の時間でないもの (JUMAN の意味情報)
数量修飾	「しめて」など、「しめて3つ」を扱う (大半は JUMAN の意味情報)
相対名詞修飾	「いくぶん」など、「いくぶん右」を扱う (JUMAN の意味情報)
用言弱修飾	「とっても」など、「とっても静かに」を扱う (大半は JUMAN の意味情報)
補文ト	「～と思う」、「～という」など (JUMAN の意味情報)
付属動詞候補 (基本)	「(書き) 損じる」など (JUMAN の意味情報)
付属動詞候補 (タ系)	「(書いて) さしあげる」など (JUMAN の意味情報)
修飾 (ニ格)	「無条件に」など (大半は JUMAN の意味情報)
修飾 (デ格)	「無条件で」など (大半は JUMAN の意味情報)
修飾 (ト格)	「一段と」など (大半は JUMAN の意味情報)
～を～に構成語	「(～を) ～に」になる可能性の高い語 (大半は JUMAN の意味情報)
換言	言い換え表現 (JUMAN の意味情報)
形容詞派生	形容詞から派生した形態素 (JUMAN の意味情報)
動詞派生	動詞から派生した形態素 (JUMAN の意味情報)

名詞派生	名詞から派生した形態素 (JUMAN の意味情報)
同義	同義語 (JUMAN の意味情報)
反義	反義語 (JUMAN の意味情報)
多義	多義であることを示す (JUMAN の意味情報)
対語	対語からなることを示す (JUMAN の意味情報)
地名	地名 (JUMAN の意味情報)
人名	人名、姓もしくは名 (JUMAN の意味情報)
地名末尾	地名の末尾に来る形態素 (大半は JUMAN の意味情報)
地名末尾外	地名の末尾の後に来る形態素 (JUMAN の意味情報)
住所末尾	住所の末尾に来る形態素 (JUMAN の意味情報)
人名末尾	人名の末尾に来る形態素 (大半は JUMAN の意味情報)
人名末尾外	人名の末尾の後に来る形態素 (JUMAN の意味情報)
組織名末尾	組織名の末尾に来る形態素 (大半は JUMAN の意味情報)
顔文字	顔文字 (JUMAN の意味情報)
長音挿入	長音挿入による非標準表記 (JUMAN の意味情報)
非標準表記	小文字化、長音化による非標準表記 (JUMAN の意味情報)
長音挿入可	長音が挿入可能 (JUMAN の意味情報)
濁音可	連濁可能 (JUMAN の意味情報)
尊敬動詞	尊敬動詞 (JUMAN の意味情報)
謙讓動詞	謙讓動詞 (JUMAN の意味情報)
丁寧動詞	丁寧動詞 (JUMAN の意味情報)
授受動詞	授受動詞 (JUMAN の意味情報)
可能動詞	可能動詞 (JUMAN の意味情報)
可能接尾辞	可能接尾辞 (JUMAN の意味情報)
動詞転	動詞の転 (JUMAN の意味情報)
標準	非標準的な形態素に対する標準形 (JUMAN の意味情報)
方言	方言であることを示す (JUMAN の意味情報)
古語的	古語的であることを示す (JUMAN の意味情報)
注釈	注釈 (JUMAN の意味情報)
省略	省略表現であることを示す (JUMAN の意味情報)
漢字部首名	漢字部首名 (JUMAN の意味情報)
内容語	意味属性を取得したり、キーワードとして用いる部分 (接頭・接尾辞に関しては JUMAN の意味情報)
準内容語	独立した基本句ではないが内容語 (JUMAN の意味情報)
自動獲得	Web・Wikipedia から自動獲得された語 (JUMAN の意味情報)
自動認識	動的未知語処理によって認識された語 (JUMAN の意味情報)
濁音化	連濁によって 1 文字目が濁音になった語 (JUMAN の意味情報)
意味分類	自動獲得語に対する意味分類 (JUMAN の意味情報)

既知語帰着	自動獲得語に対して帰着した既知語 (JUMAN の意味情報)
Wikipedia 上位語	Wikipedia から獲得した上位語 (一部は JUMAN の意味情報)
Wikipedia エントリ	Wikipedia から獲得した表現
Wikipedia リダイレクト	Wikipedia から獲得した同義表現
Wikipedia 多義	Wikipedia から獲得した表現であり、多義
読み不明	Wikipedia から獲得した表現であり、読みが不明
文頭	文頭の形態素
文末	文末の形態素
表現文末	括弧、句点など以外の、文末の通常の単語をマーク
一般並列語	並列を示す助詞、接続詞
時間辞	「年」、「月」など
時間名詞	「年」、「月」など
弱数詞	「半(個)」「多数(個)」など、数詞ではないが数詞相当の語
カウンタ	数詞に続く名詞または接尾辞。数量の文節の類似度はカウンタに対して与える
相対名詞	「左」、「東」など
副詞的接尾辞	「ほど」、「間」など
疑問詞	「何」「誰」など
否定	否定を表す形態素
類似計算	並列の類似度計算を柔軟にするため
一人称	「私は... 思う」などを扱うため
呼掛	並列でない可能性大
疑似代表表記	代表表記が与えられていない形態素に対して擬似的に生成した代表表記
複合辞	「(~を)めぐる」、「(~と)いう」など
と基本形複合辞	複合辞を構成する形態素
とタ系連用テ形複合辞	複合辞を構成する形態素
に基本形複合辞	複合辞を構成する形態素
に基本連用形複合辞	複合辞を構成する形態素
にタ形複合辞	複合辞を構成する形態素
にタ系連用テ形複合辞	複合辞を構成する形態素
を基本形複合辞	複合辞を構成する形態素
を基本連用形複合辞	複合辞を構成する形態素
をタ形複合辞	複合辞を構成する形態素
をタ系連用テ形複合辞	複合辞を構成する形態素
形態素連結	KNP の最初の処理で形態素列を連結して一つの形態素にしたことを示す
移動動詞	移動を表す動詞

形容詞的名詞	形容詞と名詞の曖昧性がある形態素
名詞的動詞	名詞と動詞の曖昧性がある形態素
記英数カ	記号, 英文字, 数字, カタカナ
漢字	漢字(「カ月」「ヶ国」なども)
かな漢字	ひらがなもしくは漢字
カタカナ	カタカナ
数字	数字
英記号	英記号
記号	記号
かな漢字	ひらがなもしくは漢字
ひらがな	ひらがな
括弧始	括弧始
括弧終	括弧終
括弧	括弧
述語区切	括弧終, 読点, 「——」, 「……」
名詞的形容詞語幹	ナ形容詞の語幹
名詞相当語	名詞に相当する形態素(名詞、名詞的形容詞語幹、名詞性接尾辞)
形副名詞	形式名詞または副詞的名詞
活用語	活用する形態素
用体変化	用言, 体言を変化させる接尾辞
接頭	接頭辞, 括弧始
付属	付属語(決まり文句の中の自立語も)
自立	接頭でも付属でもないもの
サ変	サ変名詞, 「～勝」など
サ変動詞	「する」「出来る」「直後」などの続くサ変名詞
複合←	最終的に複合語になる(前の語に続く)と判断された語
非独立接頭辞	独立した基本句ではない接頭辞, 括弧始(「お」「第」…)
特殊非見出語	形式名詞の「の」、(サ変+)「予定」: ひとつ前の形態素で辞書引き(独立基本句)
タグ単位始	新たなタグ単位(基本句)の始まり
文節始	新たな文節の始まり
文節主辞	文節の主辞
固有キー	固有表現認識のための手がかり形態素
固有修飾	固有表現を修飾する形態素
NE	固有表現認識結果
品曖	品詞の曖昧性があることを示す
原形曖昧	原形の曖昧性があることを示す
ALT	形態素の別候補

品詞変更	KNP において品詞を変更したことを示す
代表表記変更	KNP において代表表記を変更したことを示す
未知語	未定義語
複合名詞構成	「通常国会」のように副詞+名詞などで複合名詞を構成
連用形名詞化	「蒸し(器)」など名詞化する動詞連用形
連用形名詞化疑	連用形名詞化する可能性があるもの
弱動詞	動詞と、接続詞、副詞、連体詞などとの曖昧性をもつ形態素
ある曖昧	動詞と連体詞の曖昧性をもつ「ある」
照応接頭辞	照応詞となる接頭辞(「同」)
肩書同格	「肩書+人名」タイプの同格表現
地名同格	地名に関する同格表現
組織同格	組織名に関する同格表現
種別同格	「～の一種」のような同格表現

基本句 (タグ単位) ・ 文節に付与される feature

読点	形態素から伝搬
句点	形態素から伝搬
括弧始	形態素から伝搬
括弧終	形態素から伝搬
助詞	形態素から伝搬
接続詞	形態素から伝搬
連体詞	形態素から伝搬
指示詞	形態素から伝搬
副詞形態指示詞	形態素から伝搬
連体詞形態指示詞	形態素から伝搬
副詞	形態素から伝搬
感動詞	形態素から伝搬
～ない	形態素から伝搬
～ぬ	形態素から伝搬
形副名詞	形態素から伝搬
数量相対名詞修飾	形態素から伝搬
数量修飾	形態素から伝搬
相対名詞修飾	形態素から伝搬
用言弱修飾	形態素から伝搬
相対名詞	形態素から伝搬
補文ト	形態素から伝搬
的	形態素から伝搬
一人称	形態素から伝搬
疑問詞	形態素から伝搬
肩書同格	形態素から伝搬
サ変	形態素から伝搬
サ変動詞	形態素から伝搬
連体修飾	形態素から伝搬, 体言止を判定詞と解釈するか動詞と解釈するか 等で利用
括弧始 2	句内に括弧始が二つある場合
括弧終 2	句内に括弧終が二つある場合
強サ変	格解析対象となるサ変名詞
サ変動詞化	「項目を確認の後」など動詞化したサ変名詞
述語化	「～をご覧の」など動詞化した名詞

文頭	文頭の文節 (プログラムで与えられる)
文末	文末の文節 (プログラムで与えられる)
引用内文頭	引用内の文頭文節
引用内文末	引用内の文末文節
主節	文の主節
非主節	文の主節ではない
体言	体言
用言	用言
準用言	「首より下の…」や「～から～日間」などの相対名詞や時間
弱用言	弱い用言
～とみられる	「～とみられる」タイプの弱用言
機能的用言	機能的な用言
用言一部	「責任(ある)」など、用言の一部
名詞項候補	橋渡し照応の先行詞候補
先行詞候補	ゼロ代名詞の先行詞候補
候補	係り先の候補
文節内	複合名詞の主辞基本句以外に付与
SM	意味マーカ
レベル	用言の強さ (A-,A,B-,B,B',C の 6 段階)
ID	ルールの ID(タイプ)
正規化代表表記	形態素曖昧性を含む正規化した代表表記
主辞代表表記	主辞の基本句に対する正規化代表表記
主辞'代表表記	主辞が一文字の場合にその前の句も含めた正規化代表表記
用言代表表記	用言の代表表記 (格フレーム収集用)
連用要素	連用の句・節
連用節	連用修飾節
連体節	連体修飾節
格要素	用言の格要素
非格要素	格要素ではない
修飾	修飾的な要素
態	用言の態
判定詞	判定詞
体言止	体言で文が終わっている場合
サ変止	サ変名詞で文が終わっている場合
裸名詞	助詞などを伴わない裸の名詞
非用言格解析	格解析対象のサ変名詞や形容詞語幹など
状態述語	形容詞、判定詞などの状態を表す述語

動態述語	動態を表す述語 (状態述語以外)
強数量修飾	「このうち3つ」などの数量修飾
ため-せい	「ため」「せい」
補文	補文表現
区切	並列構造の障壁としての強さ. Xが読点あり, Yが読点なしの場合
並キ	並列構造の存在をしめす表現
読点並キ	並列構造の存在をしめす表現 (読点あり)
サ変スル	並列の条件で使う
カタカナ読点カタカナ	並列の条件で使う
スルナル	並列の条件で使う (「〜と」の閾値調整)
カウンタ	形態素から語彙をとってくる
一文字漢字	一文字漢字の自立語から構成される
時間	時相名詞, 時間辞, 時間名詞を含む. 例外あり
強時間	「3時に」などの強い時間表現
数量	数詞, 助数辞を含む. 例外あり
回数	「〜回」「〜度」などの回数
時数ノ	「3時の」「3つの」など. 「僕の3つの宿題」で「僕の」が「3つの」に係らないようにする
時数裸	連体修飾を制限. 「六日 (告示の)」「昨日 (告示の)」など
可能表現	形態素の可能動詞から伝搬
時制	時制の記述
未来句	明日, 来週, 3日後 など
否定表現	否定の表現
準否定表現	否定的な表現
二重否定	二重否定の表現
限定否定	「〜だけでない」
地名	地名
地名疑	地名の可能性がある句
人名	人名
人名疑	人名の可能性がある句
組織名	組織名
組織名疑	組織名の可能性がある句
固有名詞	固有名詞
住所	〜市, 〜町など
住所疑	住所の可能性がある句
住所の	連体修飾している住所句
名並終点	名詞並列の終点にボーナスを与えるため

述並終点	述語並列の終点にボーナスを与えるため
～られる	「(ら)れる」(格フレーム収集用)
～させる	「(さ)せる」(格フレーム収集用)
～もらう	(格フレーム収集用)
～ほしい	(格フレーム収集用)
～できる	(格フレーム収集用)
～にくい	(格フレーム収集用)
～たい	(格フレーム収集用)
～いただく	(格フレーム収集用)
外の関係	連体修飾が「外の関係」であることを構造評価で用いるため
ルール外の関係	「外の関係」とする表現
トモ	並列の末尾の制限のため
デモ	並列の末尾の制限のため & 疑問の判定用
モ	並列の末尾の制限のため
モ～	並列の末尾の制限のため
ト	並列の末尾の制限のため
タリ	並列の末尾ヒントのため
テモ	疑問の判定用
ガ	助詞のマーク
ヲ	助詞のマーク
ニ	助詞のマーク
デ	助詞のマーク
カラ	助詞のマーク
ヨリ	助詞のマーク
ヘ	助詞のマーク
マデ	助詞のマーク
ニテ	助詞のマーク
ハ	助詞のマーク
カ	助詞のマーク & 疑問の判定用
デハ	助詞のマーク
ニハ	助詞のマーク
ニモ	助詞のマーク
トハ	助詞のマーク
疑問	疑問表現
提題	「～は」, 「～では」
提題受	「～は」(未格, 提題)の係り先の優先度
～によると	文末に係るとする(新聞特有かもしれない)

～によれば	文末に係るとする (新聞特有かもしれない)
係	係タイプ
受	受タイプ
非主題	主題ではない
主題表現	主題を表す表現
準主題表現	主題表現に係るノ格
強調構文	強調構文の述語にマーク
複合辞	複合辞表現
連体並列条件	連体の用言の並列の末尾の制限のため
～ (複合辞+も)	「～はもちろん」の並列の末尾の制限のため
～を (含む)	「～はもちろん」の並列の末尾の制限のため
並列タイプ	並列のタイプ (AND, OR, ?, OTHER)
格解析なし	格解析対象外の用言
格要素指定	格要素の位置を指定
格要素表記直前参照	「方」など一つ前の句を格要素とするとき
省略解析なし	省略解析対象外の用言
省略格指定	省略解析の対象の格
省略解析対象指示詞	省略解析対象の指示詞
人称代名詞	人称代名詞 (「彼」「彼女」)
照応ヒント	照応先の手がかり
不特定人	不特定:人を格要素にとる
格関係	格解析の結果 (用言に付与)
格解析結果	格解析の結果 (用言に付与)
格関係 X	格解析の結果 (用言に付与、格要素ごと)
解析格	格解析結果 (格要素に付与)
解析連格	被連体修飾詞の連体修飾節に対する格解析結果 (格要素に付与)
NE	固有表現 (主辞基本句に付与)
NE 内	固有表現を構成する基本句
Wikipedia 上位語	Wikipedia から獲得した上位語
Wikipedia エントリ	Wikipedia から獲得した表現
Wikipedia リダイレクト	Wikipedia から獲得した同義表現
括弧同格	「」を使った同格表現
主題格	主格の推定
クエリ必須係り受け	検索において必須の係り受け
クエリ削除語	検索において不要な表現
係チ	係り受け可能性のチェックに利用
後方チ	係り受け可能性のチェックに利用

隣係絶対	隣に絶対係る
隣受絶対	隣を絶対受ける
敬語	敬語表現
節機能	節の機能の分類
条件節候補	条件節の候補
仮定法	仮定法の条件節
時間経過同時候補	「時間経過:同時」の候補
モダリティ	モダリティの分類
受け NONE	受けの feature がない (ERROR 検出用)
タグ単位受	基本句間の係り受けを制御
～を～に可能	「～を～に」の可能性を示す
～が～に疑	「～が～に」の疑いがある (格フレーム収集用)
～が～の疑	「～が～の」の疑いがある (格フレーム収集用)
～限	副詞的な「最低限」「最大限」「最小限」に付与