A Pragmatic Analysis of So-Called Anaphoric Islands

Gregory Ward; Richard Sproat; Gail McKoon

A PRAGMATIC ANALYSIS OF SO-CALLED ANAPHORIC ISLANDS

GREGORY WARD   RICHARD SPROAT   GAIL McKOON

Northwestern   *AT&T Bell*   *Northwestern*
University   *Laboratories*   *University*

It is commonly assumed that words are grammatically prohibited from containing antecedents for anaphoric elements, and thus constitute 'anaphoric islands' (Postal 1969). In this paper, we argue that such anaphora—termed OUTBOUND ANAPHORA—is in fact fully grammatical and governed by independently motivated pragmatic principles. The felicity of outbound anaphora is shown to be a function of the accessibility of the discourse entity which is evoked by the word-internal element and to which the anaphor is used to refer. The morphosyntactic status of the antecedent is but one factor affecting the accessibility of that entity. A series of psycholinguistic experiments support the analysis.*

INTRODUCTION

1. For over twenty years, various attempts have been made to rule out word-internal antecedents for anaphoric elements. The first such attempt is found in Postal 1969, where contrasts such as the one between 1a and 1b are discussed (p. 230):

(1) a. Hunters of animals tend to like them. [them = animals]
  b. *Animal hunters tend to like them.

To account for the deviance of examples like 1b, Postal argued that words such as *animal hunters* constitute a type of ANAPHORIC ISLAND—'a sentence part ... which cannot contain the antecedent structure for anaphoric elements lying outside' (1969:205). In particular, he proposed the following constraint on what he termed OUTBOUND ANAPHORA: for any word (W1), no anaphor could have as an antecedent another word which is either 'part of the sense of' W1 or morphologically related to W1.

While Postal's observations concerning so-called anaphoric islands were originally cited as evidence for the theory of Generative Semantics, these observations have more recently been cited as evidence for particular views of the relation between morphology and syntax. What is common to these disparate theories is the assumption that there exists some kind of GRAMMATICAL prohibition against the kind of anaphora illustrated in 1b.

In this paper we argue that outbound anaphora is not ruled out by any principle of grammar: morphemes in word-internal positions, for example, may serve as antecedents for subsequent anaphora. Our analysis presupposes a sharp distinction between syntax and pragmatics. In particular, we assume that a genuinely ungrammatical construction is ungrammatical in all (nonmetalinguistic) contexts, and cannot be 'amnestied' by pragmatic or discourse factors. Given this assumption, we maintain that outbound anaphora is fully gram-

matical and governed by independently motivated pragmatic principles. In this way, our approach is similar to that of Reinhart 1983, in which it is argued that, aside from cases of bound anaphora, the grammar need not make any special statement about the referential possibilities of anaphoric elements.

For the purposes of this study, we adopt a conventional view of the notion 'word'. We will consider a word to be any combination of a stem and affixes (normally written as one orthographic word in English), or any compound (which may consist of more than one orthographic word in English). This usage of the term is consistent with most of the work in morphology, including Matthews 1974, Aronoff 1976, Bauer 1983, and Mohanan 1986, inter alia.

We begin with a review of previous studies of anaphoric islands in general and outbound anaphora in particular, pointing out inadequacies. Next, we present our pragmatic account of outbound anaphora, and argue that the interpretability of an anaphor is a function of the relative accessibility of the discourse entity to which the anaphor is used to refer; the morphosyntactic status of the antecedent of the anaphor is only one factor which affects the relative accessibility of that entity. As part of our discussion we will review the results of a series of psycholinguistic experiments that support our analysis.

PREVIOUS LITERATURE

2.1. ANAPHORIC ISLANDS AND GENERATIVE SEMANTICS. To the best of our knowledge, Postal (1969) was the first to claim that—as he put it—reference both into and out of words is ungrammatical. Consider his examples of outbound anaphora in 2:[1]

(2) a. *Max is an *orphan* and he deeply misses *them*. (orphan = 'a child whose *parents* have died') (Postal 1969:206, ex. 3a)

b. *The best *pork* comes from young *ones*. (pork = 'meat from *pigs*') (Postal 1969:226, ex. 100b)

c. *Max wanted to *glue the boards together* but Pete wanted to *do so* with tape. (glue = '*fasten* with glue'). (Postal 1969:212, ex. 35b)

d. *McCarthyites are now puzzled by *his* intentions. (Postal 1969:213, ex. 42b)

e. *The best *wombat*meat comes from young *ones*. (Postal 1969:226, ex. 100a)

f. *Smokers really shouldn't *do so*. (Postal 1969:217, ex. 65b)

[1] In these and all subsequent examples, we shall adopt the convention of italicizing intended coreferential expressions, with the following stipulations: (i) whenever a word-internal expression is phonologically or orthographically unmodified within the containing word, we italicize just the portion of the word which corresponds to the intended antecedent (e.g. *Bush* supporters, *flut*ist, *New York*er, *smoke*r); (ii) if the containing word is not so clearly segmentable, we italicize the entire containing word (e.g. *Belgian*, *Glaswegian*, *second*). Furthermore, we shall represent greater than normal intonational prominence (where relevant) with small capitals. Finally, in our review of previous studies we shall be using the annotations of unacceptability used by the original authors (usually '*'). Elsewhere, however, we shall be using the symbol for pragmatic deviance ('#'), given our claim that outbound anaphora involves no grammatical violation.

On the basis of such data, Postal concluded that coreferential pronouns (e.g. 2a), 'identity of sense' pronouns (e.g. 2b), and the pro-VP *do so* (e.g. 2c) cannot be anaphorically related to words that constitute 'part of the meaning' of another word in the sentence. Even if a word is morphologically present within another word, Postal claimed, it still cannot serve as an antecedent for these anaphoric elements, as illustrated in 2d–f.

Postal also argued that anaphoric elements themselves may not occur as part of the sense of a word, nor may they be morphologically incorporated into a word. Such anaphora, which he termed INBOUND ANAPHORA, is exemplified in 3:

(3) a. *The grolf wanted to visit *Max*. (grolf = 'one who has written the biography of *X*') (Postal 1969:206, ex. 11a)

b. *The boy who owned a flark made fun of Max's *gorilla*. (flark = 'a device for removing the pelt of *one*') (Postal 1969:210, ex. 25a)

c. *The fact that Max plorbed Betty did not convince Pete to *kiss* her on the lips. (plorb = '*do so* on the lips') (Postal 1969:213, ex. 39a)

d. *McCarthy* was glad that *him*ites were the majority in the room. (Postal 1969:214, ex. 50a)

e. *Harry was looking for a rack for *magazines* and he found a *one*-rack. (Postal 1969:216, ex. 60b)

f. *People who *smoke* like other *do so*ers. (Postal 1969:217, ex. 69a)

In 3a–c we see that anaphors may not occur as part of the sense of a word, while in 3d–f we see that anaphors may not be morphologically incorporated in lexical items. Thus, both simple and derived morphological forms are claimed to be anaphoric islands with respect to both outbound and inbound anaphora.

As Postal noted, some of these data seemed problematic for the theory of Generative Semantics and would appear to provide good support for the alternative theory of Interpretive Semantics then under development. Recall that in Generative Semantics it was posited that a word such as *orphan* might actually be represented syntactically by the phrase *a child whose parents have died*. It was therefore something of a puzzle that one could not refer to the deceased parents with an anaphor, as illustrated in 2a. By contrast, in Interpretive Semantics words were not decomposed into underlying syntactic representations; this theory was therefore not required to explain examples of ill-formed outbound anaphora like those in 2a–c or the absence of words with the characteristics required to yield examples like those in 3a–c.

Interestingly, Postal marshaled the anaphoric-island data as evidence FOR rather than against Generative Semantics. First, while Interpretive Semantics could explain the lack of inbound anaphora in cases like 3a–c, it could not explain the absence of forms like *himite*, *oner*, or *do soer* in 3d–f without some additional constraint. Generative Semantics, however, coupled with an anaphoric-island constraint applying late in the derivation of sentences, could give a uniform account of why ALL such cases of inbound anaphora are ill-

formed. Similarly, while Interpretive Semantics could handle cases of outbound anaphora like 2a–c, it could not without additional stipulation account for those in 2d–f. For example, given that *McCarthy* is morphologically present in *McCarthyites*, there should be no reason on an interpretive account why it could not function as an antecedent for the anaphor in 2d. Again, with an additional late anaphoric island constraint, Generative Semantics could provide a uniform account of all of the examples in 2. Given these assumptions, *McCarthyites* and *orphan* are treated alike, since both would be marked as anaphoric islands late in the derivation and both would be equally 'inaccessible' to subsequent anaphora. Finally, Postal argued that a late anaphoric-island constraint was in fact required on independent grounds. He presented evidence that relational adjectives such as *American* in *the American attempt to invade Cuba* are derived from underlying full NPs (see also Levi 1978); indeed, as this example shows, the underlying NP can evidently serve as the antecedent for the deleted subject of the embedded clause *to invade Cuba*. Yet such adjectives nonetheless constitute islands, according to Postal, who offered as evidence the examples in 4 (1969:223):

(4) a. **Her* enemies were pleased by the *American* invasion of Vietnam.
   b. **America* praised the *it*an invasion of Cuba.

Thus, Postal concluded, there must be some kind of constraint that marks simple and derived words as anaphoric islands fairly late in the derivation of sentences, at least after the application of the rule converting noun phrases into relational adjectives. Given that a late anaphoric-island constraint appeared independently necessary, Generative Semantics stood in a better position than Interpretive Semantics to account for these data; only the former could readily explain parallels between words that only underlyingly 'contained' antecedents or anaphors and words that morphologically contained antecedents or anaphors. It was thus taken to be an advantage of Generative Semantics that it is only on the surface that, say, *pork* and *wombatmeat* consist respectively of one and two morphemes; the anaphoric-island constraint treats them identically with respect to outbound anaphora.

Ross attempted to pinpoint the stage in the derivation at which the anaphoric-island constraint applies, claiming that 'it is perfectly possible for pronouns to appear in the course of a derivation which refer to NPs "inside" words, as long as these pronouns do not eventually appear in surface structures' (1971:599). For example, in 5 the ellipted VP is *justify herself*, where *herself* clearly has *Britain*, part of *British*, as its antecedent (Ross 1971:599, ex. 2):

(5) I approve of *America*'s attempt to justify *herself*, but I don't approve of the British attempt (to).

To handle such data, Ross suggested that the anaphoric-island constraint is triggered only by pronouns which are present in surface structure. The fact that the implicit reference to *Britain* in 5 is possible was taken by Ross to be further support for Generative Semantics.[2]

---

[2] It is interesting that Ross appears to have overlooked the fact that the omitted *herself* does not have *America* OR *British* as a direct antecedent, at least not in the theory of transformational syntax assumed at the time (nor, for that matter, in current Government-Binding theory). Rather,

Before we proceed further with the discussion, it is worth bearing in mind two points concerning grammatical theory at the time of the early discussions of anaphoric-island phenomena. First, most researchers in generative syntax then had little interest in morphology per se; hence, there was often no attempt to distinguish cases in which an antecedent is morphologically contained within another word from cases in which the two words are merely morphologically RELATED (though see the discussion of Browne 1974 below). Second, early studies in the generative framework viewed anaphora as a relationship—either a transformational one or one involving some sort of indexing—between two positions in a syntactic structure. The view that words were anaphoric islands therefore constituted, in effect, a syntactic constraint. While we do not deny that syntax may constrain at least one kind of anaphora, namely bound anaphora, we shall assume, as argued in Reinhart 1983, that unbound pronouns are not indexed or otherwise structurally related to their antecedents. Rather, following Karttunen 1976, Grosz 1977, Morgan 1978, Webber 1979, Sidner 1979, and Grosz & Sidner 1986, inter alia, we assume that such reference is more accurately seen as a relation between language and discourse entities, which constitute part of a speaker's (continuously updated and revised) model of the ongoing discourse.

2.2. THE GRADIENT NATURE OF OUTBOUND ANAPHORA. Subsequent work on so-called anaphoric islands revealed outbound anaphora to be a gradient phenomenon, rather than the categorical one originally described by Postal.

Tic Douloureux 1971, for example, observed that certain 'unmentionable' body substances may be felicitously referred to with an anaphor even when those substances are not explicitly evoked in the preceding discourse. Consider the examples in 6, in which no explicit antecedent for the anaphor occurs (Tic Douloureux 1971:46):

    (6) a. John *bled* so much *it* soaked through his bandage and stained his shirt. (bleed = 'to emit *blood*')
        b. When Little Johnny *threw up*, was there any pencil-eraser in *it*? (throw up = 'to emit *vomit*')

To account for such data, Tic Douloureux proposed the following 'grammatical' principle (1971:48): 'Whenever a sentence has a semantic interpretation making reference to an action or event that (inferentially) results in the production of an unmentionable bodily substance, such a substance can be referred to by a pronoun *it* within the sentence...' Significantly, this principle makes no reference to any morphological or syntactic relation between anaphor and ante-

---

the antecedent for *herself* is the deleted subject of the VP *to justify herself*. given that the verb *attempt* is an EQUI-verb, and that the related noun *attempt* is an EQUI-controlling noun; in current parlance, the subject of *attempt* controls the PRO of the embedded clause. Curiously, however, while *French* can apparently control the PRO in (i), as Postal 1969 noted in connection with similar examples, an explicit anaphor—which should permit coindexing with the subject PRO—is odd in this context, as seen in (ii):

    (i) the French attempt PRO to regain the former colonies
    (ii) ?the *French* attempt PRO to regain *her* former colonies

cedent. However, as we shall see, the inferential process alluded to in Tic Douloureux's principle extends far beyond unmentionable bodily substances.

Lakoff & Ross (1972) proposed a set of principles designed to account for some of the gradations in acceptability for outbound anaphora. First, they suggested that examples of outbound anaphora are improved if the intended antecedent is morphologically related to the surface word that contains it. Thus, 7b is correctly predicted to be more acceptable than 7a (Lakoff & Ross 1972:121):

(7) a. *The *orphan* misses *them*.
    b. ?*A *guitar*ist bought *one* yesterday.

Second, they claimed that an even greater improvement can be achieved if the derived lexical item containing the antecedent does not COMMAND the pronoun.[3] Thus 8a is worse than 8b, they claimed, because in 8a the word containing the antecedent (*guitarist*) commands the pronoun (*it*), while in 8b it does not (1972:121):

(8) a. ?*The *guitar*ist thought that *it* was a beautiful instrument.
    b. ?John became a *guitar*ist because he thought that *it* was a beautiful instrument.

On the basis of these observations, Lakoff & Ross proposed the following three degrees of deviance for outbound anaphora:

(9) a. '*' if the lexical item and the antecedent are not morphologically related;
    b. '?*' if the lexical item and the antecedent are morphologically related and if the lexical item commands the pronoun;
    c. either '?' or 'ok' if the lexical item and the antecedent are morphologically related and if the lexical item does not command the pronoun.

However, it is not the case that morphological unrelatedness necessarily results in infelicitous outbound anaphora. Consider the example in 10, where the containing word *second* is clearly not morphologically related to the intended antecedent *two*:[4]

(10) This is the *second* time in *as many* weeks.

Another problem is that Lakoff & Ross's command condition 9b would assign the second degree of deviance to the naturally-occurring examples in 11:

(11) a. The *Senator Bradley* forum has been canceled due to *his* need to be in Washington for the budget vote.
         (note on poster at AT&T Bell Labs; September 26, 1990)
     b. Last night's *Sinead O'Connor* concert at the Garden will be *her* last.
         (WNBC 6:00 News; August 25, 1990)

---

[3] Node A commands node B if neither node dominates the other and if node B is dominated by the first S node above A (Ross 1986:201).

[4] As we explain in §3.4, what is required for the felicitous outbound anaphora exemplified in 10 is the existence of a well-instantiated lexical—rather than morphological—relationship between the containing word and the intended antecedent.

c. I was reading this *Peggy Noonan* book on *her* years at the White House...

(Julia Hirschberg in conversation; November 9, 1990)

In all these examples, the lexical item containing the antecedent commands the pronoun, yet none seems particularly infelicitous.

Watt (1975) discussed a number of factors that, he claimed, serve to improve the 'penetrability' of outbound anaphora. First, he noted that such anaphora is facilitated when the antecedent bears contrastive stress, as in 12 (Watt 1975:106):

> (12) All the *Nixon*ites I know are for putting all the Agnewites in cold storage till 1976; but *HE HIMSELF* doesn't care a fig.

Here, it is claimed that the contrast between Nixon and Agnew—marked prosodically by a pitch accent on *Nixon*—'exposes' the antecedent in a way the deaccented antecedent would not. Watt argued that exposed antecedents result in reduced processing effort (1975:105):

> 'In the case of an 'impenetrable', exposure of (= penetration to) the contained anaphorical antecedent would thus be possible at the point in hearing the sentence when only the antecedent had been heard, rather than, retrospectively, when the anaphor was heard, perhaps much later. A reduction of processing effort should result, and so a gain of acceptability.'

Thus, for Watt, accent on *Nixonites* in 12 serves to expose the substring *Nixon*, rendering the NP 'available' for subsequent reference. However, as noted by Wilson & Sperber (1979), Prince (1981b, 1986), Rooth (1985), Hirschberg & Pierrehumbert (1986), and Pierrehumbert & Hirschberg (1990), among others, the function of pitch accent is not to expose linguistic strings, but rather to highlight, or focus, the discourse entities to which those strings refer. Such an analysis of accent is consistent with our view of reference as a relation between language and entities in a discourse model, rather than as a relation between linguistic objects. Furthermore, we argue that what is relevant for felicitous outbound anaphora is not accent per se, but rather the relative accessibility of the discourse entity which may be evoked AS A RESULT OF a speaker's use of accent. Nonetheless, we agree with Watt that accent is relevant to the interpretation of outbound anaphora, though it is but one of many factors that contribute to the relative accessibility of discourse entities.

Another factor contributing to felicitous outbound anaphora, according to Watt, is the degree to which the anaphor is 'specific' to the particular antecedent. To illustrate, Watt offered the examples in 13 (1975:102):

> (13) a. ??Whenever Otis meets a lifelong *New Yorker* he says he thinks *it*'s the worst city in the world.
>
> b. +Whenever Otis meets a lifelong *New Yorker* he says he wouldn't live *there* on a bet.[5]
>
> c. +Whenever Otis meets a lifelong *New Yorker* he says he would never visit *such a place*.

Here Watt claimed that, as an anaphor becomes increasingly specific (i.e. from the least specific, *it*, to the most specific, *such a place*), the corresponding

---

[5] Watt used ' + ' to mean 'the antithesis of '"*"', however interpreted' (1975:101).

islands become increasingly 'penetrable'. While we disagree with Watt about the infelicity of 13a, we nonetheless agree that in general the more descriptive the anaphor, the greater the possibility of successful reference.[6]

Watt's set of conditions under which penetration into islands is more or less possible constituted the first attempt of which we are aware to describe what would now be called pragmatic factors that affect the well-formedness of outbound anaphora. However, Watt adopted the contemporary prevailing view of anaphora as essentially a relation between linguistic elements: 'The bond joining anaphor and antecedent is sensitive to whether or not both anaphor and antecedent are present in the given sentence as 'words', but this sensitivity is very mutable' (1975:101). This contrasts with the more modern (and more accurate) view of anaphora as a relation between a linguistic anaphor and its nonlinguistic referent in the discourse model.

Corum (1973) presented additional evidence in support of a gradient, rather than categorical, constraint on outbound anaphora. She argued that, in some cases, pronouns MUST be allowed to refer to an antecedent that is contained in the semantic structure of another word. She further suggested that the gradient nature of the constraint—i.e. that anaphors can refer AT ALL to items within words—is evidence for a Generative Semantic as opposed to an Interpretive approach. Browne 1974, however, argued that Corum's idea of (semantic) containment must be weakened to 'semantically related', because an anaphor's antecedent can either contain or be contained in the surface form. As evidence, Browne provided the examples in 14 (1974:620):

(14) a. Mary knows *Kurd*ish, because she is *one*.
     b. John is a *Kurd*, and his children can speak *it*.

In 14a the antecedent of *one* (*Kurd*) is semantically and morphologically contained within the word *Kurdish*, while in 14b the intended antecedent of *it* (*Kurdish*) actually contains the surface word *Kurd*. In fact, all of Browne's examples involve surface words which are both morphologically AND semantically related to the intended antecedent (cf. Lakoff & Ross's 1972 formulation concerning MORPHOLOGICAL relationship).

We note in passing that, assuming the examples in 14 are well-formed, Browne's argument has an undesirable consequence for the Generative Semantics position. If *Kurdish* is represented as 'the language spoken by Kurds' in 14a, and if *Kurd* is represented as 'people who speak Kurdish', as 14b would seem to suggest, then a representational infinite regress results.

2.3. OUTBOUND ANAPHORA AND RECENT THEORIES OF MORPHOLOGY. While the outbound-anaphora data were originally offered as evidence for Generative Semantics, such data have also been cited in support of a number of claims about morphology. For example, Levi (1978) argued that the data supported her position that complex nominals (e.g. compound nouns) are categorially

---

[6] A better example to illustrate Watt's point in 13a is presented in (i):

   (i) Whenever Otis meets a lifelong *New Yorker* he says he thinks *it*'s dirty.

Without the predicate in Watt's example (*the worst city the world*), the *it* of (i) is difficult to interpret.

nouns rather than noun phrases. More recently, anaphoric-island data have been reinterpreted in the context of the theory of Lexical Phonology and Morphology. An important principle of lexicalist theories of morphology (e.g. Pesetsky 1979, Kiparsky 1982, and Mohanan 1986, inter alia) is the LEXICAL INTEGRITY HYPOTHESIS. Under this hypothesis, syntactic processes do not have access to the internal structure of words. Movement transformations, for instance, are prevented from moving morphemes either into or out of words. According to Pesetsky 1979 (and subsequent work, e.g. Mohanan 1986), such lexical 'integrity' is derivable from an important construct of Lexical Phonology and Morphology, namely BRACKETING ERASURE. Bracketing erasure deletes word-internal brackets at certain points in the derivation of a word (at the end of each cycle, in most versions of the theory). Crucially, word-internal brackets are also deleted at the end of a word's derivation, prior to lexical insertion. Bracketing erasure thus prohibits postlexical (e.g. syntactic) processes from having access to word-internal components; no syntactic process, for example, may make reference to the morpheme *truck* in the compound *truck driver*. Hence, such a compound would be as unanalyzable as *orphan* with respect to syntactic operations.

Under the assumption that anaphora involves a syntactic relationship between word strings, Simpson 1983 noted that the existence of anaphoric islands follows from the lexical integrity hypothesis. Because word-internal components are not visible to syntactic operations, there would be no way for an anaphor to be coindexed with a word-internal antecedent. Outbound anaphora is thus predicted to be categorically ungrammatical.[7] However, as we have seen, outbound anaphora is not, contra Simpson, a categorical phenomenon. Furthermore, while Simpson's approach makes a strong (but untenable) prediction concerning cases of sentence-internal anaphora, it is unclear what prediction it would make in a case where the anaphor is in a different sentence from its (word-internal) antecedent. Compare, for example, 15a–b:

(15) a.   #Yesterday, I met this really odd *truck* driver who lives in *it*.

b.   Yesterday, I met this really odd *truck* driver. #He lives in *it*.

Assuming that intersentential coreference is not governed by syntactic coindexation, Simpson's theory rules out 15a, while making no claim about the equally infelicitous 15b.

Sproat (1985, 1988) argued that Postal's prohibition against both inbound and outbound anaphora is derivable without appealing to the notion of lexical integrity. Instead, he suggested that the constraint could be derived from considerations concerning the kinds of antecedents that anaphors may have. He argued that previous work on anaphora within generative syntax has implicitly assumed that an antecedent for a pronoun must be a maximal projection. So it has been assumed, for example, that *him* in 16 cannot be coindexed with

---

[7] Note that this is similar to Postal's 1969 notion that the anaphoric-island constraint applies late in the derivation: in both cases, a principle applies that renders morphologically complex words indistinguishable from monomorphemic words with respect to postlexical processes (including anaphora).

just the head noun *man*, but only with the maximal projection of the head noun, i.e. the NP *the large man* (Sproat 1988:294):

(16) *The large *man* had a hat with *him*.

Sproat proposed that anaphors such as pronouns or the pro-VP *do so*—both maximal projections under his analysis—must have as antecedents phrases that are likewise maximal projections. Thus, he argued, one can derive structural constraints on outbound anaphora by appealing to the prohibition on maximal projections within words in English, as evidenced by the ungrammaticality of *a [The Bronx] hater*, where a maximal projection (*The Bronx*) occurs word-internally (Fabb 1984). Under such an analysis, *truck* in *truck driver* could not serve as the antecedent for a pronoun simply because it is not of the right syntactic form. In this way, both Sproat (1985, 1988) and Simpson (1983) argued that no anaphoric-island constraint per se is necessary, with Sproat pointing out that so-called anaphoric islands do not, contra Simpson, provide evidence for the lexical integrity hypothesis. However, both Sproat's and Simpson's approaches, like Postal's original analysis, treated anaphoric islands as a categorical phenomenon, which, as we have seen, is not supported by the data.

Like Lakoff & Ross 1972, Lieber 1984 suggested that structural configuration plays a significant role in the acceptability of outbound anaphora. Appealing to Government-Binding theory (Chomsky 1981), Lieber claimed that R-expressions (i.e. nonpronominal referring expressions) may not be bound, and hence that pronouns may not c-command their R-expression antecedents.[8] This constraint, she claimed, could account for the contrast illustrated in 17 (1984:188):

(17) a.   *McCarthy*ites are now puzzled by *him*.
     b. **He* distrusts *McCarthy*ites.

Specifically, Lieber attributed the unacceptability of 17b—where *he* c-commands the R-expression *McCarthy*—to a violation of Condition C of the binding theory, which states that an R-expression may not be bound. By appealing to binding theory, Lieber attempted not only to account for the ill-formedness of 17b, but also to argue against the lexical integrity hypothesis; since, she claimed, the syntactic principles of binding theory must have access to word-internal elements in order to rule out 17b, it follows that the lexical integrity hypothesis cannot be correct.

However, the problem with Lieber's example 17b is not that *McCarthy* is c-commanded by the subject pronoun; rather, its deviance results from the fact that there is no antecedent for the anaphor in the context provided. We would not expect *he* to specify McCarthy in this example any more than we would expect *he* to specify McCarthy in, say, *he left*. In an appropriate context, Lieber's example—slightly modified—is fine. Consider the constructed example in 18a, as well as the naturally-occurring example in 18b, from a report of an interview with Salman Rushdie:

---

[8] There are various definitions of c-command. For Lieber's—and our—purposes the following definition (taken from Radford 1988:115) will suffice: X c-commands Y iff the first branching node dominating X dominates Y, and neither X nor Y dominates the other.

(18) a. After *McCarthy* had undergone a change of heart and issued a
          public apology, *he* began to distrust the very McCarthyites who
          previously had been so fiercely loyal.
     b. *He* has called editors to tell them *Rushdie* jokes ... (*New York
          Times Magazine*, 'Rushdie in hiding'; November 3, 1990, p. 68)

The felicity of these examples argues against any attempt to provide an exclu-
sively structural account of outbound anaphora.

Finally, Sproat & Ward (1987) challenged the claim that the unacceptability
of so-called anaphoric islands involving outbound anaphora is the result of a
violation of some syntactic or morphological principle.[9] They argued that prag-
matic factors such as contrast and topicality serve to increase the salience of
a referent evoked by a word-internal element to a level where outbound anaph-
ora is felicitous. In this paper we develop some of the suggestions introduced
in this earlier work, and present the results of a series of psycholinguistic
experiments that support these suggestions.

2.4. SUMMARY. Anaphoric-island data were first offered in support of the
decompositional approach of Generative Semantics. Although Postal's original
1969 formulation of the anaphoric-island condition included a categorical pro-
hibition on reference 'into and out of' words, it was soon noted (Lakoff &
Ross 1972, Watt 1975) that the conditions on well-formed outbound anaphora
were in fact gradient. The phenomenon was subsequently recast in terms of
lexical integrity, a key principle of lexicalist morphological theory. The earlier
anaphoric-island stipulation was argued to be derivable from a more general
prohibition against syntactic access to lexical structure (implemented as brack-
eting erasure in Lexical Phonology and Morphology). Sproat (1985, 1988) ar-
gued against this approach and suggested instead that there was a syntactic
condition on the kinds of phrases which could serve as possible antecedents
for anaphors.

With few exceptions, previous approaches have assumed that outbound
anaphora is to be ruled out by some morphological or syntactic principle. In
what follows we shall suggest, as in the studies of Simpson 1983 and Sproat
1985, 1988, that there is no specific anaphoric-island restriction. However,
unlike Simpson or Sproat, we shall argue that the degree to which outbound
anaphora is felicitous is determined by the relative accessibility of the discourse
entities evoked by word-internal lexical elements, and not by any principle of
syntax or morphology.[10] While some previous studies (e.g. Tic Douloureux
1971, Watt 1975) have acknowledged the importance of pragmatic factors in
the acceptability of outbound anaphora, most others have taken the alternative
position that outbound anaphora is ungrammatical, and only occasionally ame-
liorated through contextual manipulations. In the following section we reject
this 'ungrammatical-but-salvageable' view of outbound anaphora, and present
our pragmatic analysis of the phenomenon.

---

[9] The sole exception is outbound anaphora with the pro-VP *do so*, on which see §3.3 below.

[10] Nor by any principle derivable from other morphological or syntactic principles, such as lexical
integrity.

A PRAGMATIC ANALYSIS OF OUTBOUND ANAPHORA

3. As noted in §1, we shall assume that a genuinely ungrammatical construction is ungrammatical in all (nonmetalinguistic) contexts, and cannot be 'amnestied' by pragmatic or discourse factors.[11] Given such an assumption, it would be inconsistent for a construction to be ruled out by syntactic considerations and, at the same time, be acceptable under certain discourse conditions. Rather, we would maintain that such a construction is syntactically well-formed, but restricted to certain discourse contexts for pragmatic reasons.

In our study of inbound and outbound anaphora, we will concentrate on cases where the antecedent (in the case of outbound anaphora) or anaphor (in the case of inbound anaphora) is morphologically 'contained' within a word.[12] Specifically, we propose that:

(19) A. Inbound anaphora is ruled out by a grammatical principle that prohibits pronominal elements from appearing in word-internal positions.

B. Outbound anaphora is not ruled out by any grammatical principle—with the exception of outbound anaphora involving *do so* (see §3.3).

First, we claim that inbound anaphora is ungrammatical: word-internal anaphors are categorically ruled out by independently motivated morphosyntactic principles. There are a number of ways in which this prohibition could be derived, but for the purposes of this discussion we present the simplest of these (see Sproat 1985, 1988 for a different explanation). Pronouns are closed-class items, and as such do not freely allow further morphological derivation (Paul Kiparsky, personal communication, 1990). Thus forms like **himite* or **them-hater* are ruled out by the same morphological constraint that generally prevents formations like **withing* or **overer*.

Given our assumption that ungrammatical constructions cannot be amnestied by pragmatic factors, it follows that inbound anaphora should not be possible in ANY (nonmetalinguistic) discourse context. Indeed, we know of no contexts in which such anaphora is well-formed. We thus conclude that inbound and outbound anaphora are, contra Postal 1969, distinct in that only the former is governed by morphosyntactic principles. Crucially, however, inbound anaphora is not ruled out because words are anaphoric 'islands', but rather because pronouns are categorically barred from word-internal positions.[13]

Second, we claim that there is no principle of grammar that explicitly pre-

[11] For a contrasting view, see Shibatani & Kageyama (1988), who argue for an Anaphoric Island Constraint, while conceding that violations may occur as a result of 'some kind of pragmatic inference rather than by a direct coreferential relation' (1988:473, n. 12). However, they provide no criteria to distinguish between these two possibilities. As we will argue in the following discussion, such a distinction is both unmotivated and unnecessary.

[12] Examples where no morphological containment is involved, e.g. 2a, are discussed in §3.4 below.

[13] One might also point out that some languages do allow incorporated pronouns within verbs (see, for instance, Bresnan & Mchombo 1987). As far as their anaphoric behavior is concerned, incorporated pronouns in languages that have them are exactly like nonincorporated pronouns in

vents word-internal antecedents for pronominal anaphors.[14] As initial evidence, consider the naturally-occurring data in 20, drawn from our corpus of outbound anaphora (part of which is presented in the Appendix).

(20) a. For a *SYNTAX* slot, I'd rather see someone with more extensive coursework in *it*. (Judith Levi discussing various subdisciplines of linguistics; January 18, 1987)

b. Patty is a definite *Kal Kan* cat. Every day she waits for *it*. (Television advertisement for Kal Kan; January 28, 1987)

c. There's a *Thurber* story about *his* maid ... (Michael Riley in conversation; September 7, 1988)

d. We went up to *Constable* country; we stayed in the village *he* was born in. (Kenneth Sproat in conversation; October 11, 1988)

e. I refer you to the *Schachter* paper; *he*'s very proud of it ... (Mark Baker in response to a question at NELS; November 12, 1988)

f. Well, action is still needed. If we're to finish the job, Reagan's Regiments will have to become the *Bush* Brigades. Soon *he*'ll be the chief, and he'll need you every bit as much as I did. (Ronald Reagan, farewell speech, January 11, 1989, reported in Associated Press Newswire)

g. Millions of *Oprah Winfrey* fans were thoroughly confused last week when, during *her* show, she emotionally denied and denounced a vile rumor about herself. (*Chicago Tribune*, column by Mike Royko; May 22, 1989; cited in James McCawley's '1989 linguistic flea circus' as an example of reflexive usage—not as an example of outbound anaphora)

h. I had a *paper* route once but my boss said I took too long deliverin' '*em*. ('L. A. Law'; 1987)

i. I'm a *mystery-story* buff and read (and watch on PBS) a lot of *them*. (Northwestern University electronic bulletin board; January, 1989)

---

a language like English. Again, this does not affect our argument here; it seems that English MORPHOLOGICALLY rules out any kind of pronoun 'incorporation', and it is this grammatical fact which accounts for the inbound anaphora data. If English did allow incorporated pronouns, we would expect them to behave like free pronouns with respect to their anaphoric behavior, just as they do in languages that allow them.

[14] Following previous work on anaphoric islands, we shall restrict our analysis of outbound anaphora to nonepithet anaphors. However, we note that anaphoric epithet NPs, illustrated in (i) and (ii), also participate in such anaphora:

(i) The *Philadelphia* Inquirer beseeched its readers through a series of editorials last summer to stop giving to beggars, especially drug and alcohol abusers, who the paper claimed were driving away tourists and threatening the economic survival of *the city*'s downtown. (*Chicago Tribune* article, 'Beggar's bounty: Deaf ear, cold shoulder'; May 13, 1990)

(ii) Health Secretary Louis Sullivan said Monday he was outraged that 'unAmerican' protesters prevented him from being heard at an *AIDS* conference, but the incident would not reduce his commitment to fight *the disease*. (*Chicago Tribune* article, 'AIDS protest angers health secretary'; June 26, 1990)

j. In the distance, we heard the sound of an *ambulance* siren. Within a minute or so *it* arrived and stretcher bearers took the boy away. (*New York Times Magazine*, 'The tragedy of Detroit'; July 29, 1990, p. 25)

k. Officials in the Danish capital believe they've found a way to stop *bicycle* thefts—let people use *them* for free. (Associated Press Newswire; November 10, 1990)

l. I was reading this *Peggy Noonan* book on *her* years at the White House ... (= 11c)

If one takes the position that outbound anaphora violates a principle of grammar, one will have to allow for frequent pragmatic amnestying in order to accommodate the well-formedness of data such as those in 20. In the absence of any account of the conditions under which such amnestying is possible, it is not clear how to evaluate this position. Moreover, such an account would also have to explain why cases of truly ungrammatical inbound anaphora fail to be rendered acceptable under ANY circumstances. For example, if one were to argue that 20a can be amnestied because the anaphor is interpretable by some kind of 'pragmatic inference', one would have to explain why the same sort of pragmatic inference fails to salvage the following example, where there is clearly no difficulty in interpreting the anaphor:[15]

(21) *I'll eat *oysters* on occasion, but I'm really not much of a *them* lover.

On the basis of such data, we reject the view that outbound anaphora is ungrammatical and argue instead for a pragmatic analysis of the phenomenon. From this, it follows that the many examples of ill-formed outbound anaphora discussed by Postal (1969) and others are not syntactically UNGRAMMATICAL, but rather pragmatically INFELICITOUS.

Before proceeding, we first lay out some assumptions concerning the pragmatic framework that we will be adopting. As we have noted, one of the problems with previous accounts of outbound anaphora has been the assumption that anaphora—indeed, reference in general—involves a direct relation between LINGUISTIC objects. As discussed above, Postal's original formulation of the problem in terms of anaphoric islands involved morphosyntactic restrictions on possible antecedents for anaphoric elements: 'Outbound anaphora is the relation between a [sentence] chunk, part of which is interpreted as antecedent, and some anaphor outside of that chunk' (1969:206). Watt 1975 furthermore talks of 'penetrating' a word or phrase in order to arrive at a pronoun's antecedent.

In contrast, we maintain that a more adequate account of outbound anaphora

---

[15] One might argue that, on the one hand, constructions like *them lover* violate a strong morphosyntactic constraint, whereas instances of outbound anaphora violate only weak morphosyntactic constraints and are therefore more readily amnestied by pragmatic factors. While this is a possible theory, it is not clear how one would distinguish it empirically from the pragmatic approach we present below. Furthermore, we would argue that the pragmatic factors affecting the acceptability of outbound anaphora are factors that are relevant to anaphora in general; thus, the idea that outbound anaphora is even weakly ungrammatical serves no apparent purpose.

is possible once reference is viewed as a relation that holds between language and one or more entities in a constructed representation, or model, of the ongoing discourse (see Karttunen 1976, Grosz 1977, Webber 1979, and Sidner 1979, inter alia). Under this view, pronouns and other anaphors are used to refer to discourse entities rather than to linguistic antecedents. The felicity of a particular instance of anaphora, then, is a function of the relative accessibility of the discourse entity to which the anaphor is intended to refer, as well as the type of anaphor used to refer (Watt 1975). As is well known, pronouns are the most pragmatically constrained type of anaphor in that their felicitous use requires that the hearer has (or could appropriately come to have) the referent of the pronoun 'in consciousness' at the time of the hearing or processing of the utterance (see Chafe 1976, Sidner 1979, Prince 1981a, Gundel & Hedberg 1990, inter alia). That is, felicitous use of a pronominal referring expression requires that the entity to which the pronoun is being used to refer is accessible for the hearer at the time of the utterance.

We intend to demonstrate that outbound anaphora is sensitive to the same types of pragmatic constraints as are other types of pronominal reference. Specifically, we claim that word-internal morphemes may felicitously serve as antecedents for subsequent anaphora just in case the discourse entity evoked by the antecedent in question is sufficiently accessible at the time of the utterance. In those cases where the discourse entity evoked by the word-internal antecedent is not sufficiently accessible, we predict that outbound anaphora will be infelicitous.[16]

In §3.1 we discuss some of the morphosyntactic and semantic factors that affect the accessibility of discourse entities, and thus the felicity of outbound anaphora. We show that the infelicity of at least some types of outbound anaphora is derivable from various semantic and syntactic properties of words, given certain assumptions about the effects those properties have upon discourse entities introduced by word-internal morphemes. In §3.2 we consider some of the pragmatic factors that affect the felicity of outbound anaphora, and in §3.3 we argue that the VP anaphor *do so*, unlike other anaphors, is governed by morphosyntactic principles and does not participate in outbound anaphora. In

---

[16] An examination of our corpus of naturally-occurring data reveals that antecedents in word-internal positions evoke discourse entities of one of three types: a kind (in the sense of Carlson 1977), a mass term, or a specific set of one or more individuals. By far the largest class of examples in the corpus involves reference to particular individuals that are evoked by proper-name antecedents. Curiously, DiSciullo & Williams (1987:50–51) claim that words are 'referential islands' for proper names and that proper names within words are not 'truly referential'. From this, they claim, it follows that (for example) the property of admiring Nixon is not an essential property of a Nixon admirer. Thus, they argue that (i), unlike (ii), is not a contradiction (we include DiSciullo & Williams' judgments, 1987:51):

    (i)   John is a Nixon admirer in every sense except that he does not admire Nixon.

    (ii) *John admires Nixon in every sense except that he does not admire Nixon.

If one can construe a Nixon admirer as being a person with a reliable set of traits (e.g. is clean-shaven, always wears three-piece suits, and carries an attaché case), then (i) might not be construed as a contradiction. But whether or not *Nixon* in *Nixon admirer* can be used referentially is beside the point.

all three of these sections we present psycholinguistic evidence in support of our claims. Finally, in §3.4 we discuss cases of outbound anaphora whose antecedents are not morphologically present.

3.1. MORPHOSYNTACTIC AND SEMANTIC FACTORS THAT AFFECT THE FELICITY OF OUTBOUND ANAPHORA. A key factor in determining the felicity of outbound anaphora is the semantic transparency of the word containing the antecedent of the anaphor (cf. Lieber 1984). The containing word must be sufficiently transparent for the word-internal morpheme to successfully evoke an accessible discourse entity. Consider the following examples:

(22) a. Although casual *cocaine* use is down, the number of people using *it* routinely has increased. (WCBS 11 O'clock News; December 20, 1990)

b. Patty is a definite *Kal Kan* cat. Every day she waits for *it*. (= 20b)

In 22a, *cocaine use* is a semantically transparent synthetic compound: the righthand member is a deverbal nominal and the lefthand member is readily interpretable as the internal argument of the verb *use*. Thus, *cocaine use* means 'use of cocaine'. To arrive at this interpretation, a hearer must access the meanings of both *cocaine* and *use*, and it is in part this decomposition process, we claim, that renders the discourse entity *cocaine* accessible in the context of 22a. To understand the compound *Kal Kan cat* in 22b, the hearer must figure out the intended relation between cats and the substance Kal Kan. In the course of determining this relation, the hearer must access the referent of the brand name *Kal Kan* along with the denotation of the common noun *cat*. Again, such semantic decomposition serves to render accessible the relevant discourse entity.

However, it is well known that morphologically complex words tend to acquire idiosyncratic, institutionalized meanings over the course of time (Aronoff 1976, Bauer 1983). As a result, some morphologically complex words have become semantically opaque in that they can no longer be straightforwardly interpreted on the basis of their component parts. As the following examples illustrate, semantic opacity generally inhibits outbound anaphora.

(23) a. Fritz is a *cow*boy. #He says *they* can be difficult to look after.

b. Roberta is an ordained *Luther*an minister. #She's currently studying the early years of *his* life.

c. #Ironically, Paula had a *Caesar*ean while writing a book on *his* rise to power in early Rome.

d. Dom's clothes are absolutely *elephant*ine. #Indeed you could almost lose *one* in them.

Consider first the compound *cowboy* in 23a, a word that has become institutionalized. Because of institutionalization a hearer may access the meaning of the compound directly, i.e. without morphologically decomposing it. Thus *cow*, despite its morphological presence, would not generally evoke an accessible discourse entity when *cowboy* is uttered. The examples of derivational affixation in 23b–d illustrate the same point: elements within semantically opaque or institutionalized constructions do not evoke accessible discourse entities,

and thus do not generally permit felicitous outbound anaphora. In 23b, for instance, *Lutheran* is clearly related to *Luther* morphologically, yet it is to some extent only accidental that the former means 'the branch of Protestantism adhering to the views of Martin Luther'. Of course, the distinction between transparent words and opaque or institutionalized words is gradient rather than categorical. We would therefore expect word-internal morphemes to evoke discourse entities with a greater or lesser degree of accessibility depending, inter alia, upon the relative transparency of the containing word.

While semantically transparent compounds do allow felicitous outbound anaphora, it is also true that anaphora involving antecedents within compounds is, other things being equal, more difficult to construe than anaphora involving non-word-internal antecedents. One explanation for this difference may lie in the semantic difference between modifiers and predicates. First, we assume that compounds are modifier-head constructions (see, for instance, Levi 1978). That is, in the compound *Kal Kan cat*, *Kal Kan* can be said to modify *cat* in much the same way as the adjective *hostile* modifies *aunt* in the adjective-noun sequence *hostile aunt*. Let us further assume, following Wilson & Sperber 1979, that adjectives functioning as modifiers (in prenominal position, for example) are more backgrounded, i.e. less salient, than adjectives functioning as predicates. Given these assumptions, we can account for the infelicity of many instances of outbound anaphora involving compounding with the following hypothesis: discourse entities evoked by modifiers are, ceteris paribus, less accessible than entities evoked by predicates.

In fact, this hypothesized difference between modifiers and predicates has some empirical support. In an experiment reported fully in McKoon et al. 1990, it is shown that adjectives functioning as modifiers are generally less salient than the same adjectives functioning as predicates. Consider the sentences in 24, from McKoon et al. 1990:

> (24) John doesn't like to visit his relatives very much.
>> a. His intolerable aunt is hostile.
>> b. His hostile aunt is intolerable.
> He never has a very good time.

McKoon et al. (see also Rothkopf et al. 1988) found that adjectives were more available when presented in a later memory test if they had appeared in the text as predicates (e.g. *hostile* in 24a) than if they had appeared as (prenominal) modifiers (e.g. *hostile* in 24b). This finding suggests that, other things being equal, modifiers are generally less salient than predicates. In this way, we can account for the relative infelicity of outbound anaphora involving anaphors whose antecedents are functioning as compound-internal modifiers.

3.2. PRAGMATIC FACTORS THAT AFFECT THE FELICITY OF OUTBOUND ANAPHORA. In this section we discuss some pragmatic factors that affect the accessibility of discourse entities, and hence affect the felicity of outbound anaphora. We also review a series of psycholinguistic studies that provide empirical support for our analysis.

The accessibility of discourse entities is sensitive to a number of pragmatic

factors. In particular, a discourse entity seems to be more accessible (and subsequent outbound anaphora more felicitous) when the entity stands in salient opposition to some other discourse entity (see Watt 1975). Examples of such contrast are provided in 25:

(25) a. Well, action is still needed. If we're to finish the job, Reagan's Regiments will have to become the BUSH Brigades. Soon *he*'ll be the chief, and he'll need you every bit as much as I did. (= 20f)

b. For a SYNTAX slot I'd rather see someone with more extensive coursework in *it*. (= 20a)

c. Cliff Barnes: Well, to what do I owe this pleasure?
Ms. Cryder: Actually, this is a BUSINESS call, and I'd like to get right down to *it*. ('Dallas'; 1987)

In 25a then-President Reagan is contrasting his regiments with soon-to-be inaugurated President Bush's brigades. As a result of this contrast, we claim, the discourse entity corresponding to *Bush*, being in salient opposition to the discourse entity evoked by *Reagan*, is rendered more accessible. Similarly, in 25b the speaker is contrasting syntax with other subdisciplines of linguistics, and in 25c the second interlocutor contrasts business with pleasure. As is the case with contrast in general, contrast in these examples is realized intonationally with a pitch accent on the word or morpheme that evokes the discourse entity being contrasted (cf. Watt's 1975 claim—discussed in §2.2—that accent can 'expose' a word-internal antecedent).

Related to the notion of contrast is the notion of discourse topic (Chafe 1976 and Reinhart 1981, inter alia). We have observed that topical discourse entities evoked by word-internal elements facilitate outbound anaphora more than non-topical discourse entities do. Consider the following token, from a story about violence in Detroit:

(26) In the distance, we heard the sound of an *ambulance* siren. Within a minute or so *it* arrived and stretcher bearers took the boy away. (= 20j)

Here the pronoun *it* can felicitously be used to refer to a specific ambulance, which was evoked by a word-internal morpheme in the previous sentence. One of the topics of the magazine article in question was the dramatic increase of crime-related injuries in Detroit. We maintain that, in this context, ambulances are relatively topical, and this topicality renders the example in 26 felicitous.

To investigate the effects of contrast and topicality on outbound anaphora, a series of psycholinguistic experiments was recently conducted (McKoon et al. 1990). It was hypothesized that these pragmatic factors would serve to increase the accessibility of discourse entities evoked by word-internal elements, and thus facilitate outbound anaphora. Below we present an overview of the experiments, beginning with a discussion of how accessibility was manipulated and how ease of comprehension was measured.

Accessibility was manipulated in two ways: syntactically, by varying morphosyntactic structure, and pragmatically, by varying topicality and contrast.

In the first experiment, a set of 24 texts was used, each with four versions; an example is provided in Table 1. The last sentence of each version of each text contained a pronominal anaphor. In two of the four versions, the antecedent of this anaphor appeared in a nominal compound in the penultimate sentence, and in the other two versions the antecedent appeared in a verb phrase. It was hypothesized that discourse entities evoked by compound-internal antecedents would be less accessible than entities evoked by antecedents not contained in compounds, and that this difference could be attributed to the fact that the antecedent in the NP versions appeared as a modifier within the compound (see above, §3.1). Therefore, it was predicted that comprehension of the anaphor in the final sentence would be facilitated in the VP versions relative to the NP versions. In Table 1, for example, comprehension of the pronoun *they* in the final sentence was predicted to be facilitated when its antecedent *deer* appeared as a verbal argument (*hunting deer*) relative to when it appeared as a compound-internal modifier (*deer hunting*).

Compound/Non-Topical

Sam has many interests in the outdoors. He's an avid skier, and each winter he takes about a month off from work to ski in Colorado. In the summertime, he visits his parents in Montana, where he has a chance to do some mountain climbing. Lately, he's taken up deer hunting.

And he thinks that they are really exciting to track.

Compound/Topical

Sam likes the outdoor life. Having grown up in rural Kentucky, he knows a lot about nature and is an expert at fishing and shooting. He goes on hunting trips as often as he can. He used to hunt just small game, like rabbit and quail. However, lately he's taken up deer hunting.

And he thinks that they are really exciting to track.

Verbal complement/Non-Topical

Sam has many interests in the outdoors. He's an avid skier, and each winter he takes about a month off from work to ski in Colorado. In the summertime, he visits his parents in Montana, where he has a chance to do some mountain climbing. Lately, he's taken up hunting deer.

And he thinks that they are really exciting to track.

Verbal complement/Topical

Sam likes the outdoor life. Having grown up in rural Kentucky, he knows a lot about nature and is an expert at fishing and shooting. He goes on hunting trips as often as he can. He used to hunt just small game, like rabbit and quail. However, lately he's taken up hunting deer.

And he thinks that they are really exciting to track.

TABLE 1. Examples of texts with pronominal anaphors.

In addition to varying morphosyntactic structure, McKoon et al. also varied the accessibility of the referent of the antecedent in the final sentence by ma-

nipulating the contrast between the referent and other discourse entities, as well as the relation between the referent and the overall topic of the text. The texts in which the referent of the intended antecedent was designed to be topical and/or contrastive were labeled 'topical' versions. In the topical versions of the texts in Table 1, for example, the discourse is largely about fishing and hunting, and includes mention of particular animals that have been hunted: in this context, deer are relatively topical. In the nontopical versions, the discourse is about the outdoors in general with no mention of animals, and thus deer in particular are less topical. Under our view of discourse comprehension, we predicted that the topical versions would render the referent more accessible than the nontopical versions, and that this increased accessibility would facilitate comprehension of the pronoun in the final sentence.

Measuring the difficulty of comprehension for the pronoun requires a model of the comprehension processes involved (see, for instance, van Dijk & Kintsch 1983 and McKoon & Ratcliff 1989). For the purposes of this discussion, we describe only the most minimal model, sufficient to allow interpretation of our experimental results (cf. Greene et al. 1990 and Ratcliff & McKoon 1988). The first assumption of the model is that comprehension of a pronoun begins with a process that matches the grammatical features of the pronoun (i.e., in English, gender, number, and person) against the corresponding features of all the entities that have been recently evoked in the discourse model. Discourse entities will vary in the degree to which they match the features of a pronoun, depending upon the accessibility of the entities in question as well as the extent to which the semantic features of the entities correspond to the features of the anaphor. This matching process can have one of several results. If the discourse is not well constructed, there may be no entity that matches to a sufficient degree for the pronoun to be interpreted as referring to that entity. In this situation, other kinds of processing might be initiated, perhaps involving a conscious (as opposed to an automatic) search for the referent, or else the attempt at comprehension could be abandoned altogether, leaving the pronoun without an interpretation. Another possible result of the matching process would be for several candidate entities to match to a high degree, requiring additional contextual information or further processing to decide among them. Finally, if one entity matches the pronoun better than all others, this entity can be interpreted as the intended referent, with the information about the referent being combined with information about the pronoun. All other things being equal, more accessible discourse entities will be matched to a greater degree and more quickly than less accessible ones.

This model can be applied in a straightforward way to the pronouns in the final sentences of the texts used in the experiments. We assume that the grammatical features of the pronoun in a final sentence are matched against (the features of) all of the entities in the text. The most recently evoked entities will all match to some degree; however, the texts in the experiment were constructed in such a way as to rule out, by means of feature mismatches or semantic implausibility, all referents except the intended one. It is the accessibility of this referent that will presumably determine the speed and outcome

of the matching process. The more accessible the referent, the more likely it is that there will be a successful interpretation of the pronoun, and the more quickly this outcome can be achieved.

Given such a model, the experiments reported in McKoon et al. 1990 were designed to measure whether the pronouns in the final sentences were understood as referring to the intended discourse entity, and, if they were so understood, whether the speed of understanding was affected by the relative accessibility of that referent. The texts in the experiments were presented to subjects on a CRT screen. A subject initiated each text by pressing the space bar on the keyboard. This caused the first line of the text to be displayed. When the subject finished reading this line, another press of the space bar brought up the next line of the text, and so on until the final line of the text appeared. When the subject pressed the space bar after the final line of the text, a single test word was displayed on the screen. Subjects were instructed to respond 'yes' or 'no' (by pressing keys on the keyboard) according to whether the test word had or had not appeared in the text that had just been presented. For the 24 texts exemplified in Table 1, the test word was always the (intended) antecedent of the pronoun in the final sentence (e.g. *deer*), and the correct response to this test word was 'yes'. Test words for which the correct response was 'no' were presented after the final lines of filler texts.

This procedure provided two measures, as shown in Table 2. The first measure is the reading time for the final sentence containing the pronoun, and the second is the response time for the test word. The response times for the test words can be used to decide whether the pronouns were equally well understood across the four conditions. Assuming that the successful interpretation of a pronoun leaves its referent highly accessible, decisions on the test word (which corresponds to the referent) should be relatively fast and accurate. So, if the pronouns are equally well understood in all conditions, then response times to the test word should be equally fast and accurate in all conditions, exactly as shown in the results in Table 2: there are no significant differences among the response times, and accuracy rates are all above 95%. Given equal comprehension of pronouns across conditions, any differences in reading times

TEXT VERSION	READING TIMES	RESPONSE TIMES
COMPOUND/NONTOPICAL:		
... Lately, he's taken up deer hunting.	2117ms	907ms
And he thinks that they are really exciting to track.		
COMPOUND/TOPICAL:		
... However, lately he's taken up deer hunting.	1785ms	870ms
And he thinks that they are really exciting to track.		
VERBAL COMPLEMENT/NONTOPICAL:		
... Lately, he's taken up hunting deer.	1868ms	893ms
And he thinks that they are really exciting to track.		
VERBAL COMPLEMENT/TOPICAL:		
... However, lately he's taken up hunting deer.	1738ms	886ms
And he thinks that they are really exciting to track.		

TABLE 2. Results for texts with pronominal anaphors.

for the final sentences can therefore be attributed to differences in difficulty
of comprehension. McKoon et al. predicted (a) that comprehension would be
relatively more difficult for the nontopical versions than for the topical versions,
and (b) that comprehension would be relatively more difficult for the compound
versions than for the VP versions. The data confirmed these predictions. For
antecedents in both compound and noncompound structures, reading times
were significantly slower with the nontopical versions, showing a clear prag-
matic effect of topicality and contrast on both outbound and nonoutbound
anaphora. Also, for the nontopical versions, reading times were significantly
slower when the antecedents had appeared in nominal compounds than in ver-
bal complements. However, for the topical versions, there was no significant
effect of morphosyntactic structure on reading times. (Both the main effect of
topicality and the main effect of morphosyntactic structure, as well as the
interaction of the two, were significant by analyses of variance.) Apparently,
for these versions, the accessibility of the referent was already sufficiently high
that it could not be significantly increased by having the antecedent in a verbal
complement.

These results support our pragmatic account of outbound anaphora in three
ways. First, there is a significant effect of whether the intended antecedent is
word-internal or not: in the absence of topicality, reading times were slower
for the compound versions than for the VP versions. This observation is con-
sistent with the results of the experiments described in §3.1, which showed
that adjectival modifiers are generally less accessible than predicate adjectives.
Given that compounds are also instances of modifier-head constructions, we
are in a position to provide a unified account of both sets of data. All other
things being equal, modifiers—of any grammatical category—are less acces-
sible than predicates and complements. Second, the topical versions facilitated
comprehension of the anaphor; indeed, in the topical versions there was no
significant difference in comprehension between the compound version and the
VP version, suggesting that topicality and contrast might in effect make ac-
cessibility high enough to be impervious to the effects of morphosyntactic struc-
ture. Third, both syntactic versions were affected by manipulations of topicality
and contrast, suggesting that outbound anaphora is sensitive to the same types
of pragmatic factors as anaphora in general.

Our interpretation of the results from this first experiment depends crucially
on the assumption that the lack of differences in response times to a test word
across conditions indicates a lack of differences in levels of comprehension for
the pronoun across conditions. That is, we assume that the referent of the
pronoun was correctly identified in all conditions. In several follow-up exper-
iments (also reported in McKoon et al. 1990), this assumption was tested. For
these experiments a new final sentence was written for each text, in which the
pronoun was replaced by a nominal that had not previously appeared in the
text. For example, the new final sentence for the text in Table 1 was *And he
thinks bears are really exciting to track* (cf. *And he thinks they are really
exciting to track*). With the new nominal, there is no pronominal reference to
deer in the final sentence, and therefore there should be no facilitation of re-

sponse times to *deer* when it appears as a test word. That is, response times to the text word should be facilitated when the final sentence contains the pronoun, relative to when the sentence contains a new nominal, if the referent of the pronoun was actually identified during reading. This pattern, of course, should only obtain for the original test word (e.g. *deer*). With some other test word from the text (e.g. *trips*), response times should not be affected by the substitution of a new nominal for the original pronoun. The results of these follow-up experiments fully supported these predictions, thereby justifying the assumption that the test-word response times in the original results (Table 2) do indicate that the pronouns in question were understood across conditions, and that, consequently, reading times did in fact reflect comprehension difficulty.

In this section, we have argued that outbound anaphora is a fully grammatical anaphoric process of English whose felicity—like that of all grammatical phenomena—is determined by discourse context. Outbound anaphora thus contrasts sharply with inbound anaphora, which has been shown to be categorically ungrammatical. In the next section we discuss another grammatical restriction on anaphora in English.

3.3. OUTBOUND ANAPHORA INVOLVING *DO SO* AND *DO IT*. In distinguishing between 'deep' and 'surface' anaphora, Sag & Hankamer 1984 argued that surface anaphors are 'syntactically controlled' in that they require an explicit linguistic antecedent, while deep anaphors, being 'pragmatically controlled', do not.[17] Consider, for example, the contrast in 27 between the surface anaphor *do so* and the deep anaphor *do it* (examples from Sproat & Ward 1987:331):

(27) a. A: I'm going to lift this 500 lb. barbell.
          B: With your back, do you think you should {do it, do so}?
     b. [A bends down to lift a 500 lb. barbell]
          B: With your back, do you think you should {do it, *do so}?

From these examples, we see that the explicit occurrence of a (VP) antecedent is required for felicitous use of *do so*. No such morphosyntactic restriction applies to the deep anaphor *do it*; indeed, there need be no explicit antecedent at all.

Sproat & Ward 1987 noted that, contra Postal, reference to an action evoked by a verb contained within a nominal is felicitous with the anaphor *do it*, but not with *do so*.[18] Consider first the following examples of felicitous *do it* anaphora:

(28) a. Mary is a heavy *smoker*—even though her doctor keeps telling her not to *do it*.

---

[17] The terms 'deep' and 'surface' anaphora—first introduced in Hankamer & Sag 1976—are replaced in Sag & Hankamer 1984 by the (more accurate) terms 'model-interpretive anaphora' and 'ellipsis', respectively. However, the original terms are still the ones generally used in the literature to describe the distinction between the two types of anaphoric processes, even by Sag & Hankamer themselves in 1984.

[18] We assume, following Webber 1979, that verb phrases that denote actions or events can evoke discourse entities.

    b. In response to his wife's strenuous objections, Bill isn't much of
        a *sportscar race*r any more, but he still manages to *do it* every
        once in a while.

The surface anaphor *do so*, which requires an explicit VP antecedent, does not
pattern in the same way: the examples of *do so* anaphora in 29, corresponding
to the examples of *do it* in 28 above, are much worse:

    (29) a. *Mary is a heavy *smoker*—even though her doctor keeps telling
          her not to *do so*.
        b. *In response to his wife's strenuous objections, Bill isn't much of
          a *sportscar race*r any more, but he still manages to *do so* every
          once in a while.

Note that the corresponding examples of *do so* anaphora with full-VP ante-
cedents are fully acceptable, as illustrated in 30:

    (30) a. Mary *smokes* heavily—even though her doctor keeps telling her
          not to *do so*.
        b. In response to his wife's strenuous objections, Bill doesn't *race*
          sportscars very much any more, but he still manages to *do so*
          every once in a while.

Unlike other anaphors, then, *do so* is highly constrained in terms of the mor-
phosyntactic form of possible antecedents (Hankamer & Sag 1976, Sag & Han-
kamer 1984). Assuming that this constraint is a grammatical one, and given
our working assumption that truly ungrammatical violations cannot be salvaged
by pragmatic factors, it follows that no discourse context will render *do so*
anaphora felicitous with non-VP antecedents. The examples in 29 illustrate the
categorical unacceptability of such anaphora.

    This distinction between surface and deep anaphora makes a number of em-
pirically testable predictions. If we assume, following Sag & Hankamer 1984,
that deep VP anaphors such as *do it* are understood with respect to a discourse
model, then their interpretation should be sensitive to pragmatic factors, pre-
sumably the same kinds of pragmatic factors to which pronomial outbound
anaphora was found to be sensitive. Furthermore, under this assumption deep
VP anaphors should be sensitive to morphosyntactic factors only to the extent
that these factors indirectly affect the accessibility of the referent event in the
discourse model.[19] By contrast, a surface VP anaphor such as *do so*, being
sensitive to the linguistic representation of its antecedent, should be more sen-
sitive to morphosyntactic factors than to pragmatic ones.

    These hypotheses were also tested in the series of psycholinguistic experi-
ments described above (McKoon et al. 1990). The same experimental design
used to investigate pronominal anaphora was used to investigate surface versus
deep anaphora, first with the deep anaphor *do it* used in place of the pronominal
anaphor (see Table 3). The accessibility of the referent event for the VP anaphor

---

[19] Murphy 1985 and Tanenhaus & Carlson 1990 have shown that syntactic parallelism between
a deep VP anaphor and its antecedent does appear to affect comprehension difficulty for the an-
aphor. However, with the materials used in their experiments, parallelism probably affected the
discourse-level representation of the antecedent event.

Deep anaphor (do it)

Nominalized antecedent/Non-Topical

Joe does not have a very good sense of reality.
Last year, he told everybody that he was going
to go to law school. He isn't. In fact, he'll
soon be dropping out of college. Next, he said
he was dating a Vogue model. He wasn't. Now, he
claims to be a good basketball player.

But, in fact, he's never done it.

Nominalized antecedent/Topical

Joe is generally considered to be the best
athlete Central High School has ever had. He
swims; he's the star pitcher of the baseball
team; and he is a defensive end on the varsity
football team. And since he's 6'6", people
naturally assume that he's a basketball player.

But, in fact, he's never done it.

VP antecedent/Non-Topical

Joe does not have a very good sense of reality.
Last year, he told everybody that he was going
to go to law school. He isn't. In fact, he'll
soon be dropping out of college. Next, he said
he was dating a Vogue model. He wasn't. Now, he
claims to play basketball well.

But, in fact, he's never done it.

VP antecedent/Topical

Joe is generally considered to be the best
athlete Central High School has ever had. He
swims; he's the star pitcher of the baseball
team; and he is a defensive end on the varsity
football team. And since he's 6'6", people
naturally assume that he plays basketball.

But, in fact, he's never done it.

Surface anaphor (do so)

Nominalized antecedent/Non-Topical

Joe does not have a very good sense of reality.
Last year, he told everybody that he was going
to go to law school. He isn't. In fact, he'll
soon be dropping out of college. Next, he said
he was dating a Vogue model. He wasn't. Now, he
claims to be a good basketball player.

But, in fact, he's never done so.

Nominalized antecedent/Topical

Joe is generally considered to be the best
athlete Central High School has ever had. He
swims; he's the star pitcher of the baseball
team; and he is a defensive end on the varsity
football team. And since he's 6'6", people
naturally assume that he's a basketball player.

But, in fact, he's never done so.

VP antecedent/Non-Topical

Joe does not have a very good sense of reality.
Last year, he told everybody that he was going
to go to law school. He isn't. In fact, he'll
soon be dropping out of college. Next, he said
he was dating a Vogue model. He wasn't. Now, he
claims to play basketball well.

But, in fact, he's never done so.

VP antecedent/Topical

Joe is generally considered to be the best
athlete Central High School has ever had. He
swims; he's the star pitcher of the baseball
team; and he is a defensive end on the varsity
football team. And since he's 6'6", people
naturally assume that he plays basketball.

But, in fact, he's never done so.

TABLE 3. Examples of texts with VP anaphors.

was manipulated in the same way that accessibility was manipulated for the
referent of the pronoun. Topicality was varied by manipulating either the con-
trast between the referent event and other discourse events or the relation
between the referent event and the overall topic of the text. As in the other
experiments, the topical contexts were predicted to make the referent event
more accessible than the nontopical contexts, thus facilitating comprehension
of the deep anaphor in the final sentence. Morphosyntactic structure was varied
as before, with the antecedent occurring either within a nominalization or as
a verb phrase. There was no reason to believe that these two structures differed
with respect to the accessibility they contributed to the relevant event in the
discourse model; therefore, the two structures were predicted not to differ-
entially affect comprehension of a deep anaphor specifying that event.

Both predictions for *do it* were supported by the data. In the topical versions
reading times for the final sentences averaged 1504 ms, while in the nontopical

versions they averaged 1552 ms (a significant difference by analyses of variance that did not interact with morphosyntactic structure); this shows the predicted effect of pragmatic factors on deep anaphora. However, the morphosyntactic structure of the antecedent did not significantly affect reading times of the final sentences (1532 ms for the nominalization vs. 1524 ms for the verb phrase). Apparently, the two structures did not differentially affect the accessibility of the referent event.

McKoon et al. 1990 established that there were no significant differences in comprehension of the anaphors across experimental conditions in the same way as in the experiment described in §3.2, using test words taken from the antecedent of the VP anaphor (e.g. *basketball* in Table 3). As expected, response times for these test words did not differ significantly across conditions.

Next, the surface anaphor *do so* was tested by replacing the *do it* anaphors in the previous experiment with *do so* anaphors. If it is true that surface anaphors are understood with direct reference to a linguistic representation, and only indirectly with reference to events in the discourse model, then replacing *do it* with *do so* should alter the effects of the pragmatic and morphosyntactic variables that were obtained in the earlier experiment. Whereas comprehension of the *do it* anaphor was affected by the topicality of the referent event in the discourse model more than by the morphosyntactic form of its antecedent, comprehension of *do so* should be affected more by linguistic form than by topicality. Again, the results were as predicted: when the antecedent for *do so* was contained within a nominalization, reading times for the final sentences averaged 1740 ms; when the antecedent for *do so* was a verb phrase, reading times averaged 1601 ms, demonstrating a significant effect of morphosyntactic structure (by analyses of variance) that did not interact with topicality. Reading times in the topical versus nontopical versions did not differ significantly (1686 ms vs. 1654 ms, respectively), indicating that topicality had no effect on comprehension of the surface anaphor. Overall, reading times for the *do so* sentences were slower than for the *do it* sentences, but the absence of significant differences in response times to test words selected from the antecedent indicated that there were no significant differences in comprehension of the anaphors across experimental conditions.

This psycholinguistic evidence supports our claim that outbound anaphora involving the pro-VP *do so* is not rendered more felicitous by the same pragmatic factors that facilitate other types of outbound anaphora. This result is predicted from the existence of a GRAMMATICAL restriction on the antecedent of *do so* and further supports our general contention that true morphosyntactic violations cannot be amnestied by pragmatic factors.

3.4. OUTBOUND ANAPHORA WITHOUT MORPHOLOGICAL CONTAINMENT. Up to now, we have dealt primarily with outbound anaphora involving antecedents that are morphologically contained within words. In this final section we would like to consider cases in which the antecedent of the pronominal anaphor is not morphologically contained in, or in some cases even morphologically related to, the words that introduce them. Consider the examples in 31:

(31) a. 'I heard someone say,' he began, 'that you are a *New Zealand*er.
I was out *there* as a small boy.' (Ngaio Marsh, *Night at the Vulcan* (1951:207). New York: Jove)
   b. Jean is a *Frenchman*, though he hasn't lived *there* for many years.
   c. This is the *fourteen*th time in *as many* weeks.
   d. This is the *second* time in *as many* weeks. (= 10)
   e. Mary is a *physicist*: she says *it*'s an exciting field.
   f. Bill is a *linguist*: he says *it*'s an exciting field.

These data suggest that in some cases of outbound anaphora the morphological relationship between the word containing the antecedent and the antecedent itself need not be regular or even apparent. So, while *New Zealand* is clearly morphologically contained within *New Zealander*, the same cannot be said of the pair *France* and *Frenchman*. And while *fourteenth* is presumably derived from *fourteen* by suffixation of *-th*, there is no morphophonological relationship between the forms *two* and *second*. Finally, although *physicist* may be morphologically derived from *physics*, the relationship between *linguistics* and *linguist*, from a surface morphological point of view, appears to go in the opposite direction.

What the examples in 31 have in common is the fact that the link between the containing word and the intended antecedent is in each case an example of a well-instantiated LEXICAL relationship. Specifically, the pairs *New Zealand/New Zealander* and *France/Frenchman* are examples of the relationship between names of countries and names for inhabitants of those countries. This relationship is well instantiated in that it is quite generally the case that there is a term of provenance—usually unique within a given register—associated with each country name. Although there are subregularities, this relationship is by no means generally expressed in a morphologically regular fashion, as seen in 32:

(32) COUNTRY	PROVENANCE TERM
France	Frenchman
New Zealand	New Zealander
Canada	Canadian
Brazil	Brazilian
America	American
Spain	Spaniard
Thailand	Thai
Denmark	Dane

However, the SEMANTIC relationship expressed by these examples is entirely regular and predictable: all of the nouns in the righthand column refer to a person living in or originating from the corresponding country in the lefthand column. Similarly, the pairs *fourteen/fourteenth* and *two/second* (31c–d) are particular instances of the well-instantiated—indeed, completely productive— relationship between a cardinal number and its associated ordinal. Again, the morphology is irregular for some of the more common cases (*first, second, third, fifth, twelfth*), but the semantics is entirely regular. And finally, *physics/*

physicist and *linguistics/linguist* (31e–f) are examples of the relationship be-
tween a field and a practitioner in that field.

To account for such cases, we would like to suggest that outbound anaphora
is sensitive to the productivity (and semantic predictability) of the relationship
between an anaphor's antecedent and the lexical item containing that ante-
cedent. That is, *Frenchman* can evoke France in 31b precisely because the
relationship between Frenchmen and the country France is sufficiently trans-
parent due to the well-instantiated relationship of which the pair *France/
Frenchman* is an instance.[20] Similarly, *second* can evoke the number two in
31c because of the well-instantiated and semantically transparent relationship
between cardinal numbers and their associated ordinals.[21] Felicitous outbound
anaphora, then, does not appear to require a morphological relationship in the
strictest sense; a sufficiently clear and well-instantiated lexical relationship will
suffice.

The lexical relationships exemplified in 31 are reminiscent of traditional in-
flectional paradigms (see, for instance, Matthews 1974:156). In both cases,
there is a sense in which a word fills a particular 'slot' in a paradigm that
expresses some relationship between word forms.[22] In the case of the English
past-tense paradigms, for example, *compiled* fills the past-tense slot of *compile*.
Irregular forms are full-fledged members of the paradigm; the suppletive form
went is as much the past-tense form of *go* as *compiled* is of *compile*. In a similar
vein, *Frenchman* could be said to fill the provenance slot in a paradigm relating
it to the place term *France* (as in the set of paradigms in 32 above); despite
the irregular morphology, it is no less a provenance term than the regular form
New Zealander. Although the notion of paradigm has traditionally been used
in the description of inflectional morphology, there is no a priori reason for
that restriction; the lexical relationship expressed in the examples in 32 is quite
similar to the relationship among inflectional verb forms.

However, not all instances of outbound anaphora are best analyzed in terms

---

[20] A similar well-instantiated relationship seems to hold between a place and the language spoken
there. Consider the naturally-occurring token in (i):

   (i) I had *French* for eight years and I've never been *there*. (Prospective apartment renter in
       conversation; April 12, 1987)

[21] Watt 1975 also discusses the possibility of felicitous outbound anaphora with cases like *two*
and *second*, but offers a very different analysis (see §2.2 above). Watt argues that, when a 'hidden
antecedent [is] so circumscribed as perforce to be one particular word', then outbound anaphora
is possible. In the case of 'the second blow for freedom in as many weeks' (1975:111). Watt argues
that the 'number anaphor' *as many* requires a numerical antecedent, and since the 'set of possible
antecedents is so circumscribed that it has only the one member, "two" itself (1975:112)', outbound
anaphora is possible, indeed 'forced'. However, the explanation appears not to be that *as many*
forces any particular antecedent (although it certainly does do that), but that *second* is so trans-
parent. In our terms, *second* is transparently related to the cardinal number two, and therefore its
use will serve to render it sufficiently accessible for subsequent reference.

[22] While paradigm slots are usually filled by a unique word form, some slots are occasionally
filled by more than one form, e.g. the English plural forms *cacti* and *cactuses*. However, as Aronoff
(1976) and others have noted, there is a strong tendency for the existence of a filled slot to 'block'
additional forms.

of the paradigmatic lexical relationships exemplified in 31. In 22b, for example, it seems specious to analyze *Kal Kan cat* as filling a slot in a paradigm that relates things to cats liking those things; there is no well-instantiated 'cat___which___likes___x' paradigm in the English lexicon. Instead, as is the case with most of the examples discussed in this paper, outbound anaphora is felicitous in 22b because the discourse entity *Kal Kan* is sufficiently accessible to permit subsequent anaphoric reference to it, due in part to the morphological presence of the brand name *Kal Kan*. Thus, we suggest that there are in fact two sources for the contained antecedent in examples like 31a: one is the paradigmatic relationship that the containing-word/contained-word pair instantiates, and the other is the actual morphological presence of the contained word; *New Zealander* both morphologically contains *New Zealand* and is paradigmatically related to it qua provenance term.

Given this analysis, outbound anaphora is predicted to be generally infelicitous when there exists neither a morphological relationship between an anaphor's antecedent and the lexical item containing that antecedent, nor a paradigmatic lexical relationship of the kind exemplified in 31. This prediction appears to be borne out by the data. Consider again Postal's classic orphan example in 33.

(33)   #Max is an *orphan* and he deeply misses *them*. (cf. 2a)

First, it is clear that *orphan* and *parents* are not morphologically related. Second, although the words *orphan* and *parent* might be formally related, given certain assumptions about the lexicon, it is clear that they do not form part of a well-instantiated lexical relationship. So, while one can find (or construct) an appropriate provenance term for a given country or city term, there is no general pattern such that for some term *x*, there is a word meaning 'person whose *x* has died'; only a few such pairs exist in English, namely *orphan/ parent*, *widow/husband*, and *widower/wife*. It is this lack of morphological or paradigmatic lexical relationship, we claim, that renders 33 infelicitous.

However, in a more suitable context, even anaphora paralleling that in 33 is possible:

(34)   'That depends on whose mother she is,' Fitz told him. 'Mine has brown hair—hardly a bit of grey in it. Your mother's hair probably turned white in a night long ago.'
'I haven't got a mother,' said Johnny pathetically, staring at his ham sandwich. 'I'm an orphan.'
'Why, that's terrible, Johnny, when did it happen? You never told me you were an *orphan*.' Fitz was deeply concerned.
'I'm getting sort of used to it. *They* died when I was three.' (Elswyth Thane, *Ever after* (1945:155). New York: Hawthorn Books; noted by Beth Levin)

Here, in a context in which the existence of one's parents is under discussion (without an explicit mention of *parents*), subsequent pronominal anaphora is possible despite the absence of any morphological or lexical relationship.

CONCLUSION

4. Previous accounts of outbound anaphora have attempted to rule it out by means of various morphological and syntactic principles. Instead, we have argued that outbound anaphora is fully grammatical and that, like anaphora in general, its felicity is a function of the accessibility of the discourse entity to which the anaphor in question is used to refer. We have identified a number of morphosyntactic, semantic, and pragmatic factors that increase the accessibility of discourse entities—and therefore the felicity of outbound anaphora. Our analysis is supported by a series of psycholinguistic studies which show that topicality and contrast facilitate comprehension of word-internal anaphors.

APPENDIX

Below are some naturally-occurring tokens of outbound anaphora classified according to the type of discourse entity evoked. Specifically, we have classified the examples according to whether the word-internal antecedent:

- is a proper name or common noun which evokes a specific referent in the discourse corresponding to that name or noun;
- is a common noun and evokes an individual corresponding to a kind in the discourse;
- is a common noun and evokes an individual corresponding to a mass in the discourse.

On the italicization conventions for indicating coreference, see note 1.

1. SPECIFIC REFERENTS:

1. RS: Well, she got an *LSA* paper out of it.
   JH: Yes, she was *there*.
   (Julia Hirschberg and Richard Sproat in conversation; January 30, 1987)

2. A: It has something to do with *Suez* prices.
   B: Did it mean anything to you?
   A: I dunno. His father was a general *there*.
   ('Still Crazy Like a Fox'; April 5, 1987)

3. I had *French* for eight years and I've never been *there*.
   (Prospective apartment renter in conversation; April 12, 1987)

4. GW: Excuse me, sir, but what's the *tray* situation?
   CW: I'll bring *them* right out.
   (Gregory Ward and cafeteria worker, Ida Noyes Hall, University of Chicago; April 23, 1987)

5. A 508-page manuscript of nine *Mozart* symphonies written in *his* own hand in Salzburg in the 1770's, before the composer's 20th birthday, was auctioned yesterday by Sotheby's in London for $4.34 million. (*New York Times* article, 'Record Price for Mozart Manuscript'; May 23, 1987)

6. There's a *balance sheet* concern—we've never had to read *it* before. (Arno Penzias; September 29, 1987)

7. Thanks for the *Philly* dirt—I have never been *there* but if I ever do [sic] I'll let you know. (Message on electronic bulletin board; 1988)

8. Our *postscript printer* room had some water problems (under the floor) this weekend. *They* should be back up by 10:00am [sic] ... (Don Bock in email; 1987)

9. There's a *Thurber* story about *his* maid ... (Michael Riley in conversation; September 7, 1988)

10. I didn't know you had a *Joan Miller*-fan. Was this *her* office? (Michael Riley in conversation; September 12, 1988)

11. We went up to *Constable* country: we stayed in the village *he* was born in. (Kenneth Sproat in conversation: October 11, 1988)

12. You couldn't find a stronger *Dukakis*-supporter. The only way I wouldn't vote for *him* ... (Michael Riley in conversation: October 18, 1988)

13. ... that Mario Biaggi could not survive a long *jail*-sentence: that he would die *there*. (WINS: November, 1988)

14. RS: You don't know *Chinese*. I assume?
PCS: I've been *there*, but I don't speak it.
(Richard Sproat in conversation with prospective MIT Coop student)

15. *Bush* supporters would stay home, figuring *he*'d already won. (Julia Hirschberg in conversation: November 9, 1988)

16. I refer you to the *Schachter* paper: *he*'s very proud of it ... (Mark Baker in response to a question at NELS: November 12, 1988)

17. A cheer went up at *Mulroney* headquarters in *his* hometown of Baie-Comeau, Quebec, when the CBC made its first projection. (Associated Press Newswire: November 21, 1988)

18. I was an *IRS*-agent for about 24 years ... I stopped working for *them* ... (Radio ad for AARP heard December 31, 1988)

19. Well, action is still needed. If we're to finish the job, Reagan's Regiments will have to become the *Bush* Brigades. Soon *he*'ll be the chief, and he'll need you every bit as much as I did. (Ronald Reagan, farewell speech, January 11, 1989; reported in Associated Press Newswire)

20. *Museum* visitors can see through *its* big windows the 900-year-old Tower of London and the modern office blocks of the City financial district. (Associated Press Newswire: July 5, 1989)

21. 'Sometime [sic] they say, "Get away from me, I don't want to hear that *Jesus* stuff,"' he said. 'But I think deeply of *him*. He's always with me and I want other people to know he can be with them, too.' (Associated Press Newswire: August 29, 1989)

22. *Rolling Stones* fans: clear your calendars! *They*'re adding more concert dates. (WCBS 11 O'clock News: September 26, 1989)

23. Spokesmen for the federal prosecutor's office in Karlsruhe said they viewed the letter as an authentic claim of responsibility from the Red Army Faction, which had been dormant for three years until the *Herrhausen* assassination. *His* armored Mercedes was blown up by a remote-control bomb in Bad Homburg, where he lived, as he was being driven to work Nov. 30. (Associated Press Newswire: December 5, 1989)

24. Millions of *Oprah Winfrey* fans were thoroughly confused last week when, during *her* show, she emotionally denied and denounced a vile rumor about herself. (*Chicago Tribune*, column by Mike Royko: May 22, 1989; cited in James McCawley's '1989 linguistic flea circus,' as an example of reflexive usage—not as an example of outbound anaphora)

25. The *Paris* idea holds a lot of charm, 'cuz I used to live *there*, y'know. (Greg McKenna in conversation: March 14, 1990)

26. Do *parent*al reactions affect *their* children? (from Jill Burstein, uttered by one of her students: March 15, 1990)

27. 'My daughter knows the *German* people. She's been *there* [...]' (*Chicago Tribune* article, 'Holocaust revisionism: A family's strident salute': May 8, 1990)

28. 'I heard someone say,' he began, 'that you are a *New Zealand*er. I was out *there* as a small boy.' (Ngaio Marsh, *Night at the Vulcan*. 1951:207. New York: Jove)

29. JL: So, what's your *child* situation?
DS: *He*'s 4.
(Judith Levi and Deborah Schiffrin in conversation: May 22, 1990)

30. You know, this is a *Pilgrim* town here: *they* came into this harbor... (ABC 'Nightline': July 4, 1990)

31. Nancy: The whole thing—it was like suddenly being caught in a *Diane Arbus* picture. Do you ever get that feeling?
    Elliot: Hourly.
    Nancy: Me too.
    Elliot: Yeah, we never shoulda had *her* take our wedding picture.
    ('thirtysomething': July 17, 1990)

32. Another *Nixon* Summit, at *His* Library. (Title of article in *New York Times*: July 20, 1990)

33. In the distance, we heard the sound of an *ambulance* siren. Within a minute or so *it* arrived and stretcher bearers took the boy away. (*New York Times Magazine*, 'The tragedy of Detroit': July 29, 1990, p. 25)

34. The *Senator Bradley* forum has been canceled due to *his* need to be in Washington for the budget vote. (Note on poster at AT&T Bell Labs: September 26, 1990)

35. I was reading this *Peggy Noonan* book on *her* years at the White House ... (Julia Hirschberg in conversation: November 9, 1990)

36. There's no reason to become a *California* citizen, unless I'm gonna live *there*. (Ken Baime to Gregory Ward in conversation: August 8, 1990)

37. I used a *gutter* person before, but just to clean *them*. (Julia Hirschberg in conversation: October 13, 1990)

38. *Saudi* anti-aircraft guns fired on *Iraqi* planes along *their* common border. (NBC Nightly News: August 11, 1990)

39. Last night's *Sinead O'Connor* concert at the Garden will be *her* last. (WNBC 6:00 News: August 25, 1990)

40. I think if I were a *Peru*vian I wouldn't want to live *there* for the next couple of years. (John Kingston in conversation: September 6, 1990)

41. Heisenberg had bitter words to say about the lack of funds and materials, and the drafting of scientific men into the services. Excerpts from *American* technical journals suggested that plenty of technical and financial resources were available *there* for nuclear research. (Albert Speer, *Inside the Third Reich*, translated by Richard Winston and Clara Winston (1970:225–26). New York: Collier)

42. AMA: Cut *AIDS* 'protection'
    Doctors want *it* handled like other sexual diseases (Title of article in *Chicago Tribune*: December 6, 1990)

43. 'That depends on whose mother she is.' Fitz told him. 'Mine has brown hair—hardly a bit of grey in it. Your mother's hair probably turned white in a night long ago.'
    'I haven't got a mother,' said Johnny pathetically, staring at his ham sandwich. 'I'm an orphan.'
    'Why, that's terrible, Johnny, when did it happen? You never told me you were an *orphan*.' Fitz was deeply concerned.
    'I'm getting sort of used to it. *They* died when I was three.' (Elswyth Thane, *Ever After* (1945:155). New York: Hawthorn Books)

44. Our neighbors, who are sort of *New York City*-ites, they have jobs *there* ... (Ginny Beutnagel in conversation: December 30, 1990)

2. KINDS:

1. GW: So, Roger, do they even HAVE venison in New Zealand?
   RR: Oh yes. They have a real *deer* problem. *They*'ve been running around eating all the forests.
   (Gregory Ward and Roger Ratcliff in conversation: 1987)

2. I had a *paper* route once but my boss said I took too long deliverin' '*em*. ('L.A. Law': 1987)

3. ... play the *Cutlass-Supreme* Game and win *one* ... (Radio ad heard on WINS; May 20,
    1988)

4. 47th-St. Photo announces its *microwave oven* sale, just when you need *it* for your apart-
    ment ... (Radio ad heard on WINS; November 11, 1988)

5. I'm a *mystery-story* buff and read (and watch on PBS) a lot of *them*. (Northwestern Uni-
    versity electronic bulletin board; January, 1989)

6. ... the only way to solve this *homeless* problem, say those who work with *them* ... (WCBS
    11 O'clock News; January 4, 1989)

7. We asked *Saab 9000-CD* owners about *its* road-handling ... (Television ad for Saab; March
    12, 1989)

8. If you're a *small business* owner, or interested in starting *one* ... (TV ad; June 14, 1989)

9. Game show host: So, I hear you're a real *cat*-lover. *How many* do you have now? ('Jeop-
    ardy'; July 24, 1989)

10. Euripides—described by Sophocles as a *woman*-hater in his tragedies, but very fond of
    them in bed—complained that they were always having other women 'coming into the
    house gossiping.' ... (Reay Tannahill, *Sex in history* (1982:95). New York: Stein & Day)

11. CHECK *VISA* REQUIREMENTS FOR INTERNATIONAL TRAVEL
    Though many popular destinations don't require *them*—including Canada, Mexico, Eng-
    land, much of the Caribbean and Europe, Japan, Thailand, and Hong Kong—the ma-
    jority of countries still do. (Column in *Money Magazine*; January 1990, p. 144)

12. Officials in the Danish capital believe they've found a way to stop *bicycle* thefts—let
    people use *them* for free. (Associated Press Newswire; November 10, 1990)

3. MASS TERMS:

1. For a *syntax* slot, I'd rather see someone with more extensive coursework in *it*. (Judith
    Levi discussing various subdisciplines of linguistics; January 18, 1987)

2. Patty is a definite *Kal Kan* cat. Every day she waits for *it*. (TV ad for Kal Kan; January
    28, 1987)

3. There does not seem to me to be a serious *snow* problem. There is *some*, but no large
    accumulation is forecast. (Mark Liberman in email; 1987)

4. It's awfully *foggy* tonight so you people out there driving better watch out for *it*. (Heard
    on Chicago radio station; April 16, 1987)

5. Cliff Barnes: Well, to what do I owe this pleasure?
    Ms. Cryder: Actually, this is a *business* call, and I'd like to get right down to *it*. ('Dallas';
    1987)

6. Chang Ching-hui uttered a third saying when the Japanese were making so many com-
    pulsory *grain* purchases that the peasants of the Northeast had *none* left. (Aisin-Gioro
    Pu Yi, *From emperor to citizen*, translated by W. J. F. Jenner (1987:287). Oxford: Oxford
    University Press)

7. MR: How did you become an *AI* person?
    JH: I got a degree in *it*.
    (Michael Riley and Julia Hirschberg in conversation; October 4, 1988)

8. I know, you probably get eight gazillion jokes from *pragmatics* students each semester
    you teach *it*, but maybe this one you haven't seen: (Ellen Prince in email; October 5,
    1988)

9. I don't rejoice in the *stock* market going down so much, now that we started owning *some*.
    (Dan Hirschberg in conversation; November 11, 1988)

10. At the same time as coffee beans were introduced, the Arabs made changes in *coffee*
    preparation that greatly improved *its* flavor. (J. Schapira, D. Schapira and K. Schapira,
    The book of coffee and tea (1982:7). New York: St. Martin's Press)

11. Jo Ann Smith is a *beef* person. She grew up on *it* and remains a great fan of the standing rib roast. (Associated Press Newswire: July 6, 1989.)

12. 'Anyhow,' he said, 'it is nearly *Luncheon* Time.' So he went home for *it*. (A. A. Milne, *Winnie-the-Pooh* (1926:41). London: The Reprint Society)

13. 'It must be getting on for *luncheon* time,' he remarked to the Otter. 'Better stop and have *it* with us ...' (Kenneth Grahame, *The wind in the willows* (1908:92). London: The Reprint Society)

14. They're afraid it's the *Gas and Electric* man come to turn *that* off. (Interview with Baltimore politician, ABC 'Nightline': March 23, 1990)

15. Very well. But I warn you that if you continue in such foolishness you'll be the last *paleontologist* alive by the time you retire. There's no future in *it*. (Stephen Jay Gould, 'In touch with Walcott', *Natural History*, July 1990:16)

16. Although casual *cocaine* use is down, the number of people using *it* routinely has increased. (WCBS 11 O'clock News: December 20, 1990)

REFERENCES

ARONOFF, MARK. 1976. Word formation in generative grammar. Cambridge, MA: MIT Press.

BAUER, LAURIE. 1983. English word-formation. Cambridge: Cambridge University Press.

BRESNAN, JOAN, and SAM MCHOMBO. 1987. Topic, pronoun and agreement in Chicheŵa. Lg. 63.741–82.

BROWNE, WAYLES. 1974. On the topology of anaphoric peninsulas. Linguistic Inquiry 5.612–20.

CARLSON, GREG N. 1977. Reference to kinds in English. Amherst, MA: University of Massachusetts dissertation.

CHAFE, WALLACE. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. Subject and topic, ed. by Charles Li, 25–55. New York: Academic Press.

CHOMSKY, NOAM. 1981. Lectures on government and binding. Dordrecht: Foris.

CORUM, CLAUDIA. 1973. Anaphoric peninsulas. Chicago Linguistic Society 9.89–97.

DISCIULLO, ANNA MARIA, and EDWIN WILLIAMS. 1987. On the definition of word. Cambridge, MA: MIT Press.

FABB, NIGEL. 1984. Syntactic affixation. Cambridge, MA: MIT dissertation.

GREENE, STEVEN; GAIL MCKOON; and ROGER RATCLIFF. 1990. Discourse models and anaphoric reference. Evanston, IL: Northwestern University, MS.

GROSZ, BARBARA. 1977. The representation and use of focus in dialogue understanding. Berkeley, CA: University of California dissertation.

——, and CANDACE SIDNER. 1986. Attention, intentions and the structure of discourse. Computational Linguistics 12.175–204.

GUNDEL, JEANETTE, and NANCY HEDBERG. 1990. Givenness, implicature, and the form of referring expressions in discourse. Berkeley Linguistics Society 16. 442–53.

HANKAMER, JORGE, and IVAN A. SAG. 1976. Deep and surface anaphora. Linguistic Inquiry 7.391–426.

HIRSCHBERG, JULIA, and JANET PIERREHUMBERT. 1986. The intonational structuring of discourse. Proceedings of the 24th annual meeting, 136–44. New York: Association for Computational Linguistics.

KARTTUNEN, LAURI. 1976. Discourse referents. Syntax and semantics VII: Notes from the linguistic underground, ed. by James McCawley, 363–86. New York: Academic Press.

KIPARSKY, PAUL. 1982. Lexical phonology and morphology. Linguistics in the morning calm, ed. by In Seok Yang, 3–91. Seoul: Hanshin.

LAKOFF, GEORGE, and JOHN ROSS. 1972. A note on anaphoric islands and causatives. Linguistic Inquiry 3.121–25.

Levi, Judith N. 1978. The syntax and semantics of complex nominals. New York: Academic Press.

Lieber, Rochelle. 1984. Grammatical rules and sublexical elements. Papers from the parasession on lexical semantics, Chicago Linguistic Society 20.187–99.

Matthews, Peter. 1974. Morphology. Cambridge: Cambridge University Press.

McKoon, Gail, and Roger Ratcliff. 1989. Semantic association and elaborative inference. Journal of Experimental Psychology: Learning, Memory, and Cognition 15.326–38.

——; Gregory Ward; Roger Ratcliff; and Richard Sproat. 1990. Morphosyntactic and pragmatic manipulations of salience in the interpretation of anaphora. Evanston, IL, and Murray Hill, NJ: Northwestern University and AT&T Bell Laboratories, ms.

Mohanan, K. P. 1986. The theory of lexical phonology. Dordrecht: Reidel.

Morgan, Jerry. 1978. Toward a rational model of discourse comprehension. Proceedings of TINLAP-2, ed. by David Waltz, 109–14. New York: Association for Computing Machinery.

Murphy, Gregory. 1985. Psychological explanations of deep and surface anaphora. Journal of Pragmatics 9.785–813.

Pesetsky, David. 1979. Russian morphology and lexical theory. Cambridge, MA: MIT, ms.

Pierrehumbert, Janet, and Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. Intentions in communication, ed. by Phil Cohen, Jerry Morgan, and Martha Pollack, 271–311. Cambridge, MA: MIT Press.

Postal, Paul. 1969. Anaphoric islands. Chicago Linguistic Society 5.205–39.

Prince, Ellen F. 1981a. Toward a taxonomy of given-new information. Radical pragmatics, ed. by Peter Cole, 223–55. New York: Academic Press.

——. 1981b. Topicalization, focus-movement, and Yiddish-movement: A pragmatic differentiation. Berkeley Linguistics Society 7.249–64.

——. 1986. On the syntactic marking of presupposed open propositions. Papers from the parasession on pragmatics and grammatical theory, Chicago Linguistic Society 22.208–22.

Radford, Andrew. 1988. Transformational grammar. Cambridge: Cambridge University Press.

Ratcliff, Roger, and Gail McKoon. 1988. A retrieval theory of priming in memory. Psychological Review 95.385–408.

Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. Philosophica 27.53–94.

——. 1983. Anaphora and semantic interpretation. Chicago: Chicago University Press.

Rooth, Mats. 1985. Association with focus. Amherst, MA: University of Massachusetts dissertation.

Ross, John. 1971. The superficial nature of anaphoric island constraints. Linguistic Inquiry 2.599–600.

——. 1986. Infinite syntax. Norwood, NJ: Ablex.

Rothkopf, Ernest; Mary Koether; Marjorie Billington; and Barbara Biesenbach. 1988. Why are certain sentence constructions mnemonically robust for modifiers? New York: Columbia University, ms.

Sag, Ivan A., and Jorge Hankamer. 1984. Toward a theory of anaphoric processing. Linguistics and Philosophy 7.325–45.

Shibatani, Masayoshi, and Taro Kageyama. 1988. Word formation in a modular theory of grammar: Postsyntactic compounds in Japanese. Lg. 64.451–84.

Sidner, Candace. 1979. Towards a computational theory of definite anaphora comprehension in English discourse. Cambridge, MA: MIT dissertation.

Simpson, Jane. 1983. Aspects of Warlpiri morphology and syntax. Cambridge, MA: MIT dissertation.

Sproat, Richard. 1985. On deriving the lexicon. Cambridge, MA: MIT dissertation.

——. 1988. On anaphoric islandhood. Theoretical morphology, ed. by Michael Hammond and Michael Noonan, 291–301. Orlando: Academic Press.

——, and GREGORY WARD. 1987. Pragmatic considerations in anaphoric island phenomena. Chicago Linguistic Society 23.321–35.

TANENHAUS, MICHAEL K., and GREG N. CARLSON. 1990. Comprehension of deep and surface verb phrase anaphors. Language and Cognitive Processes 5.257–80.

TIC DOULOUREUX, P.R.N. 1971. A note on one's privates. Studies out in left field: Defamatory essays presented to James D. McCawley on the occasion of his 33rd or 34th birthday, ed. by Arnold Zwicky et al., 45–52. Edmonton & Champaign: Linguistic Research, Inc.

VAN DIJK, TEUN A., and WALTER KINTSCH. 1983. Strategies of discourse comprehension. New York: Academic Press.

WATT, W. 1975. The indiscreteness with which impenetrables are penetrated. Lingua 37.95–128.

WEBBER, BONNIE. 1979. A formal approach to discourse anaphora. New York: Garland Press.

WILSON, DEIRDRE, and DAN SPERBER. 1979. Ordered entailments: An alternative to presuppositional theories. Syntax and semantics XI: Presupposition, ed. by Choon-Kyu Oh and David Dinneen, 299–323. New York: Academic Press.

Gregory Ward
Department of Linguistics
Northwestern University
2016 Sheridan Road
Evanston, IL 60208