Guidelines for selection and annotation of interchangeable particle verbs

Jason Grafmiller

Last modified: May 28, 2018

Contents

1 Introduction

This document contains notes and comments on the extraction, selection, and annotation of interchangeable particle verbs, as used for the project "Exploring probabilistic grammar(s) in varieties of English around the world" (EPG).[1] Our aim is to provide a detailed account of the methods used for collecting the data in the EPG project, as well as a more general set of guidelines for future research on particle verbs in English.

For the project, we collected all instances of transitive particle verbs containing one of the 10 particles listed in (1).

(1) Particles: *around, away, back, down, in, off, on, out, over, up*

This list is based on the method used by Gries (2003:67-68) who selected the 10 verbal heads and 10 particles with the greatest combinatorial potential, based on his list of 1357 transitive particle verbs (pp. 203-210). Since the EPG study is based on data from numerous varieties, searches were not restricted by verbal head, in order to capture uses that may be idiosyncratic to specific varieties.

1.1 Data extraction

Data were extracted from the CLAWS7 tagged version of the ICE corpora using regular expression searches for tokens matching all the possible verbal heads followed by any of the particles listed above. In order to maximize recall, the initial regex searches were quite simple, matching any lexical verb followed by one of the particles listed in (1), within a 10-word window.

(2) Regex search term (note that . . . represents all the remaining particles):

```
\w+_VV[A-Z0-9]\s(\w+_\w+\s){0,10}?(around|away|...)_.*?\b
```

Precision is rather low with this method, since many uses of the forms in (2) found this way will not be true particles, but false positives are less of a concern than false negatives.

Two concerns motivated us to restrict the data to only these particles. First, automatic POS tagging of particles is quite unreliable, as machine taggers have difficulty distinguishing particles from

[1] `http://wwwling.arts.kuleuven.be/qlvl/ProbGrammarEnglish.html`

prepositions and adverbs, and many non-particles are often tagged as such. POS-tag based extraction for PVs thus fares poorly in terms of both precision and recall. While more sophisticated methods for automatic extraction of PVs have been developed (see e.g. Baldwin 2005; Kim & Baldwin 2010), these tools are usually trained on datasets of written standard varieties of British or US English, such as the Wall Street Journal Corpus. The success of these algorithms on PV extraction in other regional and/or non-standard varieties remains to be seen, and so we opted for a more careful semi-automatic approach here.

Recall improves with basic wordform searches, but the results of such searches require extensive manual post-filtering to remove the many false positives, i.e. uses of *around*, *away*, etc. that are not true particles. The extensive time demands of such manual filtering was our second concern, hence the decision to limit our data to these most frequent particles. We believe this method is sufficient to capture the bulk of PV usage in the varieties we study here, but we acknowledge that this procedure could potentially overlook novel and/or region specific usages of other particles. We leave a more fine-grained investigation of individual PVs within specific varieties to future research.

2 Selection of interchangeable tokens

After the initial extraction of the data, the easiest way to refine the set of interchangeable tokens is to identify and exclude tokens occurring in contexts which do not allow one alternate or the other. Various types of unsuitable tokens must be filtered out, either automatically or manually.

The following is a list of criteria for excluding certain classes of tokens. More challenging tokens not fitting into any of these classes should be considered carefully (see section 2.11).

2.1 Intransitive tokens

Obviously, intransitive particle verbs (3) are excluded as they have no direct object to begin with.

(3) a. On their way home, they **called in** at Port Ross, . . .

b. how’s he **getting on** anyway.

> *Note:* In automated searches, these can often be easily identified by the presence of punctuation, a conjunction, or another preposition immediately following the particle, provided the particle immediately follows the head.

2.2 Passive sentences

Passive uses of particle verbs (4) are excluded, since we cannot tell which order a speaker would have used in the corresponding active construction.

(4) a. But the idea is that the crude birth rate should be **brought down** from thirty point four per thousand …

b. lose to a spectacular pavement of Alexandrian mosaic **brought back** by crusaders, …

> *Note:* With tagged/parsed corpora, this process can be semi-automated by excluding past participle forms of the verb (e.g. those tagged with 'VVN'). However, be aware that most corpora are tagged automatically, and there are often many mistagged *-ed* verb forms. An alternative strategy is to search for cases where the particle is immediately followed by *by* (4b), but these are only a minority of passives.

2.3 Extracted direct objects

Tokens where the direct object is either an extracted *wh-* word (5) or a relative pronoun (6) are excluded. As with passive uses, we cannot be sure what the order would have been had the object been in its "basic" position.

(5) a. What did Sansa **pick up**?

b. What did Nasir bring back from the moon?

(6) a. Anyway we bought 2 garden chairs (like that) which he and Sara are **bringing over** next week, probably.

b. well first of all clive the beliefs about god that this survey **throws up**…

2.4 Doubled particles

Cases where the speaker uses the particle after the head, and then repeats it again after the direct object, should be excluded.

(7) a. And Brian my brother **take off** the crocodiles **off** them

b. I don't think they appreciated it cos they **turned up** theirs **up** louder which wasn't very helpful really.

c. They weren't quick enough to be able to write a story about it but **put in** the picture **in** anyways.

These tend to show up only in spoken data.

2.5 Modified particles and adverbials

Cases where there is an adverb immediately before the particle should be excluded as they are not interchangeable (e.g. Bolinger 1971; Fraser 1976; Gries 2003). Most often, these are adverbs such as *right* or *straight* (8), but such examples can really involve any adverbs (9).

(8) a. **threw** it *right* **down**

b. like hodds had like **put** the window *half* **up**

c. My daughter really enjoys dunking, and the chute that **sends** the ball *straight* **back** to you is a great feature.

(9) a. He loosened his fingers carefully so that they might not snag the sheer chiffon and **laid** it *gently* **down** on the corner of the bureau.

b. **Brought** his hand *slowly* **down** upon the knob.

c. You're so afraid of that old nature that you had that you want to **put** it *totally* **down**

d. She reaches up, takes off his hat, and **tosses** it *casually* **away**.

e. He **tilted** his head *slowly* **back** to stare at the hot, hard sky.

2.6 PP arguments and adjuncts

Some verbs, e.g. *put*, are associated with argument structures involving both a direct object as well as an obligatory PP argument (10), while many other verbs can take PP modifiers (11). Note that the examples in (10) are different from interchangeable examples, e.g. *put (your coat) on (your coat)*. Examples like in (10) and (11) are excluded.

(10) a. **keep** the beams *on this side*

b. **put** the book *on the table*

c. Her very dark complexion—very dark even for one of Indian ethnic origin—**threw** a mystic air *around her*.

(11) a. He **threw** it (*over the fence*)

b. Uh just **take** a stroll (*down the wo probably the boardwalk*).

c. You didn't **get** the instruction (*in your class*).

With a discontinuous alternate, a dead give away of a PP argument is that the particle is immediately followed by a determiner or pronoun. On the other hand, PP arguments with continuous uses can be trickier to detect, and usually require manual checking. These are differences of the kind illustrated in (12).

(12) a. The lift was out of order, so he decided to **run up** the stairs [NOT interchangeable]

b. You do have to be firm with lawyers to not send endess letters back and forth and **run up** the bill. [interchangeable]

2.7 Prepositional verbs

Care should be taken to distinguish between cases where words like *on* or *up* should be treated as a preposition, as in (13), or as a particle, as in (14).

(13) He **called on** the U.S. to end its economic embargo of Cuba.

(14) She **switched on** the powerful docking lamps.

In (13), *on* is part of the PREPOSITIONAL VERB *call on*, while in (14), it is part of the particle verb *switch on* (see discussion in Quirk et al. 1985: Ch. 16; Biber et al. 1999: Ch.5)).

Despite their surface similarities with "true" particle verbs, prepositional verbs are NOT interchangeable. This is the most notable distinction between the two classes, however, because we are interested in interchangeability itself, excluding tokens as prepositional verbs based solely on their perceived non-interchangeability makes the selection process circular. Fortunately, there are a number of independent ways to distinguish the two (Bolinger 1971: ; Cappelle 2005:78-81; Rodriguez-Puente 2013:68-95). I discuss the most reliable tests below, which can be used when the interchangeability of a token is in question. Since these test are (arguably) independent of interchangeability, they should be the primary criteria for inclusion of a particle verb token. Some common prepositional verbs are listed in (15).

(15) **Common prepositional verbs**

come back/in/on/off/out
get in/off/on/over
go in/on/out of/through/up
jump in/on/off/out of
keep on
look at
marvel at
work at/on

2.7.1 Pronoun test

A reliable indication of a prepositional verb is that it is perfectly acceptable to have a pronominal direct object (16a), while this is not the case with particle verbs in the continuous order (16b). This is essentially a modification of Bolinger's (1971:112) "definite NP test", and there are arguably instances in which the continuous order is acceptable with a pronominal direct object. More importantly though, the discontinuous order is never allowed with prepositional verbs (17).

(16) a. even Turkey, once one of Syria's closest allies, has **called on** it to institute reforms.

b. *She walked over to the lamp and **switched on** it.

(17) *even Turkey, once one of Syria's closest allies, has **called** it **on** to institute reforms.

2.7.2 Intervening adverb test

Only prepositional verbs allow adverbs between the head and preposition in the continuous order.

(18) a. **calling** repeatedly **on** the miracle worker for help

b. ***switching** repeatedly **on** the light

2.7.3 Preposed PP test

Only prepositional verbs allow pied-piping of the preposition and its argument.

(19) a. **On** whom are you going to **call**?

b. ***On** what are you going to **switch**?

Both are acceptable with stranded particles/prepositions however.

(20) a. Are those the people **on** whom the mayor **called** to save the city?

b. *Is this the lamp **on** which the teacher **switched** to brighten the room?

2.7.4 *it* cleft test

Only prepositional verbs allow *it* clefts with focused PPs (21).

(21) a. It was **on** the Ghostbusters that the mayor **called** to save the city.

b. *It was **on** the lamp that the teacher **switched** to brighten the room.

Be aware that this is NOT the case with focused objects of stranded prepositions/particles, which are acceptable in both cases (22).

(22) a. It was the Ghostbusters who the mayor **called on** to save the city.

b. *It was the lamp that the teacher **switched on** to brighten the room.

2.8 Phrasal-prepositional verbs

These are particle (phrasal) verbs that, according to Thim (2012: 28-29), take prepositional phrases as their complements rather than NPs.

(23) a. I just can't **keep up with** the language, Mrs Sangster said. <ICENZ:W2F-016>

b. Now, if you will excuse me, I must **look in on** father. <GLOWBE:CA>

c. The cancer **caught up with** him in 1980 ... <GLOWBE:JM>

There are also other verbs that can take direct objects, such as those in (24).

(24) a. but some doctors don't even **let** parents **in on** the whole truth.

b. but if you're one of these taxi drivers I've heard about who **fixes** foreigners **up with** prostitutes...

In most cases, phrasal-prepositional verbs should NOT be included. Some frequent NON-interchangeable verbs are listed in (25).

(25) **Common phrasal-prepositional verbs**

catch up with/on
fix up with
get on with
keep up with/on
let in on
look in/down on
put up with

Some verbs are interchangeable however, such as *mix up with*, where you have variation between [*mix* X *up with* Y] and [*mix up* X *with* Y].

(26) a. Until you **mix up** science **with** spirituality, you will never understand the truth.

b. The "Gospel of Prosperity" has **mixed** Jesus **up with** money.

2.9 Misc. types to be excluded

A host of fixed or otherwise unusable expressions specific to individual verbs, heads, or particles need to be filtered out as well. Some of these are described below.

2.9.1 Titles, names, and quotations

Tokens involving titles and names should always be excluded, as these do not alternate (27). The same goes for quotations, song lyrics (28), etc., which should also be excluded.

(27) I’ve been speaking with John Keenan he’s coordinator of “**Take** Me **Out** to the Ball Game” . . .

(28) **Take out** the papers and the trash
Or you don’t get no spendin’ cash
– “Yakety Yak”, The Coasters

2.9.2 “*way*” constructions

There’s a common construction [V X*’s way*], generally known as the *way*-construction (Jackendoff 1990), which introduces a kind of generic path argument with verbs of motion, as in (29), and possess a number of idiosyncratic formal and semantic properties (see Perek 2016 for review).

(29) a. Start at the ears and **make** your way **down**.

b. The next morning we woke up, had a quick (and small) breakfast, and **walked** our way **down** to the ferry from Stavanger to Tau!

These should be excluded, as they are not interchangeable (31). These are discontinuous tokens where the direct object contains the head noun *way* with a possessive determiner, usually pronominal.

(30) a. *Start at the ears and **make down** your way.

b. *. . . and **walked down** our way to the ferry from Stavanger to Tau!

2.9.3 Coding your ass off

Cases involving the kind of metaphoric intensification like in (31) are also excluded from the dataset. These are tokens of the general frame [V X*’s* Y *off/out*], where the head noun Y usually refers to some body part associated with the action described by the verb. Other common uses involve a small set of objects including *heart, ass,* and *butt* (32).

(31) a. I **laughed** my head **off** at Mary’s joke.

b. Justin Bieber apparently **cried** his eyes **out** after being arrested

c. **Typed** my fingers **off** this morning!

d. Babies were **screaming** their lungs **out** as rats went scurrying everywhere.

(32) a. John Small **plays** his little heart **out** for you.

b. I **work** my ass **off** all day, and these goddamned hippies close down the Brooklyn Bridge so I can't get home? That ain't right!

c. Kevin Tobin is **working** his butt **off** doing the illustrations and here is a sample of what you can expect.

They cannot be used in the continuous order (33), so do not count as particle verbs.

(33) a. *John Small **plays out** his little heart for you.

b. *I **work off** my ass all day…

c. *Kevin Tobin is **working off** his butt

Note that this does not include tokens like *worked the baby weight off*, which are interchangeable, and so should be kept in the dataset.

2.9.4 Clausal complements

Some verbs, most notably *point out, find out* and *turn out* can take either NPs or clauses, with (34) or without (35) a complementizer, as complements. These always occur in the continuous order.

(34) a. Slowly she **found out** that only this process had the potential solution for all human problems…

b. **Find out** whether that set you're considering really has the features that matter.

c. She **pointed out** how most wars are fought over limited natural resources, such as oil…

(35) a. After I **found out** my second son was autistic…

b. More astute observers than ourselves **pointed out** we were too dismissive of him in our pre-season analysis.

c. …**turns out** she was making wedding plans for Friday…

I have searched GloWbE and Google for examples of discontinuous uses of these verbs with sentential complements, and I could not find any. Examples of such verbs with sentential complements can thus be ruled non-interchangeable and therefore are excluded from the dataset. Uses of the verbs with NP direct objects should be included however.

(36) a. They have been wonderful as they worked with us to **figure out** a solution.

b. We are still trying to **find out** the cause of the food poisoning.

(37) a. . . . recklessness could have been a disadvantage to humans with their larger mental capacity to go away and **figure** a problem **out**.

b. You might be able to **find** the answers **out** at open days, but it helps to have someone tell you in advance what questions you should be asking.

2.10 Special cases to be *INCLUDED*

The following sections outline some common examples of tokens which may not seem interchangeable at first, but do in fact vary, and so should be included in the dataset.

2.10.1 'time'-away constructions

Examples of what Jackendoff (1997) calls the 'time'-*away* construction can be found in both the discontinuous (38) and continuous (39) variants. These should be retained.

(38) a. On Sundays, I would just sit in my room, enjoying the solitude and **drinking** the day **away**.

b. **Party** the morning **away** at Namur 'breakfast club'.

c. I used to cringe at the thought of them hearing me **snore** the night **away**.

(39) a. I'm not feeling too great so I'm gonna be **drinking away** the day.

b. I knew that I would **sleep away** the morning.

c. Nor did she show in the least any interest in magical researches, rolling herself in her coverings to **snore away** the night.

2.10.2 Interchangeable idioms

There are a number of expressions that might seem at first blush to be fixed idioms, but are actually interchangeable. These include expressions like *throw in the towel* (40), *put up a fight* (41), or *put on weight* (42).

(40) a. After four days of wrangling, Washington **threw in** the towel. <GLOWBE:GB>

b. After five years of frustrations Jayne **threw** the towel **in** and sold the home to Sines for 15,000. <GLOWBE:GB>

(41) a. It **put up** a fight and didn't die quickly, but rather, became enraged with its captors. <GLOWBE:CAN>

b. you don't want to **put** a fight **up** close to your wedding.

(42) a. World sailing champion Darren Choy, 16, is trying to **put on** weight to help him sail Olympic class dinghies. <GLOWBE:SG>

b. More lambs should survive the spring, while good grass growth meant they would **put** weight **on** quickly. <GLOWBE:NZ>

You should be very careful with idiomatic expressions. While semantic compositionality does indeed correlate with the use of the continuous order, it does not guarantee it. We advise always checking for alternate variants in independent sources (See section 2.11).

2.11 Checking intuitions

Ideally, reliance on subjective assessments of interchangeability should be minimized, however the status of specific instances not falling into any of the above categories can be difficult to determine even for native speakers of English. When manually filtering the data, you should keep several things in mind:

1. It can be difficult to distinguish between what is "possible" and what is merely very unlikely in a given context. When consulting your intuitions, sometimes it is helpful to change features of the direct object slightly, e.g. replacing a long and/or complex direct object with a shorter one, or by making the direct object (in)definite.
2. What is categorical in one variety of English, may be variable in others.
3. If you are uncertain about a token, it is advisable to check for alternate instances of a token using online corpora, e.g. BNC, COCA, GloWbE, or Google Advanced searches, which can be restricted to English language sites from particular regional domains (e.g. google.in for India).

For example, you might be uncertain about the interchangeability of the particle and the complex direct object *the thankless job...* in (43).

(43) a. It was Britain's ambassador Sir David Hannay who **took on** [the thankless job of explaining why Washington and London would have nothing to do with the proposal from Paris]$_{DirObj}$ <ICE-GB:S2B-012>

b. ??It was Britain's ambassador Sir David Hannay who **took** [the thankless job of explaining why Washington and London would have nothing to do with the proposal from Paris]$_{DirObj}$ **on**.

You might consider the simpler case where the direct object is much shorter, e.g. just "the job", and see how your intuitions change.

(44) It was Britain's ambassador Sir David Hannay who **took** the job **on**.

You could further check the possibility of the alternate (discontinuous) order by searching for a simplified version of the direct object with the same head. In this case, GloWbE searches for the alternate order, “[take] the job on”, found 40 some hits, and Google searches returned still more. A good rule of thumb is if 5 or more hits can be found for an alternate construction, the token should be accepted as interchangeable.

3 Identifying the Direct Object

Identifying the boundaries of the full direct object NP is not nearly as straightforward as it seems. Difficulty mostly arises with continuous tokens, where there is no overt indication of the right boundary of the direct object NP. With discontinuous tokens, it can generally be assumed that the material between the verb and the particle constitutes the full direct object NP. In this section we provide some guidelines for identifying/delimiting the direct object NP. In the examples below, the direct object is marked in square brackets.

3.1 Temporal modifiers

Temporal modifiers should not be included as part of the direct object. These include temporal adverbials like *tomorrow* and *last week* (45), as well as PPs headed by *in* or *on* (46).

(45) a. Until a judge **struck down** [the provision] last month,

b. “We **turned back** [the clock] today,” a teary Edwards said afterward.

(46) a. He **took over** [the company] in November of 1999.

b. The town seized the property for non-payment of taxes and **tore down** [the building] in 1937…

c. Tesla is going to **pick up** [the car] on Monday.

3.2 Narrow vs. wide scope locatives

Finding the right edge of a direct object in the presence of a post-modifying locative PP (marked in italics) can be tricky.

(47) a. he **picked up** the phone *in his hotel suite*,

b. He picked his keys up off the table, **turned off** the lights *in the apartment*,

c. She went to the kitchen, found a mop, and spent the next ten minutes **cleaning up** the water *in his bathroom and hall*.

Such modifiers are often ambiguous as to whether they modify the direct object itself (narrow-scope), or the VP as a whole (wide-scope). Wide-scope locatives (48) modify the location of the entire event, while narrow-scope locatives (49) are restictive modifiers that say something about the location of the direct object itself, and not necessarily the location of the event.

(48) **Wide-scope locatives**

a. With the decline of whaling in Cook Strait and the offer of work on Tucker's station a party of Tory Channel whalers under Jack Norton **set up** [a whaling station] *at Northwest Bay on Campbell Island* in 1909.

b. He went into the bathroom after **taking off** [his t-shirt and trousers] *in his room*.

c. Aside from environmental pollution, **putting up** [the aquaculture structures] *in sheltered areas* will likely result in conflict with fishermen who use the areas for fishing as well as sheltering their boats.

d. They were forced to **ride out** [the storm] *in their hotels*.

(49) **Narrow-scope locatives**

a. I **pick up** [the electricity *in the air*, the force fields of handsome men who mill around me], until my stomach feels charred and my hands are shaking.

b. Like you know asking the girl to wear some kind of a grape in front of her of her of her chest and ask the bridegroom to go **pick up** [the grape *in front of her chest*] and something like that.

c. Even I forgot that Auntie Maggie **took up** [two entire seats *at the back of the bus*]. <ICE-JM:WF2-018>

3.2.1 Identifying wide-scope modifiers

As a rule of thumb, wide-scope locatives can be used felicitously to answer the question "where did the event take place?".

- Q: Where did the party of whalers set up a whaling station?
  A: at Northwest Bay on Campbell Island. (cf. (48a))
- Q: Where did he take off his t-shirt and trousers?
  A: in his room (cf. (48b))
- And so on...

Fronting the PP is also often a useful test for wide scope modification. If the PP can be fronted felicitously, it should not be included as part of the direct object.

(50) a. I **threw in** the towel in almost every aspect of my life: I quit my job, sold the car, told my bewildered friends and colleagues I was going to commit suicide, and stopped seeing Edie. <ICE-SG:W2F-012>

b. In almost every aspect of my life, I **threw in** the towel…

(51) a. *In the air, I **picked up** the electricity… [cf. (49a)]

b. *In front of her chest, she **picked up** the grape… [cf. (49b)]

Note: Of-phrases are always considered part of the direct object.

3.3 Restrictive vs. non-restrictive relative clauses

When the direct object involves a relative clause, you must determine whether it is a restrictive or non-restrictive clause.[2] Only restrictive clauses are included as part of the direct object. For written texts, punctuation can often be a useful quide, as non-restrictive RCs are often separated by a comma, however this is not 100% reliable. Absent any punctuation, the status of an RC must be determined based on its function in the sentence.

Prototypically, a restrictive RC serves to restrict the denotation of the head noun it modifies. For example, in (52a), the set of employees who have a green card, is smaller than the set of all employees. The information expressed in the relative clause is thus an integral part of the meaning expressed by the matrix clause in that it delimits the set of employees under discussion. See Quirk et al. (1985:1245-1250) and Huddleston & Pullum (2002:1034-1035) for discussion of restrictive (integrated) RCs.

(52) **Restrictive RCs**

a. What is your experience as to whether or not they **take on** [those employees *who have a green card*]… <ICE-GB:S1B-062>

b. There are a number of employers My Lord who will **take on** [people *who have a disability*]. <ICE-GB:S1B-062>

c. He showed me how to move in slow-motion and how to breathe away, so as to avoid **setting up** [draughts *that a bugoid could sense*]. <ICE-IND:W2F-004>

d. They do not help us make verifiable generalizations about the political behaviour of organized groups, about their relationship to the state, or about their role in political systems, and it is very hard to **set out** [a grand theory *that does not include such generalizations*]. <ICE-CAN:W2A-018>

[2] Restrictive RCs are also called ‘integrated’, ‘defining’, or ‘identifying’ relative clauses. Non-restrictive RCs are called ‘supplementary’, ‘appositive’, ‘non-defining’, or ‘non-identifying’ relative clauses.

Non-restrictive RCs add supplementary information about their antecedents, information that is not needed to delimit or restrict the set(s) denoted by the head noun. In (53a) the RC *that tested the soil…* does not play a role in identifying a unique set of landers, rather it only functions to provide some extra detail about them. The supplementary nature of the RCs in (53) is reflected in the fact that the removal of the RCs does not change the identity of the antecedents, unlike in (52). By virtue of their unique identification, proper nouns are only ever modified by non-restrictive RCs. See Quirk et al. (1985:1257-1260) and Huddleston & Pullum (2002:1035) for discussion of non-restrictive (supplementary) RCs.

(53) **Non-restrictive RCs**

a. Viking one and two went to Mars orbited the planet mapped it then **sent** [landers] **down** on the surface *that tasted the soil looked for life took full colour stereo pictures*. <ICE-CAN:S2B-007>

b. Blarney Woollen Mills that successful group for instance *took on* [its first graduate] four years ago *who had taken HR in her final year from Professor Walsh who left after a year because there was a lot of hostility to graduates in that organisation* just to mention one small case example.

c. **Set up** [corps of Park Rangers] *who will enforce regulations through education whenever possible*. <ICE-JM:W2B-027>

Note: While non-restrictive RCs are rarely introduced by *that*, the relativizer itself should not be taken as diagnostic of RC status. Despite what is commonly believed, *that* can be used with non-restrictive RCs, as (53a) shows, and *which* can be used with restrictive RCs (*a date which will live in infamy*).

4 Annotation of linguistic features

The following factors were identified for each token in the dataset.

4.1 Animacy and concreteness of direct object

Following Wolk et al. (2013), **Animacy** is coded in five levels, with separate columns indicating the animacy level of each constituent.

Animacy

Code	Category	Comments	Examples
‘a’	Human & animal	Only higher animals (not e.g. ‘fish’ or ‘bugs’); includes spirits, god(s), and other agentive (human-like) supernatural entities	*Shakespeare, engineers, the horse, a sixteen-year-old girl, Mr. Kennedy, God*
‘c’	Collective	Organizations or political states/bodies when seen as having collective purpose, agenda or will	*the House of Lords, the church, parliament, another country*
		Group of animate individuals with potential variable anaphoric reference (*it/they*)	*family, multitudes, the public, a convoy, the majority*
‘i’	Inanimate	Non-temporal, non-locative inanimates: concrete and abstract, all gerunds, participles, and infinitives	*the table, oxygen, other topics, drinking*
‘l’	Locative	Places qua places, not groups of inhabitants/members, including *state/empire*; not referable by *they*	*the sea, the playground, China, the earth*
‘t’	Temporal	Noun or adverb with time reference	*yesterday, last week, March, 1986, this morning*

Concreteness was coded as a binary category cross-cutting animacy, were any referent that could be experienced with one of the five senses was coded as ‘concrete’. This included all animate (‘a’) DOs, as well as a subset of those DOs coded as inanimate (‘i’). All DOs not coded as concrete were coded as ’nonconcrete’.

Concreteness

Code	Category	Comments	Examples
‘concrete’	Concrete referent	Anything that could be experienced with one of the five senses	*his shoe, my lapel pin, the machine gun, your knee*
‘nonconcrete’	Non-concrete referent	Anything not coded as concrete	*the investigation, histories of American desires, Canada*

4.2 Definiteness of direct object

Definiteness of the direct object was coded according to the following scheme:

Definiteness

Code	Category	Comments	Examples
'def'	Definite NP	Proper nouns and any of the NP types listed in section 4.2.1	*his shoe, the polls, myself, all my money, what you don't want*
'indef'	Indefinite NP	Any of the NP types listed in section 4.2.2	*a new language, people, some elderflower cordial*

The following subsections provide explicit guidelines for annotating definiteness.

4.2.1 Definite NPs

The following are all the types of NPs that should be coded as 'def' (see Garretson et al. 2004)

- Proper nouns (see section 4.5.1)
- NP with a definite determiner
  - Articles: *the*
  - Demonstrative: *this, that, these, those*
  - Possessive: *her, his, its, my, our, their, your*
  - Quantifier: *all, both, each, either, every, most, neither*
- Definite Pronoun
  - Personal: All, including reflexives and possessives (*mine, hers*, etc.)
  - Impersonal: *each other, everybody (else), everyone (else), everything (else), one another*
  - Wh-pronouns: *which, who, whatever, whatsoever, whichever, whoever, whosoever, whosever*
- An *s*-genitive NP (*George Clooney's bushy beard*)
- Superlatives (*the sourest beer imaginable*)
- Temporal expressions
  - years (*1993*)
  - dollar amounts (*$179000, $20 million*
  - *today, yesterday, tomorrow*
  - *last* or *next* followed by *night, week, month, year*, or any noun referring to a specific day or period of time (e.g. *Easter, March, winter, term, Sunday*)

4.2.2 Indefinite NPs

The following are all the types of NPs that should be coded as 'indef' (see **?**).

- NP with an indefinite determiner
  - Articles: *a, an*
  - Quantifier: *another, any, enough, few, fewer, half, less, little, little or no, lots of, many, more, much, no, no more, no such, none one-half, one-third (. . .), one, one or more, ones, plenty of, several, some, twice*
- Indefinite pronouns
  - *any one (else), anybody (else), anyone (else), anything (else), no-one, nobody (else), nothing (else), one's, oneself, somebody (else), someone (else), something (else)*
- Bare plural NPs
- Numbers that are not years or monetary amounts
- Gerunds NOT headed by definite determiners ("screaming" in *the cause of the baby's screaming* is 'def'; but "drinking" in *gave up drinking* is 'indef')
- Any determinerless noun ending in *-tion, -ment, -sion, -ology,* or *-ism*

4.3 Information status of direct object

While there are many possible degrees of discourse accessibility, or 'givenness', that could be explored, only two levels of givenness are coded for all three constructions, based on the work by **?**: 249.

Givenness

Code	Comments
'given'	A constituent is coded as 'given' if its referent is mentioned at any time in the 100 words preceding the token in the discourse, or if it is a 1st or 2nd (or 3rd?) person pronoun
'new'	Any constituent that does not refer to a speech participant, and is not referred to in the preceding 100 words is coded as 'new'

4.4 Thematicity of direct object

Thematicity is measured as the normalized text frequency of the head noun in the direct object, i.e. number of uses of the constituent head word/lemma in a text divided by the total number of words in the text (**?**: 450-451).

4.5 NP type of direct object

To distinguish pronominality (among other things) from the effects of definiteness and givenness, we code the syntactic category of the direct object (head).

The follow classes of NP expression types were coded. Personal pronouns are marked in their own class separate from impersonal pronouns for several reasons. First, while the former are known to behave almost categorically, the latter are not so restricted.

(54) a. Cars come back at end of reception to **pick** everyone **up** and drive them home.

b. The planes come in and **pick up** everyone, . . .

Second, impersonal pronouns can be modified (55), unlike personal pronouns.

(55) Do you **pick up** everyone who hails your cab?

Finally, impersonal pronouns vary with respect to definiteness, unlike personal pronouns which are always definite. The full coding list for NP types is shown below.

NPType

Code	Category	Comments	Examples
'nc'	Common noun	Common noun	*birds, the market, wisdom, this year*
'np'	Proper noun	See section 4.5.1	*President Kennedy, Japan, the United Nations*
'pprn'	Personal pronoun	Personal pronouns, incl. possessives and reflexives.	*me, theirs, yourself*
'iprn'	Impersonal pronoun	Any definite or indefinite pronoun, incl. *wh* pronouns	*everyone, something, whoever*
'dm'	Demonstrative	Bare demonstrative	*this, that, these, those*
'ng'	Gerund	Present participle *-ing* forms (rare)	*give up drinking, hunting's purpose, Give your writing a break*

4.5.1 Proper nouns

It can sometimes be tricky to decide whether a nominal is proper or not. Here is a working test:

- An NP without a determiner (e.g., *Texas*) is proper if it cannot be changed in number or take a determiner (**Texases, *a Texas*). An NP with a determiner (e.g., *the West Indies, A*

Separate Peace) is proper if it cannot be changed in number or lose its determiner (**a West Indy, *go to West Indies*). If number or determiner alternation is possible, it is not functioning as a proper noun, and should be treated as a common noun.

- Nouns that are usually proper can be "coerced" into behaving like common nouns, as in *Do you mean the Washington on the Pacific or the Washington on the Potomac?* or *She wants to be a Shakespeare*. In these sentences, the names *Washington* and *Shakespeare* uncharacteristically occur with a determiner, and we therefore say that in this case, they are being **used** like common nouns, not proper nouns. We code such proper nouns used like common nouns as common nouns, since it is actual instances of usage that we are concerned with.

4.6 Length of the direct object

Two measures of length are used: length in orthographic words, and length in orthographic letters (graphemes). A few points about these counts are worth noting:

- For grapheme counts, spaces are included, while all punctuation is excluded.
- Hyphenated compounds are counted as 2 words, and hyphens are ignored when counting graphemes. Contractions are counted as 1 word.
- Different texts may use acronyms (*NASA, NATO*) and initialisms (*U.S.S.R., the U. N.*) in different ways, i.e. with or without full stops (periods) and with or without spaces in between characters. We considered all acronyms and initialisms to constitute 1 word only, regardless of spacing. Similarly, the word length of numbers, e.g. *1999, flight 93*, is counted as 1 word.
- Discourse markers (e.g. *of course, I think, like, um*), are counted as part of the relevant constituent in which they occur.
- Different VoEs use different spelling conventions. Some of these are inconsequential for measuring length (*analyse* vs. *analyze*), while others can potentially affect the resulting measurements (*doughnut* vs. *donut*). We did not correct for variation in spelling across varieties.

4.7 Persistence/priming

For each token, we coded a persistence measure (**?**) based on the PV variant used in the previous choice context (A or B), or 'none' when there is no preceding construction in the 100 words prior to the target construction.

For spoken dialogues, persistence is coded within and across turns, and within and across speakers. The first construction in each conversation or text is automatically to be coded as 'none'. It is important to consider only preceding tokens found in genuine choice contexts, ignoring any occurrences that have been excluded from the analysis.

4.8 Presence of directional PP

Gries (2003) shows that that the presence of a directional PP significantly increases the likelihood of the split order.

(56) a. This creates electrical potential which, after it reaches a certain amount (threshold) causes the cell to **send** a spike **down** *to its own axon*, and thus onto other neurons.

b. . . . it has what it takes to **bring** the country **back** *from the brink of economic collapse*.

c. how do you how do you **cut out** shoe leather *from the various hides* and how do you make shoes.

Prepositional phrase adjuncts fall into several classes.

- **locative PPs**:
  describe the location in which the event takes place (e.g. *in, on, at*)
- **directional PPs**:
  describe the direction the DirObj is moved to/from (e.g. *into, onto, from, to, toward, off (of)*)
- **instrumental PPs**:
  describe the means or manner in which the action is carried out. (*with*)
- **purpose PPs**:
  describe the reason why the subject carried out the action described by the verb (*to, in order to*)

We are primarily interested in the directional case(s) here.

4.8.1 Directional PPs

See section 3.2 for how do determine whether a PP is an adjunct of the VP or the DirObj.

(57) a. Monday I visited Susan, went to the Gym and **picked up** Tammy *from the airport* and went to watch a movie at Sovereign.

b. Thinking about this, she went into her mother s bedroom, picked up the make-up box from the bedside table and threw it out the bedroom door.

Locative PPs can be used more metaphorically as well.

(58) a. He said Dad was the craziest of all for marrying a monkey, and **throwing away** [his money] *on this big house in Port Dickson*. <ICE-HK:W2F-012>

4.8.2 Dative uses

Dative particle verbs (59) are also included.

(59) a. Features such as remote procedure calls and triggers enable a user using his terminal to issue a command to the database which then independently executes the procedures and **sends back** [the results]$_{theme}$ [to the user]$_{recipient}$.

b. um i think it's one of these cases that i got her to send out an email message to chairpersons and i've received some feedback.

4.9 Collocational measures

The strength of the association between the verb and the particle was measured using a number of association scores, including both symmetric (PMI, t-score, log likelihood, DICE) and asymmetric measures (ΔP and surprisal) (Evert 2008; Ellis 2006; Ellis & Ferreira-Junior 2009; Gries 2013). The inclusion of various association measures is intended to capture low level expectation-based processing effects in production. The stronger the association between the particle and the verb, the more likely the joined order should be.

Collocational measures are based on frequency tables of the kind shown in Table 1.

Table 1: Schematic table of verb and particle co-occurance frequencies.

	particle: present	particle: absent	Totals
verb: present	a	b	$a+b$
verb: absent	c	d	$c+d$
Totals	$a+c$	$b+d$	$a+b+c+d$

4.9.1 Delta P

Delta P (Ellis 2006; Ellis & Ferreira-Junior 2009; Gries 2013) is useful because it is asymmetric, unlike other measures which conflate the direction of association, e.g. *PMI*, t, G^2, and previous work has shown that pairwise associations for the two directions—$verb|particle$ and $particle|verb$—are rarely equivalent, or even close to each other (Gries 2013). With such measures then, we cannot be sure which direction is contributing to a high score. Because the particle placement alternation pertains to the post-verbal ordering of the particle and the object, it seems reasonable to assume that the verb is fixed for the purpose of modeling the probability of one order or the other. Thus we are interested mainly in the probability of the particle given the verb, though the alternative association could also play a role. Hence the use of both asymmetric measures $\Delta P_{particle|verb}$ and $\Delta P_{verb|particle}$. In simple language, for $\Delta P_{particle|verb}$ we are interested in the likelihood of a particle being used given the observed verb, and for $\Delta P_{verb|particle}$ we are interested in the likelihood of a verb being used given the observed particle.

Fortunately, the ΔP score is quite easy to calculate.

(60) a. $\Delta P_{p|v} = \mathrm{P}(particle|verb = present) - \mathrm{P}(particle|verb = absent)$

b. $\Delta P_{v|p} = \mathrm{P}(verb|particle = present) - \mathrm{P}(verb|particle = absent)$

The $\Delta P_{p|v}$ score is calculated from Table 1 as follows (see Gries 2013:143-144):

$$\Delta P_{p|v} = \frac{a}{a+b} - \frac{c}{c+d}$$

$$\Delta P_{v|p} = \frac{a}{a+c} - \frac{b}{b+d}$$

4.9.2 Surprisal

Surprisal is used here in the information theoretic sense, which refers to the log inverse of the conditional probability of a form given some context (e.g. Jaeger 2008). This measure is alternatively referred to simply as *information* (Jaeger 2010).

Two measures are calculated: the surprisal of the particle given the verb (**Surprisal.P**), and the surprisal of the verb given the particle (**Surprisal.V**)

$$\text{surprisal}(p|v) = -\log_2(P(p|v)) = -\log_2\left(\frac{P(v,p)}{P(v)}\right) = -\log_2\left(\frac{a}{a+b}\right)$$

$$\text{surprisal}(v|p) = -\log_2(P(v|p)) = -\log_2\left(\frac{P(v,p)}{P(p)}\right) = -\log_2\left(\frac{a}{a+c}\right)$$

Surprisal is inversely related to predictability, thus higher surprisal scores reflect more unlikely verb-particle combinations. This measure is also very similar to ΔP.

4.9.3 Pointwise Mutual Information

Pointwise mutual information is calculated as follows, where N is the total number of words in the corpus (i.e. $N = a+b+c+d$):

$$\text{PMI}(v;p) = \log_2\left(\frac{P(v,p)}{P(v)\times P(p)}\right) = \log_2\left(\frac{\frac{a}{N}}{\frac{(a+b)\times(a+c)}{N}}\right)$$

4.9.4 Log likelihood ratio (G^2)

The log-likelihood ratio score is popular measure of association. This measure has an advantage as it is more appropriate for sparse data.

It is calculated as:

$$G^2 = 2\sum_{i}^{j} n_{ij} \ln\left(\frac{n_{ij}}{m_{ij}}\right)$$

where n_ij represents the observed frequency of row i column j and m_{ij} the expected frequency.

4.9.5 *t*-score

The *t*-score is another symmetric measure of association that, like G^2, is less biased toward low frequency items than PMI.

$$t\text{−score} = \frac{a - \frac{(a+b)\times(a+c)}{N}}{\sqrt{a}}$$

4.9.6 Dice

From Evert (2008):

> The Dice coefficient focuses on cases of very strong association rather than the comparison with independence. It can be interpreted as a measure of predictability, based on the ratios O_{11}/R_1 (the proportion of instances of w_1 [the verb] that cooccur with w_2 [the particle]) and O_{11}/C_1 (the proportion of instances of w_2 [the particle] that cooccur with w1 [the verb]). The two ratios are averaged by calculating their harmonic mean, leading to the equation in [above]. Unlike the more familiar arithmetic mean, the harmonic mean only assumes a value close to 1 (the largest possible Dice score) if there is a strong prediction in both directions, from w_1 to w_2 and vice versa. The association score will be much lower if the relation between the two words is asymmetric.

While the Dice coefficient cannot be used to identify word pairs with strong negative association, the measure is most useful for identifying fixed expressions, thus it seems somewhat useful here.

The Dice coefficient is calculated as follows.

$$\text{DICE} = \frac{2a}{(a+c)+(a+b)} = \frac{2a}{2a+b+c}$$

4.10 Idiomaticity and verb semantics

For the classification of particle verb semantics, we follow the 3-way distinction (literal vs. metaphorical vs. idiomatic) used by Gries (2003:72; see also Bollinger 1971; Fraser 1976; Quirk et al. 1985), with the addition of a fourth category reflecting aspectual uses of the particle (Dehe 2002; Thim 2012; Rodriguez-Puente 2013).

4.10.1 Literal particle verbs

From Gries (2003):

> A sentence [is] counted as literal...if the meaning of the whole expression [is] totally predictable from the meaning of its parts (which [is] generally equivalent to the referent of the direct object undergoing a change of location in a manner specified by the verb. (72)

In simple terms, a token is only coded as literal if we can say that the object has undergone the spatial movement denoted by the particle. For example, in *She lifted the book up*, we can say that the book is/was 'up', therefore we code "L". On the other hand, with *speed up production* or *filling up the hole*, neither production nor the hole can be said to be 'up' in the spatial sense, thus we code "M". Particles in such literal uses are often replaceable by full prepositional phrases.

Examples:

(61) a. Emilio would **bring out** a small booklet for him to work on.

b. She took the blue pareu out of her bag and wrapped it around her hips, **kicked off** her sandals.

c. And the defendant **pick up** the fruit knife.

4.10.2 Metaphorical particle verbs

Metaphorical verbs have meanings that are still transparent, but are removed from the literal connotation (Rodríguez-Puente 2013:60). Often these will involve compositional "literal" verbs with abstract direct objects, reflecting some metaphorical mapping from a physical spatial domain to a more abstract one, e.g. 'Change of mental state is Change of location': *she slipped into depression* (Lakoff and Johnson 1980).

Examples:

(62) a. What you did about that you know uh **taking away the anxiety** at different times from different groups

b. The problem uhm with alcohol blackouts is that the person is intoxicated and is not able to **lay down memories**.

c. When Tartuffe's hypocrisy is finally exposed, Carver **screams out his hatred** while grovelling on a floor that Goldby has covered with peat moss.

Differentiating metaphorical (uses of) verbs from aspectual (uses of) verbs can be difficult. Like aspectual instances, a sentence can often still make sense without the particle, as in (63).

(63) He's coming under strong pressure from Dobrovolski who's **helping out his defence** . . .

The key difference here is that the inclusion of *out* does not change the aspectual characteristics, specifically the telicity, of the sentence. Consider the usual test for boundedness: *We helped them (out) for a couple hours* vs. **We helped them (out) in a couple hours*. Such cases we code as Metaphorical.

4.10.3 Idiomatic particle verbs

Idiomatic uses are those whose meaning is not predictable from the meanings of the parts alone. Neither the meaning of the verb, nor the meaning of the particle contribute individually to the meaning of the whole. However, it should be kept in mind that idiomatic sentences can involve verbs that are often literal or metaphorical, when those sentences involve specific direct objects (*make up my mind*, *roll up your sleeves*, *throw down the gauntlet*).

Many particle verbs are polysemous, and idiomaticity is something of a matter of degree, so distinguishing a metaphorical sentence from an idiomatic one can be tricky. While we as linguists may be able to devise a metaphorical path along which we could trace the origin of a particular particle verb, it is not necessarily the case that the metaphor(s) involved are currently active in users' minds.

Some examples of idiomatic particle verbs are provided in (64).

(64) a. Sometimes they **ferret out chicks** to determine breeding success.

b. Macdonald **drew up von Schoultz's will**.

c. **Lays off employees**.

d. Canada **shut out Brazil** four nothing yesterday …

4.10.4 Aspectual particle verbs

Although the particle verb is figurative or metaphorical in the sense that it does not imply spatial direction or a literal meaning, neither does the particle "form a figurative semantic unit with the verb" (Rodriguez-Puente 2013). Instead the particle alters the aspectual properties of the verb.

From Elenbaas (2007):

- "for x time" indicates duration (as a test for atelic aspect)
- "in x time" indicates an endpoint (test for telic aspect)

In cases where the particle is redundant, but the meaning is not embedded in the verb, the particle tends to imply the beginning and closing of an action. In other words, certain particles can alter a situation's telicity, converting an action into an accomplishment (Rodriguez-Puente 2013).

Examples:

(65) a. Well they'll finish off their B As at the age of twenty.

b. We drank her water so she’s going to eat up my salad dressing.

c. The management decided to close down the plant. (Brinton 1985)

In all the examples, the sentences could make sense without the particle, yet (65a) implies that their BAs will not only begin the process of being finished but will be fully completed, and in (65c), management not only decided to close the plant, but to close it down completely (Brinton 1985). For our purposes, verbs that could still make sense without the particle and for which the particle seemed to add telicity were rated as aspectual (see the following list).

(66) Some common aspectual verbs:
add up = ‘sum’ (only applies to actual numbers or monetary amounts)
barf up = ‘vomit’ (barf up a lung would be idiomatic)
buy up
chase off/away
chop up
clean off/out/up
cover up
drink up
eat up
finish up/ off
frighten off/away
gather up
pack up
pay off = ‘repay (completely); cf. idiomatic pay off = ‘bribe’
scare off/away
sell off
snuff out = ‘extinguish’

4.10.5 Hawkins’ entailment tests

As an additional aid in classification, we used a heuristic devised by Hawkins (1999) and Lohse et al. (2004) which is based on the notion of ‘lexical dependency domain’. The idea is that “the parser needs to access one category to assign certain semantic and/or syntactic properties to another” (Lohse et al. 2004: 244). The dependency between a verb and particle in a given use is assessed by two entailment tests designed to tease apart semantic compositionality. The first test assesses the independence of the interpretation of the verb (67), while the second test assesses the semantic independence of the particle (68).

(67) Verb entailment test:
If [X V NP Part] entails [X V NP], then the verb is *independent*, otherwise it is *dependent*

(68) Particle entailment test:
If [X V NP Part] entails [NP PredV NP], then the particle is *independent*, otherwise it is

dependent
PredV = predication verb (BE, BECOME, COME, GO, STAY)

These tests lead to four possible verb-particle combinations, ranging from fully compositional/transparent to fully idiomatic: V_I P_I, V_I P_D, V_D P_I, V_D P_D. Fully compositional particle verbs—V_I P_I, as in (69)—meet both entailment tests.

(69) a. Mickey **pulled down** the shades.

b. Donald **brought in** the groceries.

For *Mickey pulled down the shades* or *Donald brought in the groceries*, *Mickey pulled the shades* and *Donald brought the groceries* are both entailed. The relevant entailments also hold of the particles: *the shades* go *down* and *the groceries* go *in*. Full compositional uses are always coded as 'Literal'.

At the other end of the spectrum are particle verbs that fail both tests—both the particle and the verb depend on the other for interpretation, as in (70).

(70) a. Dracula **carried out** his chores.

b. The Pope **barfed up** a lung.

For example, (70a) entails neither *Dracula carried his chores* nor *the chores* are/go/come/stay *out*. Likewise, (70b) does not entail *The Pope barfed a lung* or *a lung* is/goes/comes/stays *up*. And so on. Cases where both the verb and the particle are dependent on each other are coded as 'idiomatic'.

Intermediate cases we coded as 'Metaphorical' or 'Aspectual'. Examples of these would be cases like (71)

(71) a. Napoleon **ate up** all the sushi.

b. Chewbacca **turned on** the stereo.

Napoleon ate up all the sushi does entail *Napoleon ate all the sushi* but not *all the sushi* is/goes/comes/stays *up*, while *Chewbacca turned on the stereo* entails *the stereo* is *on* but not *Chewbacca turned the stereo*. In the first case, the verb is semantically independent and the particle is not, while in the second cases, it is the other way around.

4.10.6 Binary classification

Due to the difficulty of coding semantics reliably, we also adopted a binary system: 'compositional' vs. 'non-compositional'. Tokens coded as 'Literal', i.e. those passing both of the entailment tests, were coded as 'Compositional'. All other cases were coded as 'Non-compositional'.

4.10.7 Full verb list

A complete list of every verb found in the EPG dataset and its semantic coding can be found at `http://tinyurl.com/popg9vn`.

4.11 Phonological features

Two phonological factors were coded for, the segmental patterns at the verb-particle boundary Gries (2011), and the degree to which the given variants diverged from optimal rhythmic alternation Shih et al. (2015); Shih (2017).

4.11.1 CV alternation

This is straightforward to code. We simply looked at the final segment of the verbform and the initial segment of the particle and coded each as either a consonant (C) or a vowel (V). This gives us 4 possible combinations: CC, CV, VC, and VV.

(72) a. *put back* (CC)

b. *pack up* (CV)

c. *pay back* (VC)

d. *pay off* (VV)

4.11.2 Rhythm

To calculate the rhythmicity of a PV token we counted the number of unstressed syllables at the boundary between the verb and particle in the continuous variant (73), and the verb and direct object in the split variant (74).

```
       x   -   |  x    -  x    -
(73)   tak.ing |  out  a  mort.gage

       x   -   |  -  x    -      x
(74)   tak.ing |  a  mort.gage   out
```

The count begins at the last primary stressed syllable of the verb form (x), and continues until the first primary stress on the particle (for the continuous variant) or the direct object NP (for the split variant). So the count in (73) is 1, and the count in (74) is 2.

This resulted in a measure of 'eurhythmy distance' (ED) for each variant, from which we calculated a measure of comparative rhythm by subtracting the split ED from the continuous ED. See Shih et al. (2015) and Shih (2017) for further examples and disussion.

(75) Split preference (Rhythm $<$ 0)

a. *handing out the stockings* [ED = 1]

b. *handing the stockings out* [ED = 2]

(76) Continuous preference (Rhythm $>$ 0)

a. *throw away books* [ED = 1]

b. *throw books away* [ED = 0]

(77) No preference (Rhythm = 0)

a. *picking up gorgeous Samsons* [ED = 1]

b. *picking gorgeous Samsons up* [ED = 1]

Stress patterns for the different varieties were obtained from the Unisyn Lexicon.[3] Where appropriate, pronunciations were adjusted to the closet variety in question, e.g. the General American accent was used as a proxy for Canadian English. Otherwise the British RP pronunciation was used as the model.

[3] `http://www.cstr.ed.ac.uk/projects/unisyn/`

References

Baldwin, Timothy. 2005. Deep lexical Acquisition of Verb–particle Constructions. *Computer Speech & Language* 19(4). 398–414. doi:10.1016/j.csl.2005.02.004.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Bolinger, Dwight. 1971. *The Phrasal Verb in English*. Cambridge, MA: Harvard University Press.

Brinton, Laurel J. 1985. Verb Particles in English: Aspect or Aktionsart? *Studia Linguistica* 39. 157–168.

Cappelle, Bert. 2005. *Particle Patterns in English: A Comprehensive Coverage*. Leuven, Belgium: K.U. Leuven Ph.D. Thesis.

Elenbaas, Marion. 2007. *The Synchronic and Diachronic Syntax of the English Verb-Particle Combination*. Nijmegen: Radboud University Ph.D. Thesis.

Ellis, N. C. 2006. Language Acquisition as Rational Contingency Learning. *Applied Linguistics* 27(1). 1–24. doi:10.1093/applin/ami038.

Ellis, Nick C. & Fernando Ferreira-Junior. 2009. Construction Learning as a Function of Frequency, Frequency Distribution, and Function. *The Modern Language Journal* 93(3). 370–385. doi:10.1111/j.1540-4781.2009.00896.x.

Evert, Stefan. 2008. Corpora and Collocations. In A. Lüdeling & M. Kytö (eds.), *Corpus Linguistics. An International Handbook*, 1212–1248. Berlin: Mouton de Gruyter.

Fraser, Bruce. 1976. *The Verb-Particle Combination in English*. New York: Academic Press.

Gries, Stefan Th. 2003. *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement*. New York: Continuum Press.

Gries, Stefan Th. 2011. Acquiring particle placement in English: A corpus-based perspective. In Pilar Guerrero Medina (ed.), *Morphosyntactic alternations in English: Functional and cognitive perspectives*, 236–263. London & Oakville, CT: Equinox.

Gries, Stefan Th. 2013. 50-Something Years of Work on Collocations: What Is or Should Be next \ldots. *International Journal of Corpus Linguistics* 18(1). 137–166. doi:10.1075/ijcl.18.1.09gri.

Gries, Stephan Th. & Anatol Stefanowitsch (eds.). 2007. *Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis*. Berlin: Mouton de Gruyter.

Hawkins, John A. 1999. Processing Complexity and Filler-Gap Dependencies across Grammars. *Language* 75. 244–285.

Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge Grammer of the English Language*. Cambridge: Cambridge University Press.

Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 1997. Twistin' the Night Away. *Language* 73(3). 534–559. doi:10.2307/415883.

Jaeger, T. Florian. 2008. Categorical Data Analysis: Away from ANOVAs (Transformation or Not) and towards Logit Mixed Models. *Journal of memory and language* 59(4). 434–446. doi:10.1016/j.jml.2007.11.007.

Jaeger, T. Florian. 2010. Redundancy and Reduction: Speakers Manage Syntactic Information Density. *Cognitive Psychology* 61(1). 23–62. doi:10.1016/j.cogpsych.2010.02.002.

Kim, Su Nam & Timothy Baldwin. 2010. How to pick out token instances of English verb-particle constructions. *Language Resources and Evaluation* 44(1-2). 97–113. doi:10/dzhcvz. 00021.

Lakoff, George & Mark Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

Lohse, Barbara, John A. Hawkins & Thomas Wasow. 2004. Domain Minimization in English Verb-Particle Constructions. *Language* 80(2). 238–261.

Perek, Florent. 2016. Recent change in the productivity and schematicity of the way-construction: A distributional semantic analysis. *Corpus Linguistics and Linguistic Theory* 14(1). 65–97. doi:10/gddr5j. 00002.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London and New York: Longman.

Rodríguez-Puente, Paula. 2013. *The Development of Phrasal Verbs in British English from 1650 to 1990: A Corpus-Based Study*. Santiago de Campostela, Spain: Universidade de Santiago de Compostela Ph.D. Thesis.

Shih, Stephanie, Jason Grafmiller, Richard Futrell & Joan Bresnan. 2015. Rhythm's Role in Predicting Genitive Alternation Choice in Spoken English. In R. Vogel & R. van de Vijver (eds.), *Rhythm in Phonetics, Grammar, and Cognition*, 207–234. Berlin: De Gruyter Mouton.

Shih, Stephanie S. 2017. Phonological Influences in syntactic alternations. In Vera Gribanova & Stephanie S. Shih (eds.), *The Morphosyntax-Phonology Connection*, 223–252. Oxford University Press. doi:10.1093/acprof:oso/9780190210304.003.0009.

Sinclair, John. 2005. *Collins Cobuild Dictionary of Phrasal Verbs*. Glasgow, GB: HarperCollins.

Thim, Stefan. 2012. *Phrasal Verbs: The English Verb-Particle Construction and Its History*. Berlin: Mouton de Gruyter.