

Referential constraints on lexical items

PIETER SEUREN

The following represents a very provisional report on work going on. The author is fully aware of the fact that there are many uncertainties, unclarities and errors in the text as presented here. Yet it seems that the observations which are made are of a certain interest, and that the ideas put forward deserve some consideration in so far as they indicate a general direction in which an explanatory solution to the problem under consideration might be found.

It is a very widespread phenomenon, in many if not all languages, that verbs are semantically decomposable into a fairly general transitive verb and a generic object. An example from English is *brew*, which is made up of *make* and *beer*, or *advise*, which is 'give advice'. Assuming this observation to be correct, we must posit a rule in the grammatical semantics of English, which incorporates the object into a lexical verb. Let us speak of *Object-Incorporation*. Sometimes, as in the case of *brew*, the incorporation is optional, since we also have *brew beer*. Further details could be worked out, such as the fact that an adjective qualifying the object ends up as an adverb: *advise someone well*, but these are not our concern here.¹

What does concern us here is the fact that the object which is the object of Object-Incorporation is always a generic NP (or a sortal NP, if you like), but never a referring expression. It is not possible to have a lexical item decomposable into, for example, '*make the beer*', where *the beer* is used to refer to some beer which either actually exists or has just been mentioned (although it may not exist).

Likewise, we detect an ambiguity in, for example:

(1) This is a joke about a horse

where either a specific horse is mentioned, which this is said to be a joke about, or the kind of joke specified that this is said to be. Only the latter meaning, not the former, corresponds to:

¹ The alternation between dative and accusative in *give advice to somebody* and *advise somebody* follows from Object-Incorporation if the indirect object is taken to derive from a bare NP preceding the object-NP. Further arguments, especially in connection with Predicate Raising, are given in Seuren (1972).

(2) This is a horse-joke

Elements cannot be incorporated into a compound noun when they have what we may reasonably call a referring function. Only sortal expressions allow this to happen to them. Many similar observations can be made. The following sentences are synonymous only in the non-referring sense of (3a):

- (3) a. I like to work in a garden
 b. I like gardening

It is in keeping with these observations (especially (2)) that referring expressions cannot occur in compound words other than in head position. An expression such as *that coathanger* has *hanger* as the head of the compound noun, and the deictic *that* relates to that and not to *coat*. It is unthinkable that any such expression could be analysed as '*hanger of that coat*'.

If rules are to be formulated at all to relate semantic representations to surface structure, they will have to bring together elements from various positions in semantic trees under one category node, since semantically composite lexical items and compound-words occur under one category node in surface structure. Various observations made by McCawley and others suggest that there is not a direct and immediate transition from semantic trees to lexicalized phrase-markers, but, rather, that the various elements making up the contents of a lexical item are first brought together under a category node, as a sub-tree, and then replaced by a lexical item or a morphologically compound word. Following a well-known terminology, we might call such sub-trees *islands*.¹ We will, in this paper, formulate two constraints on the entering of such lexical islands. Both constraints will be formulated in terms involving the notion of 'referring expression'.

The constraints are not, however, limited to the entering into lexical islands by referring expressions. Other elements are also sometimes prohibited from entering a lexical island, but the prohibition seems to depend on external referring expressions. Take, for example, the following two sentences:

- (4) a. Alcoholics are getting younger
 b. Alcoholics are rejuvenating

The (a)-sentence is ambiguous in a way the (b)-sentence is not. It either refers to one set of people, alcoholics, and says these are getting younger,

¹ It is perhaps worth speculating that such islands form the domain of morphological rules, where, for example, command relations hold not with respect to S-nodes but with respect to any category-node dominating the domain.

instead of getting older as would be expected. This reading it shares with (4b). But it can also mean, quite naturally, that, given two sufficiently distant moments in the not too remote past, those who were alcoholics at the earlier moment tended to be older than those who were alcoholics at the later moment, and that the present is one of these moments. Let us distinguish the two readings by saying that the former involves single reference, but the latter multiple reference with regard to whoever is or was an alcoholic. The lexical verb *rejuvenate* does not allow for multiple reference.

The same is true for *have forgotten* vis-à-vis *no longer remember*.¹ In (5), the (a)-sentence allows for multiple reference in the object, but the (b)-sentence does not. (a) and (b) are synonymous only on the single reference reading:

- (5) a. I no longer remember the names of my students
 b. I have forgotten the names of my students

De Rijk (1968) provides abundant evidence that there is a significant area of overlap in the ways expressions *no longer remember* and *have forgotten* are used, where the two are synonymous. A simple case is:

- (6) a. I no longer remember John's address
 b. I have forgotten John's address

Both expressions require that the subject did, at some time in the past, know or remember the object. Failing this, neither expression can be used truthfully. De Rijk quotes Shakespeare to support this point (*The Taming of the Shrew*, Act V, sc. 1, 49):

Vincentio: Come hither, you rogue. What, have you forgot me?

Biondello: Forgot you! No, Sir: I could not forget you, for I never saw you before in all my life.

Both expressions are factive: the object-*S* is taken to be true by the speaker:

- (7) a. Kissinger, conveniently, no longer remembered that he had agreed to a deadline.
 b. Kissinger had conveniently forgotten that he had agreed to a deadline.

Hence the implicit contradiction in sentences such as:

- (8) a. I no longer remember that today is my birthday.
 b. I have forgotten that today is my birthday.

¹ This observation is due to R. P. G. de Rijk. It prompted him to write 'A note on pre-lexical predicate raising' (1968).

De Rijk presents this observation not as an argument against semantic decomposition of lexical items, but as a problem which has to be solved if such a procedure is to be at all viable. George Lakoff (1971:272) proposes that lexical decomposition does not imply complete synonymy between the analysans and the lexical item, but only an inclusion of meaning: 'McCawley's conjecture...only requires that the meaning of *cease to know* be contained in the meaning of *forget*, which it is.' According to him a lexical item may mean more than its analysans. In this case, he maintains, *forget*, includes '*cease to know*', but it means '*cease to know due to a change in the mental state of the subject*'. This answer to the problem raised by de Rijk is, however, too facile. For one thing, it does not rule out the multiple reference reading for *forget* (*have forgotten*). I may cease to know the names of my students (although I used to be able to retain their names in the old days) due to a change in my mental state: my memory is no longer what it used to be. Moreover, the rider '*due to a change in the mental state of the subject*' appears ad hoc in the light of a host of parallel observations, for which this rider would provide no solution. Thus we have, apart from (4):

- (9) a. I am no longer satisfied with my teachers
- b. I have become dissatisfied with my teachers
- (10) a. I no longer have my watch with me
- b. I have lost my watch

We must, instead, try to formulate the general principle accounting for the observations made, in so far as they are related by some general principle. It has proved to be a fruitful and promising method to set up possible semantic analyses for the sentences involved, or at least to specify certain minimal properties which their semantic analyses may reasonably be thought to have, and then to state what constraints can be detected holding between semantic representations and surface structure, especially in connection with lexical items.

Let us assume the following two semantic analyses for the two readings of (5a):

- (11) a. it used to be the case that:
 - I knew x [x were the names of my then students]
 - and:
 - now:
 - not:
 - I know y [y are the names of my now students]
- b. it used to be the case that:

I knew *x* [*x* are/were the names of my now/sometime
students]
and:
now:
not:
I know *x*.

The details are not important. What counts is the minimal claim that there are two propositions with verb *know* and the subject *I* in both (a) and (b), and that these two readings are distinguished principally by the occurrence of a separate referring expression ‘*y*[*y* are the names of my now students]’ in the second proposition of one, but not the other reading.

Both (a) and (b) can be taken to reduce to:

(12) I used to know the names of my students and now I do not know them

or a suitably abstract form corresponding to this. (Notice that the pronoun *them* in (12) is ambiguous between single and multiple reference.) A further reduction could lead to (5a).¹

For (5b), which contains the lexical verb *have forgotten*, let us assume the formation of a lexical island of something like the form ‘*used to know and then (now) not know*’, through application of a particular form of Conjunction Reduction. We would expect both (11a) and (11b) to reduce to a structure corresponding to:

(13) I [used to know and now not know] the names of my students

after which the lexical item *have forgotten* could be substituted for the island. It appears, however, that reduction to (13) is possible only for (11b), not for (11a). This is borne out by the observation made in (5), and also by the fact that:

(14) I used to know but don’t know any longer the names of my students does not allow for a multiple reference reading. It is reasonable to assume that *used to know but don’t know any longer* forms one single constituent in (14), labelled ‘V’, and, thus, an island. Apparently, the formation of such an island is blocked for (11a), but not for (11b). Since the critical difference between the two is the occurrence of a separate referring expression in (a), one is led to infer that it is that expression, occurring in that position, which is responsible for the blocking.

¹ Disregarding the difference between *know* and *remember*, which seems irrelevant in this context.

This inference is confirmed by an analysis of (4). The NP *alcoholics* occurring in both sentences of (4) is of a category which is still poorly understood. It is not even certain whether or not such NPs should be deemed to contain a quantifier. If they do, the logical properties of the quantification are unclear. If they do not, the mode of reference remains problematic. I shall assume that such NPs do not contain quantification but rather a variety of the definite article. Not only does this assumption make the present analysis more regular; it is also in keeping with the fact that ordinary, canonical forms of quantification applied to *alcoholics* in (4a) do not lead to multiple reference ambiguity:

- (15) a. All alcoholics are getting younger
 b. Some alcoholics are getting younger

The definite article, on the other hand, does lead to multiple reference:

- (16) The alcoholics of most countries are getting younger

In order to analyse (4) we must have some theory of the comparative. Whether or not the comparative-analysis adopted here (and mainly taken from Seuren (1973)) is correct is largely immaterial to the present argument, as long as it is agreed that two propositions are involved in the comparative, the second of which is usually shortened. Let us now consider the following semantic analyses for the two readings of (4a):

- (17) a. for all moments i, j (i and j are recent; i precedes j):
 there is an extent e such that:
 at j :
 x [x be alcoholics] be young to e
 and:
 at i :
 not:
 y [y be alcoholics] be young to e
- b. for all moments i, j (i and j are recent; i precedes j):
 there is an extent e such that:
 at j :
 x [x be alcoholics] be young to e
 and:
 at i :
 not:
 x be young to e

Let there be a lexical island *be young to some extent e at j and at i not* (with *i* and *j* defined as in (17)), for which the verb *rejuvenate* can be substituted, then the formation of such an island and the subsequent insertion of *rejuvenate* must be blocked for (17a), which contains a separate referring expression in the second proposition of the *and*-conjunction. The parallel with (11a) is obvious.

Can we now formulate a principle, or constraint, which could account for all observations made so far? It seems the following might be a viable starting point:

- A. If a referring expression *R* commands a node *A* in the semantic representation of a sentence, then any lexical island absorbing *A* must still be commanded by a constituent *R'* whose head is derived from *R* or from any expression making the same reference.

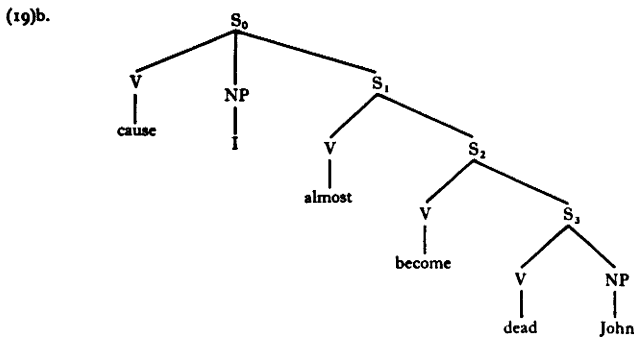
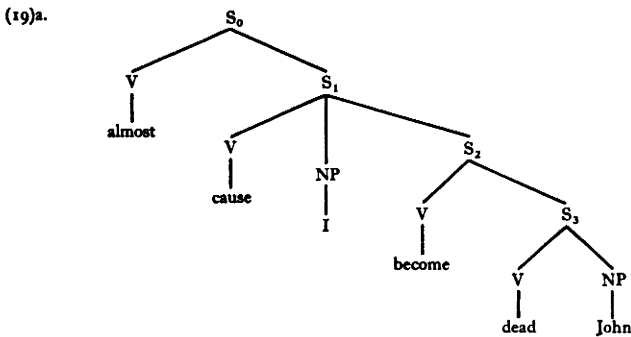
In (11a) and (17a) the referring expressions under the variable *y* command the predicates *know* and *be young*, respectively. For (11a) a lexical island of anything like the form *used to know and now not know* would entail the loss, through deletion, of the second referring expression relating to names of students, i.e. the expression under the variable *y*. Such a lexical island can therefore not be formed for (11a). A similar argument applies to (17a). There is no obstacle to such an island, however, when the sentence is derived from (11b). The island will then be commanded by the referring expression under *x*, which makes the same reference as the deleted *x*. The same, again, applies to (17b).

Constraint *A* prevents the incorporation of a referring expression such as *the beer* into a lexical verb which would then stand for '*make the beer*', since the referring expression would no longer command the verb *make*, as it did in the semantic representation. Likewise, if *a garden* in (3a) is a referring expression it commands *work* (or its semantic forbear) in semantic structure. This prevents it from being incorporated into a lexical verb *gardening*, since it would then no longer command the lexical island of which *work* has become a part. If, as seems reasonable, we extend this explanation to cover compound nouns such as *horse-joke* in (2), we must stipulate that the ordinary command-relation, as mentioned in constraint *A*, which is defined with respect to *S*-nodes, breaks down inside compounds. It might be suggested that the ordinary notion of *S*-command ceases to apply inside lexical islands generally (see p. 85, n. 1), and that a notion of *C*-command takes over (where *C* is any category-node dominating the lexical island). Further support would have to be derived from the study of morphological processes, however, for such a proposal to be made in seriousness.

There is another category of cases where the difference between referring and non-referring expressions seems to be crucial for the formation of lexical islands. As is well-known, McCawley proposes, in various papers, to derive the verb *kill* from an underlying structure containing at least the elements 'cause to become dead'. This analysis has been both confirmed and disconfirmed. It was confirmed by Morgan (1969) who observed that a sentence such as:

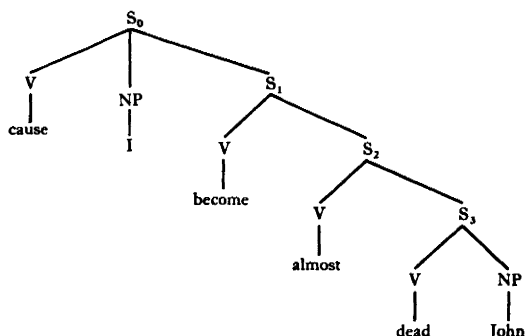
(18) I almost killed John

has a threefold ambiguity. It can mean that the speaker almost brought himself to killing John, or the speaker did something which would have killed John if he had been less lucky, or that the speaker reduced John to such a state that he almost died.¹ The three readings correspond to different semantic trees, where *almost* modifies *cause*, *become*, or *dead*, respectively:



¹ It is worth noting that (18) translates differently into French according to the reading it has. The first reading corresponds to something like *J'aurais tué Jean*; the second is adequately rendered as *J'ai failli tuer Jean*, whereas in the third reading the sentence translates as *J'ai presque tué Jean*.

(19)c.



McCawley's analysis was disconfirmed, however, by Fodor (1970) in an article entitled 'Three reasons for not deriving "kill" from "cause to die"'. Fodor's first reason is perhaps less compelling than the other two. He gives the following two sentences:

- (10) a. Floyd melted the glass though it surprised me that it would do so
 b. *John killed Mary and it surprised me that she did so

Assuming that *melt* in (a) is a causative verb, he infers that *do-so* replacement is permissible for the verb under 'cause', i.e. intransitive *melt*, and that the ungrammaticality of (b) shows that *kill* cannot be a causative verb. This argument, however, depends crucially on Fodor's judgement that (a) is grammatical, but not (b). But although the ungrammaticality of (b) is beyond doubt, most speakers disagree with Fodor on (a), which is generally considered ungrammatical.

His other two arguments are more forceful. He observes that:

- (21) John caused Bill to die on Sunday by stabbing him on Saturday.

is a good English sentence, but that insertion of *kill* turns (21) into the very dissimilar:

- (22) John killed Bill on Sunday by stabbing him on Saturday.¹

He observes, moreover, that in (23) there is an ambiguity as to the semantic subject of *swallowing*:

- (23) John caused Bill to die by swallowing his tongue.

Here either John or Bill swallowed his tongue, which caused Bill's death.²

¹ Fodor gives (22) as ungrammatical. This does not seem correct, since it is a perfectly good way of expressing the contradictory proposition which it does express.

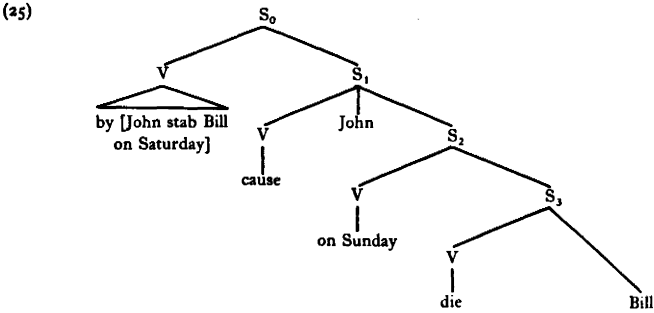
² There is a further ambiguity according to whether *his* is or is not reflexive. Pragmatic difficulties arising in connection with two of the four possible readings can be left out of account here.

But if *kill* is inserted here, only *John* can be the semantic object of *swallowing*:

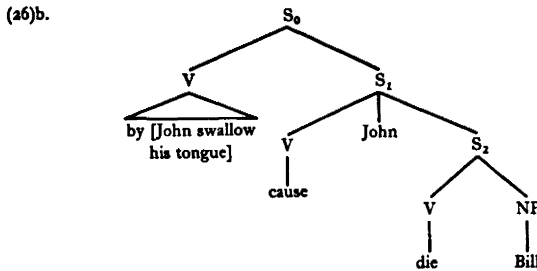
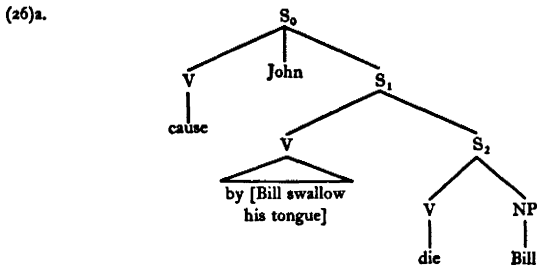
(24) John killed Bill by swallowing his tongue.

These observations are no doubt correct. And they put a serious difficulty in the way of McCawley's analysis of *kill*.

Let us, in accordance with general assumptions adhered to in this paper, consider temporal adverbials such as *on Sunday* and *by*-phrases as in (21)–(24) to be operators in semantic structure. A partial analysis of (21) will then be:

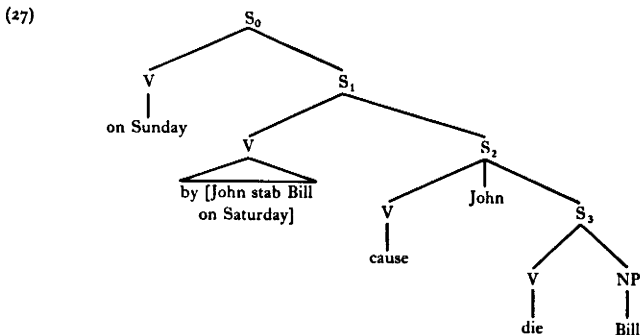


Likewise, we can give the following partial analysis for the two readings of (23):



We agree in principle with Fodor when he argues that the subject of the by-phrase is deleted when it is co-referential with the subject of the S immediately underneath it. We thus see that *John* is deleted in (25) and (26b), and Bill in (26a).

We now notice that *kill* can be used only in (26b), not in (25) or (26a). We notice furthermore, that, according to the same principles of analysis, (22) is analysed as follows:¹



It would thus seem as though *kill* can only be inserted when its component elements ('*cause*', '*become*', '*dead*') occur, in semantic representation, in subsequent Ss, without alien elements intervening. This restriction cannot be correct, however, since (19) shows that *almost* does occur as an alien element in between the component parts of *kill* without preventing the insertion of *kill*. I would suggest that the answer lies in the referring function of the alien elements (25) and (26a), and, generally, in all cases where such elements block the insertion of a lexical item. *Almost*, on the other hand, is not a referring expression: its semantic function is purely logical. In order to account for the seemingly contradictory observations made, the following constraint on lexical insertion can be formulated:

- B. No material can be made part of a lexical island when, thereby, it moves into or out of the scope of a referring operator (or an operator containing a referring expression).

Constraint B has a certain explanatory power, since it accounts for the following observations. The verb *miss* occurs in the sense of '*not hit*':

- (28) a. I missed the target
b. I did not hit the target

¹ Since, as was noted in p. 92 n. 1, sentence (22) is in some ways contradictory, the analysis presented in (27) can only be partial and incomplete: it does not show up the inherent contradiction.

Yet, although we have (29a), insertion of *miss* results in the ungrammatical (29b).

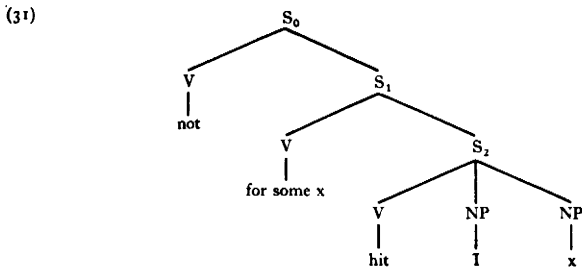
- (29) a. I did not hit anything
 b. *I missed anything
 c. I missed everything

The correct semantic counterpart of (29a) is (29c).

The verb *refuse* occurs in the sense of 'not accept' and 'not give'. These two senses can be seen to be closely related: both are reducible to 'not allow to have'. When the subjects are equal the result is 'not accept', otherwise *refuse* corresponds to 'not give'.¹ Here again, we see that *anything* must be replaced by *everything* when the lexical verb *refuse* is used:

- (30) a. He did not accept anything
 b. *He refused anything
 c. He refused everything

Analysing (29a) according to the principles adopted and investigated here, we find:

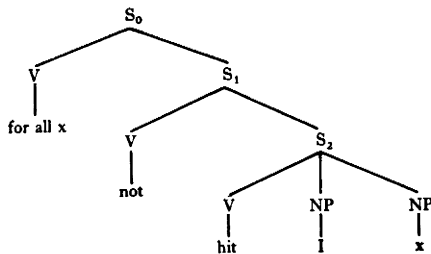


The element 'for some *x*' is considered a referring operator. Constraint *B* now prevents the formation of a lexical island of which both *not* and *hit* would form part, since material would have to be moved across the scope boundaries of the referring operator. The verb *miss* cannot be inserted here, but it can be inserted when we have a logically equivalent² semantic phrase-marker:

¹ *Give* is analysable as 'cause to have'. 'Cause not to have' is logically equivalent to 'not allow to have', and gives a more precise analysis for the one sense of *refuse* than 'not give'.

² The equivalence is not complete: (32) presupposes the existence of certain things to be hit, whereas (31) does not, or not necessarily. In fact, (29a) is ambiguous from this point of view, and only in the presupposition-loaded sense is it equivalent to (29c). See Seuren (forthcoming) for an elaboration of this presuppositional difference.

(32)



The same applies, of course, to (30).¹

A final word remains to be said about the notion 'referring expression' which plays a crucial role in the formulation of the two constraints on lexicalization formulated above. This notion has been used here in a two-fold opposition. On the one hand it is set against generic or sortal expressions (as *milk* in *Drinking milk is something I like*). On the other hand it is opposed to the notion of purely logical, or truth-functional, expressions, such as *almost*, *probably*, *not*, etc. It is clear that more is needed to make the concept of referring expression precise and satisfactory. It is equally clear, however, that an attempt at clarifying that notion would take us very far afield, further than seems desirable for the purposes of this paper.

It is felt, however, that there is some semantic property common to those expressions which have been labelled 'referring expressions' here. Once this is recognized, certain regularities emerge which appear to have an explanatory yield in the context of grammatical semantics. The term *referring* has been used since the semantic property in question seems very intimately connected with what has been recognized as 'reference' in philosophical literature. There, however, the notion is also far from clear and continues to give rise to new problems and new attempts at solving them. (The literature is so abundant that any reference would be arbitrary.) One way in which the ideas presented in this paper might be helpful consists in using the two lexical constraints as tests for problematic cases. When it is felt that the results enhance our insight and reveal further regularities, we have some reason for encouragement.

As an example we might consider the case of proper names. It has often been said that proper names have no 'meaning' but only 'reference' (no 'connotation' but only 'denotation'). This purely referential definition of proper names has also, however, been denied quite frequently. If proper

¹ It should be clear that the constraints formulated above are of a transderivational, or global nature, in Lakoff's sense. Semantic structure keeps plaguing later derivational stages, in this case lexical island formation.

names are indeed purely referential, then our constraints could predict that they can *never be incorporated into* a lexical island. We observe, however, that they can:

- (33) a. Fred is a Nixon-admirer
- b. Charlie is a Stalinist
- c. Albert has been through a process of complete Spiro-Agnewification

We are now faced with a dilemma: are our constraints incorrect, or is the theory at fault according to which proper names have no other semantic function than reference? A closer scrutiny of the semantic subtleties of the sentences in (33) may well help us to decide here. We might ask if the names *Nixon*, *Stalin* and *Spiro Agnew* in (33a, b, and c), respectively, actually refer to the well-known public figures who bear or bore those names. Or do they fulfil the function of sortal specification? It is known, of course, that proper names are sometimes used as sortals:

- (34) Frank is a second Nixon

We notice, furthermore, that when the name *Nixon* is unambiguously used to refer to that person, its incorporation into a lexical island strikes us as highly unnatural:

- (35) a. Two weeks ago John was Nixon's guest
- b. *Two weeks ago John was a Nixon-guest

It does seem appropriate to say that in (33a) it is specified what kind of admiration Fred has, more or less as in:

- (36) To admire a man like Nixon is a crime of opinion

Nor should it pass unobserved that (33c) does not mean that Albert is now identical with Spiro Agnew: this sentence does not entail an identity crisis.

Whatever the answers to questions such as these are, the proposals put forward in this paper provide us with more powerful tools to develop and test hypotheses and to refine our semantic notions than was hitherto the case.

REFERENCES

- De Rijk, R. P. G. (1968). 'A note on prelexical predicate raising', unpublished. To appear in Seuren, ed. (forthcoming).

- Fodor, J. A. (1970). 'Three reasons for not deriving "kill" from "cause to die"', *Linguistic Inquiry*, 1:4, 429-38.
- Kiefer, F. and Ruwet, N., eds. (1973). *Generative Grammar in Europe*. Foundations of Language Supplementary Series. Dordrecht: Reidel.
- Lakoff, G. (1971). 'On generative semantics'. In Steinberg and Jakobovits (1971), 232-96.
- Morgan, J. L. (1969). 'On arguing about semantics,' *Papers in Linguistics*, 1:1, 49-70.
- Seuren, P. A. M. (1972). 'Predicate Raising and Dative in French and sundry languages.' Unpublished.
- (1973). 'The comparative'. In Kiefer and Ruwet, eds. (1973), 528-64.
- (forthcoming). 'Negative's travels'. In Seuren, ed. (forthcoming).
- ed. (forthcoming). *Semantic Syntax*. Oxford: Oxford University Press.
- Steinberg, D. D. and Jakobovits, L. A. eds. (1971). *Semantics. An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*. Cambridge: Cambridge University Press.