

running head left: A. Goldberg

running head right: Corpus evidence of the viability of statistical preemption

Cognitive Linguistics (2011) Vol. 22 1: 131-154.

Corpus evidence of the viability of statistical preemption*

ADELE E. GOLDBERG

Abstract

The present paper argues that there is ample corpus evidence of statistical preemption for learners to make use of. In the case of argument structure constructions, a verb_i is preempted from appearing in a construction A, CxA, if and only if the following probability is high: $P(CxB|\text{context that would be suitable for CxA and verb}_i)$. For example, the probability of hearing a preemptive construction, given a context that would otherwise be well-suited for the ditransitive is high for verbs like explain that overwhelmingly appear in the dative, and low for verbs like tell that readily appear in the ditransitive. Strength of statistical preemption is determined both by this probability, and by the frequency (ln (F)) of a verb in a preemptive construction when the context is at least as well suited to the preempted construction. The critiques of preemption by Stefanowitsch (this volume, 2008) are countered by arguing that the relevant probabilities were not considered. Moreover, we find evidence that constructions are somewhat less constrained when yoked to non-alternating verbs, as Stefanowitsch (this volume) suggests should be the case.

Keywords: construction learning, dative, ditransitive, statistical preemption,

1. Introduction

How do speakers learn what not to say? Sentences do not come marked as unacceptable. Nonetheless, there are perfectly sensible formulations that are judged ill-formed by native speakers. These include examples such as (1) to (3):

- (1) ??*She explained her the news.*
(cf. *She told her the news.*)
- (2) ??*She considered to go to the farm.*
(*She wanted to go to the farm.*)
- (3) ?*She saw the afraid boy.*
(*She saw the scared boy.*)

We should acknowledge at the outset that expressions such as those in (1) to (3) certainly do occasionally occur in corpus data (e.g. Fellbaum 2005). But it is clear to native speakers and it can be demonstrated with corpus evidence that *explain* strongly prefers the prepositional dative (4), *consider* prefers a gerundive verb phrase (5) and *afraid* prefers a predicative rather than attributive use (6):

- (4) *She explained the news to her.*
- (5) *She considered going to the farm.*
- (6) *She saw the boy who was afraid.*

Semantic considerations (Ambridge et al. 2009) do not provide an explanation of the ill-formedness of (1) to (3), since the meanings of the words are perfectly semantically compatible with the constructions, as indicated by the paraphrases provided in parentheses (cf. also Pollard and Sag 1987). In fact (1) to (3) illustrate just the sort of errors that advanced non-native speakers often make. How do native speakers learn to systematically avoid such formulations?

It might be tempting to believe that the simple non-occurrence of a verb in an argument structure construction is sufficient to rule it out (Braine and Brooks 1995). This proposal, that formulations are only acceptable if they are heard with some

regularity — are *entrenched* — faces a problem once we recognize that many verbs are well entrenched in only one argument structure construction and yet they are freely available for use in a different one. This is clear in examples (7) to (10).

- (7) *I actually had a moth go up my nose once. I...coughed him out of my mouth.*
bikeforums.net/archive/index.php/t-292132
- (8) *“Sarah ... winked her way through the debates.”*
pcneedtogo.blogspot.com
- (9) *“She’d smiled herself an upgrade.”*
 Douglas Adams, *Hitchhiker’s Guide to the Galaxy*
- (10) *“Tim ... sneezed the milk out of his nose.”*
www.zoackkennel.com/tims-story.html

Each of these verbs, *cough*, *wink*, *smile*, and *sneeze* is very frequent (entrenched) in the intransitive construction, and only exceedingly rarely, if ever, witnessed in the various transitive constructions in (7) to (10). Therefore an account based on entrenchment would seem to inappropriately rule out the sentences in (7) to (10) as unacceptable (Goldberg 2006). Stefanowitsch’s (2008) proposal for addressing this issue is discussed in section 4, where we see that it falls prey to the same sort of counterexamples given in (7) to (10).

For challenges faced by other attempts to explain examples like (1) to (3), see excellent overviews by e.g. Bowerman (1996), Pinker (1989) and Stefanowitsch (this volume). Boyd and Goldberg (to appear) also review issues related to (a) entrenchment, (b) direct negative evidence, (c) underlying features, (d) the semantic proposal of Ambridge et al. (2009), and (e) the role of recasts or reformulations (e.g. Saxton 1997; Chouinard and Clark 2003).

2. Statistical preemption

Consider for a moment, the more easily tractable case of morphology. How is that we know we should use *went* instead of **goed*? Clearly it is because we consistently hear *went* in contexts

where *goed* would have been at least as appropriate: this is statistical preemption. The preemption process is straightforward in these cases because the actual form serves the identical semantic/pragmatic purpose as the preempted form.

Notice that speakers do not avoid forms simply because an alternative form is highly entrenched. If we invent a new verb, *kleb*, meaning “to wash” and for some pragmatic reason (say it is learned at camp), it is only used in the future tense (e.g. *I will kleb my face later*). Intuitions are clear that as soon as the situation arises in which the past tense is appropriate, speakers will be perfectly comfortable using it (*yes, I've klebbed!*). The past tense is not preempted by the more entrenched future use, because the future was not used in contexts that were at least as appropriate for the past tense.

The role of statistical preemption when considering phrasal forms requires discussion, since distinct phrasal constructions are virtually never semantically and pragmatically identical, the way *went* and the hypothetical *goed* are. Any two phrasal constructions differ either semantically or pragmatically (or both) (Bolinger 1977; Clark 1987; Goldberg 1995), and so both constructions often happily coexist for particular lexical items.

For example, the dative and ditransitive have overlapping, but distinct semantic and information structure properties. The differences have been documented time and again by many researchers (Arnold et al. 2000; Bresnan et al. 2005; Collins 1995; Erteschik-Shir 1979; Goldberg 1995; Goldberg 2006; Green 1974; Hovav and Levin 2005; Oehrle 1975; Thompson 1995; Wasow 2002). The differences are subject to some dialect differences and gradability, yet it is possible to predict with high probability which construction will be preferred in a given context, for a given dialect (Bresnan and Hay 2007; Bresnan and Ford 2010). Different formulations of the differences between the two constructions have been used; often, the emphasis has been on the greater restrictiveness of the ditransitive. For example, the dative is preferred when conveying caused motion to a place instead of a recipient (11a to 11b), and the dative is preferred in standard dialects of American and British English if the theme is a third person pronoun (12a to 12b):

- (11) a. ?? *She sent the moon a package*¹. ditransitive
 b. *She sent a package to the moon.* dative

- (12) a. ??*She sent him it.* ditransitive
 b. *She it to him.* dative

But knowledge that the prepositional paraphrase is licensed as in (13), based on positive evidence, should not in any simple way prevent the use of the ditransitive in (14), because a large number of verbs do freely appear in both constructions (Bowerman 1996, Pinker 1989):

(13) *She explained the story to someone.*

(14) ??*She explained someone the story.*

A statistical form of pre-emption can play an important role in learning to avoid expressions such as (1) to (3) in the following way (Goldberg 1993; 1995; 2006; cf. also Pinker 1984: 400). In a situation in which a construction A (CxA) might have been expected to be uttered with a given verb, *verb_i*, the learner can infer that CxA is not after all appropriate for that verb, if a different construction, CxB, is consistently witnessed instead. That is, CxB statistically preempts CxA for *verb_i*, to the extent that the probability in (15) approaches 1. I address the important factor of *confidence* of statistical preemption in section 4.

- (15) The probability of CxB statistically preempting CxA for a particular verb, *verb_i*:

$$P(\text{CxB} \mid \text{a discourse context in which the learner might expect to hear CxA}[\textit{verb}_i])^2$$

This probability is equivalent to the following:

$$P(\text{CxB} \mid \text{a discourse context at least as suitable for CxA, and } \textit{verb}_i)$$

For example, if we assume that *explain* does not readily occur in the ditransitive construction because it is statistically preempted by the dative construction, we predict the probability in (16) to be high:

- (16) P(dative \mid a discourse context at least as suitable for the ditransitive and *explain*)

As researchers, we need to operationalize how to count “discourse contexts that are at least as suitable for the ditransitive”. It turns out there is a simple proxy for this number: the total number of ditransitive *and* dative uses of a given verb, when the semantics and information structure of the ditransitive are satisfied. That is,

$$\begin{aligned}
 (17) \quad & P(\text{dative} \mid \text{context at least as suitable for ditransitive and} \\
 & \text{verb}_i) \\
 & \approx \\
 & P(\text{dative} \mid \text{verb}_i \text{ and } [\text{dative}_{\text{with relevant restrictions}} \text{ or ditransitive}])
 \end{aligned}$$

In order to calculate the probability in (17), we need to find contexts in which the ditransitive is at least as appropriate as the dative construction. The particular well-known restrictions on the ditransitive outlined do not suggest such contexts. Instead, they provide contexts where a learner might expect to hear a dative rather than a ditransitive. One clear context in which the ditransitive *is* preferred over the dative is when the recipient is pronominal and the theme is not. The dative is certainly *allowed* in such contexts, but if we look at corpus data, we find the following preference quite clearly:

- (18) a. *She told me the news.*
 b. #*She told the news to me.*

The Corpus of Contemporary American English (COCA) is a 400 million word tagged corpus, freely available on-line at <http://view.byu.edu/>, thanks to Mark Davies. Table 1 shows that the probability of witnessing a dative construction in this corpus, given a pronominal recipient and a non-pronominal theme, is only, on average, .04, for verbs that alternate:

Table 1: *Verbs that alternate: comparison of number of dative and ditransitive occurrences when the recipient argument is pronominal and the theme is not. COCA corpus.*³

Alternating verbs:	A Dative: ⁴ [v] [lexical NP] to [definite pronoun]	Ditransitive: [v] [definite pronoun] [lexical NP]	P (dative dative with relevant restrictions + ditransitive)
Tell	36	3713	< .01
Give	111	7982	.01
Show	35	932	.04
Send	146	1098	.12
Sell	40	152	.21
Bring	111	415	.21
Read	81	275	.23
Lend	7	176	.04
Total	567	14743	.04 (Average)

We should note that dative uses with pronominal recipients and non-pronominal themes are at times fully acceptable. There are interacting factors that lead to the use of one construction over another, including focus structure and structural priming (e.g. Bates and MacWhinney 1987; Bresnan 2007). For example, if the recipient argument is contrastive, the dative is perfectly natural as in (19):

(19) *She told the news to ME.*

Nonetheless, Table 1 demonstrates that the ditransitive is generally preferred over the dative in contexts of pronominal recipients and non-pronominal themes when a verb alternates.

Many researchers have noted that some verbs are less available for the ditransitive construction than others (e.g. Pinker 1989; Levin 1993; Goldberg 1995). “Non-alternating” verbs should have a higher probability of occurring in the dative than alternating verbs, given an information structure that favors the ditransitive. In just this way, we find that verbs traditionally classified as non-alternating display a quite different distribution (Table 2) than alternating verbs.

Table 2: *Non-alternating verbs: comparison of number of dative and ditransitive occurrences when the recipient argument is pronominal and the theme is not. Data from the COCA corpus.*

Non-alternating verbs:	Dative: [v] [lexical NP] to [definite pronoun]	Ditransitive: [v] [definite pronoun] [lexical NP]	P (dative dative_{with} relevant restrictions + ditransitive)
Explain	120	1	.99
Whisper	16	1	.94
Transfer	20	0	1.0
Return	74	11	.88
Entrust	13	0	1.0
Deliver	33	18	.65
Present	43	37	.53
Repeat	26	0	1.0
Total	345	69	.83 (Average)

Thus the data required for statistical preemption to work is readily available in corpora of sufficient size. For 6 out of 8 of the non-alternating verbs, more than 85% of the time that a ditransitive might have been expected, a dative is witnessed instead. Collectively, the distribution of alternating and non-alternating verbs is statistically distinct $\chi^2(1) = 4678.24, p < .0001$.

The astute reader might notice that two of the verbs, *deliver* and *present* appear almost half as often in the ditransitive as they do in the dative when there is a pronominal recipient and non-pronominal theme. This would seem to predict that ditransitive uses of these verbs would be relatively more acceptable than ditransitive uses of other verbs in Table 2, and this might well be the case. Whether a verb alternates is a matter of degree, and it is possible that verbs traditionally classified as non-alternating were in fact misclassified. At the same time, the distribution of even *deliver* and *present* is distinct from the average distribution of verbs that are well-known to alternate: *deliver*: $\chi^2(1) = 503.95, p < .0001$; *present*: $\chi^2(1) = 523.65, p < .0001$. This indicates that these verbs are at least more disfavored for the ditransitive than the average verb in Table 1. Moreover, there is a different preempting construction that is overwhelmingly preferred to either the dative

or the ditransitive in the case of *present*. We return to a discussion of this in section 7.2.

3. Challenges to statistical preemption

Stefanowitsch has argued against preemption being an operative notion in two recent papers (2008, this volume). Both analyses are worth revisiting.

Stefanowitsch (this volume) states his critique as an argument against “preemption by contextual clues,” but all preemption is via contextual clues since the basic idea is that construction B preempts another construction A by systematically appearing in discourse contexts that would be at least as suitable for construction A.

Stefanowitsch considers a data set of 100 sentences, and aims to investigate whether corpora contain the statistics necessary for statistical preemption on the basis of discourse functions to be viable. He suggests that there is no evidence that corpora do contain the relevant information. Specifically, he considers the probability of particular information structure distributions, given the dative construction, for verbs that alternate with the ditransitive and those that do not. That is, he investigates the probability of certain information structure properties, given the dative, as represented in (20). Stefanowitsch (this volume):

(20) $P(\text{context at least as well suited to the ditransitive} \mid \text{dative})$

Only dative examples are considered, not the combination of dative and ditransitive examples. But as we saw, the relevant probability is actually the probability of a dative, given certain information structure properties.

(21) $P(\text{dative} \mid \text{context at least as well suited to the ditransitive})$

To understand the difference between the two conditional probabilities in (20) and (21), it is helpful to think of a concrete situation. Only rarely when it has rained does a rainbow appear; therefore the probability of a rainbow, given rain (22) is low. At the same time, however, every time a rainbow occurs, there has been rain. Therefore the probability of rain, given a rainbow (23) is very high.

(22) Probability (rainbow | rain)

(23) Probability (rain | rainbow)

Since Stefanowitsch (this volume) does not provide data on the number of contexts in which a ditransitive construction is used, his data do not directly address whether learners have the relevant evidence for statistical preemption to be viable. We have already seen that such evidence is available to learners in corpus data of sufficient size.

We return in section 7 to the secondary but intriguing question that Stefanowitsch's paper (this volume) raises, which is whether the dative construction is used with fewer restrictions when appearing with non-alternating verbs than when it appears with alternating verbs.

4. Strength of preemption

Stefanowitsch (2008) aims to correlate the “strength” of statistical preemption with speakers' judgments of ill-formedness. Finding only a non-significant correlation between the two measures, he concludes that evidence for statistical preemption is lacking.

He suggests measuring the “strength of preemption” by simply determining how strongly a given verb is associated with a particular preempting construction. In particular, he assumes that preemption should be stronger when the verb *only* or predominantly occurs in a potentially preempting construction. Thus for example, *consider* used with a gerundive VP (*considered going*) would only weakly preempt *consider* appearing with an infinitive VP (**considered to go*) if it were found that the majority of uses of *consider* did not involve a VP complement at all, but were instead, for example, simple transitives.

However, the prediction is inaccurate, and in fact the proposal misses the point of preemption: only the contexts in which the semantic and information structure properties satisfy the potentially preempted construction are relevant for statistical preemption. In the case of *consider*, only contexts in which the infinitive VP would be appropriate, but the gerundive VP is used instead are relevant. The verb may appear more often in some other context and corresponding construction(s), but these contexts are not relevant to the strength of preemption. **Two forms must**

compete for being used in the same context for one to preempt the other.

The notion of “strength of preemption” makes sense, but it does not depend on how associated a verb is with the preempting construction. Instead, two factors are relevant. Strength of preemptive evidence can be determined by how close the probability in (15) comes to being equal to 1, together with the frequency of preemptive expressions.

4.1. Strength of preemption: probability and frequency of a verb in a preemptive construction.

Suppose that the first time a learner hears *explain*, she expects to hear it used ditransitively, but instead hears it datively. At that moment, the probability of witnessing *explain* in a preemptive context is 1, but only a single case has been witnessed. Clearly, the learner should not infer from a single exposure that the ditransitive were preempted for *explain*. On the other hand, if a learner hears *explain* used datively 100 times without ever hearing it used in the ditransitive, the probability hasn’t changed — it is still one — but the *confidence* of preemption should clearly be increased. Frequency plays a role in the process of statistical preemption exactly because preemption is statistical. Only upon repeated exposures to one construction in lieu of another related construction can the speaker learn that the second construction is not conventional. This requires that a given pattern occur with sufficient frequency.

We should note that the confidence of one construction preempting another is not a simple linear function of frequency: it is not likely that confidence doubles when a person hears their second example, or that their confidence increases two fold when exposed to 2000 as opposed to 1000 examples. We can capture the fact that confidence grows more slowly by appealing to the natural log function.

Thus we can separate the two factors that determine the strength of preemption as follows: Probability (24), and Confidence (25):

Probability of CxB statistically preempting CxA for verb_i:
(24) $P(\text{CxB} | \text{contexts in which CxA is at least as appropriate})$

Confidence of statistical preemption for verb_i:

(25) $\ln F(\text{CxB when CxA would be at least as appropriate})$

As represented in (25), confidence can be determined by the natural log of the frequency of CxB appearing when CxA would be appropriate. Thus the probability and confidence of statistical preemption for the verbs in Table 3 is as follows:

Table 3: *Strength of statistical preemption: a function of probability and confidence (based on frequency) of the preempting construction.*

Non-alternating verbs:	P (dative l context appropriate to ditrans)	Confidence of statistical preemption
Explain	.99	$\ln(120) = 4.78$
Whisper	.94	$\ln(16) = 2.77$
Transfer	1.0	$\ln(20) = 2.99$
Return	.88	$\ln(74) = 4.30$
Entrust	1	$\ln(13) = 2.56$
deliver	.65	$\ln(33) = 3.49$
present	.53	$\ln(43) = 3.76$
repeat	1.0	$\ln(26) = 3.26$

To summarize, we have seen that evidence that can be used by speakers to statistically preempt a non-occurring construction is available in corpora. However, the availability of the relevant data does not tell us that speakers actually make use of it. In fact, there is experimental evidence that they do.

5. Experimental evidence of statistical preemption in the domain of syntax

Statistical preemption of phrasal patterns has not received a great deal of attention in the experimental literature, except in a few notable studies. For example, Brooks and Tomasello (1999) found that children aged six or seven were less than half as likely to productively produce a novel verb in a transitive frame when the verb had been modeled in both an intransitive and periphrastic causative construction, than when it was only modeled in the simple intransitive. For example, if the child had heard both *The ball is tammimg*, and *He's making the ball tam*, then they were less likely to respond to "what's the boy doing"? with *He's tammimg*

the ball, than they were if only the simple intransitive had been witnessed.

It seems that hearing the novel verb used in the periphrastic causative provided a readily available alternative to the causative construction, statistically pre-empting the use of the latter. That is, hearing a periphrastic causative in a context in which the transitive causative would have been at least equally appropriate led children to avoid generating a transitive causative in a similar contextual situation (cf. also Brooks and Zizak, 2002).

Boyd and Goldberg (to appear) extend the experimental study of preemption via an investigation of the case of “a-adjectives.” A-adjectives are adjectives that begin with an unstressed schwa and can be morphologically segmented into *a-* plus a semantically related stem (e.g. *a-live*, *a-sleep*). Relevantly, these adjectives disprefer appearing prenominally as shown in (3) and (26):

(26) ??*the asleep boy*.

The distribution is motivated by the fact that the majority of *a-* adjectives historically were prepositional phrases, and as prepositional phrases, they could not be expected to appear prenominally. However, speakers today are generally unaware of the historical facts and so the question arises as to how the restriction can be learned. Boyd and Goldberg examined adult naturalistic production in three experiments, all of which required participants to describe scenes in which one of two animals with different adjective labels moved to a star. We used four classes of adjectives: real *a-* adjectives; nearly synonymous real non-*a-* adjectives; nonsense *a-* adjectives, and nonsense non-*a-* adjectives.

The task resulted in either a relative clause or prenominal (attributive) use of the target adjective (e.g. 27 or 28).

(27) Prenominal: *The sleepy/??asleep/?adax fox*.
(judgments based on data from Experiment 1 of Boyd and Goldberg)

(28) Relative Clause: *The fox that's sleepy/asleep/adax*.

The first experiment established that real *a-* adjectives (e.g. *asleep*) strongly disprefer prenominal use relative to non-*a-* adjectives (e.g. *sleepy*). In addition, *novel a-* adjectives (e.g. *adax*)

disprefer prenominal use relative to non-*a* adjectives (e.g. *chammy*) to a significant extent as well. This indicates that participants tentatively assimilate never-before-seen *a*-adjectives to the category of familiar *a*-adjectives. Categorization was tentative insofar as real *a*-adjectives were much less likely to occur preminally than novel *a*-adjectives were.

A second experiment investigated the role of statistical preemption. It was found that in fact witnessing two of the four novel *a*-adjectives used in a preemptive relative clause context just three times each dramatically decreased prenominal uses so that *all* four novel *a*-adjectives behaved indistinguishably from familiar *a*-adjectives in avoiding prenominal uses. Non-*a*-adjectives were unaffected. This result is striking because it not only demonstrates the effectiveness of preemption, but it also indicates that speakers are able to generalize evidence gleaned from statistical preemption to other members of the same category. In the context of the present work, this suggests, for example, that strong evidence that verbs of manner of speech are statistically preempted in the ditransitive might be generalized to new verbs of manner of speech, without having to witness the new verbs themselves in a preempting construction.

A final experiment showed that learners rationally disregard *pseudo*-preemptive input. Speakers did not display an increased avoidance of prenominal uses when exposed to pseudo-preemptive contexts like (29), presumably because they rationally attributed *adax*'s appearance in the relative clause to the complex adjective (cf. 30), rather than to *adax*.

(29) *The hamster, adax and proud of itself, moved to the star.*

(30) **The proud of itself hamster moved to the star.*

Productions in the last experiment patterned like those in the first experiment where no preemptive context was provided. Fillers were used to obscure the goal of the experiment and to guard against the effects being a simple result of structural priming. Debriefing confirmed that speakers were unaware of the manipulations.

Collectively, the Boyd and Goldberg (to appear) experiments go some way toward establishing how speakers are able to learn arbitrary distributional restrictions in their language — i.e., how they learn what not to say. Learners categorize their

input, tentatively generalizing restrictions to new members of a category. Familiar formulations can statistically preempt other formulations that have at least as appropriate functions in a given context. Providing more evidence that speakers categorize restrictions, results demonstrated that speakers extended the information gained from preemptive contexts to other instances of the same category. At the same time, speakers use statistical preemption wisely: they are impressively adept at ignoring alternative formulations when those formulations can be attributed to some irrelevant factor.

6. Limits of statistical preemption

Proponents of statistical preemption have emphasized that it cannot be the only mechanism by which we learn to avoid certain formulations. For example, Goldberg (1995: 126) points out that low frequency or novel non-alternating verbs are not addressed by statistical preemption since preemption presupposes hearing the verb multiple times. As noted, the Boyd and Goldberg (to appear) finding that speakers can generalize evidence from statistical preemption to other words of the same category is intriguing, but more work needs to be done to determine if this provides a general solution for low-frequency or novel words.

Statistical preemption does not provide a solution for cases where the target construction is so low frequency that it could not be expected to appear, or cases in which there is no construction that is closely enough related to the target construction (Goldberg 1995: 127). For this reason, *categorization* (in particular semantic and morpho-phonological classes, similarity, and type variability) has been invoked as an additional mechanism (Goldberg 1995: 126–140, Goldberg 2006: Chapter 5; Suttle and Goldberg, to appear). As Boyd and Goldberg (to appear) emphasize, a combination of categorization and statistical preemption may provide a general solution to semantically arbitrary restrictions of the sort we set out to address, as exemplified by (1) to (3).

7. Are non-alternating verbs more free in their distribution?

At the same time that Stefanowitsch's (this volume) critique fails to undermine the existence of corpus evidence for statistical preemption, it does raise a very intriguing question. Almost all, if not all of the work comparing the information structure of the

dative and ditransitive has focused on verbs that alternate, insofar as there has been an attempt to control for semantic differences (e.g. Bresnan et al. 2005; Goldberg 2006; Wasow 2002). Analyses have become quite sophisticated in determining which of two alternates actually appear (Bresnan et al. 2005). But what is a poor verb to do if it is not allowed to alternate?

Stefanowitsch (this volume) suggests the following hypothesis:

(A) The single construction that non-alternating verbs occur in should be used in a wider range of contexts than it is for verbs that alternate. (🍏🍏🍏page no to appear🍏🍏🍏)

Although Stefanowitsch ultimately argues that this prediction is unsupported, the hypothesis is informally confirmed by a comparison of (31a) and (31b):

- (31) a. *She explained the problem to me.*
 b. #*She told the problem to me.*

Intuitively, 31a is completely natural while 31b sounds somewhat odd in a neutral context.

In order to systematically investigate (A), we can ask, if we consider only instances of the dative construction, do we find a greater proportion of contexts otherwise well suited to the ditransitive for verbs that do not alternate than for those that do?

We will see that there is in fact evidence supporting (A). But Stefanowitsch goes further and claims that the single construction that a non-alternating verb occurs in “should be distributed randomly across information-structural conditions” (Stefanowitsch, this volume), as if constructions with non-alternating dative verbs should have no information structure constraints whatsoever. We will see that this secondary claim does not follow from (A), because non-alternating verbs generally have other possibilities available.

In order to investigate the claim in (A), we need to consider dative uses of alternating and non-alternating verbs, as Stefanowitsch does. But we need to ask, do we find relatively more dative uses *that would be at least as suitable for the ditransitive* when a verb does not have the option of appearing in the ditransitive?

7.1. *A closer look at Stefanowitch's 100-sentence corpus*
 (Stefanowitsch, this volume)

Stefanowitsch analyzes 50 sentences that include 8 alternating verbs and another 50 sentences that include 14 non-alternating verbs, culled from the ICE-GB corpus of one million words of spoken and written British English from the 1990s. He finds no significant differences in the factors he examines including animacy, givenness, or heaviness when comparing the dative arguments of alternating verbs to those of non-alternating verbs.

Because the data set is so small, it is well worth taking a close look to be sure all of the examples are relevant. In particular, we want to see whether there are relatively more instances of the dative with properties that would otherwise be well suited for the ditransitive in the case of non-alternating verbs when compared with alternating verbs.

It turns out that seven examples do not convey the semantics of “transfer” so they do not satisfy the semantic constraints on the ditransitive; these all involve non-alternating verbs, and are listed in (32a to 32g):

- (32) a. *she can also introduce you to the Parish Council.*
 b. *I'll introduce you to them then.*
 c. *to entrust all their savings to the ups and downs of equity investment.*
 d. *pushing himself to the limit.*
 e. *transfer the load directly to the pier.*
 f. *Transfer your hands back to A and B and repeat.*
 g. *Transfer your hands to C and D and repeat Steps 1-4.*

Notice 32a and 32b (with *introduce*) do not involve intended or metaphorical transfer, and the rest involve locative goals, not animate recipients.

There are therefore are $50-7=43$ relevant examples of non-alternating verbs and 50 examples of alternating verbs. We know that in the case of alternating verbs, the ditransitive construction is preferred over the dative construction when the recipient argument is pronominal and the theme is not (recall Table 1). This is also consistent with Stefanowitsch's data set in that only 2 out of 50 examples (4%) of alternating verbs appear with this information structure. These two examples are provided in 33a and b:

- (33) a. *send the demand to me for checking.*
 b. *she used to read Keats to herself.*

We have seen that non-alternating verbs more freely appeared in the dative when they had this same information structure (recall Table 2). This is also confirmed in Stefanowitsch's data in that seven of the 43 relevant non-alternating examples — 16% — appear with a pronominal recipient and a non-pronominal theme:

- (34) a. *does that convey anything to you.*
 b. *return the papers to me.*
 c. *who was saying all these outrageous things to you.*
 d. *I'm not saying anything to you about...*
 e. *said this to you.*
 f. *he didn't say anything to me.*
 g. *he presents his feelings to us.*

That is, verbs that do not alternate are four times more likely to appear in a dative with a pronominal recipient and a non-pronominal theme than verbs that do alternate, when sample size is controlled for in S's data set. The small numbers make a chi-square test inappropriate, but Fisher's exact test (1 tailed) demonstrates a marginal effect, $p = .08$, indicating a trend toward a different information structure profile of dative for verbs that do not alternate as compared with verbs that do. This is consistent with the intuitive judgments offered in (31) and repeated:

- (35) a. *She explained the problem to me.*
 b. *#She told the problem to me.*

The facts become even more clear when one considers a larger data set. We can consult the COCA corpus, and compare the single case of *tell* vs. *explain*, since this case allows us to hold general semantics roughly constant: both are verbs of communication, and both are routinely expressed with agent, content and recipient roles expressed or understood.

Despite the fact that *tell* is five times more frequent overall in the COCA corpus (166K vs. 31K), it is *explain* that is almost four times more likely to appear in the dative with pronominal recipient and non-pronominal theme (120 vs. 36).

Table 4: *Comparison of number of dative uses when the information structure suits the ditransitive (i.e., recipient argument is pronominal and the theme is not.) COCA corpus.*

	Dative: [v] <lexical NP> [to <definite pronoun>]	Other uses(assuming three argument sense)	Overall frequency of V
Tell	36	166,591	166,627
Explain	120	31,766	31,886

Here we see that the distribution of datives is significantly different for *tell* and *explain*: $\chi^2(1) = 428.91, p < .0001$. Thus the hypothesis that Stefanowitsch set out to test appears to be true, at least as indicated by his own data, and confirmed for the case of *tell* vs. *explain* when the larger COCA corpus is used.

An anonymous reviewer suggests that I perform the same searches reported in Table 4 for all of the verbs in Tables 1 and 2, but this is not easy to do, because we are only interested in uses of each verb that involve a theme and a recipient, since again, the semantic properties of the ditransitive have to be satisfied. Yet many of the verbs in Tables 1 and Table 2 are frequently used without an animate recipient. For example, in order to compare *transfer* and *send*, we would need to know the *relevant* base-rates of these verbs. Note that *transfer* can readily occur intransitively, or transitively with an inanimate goal (*She transferred (the box) to Chicago*), and *send* can be used transitively with an inanimate goal (*She sent the box to Chicago*). These uses are not relevant to the proportion of datives used despite conditions being better suited to the ditransitive, because these uses don't fit the semantic requirements on the ditransitive construction. But to categorize all uses of each verb in a 400 million-word corpus by hand was prohibitively time consuming.⁵

Note that in the case of *explain* and *tell* I have assumed that the overwhelming majority of both verbs involve a theme (something told or explained) and an animate recipient of the information. This assumption allows the data in Table 4 to be usefully analyzed.

8. Other options are available to non-alternating verbs

At the same time, notice that instances of (V NP PP) datives, with pronominal recipients and non-pronominal themes, are relatively rare (Table 3). And yet since recipients are animate and animates are often pronominal, and since themes are typically inanimate and inanimates are somewhat less likely to be pronominal, it is likely that the relevant contexts are themselves not altogether uncommon.

We know that English speakers overall are much more likely to produce shorter arguments before longer arguments, animate arguments before inanimate arguments, and pronominal arguments before non-pronominal arguments (Wasow 2002; Bresnan et al. 2005). Non-alternating verbs *do* appear relatively more often in the dative than alternating verbs in ways that buck these trends, but these verbs are not renegades entirely in their distribution; V NP PP datives, even with non-alternating verbs, are fairly rare in this circumstance.

8.1. “Heavy NP shift”

Interestingly, English offers speakers a reasonable alternative solution when speakers are faced with a pronominal recipient, a non-pronominal theme, and a non-alternating verb. Speakers may order the arguments using a “heavy NP shifted” version of the dative construction to produce: V PP NP. This allows the general trends of the language to be respected. It turns out that the heavy NP shifted version of the dative is more common among verbs that do not alternate, when the information structure would otherwise suit the ditransitive. For example, despite the fact that *tell* is overall more than five times as likely to occur as *explain* overall, it is *explain* that is 53 times more likely to occur with a pronominal recipient in heavy NP shift construction (106 vs. 2).

Table 5: *Comparison of alternating tell with non-alternating explain in the “heavy NP shifted” dative. COCA corpus.*⁶

Heavy NP shift Dative: [v] [to <definite pronoun>] <lexical	Overall frequency of V (assuming three argument sense)
--	--

	NP>	
Tell	2	166,627
Explain	106	31,886

It turns out that non-alternating verbs have other options as well. In fact, there are three other logical possibilities that present themselves beyond the situation already considered in (A):

(A) The single construction that non-alternating verbs occur in should be used in a wider range of contexts than it is for verbs that alternate.

(B) There may be a third (fourth or fifth) construction that is used when the non-occurring alternant construction would otherwise be preferred.

(C) Non-alternating verbs may fail to be appropriate in the situations that would strongly prefer the alternant construction, due to a semantic restriction on their meaning.

(D) A different verb may be used in the appropriate construction instead.

These possibilities are not mutually exclusive, and in fact there is reason to believe that each factor plays a role. We consider them each in turn.

8.2. *Another preempting construction.*

Completely monogamous verbs are quite rare. Dative verbs that do not combine with the ditransitive often appear in still other constructions. For example, many “non-alternating” verbs, including *present*, *provide*, *entrust* — which are among the ones Stefanowitsch (this volume) considers — readily appear with a direct object recipient and a prepositional theme as in the following examples from the COCA:

- (36) *These findings present us with the challenge of making these activities...*

- (37) *provide them with the tools to perform confidently when interviewed..*
- (38) *If you entrust me with the presidency, I will fight for you.*

The “provide with” construction potentially offers evidence of statistical preemption in exactly the way that the dative does. The only difference is that only the dative has traditionally been assumed to “alternate” with the ditransitive construction. But traditional alternations such as the dative-ditransitive need not take on an outsized role in our theorizing (Goldberg 2002). The data in Table 6 provide straightforward evidence that the “provide with” construction statistically preempts the ditransitive construction, insofar as the probability of witnessing the “provide-with” construction when a ditransitive would otherwise be appropriate approaches 1 for these verbs.

Table 6: Comparison of the “provide with” and ditransitive constructions when the recipient is pronominal and the theme is not.

“non”- alternating verbs that alternate with “provide with” construction	A. “Provide with”: [v] [definite pronoun] with [lexical NP]	B. Ditransitive [v] [definite pronoun] [lexical NP]	C. P(A A or B)
	e.g. “she provided him with a pencil”	e.g. ?“she presented him a pencil”	
<i>provide</i>	1541	7	.99
<i>entrust</i>	46	0	1
<i>present</i>	613	37	.94

8.3. Non-occurrence of the ditransitive for semantic reasons.

Option C raised another possibility. Some verbs do not alternate because their semantics is inappropriate. For example, once we recognize that the prepositional dative is part of a broader “caused-motion” construction, it is clear that many verbs that convey motion, but not transfer or means of transfer, only occur in the “caused-motion” construction. *Put* is such a case:

- (39) a. *She threw the blanket to Paul.*
 b. *She threw the blanket to the pole.*
 c. *She threw the blanket on the pole.*
 d. *She put the blanket on the pole.*

- (40) * *She put the pole the book.*

Speakers do not require preemptive evidence to learn not to produce (40), because the semantics of *put* makes it ill-suited for the ditransitive. The semantic properties of the verb must be compatible with the semantic constraints on the construction in order for the verb to appear in the construction (Ambridge et al. 2009; Goldberg 1995; Gropen et al. 1991). The “restriction” on *put* is not semantically arbitrary, so it is a different type than we set out to address (cf. 1 to 3).

8.4. Another verb statistically preempts a particular verb in a given construction.

There is one final situation to be considered, that outlined in option (D). Another verb may be used if a particular construction is appropriate in a given discourse but a particular verb does not occur in that construction. It is difficult to quantify how often this happens, since it is hard for both researchers and learners to tell when a particular verb is not used but would have been appropriate. At the same time, this surely does happen, and when it does, it provides evidence relevant for statistical preemption. For example, it is widely accepted that *kill* preempts the causative use of *die*, because *kill* is consistently heard in the transitive construction when one might have otherwise expected to hear *die* (McCawley 1978).

(41) **The man died the duck.*

(42) *The man killed the duck.*

That is, the probability of witnessing *kill*, given contexts in which someone is understood to die by means of direct causation — i.e., contexts where the simple causative and the verb *die* would otherwise be appropriate — is quite high:

(43) $P(\text{transitive } kill \mid \text{direct causation and } die)$

9. Conclusion

To conclude, just the sort of corpus evidence that would be required for statistical preemption to be viable is available to learners. This is clear once we recognize that the relevant probability for construction CxB to preempt construction CxA is:

$P(\text{CxB} \mid \text{context that would be at least as suitable for CxA})$

When this probability is high, there is evidence for statistical preemption. Learners' *confidence* that CxB statistically preempts CxA increases logarithmically, as the frequency increases of CxB being witnessed in a context that would otherwise be at least as suitable for CxA. Boyd and Goldberg (to appear), moreover, find that speakers may be able to extend evidence gleaned from statistical preemption to other members of the same morpho-semantic-phonological class.

Future research is required to determine whether confidence

must merely be above some threshold, or whether confidence actually continues to increase logarithmically as the number of tokens increases. We also need to know how stable the probability and relative confidence levels are across corpora. Ultimately it will be important to determine how probability and confidence combine to yield speaker judgments of ill-formedness. Another question concerns the age at which learners receive sufficient input to make statistical preemption viable. A non-trivial amount of data is required, and so we might expect the effects of statistical preemption of words in particular constructions to emerge relatively late (see Brooks and Tomasello 1999 for relevant age effects).

As Stefanowitsch (this volume) suggests should be the case, there *is* evidence that verbs that are arbitrarily yoked to a single construction display a greater willingness to exploit that single construction in a broader range of discourse contexts. But constructions do not end up varying *dramatically* in their information structure properties because “non-alternating” verbs generally have other options that allow them to avoid extending a construction much beyond its normal comfort zone. In particular, there are sometimes ways to use a target construction without violating its general constraints (e.g. “heavy NP” shifted version of the dative); there often exists a third construction that can be used instead (e.g. the “provide with” construction); and speakers sometimes use a distinct verb that *can* readily appear in the better-suited construction (e.g. causative *kill* instead of *die*). Thus languages tend to find reasonable solutions to multiple interacting constraints.

While much of language is semantically and historically motivated, there remain pockets of idiosyncrasy that speakers must learn. An important fact about language is that constructions are often in competition with one another when speakers produce (and comprehend) utterances (Bates & MacWhinney 1987). Statistical preemption of one construction by one or more other constructions provides an important factor in the learning of arbitrary distributional restrictions.

Received 21 May
Revision received 27 July

Princeton University

Notes

* I am grateful to Lisa Goldman for checking the COCA corpus data I collected in order to remove misclassified instances of the dative or ditransitive. I would also like to thank Anatol Stefanowitsch for graciously sharing his paper and his data set with me, and Graeme Trousdale, Thomas Hoffmann, and an anonymous reviewer for helpful comments on an earlier draft. I am grateful to Laura Suttle and Devin Casenhiser for helpful discussions. Finally, I owe Mark Davies a special debt of gratitude for making large and easily searchable corpora such as COCA available for free. The research was supported by NSF grant # 0613227. Email <adele@princeton.edu>

¹ This sentence is only acceptable if *the moon* is construed to refer to some people at the moon.

² Recall that the conditional probability $P(B|A)$ is the probability of B, given A: $P(B \& A)/P(A)$.

³ It might be tempting to think that these numbers indicate that the *ditransitive* construction should actually preempt the dative construction, for alternating verbs on the basis of Table 1, but in order to conclude that, we would need to calculate, for each verb:

$$P(\text{ditransitive} \mid \text{dative is at least as appropriate}) \approx \\ P(\text{ditransitive} \mid \text{ditransitive}_{\text{with relevant restrictions}} + \text{dative})$$

As the relevant restrictions on the dative are not the same as those on the ditransitive, this is not equal to $1 - P(\text{dative} \mid \text{dative}_{\text{with relevant restrictions}} + \text{ditransitive})$.

⁴ Searches were performed on COCA as follows:

[v] [pp*] [a*] [n]	Ditransitive
[v] [a*] [n] to [pp*]	Dative

[v] allows for all forms of the verb, pp* searches for all definite pronouns, a* searches for articles and n searches for all nouns. These formulas do not capture every non-pronominal noun (since bare nouns are non-pronominal), but they were used for ease of replication. We also removed by hand examples in which the second noun phrase was an adjunct or the first NP returned from a “ditransitive” search was an inanimate pronoun (*it*): (e.g. *I wish I could explain it the way Mary did.* was not counted as a ditransitive).

⁵ It would certainly be worth investigating how general this phenomenon is. I find the following contrasts to pattern the same as *explain* vs. *tell*. In doing a more thorough investigation, one needs to be careful to control for pragmatic felicity. Word frequency is also a issue, since we know that higher frequency words are generally judged better; this factor would work against the hypothesis, since alternating verbs tend to be higher frequency verbs.

- a. *She returned the shirt to me.*
 b. #*She gave the shirt to me.*

- a. *They transferred the film rights to her.*
 b. #*They sent the film rights to her.*

⁶ The probabilities of a shifted or non-shifted dative, given that the context would otherwise be as suitable for a ditransitive should ultimately be combined to provide evidence of statistical preemption, as follows:

P (dative (shifted or not) | context at least as suitable for the ditransitive)

This would yield a probability of .99 for *explain* (240/241) and a probability of .01 for *tell* (72/3782).

References

- Ambridge, Ben, Julian M. Pine, Caroline F. Rowland, Rebecca L. Jones & Victoria Clark. 2009. A Semantics-Based Approach to the “No Negative Evidence” Problem. *Cognitive Science* 33(7). 1301–16.
- Arnold, Jennifer E., Thomas Wasow, Anthony Losongco & Ryan Ginstrom. 2000. Heaviness vs Newness: the effects of structural complexity and discourse status on constituent ordering. *Language* 76(1). 28–55.
- Bates, Elizabeth & Brian MacWhinney. 1987. Competition, variation, and language learning. In Brian MacWhinney (ed.), *Mechanisms of language acquisition*. 157–193. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bolinger, Dwight. 1977. *Meaning and form*. London: Longman.
- Bowerman, Melissa. 1996. Argument Structure and Learnability: Is a Solution in Sight? In Jan Johnson, Matthew L. Juge & Jeri L. Moxley (eds.), *Proceedings of the twenty-second annual meeting of the Berkeley Linguistics Society: General session and parasession on the role of learnability in grammatical theory*, 454–468. Berkeley: Berkeley Linguistics Society.
- Boyd, Jeremy K. & Adele E. Goldberg. to appear. Learning what not to say: categorization and preemption in a-adjective production. *Language*.
- Braine, Martin D. S. & Patricia Brooks. 1995. Verb argument structure and the problem of avoiding an overgeneral grammar. In Michael Tomasello & William Edward Merriman (eds.), *Beyond names for things: Young children’s acquisition of verbs*. 353–376. Hillsdale, NJ: Lawrence Erlbaum.
- Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston and Wolfgang Sternefeld (eds.), *Roots: Linguistics in search of its evidential base* (Studies in Generative Grammar 96), 75–96. Berlin & New York: Mouton de Gruyter.

- Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: processing dative constructions in American and Australian varieties of English. *Language* 86(1). 186–213.
- Bresnan, Joan & Jennifer Hay. 2008. Gradient grammar: An effect of animacy on the syntax of *give* in New Zealand and American English. *Lingua* 118(2). 245–259.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayan. 2005. Predicting the dative alternative. Paper presented at the KNAW Academy Colloquium: cognitive foundations of interpretation, Amsterdam.
- Brooks, Patricia J. & Michael Tomasello. 1999. How children constrain their argument structure constructions. *Language* 75(4). 720–738.
- Brooks, Patricia & Otto Zizak. 2002. Does preemption help children learn verb transitivity? *Journal of Child Language* 29(4). 759–781.
- Chouinard, Michelle M. & Eve V. Clark. 2003. Adult reformulations of child errors as negative evidence. *Journal of Child Language* 30(3). 637–669.
- Clark, Eve V. 1987. The principle of contrast: A constraint on language acquisition. In Brian MacWhinney (ed.), *Mechanisms of language acquisition*. 1–33. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Collins, Peter. 1995. The indirect object construction in English: an informational approach. *Linguistics* 33(1). 35–49.
- Erteschik-Shir, N. 1979. Discourse constraints on dative movement. In Suzanne Laberge & Gillian Sankoff (eds.), *Syntax and semantics*. 441–467. New York: Academic Press.
- Goldberg, Adele E. 1993. Another look at some learnability paradoxes. Paper presented to the Proceedings of the 25th Annual Stanford Child Language Research Forum. Stanford, 1993.
- Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: Chicago University Press.
- Goldberg, Adele E. 2002. Surface generalizations: An alternative to alternations. *Cognitive Linguistics* 13(4). 327–356.
- Goldberg, Adele E. 2006. *Constructions at work: the nature of generalization in language*. Oxford: Oxford University Press.
- Green, Georgia M. 1974. *Semantics and syntactic regularity*. Bloomington, Indiana: Indiana University Press.
- Gropen, Jess, Pinker Steven, Michelle Hollander & Richard Goldberg. 1991. Syntax and semantics in the acquisition of locative verbs. *Journal of Child Language* 18(1). 115–151.
- Hovav, Malka Rappaport & Beth Levin. 2005. Are dative verbs polysemous? Paper presented to the Linguistics Colloquium, Princeton University, February 17, 2005.
- Levin, Beth. 1993. *English verb classes and alternations*. Chicago: Chicago University Press.
- McCawley, James. 1978. Conversational implicature and the lexicon. In Peter Cole (ed.) *Syntax and semantics 9: Pragmatics*. 245–259. New York: Academic Press.
- Oehrle, Richard. 1975. The Grammatical Status of the English Dative Alternation. Cambridge, MA: MIT PhD dissertation.
- Pinker, Steven. 1984. *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Pinker, Steven. 1989. *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

- Pollard, Carl Jesse & Ivan Sag. 1987. *Information-based syntax and semantics*. Stanford: CSLI.
- Saxton, Matthew. 1997. The Contrast Theory of negative input. *Journal of Child Language* 24(1). 139–161.
- Stefanowitsch, Anatol. 2008. Negative entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics* 19(3). 513–531.
- Suttle, Laura and Adele E. Goldberg. to appear. The partial productivity of constructions as induction. *Linguistics*.
- Thompson, Sandra. 1995. The iconicity of “dative shift” in English: Considerations from information flow in discourse. In Marge E. Landsberg (ed.), *Syntactic iconicity and linguistic freezes: the human dimension* (Studies in Anthropological Linguistics 9). 155–175. Berlin & New York: Mouton de Gruyter.
- Wasow, Thomas. 2002. *Postverbal Behavior*. Stanford: CSLI.