Verbs describing routines facilitate object omission in English

Lelia Glass[*]

Abstract. Which normally-transitive English verbs can omit their objects (*I ate*), and why? This paper explores three factors that have been suggested to facilitate object omission: (i) how strongly a verb **selects** its object (Resnik 1993); (ii) a verb's **frequency** (Goldberg 2005); (iii) the extent to which the verb is associated with a **routine** – a recognized, conventional series of actions within a community (Levin & Rapaport Hovav 2014; Martí 2015). To operationalize (iii), this paper leverages the assumption that a given verb may be more strongly associated with a routine in one community than another. Comparing writings across communities, this paper offers corpus and experimental evidence that verbs omit their objects more readily in the communities where they are more strongly associated with a routine. Object-omitting uses of verbs are analyzed, following other work, as intransitive aspectual activities describing an agent's routine actions; so the hearer's task is not to recover a missing object, but to recognize the routine described by the verb. More broadly, the paper explores how the meaning and syntactic potential of verbs are shaped by the practices of the people who use them.

Keywords. lexical semantics; verbs; argument structure; computational/corpus linguistics; social media; language variation

1. Introduction: Selection, frequency, routine. It is a longstanding question in lexical semantics which normally-transitive verbs in English can omit their objects to describe an event with an unexpressed theme, which cannot, and why[1]. Even near-synonyms differ in how natural they sound with their objects omitted (1) (Fillmore 1986; Rice 1988; Mittwoch 2005; Gillon 2012), making it difficult to explain omission in terms of meaning, and leading some researchers to characterize this phenomenon as partly or fully arbitrary (Fillmore 1986; Ruppenhofer 2004; Gillon 2012).

(1)   a.   I ate ___ vs. ?I devoured ___
      b.   I drank ___ vs. ?I guzzled ___
      c.   I wrote ___ vs. ?I penned ___
      d.   I raked ___ vs. ?I bagged ___

This paper focuses mainly on cases where the omitted object is said not to require a discourse antecedent, rather than anaphoric cases such as *I noticed* ((Fillmore 1986; Condoravdi & Gawron 1996); but, since this distinction becomes fuzzy in practice, the paper does not attempt to separate so-called indefinite/existential (*I ate*) from definite/anaphoric object omission (*I noticed*) in corpus data.

Joining those who strive for a predictive explanation of which verbs allow object omission to what extent and why (Resnik 1993, 1996; Cote 1996; Rappaport Hovav & Levin 1998; Martí 2015), this paper disentangles three factors suggested to facilitate object omission:

---

[1]Other words for this phenomenon include *object drop, unspecified object, null object, null argument, null complement, null instantiation, diathesis alternation, implicit object*; see Resnik (1993).

- **Selection:** how much statistical information a verb provides about the taxonomic nature of its object (Resnik 1993, 1996).

- **Frequency:** the per-million-word frequency of a verb in a corpus (Goldberg 2005).

- **Routine:** the degree to which a verb describes a series of recognized, conventional actions within a community (Levin & Rapaport Hovav 2014; Martí 2015).

Striving to make the notion of routine empirically tractable, the paper offers evidence consistent with the claim that verbs more readily omit their objects in the communities where they are more strongly associated with a routine. Fitness enthusiasts are more likely than others to omit the object of *lift* (weights), home-brewers to omit the object of *bottle* (beer), and so on.

1.1. SELECTION. In a rigorous corpus study which inspires this one, Resnik (1993, 1996) shows that object omission is more available for verbs providing more statistical information about their object, using the WordNet taxonomy of Miller et al. (1988). *Eat* provides the information that its object is some sort of food (it 'selects' food objects). Resnik defines 'selection' (2) using information theory: how much information does the verb provide about the taxonomy of its object? When predicting the WordNet class of a random object of a random verb, how much uncertainty is reduced by knowing the particular verb?

(2)   selectional strength S of a verb v is defined as:
$S(v) = \sum_c P(c|v) \log \frac{P(c|v)}{P(c)}$
where v = verb, c = WordNet class

Using this definition, *eat* strongly selects food objects; *enjoy* has much weaker selectional strength because it appears with objects from many different taxonomic classes. For Resnik, *eat* can omit its object more easily than *enjoy* can because *eat* provides more information which can be used to recover its missing object. But this analysis struggles to explain why, for example, *devein* – which strongly selects *shrimp* objects in corpora (Ruppenhofer 2004) – cannot easily omit its object (*?I deveined today*).

To test the analysis at a broader scale, I used a dependency parser (Honnibal & Johnson 2015) and a list of normally-transitive verbs from Levin (1993) – a resource which lists and groups English verbs comprehensively – to find uses of transitive verbs with and without objects in a corpus of Reddit comments drawn from general-interest subreddits in January 2014 (a total of 22.8 million words from r/AskReddit, r/news, r/worldnews, r/explainlikeimfive, r/Showerthoughts, and r/LifeProTips). Reddit is an American social media site where users post content and discuss it in threads, providing a large dataset of English conversational writing. It is organized into subreddits – some of general interest (r/AskReddit, r/news), some more specialized (like r/Homebrewing, for beer enthusiasts); I consider the general-interest subreddits here, returning to the specialist subreddits below (§2).

While the dependency parser reliably indicates direct objects with a 'dobj' dependency, there is no equally reliable indicator for the *absence* of an object. Among cases where a transitive verb is used *without* a 'dobj' dependency, only a small percentage truly qualify as object omissions; the rest are false positives (Table 1).

These false positives were addressed using a blend of automated and human methods. I automatically excluded all sentences containing passives; all verbs with particles, prepositional

false positive	example	false positive	example
passive/middle	*it was seen, it washes well*	conjunction	*buy and sell stocks*
verb with non-noun complement	*wrote that . . . , trust in God*	non-verb incorrectly tagged as verb	*Best Buy, nutrition shakes*
particle verb	*sign up for a race*	'tough' construction	*hard to handle*
inchoative	*my eyes burn*	relative clause	*the sandwich I ate*

Table 1. Among transitive verbs used without direct-object nouns as complements, there are many false positives for object omission, removed using a blend of automation and human labor.

phrases, clausal, or adverbial complements as dependents; all verbs inside relative clauses; all verbs used as auxiliaries; and all verbs related to anything else by a conjunction. To avoid biasing the data, these exclusions applied equally to verbs with a 'dobj' dependent as well as those without. In an attempt to remove inchoative uses of causative verbs (*my eyes burn*), I used a semantic role labeling tool (He et al. 2017) available through the AllenNLP tool suite (Gardner et al. 2018), which automatically labels a verb's agent argument as 'arg0' and its patient argument as 'arg1', and excluded inchoative uses of verbs with a patient 'arg1' but no agent 'arg0' – although this tool still lets many inchoative uses slip through undetected.

After these automatic exclusions, we find a fairly accurate set of sentences where transitive verbs are used with direct objects; and a set of sentences with no 'dobj' dependency which still contains more false positives than true positives for object omission. I then personally read these sentences to remove those false positives by hand. The result is a large, trustworthy dataset recording, for 832 different verbs, the percent of the time that each verb omits its object: *lift* appears with an object 200 times (*I helped her **lift the box***), without an object 17 times (*you probably could if you were **lifting** correctly*), so it omits its object 7.8% of the time.

Resnik predicts a positive correlation between a verb's selectional strength and its propensity to omit its object. To test this claim, I used each verb's lemmatized direct objects to compute its selectional strength, and then tested the correlation between a verb's selectional strength and the percent of the time that it omits its object. Figure 1 plots the percentage of object-omitting occurrences of a verb as a function of Resnik's metric for selectional strength. In linear regression models, selection has a (marginally) significant p-value (intercept = $1.15, \beta = 0.57$, standard error = $0.27, p = 0.035, R^2 = 0.005$) – but p values so close to 0.05 are not trustworthy, especially with over 800 data points, and the R^2 value of 0.005 means that the model explains less than one percent of the data overall. In sum, contrary to Resnik's findings which used data from 34 verbs, these data do not offer convincing evidence that selection facilitates object omission.

1.2. FREQUENCY. The frequency of a verb, of course, is the number of times that verb occurs (per million words) in a lemmatized corpus. Goldberg (2005) notes that in minimal pairs such as (1) (*eat* vs. *devour*), the member of the pair that allows omission is the more frequent one. Following Fellbaum & Kegl (1989), she also observes that generic and habitual sentences (those describing general or recurring patterns of events) facilitate object omission more than episodic sentences (those describing particular episodes) – *that dog bites* is said to sound more natural than *that dog bit yesterday* (the 'characteristic property' verb alternation of Levin 1993;

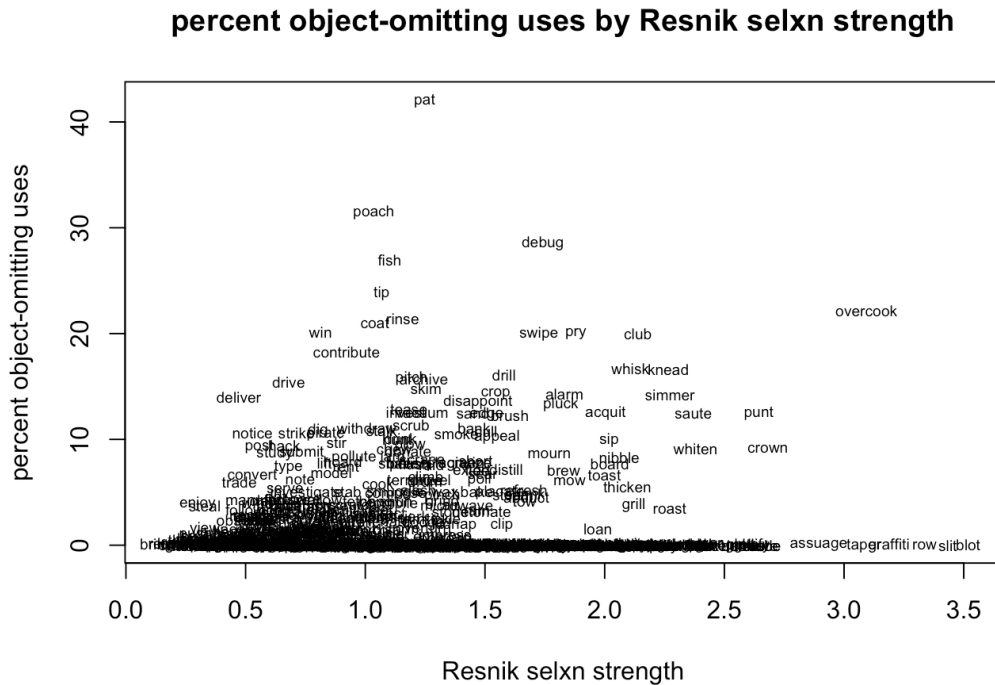**percent object-omitting uses by Resnik selxn strength**

Figure 1. Percentage of object omissions in 22.8 million words from general-interest subreddits, plotted as a function of Resnik's metric for selectional strength.

c.f. the idea from Hopper & Thompson 1980 that the most prototypically 'transitive' sentences describe particular episodes). Extrapolating from Goldberg's brief remarks, the suggestion is that if a verb appears frequently in generic or habitual contexts, it is likely to omit its object in those contexts, a pattern which would diachronically generalize to episodic contexts as well. The prediction is that a verb's frequency – both in generic/habitual contexts and overall – should positively correlate with object omission.

In contrast, Ruppenhofer (2004) (Chapter 4) studies 34 verbs, binary-classified into those said to allow or disallow object omission, and finds no association with their frequency: 'lemmas with very high token frequency allow' object omission, like *eat*, 'as well as lemmas with very low token frequency,' like *rake*. He concludes that object omission is idiosyncratic.

Using the Reddit data – instances of 832 transitive verbs used with and without objects in general-interest subreddits – described above (§1.1), Figure 2 visualizes the percentage of object-omitting occurrences of each verb as a function of its per-million-word frequency. Consistent with Ruppenhofer (2004), there is no relationship between a verb's frequency and the percent of the time that it omits an object. (A linear regression on these data has an intercept of 1.84; for per-million-word frequency, $\beta = -0.0009$, standard error $=0.0016, p = 0.56, R^2 = 0.0004$, confirming that there is no relationship between these variables; furthermore, an interaction model finds no interaction between selection and frequency). In sum, in these data, a verb's frequency does not predict the percent of the time that it omits its object.

1.3. ROUTINE. A routine is a series of actions that are well-known and conventional within a community. Stepping back, routines are rooted in the work of Schank & Abelson (1977),
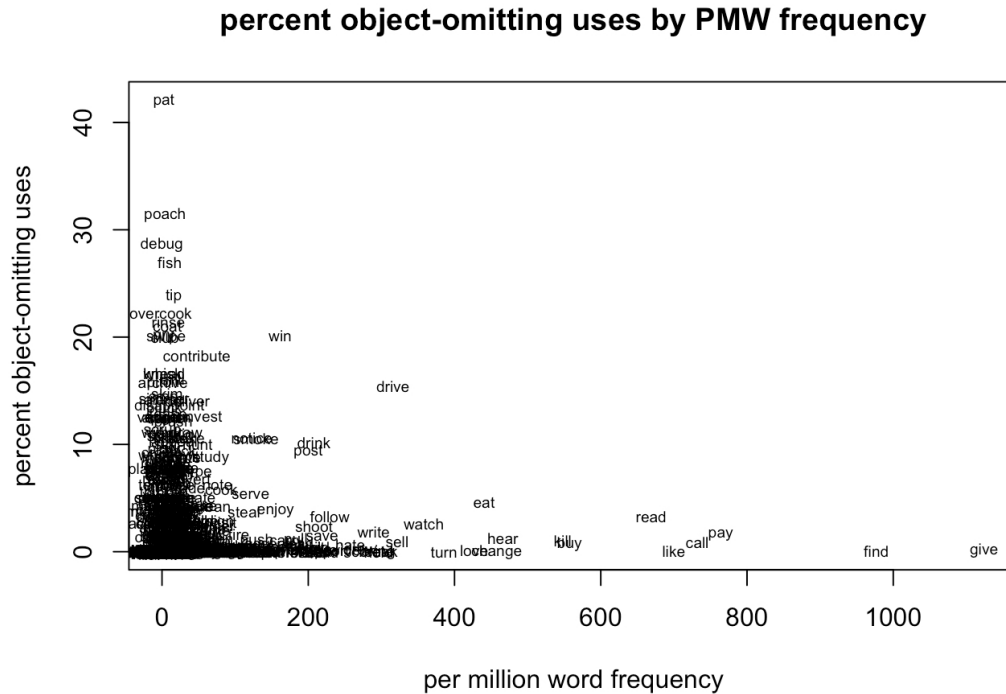
percent object-omitting uses by PMW frequency

Figure 2. Percentage of object-omitting uses of 832 normally-transitive verbs in in 22.8 million words from general-interest subreddits, plotted as a function of per-million-word frequency.

who observe that human reasoning leverages rich background knowledge of 'scripts' for how situations normally unfold. A 'restaurant' script includes artifacts such as tables, chairs, and menus; and events – some of which can be described by object-omitting verbs – such as ordering, eating, and paying. While not strictly entailed, these default expectations ground our understanding of discourses involving restaurants.

Fillmore (1976) contemporaneously argues that language understanding requires not only a grammar and a lexicon but also 'a description of the cognitive and interactional FRAMES in terms of which the language user interprets [their] environment, formulates [their] own messages, understands the messages of others, and accumulates or creates an internal model of [their] world.' A frame (synonymous with a cultural script) 'identifies an experience as a type and gives structure, coherence, and meaning to the points and relationships, objects and events, within the experience.' Frames exist in the mind independent of language, but words evoke frames, particularly verbs; *buy* and *sell* evoke different perspectives on the frame of commercial transactions, which involve a buyer, a seller, goods, a price, and so on[2].

Against this background, a *routine* is a community-wide script for a series of goal-directed actions of an agent. Even comparing verbs describing similar events, some are much more strongly associated with a routine than others; *exercise* evokes routines such as going outdoors or to a gym, wearing workout clothes, and moving around in various ways for a sustained period in order to improve one's health and mood; while *jump* describes a way of moving one's

---

[2]Similar ideas are echoed in the work of Pustejovsky (1995), who claims that a lexicon should include rich, structured encyclopedic knowledge.

body but is not associated with a routine for most speakers. (But, to foreshadow §2 and §3, *jump* may well be associated with a routine for certain athletes, such as ice skaters.)

Transitive verbs describing routines have been suggested to facilitate object omission. Martí (2015) analogizes object omission to the incorporation of object nouns into verbs (*berry-pick*), which Mithun (1984) argues is cross-linguistically only available for a community's 'institutionalized,' 'name-worthy' actions; incorporations such as *reindeer-hunt* arise in communities where that action is routine, and novel incorporations (*ladder-climb*) lead hearers to imagine that the action is somehow conventionalized, for example as part of a fitness test. Analyzing object omission as a form of silent noun incorporation, Martí claims that object-omitting verbs also describe routines.

Levin & Rapaport Hovav (2014) focus on *clean*, which they analyze as a 'result' verb when it conveys that its agent causes its theme to undergo a change-of-state (*I cleaned the table*). 'Result' verbs cannot omit their objects, they say, because the sub-event of the theme changing in state requires its argument to be syntactically realized (Rappaport Hovav & Levin 1998; Rappaport Hovav 2008). The puzzle is why *clean* actually can omit its object: *I cleaned yesterday*. They propose that *clean* is polysemous between a causative change-of-state requiring an object, and an aspectual activity describing routine actions of an agent typically leading to such changes of state (vacuuming, sweeping), in which the object can be omitted. Because changes of state commonly co-occur with the routines that bring them about, *clean* comes to polysemously describe both of them, and can omit its object when it describes a routine.

This paper aims to test, expand, and defend that analysis. Generalizing the proposal of Levin & Rapaport Hovav (2014), I analyze object-omitting uses of verbs such as *lift* as systematically polysemous, distinct in meaning from transitive uses of the same verb. Object-omitting uses are analyzed using the decompositional template of Levin & Rapaport Hovav (2014) as 'manner' verbs (3) describing aspectual activities, with the 'manner' portion filled in by the routine actions associated with the particular verb; and in a neo-Davidsonian framework ((4); Davidson 1967) as having no semantic theme argument.

(3)   [x ACT <MANNER>]   (Rappaport Hovav & Levin 1998)

(4)   Alice lifted.
      $\exists e[lift(e) \wedge agent(e, Alice)]$

To interpret an object-omitting use such as (4), the hearer's main task is not to recover any particular missing object (as argued by e.g., Cote 1996), but instead to infer the routine associated with the verb – to understand that, for strength-training enthusiasts, *lift* describes a convention of exercising with weighted equipment to improve one's strength. To the extent that such a routine is familiar to the hearer, the object-omitting use of the verb can be interpreted. As for the much-discussed contrast between *eat* and *devour*, this analysis posits that *eat* is associated with a routine in our culture (involving conventional foods, utensils, and mealtimes), whereas it is less clear what routine would be described by an object-omitting use of *devour*.

By analyzing object-omitting uses of verbs as intransitive (an idea dating back to Fodor & Fodor 1980), this analysis captures the fact that omitted objects vary widely: in terms of how they would be paraphrased; how easily or specifically they can be inferred; and even whether they exist at all (in the context of martial arts drills, *I kicked* may describe an event with no patient) – facts which would be difficult to capture if all omitted objects had to share a single,

	with object	without object
generalist	**200** Because I tore my ACL **lifting a box**.	**17** You probably could if you were **lifting** correctly.
strength	**347** I work out for the sole intention of self improvement, I like **lifting heavy things**.	**139** Damn son, how long have you been **lifting**?

Table 2: Fisher test showing that the proportion of object-omitting uses of *lift* differs significantly between generalist and strength-specialist subreddits.

silent semantics. Instead, true to what is pronounced, there is no semantic object, but one may be conceptually inferred (Rice 1988) depending on the routine associated with the verb.

For this analysis to be tested, the challenge is that it is not obvious how the elusive social/mental concept of routine can be measured empirically, nor how routine should overlap with or come apart from selection and frequency. Facing that challenge, this paper leverages the assumptions (i) that routines are gradient: in a given community, some actions are more or less routine than others; and (ii) that routines vary across communities (Mithun 1984); people who pursue specialized interests may engage in routines not shared by outsiders. Then we can compare across communities where a verb is more or less strongly associated with a routine.

Thus §2 compares the writings of specialist communities to those of generalist communities on Reddit, and §3 compares experimental items reporting the speech of specialized hobbyists to the speech of people who do not regularly engage in those hobbies. Both studies are argued to support the *comparative* claim that verbs omit their objects *relatively more often* in the communities where the actions they describe are *relatively more* routine.

2. Comparing across communities on Reddit. To test the claim that a verb omits its object more readily in a community in which it is more strongly associated with a routine, I return to the Reddit data introduced above (§1.1), this time comparing across subreddits with different conventions and routines.

In addition to the data from general-interest subreddits, I further collected uses of transitive verbs with and without objects from 82 specialist subreddits grouped into 15 diverse specialties (for example, I grouped r/bodybuilding, r/xxfitness, r/weightlighting, and r/strength_training into a 'strength' specialty). Because many specialty subreddits are small by word-count, grouping the thematically similar subreddits offers more data. Table 2 (an excerpt) illustrates subreddits grouped by speciality, as well as the total number of words in each specialty.

Using the process for identifying uses of verbs with and without objects described above, including the labor-intensive hand-correction (§1.1), I compared the proportion of uses of each transitive verb with and without an object in each specialist corpus compared to the generalist corpus (which comprises AskReddit, worldnews, and so on). Table 2 provides the count data for all occurrences of the verb *lift* with and without an object in both the generalist corpus (where it occurs 200 times with an object, 17 times without) and the specialist corpus dedicated to strength training (where it occurs 347 times with an object, 139 times without).

For each verb and each specialist corpus, such a two-by-two table was used in a Fisher Exact Test (a more-exact cousin of the χ^2 test) to determine whether the proportion of object omissions differs significantly between the specialist corpus and the generalist one. In Table 2,

specialty	subreddits	verbs omitting obj signif. more
generalist 22.8 million words	AskReddit, worldnews, news, explainlikeimfive, LifeProTips, Showerthoughts	n/a; 'generalist' is the reference to which all 'specialist' corpora are compared
ballsports 16.5 million	nfl, nba, soccer, CFB, Baseball, MLS, cricket, ultimate, lacrosse, football, Basketball	*bat, block, catch, cover, deliver, drink, execute, hear, hold, like, miss, pass, press, produce, reach, recall, receive, sell, shoot, shovel, strike, throw, visit*
finance 5.8 million	personalfinance, investing, business, stocks, smallbusiness, RealEstate, wallstreetbets, realtors	*bite, buy, call, close, cook, eat, file, hold, invest, purchase, rent, save, sell, smoke*
beverages 2.7 million	Homebrewing, beer, tea, wine, CraftBeer	*age, boil, bottle, chill, clean, cool, distribute, enjoy, fill, filter, freeze, mash, open, serve, shake, stir, taste, transfer*
photo 2.3 million	photography, astrophotography, photocritique, WeddingPhotography	*buy, guide, order, purchase, shoot, stack, track*
strength 1.5 million	bodybuilding, xxfitness, weightlifting, strength_training	*cut, eat, gain, lift, press, push, smoke, stretch, wash*
martial arts 1.3 million	bjj, martialarts, MuayThai, karate, kungfu, taekwondo, kravmaga	*hit, kick, pass, punch, stretch, strike, tap*
(…& more)	(…and more)	(…and more)
words=72m	n=88 (6 generalist, 82 specialist)	n=134

Table 3. An excerpt illustrating the 'generalist' and 'specialist' Reddit corpora; the subreddits they comprise; their total word count; and some of the 134 verbs significantly more often omitting objects in the specialist corpora compared to the generalist one.

lift indeed omits its object more often in the strength-specialist corpus ($p < 0.001$).

Ultimately, this process yields a dataset of 134 verbs across 15 specialist corpora found to omit their objects proportionally more often at a statistically significant level ($p < 0.05$) in those specialist corpora. Table 2 excerpts these data, available in full at https://osf.io/kxe3u/.

Impressionistically, some of the verbs in Table 2 indeed seem to describe community-specific routines, while others are more debatable. Some verbs polysemously take on a narrower sub-sense within a given community: for photographers, *shoot* specifically involves photographs; for martial artists, *tap* specifically means tapping on a mat to end a fight. Such polysemy is consistent with the claim that verbs are more disposed to omit their objects in the communities in which they are more strongly associated with a routine: if a verb describes a routine in a particular community, then it takes on a narrower sense to describe that routine.

Other verbs already robustly omit their objects in generalist subreddits (perhaps because they describe mainstream routines?), but omit their objects even more often in certain specialist subreddits: in r/personalfinance, people budget for *eating* out or *cooking* at home.

Some object-omitting uses of verbs blur the boundary between 'definite/anaphoric' and 'indefinite/existential', illustrating why I did not attempt to separate these: for sports fans, do

catch, throw, and *pass* refer to a particular discourse-given ball, or to balls in general? For those interested in finance, does *file* refer anaphorically to one's specific taxes or existentially to taxes in general?

In sum, the object-omitting verbs that emerge from specialist subreddits are quite heterogeneous. But, taken together, these data can be used to explore the claim that a verb is more likely to omit its object in a community where it is more strongly associated with a routine.

Quantitatively, there is no metric for routine. But one might expect that if a verb is more strongly associated with a routine in Community A compared to Community B, then it might have *comparatively higher* selection in Community A, because one might expect it to appear more regularly with the objects involved in the routine that it describes (generalists might discuss *lifting* all sorts of things, while strength enthusiasts might tend to discuss *lifting weights* in particular). One might also expect that verb to be *comparatively more* frequent in Community A, because one might expect an action to be discussed more frequently where it is more routine (if lifting weights constitutes a routine for strength enthusiasts, then they might use the verb *lift* more often than other people). In other words:

(5) **The Comparative Routine → Selection Hypothesis**: If a verb is more strongly associated with a routine in Community A compared to Community B, that verb should more-strongly select its object in the speech/writings of Community A.

(6) **The Comparative Routine → Frequency Hypothesis**: If a verb is more strongly associated with a routine in Community A compared to Community B, that verb should occur more frequently in the speech/writings of Community A.

(7) **The Comparative Routine → Omission Hypothesis**:
   a. If a verb is more strongly associated with a routine in Community A compared to Community B, that verb should omit its object more often in the speech/writings Community A.
   b. If a verb omits is object more often in the speech/writings of Community A compared to Community B, then that verb should be more strongly associated with a routine in Community A – and thus, by (5) and (6), should be more strongly select its object and occur more frequently in the speech/writings of Community A.

To test (7-b) with respect to selection, Figure 3 plots the selection (per Resnik's metric) of these verbs in both the generalist corpus and in the speciality corpus where they omit their objects significantly more often[3]. According to a Wilcoxon test for paired samples in R (Wilcoxon 1945; a non-parametric alternative to a paired t-test), consistent with (7-b), these verbs have systematically higher selection in the specialist subreddits in which they omit their objects significantly more often.

To test (7-b) with respect to frequency, Figure 4 plots the per-million-word frequency of these verbs in both the generalist corpus and in the speciality corpus where they omit their objects significantly more often[4]. These results are also highly significant in a Wilcoxon test

---

[3]Median Resnik selection is 0.62 in generalist data, 0.76 in specialist data; mean Resnik selection is 0.73 in generalist data, 0.87 in specialist data. Comparing Resnik selection between generalist and specialist subreddits in a Wilcoxon test, $V = 2120, p < 0.0001$ – a significant difference

[4]Median per-million-word-count is 84 in generalist data, 203 in speciality data; mean per-million-word count is 154.4 in generalist data, 365.2 in specialist data.

Resnik selxn strength of verbs in generalist vs specialty subreddits

Figure 3. Comparing the selection of 134 verbs in both generalist subreddits and in the specialist subreddits in which they omit their objects significantly more often.

per-million-word counts of verbs in generalist vs specialty subreddits

Figure 4. Comparing the per-million-word frequency of 134 verbs in both generalist subreddits and in the specialist subreddits in which they omit their objects significantly more often.

for paired samples: consistent with (7-b), these verbs are significantly more frequent in the specialist subreddits where they also omit their objects significantly more often ($V = 1488, p < 0.0001$).

Greatly expanding the data considered in the literature on object omission, these results are consistent with the claim that verbs omit their objects more often in the communities where they describe routines. But with no clear way of measuring routine in a corpus, this evidence

is necessarily indirect. Therefore, §3 uses an experimental setup to provide more controlled, complementary evidence for the same claim.

3. Experiments. I present two experiments, one crossing routine with selection (Experiment 1) and one crossing routine with frequency (Experiment 2), both consistent with the hypothesis that routine, disentangled from selection and frequency, facilitates object omission[5]:

(8)   **The Comparative Routine → Omission Hypothesis**: If a verb is more strongly associated with a routine in Community A compared to Community B, that verb should (be judged by experimental participants to) omit its object more easily in the speech/writings Community A.

3.1. EXPERIMENT 1: ROUTINE VS. SELECTION. Experiment 1 uses experimental items of the form in (9), each one randomly assigned to one of four conditions crossing high vs. low routine with high vs. low selection. In all conditions, a narrator tells a three-sentence story about their weekend using a past-tense target transitive verb V with an object; then participants rate on a five-star Likert scale the likelihood of that narrator using a past-tense, episodic, object-omitting form of V.

(9)   **Experiment 1: Routine vs. Selection**
Your friend Caroline tells you about her weekend:

	Selection low	**Selection high**
***Routine* low**	*I went for a bike ride. Then, just because my friend wanted me to try it,* I bottled beer and I bottled lotion. *Then I went fishing.*	*I went for a bike ride. Then, just because my friend wanted me to try it,* I bottled beer and I bottled more beer. *Then I went fishing.*
***Routine* high**	*I worked on my home-brewing project. Like every weekend,* I bottled beer and I bottled lotion. *Then I grew some hops.*	*I worked on my home-brewing project. Like every weekend,* I bottled beer and I bottled more beer. *Then I grew some hops.*

The next time Caroline tells you about bottling beer the day before, how likely do you think she is to say the following? 'I bottled yesterday.' ☆☆☆☆☆

The verb V (*bottle*) is one that does not normally omit its object. Moreover, past-tense uses of verbs are particularly resistant to object omission (Goldberg 2005) – and indeed, past-tense uses are under-represented (while progressive uses are over-represented; c.f. Hopper & Thompson 1980) among the object-omitting uses of verbs found in my Reddit studies. This setup was intended to give the stimulus sentence (*I bottled yesterday*) a low baseline rating, so that the manipulations predicted to improve this rating would be most visible.

The target verb V is intended to be more strongly associated with a routine for the narrator in the high-routine condition than the low-routine condition. In high-routine items, all three sentences of the story revolve around the same hobby, which the narrator describes as a regular commitment, conveying that the action described by V is a routine for them. In low-routine items, each sentence of the story describes a different activity, and the action described by V is pursued haphazardly at the suggestion of a friend.

The target verb V appears twice in all conditions, meaning that its frequency is held con-

---

stant and providing two data points about the objects it combines with. In the high-selection condition, the verb appears twice with the same object (consistent objects lead to higher selectional strength), while in the low-selection condition, it appears once with one object and a second time with another object meant to differ taxonomically from the first one (lower selectional strength). This design tests how the Likert rating of object omission is affected by selection, routine, and their interaction.

I turn next to the second experiment, then discuss the results of both experiments together.

3.2. EXPERIMENT 2: ROUTINE VS. FREQUENCY. Experiment 2 parallels the design of Experiment 1, but this time crossing routine with frequency. Experimental items follow the form in (10), each one randomly assigned to one of four conditions crossing high vs. low routine with high vs. low frequency. Again, the narrator tells a three-sentence story about their weekend; then participants rate on a five-star Likert scale the likelihood of that narrator using a past-tense, episodic object-omitting form of the target verb V.

The high vs. low routine conditions mirror the setup in Experiment 1: the narrator either indicates that they participate in the action described by the target verb regularly or haphazardly. As for frequency, the target verb V appears with an object once in the low-frequency condition, and twice with that same object in the high-frequency condition. This design tests how the Likert rating of object omission is affected by frequency, routine, and their interaction.

(10)   **Experiment 2: Routine vs. Frequency**
Your friend Caroline tells you about her weekend:

	Frequency low	**Frequency high**
Routine **low**	*I went for a bike ride. Then, just because my friend wanted me to try it,* I bottled beer. *Then I went fishing.*	*I went for a bike ride. Then, just because my friend wanted me to try it,* I bottled beer and I bottled more beer. *Then I went fishing.*
Routine **high**	*I worked on my home-brewing project. Like every weekend,* I bottled beer. *Then I grew some hops.*	*I worked on my home-brewing project. Like every weekend,* I bottled beer and I bottled more beer. *Then I grew some hops.*

The next time Caroline tells you about bottling beer the day before, how likely do you think she is to say the following? 'I bottled yesterday.' ☆☆☆☆☆

In both experiments, eight items like (9)–(10) were presented on Qualtrics in a random order and interspersed with ten fillers, some of which asked about intransitive verbs (*run*), some about verbs that commonly omit their objects (*paint*), and some which asked out-of-the-blue (without a story from the narrator) about verbs that do not typically omit their objects (*lock*). Participants (n=100 for Experiment 1, 98 for Experiment 2) using United States I.P. addresses were recruited via Amazon's Mechanical Turk with an official IRB exemption from my institution; after being advised that they would be paid regardless of their answer, participants indicated their native language to be English.

3.3 RESULTS. For both experiments, data were analyzed using mixed-effects linear regression model in R (lmer; Bates et al. 2015), predicting an item's Likert rating as a function of selection(Experiment 1)/frequency(Experiment 2), routine, and their interaction, with random intercepts for both items and participants. In these models (Table 4), no effect is found for se-

Exp1	lo routine	hi routine
lo selxn	2.37 (intercept)	2.78 ($\beta = 0.41, t = 4.10, p < 0.001$**)
hi selxn	2.38 ($\beta = 0.01, t = -0.09, p = 0.92$)	2.74 ($\beta = -0.05, t = -0.31, p = 0.76$)

Exp2	lo routine	hi routine
lo freq	2.0 (intercept)	2.42 ($\beta = 0.42, t = 4.38, p < 0.001$**)
hi freq	2.13 ($\beta = 0.13, t = 1.36, p = 0.17$)	2.47 ($\beta = -0.08, t = 0.61, p = 0.55$)

Table 4: Model's predictions for Likert rating of object omission of 'rating' based on selection(Experiment 1), frequency (Experiment 2), routine, and their interaction. High-routine items have significantly higher Likert ratings than low-routine items.

lection, nor for frequency, nor for any interaction of either one with routine[6]. The consistent finding is that high-routine items are rated significantly higher than low-routine items.

In sum, these experimental results are consistent with the hypothesis (8) that routine, separated from selection and frequency, facilitates object omission. From the null results found for selection and frequency, nothing positive can be concluded; these factors may still facilitate object omission in ways that this experiment did not identify. But the experiment does identify a striking effect of routine: consistent with the Routine $\rightarrow$ Omission Hypothesis, a verb is judged to omit its object more easily when the speaker indicates that the verb is more strongly associated with a routine for them.

4. Conclusion. Object omission has attracted decades of interest because it requires us to disentangle syntax, word meaning, discourse, and world knowledge (Cote 1996); and because our desire for far-reaching explanations conflicts with the messiness of the data – where we observe differences within near-synonyms as well as across languages. Aiming to find some regularity in this realm, this paper has tested and defended the claim (Levin & Rapaport Hovav 2014; Martí 2015) that verbs facilitate object omission if they describe routines. A corpus study finds verbs omitting their objects more often in specialist communities compared to generalist ones, and shows that these verbs also have stronger selectional strength and higher frequency in the communities where they omit their objects more often. Experiments show that when routine is crossed with selection and with frequency, it is routine that improves Likert ratings of object omission. Taken together, these studies provide evidence consistent with the claim that verbs describing routines facilitate object omission. Since routines vary across communities and time, this analysis also explains why object omission may be judged differently across speakers, and makes synchronic and diachronic predictions about where it may emerge.

Of course, the demonstrated effect of routine explains at most one part of a complex phenomenon. Many other elements play a role too: the subject/agent, the omitted object/theme, the tense/aspect/modality and structure of the full sentence, and the larger discourse (Hopper & Thompson 1980; Cote 1996; Goldberg 2005; Mittwoch 2005).

Methodologically, experiments allow us to isolate variables which may be correlated in corpora (selection, frequency, routine), and to control for tense, aspect, modality, and surrounding discourse. Illustrating the value of natural language processing tools for theoretical lin-

---

[6]These results – that high-routine items have systematically higher ratings than low-routing items – also replicate, for both experiments, in an ordinal regression, which treats Likert stars as ordered but not linear. The data and code are available on the Open Science Framework: https://osf.io/kxe3u/

guists, the corpus study provides evidence that the experimental findings extend to real life: indeed, specialists often omit the objects of verbs describing their unique routines. A structured social media corpus is used to compare across communities where verbs are arguably more or less associated with routines – perhaps a new front in the study of cross-community differences in lexical semantics. Zooming out, this paper offers evidence that language, down to its structure, is fundamentally social. The syntax of verbs is argued to be shaped by the practices of the people who use those verbs.

References

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Condoravdi, Cleo & Jean Mark Gawron. 1996. The context-dependency of implicit arguments. In Makoto Kanazawa, Christopher Piñón & Henriëtte de Swart (eds.), *Quantifiers, deduction, and context*, Stanford: CSLI (Center for the Study of Language and Information) Publications.

Cote, Sharon Ann. 1996. *Grammatical and discourse properties of null arguments in english*. Philadelphia: University of Pennsylvania dissertation.

Davidson, Donald. 1967. The logical form of action sentences. In Nicholas Rescher (ed.), *The logic of decision and action*, 81–95. Pittsburgh, PA: University of Pittsburgh Press. https://doi.org/10.1093/0199246270.003.0006.

Fellbaum, Christiane & Judy Kegl. 1989. Taxonomic structures and cross-category linking in the lexicon. In Ken deJong & Yongkyoon No (eds.), *Eastern States Conference on Linguistics*, vol. 6, 93–104. Department of Linguistics at The Ohio State University.

Fillmore, Charles. 1986. Pragmatically controlled zero anaphora. In *Proceedings of Berkeley Linguistics Society*, vol. 12, 95–107. Berkeley: Berkeley Linguistics Society. https://doi.org/10.3765/bls.v12i0.1866.

Fillmore, Charles J. 1976. Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the origin and development of language and speech*, vol. 280 1, 20–32.

Fodor, Jerry A. & Janet Dean Fodor. 1980. Functional structure, quantifiers, and meaning postulates. *Linguistic Inquiry* 11(4). 759–770.

Gardner, Matt, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson Liu, Matthew Peters, Michael Schmitz & Luke Zettlemoyer. 2018. AllenNLP: A deep semantic natural language processing platform. *Proceedings of the Association for Computational Linguistics*. https://doi.org/10.18653/v1/w18-2501.

Gillon, Brendan S. 2012. Implicit complements: A dilemma for model theoretic semantics. *Linguistics and Philosophy* 35(4). 313–359. https://doi.org/10.1007/s10988-012-9120-2.

Goldberg, Adele E. 2005. Constructions, lexical semantics, and the correspondence principle: Accounting for generalizations and subregularities in the realization of arguments. In Nomi Erteschik-Shir & Tova Rapoport (eds.), *The syntax of aspect: Deriving thematic and aspectual interpretation*, 215–236. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199280445.001.0001.

He, Luheng, Kenton Lee, Mike Lewis & Luke Zettlemoyer. 2017. Deep semantic role label-

ing: What works and what's next. In *Proceedings of the Association for Computational Linguistics*. https://doi.org/10.18653/v1/p17-1044.

Honnibal, Matthew & Mark Johnson. 2015. An improved non-monotonic transition system for dependency parsing. In *Proceedings of the 2015 conference on Empirical Methods in Natural Language Processing*, 1373–1378. https://doi.org/10.18653/v1/d15-1162.

Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251–299. https://doi.org/10.2307/413757.

Levin, Beth. 1993. *English verb classes and alternations*. Chicago: University of Chicago Press.

Levin, Beth & Malka Rapaport Hovav. 2014. Manner and result: A view from 'clean'. In Rob Pensalfini, Myfany Turpin & Diana Guillemin (eds.), *Language description informed by theory*, 337–357. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.147.14lev.

Martí, Luisa. 2015. Grammar versus pragmatics: Carving nature at the joints. *Mind & Language* 30(4). 437–473. https://doi.org/10.1111/mila.12086.

Miller, George, Christiane Fellbaum, Judy Kegl & Katherine Miller. 1988. WordNet: An electronic lexical reference system based on theories of lexical memory. *Revue quebecoise de linguistique* 17(2). 181–212. https://doi.org/10.7202/602632ar.

Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60(4). 847–894. https://doi.org/10.2307/413800.

Mittwoch, Anita. 2005. Unspecified arguments in episodic and habitual sentences. In Nomi Erteschik-Shir & Tova Rapoport (eds.), *The syntax of aspect: Deriving thematic and aspectual interpretation*, 237–254. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199280445.003.0011.

Pustejovsky, James. 1995. *The generative lexicon*. Cambridge: MIT Press.

Rappaport Hovav, Malka. 2008. Lexicalized meaning and the internal structure of events. In Susan Rothstein (ed.), *Theoretical and crosslinguistic approaches to the semantics of aspect*, 13–42. Amsterdam: John Benjamins. https://doi.org/10.1075/la.110.03hov.

Rappaport Hovav, Malka & Beth Levin. 1998. Building verb meanings. In Miriam Butt & Wilhelm Geuder (eds.), *The projection of arguments: Lexical and compositional factors*, 97–134. CSLI (Center for the Study of Language and Information) Publications, Stanford.

Resnik, Philip. 1993. *Selection and information: A class-based approach to lexical relationships*: University of Pennsylvania dissertation.

Resnik, Philip. 1996. Selectional constraints: An information-theoretic model and its computational realization. *Cognition* 61. 127–159. https://doi.org/10.1016/s0010-0277(96)00722-6.

Rice, Sally. 1988. Unlikely lexical entries. In *Proceedings of the Berkeley Linguistics Society*, vol. 14, 202–212. Berkeley: Berkeley Linguistics Society. https://doi.org/10.3765/bls.v14i0.1797.

Ruppenhofer, Josef. 2004. *The interaction of valence and information structure*. Berkeley: University of California dissertation.

Schank, Roger C. & Robert Abelson. 1977. *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, New Jersey: Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203781036.

Wilcoxon, Frank. 1945. Individual comparison by ranking methods. *Biometrics* 1. 80–83. https://doi.org/10.2307/3001968.