

Categories & Features of the XIP German Grammar

(september 2006, S. Maurel)

1. Introduction

The purpose of this document is to list the existing categories and features that are used by the XIP German grammar (by september 2006).
A category in a XIP grammar must be declared as a set of features (attribute-value pairs).

2. List of categories

Following is the list of categories, with their main feature in square brackets.

in: categories_tagger.xip :

```
top = [cat=0].                //acts as root to chunking trees

adj = [adj=+,verbal=-].
adv = [adv=+,verbal=-].
conj = [conj=+,verbal=-].
det = [det=+,verbal=-].
fm = [fm=+,verbal=-].
itj = [itj=+,verbal=-].
noun = [noun=+,verbal=-].
num = [num=+,verbal=-].
ptcl = [ptcl=+,verbal=-].
Padj = [padj=+,verbal=-].
prep = [prep=+,verbal=-].
Postp = [postp=+,verbal=-].
pron = [pron=+,verbal=-].
punct = [punct=+,verbal=-].
trunc = [trunc=+,verbal=-].
verb = [verb=+,verbal=+].
symbol = [symbol=+,verbal=-].
```

in: categories.xip :

```
vfin = [ verb=+ ].
infzu = [verb=+ ].
ap = [ adj=+ ].
apl = [adj=+ ].
np = [ np=+ ].
pp = [ pp=+ ].
```

3. List of features

Following is the list of features, with their possible values in curly brackets.

in: features_tagger.xip:

```
// main categories

cat:{0},

adj:{+},    // adjective
adv:{+},    // adverb
conj:{+},   // conjunction
det:{+},    // determiner
itj:{+},    // interjection
noun:{+},   // noun
num:{+},    // numeral
padj:{+},   // adjective-like pronoun
pref:{+},   // verb prefix
prep:{+},   // preposition
postp:{+},  // postposition
pron:{+},   // pronoun
ptcl:{+},   // particle
punct:{+},  // punctuation
trunc:{+},  // truncated word
verb:{+},   // verb

fm:{+},     // foreign material

// spelling information
newspell:{+}, // new spelling           :: Fluss
oldspell:{+}, // old spelling             :: Fluß
err:{+},      // erroneous spelling    :: erstmal [should be "erst mal"]

// lexical source
lex:{+},
guessed:{+},
openclass:{+},

// ranking (for compounds)
rank:{1,2,3,4,5,6,7,8,9},

// filter for depreciated readings
filter:{+},
filter2:{+},
filtertrunc:{+},

// for special symbols
spec:{+},

/ morphology \

agr: [
  gender:    [ masc:{+}, fem:{+}, neut:{+} ],
  number:    [ sg:{+}, pl:{+}],
  case:      [ nom:{+}, acc:{+}, dat:{+}, gen:{+} ],
  decl:      [ strong:{+}, weak:{+} ]
],

nagr: [
  mnom:{+}, macc:{+}, mdat:{+}, mgen:{+},
  fnom:{+}, facc:{+}, fdat:{+}, fgen:{+},
  nnom:{+}, nacc:{+}, ndat:{+}, ngen:{+},
  xnom:{+}, xacc:{+}, xdat:{+}, xgen:{+}
],
```

```

person:      [ P1:{+}, P2:{+}, P3:{+} ],

vform:      [ finite:{+}, indic:{+}, subjc:{+}, nursubjc:{+}, imp:{+}, ppast:{+},
              ppres:{+}, inf:{+}, zu:{+,-} ],
vtype:      [ vfull:{+}, aux:{+}, modal:{+}, trans:{+,-} ],
tense:      [ pres:{+}, past:{+} ],
degree:     [ pos:{+,-}, comp:{+,-}, super:{+,-} ],

            invar:{+},
            pred:{+,-},
            attr:{+},
            v1:{+,-},
            v2:{+,-},

pprzu:{+},           //for adjectival participles

//for determiners and pronouns
def:{+},             // definite
indef:{+},           // indefinite
dem:{+},             // demonstrative
pers:{+},            // personal (pronoun only)
poss:{+},            // possessive
quant:{+},           // quantifying ("viel")
recp:{+},            // reciprocal ("einander")
refl:{+},            // reflexive ("sich")
rel:{+},             // relative
predet:{+},          // pre-determiner ("solch [ein ...]")
wh:{+},              // interrogative or relative

//for prepositions
preprep:{+},         // e.g. "bis" before "auf", "zu", "in", ...
crc:{+},             // e.g. "[von morgen] an"
padv:{+},

//for conjunctions:
coord:{+},
coord1:{+},
coord2:{+},
subord:{+},
infzu:{+},
sadv:{+},

// for numerals
card:{+},
dig:{+},
dec:{+},
roman:{+},
ord:{+},
fract:{+},

// for nouns
common:{+},
lang:{+},
day:{+},
month:{+},
time:{+},
date:{+},
prod:{+},
title:{+},
unit:{+},
nadj:{+},           // derived from adjective           ::[der] Kranke

// for proper nouns:
proper:{+},
geo:{+},

```

```

continent:{+},
country:{+},
lake:{+},
region:{+},
river:{+},
sea:{+},
state:{+},
town:{+},
name:{+},
  firstname:{+},
  lastname:{+},
bus:{+},
org:{+},
doc:{+},
part:{+},

premod:{+},
postmod:{+},

// particles,
  neg:{+},

// punctuations:
  comma:{+},          // ","
  sent:{+},           // sentence final punctuation ".", "!", "?"
  semicolon:{+},      // ";"
  colon:{+},          // ":"
  dash:{+},           // "-"
  quote:{+},          // quotation mark
  pleft:{+},          // opening parenthesis
  pright:{+},         // closing parenthesis
  quest:{+},          // "?"
  exclam:{+},         // "!"
  sent_sep:{+},       // separator that may be followed by capital letter
  paren:{+},
  left:{+},
  right:{+},
  skip:{+},           // punctuation to ignore (dash, quoate, ...)

// symbols:
  curr:{+},

// OTHER:
  colloq:{+},         // colloquial
  arch:{+},           // archaic, obsolete
  abbr:{+},
  acron:{+},
  elid:{+},
  elid_s:{+},
  email:{+},
  expr:{+},
  foreign:{+},
  init:{+},
  letter:{+},
  symbol:{+},
  sgml:{+},
  url:{+},
  fuge:{+},
  pronom:{+},
  polite:{+},
  vpref:{sep,unsep},
  req_pref:{+,-},

// added in lexicon_tagger.xip:
  prefadv:{+},

```

```

// caractere, symbols, etc.
  upp:{+},
  low:{+},

// mark first and last word in a sentence
  firstword:{+},
  lastword:{+},

// filter for words that should not appear at the beginning
  notfirst:{+},
// other filters
  verbal:{+,-},

// XIP internal feature for capitalized words
  cap:{+},
  allcap:{+}cat:{0},

```

in: features.xip :

```

//for passive subject
  pass:{+},

// added in lexicon_v2.xip:
  imperso:{+},
  novmod:{+},

// added for categories:
  np:{+},
  pp:{+},
  ap:{+},
  bg:{+},

// added for entities
  human:{+,-}

```