

Manual for particle verbs dataframe

Jason Grafmiller

Last modified: April 6, 2018

1 Variety

The ICE component of the token

2 TokenID

Unique identifier for the individual token

3 FileID

ID of the ICE corpus file in which the token was found

Format: VARIETY:FILENAME

4 TextID

ID of the ICE corpus text in which the token was found. Individual files in ICE can have multiple texts

Format: VARIETY:FILENAME:TEXTNUMBER

5 LineID

ID of the line in the text in which the token sentence was found.

Format: VARIETY:FILENAME:TEXTNUMBER:LINENUMBER

6 SpeakerID

ID of the speaker of the sentence. Speakers in spoken texts are indicated with capital letters. Authors of written texts have no IDs (use TextID).

Format: VARIETY:FILENAME:TEXTNUMBER:SPEAKERID

7 Corpus

The corpus the token was taken from.

Levels: ICE, GloWbE

8 GenreFine

The 14-level ICE or GloWbE sub-register in which the token was found.

Levels: See ICE documentation

9 Register

The custom register level of the observed token crossing formality with mode

Levels: 'spok.informal', 'spok.formal', 'writ.informal', 'writ.formal', 'online'

10 Mode

The mode ('spoken', 'written') of the token.

11 comment

General comments about token.

12 SentencePlain

The plain text version of the sentence containing the particle verb token.

13 WholeConstructionPlain

The plain text version of the VP containing the particle verb token.

14 Verb

The lemma of the verbal head (*pick* in *pick up*)

15 Particle

The particle

16 VerbPart

The full particle verb lemma

17 DirObjPlain

The plain text version of the direct object NP

18 DirObjHeadPlain

The plain text version of the direct object NP head

19 Response

The variant order

Levels: 'continuous', 'discontinuous'

20 DirObjWordLength

Length of direct object NP in words

21 DirObjLettLength

Length of direct object NP in orthographic characters

22 DirObjExprType

Syntactic category of the direct object NP

Levels: 'nc' = common noun; 'np' = proper noun; 'iprn' = impersonal pronoun; 'vp' = gerund (-ing)

23 DirObjDefiniteness

Definiteness of the direct object NP

Levels: 'def', 'indef'

24 DirObjGivenness

Givenness of the direct object NP

Levels: 'given', 'new'

25 DirObjConcreteness

Givenness of the direct object NP

Levels: 'Concrete', 'Nonconcrete'

26 Semantics

Semantic compositionality of the verb-particle

Levels: 'compositional', 'non-compositional'

27 DirectionalPP

Is there a postmodifying directional PP present?

Levels: 'yes', 'no'

28 DirObjThematicity

Normalized frequency of direct object head lemma in its text (per 2000 words)

29 PrimeType

The response type of the preceding particle verb, if any

Levels: 'Continuous', 'Discontinuous', 'none'

30 VerbFreq

Frequency of verb lemma in ICE/GloWbE

31 PartFreq

Frequency of particle lemma in ICE/GloWbE. Includes all uses of the wordform, e.g. **all** uses of *up* in ICE/GloWbE

32 VerbPartFreq

Frequency of Verb-Particle lemma combination in ICE/GloWbE

33 PartOnset

The segmental type of onset of the particle

Levels: 'C' = consonant, 'V' = vowel

34 CValternation

The segmental pattern of the final segment of the verb and initial segment of the particle. E.g., *pick up* = CV, *put down* = CC, *throw down* = VC, and so on.

35 VerbForm

The inflected form of the verb as found in the text

36 VerbFormStress

The lexical stress pattern of the verbform syllables.

Notation: 'P' = primary stress; 'U' = unstressed; 'S' = secondary stress

37 PartStress

The lexical stress pattern of the particle syllables.

Notation: 'P' = primary stress; 'U' = unstressed

38 VerbPartStress

The lexical stress pattern of the verbform and particle syllables in the continuous order.

Notation: 'P' = primary stress; 'U' = unstressed; 'S' = secondary stress; '/' = word boundary

39 Clash

Does the continuous variant result in a stress clash across the verb-particle boundary (*conNECT UP*)?

40 ContRhythm

Degree of divergence from optimal rhythmic alternation (PUPU...) from the verb to the next word in the continuous variant. 0 = optimal rhythm.

41 DiscontRhythm

Degree of divergence from optimal rhythmic alternation (PUPU...) from the verb to the next word in the discontinuous variant.

42 Surprisal.P

Shannon information of the particle given the verb (see annotation guidelines for more details).

Calculation: Log inverse of the conditional probability of the particle (p) given the verb (v), i.e.

$$I(p|v) = -\log_2(P(p|v)) = -\log_2\left(\frac{\text{VerbPartFreq}}{\text{VerbFreq}}\right) \quad (1)$$

43 Surprisal.P

Shannon information of the particle given the verb (see annotation guidelines for more details).

Calculation: Log inverse of the conditional probability of the particle (p) given the verb (v), i.e.

$$I(p|v) = -\log_2(P(p|v)) = -\log_2\left(\frac{\text{VerbPartFreq}}{\text{VerbFreq}}\right) \quad (2)$$

44 Surprisal.V

Shannon information of the particle given the verb (see annotation guidelines for more details).

Calculation: Log inverse of the conditional probability of the particle (p) given the verb (v), i.e.

$$I(v|p) = -\log_2(P(v|p)) = -\log_2\left(\frac{\text{VerbPartFreq}}{\text{PartFreq}}\right) \quad (3)$$