J. Linguistics, 19 (1983), 29–58. Printed in Great Britain

Constraints on the agentless passive

E. JUDITH WEINER AND WILLIAM LABOV

Temple University and University of Pennsylvania

(Received 18 February 1981)

This paper is a quantitative study of the factors that determine the selection of passive constructions over active ones by English speakers.[1] By examining a large body of passives used in spontaneous speech, together with the sentences that show an opposing choice, we are able to throw light on the crucial question of which syntactic and which semantic features of the environment act to constrain the choice and whether syntactic or semantic factors predominate in this case. In the course of the analysis, we will also have something to say about the social factors that have been reported to determine the use of the passive.

Do active and passive mean the same thing? The active–passive relationship has long been recognized as a highly productive pairing of syntactic forms and has played a major rôle in early statements of transformational grammar. In recent years, limitations of this productivity have become a major focus, and such problems as the difficulty of drawing a sharp line between passivizable and non-passivizable verbs have led a number of linguists to exclude the passive from the array of movement transformations (Bresnan, 1978). Differences in intuitive reactions to active and passive forms with quantifiers, first noted by Chomsky (1957: 100–1), have led many to argue that there is no meaning equivalence of active and passive. Thus for the pair

(1) Everyone likes someone
(2) Someone is liked by everyone

G. Lakoff states that 'in my speech, though not in that of all speakers of English, (1) and (2) have different meanings' (1970: 14–15). Katz & Postal

[1] This analysis is the product of the research project on Linguistic Change and Variation, supported by the National Science Foundation under NSF SOC7500245. The original framework for analysis was developed by W.L. for a study of the passive by members of the class on Sociolinguistic Analysis (Linguistics 561) at the University of Pennsylvania, who developed the first body of data referred to below. E.J.W. developed the present coding system and reanalysed the initial materials along with other bodies of speech drawn from the records of the project and ran the variable rule programmes. The final analysis is the joint work of both authors. We are particularly indebted to Ivan Sag for many detailed comments and corrections, though he cannot be considered responsible for any failure on our part to close the gap between the empirical study of the passive and most recent thinking in formal syntax. We are also indebted to Susumu Kuno, Beatriz Lavandera, and Anthony Kroch for their critical reactions, which we hope we have responded to, at least in part.

(1964: 72–3) were not convinced by Chomsky's original example and found both meanings in active and passive; we have the same reaction to these.

Yet there cannot be any doubt that some passive constructions do not mean the same as their corresponding actives. The latest attack on the problem of explaining these differences is that of McConnell-Ginet (1982), who deals with the semantic differences produced by adverbial modification:

(3) Reluctantly, Joan instructed Mary.
(4) Reluctantly, Mary was instructed by Joan.

McConnell-Ginet concludes that there are several different types of adverbial modification involved, and that the passive's two verbs offer more possibilities for such modification than the active does. In spite of these potential differences in the meaning of active and passive in particular contexts, our observations of the use of this option in spontaneous speech convince us that the choice of the agentless passive and generalized active is fundamentally a syntactic one, and that active and passive normally have the same meaning in a truth-conditional sense. Many of the differences between active and passive that have been observed seem to us to be differences in focus or emphasis, characteristic effects of the re-ordering of sentential elements. It seems to us preferable to restrict the term 'meaning' more narrowly to designate the coupling of a given sentence with a given state of affairs.[2]

One strategy that we might follow is to say that we are using 'rough semantic equivalence' and admit that a more precise sense of the meaning of active and passive would require us to register every possible inference that one individual or another would be likely to make in interpreting one or the other form. But it seems to us that this concession to an idealistic semantics is needlessly unrealistic. If we isolate words from their use, we can show that there is no such thing as a precise synonym, since all words have slightly different privileges of occurrence when we consider every possible context. But in practice, the need for stylistic variation leads all speakers and writers of English to substitute one word for the other with the expectation that any differences that might arise in other contexts will not affect interpretation in that one. In our study of active and passive, we are concerned with the information that is transmitted by this choice in everyday life. In our data there is ample evidence (see below) that the two forms are used interchangeably

[2] In this respect, we are continuing in the tradition of Weinreich's efforts to restrict 'meaning' to significata that are shared throughout the speech community (see his review of Osgood *et al.*, 1958). Recent responses to evidence on the unreliability of quantifier dialects have admitted that both meanings can be found but claimed that individuals idiosyncratically 'prefer' one reading or the other. In an unpublished paper, 'Dictionaries of the future: A set of parameters for descriptive semantics' (to appear), Weinreich pointed out that descriptive semantics must deal with the 'constant, institutionalized aspect of the meaning of signs . . . without denying the existence of a non-institutionalized margin of meaning'.

to refer to the same states of affairs. This does not rule out the possibility that there are some contexts where this is not so.

An earlier investigation of the choice of the *get* vs. *be* passive (Labov, 1975) demonstrated that these function as semantic equivalents in general contexts even though they were semantically differentiated in the special context of purpose clauses. Reactions to *He got arrested* and *He was arrested* in context showed the same proportion of interpretations of the subject as the agent (= 'He got himself arrested'). But *He got arrested to test the law* gave a much higher proportion of such interpretations than *He was arrested to test the law*. We can infer that there will be contexts where active and passive lead to different semantic interpretations of states of affairs and ultimately it will be important to locate these contexts, but these will undoubtedly be a small subset of the total range of uses and not likely to affect our search for the general constraints on the choice of active vs. passive.

The move to restrict rather than expand the term 'meaning' is consistent with sociolinguistic studies which obtain objective evidence of the stylistic and social differentiation of linguistic forms. At the same time, we must not conclude that all linguistic variation subserves the social functions of identifying the social position of speakers and adjusting to the social position of the listeners and audience.[3] Variable elements are found at all levels of linguistic structure. Some variation is the result of articulatory constraints on grammatical processes; some reflects a variable recognition of grammatical boundaries; some appears to be the residue of historical processes which persists long after the social conditions that gave rise to it have disappeared.

There is no reason to confine the study of variation to alternative ways of saying the same thing, although this kind of variable has been the chief focus in the development of variable rules. The *Heidelbergforschungsprojekt* has made great strides in the probabilistic weighting of context-free phrase structure rules by speakers of pidgin German (1976) and these are essentially meaningful choices. Studies of the development of phrase structure in the early stages of language learning (Bloom, 1970) inevitably involve variation of a meaningful kind. However, it is clear that the sharpest analytical conclusions on the conditioning factors that constrain linguistic change and variation can be made when form varies but meaning is constant, rather than when both are varying together. Theoretically, it should be possible to draw equal profit from cases where a single form is used with several meanings. But the possibility of accurate measurement is less immediate with semantic variation. We obviously have a much better chance of getting intersubjective agreement in identifying formal variants than semantic variants.

We therefore approach the passive with an eye to a bold simplification of the problems of meaning. We will treat active and passive as truth-

[3] See Lavandera (1978) and the response of Labov (1978).

conditionally equivalent and used on the whole to refer to the same states of affairs.

Earlier reports of social differences in the use of active and passive. The only previous data we have on the use of the passive in spontaneous speech is the result of its use as one of the defining features of Basil Bernstein's 'elaborated code'. A high proportion of 'passive verbs' out of all finite verbs is said to be characteristic of the elaborated code available to middle-class speakers, as opposed to 'restricted codes' which show lower percentages.[4] In Bernstein's latest presentation (1971) a comparison of two middle-class groups and three working-class groups shows that more passives are used by the former at a significance level of beyond 0.02 (p. 101).

Lawton (1968) gives a number of other results on the passive that indicate that there is more variation in this indicator than others used by these researchers. There was no significant difference in the use of passive verbs by 12-year-olds but there was a difference among 15-year-olds (5 per cent for the working class, 12 per cent for the middle class – p. 125).

Bernstein's interpretation of the use of the passive by middle-class speakers is that this is a prominent indication of the fact that these speakers do more syntactic planning than working class speakers. There is no linguistic discussion accompanying these studies that distinguishes various types of passives, or more importantly, isolates the syntactic choice between active and passive.

The notion that the passive involves more syntactic planning was perhaps not as naïve when it was first formulated as it appears today. The general consensus that has emerged from psycholinguistic experimentation is that the production and perception of sentences cannot easily be coupled with the formal representations of grammar now available (Fodor *et al.*, 1974); it is even less likely that we can make correlations with a planning process that precedes the production. But in any case, the Bernstein data does not represent a controlled study of a syntactic choice: the use of the passive is not compared to the cases where the choice might have been made but was not.

If we consider all possible finite verbs, a large number will be actives with two arguments. In spontaneous speech, the proportions of passives with two-argument predicates is vanishingly small in all data examined so far (see below). The vast majority of passive constructions has only one argument, that is, no agent is present. The comparison of passive verbs to all finite verbs is therefore a comparison of sentences with one argument to sentences with one, two and three arguments. Differences in the proportion of passives used by speakers may therefore reflect what is being talked about and the amount of detail being provided more than any syntactic choice.

---

[4] The total amount of speech studied for each group was 1800 words. This is a very small body of speech compared to those used in sociolinguistic analysis.

The Bernstein-type studies are also characterized by the formal nature of the experimental setting. In one study, groups were assembled to discuss the abolition of capital punishment (Bernstein, 1971).[5] There is evidence in other studies of a differential reaction of speakers from various social backgrounds to experimental situations of this sort (Labov, 1969b). Van den Broeck's study of Maaseik, Belgium (1977), for example, showed that middle-class speakers used more passives than working-class speakers in formal style, but when the social context was switched to one that favoured the vernacular, the relation of the two groups was reversed in this respect.

The comparison is close enough to allow us to conclude that the experimental techniques of Bernstein do not reflect differences in the availability of particular constructions to his subjects, but rather different modes of repression of a linguistic competence that may emerge freely in other contexts.

Another interesting aspect of Bernstein's portrait of the 'restricted code' is the critical focus on the pronoun *they* without a specific referent:

> Inasmuch as referents are not finely differentiated then the global term 'they' will be adopted as a general label. The non-specificity implied by 'they' is a function of the lack of differentiation and the subsequent concretizing of experience which characterizes a restricted code as a whole. (1971: 110)

There is no connection drawn here between the use of 'they' and the passive, though as it will appear below, the use of such generalized pronouns is a major alternative to the agentless passive. It seems clear that the formal context in Bernstein's experiments did lead to a greater use of the passive by middle-class speakers but he did not realize that the working-class speakers' use of generalized pronouns was simply their alternative way of saying the same thing. The absence of a linguistic analysis among writers with an educational orientation often leads to attribution of conceptual limitations to non-standard speakers.

Our own analysis will begin with a description of the passive as a linguistic variable and a description of our sample (Section 1). We will then present a quantitative analysis of the external sociolinguistic constraints on the passive (Section 2.1) before proceeding to our primary topic, internal constraints on the choice of active or passive (Section 2.2).

1. Definition of the variable

In order to study any variable phenomenon systematically, it is necessary to define the basic alternation: what is varying with what? At first glance, this seems a simple matter for the passive: passive alternates with active. *The*

[5] Recognizing the fact that working-class subjects may be ill at ease, Bernstein held several practice sessions before the recorded session.

neighbourhood was settled by Germans alternates with *Germans settled the neighbourhood.* But such agent passives are rare events. In spontaneous conversations that we have examined, they make up only 1 to 2 per cent of the total number of passivizable transitive verbs with subject and object noun phrases. On the other hand, passives are not as rare as these figures would imply. They are quite common in conversation and spontaneous speech, but the great majority of them are agentless passives. Early in this study, the decision was made to focus on the agentless passive as the primary topic for variable analysis.[6]

The problem of defining the alternant of agentless passives seems quite obscure at first. Discussions of the English passive usually do not present a syntactic option if the subject is missing.[7] As we examine colloquial spoken English, however, we find a very rich alternation of truth-conditionally equivalent forms:

> Cause we have boundaries in this school. Like out at like, the w – like you know, Lower Merion's allowed to smoke in the halls 'n' do all that crap, right? Over here, if th – I don't care if they never allow you to smoke in the halls.
>
> [A1061, 16, W, M, WC][8]

In this passive, the first sentence with *allow* might have taken the form, *they allow Lower Merion to smoke in the halls*, and the second, *I don't care if you're never allowed to smoke in the halls.* The pronoun *they* is not a purely grammatical formative (i.e. a dummy element) as in the case of *there* in *There's a man at the door*, or *it* in *It's a shame he does that. They* has a limited semantic value in that it seems to exclude the first person and perhaps the second singular as well. The choice of a particular generalized pronoun is nearly automatic in many contexts. We were able to make one systematic small-scale observation of this when thieves broke into a locked closet in our building that was used by the alumni relations staff for storing liquor. Weiner took the occasion to question people in the vicinity who were milling around the broken door. In response to 'What happened?' she received five agentless passives, such as 'The liquor closet got broken into' and five active responses of the form 'They broke into the liquor closet.' Three of the nonspecific subjects were *they* and two were *somebody*. Note that there is no information

---

[6] Langacker and Munro come to a similar decision in their analysis of the passive (1975). In Horgan's study of acquisition of the full passive (1978), it was necessary to resort to experimental techniques to elicit the forms needed, since so few passives with agents occurred in spontaneous speech. She cites Harwood (1959), who found no instances of full passives in over 12,000 utterances of 5-year-olds. This is as true of adults as of children. Brown (1973) found no full passives in 2100 utterances of parents studied.

[7] Excluding, of course, the verbs where ergative-type constructions are available as in *The door opened.*

[8] Quotations are identified by tape number, age, ethnicity (W = white, B = black), sex (M/F) and social class (MC = middle class, WC = working class).

here on singular vs. plural, but merely an exclusion of the speaker from the class of possible referents.

We can therefore include *they* with the class of [−definite] pronouns *somebody*, *someone*, *people*, etc. But a further differentiation must be made. A [−definite] pronoun indicates that a referent is not known to the hearer. Yet given this situation, there are still two distinct possibilities in regard to the *speaker's* knowledge: the speaker may have a particular referent in mind ('a particular somebody') or may not ('nobody in particular'); this applies equally well to all of the [−definite] pronouns. The dimension [±definite] refers to common assumptions about hearer's knowledge of the referent; we will use the feature [±specific] to refer to the second dimension that bears on the speaker's knowledge.

Given a [−definite] pronoun, [+specific] indicates that the speaker has knowledge of a particular referent in mind; [−specific] that he does not. The analytic task is therefore to examine each [−definite] pronoun to see if it is [+specific] or [−specific]. The 'semantically empty' pronouns that are the subjects of agentless passives are [−definite, −specific]. Pronouns such as *they*, *you*, *we* are [+definite] if they have a specific referent, but are otherwise [−definite, −specific]. *You* differs from *they* in that it more easily allows the possibility of first and second persons and is the closest equivalent to the dummy subjects German *man*, formal French *on*, and formal English *one*.

I think you expect a lot of your children.

[A907, 41, W, F, LWC]

In rare cases, *he* or *we* or perhaps even *I* can be used without specific reference, as can any personal pronoun.

The category of [−specific] noun phrases is usually referred to as general or impersonal. The major equivalent in English of the impersonal *man* or *on* is the formal *one*, but in our study of spontaneous speech we did not find a single example of the impersonal subject *one*.[9] The [−specific] nouns formed with *some-*, *any-*, and *every-*, along with *people*, have been commonly recognized as possible subjects of agentless passives, but of the 961 such sentences we found, there were only 69 of these, or 7 per cent of the total. *You* and *they* accounted for 79 per cent, and this was remarkably uniform across ages and social groups.[10] Because impersonal *you* and *they* are characteristic of colloquial speech, their syntactic rôle as subjects of agentless passives has been largely ignored in the past and even stigmatized as the

[9] This underlines the fact that the interviews we are studying are closer to colloquial speech than other types of speech that might be analysed. Since it will appear that passive sentences are associated with more formal contexts, a higher percentage of passive may appear in such formal productions.

[10] One group conversation showed a relatively high frequency of impersonal *we* (24 per cent) but most groups adhered closely to a norm of 85 per cent for *you* and *they*. There are sizeable differences in the use of *you* vs. *they*, apparently determined by factors outside of the scope of this study.

improper use of pronouns without referents.[11] Yet they are obviously the major choice for the active alternant of agentless passives. We will refer to the entire category of pronouns used with the feature [−specific] as GENERALIZED PRONOUNS and to [−specific] *you*, *they*, etc. as GENERALIZED PERSONAL PRONOUNS.

In discussions of agentless passives in the framework of the 'extended standard theory', it has been pointed out that the lexical form *somebody* cannot be considered the underlying subject form, since there are cases where actives with *somebody* differ in meaning from the passives. This objection will of course be resolved if we are able to distinguish consistently between [+specific] and [−specific] uses of *somebody*. Another type of confusion arises when using generalized forms which tend to exclude some possibilities of personal responsibility – *they* in particular. The most obvious cases involve negative acts that the speaker might want to deny, like stealing, and here the use of *They broke in*, or *Somebody broke in* might well seem to exclude the speaker as a possible referent. It is not so clear with neutral acts, as in *They always made their wine* [59, Irish, W, M, WC]. The use of *they* does not exclude first person participation here necessarily, or at any rate no more than using *people* or *some people*. Conversely, the use of the passive in *The closet was broken into* might be taken to suggest the lack of the speaker's participation just as strongly as *They broke into the closet*.

Because the generalized pronouns are not formally recognized like the impersonal pronouns of French and German, there is a tendency for their semantic contributions to be exaggerated in introspective reactions. On the other hand, educationally orientated writers, drawing on different kinds of intuitions, tend to recognize and stigmatize the absence of referential content, as noted above. However, close study of their use in spontaneous speech reveals little evidence for the influence of semantic distinctions in the choice of active or passive, and on the other side, little support for the notion that generalized pronouns are characteristic of uneducated or uncultivated speech. The choice of agentless passive vs. active with generalized subject pronoun appears to be used by all speakers of English as two alternative ways of saying the same thing. For brevity, we will refer to this choice as agentless passive vs. generalized active, or in this paper, passive vs. active. Our use of the term 'agentless' does not refer to sentences that could not have an agent like *They went home*, since no passive choice is possible here, but to sentences and contexts where the existence of an agent is implied but no information on his or her identity is expressed.

We are now able to specify where the variable occurs and where it does

---

[11] As noted above, Bernstein and similar writers have argued that the use of 'exophoric' pronouns like *they* is a consequence of the lack of differentiation in the 'restricted code', part of the pattern which 'inhibits generalizing ability at the higher ranges' (1971: 81). Closer study of the full range of impersonal pronouns might have produced a somewhat different result in their analysis.

not. A first approximation to the overall envelope of these possibilities is: all constructions without agents which may be analysed as transitive verbs with realized objects in their underlying structure. In such a construction, the occurrence of the passive variable will lead to an agentless passive; the non-occurence to an active sentence with a [−specific] subject.

An empirical study of the choice between active and passive must remain independent of any current formalism for representing the productive relationship of active and passive. As in the study of contraction and deletion of the copula (Labov, 1969*a*), we can recognize clearly the set of variants, but cannot say where in the grammar this variation is to be located. Formal arguments are multiform and indecisive. An early transformational model (Chomsky, 1957) would suggest a variable passive transformation; a later version (Chomsky, 1965) would require setting constraints on the phrase structure rule that brackets passive with manner adverbial. More recent attacks on the passive transformation leave the relationship between active and passive a matter of semantic interpretation (Bresnan, 1978), but it is not yet clear to us how such approaches can be used to incorporate within the grammar the choice of active vs. passive that we have described here.

Avoiding the instability of these formalisms, we prefer to move towards a theory with more solid foundations, which will have the capacity to incorporate the unlimited data base drawn from the use of language in everyday life. The passive variable as we are using the term here is comparable to the phonological variables first introduced in sociolinguistic studies (Labov, 1966) that were independent of the selection of a base form. Given a mutually exclusive choice of two possible ways of saying the same thing – active or passive – what are the factors that determine this choice, in so far as it is determined? The answers will illuminate our understanding of the active and passive variants. Knowledge of their cognitive and social significance, if any, and the kinds of linguistic pressures they respond to will lead towards a more confident assignment of the variation to a specific place in the grammar. A variable syntactic rule could be written, but a choice among the various alternatives cannot be intelligently made without a number of parallel studies of related rules of actives with agents specified, left dislocations, predicate adjective constructions, and so on. Here our first task is to report the decisions already made in defining and delimiting the variable at hand.

One fundamental problem in the analysis of spontaneous speech is to discriminate transitive verbs from intransitive or other verb types that do not take direct objects. Following many other analyses of *have* (e.g. Bach, 1967), we would exclude this from the set of transitive verbs, along with the obvious cases of verbs of measure such as *cost* and *weigh*, and verbs of symmetry such as *marry* and *resemble* whose subjects are semantically symmetrical with their objects and do not qualify as agents or patients.

We use the word 'transitive' here in a broad sense to include verb plus particle or even verb plus preposition when the clauses were judged to have a passive by the coder. For example, *They looked at the room* ↔ *The room*

https://doi.org/10.1017/S0022226700007441 Published online by Cambridge University Press

was looked at. The boundary between transitive verb and intransitive verb plus preposition remains as an area of vagueness to be investigated by empirical means. The proportion of such cases that were decided on intuitive grounds was approximately 15% of the total.

Clauses containing verbs with sentential objects were also systematically excluded. While a sentence like *They say that times are hard*, with generalized *they*, could conceivably have the alternate *That times are hard is said* (*to be the case*), we found in accordance with our intuitions that extraposition was categorical when the passive alternant was produced, and extraposed sentences like *It is said that times are hard* involve changes of surface structure that are incompatible with the constraints to be considered. Our examination of internal constraints on the passive will give particular attention to the consequences of placing subject or object in preverbal or postverbal position in relation to parallel or non-parallel placement of coreferents in preceding sentences. Once extraposition has applied, the underlying object of the agentless passive is no longer in initial, preverbal position and the effect of such parallel structure is eliminated.

We also excluded the 'quasi-modals' *want to, begin to, stop*, etc., in clauses with EQUI-NP deletion applied, such as *They began to bother Mary*, though *Mary began to be bothered* is quite acceptable. The change of subjects for both verbs led to a higher probability of semantic differentiation.[12]

It was also necessary to discriminate forms with *-ed* or *-en* which were clearly passives from homonymous forms which were adjectives. There are a number of ways to discriminate adjectives from passive participles, most obviously, their compatibility with adverbial suffixes such as *-ly*; *happily* but not *instructedly*. But such potential compatibility does not prove that a given form is an adjective and not a homonymous past participle. Thus *distracted* does accept *-ly* but it may not be a predicate adjective in *He was distracted by this*. The intuitive criterion which we applied in practice appears to be an effective test of verbal status: if the subject of the agentless passive can be transformed into an active with a generalized pronominal subject, it is included in the variable. We exclude from the definition of the variable those *-ed* or *-en* forms which do not correspond to an active verb with a generalized *they* or *you*. Conversely, if a pronominal subject can be transformed into an agentless passive without radical change of meaning, it is included in the analysis.

We have stressed that the quantitative analysis of this syntactic variable

[12] A reader for this journal points out that there is a potential ambiguity in *Mary began to be bothered* that is not found in *They began to bother Mary*. In the first sentence, *began* may be understood as taking a sentential subject, 'It began to bother Mary', but this is not possible in the second. Though we excluded this case for other reasons, this is one of a number of cases where the generalized pronoun does not alternate freely with the agentless passive. Whenever the subject understood is a single situation that might be referred to by *it*, *they* is not appropriate. As noted above, the test for the variable excluded such cases, since the subject of the passive could not be transformed into an active with a generalized pronominal subject without a radical change in meaning.

is not conceptually different from phonological analysis. The recognition of a phonological zero such as *I pass' here yesterday*, or even *I went pas' him* is not a mechanical process. The whole sentence (and discourse) must be understood first, and the appropriateness of a past or a present meaning must be decided subsequent to semantic processing. As in the case of the passive, the analysts must use their full knowledge of the language to recognize the occurrence of the variable. A formal statement of this procedure would be dependent on a completed formal analysis of the grammar as a whole.

The definition of the variable included in the passive alternant includes both the *get* and the *be* forms. Though there is a clear semantic difference between *John got arrested to test the law* and *John was arrested to test the law* (R. Lakoff, 1971), experiments conducted by Labov show that the basic sentences without purpose clauses are interpreted as semantically equivalent (1975). We excluded *get* where it had a causative function. The remaining *get* sentences were tabulated separately, but included with the *be* forms as instances of application of the variable rule. The selection of the passive as the output of the variable rule has no substantive significance at this point, and the same results would be obtained if the generalized active forms had been considered the output.

We began our analysis with a study of 21 speakers from working-class white neighbourhoods in Philadelphia. The agentless sentences in their interviews had already been extracted with considerable context in an earlier exploratory study.[13] These data were re-coded within our current framework to yield 825 tokens. To this initial group we added two of the central figures from our studies of working-class Philadelphia neighbourhoods, and 11 speakers from two suburban middle class neighbourhoods, giving 482 more items.[14] We then decided to add a group of speakers from a radically different dialect in order to assess the range of differences in the use of agentless passives among English dialects. We studied ten members of the Jets, a black adolescent group from Harlem (Labov *et al.*, 1968), extracting all agentless sentences from group sessions as well as individual interviews, yielding an additional 203 tokens.

The variable rule programme of Cedergren and Sankoff was used to analyse the influence of a number of external and internal constraints on the choice of active or passive in the 1489 agentless sentences. The version of the programme used here is Varbrul II (Sankoff & Labov, 1979), where the various contributions to the output probability in any given case are represented as

$$\frac{P}{1-P} = \frac{p_0}{(1-p_0)}\frac{p_1}{(1-p_1)}\frac{p_2}{(1-p_2)}\cdots\frac{p_n}{(1-p_n)},$$

[13] Interviews and initial coding were done by members of a class on the study of the speech community, Linguistics 560, at the University of Pennsylvania, whose contribution we gratefully acknowledge.

[14] The working-class interviews were carried out by Anne Bower, and the middle-class interviews by Arvilla Payne, of the project on Linguistic Change and Variation.

where P is the over-all probability of the rule applying, p_0 is an input probability, and p_1 through p_n are the contributions to the over-all probability from environmental factors. If a factor has no influence on the rule, it will show a value of 0.5 (since $0.5/(1-0.5) = 1$). Values over 0.5 favour the application of the rule and values below 0.5 disfavour it.

2. External constraints on the choice of active or passive

As mentioned above, the generalized pronouns are characteristic of colloquial English, and it has long been recognized that the excessive use of the passive is a mark of formal written style, in scientific as well as literary writing. The stylistic analysis of our materials followed the distinction between casual and careful speech developed in previous studies (Labov, 1966, 1972: ch. 3; Cedergren, 1973) but with the particular coding developed in the study of the Philadelphia speech community.[15] The results showed 639 agentless sentences in careful speech and 850 in casual. The over-all percentages showed the favouring of active over passive in the expected direction, with 40 per cent passive in careful speech, though the difference is not a large one. Table 1 adds the variable rule probabilities, taken from an analysis including 22 environmental factors in seven groups. The probabilistic weights for style reflect the percentages: 0.54 for careful speech and 0.46 for casual.

Style	Tokens	Passive (%)	VARBRUL weights
Careful	639	40	0.54
Casual	850	32	0.46

Table 1

Effect of style on the choice of passives in agentless sentences

Small differences of this type can be related to standard statistical measures of confidence by comparing the total likelihood of a variable rule analysis that distinguishes casual from careful speech with a second analysis that ignores this factor. The difference in the sums of the log likelihoods for each cell, for each analysis multiplied by -2, is equal to chi-square, with degrees

[15] The coding system used here differs from that first developed in Labov 1966 in that channel cues are not utilized. A series of discrete decisions mark as contexts for casual speech personal narratives, discussions of kids' games, group interaction and tangential movements by the speaker; responses to questions, oratorical ('soapbox') style, and discussions of language are coded for careful speech, as well as the general unmarked body of consultative conversation.

of freedom equal to the difference in the degrees of freedom of the two analyses.

$$\chi^2 = -2\,(\mathrm{SUM}(\mathrm{LL}_1) - \mathrm{SUM}(\mathrm{LL}_2)).$$

Thus chi-square for this stylistic result is 5.6, which is significant at the 0.02 level for one degree of freedom. The same stylistic coding shows large and regular differences with such sociolinguistic variables as (ING), (DH) and negative concord in every subsection of our Philadelphia sample. By comparison, the choice of active and passive is not an important stylistic factor in spontaneous speech, though we have no doubt that more formal language will show a stylistic preference for the passive.

The choice of active or passive is a well established variable in English, and if it had strong social significance, we would expect that the sex of the speaker would play a major rôle. In all previous studies of stable variables, it has been found that women use more of a prestige form and less of a stigmatized form, particularly in careful speech (Labov, 1966; Shuy, Wolfram & Riley, 1966; Trudgill, 1972). When we look at the over-all percentages, women do use a small percentage more of passives than men, but the multivariate analysis shows clearly that this is the result of other distributional factors. Table 2 shows the somewhat surprising fact that sex has no effect at all on the variable, with both men and women at exactly 0.50. The difference in log likelihood, when we remove sex as a factor, is precisely nothing – 0.00.

	Tokens	Passive (%)	VARBRUL weights
25 males	759	32	0.50
19 females	730	39	0.50

Table 2

Effect of sex on the choice of passives in agentless sentences

We then turn to social class, where previous reports (particularly those of Bernstein and his students (Lawton, 1968: 109, 118, 125)) might lead one to expect to find that our working-class speakers use fewer passives than middle-class speakers. This is not the case. In Table 3, both the raw percentages and the variable rule programme output indicate that the white working-class Philadelphians use the passive alternant more than the middle-class whites at p_i of 0.58, while the black adolescents from New York are most similar to the middle-class groups. When we examine the log likelihoods as outlined above, it appears that both of these differences are significant at the 0.01 level: working-class speakers use the passive significantly more than

https://doi.org/10.1017/S0022226700007441 Published online by Cambridge University Press

	Tokens	Passive (%)	VARBRUL weights
23 white working-class Philadelphians	982	40	0.58
11 white middle-class Philadelphians	294	27	0.46
10 black adolescent New Yorkers	213	27	0.45

Table 3

Effect of class and ethnicity on the choice of passives in agentless sentences

middle class. When we collapse white working-class and middle-class speakers (Table 10, Run 6), the chi-square is 4.8; when we collapse white and black working class speakers (Table 10, Run 5), the chi-square is 9.7. Though these significant effects are only moderate in size, they reverse the expectations created in the literature reviewed above. A further understanding of the result can be obtained from a comparison with Van den Broek's study of class differences in Belgium (see below).

We can also examine age distributions. Though we have no reason to suspect change in progress, there may be a difference between the 19 teenagers in our sample and the 23 adults. Table 4 shows that the adults do favour the

	Tokens	Passive (%)	VARBRUL weights
19 adolescents	418	28	0.46
23 adults	1071	38	0.54

Table 4

Effect of age on the choice of passives in agentless sentences

passive by the same differences in p_i scores that we saw in the style group: 0.54 for adults as against 0.46 for adolescents. But here the effect is not significant even at the 0.05 level: log likelihood difference when age is removed (Table 10, Run 4), is only 2.1.

Though age and social class are noticeable effects, they do not indicate that external factors have a sizeable influence on the choice of active or passive

https://doi.org/10.1017/S0022226700007441 Published online by Cambridge University Press

in agentless sentences. By comparison, such stable sociolinguistic variables as (DH) or (ING) show massive external p_i (Labov, 1966; Labov *et al.*, 1980). Whatever the passive is, it does not appear to be a prominent sociolinguistic variable.

There is, however, one area of social variation connected with the passive which is quite striking: the choice of auxiliary. Table 5 shows the distribution of *get* vs. *be* for adults and teenagers, male and female. Adults show a preponderant use of *be*, as do female teenagers to a lesser extent; male

		Auxiliary	
Sex/age group	Tokens	Be* (%)	Get (%)
Adults			
Female	265	77	20
Male	170	78	20
Adolescents			
Female	30	63	37
Male (total)	87	32	66
Black	56	25	75
White	31	48	52

* The remaining percentages are for *have* passives with numbers too small to be considered for social distribution.

Table 5

Distribution of ⟨get⟩ and ⟨be⟩ passives by sex and age

teenagers are significantly different from all other groups in their heavier use of *get*, and this tendency is stronger among blacks than whites. These results parallel those of Feagin in her study of white working-class adults and adolescents in Anniston, Alabama (1979). In Anniston, we find a shift from less than 30 per cent *get* for adults to close to 80 per cent for adolescents (p. 97). Both teenage boys and girls showed this high use of *get*. A shift to the *get* passive appears to be one of the most active grammatical changes taking place in English; and at least in the North, it seems to be also a stigmatized sociolinguistic variant which is used more by males than females.

3. Internal constraints

Our attention is then drawn to internal factors that may help us to gain a better understanding of the linguistic significance of this massive variation.

The concept of GIVEN vs. NEW has been put forward as a major explanatory device for the passive as well as other left and right movements of major constituents. It is generally considered that in the unmarked case, GIVEN information precedes NEW information:

> In the unmarked case the new is, or includes, the final lexical item, so that the unmarked sequence, excluding anaphoric elements, is given preceding new; but the focus can appear at any point in the information unit. The constituent specified as new is that which the speaker marks out for interpretation as non-derivable information, either cumulative to or contrastive with what has preceded; the given is offered as recoverable anaphorically or situationally. (Halliday, 1967: 211)

and in the passive, the logical object brought to the early subject position represents GIVEN information:

> . . . if the verb of a sentence is an action-process, its patient noun root will convey new information and its agent noun root old information. (Chafe, 1970: 219)

Neither of the writers cited here consider the choice between agentless passive and generalized active. But a choice is implied in their consideration of the 'semantic functions' of the passive. Chafe presents the passive as a way of achieving two functions simultaneously. One is to allow the verb to appear without an agent or an experiencer.

> The other function of the passive inflection is to change the order of priorities for the distribution of new information. (1970: 220)

Given the possibility we are considering here of two ways of presenting transitive verbs without agents, it is clear that Chafe's analysis would show that the choice of agentless passive vs. generalized active is determined by the GIVEN vs. NEW status of the two referents: the action or experience referred to by the verb on the one hand and the referent of the patient noun on the other.

Experimental evidence. There has been a considerable amount of experimental work in search of objective correlates of the subtle differences between active and passive sentences. Psycholinguists have not thought in terms of GIVEN vs. NEW but rather in terms of a wide variety of cognitive approaches such as 'conceptual focus' (Tannenbaum & Williams, 1968), 'salience' and 'focus of attention' (Turner & Rommetveit, 1968), or 'importance' (Johnson-Laird, 1968). Most of the experimental approaches have contrasted active and passive sentences with agents expressed. Johnson-Laird (1968) showed that sentences such as 'Blue is followed by Red' tended to attribute greater importance to their surface subjects than sentences such as 'Red follows Blue', as reflected by association of the passive sentences with larger blue areas in coloured rectangles. Turner & Rommetveit (1968) showed that when

children were cued by pictures to remember a sentence, there were more correct recalls when the picture focused attention on the surface subject of the active or passive, but when the picture focused attention on the object of the sentence, there was a significant tendency to reverse the voice of the sentence.

Tannenbaum & Williams (1968) focused attention on subject or object by a device which is somewhat closer to the effects observed in spontaneous speech. First-year junior high-school students were presented with preamble texts of six sentences with the noun phrase of conceptual focus in consistently active or passive position, and they were asked then to form an active or a passive sentence describing a picture in which that referent was seen as the agent or patient of an action. Latency times for the completion of active sentences were always less than for passives; but the difference between actives and passives was maximum for cases where the preceding text focused on the logical subject, and greatly reduced when the text focused on the logical object. Thus a preceding string of six references to a patient facilitated the formation of passive sentences with that referent in initial position. Tannenbaum & Williams also found a smaller effect of passive vs. active form of the sentences in the preamble, when the passive preambles reduced the over-all latency of passive productions, but not significantly. The Tannenbaum & Williams experiment appears to support the predominance of the GIVEN vs. NEW effect for the active–passive contrast with agents expressed.

Recent discussions of GIVEN vs. NEW have taken on an increasingly subjective character. Chafe (1974) presents GIVEN as that which can reasonably be assumed to exist in the consciousness of the addressee at some point in a discourse. Information may then not be new in the technical sense that the information was objectively absent. Chafe argues that it can be present in a speaker's memory but not be in his consciousness at a certain time (p. 112). He points out that the duration of any item's GIVEN status may be assured for only one sentence (p. 129). Kuno (1972) deals with the opposition of 'old, predictable information' and 'new, unpredictable information', but he does not indicate any confidence in objective criteria for this distinction.

In proceeding with an objective study of variation in the spontaneous use of passive and generalized active, we will naturally want to consider the ways in which such subjective factors might interfere with the validity of an objective coding procedure. First there is the matter of how far back we would want to search for evidence of GIVEN status. If the referent of a given sentence occurred in an immediately preceding sentence, it can reasonably be assumed that it is available to a speaker as a 'given'. But if we were to confine our attention to only this sentence we would be losing a great many GIVENS in the sentences immediately preceding. To search through an entire conversation would assuredly locate many elements as GIVEN which are not immediately in the foreground of the speaker's attention. A decision to search for a fixed

number of preceding sentences is necessarily arbitrary, but the validity of that decision can be examined in a multivariate analysis by examining successively the effect of being GIVEN in sentences successively further removed. It seems clear that GIVEN status is not discrete but gradient, and we would expect to find that the effect is strongest for the immediately preceding sentence and diminishes gradually as we recede into the past of the preceding discourse. Any valid analysis should be able to demonstrate an effective vanishing point where the inclusion of all preceding sentences would make no difference at all.

The other subjective effect is that the element may not be 'given' at all in an objective sense, i.e. not occur in the surface structure, but be present in the speaker's consciousness as part of a long chain of inference from the surface structure. We can minimize this effect by not limiting ourselves to a search for forms with identical surface structure, but accepting a very broad notion of coreferential noun phrase. Two noun phrases are coreferent if their intended referent is the same discourse entity. Some subtle forms of reference of inference will still escape us. But it seems to us that this will apply equally well to the agentless passive and the generalized active. In some senses, the referents of 'they' or 'you' may be considered part of a hidden agenda. There is no way that we can predict what the effect of such subjective factors may be on the whole, but in any case we will be cautious in the interpretation of our results and be sure to rest our case on robust and consistent results of objective coding.

An objective and relevant definition of 'sentence' would seem to be crucial. The most abstract extraction of underlying sentences would hardly do, since many of these constructs will have elided one or the other of the two constituents we are interested in and will not be relevant to the choice of active or passive. On the other hand, the notion of a sentence in surface structure defined by intonational contours is difficult to determine reliably in spontaneous speech and not necessarily related to the organization of form and meaning that we are concerned with. The 'sentences' we will consider as sites for the location of relevant noun phrases will be finite clauses. The clause will thus prove a more useful unit for our purposes.

We therefore proceed with a definition of GIVEN as follows: the logical object of an agentless clause is defined as GIVEN if any noun phrase coreferential with it was present in any one of the preceding five clauses, irrespective of the termination of speaker turns. Thus in the sentence

(5) If they asked ya to stay wash the windows, you washed the windows (56, W, F, WC).

the windows is GIVEN in the clause preceding *you washed the windows*. A GIVEN noun phrase is therefore one that has a coreferential noun phrase anywhere in the preceding five clauses. When we examine the over-all distribution of this factor among active and passive agentless clauses, our first impression

is that it has a strong influence in promoting the passive. When the logical object of the clause is GIVEN in the preceding context, it occurs in initial position as subject of a passive construction 41 per cent of the time; when it is NEW, or not GIVEN, this happens only 25 per cent of the time. The other 75 per cent of NEW noun phrases are located in post-verbal position as the objects of active verbs.

Before we conclude that GIVEN vs. NEW is the dominant factor in determining the choice of active or passive, it would be wise to consider syntactic factors which intersect with but are distinct from this one. Our early studies of the agentless passives drew attention to the effect of surface structure, whether or not the logical object referred backward to coreferential noun phrases in subject position. Such parallelism of surface structure seems to be a powerful factor in determining the choice of active or passive, and passives are favoured when the logical object moves into a position parallel with its co-referents.

The over-all distribution of this constraint proves to be somewhat larger than the GIVEN vs. NEW effect. When coreferential noun phrases were located in subject position in preceding clauses, the logical object of the agentless clause appeared in parallel subject position 58 per cent of the time, that is, the passive alternant was chosen. When this was not the case, the passive choice was realized only 29 per cent of the time.

Thus the GIVEN-NEW distinction produced a 16 per cent difference in the choice of a passive; the parallel vs. non-parallel surface structure produced a 29 per cent difference. But such across-the-board comparisons can be misleading unless the data are evenly distributed among all possibilities. The variable rule programme allows us to compare the effect of both factors in a single analysis; Table 6 indicates that the weight contributed by the choice of passive by GIVEN is 0.54 as against 0.46 for NEW; whereas parallel vs. non-parallel surface structure gives us 0.62 vs. 0.38. The chi-squares derived

	Tokens	Passive (%)	VARBRUL weights	χ^2
Given	955	41	0.54	
New	534	25	0.46	25.27***
Parallel SS	350	58	0.62	
Non-parallel SS	1139	29	0.38	71.61***

* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

Table 6

Given vs. new and parallel vs. non-parallel surface structure as constraints on the passive

https://doi.org/10.1017/S0022226700007441 Published online by Cambridge University Press

from log-likelihood comparisons are both highly significant, though the parallel structure figure is at a higher order of magnitude.

However, this comparison may not evaluate the effect of GIVEN vs. NEW in full, since a co-referential noun phrase in the immediately preceding clauses may have the strongest effect. We therefore distinguished each GIVEN by how far back we had to go to find the coreferential noun phrase which established that status:

Given [G]

0 not given at all in the preceding five clauses
1 given in the immediately preceding clause
2 not given in the preceding clause but given in the second preceding clause
3 not given in the two preceding clauses but given in the third preceding clause
4 not given in the three preceding clauses but given in the fourth or fifth preceding clause

At the same time, we extended a similar analysis to the factor group of parallel surface structure. But parallelism is disrupted by any intervening non-parallel clauses, and a more detailed analysis of this group must consider the number of consecutive parallel clauses preceding the one in question. In other words, the stronger parallel structure cases are all subsets of the basic conditions that the immediately preceding clause must contain a parallel subject.

Surface Subject [SS]

0 is not coreferential with subject of preceding clause
1 is coreferential with subject of preceding clause, but not the clause before it
2 is coreferential with subjects of two preceding clauses only but not the third preceding clause
3 is coreferential with subjects of three preceding clauses, but not the fourth preceding clause
4 is coreferential with four or more preceding clauses

Table 7 shows the result of this more detailed variable rule analysis. The GIVEN vs. NEW effects show a fairly regular gradation of weights, but in a direction opposite to our first expectation. When the logical object is given in the immediately preceding clause, there is no effect. On the contrary, this category (Given 1) has a weight of 0.39, slightly less than not GIVEN at all (Given 0). It is only when the GIVEN is located in the second clause preceding that there is a sizeable effect, of 0.58. A coreferential noun phrase located in the third or fourth preceding clause is only slightly higher, 0.60, and in the fourth preceding clause the figure falls off to 0.53.

If the effect of GIVEN status were a powerful determinant of the passive form, it would follow that the largest figure would be Given 1, and Given 2, 3 and 4 would fall off in that order. However, the significance figures show that the

	Tokens	Passive (%)	VARBRUL weights	χ^2
Given				
0	534	25	0.40	.
1	616	42	0.39	25.27***
2	146	43	0.58	18.41***
3	61	47	0.60	1.13
4	132	36	0.53	0.88
SS				
0	1139	29	0.21	.
1	214	51	0.45	71.61***
2	57	70	0.63	10.41*
3	37	67	0.67	0.87
4	42	64	0.56	0.85

* $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

Table 7
Surface structure constraints on the passive

effect of GIVEN status is concentrated almost entirely in Given 2. The chi-square opposite each figure shows the significance of the effect of separating each factor, along with the factors below it, from the factor above. Thus the effect of distinguishing all Given 1, 2, 3, 4 from Given 0 is sizeable, with a χ^2 of 25.27; but the largest part of that is based on the effect of Given 2. When Given 2, 3, 4 are separated from 1 and 0, there is a very significant effect, with χ^2 of 18.41. But there is no additional significance when Given 3 and 4 are separated from Given 2.

Thus we obtain a sizeable improvement in likelihood of the analysis by distinguishing Given 0 and 1 from cases where the coreferential noun phrase was in a second preceding or earlier clause; but there are no significant losses or gains from making further distinctions between Given 2 and Given 3, or Given 3 and Given 4.

On the other hand, if the coreferents located in the third and fourth preceding clause had NO effect in encouraging the choice of passive, we would expect that the effect of Given 2 would increase when we separated out these cases. But this does not happen: the weight contributed by a category that includes GIVEN in the 2nd, 3rd or 4th preceding clauses is 0.61; when we separate Given 3 and Given 4 from Given 2, Given 2 does not rise but remains practically the same. It follows that the effect of GIVEN in the third and fourth

https://doi.org/10.1017/S0022226700007441 Published online by Cambridge University Press

preceding clauses is not to be ignored, but has about the same effect as Given 2.

We cannot account fully for the location of the GIVEN effect in the second preceding clause, though it may well be connected with the fact that many of the immediately preceding clauses were located in the same surface sentence and the passive verb located in a sentential complement. The present analysis does not distinguish effects across clause boundaries from effects across surface sentence boundaries; the result of Table 6 clearly calls for such a distinction in further studies of the passive.

When we turn to the analysis of parallel structure, we find a different picture. The effect on the passives of SS1, only one preceding parallel clause, is considerably greater than SS0, and SS2 and SS3, with longer parallel strings, are progressively greater. The improvement peaks at SS3 and falls off somewhat at SS4. The chi-square of 71.61 indicates the total improvement in likelihood obtained by distinguishing all cases of parallel structure from cases without parallel structure. When we separate strings of two and more parallel clauses, there is a moderate improvement of likelihood, with chi-square at 10.41. But if we then separate out strings of three or more, the effect is not significant, ($\chi^2 = 0.87$). It follows that strings of four or more are also not going to show a significant effect. The Varbrul weights for strings with 2, 3 and 4 preceding parallels are all approximately in the same range, and this effect seems to show that nothing is gained in the way of explanation by examining sentences even more remote.

To make a finer comparison of GIVEN vs. NEW and parallel structure we must consider exactly what the relations of these two factor groups are. From the definitions just given, it is apparent that all the cases of parallel structure correspond to a single category in the GIVEN factor group: that is, Given 1. Since parallel structure requires a coreferential noun phrase in the immediately

	Tokens	Passive (%)	VARBRUL weights
G1(P)			
Given with parallel structure	360	58	0.54
G1(~ P)			
Given with non-parallel structure	277	23	0.36

Table 8

Given with parallel structure vs. given with non-parallel structure as constraints on the passive

https://doi.org/10.1017/S0022226700007441 Published online by Cambridge University Press

preceding clause, then every case of parallel structure falls into the first Given category. On the other hand, there are cases of coreferential noun phrases in the immediately preceding clause that are non-parallel: that is, SSo. This is obviously the case when the logical object occurs as the surface object of the preceding clause.

Figure 1 sums up the interrelations of these two factor groups: it shows a four-cell table with the intersection of GIVEN vs. NEW and parallel vs. non-parallel surface structure. The upper right quadrant is empty: if the

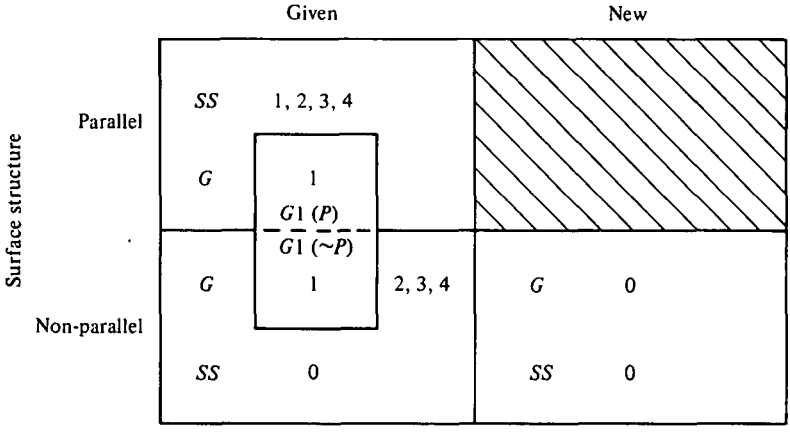

Figure 1. Interrelationship of given vs. new and parallel vs. non-parallel structures.

logical subject is not GIVEN, it cannot be parallel. In the upper left quadrant are the parallel subcategories of SS1, 2, 3, 4 which all co-occur with given status: necessarily all G1 as well. In the lower left there is only SSo, non-parallel, but here we can have Given 1, 2, 3, OR 4. The crucial category, which permits comparison of the two effects, is G1. We can examine cases of G1 WITH parallel structure (the upper half of the rectangle) or WITHOUT (the lower half). We therefore ran the variable rule analysis dividing G1 into the two subtypes shown here, G(P) and G(~ P). Within a fixed given status, we can compare the effect of parallel structure. Table 8 shows the result. First we notice that the effect of parallel structure is sizeable: Given 1 with parallel structure is 0.54, and Given 1 non-parallel is only 0.36, at about the level of the not-given category.

Figure 1 also shows that a second close comparison can be made: with this differentiation of two kinds of G1, we can make a better comparison of the effect of GIVEN status, that is, the entire lower left vs. the lower right quadrant. Table 9 shows this result. Given 1(P) is about the same, 0.56. The crucial comparison is between Go and the combined category G(~ P), which includes G1(~ P) and G2, 3, 4. The comparison of G(~ P) with Go gives us

https://doi.org/10.1017/S0022226700007441 Published online by Cambridge University Press

	Tokens	Passive (%)	VARBRUL weights
G(~ P)			
All given non-parallel	616	33	0.52
Go			
New	534	25	0.44
G1(P)			
Given parallel	360	58	0.56

Table 9

Effect of given in non-parallel sentences vs. new on choice of the passive

a strict measure of the effect of given, and it is in the expected direction: 0.52 vs. 0.44. However, this is among the smaller of the effects we have found, and it does not support the notion that given vs. new is a powerful determinant of the choice of active vs. passive agentless clauses.

The converse conclusion is that the ordering of surface syntax across clauses is the predominant linguistic influence on this choice. This result reinforces our earlier decision to treat the choice of active and passive as two alternative ways of saying 'the same thing', since it is conditioned by formal, syntactic factors far more than the influence of given vs. new. One might argue that parallel strings are also characteristic of semantically significant choices: there is a cognitively determined tendency to keep talking about the same thing.

Parallel surface structure is a 'mixed' category, involving coreference as well as surface order. One way of approaching this question is to consider what would be the effect of a purely syntactic condition which intersects with parallel surface structure: whether or not the given agentless clause was preceded by a passive one, irrespective of coreference. Table 10 shows the analysis of Run 1 which includes this factor, along with the basic divisions of given vs. new and parallel structure. These remain unchanged except for a few hundredths. But the preceding passive proves to be an independent and powerful conditioning factor: only 126 cases had a preceding passive (and this includes a preceding passive anywhere in the preceding five clauses). But this subcategory shows the highest percentage of passives: 72 per cent. The variable rule weight contributed by a preceding passive is 0.69, and the gain in likelihood produced by adding this factor to our analysis is very large: chi-square = 55. Since a preceding passive could occur with any of the combinations of given/new or parallel/non-parallel SS it is understandable that adding this variable factor does not disturb the earlier analyses.

	Full analysis	Factors merged in successive runs				
		Style	Sex	Age	Race	Class
Run no.	1	2	3	4	5	6
Input p_0	0.74	.	.	.	.	.
Given status						
0 New	0.40	.	.	.	.	.
1 In preceding	0.39	.	.	.	.	.
2 2nd preceding	0.58	.	.	.	.	.
3 3rd preceding	0.67	.	.	.	.	.
4+ 4th preceding	0.53	.	.	.	.	.
Parallel SS						
0 Not parallel	0.21	.	.	.	.	.
1 Preceding	0.45	.	.	.	.	.
2 Preceding	0.63	.	.	.	.	.
3 Preceding	0.67	.	.	.	.	.
4+ Preceding	0.56	.	.	.	.	.
Preceding passive						
Yes	0.69	.	.	.	.	.
No	0.31	.	.	.	.	.
Style						
Careful	0.54	—	.	.	.	.
Casual	0.46	—	.	.	.	.
Sex						
Male	0.50	.	—	.	.	.
Female	0.50	.	—	.	.	.
Age						
Adolescents	0.46	.	.	—	−03	−02
Adults	0.54	.	.	—	+03	+02
Class/Ethnicity						
White MC	0.46	.	.	.	−04	−03
White WC	0.58	.	.	.	#	.
Black WC	0.45	.	.	.	.	#
No. of cells	200	129	138	148	184	142
Chi square	.	5.6	0.00	2.1	4.8	9.7
$p <$	.	0.02	.	.	0.05	0.01

—, Factor group eliminated.
#, Factor merged with factor above.
$+i$, $-i$, Greater or less than Run 1 by $i/100$.
., Does not differ from Run 1 by more than 0.1.
$\chi^2 = -2(\log \text{likelihood}_1 - \log \text{likelihood}_2)$

Table 10

VARBRUL probabilities for the passive: full analysis and effects of successive mergers of external factors

https://doi.org/10.1017/S0022226700007441 Published online by Cambridge University Press

The fact that it is a powerful factor reinforces our conclusion that the choice of agentless passive is conditioned by syntactic considerations.

These findings coincide with other recent studies that indicate the limitations of the information factor in determining sentence structure. Linde (1974) did not find that given vs. new was a significant factor in determining the choice of dummy sentences in apartment lay-outs, e.g. *There is a living room on the right* vs. *A living room is on the right*. Poplack (1980) has carried out a massive variable rule analysis of the informational constraints on the deletion of final (s) in Puerto Rican Spanish. Although the plural inflection /s/ is most likely to be found in the first element of a noun phrase, its presence or absence on succeeding elements is determined by a sequential effect similar to the effect of a preceding passive in this study. /s/ tends to follow /s/, and zero tends to follow zero. An informational analysis would have predicted that a preceding /s/ would favour the presence of a following zero.

The given/new distinction has recently been re-examined by Prince, with an eye to the structure actually used in discourse (1979). She provides a finer subcategorization of the possibilities, with a three-way division of EVOKED (textually or situationally), INFERRED, and new (BRAND-NEW or UNUSED). BRAND-NEW is in turn categorized as anchored or unanchored, depending on whether the NP is linked to the rest of the discourse through another NP, as with a relative clause. It seems likely that the effect of given/new shown in our results would be sharpened if the data were re-analysed with a factor group that registered these distinctions.

Silva-Carvalan's recent variable rule analysis of the postposing of subjects in Mexican–American Spanish shows no effect of given vs. new (1977). However, she reports a powerful effect of given vs. new on the presence or absence of pronoun subjects. Given status favoured the deletion of a pronoun with a weighting of 0.69, as compared to 0.66 for the effect of parallel subject. It appears that given vs. new may be a more powerful determinant of the presence or absence of an element than the ordering of sentence constituents. It is also possible that the effect will be stronger if it is exerted on the choice of placement of two noun phrases in two-argument sentences, rather than a single noun phrase vs. verb in one-argument constructions.[16] An extension of the present analysis to the contrast of active and passive in two-argument sentences, with agents expressed, may throw more light on the status of the informational component in determining the structure of sentences in spontaneous speech.

4. THE INDEPENDENCE OF INTERNAL AND EXTERNAL CONSTRAINTS

Table 10 also shows the results of a systematic testing of each factor for significance by examining differences in log likelihood when factors are

[16] A point made by Susumu Kuno in the discussion of the oral presentation of this paper.

	Factors merged in successive runs									
	Full analysis	Given/new				Parallel SS				Pre. pass
Run no. ...	1	7	8	9	10	11	12	13	14	15
Input p_0	0.74		−04	−11	−08	−02	−07	−11	−23	−13
Given status										
0 New	0.40	+02	+05	+06	—	.	.	.	−04	.
1 In preceding	0.39	+02	+05	+15	—	.	.	.	+15	.
2 2nd preceding	0.58	+02	+03	#	—	.	.	.	−04	.
3 3rd preceding	0.60	−03	#	#	—	.	.	.	.	.
4 4th+ preceding	0.53	#	#	#	—	.	.	.	−03	.
Parallel SS										
0 Not parallel	0.21	.	.	+06	+04	+03	+07	+13	—	.
1 Preceding	0.45	.	.	−02	.	+03	+06	+21	—	.
2 Preceding	0.63	.	.	−02	.	−03	+07	#	—	.
3 Preceding	0.67	.	.	−02	.	−02	#	#	—	−03
4+ Preceding	0.56	.	.	−02	.	.	.	.	.	+02
Preceding Passive										
Yes	0.69	.	.	.	.	.	.	.	.	—
No	0.31	.	.	.	.	.	.	.	.	—
Style										
Careful	0.54	.	.	.	.	.	.	.	.	.
Casual	0.46	.	.	.	.	.	.	.	.	.
Sex										
Male	0.50	.	.	.	.	.	.	.	.	.
Female	0.50	.	.	.	.	.	.	.	.	.
Age										
Adolescents	0.46	.	.	.	.	.	.	.	.	.
Adults	0.54	.	.	.	.	.	.	.	.	.
Class/ethnicity										
White MC	0.46	.	.	.	.	.	.	.	.	.
White WC	0.58	.	.	.	.	.	.	.	.	.
Black WC	0.45	.	.	.	.	.	.	.	.	+03
No. cells	200	185	164	137	109	191	178	156	122	148
Chi square	.	0.9	1.1	18	25	0.9	0.9	10	72	55
$p <$				0.001	0.001			0.02	0.001	0.001

—, Factor group eliminated.
#, Factor merged with factor above.
$+i$, $-i$, Greater or less than Run 1 by $i/100$.
., Does not differ from Run 1 by more than 0.01.
χ^2, $-2(\log \text{likelihood}_1 - \log \text{likelihood}_2)$.

Table 11

VARBRUL analysis of the passive: full analysis and effects of successive mergers of internal factors

https://doi.org/10.1017/S0022226700007441 Published online by Cambridge University Press

removed. In addition to the full analysis, five columns show the results of successively eliminating, one at a time, the external variables: style, sex, age, race, and social class. In these columns, a dash in each member of a factor group indicates that for that run the entire factor group has been removed. The symbol # indicates that a particular factor has been merged with the one listed directly above it. When one of the other factors in a run show a difference from the original Run 1 of more than 0.01, that difference is shown; otherwise, the space is left blank.

It is immediately apparent that there are no significant changes in the internal factors as the result of eliminating an external factor: all of the perturbations occur in the class and age factor groups. Table 11 shows the corresponding eliminations of internal factors. Here there is one change in an external factor: the weight for black working class adolescents increases by 0.03 when the parallel structure group is removed entirely. All of the other changes occur in the internal factors. There are sizeable shifts in the weights of the given/new group when parallel structure factors are eliminated and vice versa.

This result indicates a substantial independence of the two sets of constraints. All sections of the population appear to treat the passive/active choice in the same way, and conversely, the same constraints are found throughout the speech community. This independence of external and internal constraints is an empirical result that is quite different from the assumption of independence of environmental constraints within the Cedergren/Sankoff programme, where it functions simply as a null hypothesis (Sankoff & Labov, 1979). The separation of the two sets of constraints confirms other indications that social factors operate primarily upon surface patterns rather than abstract syntactic alternatives. Table 5 showed considerable social differentiation of the choice of the surface formative *get* vs. *be*, in contrast to the more abstract alternation of generalized active and agentless passive.

This multivariate analysis of the passive/active alternation has disengaged several intersecting forces that strongly determine the choice of one or the other of these ways of saying the same thing. The distribution of information in discourse is not without influence, but it is a relatively minor factor compared to the more mechanical tendency to preserve parallel structure: first in the succession of passive constructions, second in the retention of the same structural position for the same referent in successive sentences. There is undoubtedly a stylistic factor operating, which would appear more strongly if we had included more formal speech and written materials. All of these conditions on the selection of active vs. passive are general features of the English language, used in much the same way by the very different sub-sections of the speech communities that we studied.

REFERENCES

Bach, E. (1967). *Have* and *be* in English syntax. *Lg* **43**. 462–485.

Baltin, M. (1977). Quantifier-negative interaction. In R. Fasold and R. Shuy (eds.), *Studies in language variation*. Washington, D.C.: Georgetown U.P.

Bernstein, B. (1971). *Class, codes and control*. New York: Schocken.

Bloom, L. (1970). *Language development: form and function in emerging grammars*. Cambridge: MIT Press.

Brown, R. (1973). *A first language*. Cambridge: Harvard University Press.

Bresnan, J. (1978). A realistic transformational grammar. In Halle, M., Bresnan, J. & Miller, G. (eds.), *Linguistic theory and psychological reality*. Cambridge: MIT Press. 1–59.

Cedergren, H. (1973). The interplay of social and linguistic factors in Panama. Unpublished Cornell University dissertation.

Chafe, W. (1970). *Meaning and the structure of language*. Chicago: University of Chicago Press.

Chafe, W. (1974). Language and consciousness. *Lg* **50**. 111–133.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.

Feagin, C. (1979). *Variation and change in Alabama English*. Washington, D.C.: Georgetown University Press.

Fodor, J., Bever, T. & Garrett, M. (1974). *The psychology of language*. New York: McGraw-Hill.

Halliday, M. A. K. (1967). Notes on theme and transitivity in English, Part 2. *JL* **3**. 199–244.

Harwood, F. (1959). Quantitative study of the speech of Australian children. *L&S* **2**. 236–270.

Horgan, D. (1978). The development of the full passive. *JChL* **5**. 65–80.

Johnson-Laird, P. (1968). The interpretation of the passive voice. *Quarterly J. of Experimental Psych.* **20**. 69–73.

Katz, J. & Postal, P. (1964). *An integrated theory of linguistic descriptions*. Cambridge: M.I.T. Press.

Klein, W. & Dittmar, N. (1979). *Developing grammars: the acquisition of German syntax by foreign workers*. New York: Springer-Verlag.

Labov, W. (1966). *The social stratification of English in New York City*. Washington. D.C.: Center for Applied Linguistics.

Labov, W. (1969a). Contraction, deletion, and inherent variability of the English cpula. *Lg* **45**. 715–762.

Labov, W. (1969b). The logic of non-standard English. *GURT* **22**. 1–44.

Labov, W. (1975). *What is a linguistic fact?* Lisse: Peter de Ridder Press.

Labov, W. (1978). Where does the sociolinguistic variable stop? A response to Beatriz Lavandera. *Working Papers in Sociolinguistics*. Austin: Southwest Educational Development Laboratory.

Labov, W., Bower, A., Dayton, E., Hindle, D., Kroch, A., Lennig, M. & Schiffrin, D. (1982). *Social determinants of sound change*. Final Report to NSF on SOC75-00245. Philadelphia: U.S. Regional Survey.

Labov, W., Cohen, P., Robins, C. & Lewis, J. (1968). *A study of the non-standard English of Negro and Puerto Rican Speakers in New York City*. Cooperative Research Report 3288. Vols. I and II. Philadelphia: U.S. Regional Survey.

Lakoff, G. (1970). *Linguistics and natural logic*. Ann Arbor: University of Michigan.

Lakoff, R. (1971). Passive resistance. *PCLS* **7**. 149–162.

Langacker, R. W. & Munro, P. (1975). Passives and their meaning. *Lg* **52**, 789–830.

Lavandera, B. (1978). Where does the sociolinguistic variable stop? *LiS* **7**. 215–238.

Lawton, D. (1968). *Social class, language and education*. London: Routledge, Kegan Paul.

Linde, C. (1974). The linguistic encoding of spatial information. Unpublished Columbia University dissertation.

McConnell-Ginet, S. (1982). Adverbs and logical form: a linguistically realistic theory. *Lg* **58**. 144–184.

Poplack, S. (1980). The notion of the plural in Puerto Rican Spanish: competing constraints on /s/ deletion. In W. Labov (ed.), *Locating language in time and space*. New York: Academic Press. 55–68.

Prince, E. (1979). On the given/new distinction. *PCLS* **15**.

Sankoff, D. & Labov, W. (1979). On the uses of variable rules. *LiS* **8**. 189–222.
Shuy, R., Wolfram, W. & Riley, W. (1966). A study of social dialects in Detroit. Final Report, Project 6-1347. Washington, D.C.: Office of Education.
Silva-Carvalan, C. (1977). A quantitative study of subject deletion and subject placement in spoken Spanish. Paper given at the Winter Meeting of the LSA, Chicago.
Tannenbaum, P. & Williams, F. (1968). Generation of active and passive sentences as a function of subject or object focus. *JVLVB* **7**. 246–250.
Trudgill, P. (1972). Sex, covert prestige and linguistic change in urban British English. *LiS* **1**. 179–195.
Turner, E. & Rommetveit, R. (1968). Focus of attention in recall of active and passive sentences. *JVLVB* **7**. 543–548.
Van den Broek, J. (1977). Class differences in syntactic complexity in the Flemish town of Maaseik. *LiS* **6**. 149–182.
Weinreich, U. (1958). Travels through semantic space. Review of Osgood, C. E., *et al.*, *The measurement of meaning*. *Word* **14**. 374–379.