J. A. Fodor

Three Reasons for Not Deriving "Kill" from "Cause to Die"[1]

0. Lakoff (1965) has suggested that sentences like (1) derive from deep structures like (2). Prima facie, there would appear to be a variety of types of facts which support this derivation. Notice, for example, that in many dialects one or more versions of (3) are ambiguous in just the way that the derivation of (1) from (2) would predict:

(1) Floyd melted the glass.
(2) (Floyd caused (the glass melt))
(3) Floyd melted the glass $\left\{\begin{matrix}\text{and that}\\ \text{and it}\\ \text{which}\end{matrix}\right\}$ surprised me.

namely, what is said to have surprised me can be either that Floyd melted the glass or that the glass melted. Now, if (2) is the deep structure of (1), these ambiguities are easily explained. On one reading of (3), the pro-forms have replaced the constituent (Floyd caused it (the glass melt)), and, on the other reading, they have replaced only the subordinated constituent (the glass melt). The ambiguity of (3) thus appears to provide strong evidence that (1) is not the simple sentence that it seems to be; in particular, that (1) has two sentoids in its deep structure.

Further arguments might be alleged for the same view; thus, the second *it* in (4),

(4) Floyd melted the glass though it surprised me that he was able to bring it about.

like the pro-forms in one reading of (3), appears to refer to the glass melting rather than to Floyd melting the glass, and this would follow naturally from the view that (the glass melt) is a constituent of the deep structure of (1). Finally, it may be noted that (1) has two associated "do so" forms, namely, (5) and (6):

(5) Floyd melted the glass though it surprised me that he would do so.
(6) Floyd melted the glass though it surprised me that it would do so.

Once again, the natural treatment would appear to require that (2) be the deep

[1] I am deeply indebted to Janet Dean Fodor for the discussions which went into planning this paper.

structure of (1) since, on that account, we could say that (5) is a case of "do so" replacing a matrix VP and (6) is a case of "do so" replacing a constituent VP.[2]

If (1) is derived from (2), the derivation presumably involves two transformations: *predicate raising* of the embedded predicate, which permits us to map (2) onto the intermediate form (7),

(7) (Floyd ((caused to melt) (the glass))

and *lexicalization*, which permits the transformational substitution of a word for a phrase (in the present case, "(cause to $\text{melt}_{\text{Vitr}}$) ⇒ $(\text{melt}_{\text{Vtr}})$."[3] If this analysis is correct, it follows, first, that there are transformations which derive words from underlying phrases, and, second, that at least one prima facie simple sentence containing a surface transitive main verb derives from a complex deep structure containing an intransitive embedded verb. The derivation of (1) from (2) thus provides a direct precedent for such derivations as "break_{tr}" from "cause to $\text{break}_{\text{itr}}$", "$\text{tear}_{\text{tr}}$" from "cause to tear_{itr}", etc. But, moreover, it indirectly supports such dramatic derivations as "kill_{tr}" from "cause to die", etc. The suggestion is that, from a semantic point of view, (8) is related to (9) and (10)

(8) John caused Mary to die.
(9) John killed Mary.
(10) Mary died.

in very much the same way that (i) is related to (1) and (11). On the proposed analysis the syntax mirrors these relations by deriving (9) from (12) in a way that strictly parallels the derivation of (1) from (2).

(11) The glass melted.
(12) (John caused (Mary die))

This argument is obviously appealing; no one could fail to be pleased if the theory of syntax were to support the theory of entailment in this surprising way. To suppose that word-to-phrase synonymies can, in any important class of cases, be handled as

[2] It is interesting to note that the ambiguity of (3) and the wellformedness of (4) and (6) constitute counterexamples to the claim that words are "anaphoric islands" since, on the present analysis, these sentences contain pro-forms which refer to deleted clauses in the transformational source of a word.

[3] Strictly speaking, *predicate raising* and *lexicalization* are supposed to operate not on phrases but on abstract semantic representations. Thus, we must imagine that what appears in the structural index of the rule just cited is not the word "cause", but a semantic feature bundle which will be transformed into "cause" in the derivation of such sentences as (i):

(i) Floyd caused the glass to melt.

This is, however, a refinement which we shall ignore. The arguments to be presented are indifferent to it so long as it is assumed that the feature bundle which underlies the English word "melt_{tr}" differs from the feature bundle which underlies the English word "cause" only by the addition of those features which underlie "melt_{itr}". If that assumption were to be abandoned, the theory would have failed to assign (i) and (1) the same base structures, thus providing a counterexample to the claim that synonymous sentences have identical transformational sources.

instances of *syntactic* relations is to suggest an answer for one of the deepest problems in semantics, since any semantic theory must provide a disciplined technique for representing such synonymies. (Notice, for example, that almost all dictionary entries report synonymies between (defined) words and (defining) phrases.) The present suggestion is that, for at least a wide variety of cases, words are related to their defining phrases in the sorts of ways in which surface structures are related to their transformational sources.

Nevertheless, it seems to be quite certain that the deep structure of (1) is not (2) and that (9) is not derived from (12). In the next three sections, I shall present arguments which seem to me to establish this conclusively. In the final section, I shall suggest a principle which apparently does explain the ambiguity of (3) and the well-formedness of (4)–(6).

1. I remarked above that part of the interest of the derivation of (1) from (2) is that it appears to provide a precedent for the derivation of (9) from (12). This appearance is, however, illusory; the distributional characteristics of "kill/cause to die" are, in fact, different from those of "melt/cause to melt" in ways that militate against handling these pairs symmetrically in the syntax.

To see why this is so, we must first consider the analysis of (9) which derives it from (12) (i.e. from the same deep structure that directly underlies (8)). It is noteworthy, to begin with, that corresponding to (8) we have both (13), in which "do so" replaces the matrix VP "caused Mary to die", and (14), in which "do so" replaces the VP in the constituent sentence "Mary die".

(13) John caused Mary to die and it surprised me that he did so.
(14) John caused Mary to die and it surprised me that she did so.

Now, if both "cause Mary to die" and "Mary die" are constituents in the deep structure of (9), we might expect that the *do-so* transformation should operate on (9) to produce both (15), which is in fact wellformed, and (16) which, however, is not.

(15) John killed Mary and it surprised me that he did so.
(16) *John killed Mary and it surprised me that she did so.

In short, it argues against the presence of a constituent "Mary die" in the deep structure of (9) that there is no wellformed sentence (16) in which that constituent has been replaced by "do so".

This argument is not, however, decisive against the analysis of "kill" as transformationally derived from "cause to die". For it might be argued either that there are special restrictions on the *do-so* transformation which prevent it from applying to (9) in such a way as to generate (16) or, more interestingly, that (16) is precluded by the nature of the ordering relations between the *do-so* transformation and *lexicalization*. The

first line of reply leads to an ad hoc complication of the conditions upon the *do-so* transformation and so need not concern us. The second, however, merits discussion.

Suppose we assume that the transformation which introduces "do so" operates "post lexically" (i.e. *after* the transformation which collapses "cause to die" into "kill"). Since the latter transformation must be optional if we are to account for the wellformedness of (8), this way of ordering the rules permits either of two derivational routes. One possibility is that the *do-so* transformation operates under identity either with the matrix or with the embedded clause in (12) producing, respectively, sentences like (13) and sentences like (14). Alternatively, a lexicalizing transformation collapses (12) into (9) and *do-so* applies under identity with the resulting form to produce (15) but not (16). In short, ordering the *do-so* rule posterior to the transformation which converts "cause to die" into "kill" guarantees that we will never produce the ungrammatical form (16) in which it is *both* the case that "kill" has replaced (cause to die) *and* that *do so* has operated on the verb phrase of the embedded sentence.

So far so good. But, unfortunately, this solution, which works for the relation "kill/cause to die", fails for the relation "$\text{melt}_{\text{tr}}/\text{melt}_{\text{itr}}$". For as we have seen, there *are* two "do so" forms corresponding to sentences like (1), e.g. sentences like (5) and (6). If the base structure of (1) is assumed to be (2), then *do-so* must have replaced the matrix verb phrase "cause it = S" to produce (5) and it must have replaced the subordinated verb phrase "melt" to produce (6). However, it was precisely in order to avoid having *do-so* operate under identity with the subordinated verb phrase in (12) that we assumed that the *do-so* transformation must be ordered after *predicate raising* and *lexicalization*. In short, the transformational ordering required for the "cause to die/kill" case makes the wrong prediction for "$\text{melt}_{\text{itr}}/\text{melt}_{\text{tr}}$" and vice versa. The *do-so* transformation must be assumed to precede *lexicalization* if we are to account for one kind of case, but must be assumed to follow it if we are to account for the other.[4]

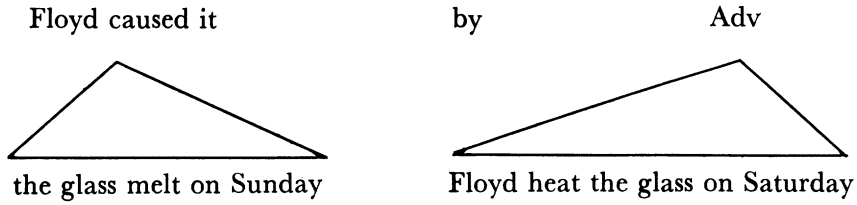
2. Thus far I have argued that the behavior of "do so" militates against assigning a deep structure containing an abstract causative verb to *both* "kill" sentences and "melt_{tr}" sentences. In this section and the following one, I want to argue that, for independent reasons, this sort of base structure ought not be assigned to *either* type of sentence.

The surface form (17) is clearly wellformed and surely derives from a deep structure like (18).

(17) (Floyd (caused (the glass to melt on Sunday))) (by (heating it on Saturday))

[4] This asymmetry between "melt_{tr}" and "kill" *vis à vis do-so* transformation has been independently noticed by Bouton (1969).

(18)



But now, if we are to recognize a transformation which reduces "cause to melt_{itr}" to "melt_{tr}", something will have to be done to prevent that transformation from acting on (18) to produce the ungrammatical (19).

(19) *Floyd melted the glass on Sunday by heating it on Saturday.

The point is, roughly, that one can cause an event by doing something at a time which is distinct from the time of the event. But if you melt something, then you melt it when it melts. If, despite this consideration, we wish to assume an abstract verb "cause" in the deep structure of sentences like (1), we will have to stipulate that *predicate raising* cannot apply to structures like (18) unless any time adverb on the embedded sentence matches any time adverb in the "by" phrase of the matrix sentence. While adverb matching of one sort or another may prove to be a general constraint upon embedding transformations, the particular condition just mentioned is, so far as I know, without precedent in the grammar. It appears to be an artifact of the decision to derive surface verbs like "melt", which are intrinsically constrained with respect to their time adverbs, from the same source that underlies surface verbs like "cause", which are intrinsically relatively free with respect to their time adverbs.[5]

It may be added that, though we have developed this argument in terms of "melt_{tr}/cause to melt_{itr}", the same points apply, mutatis mutandis, to "kill/cause to die". In particular, (20), but not (21), is wellformed.

(20) John caused Bill to die on Sunday by stabbing him on Saturday
(21) *John killed Bill on Sunday by stabbing him on Saturday.

3. The following argument, like the one in Section 2, purports to show that neither (1) nor (9) derive from complex base structures. Since it concerns "instrumental" (or "means") adverbials, and since such adverbials are, by and large, selectionally restricted to verbs that take animate subjects, I shall develop the argument for "kill" rather than "melt_{tr}". The reader can check for himself, however, that, barring selectional restrictions, it works for the latter case as well as the former. Indeed, I think it provides a

[5] I say these constraints are "intrinsic" because they appear to vary from verb to verb. Thus a verb like "qualify" seems, for no very obvious reason, to operate on the same pattern as "cause", vide (ii).

(ii) Floyd qualified for his degree in March by submitting his thesis in December.

general argument against the existence of any transformation which derives structures of the form NP V (NP) from structures containing subordinated sentoids.

To begin with, notice that there are a variety of types of instrumental adverbial phrases which share NP with the verbs they modify, and that they invariably share the subjects of those verbs. Thus, (22) means that John (rather than Mary) used the telephone,

(22) John contacted Mary by using the telephone.

and there is no surface structure corresponding to (23).[6]

(23) *(John contacted Mary by (Mary use the telephone))

The generalization that instrumental adverbs NP share with the deep underlying *subjects* of *their* verbs holds, too, for instrumental adverbs on embedded verbs. Thus, in (24), as in (22), it is John who does the phoning.

(24) I believe that John contacted Mary by using the telephone.

So, it appears that a necessary condition upon the wellformedness of sentences with NP-sharing instrumental adverbs is that the shared NP be the deep subject of the modified verb. Moreover, this appears to be a *sufficient* condition for this kind of NP sharing. That is, given that an NP is the deep subject of a verb, it can ipso facto be shared by an instrumental modifier of that verb (i.e. it can be the implicit subject of an instrumental modifier of that verb). This is apparently quite independent of the surface position that the NP comes to occupy.[7]

[6] Not all adverbial phrases that NP share work this way. For example, we have both (iii) and (iv) (compare (22) but *(23)).

(iii) John bit Mary after Mary ate dinner.
(iv) John bit Mary after John ate dinner.

What is striking about NP-sharing time adverbs is not only that they can NP share with deep objects, but also that *equi-NP* deletion seems to be governed by *surface* subjects. Thus, in (v) it was John who used the telephone, but in (vi) it was Mary.

(v) John contacted Mary after using the telephone.
(vi) Mary was contacted by John after using the telephone.

In both respects, NP-sharing time adverbials contrast sharply with noun-sharing instrumental adverbials. (Cf. fn. 7 et supra.) I am indebted to S. J. Keyser for having pointed out to me these curiosities in the behavior of time adverbs.

[7] Subject to the condition that in some dialects some such adverbial phrases cannot appear in full passives. Thus *(vii):

(vii) *Mary was hit by John by using a hammer.

That this exception is unsystematic (possibly the consequence of a surface constraint against iterating "by"-phrases) is suggested by a variety of considerations. For example, (viii) is wellformed and follows the rule that it is the deep *subject* that is shared.

(viii) Mary was hit with a hammer by John.

Similarly, in the short passive (ix),

(ix) Mary was found by using radar.

although the deep subject of 'find' has been deleted, it is nevertheless the deep subject of 'find' (i.e. the unspecified someone who found Mary) who is understood to have used the radar. In short, it appears that the difference

If this is true it is important, since it provides us with a test for determining whether an NP has ever been the subject of a verb in the deep structure of a sentence. Namely, barring context restrictions, an NP is the deep subject of a verb in a sentence if and only if it can be shared with an instrumental adverbial phrase in that sentence.

We can validate this test by showing that it is in fact satisfied by NPs wherever we have independent motivation for assuming them to be deep subjects of some verb. Thus:

a. An NP which has been promoted from deep subject of an embedded verb to surface object of a higher verb can nevertheless be the implicit object of an instrumental adverbial on the embedded verb, as in (25).

(25) John expected Mary to treat her cold by taking aspirin.

b. An NP can be shared by an instrumental adverb despite nominalization of the sentoid of which the NP is subject, as in (26).

(26) John's breaking windows by using a hammer surprised us.

c. An NP can be shared with an instrumental adverbial on an embedded verb after having been promoted from deep subject of that verb to superficial subject of a higher verb, as in (27).

(27) John seems to break windows by using a hammer.

d. A subject NP can be shared with an instrumental adverb even after deletion of its verb by "gapping", as in (28).

(28) John breaks windows by using a hammer and Bill by using a brick.

e. Finally, an (implicit) subject NP can be shared with an instrumental adverbial even after the former has been deleted, as in (ix).

All this strongly suggests the correctness of the claim that a structurally necessary and sufficient condition for a NP being shared with an instrumental adverbial is that the NP be the deep subject of the verb that the instrumental modifies.[8] However, we will have to abandon this generalization if we permit derivations like (9) from (12) or, mutatis mutandis, (1) from (2).

Notice that there is a deep structure, (29), which transforms into the sentence (30).

(29) (John caused (Bill die)) (by (Bill swallows Bill's tongue))

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

(30) is ambiguous, just as the principle that any deep subject can be shared by an

between NP sharing time adverbs (see fn. 6 above) and NP sharing instrumentals is intrinsic; the former NP share with deep subjects *or* deep objects and *equi-NP* delete under identity with surface subjects. The latter NP share *only* with deep subjects.

[8] This formulation is used advisedly; for example, it is not satisfied by NPs that are subjects of predicate adjectives. Thus, (x) is wellformed if it is read as a short passive with deleted subject and main verb "break". But it is ungrammatical if it is read as *NP be Pred Adj*.

(x) The vase was broken by using a hammer.

instrumental adverbial requires. (That is, (30) has the source (31) as well as the source (29).)

(31) (John caused (Bill die)) (by (John swallows Bill's tongue))

Now, if we suppose that *predicate raising* and *lexicalization* are transformations, we can derive not only (30), but also (32) from (29).

(32) John killed Bill by swallowing his tongue.

But this will not do since (32), unlike (30), is clearly univocal and it is clearly John rather than Bill who does the swallowing.

In short, "Bill" cannot be shared with an instrumental adverbial in (32) despite the fact that "Bill" is the subject of a verb in (29) and, ex hypothesis, (29) is a deep structure source of (32). Hence, if we want to save *lexicalization* and *predicate raising* as transformations, we shall have to do so at the price of abandoning the generalization (which appears to be firmly supported by (ix) and (25)–(28)) that all deep subjects can be shared by instrumental adverbials. On the present evidence, then, it looks as though "Bill" is not the subject of any verb in the deep structure of (32) but simply the object of "kill"; an undramatic but thoroughly intuitive conclusion.

One further point. Suppose we (gratuitously) decide to save *lexicalization* and *predicate raising* at the cost of abandoning the generalization about NP sharing in instrumental adverbials. We can do this by accepting the constraint that *predicate raising* and/or *lexicalization* fails when the embedded verb is modified by an instrumental, or by any other kind of adverb that NP shares in the way that instrumentals do. What is worth noticing is that we *cannot* do it by a wholesale prohibition against *predicate raising* (or lexicalizing) modified embedded verbs. For, such a wholesale prohibition would block not only the derivation of (32) from (29), but also the derivation of (33) from (34),

(33) John cooked the meat slowly.
(34) (John caused (the meat cook slowly))

and that won't do since (33) *does* have a reading on which it means what (34) predicts. In short, any attempt to save *predicate raising* and *lexicalization* would have to include in one or the other of their structural indices an enumeration of the kinds of modifiers that let an embedded verb go through and the kind that require it to block. I think that the rational conclusion at this point is that the game's not worth the candle.

4. It is worth pausing to reflect upon the moral of the arguments in Sections 2 and 3. *Lexicalization* is a transformation which purports to derive words from phrases. But phrases are, ipso facto, syntactically complex objects in a way that words ipso facto are not. We might thus expect that phrases will exhibit distributional characteristics which differ from those even of words with which they are synonymous. That is, it is simply because they have internal syntactic structure that phrases can interact with

syntactic rules in ways that can prove embarassing for *lexicalization*. In the examples we have been investigating, the phrases offered as candidates for *lexicalization* permit of modifiers. The corresponding words resist some of these modifiers simply because they lack internal structures on which to hang them. We can have two time modifiers on (17) simply because there are two verbs capable of receiving them.[9] But there is only one verb available for modification in (1); hence either we must resist the temptation to lexicalize structures like (18) or we must specify ad hoc that *lexicalization* goes through only when certain identity conditions are satisfied by any time adverbs that the clauses in (18) may happen to have picked up.

Analogously, there are two instrumental adverb positions in (8) but only one in (9). This follows simply from the fact that (8) is a two verb sentence while (9) is a one verb sentence. But this formal difference between the structures produces an embarassment of adverbs if we attempt to lexicalize (8) into (9). In particular, *lexicalization* predicts surface structures in which instrumental adverbs are inherited from both the matrix and the embedded sentence in (8), and such structures do not exist.

In short, even where a phrase and a word are synonymous, the former will characteristically exhibit degrees of syntactic freedom unavailable to the latter; two verb sentences are, ipso facto, and independent of their meaning, different in their behavior from one verb sentences. There is thus a dilemma. Either *lexicalization* carries these unwanted degrees of freedom over into surface structure, thereby predicting sentences which are in fact ungrammatical, or special, ad hoc constraints have to be instituted to insure that *lexicalization* does not apply to phrases in which these degrees of freedom have been exploited. This seems to me to be a principled reason for doubting that there are transformations which map phrases onto words.

5. I think the arguments presented in Sections 2–4 are fairly decisive against the transformational analysis of verbs like "melt", and "kill". But it remains to try to find some alternative explanation of the fact that (3) is ambiguous and that we have the two "do so" forms (5) and (6). It was these facts that provided the strongest evidence for the transformational derivation of "melt_{tr}" from "cause to melt_{itr}".

It is tempting to argue that we are faced not with a fact about "causal" verbs but rather with a fact about "pro-" forms like "do so" and "it". In particular, I suspect that the widely held view that such forms enter into surface structure only as a result of deletion under identity (i.e. that they must invariably "refer back to" or "replace" material actually present in the base structure of a sentence) may be false. In certain cases, pro-forms may refer to material which is semantically and phonologically related to (but not identical with) material actually contained in the sentence in which the

[9] Only two because, I think, there is a context restriction which prohibits "cause" from accepting the sorts of time adverbs that "melt", "heat_{vtr}", "persuade", etc. accept. Compare (xi) and (xii):

(xi) *(What John did on Tuesday (caused on Wednesday (the glass melt on Thursday)))

(xii) What John said on Tuesday persuaded Bill on Wednesday to leave for Moscow on Thursday.

Roughly, "cause" names a relation between events, and relations have no dates.

pro-form occurs. Thus, given a verb V which, like "melt", "break", etc. (but *un*like "kill", etc.) has the same phonological shape in its transitive and intransitive forms, and where "x V y" entails "y V_{itr}", it will generally be found that one or more of the pro-forms ("it", "do so", etc.) can "refer" to the intransitive form even when the pro-form occurs in transitive sentences like (3). It is this "loosening" of the conditions on the reference of pro-forms which explains why (3) is ambiguous and why both (5) and (6) are wellformed.

It may be maintained that this explanation requires an ad hoc exception to the otherwise tidy rule that pro-forms can refer only to material actually present in the sentences in which they occur. That *is* a nice rule, but, unfortunately, it has to be abandoned in any case. Sentences containing logically "symmetrical" verbs (i.e. where "x V y" entails "y V x") behave very much like "melt", "break", etc. in respect to pro-forms. Thus, we have not only (35) but also (36).

(35) John married Mary though we were surprised that he was willing to do it.
(36) John married Mary though we were surprised that she was willing to do it.

In the latter case, "it" refers to "Mary marries John", which surely is *not* a constituent of that sentence.[10]

In short, given a structure A, which is not present in, but which bears appropriate phonological and semantic relations to, items in the deep structure of sentence S, pro-forms in S may refer to A. It seems pretty likely that this is true, but how one goes about saying that it is in the framework of generative syntax is a problem I don't know how to solve.[11]

References

Bouton, L. (1969) "Identity Constraints on the Do-so Rule," *Papers in Linguistics* 1, 231–247.
Lakoff, G. (1965) *On The Nature of Syntactic Irregularity*, in *Mathematical Linguistics and Automatic Translation, Report No. NSF-16*, the Computation Laboratory of Harvard University, Cambridge, Mass.

Department of Psychology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

[10] I know of only one clear counterexample to this claim about symmetrical verbs, namely (xiii). The fact that (xiii) is ungrammatical, together with the fact that (36) is considerably better than (xiv) suggests that the generalization ought to be restricted to "volitional" symmetrical verbs. Why this should be the case, I have no idea.

(xiii) *John resembles Mary though it surprised us that she does.
(xiv) ?John married Mary though we were surprised that she did.

[11] Theorists who feel strongly about saving the theory that pro-forms invariably enter surface structure as the consequence of deletion might want to play with the possibility of liberalizing the deletion conventions. Thus, (6) might derive from (xv) under a deletion rule whch permits transitives to delete intransitives, and (36) might derive (xvi) under a deletion rule which permits NP_1 V NP_2 to delete the V NP_1 in NP_2 V NP_1.

(xv) Floyd melt the glass though that the glass would melt surprise me.
(xvi) John marry Mary though that Mary was willing for/to Mary marry John surprise me.