

# Usage Preferences: The Case of the English Verbal Anaphor *do so*

Philip Miller

EA 3967 CLILLAC-ARP, Université Paris Diderot

Proceedings of the 20th International Conference on  
Head-Driven Phrase Structure Grammar

Freie Universität Berlin

Stefan Müller (Editor)

2013

CSLI Publications

pages 121–139

<http://csli-publications.stanford.edu/HPSG/2013>

Miller, Philip. 2013. Usage Preferences: The Case of the English Verbal Anaphor *do so*. In Müller, Stefan (Ed.), *Proceedings of the 20th International Conference on Head-Driven Phrase Structure Grammar*, Freie Universität Berlin, 121–139. Stanford, CA: CSLI Publications.



## Abstract

In this paper I introduce the notion of Usage Preferences (UPs), which are statistically significant preferences in usage which can concern any aspect of linguistics. I suggest that multiple violations of UPs can have additive effects, causing grammatical sentences to be judged as unacceptable. A new judgment on sentences is proposed, the downarrow ( $\downarrow$ ) to mark sentences that are taken to be grammatical but unacceptable due to UP violations. I illustrate the idea of UPs on the basis of a discussion of the English verbal anaphor *do so*, involving both a corpus analysis and two acceptability experiments. This leads to a discussion of the relationship between grammaticality and acceptability and to remarks on the methodological importance of taking UPs into account both in linguistic theorizing and in the construction of acceptability experiments.

## 1 Introduction: Usage Preferences

Most syntacticians will have been confronted, at some point, with the following paradoxical situation. One reads a paper that proposes a constraint C on a construction X and illustrates its relevance by exhibiting occurrences of X that violate C and that indeed appear to be quite unacceptable. Yet corpus research provides examples of X violating C which appear to be perfectly acceptable. This situation can be illustrated with the verbal anaphor *do so* as construction X and the exclusion of stative antecedents as constraint C. This constraint was first suggested by Lakoff & Ross 1976 and appears to have been generally accepted. Culicover & Jackendoff 2005 propose a stronger version, namely that *do so* does not allow non-action antecedents. They provide the following examples to illustrate their claim:<sup>1</sup>

- (1) a. \*Robin dislikes Ozzie, but Leslie doesn't do so. [Stative, C&J:284, their (2a), their judgment]  
b. ?\*Robin fell out the window, but Leslie didn't do so. [Non-action event, C&J:284, their (2b), their judgment]

However one can easily find attested examples of *do so* with stative antecedents. This was first pointed out by Michiels 1978 (a paper which apparently went com-

---

<sup>1</sup>Preliminary versions of the corpus data reported here were presented at the “Linguistic Evidence” conference in Berlin on April 5 2013 and at the “Structure and Evidence in Linguistics” workshop in Stanford on April 29 2013. I would like to thank participants in both these venues for their comments. I would like to thank Gabriel Flambard, Barbara Hemforth, Anne Jugnet, and Geoff Pullum for discussion and comments. I would also like to thank Corey Cusimano for setting up the experiments on the Ibex platform and running them on Amazon’s mechanical turk. Special thanks to Emilia Ellsiepen and to Barbara Hemforth for help with the statistical analysis. All remaining errors are my sole responsibility.

<sup>1</sup>In order to clarify the interpretation of examples I underline the antecedent and double underline the anaphor.

pletely unnoticed) who, embarrassingly, cited a series of examples from articles and books written by linguists, among which the following:

- (2) The basic idea is that whenever the relation of complementary distribution holds between phones belonging to a common phoneme, it does so because the phonetic value of the phoneme depends upon the phonetic environment in which it occurs. [Stative, in Fodor, Bever and Garret, *The Psychology of Language*, cited by Michiels 1978:175]

It appears to be completely impossible to explain such an example away as a speech error. It is in fact completely natural, and would most likely not be noticed at all when reading the paragraph of the book from which it is excerpted.

More recently, Houser 2010 has corroborated the acceptability of *do so* with stative antecedents through large-scale corpus investigation and an acceptability experiment. We are therefore confronted with a paradox: why do constructed examples of *do so* with a stative antecedent, such as (1) seem to be ungrammatical when examples of apparently the same type are attested in spontaneous usage of language and felt to be perfectly acceptable? The goal of this paper is to try to explain this apparently contradictory situation in terms of USAGE PREFERENCES (UPs).

Usage Preferences are statistically significant preferences in usage which can concern any aspect of linguistics, e.g. syntax, lexical semantics, compositional semantics, discourse pragmatics, register etc. In general, it appears that a single violation of a UP has little effect on acceptability. On the other hand I suggest that violations of UPs can have additive effects, causing strong unacceptability. I will illustrate the idea of Usage Preferences with respect to *do so* and suggest that the difference between (2) and (1a) is that the former violates one UP on *do so* whereas the latter violates three of them.

## 2 Usage Preferences for finite *do so*

English has two central types of verbal anaphors. On the one hand, there are a series of complex verbal anaphors based on main-verb *do*, among which *do it*, *do this*, *do that*, *do so*. On the other hand there is Post-Auxiliary Ellipsis (PAE, more commonly known as VP Ellipsis; an alternate analysis is possible where the auxiliary is taken to be a pro-predicate, see e.g. Schachter 1978, Hardt 1993). There have been a huge number of studies on PAE,<sup>2</sup> but far fewer on *do so* (see however Hankamer & Sag 1976, Michiels 1978, Culicover & Jackendoff 2005, Houser 2010). In this paper, I will focus on *do so*, providing both corpus data and data from psycholinguistic experiments.

<sup>2</sup>Among which Sag 1976 (who initially proposed Post-Auxiliary Ellipsis as a more appropriate label), Hardt 1993, Johnson 2001, Kehler 2002.

## 2.1 Corpus data on *do so*

Houser 2010 provides corpus data on *do so* based on the American National Corpus (<http://www.americannationalcorpus.org/OANC/index.html#>). Miller 2011 provides additional corpus data from the COCA corpus (Davies 2008-, <http://corpus.byu.edu/coca/>). On the basis of the results of these studies, the following UPs can be proposed for finite *do so*:

- (3) UP1 Finite *do so* very strongly prefers to occur with non-stative antecedents. (98% of cases according to Houser 2010)
- UP2 Finite *do so* very strongly prefers to occur referring to the same state of affairs as its antecedent and hence with the same subject as its antecedent. (98% of cases according to Miller 2011)
- UP3 Finite *do so* prefers to occur with a non-contrastive adjunct. (83% of cases according to Miller 2011)

We thus have three Usage Preferences which can be satisfied or not, giving us eight possible combinations, which are illustrated in (4), using attested examples when they are available.<sup>3</sup>

- (4) a. What is most important, in the end, is to make sure that the president makes the right decisions, that he does so in a timely manner, and that they are implemented effectively. (COCA) [UP1+, UP2+, UP3+]
- b. Nathan immediately bends down to pick them up but is jostled as he does so and stumbles, breaking his fall with his right hand ... (COCA) [UP1+, UP2+, UP3-]
- c. We assume that logical thought, syllogistic analytical reason, is the necessary, right thought—and we do so because this same thought leads us to think this way. (COCA) [UP1-, UP2+, UP3+]
- d. ... a story about someone who had paid the tolls for the car behind them as a random holiday gift. DH thought that this was just cool and fun so he did so at the tollbooths we encountered in our travels. (Houser 2010:135 [UP1+, UP2-, UP3+])
- e. He delighted in Mr. Spitzer's downfall—and continues to do so. (COCA, note that this is **nonfinite do so**) [UP1-, UP2+, UP3-]
- f. Soon after BMG began restructuring its businesses, some of its rivals did so too. (COCA) [UP1+, UP2-, UP3-]
- g. ↓Mary assumes that logical thought is necessary. Peter does so because this same thought leads us to think that way. [UP1-, UP2-, UP3+]
- h. ↓Mary assumes that logical thought is necessary. Peter does so too. [UP1-, UP2-, UP3-]

<sup>3</sup>In order to clarify examples, I wavy underline the noncontrastive adjunct when it is present.

Examples with violations of only one of UP1 or UP2 (as in (4a-f)) can be found in corpora, with no intuitively clear decrease in acceptability. However, I have not been able to find examples violating both preferences and there is an intuitively strong decrease in acceptability in constructed examples of this type (as in (4g,h)). On the other hand, one can find examples violating either of UP1 or UP2, conjointly with a violation of UP3 (as in (4e,f)), and this does not seem to lead to an intuitively clear decrease in acceptability.<sup>4</sup> I therefore suggest that UP1 and UP2 are STRONG Usage Preferences whereas UP3 is a WEAK Usage Preference. This appears to correlate with the much stronger statistical strength of UP1 and UP2 in corpus data (98% as opposed to the 83% found for UP3). Violation of two strong UPs thus appears to lead to unacceptability.

## 2.2 A functional explanation for UP3

Let us now have a closer look at UP3. Consider what happens to examples like (4a,c), where UP2 is satisfied, and UP3 is as well, if one removes the non-contrastive adjunct, as illustrated in (5a,b):

- (5)
- a. What is most important, in the end, is to make sure that the president makes the right decisions, that #he does so, and that they are implemented effectively.
  - b. We assume that logical thought, syllogistic analytical reason, is the necessary, right thought—and #we do so.

Because the sentence with the anaphor refers to the same state of affairs as its antecedent, removing the noncontrastive adjunct leads to simple tautologous repetition of the previous content and thus to infelicity because of the violation of Grice's maxim of quantity.

On the other hand, an example like (4b), which does not respect UP3 (but does respect UP1 and UP2), does not have the same tautologous status because a second reference to the same state of affairs is made in order to temporally locate another event. All of the examples which do not satisfy UP3 in Miller 2011's sample are in fact of this type, and it is very clearly the overwhelmingly most frequent case of UP3 violations in corpus data. We thus see that UP3 has an obvious functional explanation, given UP2.

## 2.3 Motivations for UP1 and UP2?

By contrast with UP3, UP1 and UP2 seem much less obviously amenable to some sort of functional explanation. UP1 might be thought to stem from the fact that *do so* contains main verb *do*, which is not a stative verb (except in certain idiomatic uses), as has often been suggested in the literature. This idea provides an intuitively

<sup>4</sup>As discussed in the next section, certain violations of UP3 can lead to strong unacceptability because they violate the maxim of quantity.

satisfying explanation for the fact that *do it*, *do this*, and *do that* are clearly much more strongly unacceptable with stative antecedents, as shown by the variants of (2) and (4c).

- (6) a. ... whenever the relation of complementary distribution holds between phones belonging to a common phoneme, #it does it/this/that because ...
- b. We assume that logical thought, syllogistic analytical reason, is the necessary, right thought—and #we do it/this/that because ...

The problem is that if *do so* were synchronically compositional, one would expect to get as strong a resistance to stative antecedents as one finds with *do + it/this/that*, which *can* be argued to be compositional. So at best, UP1 can be explained as a synchronic dispreference inherited from a putative previous situation where *do so* was compositional. As for UP2, it is hard to see what kind of functional (or other) explanation might explain it. In particular, an analysis in terms of performance difficulties seems hard to imagine.

To conclude, it appears that UPs may not always be synchronically motivated. Of course, stating a UP is always an invitation to attempt to provide some sort of more general explanation for it (as was suggested above for UP3). But we must be ready to accept that some UPs might lack any relevant motivation, at least provisionally, and still have additive effects on acceptability.

### 3 UPs and grammaticality

#### 3.1 Acceptability and grammaticality

If one assumes that there exists a specifically linguistic set of cognitive capacities (innate or not), which we can call linguistic competence and which is not reducible to more general cognitive capacities, then one of the central goals of linguistic theory must necessarily be to provide a model of this linguistic competence. In order to do this, we need to have hypotheses about what it is we are supposed to model. If we are working in syntax and semantics, it means that we need to have hypotheses about what are the syntactically and semantically well-formed strings of the language. This requires, in turn a set of hypotheses about performance.

In early work, Chomsky suggested that one should base one's theory of competence on an analysis of the cases where grammaticality and ungrammaticality were obvious and that "in many intermediate cases we shall be prepared to let the grammar itself decide, when the grammar is set up in the simplest way so that it includes the clear sentences and excludes the clear non-sentences" (Chomsky 1957:14). However, very quickly, generative linguists began to use intuitive judgments of grammaticality on sentences for which such judgments were far from obvious as crucial evidence in arguing for one theory over another, rather than "letting the grammar itself decide" in such cases. The dangers of this methodology

have been clearly pointed out by numerous authors, as early as Schütze 1996 (see Gibson & Fedorenko 2013 for a recent discussion and Sprouse & Almeida 2013 for an opposing point of view.)

Many linguists (including many in the HPSG community) have tried to improve the reliability of grammaticality judgments, both by using corpus evidence and psycholinguistic experiments, in order to provide a more solid basis for establishing the domain of well-formed items that we have to model. These studies have made it clear that acceptability is a gradient notion. The question remains open as to how the gradience of acceptability connects to grammaticality. Possible hypotheses (these are not mutually incompatible) include:

- i. Grammatical sentences can be less than fully acceptable and can even be completely unacceptable.
- ii. Grammaticality is a gradient property.
- iii. Ungrammatical sentences may be acceptable.

The first of these positions is as old as generative grammar and is commonly accepted as necessary. The classical example of this type is that performance limitations can make a grammatical sentence unacceptable. The second is upheld by many proponents of probabilistic theories of grammar, corpus linguistics and cognitive grammar.<sup>5</sup> The third position has been proposed by Lyn Frazier and her collaborators (see e.g. Arregui et al. 2006), who call it the ‘recycling hypothesis’). They claim that hearers can ‘repair’ ungrammatical sentences and that the degree of perceived acceptability will depend on the complexity of the repair process (see Kertz 2013 and Miller & Hemforth 2013 for arguments against this position). This leads them to conclude that even very simple ungrammatical sentences can be judged to be acceptable.<sup>6</sup>

I follow Schütze 1996, Pullum & Scholz 2001 and Gibson & Fedorenko 2013 in taking position (i) and considering that there is a well-defined, non gradient notion of grammaticality but that various factors may intervene, making grammatically well-formed structures unacceptable. Specifically, I would like to argue here that multiple violations of UPs can lead to this result.

### 3.2 The down arrow (↓) judgment

Let us come back to the grammaticality status of the sentences in (4). I claim that all of these examples must be considered to be grammatical. If one does not accept

---

<sup>5</sup>See e.g. Pullum & Scholz 2001:26ff for discussion and references. They argue that grammaticality cannot be gradient—either grammatical constraints are satisfied or they are not—but that ungrammaticality is gradient: the more rules and/or constraints are violated, the higher the degree of ungrammaticality and consequently of unacceptability. This position seems entirely plausible to me.

<sup>6</sup>Gibson & Thomas 1999, on the other hand, show that performance mistakes can in some cases lead native speakers to judge complex ungrammatical sentences to be acceptable. This is compatible with the position taken here.

this conclusion, one of the two following positions must be assumed. Either one considers UP1, UP2, and UP3 to be constraints on grammaticality, so that all of the sentences in (4) where any of these UPs is violated, i.e. all of them except (4a), must be considered as ungrammatical. Since there does not seem to be anything especially complex about them which might lead to a performance mistake, one must adopt some position similar to Frazier's recycling hypothesis to explain why some of these are acceptable. Or one must assume that combining violations that do not individually lead to ungrammaticality can lead to ungrammaticality. This position seems to necessitate some form of gradient notion of grammaticality, which we have rejected.

I therefore suggest that we need a new type of judgment characterizing sentences which do not violate any principles of grammar but which are unacceptable because they cumulatively violate different UPs. I suggest using the down arrow ( $\downarrow$ ) to mark these sentences, as was done in (4g,h).

### 3.3 Methodological importance of UPs

Lakoff & Ross 1976 were the first to discuss *do so* in the generative literature. Because the prototype examples for PAE in the literature were cases with contrastive subjects and *too* (e.g. *Mary likes apples and Jane does too*) and *do so* was thought to be a variant of PAE, Lakoff and Ross's invented examples of *do so* typically involved contrasting subjects. Specifically, out of 33 example sentences with *do so*, 27 have contrasting subjects. Among these are ALL of the sentences that they use to argue that *do so* cannot have stative antecedents, as in (7a). This should be contrasted with the variant of their sentence given in (7b), which satisfies UP1 and UP3, and which intuitively seems much more acceptable.

- (7) a. \*Bill knew the answer, and Harry did so, too. (p.105, (8), their judgment).
- b. Bill knew the answer. He did so because he had read an article on the subject in the paper the day before.

Since Lakoff & Ross 1976, this unnatural pattern of usage has made its way into many articles and textbooks, in arguments for VP constituency and for the complement/adjunct distinction, e.g. Radford 1988:234, (23), (24); Haegeman 1991:81-82, (14), (15); Haegeman & Guéron 1999:69, (123), (124), (125); Sobin 2008 (out of 32 examples of *do so*, 26 have contrasting subjects). In all of these cases, it can be argued that the grammaticality judgments (and the results of the tests that are based on them) are strongly compromised because the baseline of the examples does not respect UP2, so that any further preference violation can lead to strong intuitions of unacceptability, possibly independently of any grammaticality problems.

Similarly, not taking UPs sufficiently into account in psycholinguistic experimentation can lead to noise in the results, making them much less statistically



significant. For instance, the materials set up by Houser 2010 to test the acceptability of stative antecedents for *do so* do not control for the identity of subjects and states of affairs, so that some of his stimuli satisfy UP2, but others do not. Similarly, some stimuli have a non contrastive adjunct, satisfying UP3, and others do not. From his discussion of his results, it appears that Houser is a bit disappointed by the weakness of the effects he finds. It is likely that part of this is due to the noise created by not paying sufficient attention to usage preferences.

## 4 Acceptability experiments for *do so*

### 4.1 Experiment 1

The UPs for *do so* discovered through corpus research, stated above in (3), suggest clear predictions for acceptability experiments. Stimuli respecting all UPs should be the most acceptable; stimuli violating several UPs should be less acceptable than those violating only one; and stimuli violating the two strong UPs (UP1 and UP2) should be less acceptable than those violating a strong UP and UP3 (which we suggested was weak). In order to test the predictions an acceptability experiment was run on Amazon's Mechanical Turk. Items were constructed on the basis of three binary factors, corresponding to UP1, UP2, and UP3:

- I. Eventive vs. stative antecedent [Evt/St]
- II. Same subject vs. different subject [SSubj/DSubj]
- III. Non contrastive adjunct vs. no adjunct [Adj/NoAdj]

Here is a typical item in its eight conditions:

- Event
  1. The President of the Senate obtained bipartisan support on this issue. He did so thanks to hours of painstaking negotiations with influential members. [Evt, SSubj, Adj]
  2. The President of the Senate obtained bipartisan support on this issue. The press has reported that he did so. [Evt,SSubj,NoAdj]
  3. The President of the Senate obtained bipartisan support on this issue. The Speaker of the House did so as well, thanks to hours of painstaking negotiations with influential members. [Evt,DSujb,Adj]
  4. The President of the Senate obtained bipartisan support on this issue. The Speaker of the House did so as well. [Evt,DSujb,NoAdj]
- State

5. The President of the Senate has bipartisan support on this issue. He does so thanks to hours of painstaking negotiations with influential members. [St,SSubj,Adj]
6. The President of the Senate has bipartisan support on this issue. The press has reported that he does so. [St,SSubj,NoAdj]
7. The President of the Senate has bipartisan support on this issue. The Speaker of the House does so as well, thanks to hours of painstaking negotiations with influential members. [St,DSubj,Adj]
8. The President of the Senate has bipartisan support on this issue. The Speaker of the House does so as well. [St,DSubj,NoAdj]

## Methods

40 items were constructed. They were distributed across 8 lists following a Latin Square design, randomly mixed with 56 distractors.<sup>7</sup> 160 participants were asked to judge the acceptability of the second sentence (explained in terms of naturalness of the second sentence as a continuation of the first) on a 7 point scale. The experiment was run using Amazon's mechanical turk and the Ibex platform for online experiments.

We modelled the data using linear mixed effect models with EVENT, SUBJECT and ADJUNCT as fixed effects and random effects (including intercept and slope) for Participant and Item. The contribution of each interaction and main effect was assessed using likelihood-ratio tests.

## Results

Acceptability ratings per condition, with error bars, are presented in figure 1.

There was a significant main effect of event ( $\text{Chi}^2(1) = 44.17, p < .001$ ), indicating that events were rated higher than states. The main effect of subject was also significant ( $\text{Chi}^2(1) = 9.02, p < .01$ ). Here, sentences with different subjects were rated higher than those with the same subject. There was also a significant main effect of adjunct ( $\text{Chi}^2(1) = 5.18, p < .05$ ) indicating that sentences with adjunct were rated higher than those without adjunct

The two-way interaction between event and subject was significant ( $\text{chi}^2(1) = 10.07, p < .01$ ), which was due to a smaller difference between event and state within the different subject condition than within the same subject condition. The two-way interaction between event and adjunct was also significant ( $\text{chi}^2(1) = 4.44, p < .05$ ), which was due to a smaller difference between event and state within the adjunct conditions. The two-way interaction between subject and adjunct was highly significant ( $\text{chi}^2(1) = 79.02, p < 0.001$ ). Here pairwise comparisons revealed that within

<sup>7</sup>In order to maximize the dispersion of the acceptability judgments on the materials of this experiment, in which none of the conditions are strongly unacceptable, distractors were chosen so as not to contain anything strongly unacceptable.

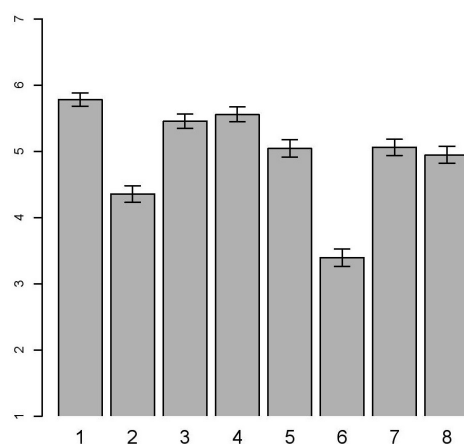


Figure 1: Acceptability ratings per condition

the adjunct condition, same subject was rated slightly higher than different subject, while within no adjunct, different subject was rated higher than same subject. This is illustrated in figure 2.

The three-way interaction was not significant ( $\chi^2(1) < 1$ ), which indicates that the two-way interactions were independent of the third factor.

## Discussion

Results for the event vs. state factor were as expected. Mean acceptability for the four event conditions was 5.29 whereas for the four state conditions it was 4.6. This should be compared with Houser 2010:66 who finds a median acceptability rating of 6 for activity predicates and of 3 for states (also on a 7 point scale). The difference may be due in part to the choices of stative predicates and also to the fact that the stimuli used here were more natural than those used by Houser. It may also be due to the level of education of the participants, as discussed below.

On the other hand, results for the same subject vs. different subject factor were completely unexpected. Sentences with same subjects were judged less acceptable on average (4.64) than those with different subjects (5.25). Several hypotheses can be considered to explain the difference between these results and the expectations stemming from the corpus data.

First, it may be the case that the sample studied by Miller 2011 was not representative. Further corpus analysis is required to see if this is in fact the case. Though a detailed analysis is beyond the scope of this paper, it was possible to use search heuristics on the COCA to corroborate the idea that *do so* disprefers contrasting subjects. Figure 3 provides the number of occurrences of *do so*, *do it* and PAE with *do*, followed by an optional comma and either *too* or *as well* and a period.<sup>8</sup> The second column provides the raw number of occurrences in the COCA;

<sup>8</sup>In order to ensure that only finite forms were being counted, only the forms *does* and *did* were

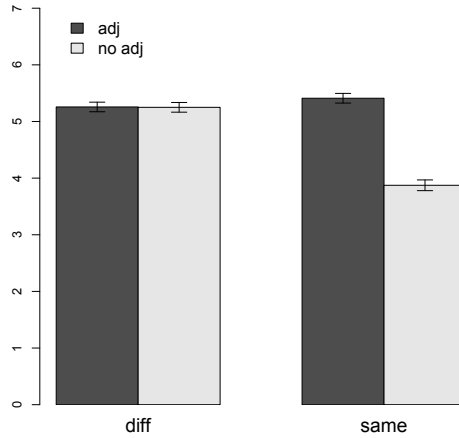


Figure 2: Acceptability ratings for different and same subjects with and without an adjunct

Pattern	N of occ	%	overall % in COCA
does/did so (,) too/as well .	18	2.65%	12%
does/did it (,) too/as well .	33	4.85%	8%
does/did (,) too/as well .	697	92.5%	80%
Total	748	100%	100%

Figure 3: Comparative frequency of *do so*, *do it* and PAE with contrasting subjects in the COCA

the third column gives the percentage of occurrences of each anaphor in the *too/as well* context and the fourth column gives the percentage of occurrences of each anaphor overall in the COCA (as reported in Miller 2011). It appears that PAE is the preferred construction in these contexts, occurring more frequently than overall in the COCA. On the other hand, both *do so* and *do it* are less frequent in this context than overall. Specifically, it appears that *do so* is four to five times less frequent in this context than it is overall in the COCA. This discussion is only meant to be suggestive. It includes a bit of noise and covers only one subcase of the different subject condition. Further detailed corpus analysis is required to shed more definitive light on the situation.

Second, the sentence pattern used in conditions 2 and 6 (the same subject/no adjunct conditions) appears to be a highly unnatural use of *do so*, and these stimuli were judged by far the least acceptable (mean ratings of 4.35 for condition 2 and of 3.39 for condition 6). On the other hand, both sentence patterns used in the

---

included in the search.

different subject conditions appear to be equally acceptable. It is thus tempting to neutralize conditions 2 and 6. This could be done by replacing the results for conditions 2 and 6 by conditions 1 and 5 respectively. This leads to a mean of 5.41 for the same subject condition and of 5.25 for the different subject condition. Or one might simply prefer to reduce the comparison to the conditions with adjuncts, which gives essentially the same results (5.41 for the same subject condition and 5.26 for the different subject condition). Better test materials must be constructed, avoiding the pattern used in conditions 2 and 6, in order to see if these hypotheses are on the right track. Specifically, this condition should have been tested using examples of the type illustrated in (4b), which, as mentioned above, are the typical instances of this case found in corpora.

A third hypothesis involves consideration of the level of education of the participants. In this experiment, I did not think to include a question on this. However a question on level of education was asked in a subsequent experiment and showed that 65.6% of participants had completed elementary school, .6% had completed junior highschool and 33.8% had completed highschool. There were no college graduates. This is potentially very important as Miller 2011 shows that register is a very significant factor in the use of *do so*: it is 13.5 times more frequent in the academic part of the COCA than in the spoken and fiction parts. It is thus possible that the typical participants in Amazon mechanical turk experiments are simply not sufficiently familiar with the usual use of *do so* to make the same acceptability judgments as speakers familiar with academic English. This might explain an overall bias towards high acceptability in this experiment (since *do so* might overall be interpreted as a marker of higher register and thus an example of “good speech”). It may also partly explain the difference between judgments of *do so* with stative antecedents in this experiment and in Houser’s: since his participants were college undergraduates they can be presumed to be more familiar with academic English and might thus be more sensitive to UP1. On the other hand, the probable lack of familiarity with academic English of the Amazon mechanical turk participants might lead to a reduced capacity to discriminate between natural and unnatural uses.

In the light of this suggestion, one might wonder why there would be a difference between the same/different subject factor (UP1) and the event/state factor (UP2). Why should Amazon mechanical turk participants be sensitive to the former, but not to the latter? A hypothesis here might be linked to the presence of main verb *do* and its eventive semantics (cf. section 2.3). This is shared by *do it*, *do this* and *do that*, which are register neutral. The relevance of the distinction between stative and eventive can thus be expected to be at least partly accessible to speakers unfamiliar with academic English, given the obvious main verb status of *do* in *do so*. On the other hand, the same subject constraint is not shared by the other verbal anaphors based on main verb *do* and speakers unfamiliar with academic English might be completely unaware of it. As mentioned in section 2.3, UP2 does not seem to have any obvious independent motivating factors that would allow it to be inferred by speakers unfamiliar with the usual uses of *do so*. In order to evaluate

this hypothesis, similar experiments will have to be conducted on a population of participants who are users of academic English.

## 4.2 Experiment 2

In the presentation of the results of experiment 1 at the HPSG conference, I suggested that register compatibility might be one reason for the difference in perceived acceptability between the examples proposed by Lakoff & Ross 1976 and Culicover & Jackendoff 2005 (cf. (7a) and (1a,b)) and the similar experimental materials of conditions 3,4,7,8 with contrasting subjects. Indeed, in the experimental materials, I chose to use *as well*, which is marked for higher register and thus very compatible with *do so*, rather than *too*, which is not clearly register marked.<sup>9</sup> Participants at the HPSG conference and other linguists I consulted felt that replacing *too* by *as well* led to a significant increase in the acceptability of the Lakoff & Ross and Culicover & Jackendoff examples.

In order to test the relevance of these observations, an experiment was run on Amazon's mechanical turk. Items were constructed on the basis of two binary factors:

- I. Event vs. state [Evt/St]
- II. *too* vs. *as well* [too/as well]

Here is a typical item in its four conditions:

1. A.—Kate read the book. B.—Karen did so too. [Evt/too]
2. A.—Kate read the book. B.—Karen did so as well. [Evt/as well]
3. A.—Kate owned an apartment. B.—Karen did so too. [St/too]
4. A.—Kate owned an apartment. B.—Karen did as well. [St/as well]

## Methods

20 items were constructed. They were distributed across 4 lists following a Latin Square design, randomly mixed with 60 distractors. 80 participants were asked to judge the acceptability of the B's response (explained in terms of its naturalness as a response to A's initial statement) on a 7 point scale.

<sup>9</sup>A more detailed look at the corpus data reported in the table presented in figure 3 suggests that this is on the right track. The following table distinguishes the cases of *too* and *as well* initially grouped in column 2:

Pattern	N of occ	<i>too</i>	<i>as well</i>
does/did so (,) too/as well .	18	5	13
does/did it (,) too/as well .	33	30	3
does/did (,) too/as well .	697	629	68
Total	748	664	84

It appears that while *as well* is on the order of 10 times less frequent than *too* with PAE and *do it*, it is more frequent with *do so*.

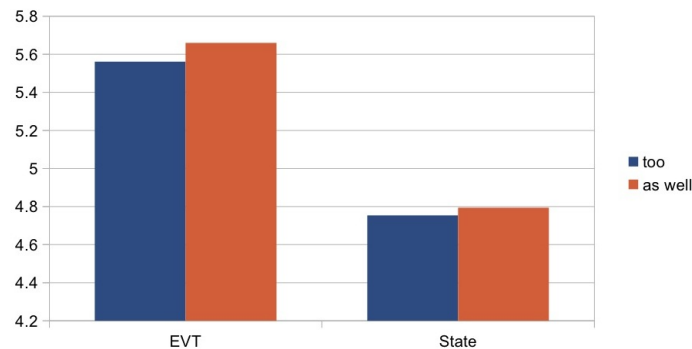


Figure 4: Acceptability ratings for Evt/St and *too* vs. *as well*

## Results

The results are shown in figure 4. Once again items with event antecedents were judged to be significantly more acceptable (mean: 5.6) than with state antecedents (mean: 4.78). This effect was highly significant ( $\chi^2(1)=17.245, p<.001$ ). However there was no significant main effect of *too* vs *as well* ( $p>.24$ ).

## Discussion

The experiment further corroborates the relevance of the event/state distinction for antecedents of *do so*. However, contrary to expectations, no effect of the choice between *too* and *as well* was found. As with experiment 1, this might be a consequence of the probable lack of familiarity of the participants with academic English. Further experiments on subjects who *are* familiar with academic English are necessary to corroborate this conjecture.

## 5 Some remarks on nonfinite *do so*

The corpus data and experiment reported in Houser 2010 shows that the effect of stativity on acceptability is reduced when *do so* is in a nonfinite form. Houser follows up on a suggestion of Huddleston et al. 2002 that this is due to the impossibility of PAE in these contexts.<sup>10</sup>

Miller 2011 points out that the proportion of nonfinite over finite uses of *do so* in the COCA can be estimated to be much higher than for other common verbs, so that nonfinite forms are strongly overrepresented. Specifically, *to do so* is four times more frequent in the COCA than one would expect, given the frequency of

<sup>10</sup>Certain varieties of British English allow what is apparently auxiliary *do* to appear in nonfinite contexts, e.g. (i) So far, everything that could go wrong has done. (*The Guardian*, 26/05/2001, I thank L. Haegeman for this example); (ii) —Does Mr Charley Newton live here? —He might do (BBC, *Westway*, 23 nov 2001). Such examples are impossible in American English.

finite *do so* and the average proportion of finite to non finite uses for high-frequency verbs.

This suggests that we are confronted here with a typical Saussurean systemic effect: for two constructions A (PAE) and B (*do so*) that are in competition in an environment E (finite clauses), if B can occur in a environment E' that precludes A (nonfinite clauses), UPs that favor A over B in E will lead to reduced unacceptability if they are violated in E'. Houser 2010 provides another example of this type of Saussurean systemic effect illustrated in the following constructed example:

- (8) a. The students who know French best do so because they lived in France for a year. (Houser 2010:4,(8b))
- b. \*The students who know French best do because they lived in France for a year.

He points out that when the antecedent of *do so* is in a relative clause on the subject of *do so* then alternation with PAE is once again impossible. His experimental results show a significant improvement in the acceptability of sentences with stative antecedents in this configuration.

## 6 Conclusion

In this paper, I have tried to argue that it is necessary to take Usage Preferences into account when analyzing linguistic phenomena. Since these preferences govern verbal behavior, they must be taken to be part of linguistic competence, and as such are a necessary part of any complete analysis of a speaker's knowledge of language. On the other hand, I have argued that violating UPs does not affect grammaticality. However, it appears that acceptability judgments *can* be affected by not respecting UPs and, more specifically, that the effects of UP violations can be additive (as it has been argued that UPs do not necessarily result from performance difficulties, this provides a further case of additive effects beyond those linked to performance problems, such as those discussed in Hofmeister et al. 2014)). It is therefore methodologically very important to take Usage Preferences into account, both when using introspective acceptability judgments and when setting up experimental materials. One of the flaws that make the results of experiment 1 reported above difficult to interpret is precisely that I did not follow my own advice, using a dispreferred sentence type as materials in two conditions.

I have also suggested that UPs might be divided into strong and weak UPs. One might speculate that such a binary distinction is not sufficient and that what is needed is continuous statistical weighting of UPs, which might be directly linked to their statistical strength in language use. The effects on acceptability of multiple violations of UPs might then be predictable on the basis of some combinatorics based on the weighting of the UPs involved, the exact nature of which is left for further investigation.



The line of investigation proposed here can be understood as a strategy for dealing with the gradience of acceptability judgments (see e.g. Keller 2001, Sorace & Keller 2005) on a basis similar to much current work on probabilistic approaches to grammar (see e.g. Bod et al. 2003, Manning 2003, Bresnan 2007). However, contrary to much work that takes grammaticality itself to be a gradient property, I have tried to suggest an approach which accounts for gradience of acceptability based on a nongradient conception of grammaticality, explaining differences in acceptability between equally grammatical sentences on the basis of a calculus of UPs.

Another important conclusion of the study is the importance of register (which might itself be modeled in terms of UPs) and of the fact that different groups of speakers might not be equally competent in all registers (an idea which will be obvious to sociolinguists).

The idea of UPs may not at first sight be appealing to many specialists of theoretical syntax and semantics. There seems to be something unsatisfying about simple stipulatory statements about usage. Obviously, anyone would prefer to be able to explain any observed UP on the basis of more general properties (as was suggested for UP3 in section 2.2 above). This leads to two observations. First, in order to even raise the question of explaining a UP it is necessary to have observed it. This in and of itself makes the study of such UPs relevant, as they raise intriguing questions for theoretical studies. Second, even if it turns out that a UP is not amenable to any more general synchronic explanation, it is still a part of linguistic competence, and as such must be described in any complete theory of knowledge of language. It is also crucial to take them into account in order to avoid mistakenly classifying as ungrammatical structures that are in fact only dispreferred.<sup>11</sup>

## References

- Arregui, Ana, Charles Clifton Jr., Lyn Frazier & Keir Moulton. 2006. Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language* 55. 232–246.
- Bod, Rens, Jennifer Hay & Stefanie Jannedy. 2003. *Probabilistic syntax*. Cambridge, MA: MIT Press.
- Bresnan, Joan. 2007. A few lessons from typology. *Linguistic Typology* 11(1). 297–306.
- Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.

---

<sup>11</sup>See Payne et al. 2013 who argue that one must distinguish being ungrammatical and being a dispreferred alternate of another construction, arguing that a classical unacceptable example like *the one of physics* is simply dispreferred with respect to *the physics one* and that it is the existence of a preferred alternate which causes the dispreferred one to be felt to be unacceptable.

- Culicover, Peter & Ray Jackendoff. 2005. *Simpler syntax*. Oxford: Oxford University Press.
- Davies, Mark. 2008-. The corpus of contemporary american english: 450 million words, 1990-present. <http://corpus.byu.edu/coca/>.
- Gibson, Edward & Evelina Fedorenko. 2013. The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes* 28. 88–124.
- Gibson, Edward & James Thomas. 1999. Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes* 14. 225–248.
- Haegeman, Liliane. 1991. *Introduction to government and binding theory*. Oxford: Blackwell.
- Haegeman, Liliane & Jacqueline Guéron. 1999. *English grammar: A generative perspective*. Oxford: Blackwell.
- Hankamer, Jorge & Ivan A. Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7. 391–426.
- Hardt, Daniel. 1993. *Verb phrase ellipsis: Form, meaning, and processing*: University of Pennsylvania dissertation. Distributed as IRCS Report 93-23.
- Hofmeister, Philip, Laura Staum Casasanto & Ivan A. Sag. 2014. Processing effects in linguistic judgment data: (super-)additivity and reading span scores. *Language and Cognition* to appear.
- Houser, Michael John. 2010. *The syntax and semantics of do so anaphora*. Berkeley, CA: University of California, Berkeley dissertation.
- Huddleston, Rodney, Geoffrey K. Pullum et al. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Johnson, Kyle. 2001. What VP ellipsis can do, and what it can't, but not why. In Mark Baltin & Chris Collins (eds.), *The handbook of contemporary syntactic theory*, 439–479. Cambridge: Blackwell.
- Kehler, Andrew. 2002. *Coherence, reference, and the theory of grammar*. Stanford: CSLI Publications.
- Keller, Frank. 2001. *Gradience in grammar: experimental and computational aspects of degrees of grammaticality*. Edinburgh: University of Edinburgh dissertation.
- Kertz, Laura. 2013. Verb phrase ellipsis: The view from information structure. *Language* 89. 390–428.

- Lakoff, George & John Robert Ross. 1976. Why you can't do so into the sink. In James D. McCawley (ed.), *Syntax and semantics 7: Notes from the linguistic underground*, 101–111. New York: Academic Press. First semipublished in 1966.
- Manning, Christopher D. 2003. Probabilistic syntax. In Rens Bod, Jennifer Hay & Stefanie Jannedy (eds.), *Probabilistic linguistics*, 289–341. Cambridge, MA: MIT Press.
- Michiels, Archibald. 1978. A note on the relation between agent and stativity. *Neophilologus* 62. 172–177.
- Miller, Philip. 2011. The choice between verbal anaphors in discourse. In I. Hendrickx, S. Lalitha Devi, A. Branco & R. Mitkov (eds.), *Anaphora processing and applications: 8th discourse anaphora and anaphor resolution colloquium, daarc 2011*, vol. 7099 Lecture Notes in Artificial Intelligence, 82–95. Berlin: Springer.
- Miller, Philip & Barbara Hemforth. 2013. Verb phrase ellipsis with nominal antecedents. Ms. Université Paris Diderot.
- Payne, John, Geoffrey K. Pullum, Barbara C. Scholz & Eva Berlage. 2013. Anaphoric *one* and its implications. *Language* 89. to appear.
- Pullum, Geoffrey K. & Barbara C. Scholz. 2001. Model-theoretic and generative-enumerative syntactic frameworks. In Philippe de Groote, Glyn Morrill & Christian Retoré (eds.), *Logical aspects of computational linguistics*, 17–43. Berlin: Springer.
- Radford, Andrew. 1988. *Transformational grammar: A first course*. Cambridge: Cambridge University Press.
- Sag, Ivan A. 1976. *Deletion and logical form*. Cambridge, MA: MIT dissertation.
- Schachter, Paul. 1978. English pro predicates. *Linguistic Analysis* 4. 187–224.
- Schütze, Carson T. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.
- Sobin, Nicolas. 2008. *Do so* and VP. *Linguistic Inquiry* 39. 147–160.
- Sorace, Antonella & Frank Keller. 2005. Gradience in linguistic data. *Lingua* 115. 1497–1524.
- Sprouse, Jon & Diogo Almeida. 2013. The empirical status of data in syntax: A reply to Gibson and Fedorenko. *Language and Cognitive Processes* 28. 222–228.