

SYNTACTIC FORM AND DISCOURSE ACCESSIBILITY

GREGORY WARD

ANDREW KEHLER

Northwestern University

University of California San Diego

Abstract

One of the central issues in studies of reference is the relationship between morpho-syntactic form and the accessibility of discourse referents. However, most of the work in this area has been concerned primarily with reference to entities; considerably less work has addressed the relationship between syntactic form and the discourse accessibility of events. In this paper, we consider forms of event reference that involve what Bolinger (1972) termed ‘identifier *so*’, including the *do so* construction. In particular, we consider those referring expressions whose antecedents are not (syntactically-matching) VPs. In so doing, we respond to and counter a recent criticism of our previous work by Fu et al. (2001) and discuss several factors that appear to affect the accessibility of events evoked by nominalizations. Our account is based on a corpus of naturally-occurring and felicitous examples with *do so* that are apparently disallowed under the Fu et al. account.

1. Introduction

One of the central issues in studies of reference – be they theoretical, computational, or psycholinguistic – is the relationship between morpho-syntactic form and the accessibility of discourse entities. For instance, work in theoretical linguistics concerned with COGNITIVE OR INFORMATION STATUS (Chafe 1976; Prince 1981; Ariel 1990; Gundel et al. 1993; Lambrecht 1994; inter alia) has attempted to account for the felicity of a particular referential form in a given discourse context and, in particular, the factors that affect the accessibility of referents in that context. As is well known, one such factor is the syntactic

position in which a referring expression occurs, e.g. the oft-noted claim that entities referenced in subject position are typically more salient than those mentioned in positions that are lower on the obliqueness hierarchy (objects, arguments, adjuncts, etc). Similarly, computational linguists have sought to determine the factors that contribute to accessibility as a basis for developing algorithms for pronoun resolution (Brennan 1987; Lappin & Leass, 1994; Strube 1998; Mitkov 2002; inter alia). Finally, psycholinguists have used a variety of experimental methods to tease apart the seemingly competing factors that determine how people assign referents to pronouns, often with contradictory results (Crawley et al. 1990; Smyth 1994; Stevenson et al. 1994; Chambers & Smyth 1998; inter alia).

It is safe to say that most of the work in this area has been concerned primarily with reference to entities; considerably less work has addressed the relationship between syntactic form and the discourse accessibility of events. These two areas of inquiry differ in an important respect. When considering the accessibility of entities evoked by nominals, the primary concern has been the syntactic position in which that nominal occurs. In the case of events, however, one must also consider the effect of the particular syntactic form used to evoke the event itself: whether it was evoked by an active voice clause, a passive voice clause, or any of a variety of more marked sentential constructions, including gerunds and even full-fledged NP nominalizations. As we will see, this choice of form may also affect the accessibility of the event in question.

A comprehensive account that considers the relationship between the full range of syntactic forms available for evoking events in discourse and the full set of referring expressions available for subsequent reference to those events would take us well beyond the scope of this paper. Instead, we will focus here on a particular class of referring expressions: those that involve what Bolinger (1972) termed ‘identifier *so*’. Examples of identifier *so*, used preverbally and as part of the *do so* construction, are shown in (1) and (2):

- (1) “And with complete premeditation [they] resolved that His Imperial Majesty Haile Selassie should be strangled because he was head of the feudal system.” He was *so strangled* on Aug. 26, 1975, in his bed most cruelly. (*Chicago Tribune* 12/15/94)

- (2) As an imperial statute the British North America Act could be amended only by the British Parliament, which *did so* on several occasions. (Groliers *Encyclopedia*)

Previous accounts of *so* anaphora have treated it as a form of SURFACE ANAPHORA (Hankamer & Sag 1976; inter alia) which requires the presence of an appropriate syntactic VP antecedent for its interpretation. In such a model, the issue of morpho-syntactic form and relative accessibility simply does not arise: either there is a suitable VP antecedent available or there is not. However, examples like (2) are problematic for this view, since an active voice occurrence of *did so* is used felicitously to refer to an event evoked from a passive clause. As such, the active voice VP required by a surface anaphoric account – *amend the British North America Act* – does not occur in the prior discourse and is thus unavailable as an antecedent. In what follows, we will primarily be interested in antecedents of *so* anaphora that involve such mismatches.

In fact, in our previous work (Kehler & Ward, 1995; 1999) we have argued that *do so* does not qualify as a surface anaphor in that it imposes no restrictions on the particular syntactic form of its antecedent. We have argued instead that, like other forms of anaphora, *do so* is interpreted in terms of purely semantic referents within the hearer's discourse model. However, in a recent paper, Fu et al. (2001) take issue with this account and maintain that *do so* is in fact a surface anaphor. Moreover, they claim that the felicity of an antecedent expression in conjunction with the anaphor *do so* is a reliable diagnostic for the presence of a VP in that expression, and indeed use this diagnostic to argue for a particular syntactic analysis of process nominalizations.

In this paper, we extend our previous analysis and respond to Fu et al.'s proposal by demonstrating that compatibility with *do so* anaphora under closer examination does not provide evidence for their syntactic analysis. Indeed, using another class of nominalizations not addressed by Fu et al. (in addition to the evidence cited in Kehler & Ward 1999), we show that reference with *do so* is sensitive to the relative salience of the evoked event, with the morpho-syntactic form of the antecedent being but one of many factors that determine the felicity of such reference. As such, the question of when an event associated with a nominalization is sufficiently accessible for subsequent

reference with identifier *so* is considerably more complex than the state of affairs described by Fu et al.

In the next section, we briefly review the account of *so* anaphora from Kehler & Ward (1995, 1999), which aims to provide a unified and compositional account of both preverbal *so* and *do so* despite their curious and idiosyncratic anaphoric properties. In Section 3, we discuss the alternative account put forth by Fu et al., and respond to and counter their criticism of our earlier account. We follow up this analysis in Section 4 with a discussion of a corpus of naturally-occurring and felicitous examples of *do so* with nominalized antecedents that are disallowed under the Fu et al. analysis. These examples bear a striking resemblance to acceptable examples involving so-called ANAPHORIC ISLANDS at the nominal level per the analysis of Ward et al. (1991); in both cases felicity of reference cannot be accounted for by appeal to morphosyntactic considerations alone. Among the variety of factors that appear to affect the accessibility of events evoked by such nominalizations, we discuss three that stand out in our data: semantic transparency, modification, and genericity.

2. *Properties of Identifier ‘So’*

Identifier *so* is used to refer to a contextually salient event of the type denoted by the verb it modifies.¹ It may appear in either preverbal or postverbal position, as illustrated in (3a–b) respectively, or in postverbal position as part of the *do so* construction, illustrated in (3c).

¹ As such, none of the following uses of *so* are identifier:

- preposed propositional *so*: So it seems. So you say. So it is.
- postverbal propositional *so*: I think/suppose/say/believe so.
- veridical *so*: Is that so?
- consequential *so*: A: He’s a pig. B: So you’re **not** going out with him after all?
- particle *so*: So, how long have you been at Northwestern?
- ‘queer’ *so*: ‘I’ve come to the conclusion,’ he told me, ‘that I’m not really “so” at all. I much prefer girls.’ At this date the cant word among homosexuals for their proclivities was ‘so’. (OED)
- additive *so*: Bill gave a speech and so did Hillary.
- intensifier *so*: There are **so** many uses of ‘so’!
- generation X *so*: I’m like, so going out with him.
- and so on and so forth...

- (3) Secretary Powell spent two months lobbying the United Nations very hard on Iraq.
- a. By so lobbying, he was able to say that the U.S. had at least tried to get a war resolution passed.
 - b. By lobbying so, he was able to say that the U.S. had at least tried to get a war resolution passed.
 - c. By doing so, he was able to say that the U.S. had at least tried to get a war resolution passed.

As we have argued elsewhere (Kehler & Ward 1999), the two positional variants of identifier *so* display quite different properties. For example, only postverbal *so* permits exophoric reference, as shown in (4a–b):

- (4) [Andy is holding a newborn baby with one hand behind his head, and shows Gregory]
- a. Andy: By holding him *so*, you add support to his developing neck muscles.
 - b. Andy: #By *so* holding him, you add support to his developing neck muscles.
 - c. Andy: #By *doing so*, you add support to his developing neck muscles.
 - d. Andy: By holding him *this way*, you add support to his developing neck muscles.

Moreover, although it appears postverbally, the *so* of *do so* has precisely the same semantic and pragmatic properties of preverbal identifier *so*. For instance, as shown by (4b–c), both preverbal *so* and *do so* require that the referent event be LINGUISTICALLY EVOKED, that is, explicitly introduced in the discourse via a linguistic expression. In contrast, postverbal *so* is simply a manner adverbial anaphor and, like other such anaphors, permits situational evocation (compare 4a and 4d).

Another distinction between preverbal *so* and *do so* on the one hand and postverbal *so* on the other is that only the latter specifically requires an evoked manner. Consider again example (1), repeated below as (5).

- (5) “And with complete premeditation [they] resolved that His Imperial Majesty Haile Selassie should be strangled because he was head of the feudal system.” He was *so strangled* on Aug. 26, 1975, in his bed most cruelly.

Replacing *so strangled* with *strangled so* in this passage results in a decidedly odd reference, in the same way as the manner adverbial *in that way* does:

- (6) ?? “And with complete premeditation [they] resolved that His Imperial Majesty Haile Selassie should be strangled because he was head of the feudal system.” He was *strangled so/in that way* on Aug. 26, 1975, in his bed most cruelly.

This oddness presumably results from the fact that there is no evoked manner associated with the strangling event; modifying the passage to include one explicitly (e.g., “with a rope”) results in perfect felicity for both postverbal *so* and the adverbial *in that way*:

- (7) “And with complete premeditation [they] resolved that His Imperial Majesty Haile Selassie should be strangled with a rope. He was *strangled so/in that way* on Aug. 26, 1975, in his bed most cruelly.

Crucially, however, both *do so* and *so doing* pattern with preverbal *so* in not requiring an evoked manner:

- (8) “And with complete premeditation [they] resolved that His Imperial Majesty Haile Selassie should be strangled because he was head of the feudal system.” And they in fact *did so*, on Aug. 26, 1975, in his bed most cruelly.

- (9) “And with complete premeditation they strangled His Imperial Majesty Haile Selassie because he was head of the feudal system.”
By so doing, they forever changed the course of Ethiopian history.

Based on this evidence, we take preverbal *so* and the *so* of *do so* to be positional variants of the same event anaphor.² We discuss these two forms further in the sections that follow.

2.1 Preverbal ‘*so*’

Examples of the preverbal *so* construction are provided in (10)–(12).

- (10) “And with complete premeditation [they] resolved that His Imperial Majesty Haile Selassie should be strangled because he was head of the feudal system.” He was *so strangled* on Aug. 26, 1975, in his bed most cruelly. (=1)
- (11) In fact, in substantiating these fears, Judge Bork again essentially concedes that economic freedom is a component of the Constitution: “We already have clauses that could be used to protect economic freedom – and were *so used*.” (*Wall Street Journal*)
- (12) In fact, it is interesting that, in English, at least, there is virtually no marking of an NP with respect to the Discourse-status of the entity it represents. Of course, if an NP is indefinite and is thereby understood as evoking something Hearer-New, we can infer Discourse-New. However, if it is not *so marked*, then, with one exception, we cannot tell from its form whether it has occurred before in the discourse. (Prince 1992:304)

² This is not to say that preverbal *so* and *do so* are interchangeable; the former is considerably more restricted than the latter. There appears to be a very general constraint at play in that the referent has to be more specific than the denotation of the verb modified by *so*, although this need not be a manner as with postverbal *so*. Hence, example (5) would be odd without the appearance of the *because* clause. In the case of *do so*, however, this specificity constraint is met automatically, since any event is more specific than a generic ‘doing’.

The information status of the referent event associated with the preverbal *so* construction is constrained in a number of ways. First, the referent event associated with preverbal *so* must be SALIENT in the mental model of the hearer; compare (13) with (10):

- (13) With complete premeditation, they resolved that His Imperial Majesty Haile Selassie should be strangled because he was head of the feudal system. They also resolved to commit a variety of other violent acts, although those would come somewhat later. Obviously, these people were very prone to violence.
- a. #Selassie was so strangled on Aug. 26, 1975, in his bed most cruelly.
- b. Selassie was strangled on Aug. 26, 1975, in his bed most cruelly.

The substantial material that intervenes between the first mention of the strangling event and the subsequent reference to it in (13a) renders the reference infelicitous. Note that the salience criterion implies that the event must be HEARER-OLD in the sense of Prince (1992); that is, felicitous use of preverbal *so* requires that the speaker have evidence that the hearer is familiar with the event at the time of the utterance.

Second, as we have already argued, the referent must be linguistically evoked; that is, reference with identifier *so* cannot be exophoric, as shown by (14) (in addition to (4b–c)):

- (14) [A and B together have just witnessed Haile Selassie being murdered by strangulation]
- a. A: #He was so strangled most cruelly.
- b. A: He was strangled most cruelly.

Finally, as with other anaphoric expressions, the referent event associated with preverbal *so* may be INFERRABLE in the sense of Prince (1981, 1992):

- (15) Regarding a possible Elvis Presley stamp, Postmaster General Frank notes that anyone *so* honored must be “demonstrably dead” for 10 years. (*Wall Street Journal*)

Here, there is no honoring event that is explicitly introduced into the discourse. Instead, a hearer must reason from the mention of *a possible Elvis Presley stamp* that putting someone’s picture on a stamp constitutes a kind of honoring event. Although the chain of reasoning required for this interpretation is quite complex, reference to the inferred event with *so* is fully felicitous.

2.2 ‘Do so’

As argued above, preverbal identifier *so* is related to the *do so* construction, illustrated in (16):

- (16) Sam sold his stock on insider information, and Martha did so too.

Previous accounts of this intensively investigated construction have noted its seemingly idiosyncratic syntactic and anaphoric properties (Lakoff & Ross 1966; Anderson 1968; Bouton, 1970; Halliday & Hasan 1976; Hankamer & Sag 1976; Sag & Hankamer 1984; Quirk et al. 1985; Miller 1990; Ward et al. 1991; Cornish 1992; Fu & Roeper 1993; Dechaine 1994; Fu et al. 2001). Here we will only discuss those properties of the construction that bear directly on our analysis.

First, it is clear that the *do* of *do so* is main verb *do* and not auxiliary *do* (Hankamer and Sag 1976; Sag 1976; Quirk et al. 1985; Miller 1990; Dechaine 1994; inter alia).

- (17) a. Hillary did so.
 b. *Did Hillary so?
 c. Did Hillary do so?
 d. Hillary did.
 e. Did Hillary?

As (17a–e) show, the main verb *do* of *do so* does not undergo auxiliary inversion, unlike auxiliary *do* (Miller 1990). Likewise, the main verb *do* permits *so*, unlike auxiliary *do*:

- (18) a. Dubya filed a lawsuit, and Al did too. [auxiliary *do*]
 b. Dubya has filed a lawsuit, and Al has too.
 c. Dubya will file a lawsuit, and Al will too.
- (19) a. Dubya filed a lawsuit, and Al did so too. [main verb *do*]
 b. *Dubya has filed a lawsuit, and Al has so too.
 c. *Dubya will file a lawsuit, and Al will so too.

Here, we see that it is the main verb *do*, and not the auxiliary form of *do*, that co-occurs with *so*. Furthermore, *do so* (and its participial variant *so doing*) is more restricted in its use than auxiliary *do* (Lakoff & Ross 1966; Anderson 1968; Bouton, 1970; Hankamer & Sag 1976; Quirk et al. 1985; Miller 1990; Dechaine 1994; inter alia).

- (20) a. Al knows French, and Tipper does too. [auxiliary *do*]
 b. Al knows French, and so does Tipper. [auxiliary *do*]
 c. #Al knows French, and Tipper does so too. [main verb *do*]
 d. #Al knows French, and in so doing, is not popular with Republicans. [main verb *do*]

In these examples, we see that a strongly stative verb like *know* permits auxiliary *do* with ellipsis, but disallows main verb *do*.³

As for the status of the *so* of *do so*, one might be tempted to categorize it as an NP given its superficial similarity to other event anaphors, such as *do it*

³ Hankamer and Sag (1976) suggest that the key distinction is between active (nonstative) antecedents (which are said to be compatible with *do so*) and stative VP antecedents (which are not), but as pointed out by Quirk et al. (1985) and Dechaine (1994), a more fine-grained categorization is required. We thank an anonymous reviewer for helpful comments with respect to this issue.

and *do that*. However, as others have noted (Bouton 1970; Hankamer & Sag 1976; Quirk et al. 1985), there is strong distributional evidence that the *so* is categorially an adverb. First, note that it does not passivize like NPs do:

- (21) a. *...and so was done by Hillary.
 b. ...and it was done by Hillary.
 c. ...and that was done by Hillary.

Second, unlike NPs, it does not cleft:

- (22) a. It is that which Hillary did.
 b. *It is so which Hillary did.
 c. What Hillary did was that.
 d. *What Hillary did was so.

Given these facts, we conclude that the *do* of *do so* is an intransitive main verb and the *so* is an adverbial modifier.

In the analysis presented in (Kehler & Ward 1999), all of the previously discussed properties of *do so* are captured by a compositional account in which the *do* of *do so* denotes the most semantically general type of event, and the *so* marks the information status of that event as both salient and discourse-old, the latter of which excludes situationally-evoked referents. Seen in this way, *do so* (and its variant *so doing*) are simply forms of standard hyponymic reference (cf. Miller 1990), as can be seen by considering the following progression of examples:

- (23) The hit man dispensed with his mob boss by shooting him in broad daylight, with plenty of witnesses around.
 a. By so shooting him, the hit man established himself as his victim's likely successor.
 b. By so murdering him, the hit man established himself as his victim's likely successor.

- c. By doing so, the hit man established himself as his victim's likely successor.

Continuations (23a–c) illustrate *so* anaphora using the same verb (*shooting*), a more general hyponym (*murdering*), and ultimately the most general hyponym (*doing*), respectively. This progression from specific to general event type directly parallels the situation for reference to entities with nominal anaphors, as illustrated by the different referential options given in (24).

- (24) Chris took his poodle to the vet. The poodle / the dog / the animal was in a lot of pain.

Given this analysis, we can account for the fact that *do so* is restricted in its range of reference: stative events like *know* and *own* are simply not 'doings' as they do not involve agency (cf. 20a–d). Thus, as a form of hyponymic reference to a general type of doing event, use of *do so* is incompatible with such predicates. In this way it parallels its nominal counterparts *do it* and *do that*, although for those referring expressions the constraint arises from semantic restrictions that transitive *do* places on its direct object.

Moreover, this analysis accounts for the fact that *do so* is not a surface anaphor in the sense of Hankamer & Sag (1976), i.e. that it does not require an antecedent of any particular syntactic form. Instead, *do so* again patterns referentially like its nominal counterparts *do it* and *do that* in that they are all used to refer to events in (the speaker's representation of) the hearer's mental model of the discourse.

Indeed, our account correctly predicts the existence of mismatches between the morpho-syntactic form of the anaphor *do so* and that of its antecedent. In what follows, we briefly present several classes of naturally-occurring examples of anaphor-antecedent mismatches. The first class consists of mismatches between the voices used each clause; consider (25–26):

- (25) Section 1 provides the examples to be derived by Gapping, and a formulation of Gapping capable of *doing so*. [= deriving the examples] (text of Neijt 1981)

- (26) As an imperial statute the British North America Act could be amended only by the British Parliament, which *did so* on several occasions [= amended an imperial statute]. (=2)

In these examples, the passive voice of the antecedent sentence does not match the active voice of the anaphor; under a surface anaphoric account of *do so*, such mismatches are predicted to be ill-formed. Likewise, in our second class of examples (27–28), the antecedent expression has been nominalized, and hence (under most accounts; see Section 3) there is not the requisite VP antecedent available as required by a surface anaphoric account:

- (27) The defection of the seven moderates, who knew they were incurring the wrath of many colleagues in *doing so*, signaled that it may be harder to sell the GOP message on the crime bill than it was on the stimulus package. [= defecting] (*Washington Post*)
- (28) Even though an Israeli response is justified, I don't think it was in their best interests to *do so* right now. [= respond] (token provided by Dan Hardt)

Finally, in examples (29–30) the antecedents are 'split':

- (29) The survey results, released by county officials this week, also showed that most of the teenagers who drank alcohol, smoked marijuana or had sex started *doing so* between the ages of 13 and 16. [= drinking alcohol / smoking marijuana / having sex] (*Washington Post*)
- (30) Fortunately, the first person to die in 1990 and the first couple to file for divorce in 1990 were allowed to *do so* anonymously. [= die / file for divorce] (text of Roeper 1990, cited by J. McCawley's "1990 Linguistic Flea Circus" and discussed by Dalrymple et al. 1991)

Again, we see that the conjoined VP required under a surface-anaphoric account of *do so* is not available here; a (presumably quite dubious) cut-and-paste

operation would be necessary to combine two VPs from separate sentences and add the appropriate connective (i.e., *or* in (29) and *and* in (30)) to arrive at an antecedent of an appropriate form.

To summarize thus far, *do so* is a compositional anaphoric construction consisting of intransitive main verb *do*, denoting the most general of event types, and identifier *so* marking the information status of that event. As our corpus of naturally-occurring data reveals, these constructions are not restricted to antecedents of a particular syntactic form.

3. *A Recent Syntactic Proposal*

Despite the existence of such mismatches, Fu et al. (2001), henceforth FRB, attempt to resurrect the notion that *do so* is a surface anaphor. While aware of the problems for such an analysis raised by our earlier work (Kehler & Ward 1995), they maintain that *do so* categorically requires a syntactic VP antecedent, and use this assumption to argue that PROCESS NOMINALIZATIONS such as those in (27–28) contain a VP in their syntactic representations.⁴ While space does not allow us to present their analysis in detail, we will briefly discuss three problems that we see with their argumentation and the conclusions they draw from it.

3.1 *Comparison with ‘Deep Anaphora’*

FRB cite the contrast between (31–32) as evidence that *do so* requires a underlying syntactic VP as an antecedent:

- (31) His removal of the garbage in the morning and Sam’s doing so in the afternoon were surprising. (= FRB’s 42b)
- (32) *Kim’s accident in the morning and Sue’s doing so in the evening were not coincidences. (= FRB’s 43b)

⁴ Process nominalizations are deverbal nouns that denote an event of the type associated with the nominalized verb. For instance, *defection* in (27) denotes a ‘defect’ event and likewise *response* in (28) denotes a ‘respond’ event. Process nominalizations are thus distinguished from RESULT nominalizations (e.g., *invention*, used to refer to an object itself and not the act of inventing it), and ROLE nominalizations (e.g., *murderer* denotes the agent of a murder event and *employee* denotes the theme of an employ event; see below).

Both *removal* in (31) and *accident* in (32) are event-denoting, but only (31) is acceptable with *doing so*. FRB conclude that the difference must therefore be syntactic: The nominalization *removal* must contain a VP in syntax to license *do so*, whereas *accident*, which is not a nominalization, does not.

If this is the reason for the contrast, however, then the contrast should disappear when *do so* is replaced by an indisputably non-surface anaphor such as *do it*, which imposes no requirement for a syntactic VP antecedent. However, the judgments in fact stay the same with this replacement:

- (33) His removal of the garbage in the morning and Sam's doing it in the afternoon were surprising.
- (34) *Kim's accident in the morning and Sue's doing it in the evening were not coincidences.

Thus, while there is a contrast in event accessibility between nominalizations and event-denoting nouns, this contrast cannot be a direct result of a putative requirement that there be a VP antecedent for *do so*. If that were the case, deep anaphors like *do it* would not show precisely the same contrast. Thus, the distinction between (31) and (32) offers no evidence (one way or the other) bearing on the issue of whether nominalizations contain a VP in syntax.

3.2 *Non-Process Nominalizations*

By design, FRB's analysis applies to process nominalizations, which presumably excludes other types such as role nominalizations. But a corpus search reveals many examples of felicitous, naturally-occurring examples of precisely this type:

- (35) One study suggests that almost half of young female smokers do so in order to lose weight. [= smoke]⁵
- (36) The majority of horse riders do so purely for leisure and pleasure. [= ride horses]⁶

⁵ www.ustrek.org/odyssey/semester1/111800/111800madwomen.html

⁶ www.league.uk.com/news/media_briefings/2002/may_2002/17_may_02_a_livery_yard.htm

- (37) AmericaNet.Com, its officers, directors or employees are not responsible for the content or integrity of any ad. Sellers/buyers/subscribers/investors do so at their own risk. [= sell/buy/subscribe/invest]⁷
- (38) Data from the Retirement Survey reveals that 5% of early retirees do so because of the ill health of others. [= retire early]⁸

In response, FRB could conceivably argue that role nominalizations also have an underlying VP structure. But for many such cases, such an analysis is simply untenable:

- (39) # My computer does so faster than yours. [= compute]
- (40) # The boat's propeller failed to do so, and now we're stuck. [= propel]

The problem with this aspect of FRB's analysis is that their distinction is a categorical one: a VP is either present in the syntax or it is not. Moreover, we know of no independent evidence to the effect that some role nominals incorporate VP syntax and others do not, nor do we believe that such evidence exists. However, what we do find is that role nominalizations display a clear gradience with respect to compositionality, and we will argue in Section 4 that this compositionality in part determines the accessibility of the referenced event.

3.3 *Other Syntactic Mismatches*

As we observed earlier in Section 2.2, the data that challenge the notion that *do so* requires a syntactic antecedent are not limited to cases involving nominalized antecedents. For instance, several previously cited examples – (25–26) – involve syntactic mismatches in which there is no plausible VP antecedent. Consider (26), repeated below as (41):

⁷ www.americanet.com/Classified/sendad.html

⁸ www.npi.org.uk/reports/Active_Ageing.pdf

- (41) As an imperial statute the British North America Act could be amended only by the British Parliament, which *did so* on several occasions [= amended an imperial statute].

About this particular example, FRB say:

it is not clear that it seriously jeopardizes the claim that *do so* requires a VP/V' antecedent...That in the first conjunct the direct object is occupied by a trace, rather than a full NP may very well turn out to be immaterial for the licensing of the anaphor *do so*. (2001:572–573)

However, FRB do not provide the details necessary to adequately evaluate this possibility and its ramifications.⁹ Further, they never make explicit their assumptions about how *do so* is interpreted in the first place. The question requires attention because in a standard surface anaphoric theory (Hankamer & Sag 1976), the requirement for a matching syntactic antecedent results from a deletion process that applies during production (or, alternatively a reconstruction process that applies during interpretation). As such, it needs to be explained why a form like *do so* would impose a syntactic requirement in light of the fact that it is not associated with an ellipsis site.

The only argument FRB provide in support of their response to the well-formedness of example (41) is an alleged contrast with adjectival antecedents, which, they claim, are presumably worse because they do not involve a trace. Consider (42):

- (42) ?? This act turned out to be amendable, and the British Parliament did so in its last session.

But a considerably more acceptable variant is readily constructed:

⁹ Their phrase “may very well turn out” suggests that these details have not been worked out by them, nor will we attempt to do so here. In any case, it is certainly not obvious to us how such an analysis could be made to be consistent with both FRB’s goals and the broad range of facts in the literature on ellipsis and event reference, including constructions such as antecedent-contained deletion which require that constituents be reconstructed with traces left intact.

- (43) After the British Parliament found out that the act was amendable, they elected to do so at their very first opportunity.

Thus, although more details about their syntactic account will be necessary before it can be fully evaluated, we find it doubtful that such account can be made to handle the relevant set of facts through independently-motivated mechanisms.¹⁰

FRB also do not address cases in which *do so* is felicitous with a split antecedent – such as examples (29), (30), and (37) – in which a suitable antecedent is not available:

- (44) Fortunately, the first person to die in 1990 and the first couple to file for divorce in 1990 were allowed to *do so* anonymously. (=30)

Again, it is hard to see how such examples could be accommodated in a purely syntactic treatment. On the other hand, our analysis predicts the patterning of such reference with pronominal reference, which is likewise compatible with split antecedents:

- (45) The first person of the year to die is usually listed in the newspaper, and so is the first couple to file for divorce. In a rare show of respect, this year *their* names were kept private.

As it is clear that pronouns are not surface anaphors, FRB's analysis fails to capture the analogous behavior of the two forms.

Lastly, FRB criticize us for failing to note an alleged parallel between *do so* anaphora and adverbial modification, based on examples such as the following:

¹⁰ FRB cite other examples in which deverbal adjectives are unacceptable as antecedents of *do so*, concluding that “the slight improvement of the *-able* cases [i.e., (42)] may have to do with whether the verbal meaning is preserved” (573, fn. 24). We agree, and take such data to provide evidence for our analysis (see Section 4, where we make similar arguments about felicitous cases involving role nominalizations), whereas such gradations in the data appear to call into question the categorical predictions made by FRB's analysis.

- (46) a. Kim's explanation of the problem to the tenants thoroughly (did not prevent a riot). (= FRB's 1a)
- b. The occurrence of the accident suddenly (disqualified her). (= FRB's 2a)

However, the vast majority of our informants reject these sentences out of hand, on the relevant readings in which the adverb modifies the nominalization and not the matrix verb. Therefore, we consider the fact that our account fails to establish such a parallel to be a feature of the analysis, and not a drawback. On the other hand, we find that FRB's analysis fails to capture a different generalization: that the *so* in *do so* is the same (identifier) *so* found in the productive *so+V* construction, which, as we pointed out earlier, clearly does not require a syntactically-matching antecedent (cf. 15).

4. *Nominalizations, Accessibility, and Anaphoric Islands*

One of the conclusions we reached in the last section is that the sharp line drawn by FRB's account between process and nonprocess nominalizations – the former which are claimed to be possible antecedents for *do so* and the latter not – does not reflect the distribution of felicitous reference one actually finds in the data. Instead, what we have found through our corpus study is that there are examples of both types that range from perfectly felicitous to totally unacceptable. An adequate account will therefore have to allow either type of nominalization to serve as an antecedent *in principle* – as our analysis does – and will accordingly have to explain the gradient felicity of the data through other means. In our discourse-based analysis, the felicity of *do so* with a nominalized antecedent boils down to the question of whether the nominalization renders its underlying event sufficiently salient, and not whether a syntactic VP antecedent exists in the context.

The debate between syntactically-based and discourse-based analyses of *do so* is strikingly reminiscent of a previous debate in the literature concerning the existence of so-called ANAPHORIC ISLANDS as they pertain to nominal-level reference. Arguing against the existence of a purely structural constraint prohibiting reference to entities evoked from “word-internal” positions (Postal 1969), Ward et al. (1991) found that reference to such entities is indeed possible

under the right pragmatic conditions, providing numerous naturally-occurring examples such as (47) and (48).

- (47) Although casual *cocaine* use is down, the number of people using *it* routinely has increased. (= Ward et al.'s 22a)
- (48) Patty is a definite *Kal Kan* cat. Every day she waits for *it*. (= Ward et al.'s 20b)

An analysis of a corpus of naturally-occurring uses of *do so* with role-nominalized antecedents (including examples for over 25 different verbs) shows that such reference patterns directly with the anaphoric island data discussed by Ward et al. While various pragmatic factors may conspire to render a particular use of either an entity anaphor (e.g. pronouns) or an event anaphor (e.g. *do so*) infelicitous, those same factors in another context can also permit such usages, and thus in neither case can their occurrence be ruled out by syntactic considerations alone.¹¹

The most striking commonality between anaphoric island violations at the nominal level and reference to role-nominalized events with *do so* is the central role of the SEMANTIC TRANSPARENCY (Ward et al. 1991) or ANALYZABILITY (Langacker 2000) of the antecedent. Langacker defines analyzability as “the extent to which speakers are cognizant of the presence and the semantic contribution of component symbolic elements” (2000:127). To use his example, if we were to coin the term *flinger*, the hearer must use the meanings of its morphemes to derive the word meaning as “something that flings”; thus it is fully analyzable. On the other hand, the present-day meanings of *computer*, *freezer*, and *propeller* go well beyond the meanings “something that computes/freezes/propels”, to the point where the corresponding underlying events almost certainly receive considerably less activation upon mention. Indeed, the underlying events for *ruler*, *pliers*, and *plumber* probably do not get activated at all by the mere mention of these words. As noted by various authors (Aronoff, 1976; Bauer 1983; Langacker, 2000), there is a long-term

¹¹ Ward et al. (1991) ultimately drew a different conclusion about *do so* anaphora, stating that “it follows that no discourse context will render *do so* anaphora felicitous with non-VP antecedents” (p. 462), a conclusion not supported by the current corpus-based study.

tendency for words to lose their analyzability as they gain conventionalized meanings.¹² We would therefore expect role nominalizations to activate their underlying event representations to varying degrees – very much so for *flinger* and perhaps not at all for *ruler* – and thus the acceptability of using *do so* to refer to such underlying events would vary accordingly.¹³

This pattern appears to be borne out by our corpus. Examples (49)–(52) demonstrate felicitous uses of *do so* with highly transparent role nominalizations:

- (49) Although most collectors do so for the sheer fun of the hobby, the question “What’s the value of my collection?” does arise. [= collect]¹⁴
- (50) Most antler hunters do so recreationally, says Mr. Hovinga. [= hunt antlers]¹⁵
- (51) Residents should include contact information so that respondents may do so directly. [= respond]¹⁶
- (52) Users of information from any Applied Discovery web site do so at their own risk. [= use information from any Applied Discovery web site]¹⁷

That is, *smokers* are people who smoke; *users of information from any Applied Discovery web site* are people who use information from any Applied Discovery web site. On the other hand, our searches yielded no comparable examples with the nominalizations *computers*, *freezers*, *propellers*, or others

¹² As Langacker notes, the effect is similar to the fading of metaphors, in which ultimately speakers are unaware of the metaphorical basis for a word or phrase.

¹³ Note, therefore, that the felicity of *do so* anaphora for a given nominalization might actually change over time, if that nominalization is evolving toward a conventionalized meaning.

¹⁴ www.leuchtturm.com/en/prod/en_overview.htm

¹⁵ www.sublette.com/scj/v4n26/v4n26s2.htm

¹⁶ www.artists-in-residence.com/parlor/messages/9/9.html?0

¹⁷ www.applieddiscovery.com/termsConditions/default.asp

that are similar in terms of their (low) degree of semantic transparency. That is, we found no cases like (53) and (54) despite our best efforts.

(53) #Most computers do so quickly these days.

(54) #Today's boat propellers do so with great force.

In sum, the transparency of a nominalization affects the extent to which the representation of the underlying event is activated, which in turn determines the extent to which this representation can be accessed with an event referential form such as *do so*.

If accessibility is indeed the operative notion for establishing the felicity of event anaphors, then we would expect other factors besides transparency to also affect accessibility, perhaps in more subtle ways. One such factor is the effect of modifiers on a role nominalization. Adjectival modifiers of role nominalizations, for instance, can describe properties of either the entity denoted by the nominalization or the underlying event. Consider the following example from our corpus, in particular the last sentence:

(55) Gulden ignored a race official and jumped the tape marking the finish area to shake hands with his runners. He was the only coach in the shutes, a coach whose instincts have always told him this moment is important. After 25 years, he is remarkably more tenacious than ever. Other coaches show up at meets in jacket and tie, assigning their assistants to points on the course. Gulden shows up in sweats and puts in a few thousand meters himself, running from point to point. The greatest teachers do so by example.¹⁸

Here, the adjective is used to modify the underlying teaching event: *greatest teachers* are people who teach the greatest. On the other hand, a *tall teacher* describes a teacher who is tall, and not someone who teaches in a tall way. As

¹⁸ www.departments.bucknell.edu/pr/BucknellWorld/1995-1/gulden.html

such, we see a distinction in the accessibility of the event depending on whether the accompanying adjective modifies it or not:

- (56) a. The greatest teachers do so by example.
 b. ?The tallest teachers do so by example.

It would appear that the adjective in (56a) increases the accessibility of the teaching event enough to support subsequent reference with *do so*, whereas *tall* in (56b) causes attention to be placed on the teacher as an entity, thereby reducing the accessibility of the underlying event.¹⁹

Finally, an examination of our corpus suggests that another factor affecting the accessibility of events is the genericity of the event in question. That is, we have found that generic (or quantified) role nominalizations evoke events that are more accessible than those evoked by non-generic (or non-quantified) nominalizations. We hypothesize that the reason for this preference is because generics are typically used in situations in which the underlying event plays a key role in the main assertion of the sentence, hence highlighting the event. For instance, the pragmatic force of a sentence like (57a) centers on the fact that the people John has to dine with smoke. While non-generic nominalizations can also be used this way (57b), they also commonly serve other purposes in which the event is less central, such as to merely single out a referent of an NP amongst alternatives (57c).

- (57) a. John often has to dine with smokers.
 b. John had to dine with a smoker yesterday – poor John.
 c. John dined with that smoker over there yesterday.

As such, a non-generic role nominalization may not create the same expectation for the centrality of the event that a generic does. While our searches cannot be considered exhaustive, our preliminary analysis suggests that non-generic examples are fewer, although importantly they do exist:

¹⁹ It is worth noting that even though *teachers* in (56a) is a suitable antecedent for *do so*, it clearly does not allow the type of adverbial modification that FRB would predict if their analysis were extended to role nominalizations: **The teacher of the students greatly (is tall)*.

- (58) While it is certainly sad that he died early, it has no lasting eternal consequences. This is quite the opposite for the killer, however. Assuming that the killer did so maliciously and not accidentally, there are serious eternal consequences for his act. [= kill]²⁰

What these three factors – semantic transparency, modification, and genericity – have in common is that they all affect the extent to which the event that underlies the use of a role nominalization becomes activated and accessible during semantic interpretation. This, in turn, affects the degree to which that event is available for subsequent reference with an event anaphor. We suspect that there are other such factors as well for which we would make the same prediction: the more accessible the underlying event, the more felicitous subsequent reference to it using *do so* will be. These findings, taken together with the results of Ward et al.'s (1991) analysis of anaphoric islands, indicate that both entity and event reference are governed by essentially pragmatic – and not morphosyntactic – factors.

5. *Conclusions and Future Research*

We have shown that referential forms that involve identifier *so* – including the *do so* construction – do not impose any purely syntactic restrictions upon their antecedents. Instead, like other event-referential expressions, they are used to refer to events that have been evoked in the hearer's discourse model. The broad range of syntactic forms that can license identifier *so* anaphora – including role nominalizations – seriously calls into question the claim of Fu et al. (2001) that *do so* is a diagnostic for the existence of an underlying verb phrase in the syntactic representation of process nominalizations.

On the other hand, there is a connection between syntactic form and discourse accessibility and, in turn, between discourse accessibility and felicity of reference with *do so*. We have provided a (nonexhaustive) set of factors that affect the accessibility of events underlying a class of antecedents that are disallowed by FRB's approach – role nominalizations – which in certain circumstances allow subsequent reference with *do so*. However, many important questions remain concerning the role of these and other (as yet

²⁰ www.frontpage2000.nmia.com/~nahualli/LDStopics/Theology/4school.htm

unidentified) factors in determining the salience of events evoked by various linguistic forms. Whatever those factors ultimately turn out to be, we hope to have demonstrated the crucial role of discourse factors in the interpretation of event anaphora.

References

- Anderson, Stephen R. 1968. "Pro-sentential Forms and Their Implications for English Sentence Structure". Technical Report NSF-20, Aiken Computation Laboratory, Harvard University.
- Ariel, Mira. 1990. *Accessing Noun Phrase Antecedents*. London and New York: Routledge.
- Aronoff, Mark. 1976. *Word Formation in Generative Grammar*. Cambridge: MIT Press.
- Bauer, Laurie. 1983. *English Word-Formation*. Cambridge: Cambridge University Press.
- Bolinger, Dwight. 1972. *Degree Words*. The Hague and Paris: Mouton.
- Bouton, Lawrence P. 1970. "Do So: Do + Adverb". *Studies Presented to R. B. Lees by his Students* ed. by J. M. Sadock, 17–38. Edmonton: Linguistics Research.
- Brennan, Susan E., Marilyn W. Friedman, and Carl J. Pollard. 1987. "A Centering Approach to Pronouns". *Proceedings of the 25th Meeting of the Association for Computational Linguistics*, 155–162.
- Chafe, Wallace L. 1976. "Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View". *Subject and Topic* ed. by Charles N. Li, 25–55. New York: Academic Press.
- Chambers, Craig C. and Ron Smyth. 1998. "Structural Parallelism and Discourse Coherence: A Test of Centering Theory". *Journal of Memory and Language* 39:593–608.
- Cornish, Francis. 1992. "So Be It: The Discourse-Semantic Roles of So and It". *Journal of Semantics* 9:163–178.

- Crawley, R. A., R. J. Stevenson, and D. Kleinman. 1990. "The Use of Heuristic Strategies in the Interpretation of Pronouns". *Journal of Psycholinguistic Research* 19:245–264.
- Dalrymple, Mary, Stuart M. Shieber, and Fernando Pereira. 1991. "Ellipsis and Higher-Order Unification". *Linguistics and Philosophy* 14:399–452.
- Dechaine, Rose-Marie. 1994. "Ellipsis and the Position of Subjects". *Proceedings of NELS 24*, 47-63.
- Fu, Jingqi and Thomas Roeper. 1993. "The VP within Nominalizations: Evidence from Adverbs and the VP Anaphor *Do-So*". Paper presented at the Syntax-Morphology Connection, MIT.
- , -----, and Hagit Borer. 2001. "The VP within Process Nominals: Evidence from Adverbs and the VP Anaphor *Do-So*". *Natural Language and Linguistic Theory* 19:549–582.
- Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski. 1993. "Cognitive Status and the Form of Referring Expressions in Discourse". *Language* 69:274–307.
- Halliday, M.A.K. and Ruqaiya Hasan. 1976. *Cohesion in English*. English Language Series, Title No. 9. London: Longman.
- Hankamer, Jorge and Ivan Sag. 1976. "Deep and Surface Anaphora". *Linguistic Inquiry* 7:391–426.
- Kehler, Andrew and Gregory Ward. 1995. "On the Anaphoric Status of *Do So*". Paper presented at the Annual Meeting of the Linguistic Society of America, New Orleans.
- , 1999. "On the Semantics and Pragmatics of 'Identifier *So*'". *The Semantics/Pragmatics Interface from Different Points of View (Current Research in the Semantics/Pragmatics Interface Series, Volume I)* ed. by Ken Turner, 233–256. Amsterdam: Elsevier.
- Lakoff, George and John Robert Ross. 1966. "Criterion for Verb Phrase Constituency". Technical Report NSF-17, Aiken Computation Laboratory, Harvard University.

- Lambrecht, Knud. 1994. *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.
- Langacker, Ronald W. 2000. *Grammar and Conceptualization*. Berlin: Mouton De Gruyter.
- Lappin, Shalom and Herbert Leass. 1994. "An Algorithm for Pronominal Anaphora Resolution". *Computational Linguistics* 20:535–561.
- Miller, Philip. 1990. "Pseudogapping and *Do So* Substitution". *Papers from the 26th Regional Meeting of the Chicago Linguistic Society*, 293–305.
- Mitkov, Ruslan. 2002. *Anaphora Resolution*. London: Longman.
- Neijt, Anneke. 1981. "Gaps and Remnants – Sentence Grammar Aspects of Gapping". *Linguistic Analysis* 8:69–93.
- Postal, Paul. 1969. "Anaphoric Islands". *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society*, 205–239.
- Prince, Ellen F. 1981. "Toward a Taxonomy of Given-New Information". *Radical Pragmatics* ed. by Peter Cole, 223–255. New York: Academic Press.
- , 1992. "The ZPG Letter: Subjects, Definiteness, and Information-Status". *Discourse Description: Diverse Analyses of a Fundraising Text*, ed. by Sandra Thompson and William Mann, 295–325. Amsterdam & Philadelphia: John Benjamins.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. New York: Longman.
- Sag, Ivan and Jorge Hankamer. 1984. "Toward a Theory of Anaphoric Processing". *Linguistics and Philosophy* 7:325–345.
- Smyth, R. 1994. "Grammatical Determinants of Ambiguous Pronoun Resolution". *Journal of Psycholinguistic Research*, 23:197–229.
- Stevenson, Rosemary J., Rosalind A. Crawley, and David Kleinman. 1994. "Thematic Roles, Focus, and the Representation of Events". *Language and Cognitive Processes* 9:519–548.

- Strube, Michael. 1998. "Never Look Back: An Alternative to Centering". *Proceedings of the 17th International Conference on Computational Linguistics and the 36th Annual Meeting of the Association for Computational Linguistics (Coling-ACL '98)*, Montreal, Canada.
- Ward, Gregory, Richard Sproat, and Gail McKoon. 1991. "A Pragmatic Analysis of So-called Anaphoric Islands". *Language* 67:439–474.