Title
Form and substance in language universels

ISBN
9783110097979

Author
Hyman, Larry M

Publication Date
1984-01-31

DOI
10.1515/9783110868555.67

Peer reviewed

Form and substance in language universals*

LARRY M. HYMAN

Introduction

In a workshop such as this one, it is impossible to consider explanations for language universals without addressing the nature of explanation itself — either as it pertains to linguistics or in general. While philosophical and methodological questions have not been a major concern of mine — or probably of most linguists — we each operate with at least an implicit set of assumptions which allow us to either accept or reject something as an explanation, or an analysis of a particular corpus of data as 'explanatory'. When asked to reflect on the nature of explanation and on possible explanations of language universals, I am inclined to respond as most of my colleagues whom I have interrogated on the subject, and distinguish between 'internal' vs. 'external' explanations. If the problem to be accounted for is a syntactic one, an internal explanation will propose an account in terms of the nature of syntax itself, while an external explanation will attempt to relate the syntactic problem to phenomena outside the realm of syntax (e.g. semantics or pragmatics). Similarly, if the problem is a phonological one, an internal explanation will construct a theory of phonology to account for it, while an external explanation will seek a relation with, say, articulatory or perceptual phonetics.

There is a belief among certain linguists that only an external explanation is a true explanation. That is, a theory of syntax does not 'explain' syntax and a theory of phonology does not 'explain' phonology. Since internal explanations involve the construction of formal models, while external explanations normally do not, the internal/external dichotomy is sometimes referred to as one between formal vs. functional explanations. This opposition is useful, however, only to the extent that there is a clear distinction or break between the two kinds of explanation. Unfortunately, there is disagreement on the meaning of 'functional' as applied in this context. While everyone would agree that explanations in terms of communication and the nature of discourse are functional, it became

evident in different presentations at this workshop that explanations in terms of cognition, the nature of the brain, etc., are considered functional by some but not by other linguists. The distinction appears to be that cognitive or psycholinguistic explanations involve formal operations that the human mind can vs. cannot accommodate or 'likes' vs. 'does not like', etc., while pragmatic or sociolinguistic explanations involve (formal?) operations that a human society or individual within a society can vs. cannot accommodate or likes vs. does not like. Since psycholinguistic properties are believed to reflect the genetic makeup of man, while sociolinguistic properties reflect man's interaction with the external world, some definitions hold that only the latter truly has to do with 'function'. In the other view, any explanation which relates grammar to anything other than grammar is 'functional'.

Since the term 'functional' is thus vague, if not 'loaded', I shall not use it in this paper. Instead I shall refer to internal vs. external explanations, as defined above. It is of course no accident that internal explanations involving formal models are believed by their defenders to be related to cognition, while external explanations involving interactions between grammar and real speaker/hearers are believed by their defenders to be related to communication. It is an unfortunate and unnecessary consequence of this artificial division that the former group rarely addresses the communicative input, as the latter group has in recent years abandoned the quest for a formal model of grammar and cognitive processing. What I shall therefore attempt to demonstrate in this paper is that these views, far from being contradictory of one another, together provide a fruitful avenue for the pursuit of explanations of language universals.

1. The problem: (i) phonology

Over the past ten years or so I have been intrigued by a puzzling recurrent pattern which can be summarized as in (1).

(1) a. Language A has a [phonological, phrase-structure, transformational] rule R which produces a discrete (often obligatory) property P;
    b. Language B, on the other hand, does not have rule R, but has property P in some (often nondiscrete, often nonobligatory) less structured sense.

Let me first illustrate this pattern frequently obtaining between languages by means of two phonological examples I have addressed in previous

publications. In Hyman (1975) I considered languages which have phonological rules such as in (2).

(2) a. $V \rightarrow [+\text{nasal}] / N\_$ (perseverative assimilation)
b. $V \rightarrow [+\text{nasal}] / \_ N$ (anticipatory assimilation)

A language may have a rule which nasalizes a vowel after a nasal consonant or before a nasal consonant. A language may have both processes. It is usually quite obvious to the investigator when a language has either or both of these rules. On the other hand, it is less obvious that languages not having the phonological rules in (2) often allow for slight nasalization of vowels in the context of nasal consonants. These low-level, detail, or 'n-ary' effects are measurable and are due to the physiological properties of the speech organs. Speakers thus seem to find it more convenient or at times less difficult not to worry about synchronizing the raising and lowering of the velum with the change of vowel and consonant articulations. The result, of course, is that the velum may stay down too long, in which case we get perseverative nasalization (2a), or it may go down too soon, in which case we get anticipatory nasalization (2b). Languages having rules such as in (2) seem therefore to formally institutionalize what the vocal tract would like to do, if left to its own devices.

Or take a second phonological example, one which is of great historical interest in how languages develop tonal contrasts. It is well known that voiced obstruents have an intrinsic pitch-depressing effect on a following vowel. It is not surprising, then, that the English words in (3) tend to have the indicated pitch contours with a high-low declarative intonation:

(3) a. pin [˥˩] b. bin [˦˩]

While the pitch fall may start immediately after the release of the initial voiceless obstruent in *pin* in (3a), in (3b) the voiced obstruent in *bin* causes there to be a lowering of pitch before the high-low fall. While phoneticians do not all agree on the causes of this effect of voiced obstruents on pitch (see Ohala 1973, 1978; Hombert 1978), they do agree that the explanation lies in the operations affecting the larynx where voicing contrasts and pitch distinctions are produced. In Hyman (1977a) I talked about the three stages of historical development a language may undergo in the development of a new tonal contrast, as indicated in (4).

(4)	Stage I >	Stage II >	Stage III	
	pá [˥˥]	pá [˥˥]	pá [˥˥]	(´) = high tone
	bá [˩˥]	bă [˩˧]	pă [˧˥]	(ˇ) = rising tone
	'intrinsic'	'extrinsic'	'phonemic'	

I start with a language which has a high vs. low tonal opposition and address what happens to high-tone syllables which begin with voiced obstruents. In stage I there is a slight lowering effect caused by the initial /b/. As in the case of low-level nasalization, the pitch lowering on the vowel /a/ is an INTRINSIC by-product of a neighboring segment and not part of the phonological tone (cf. Mohr 1971). In stage II, however, the lowering effect has been exaggerated beyond the degree we would expect from universal phonetics, to such an extent that its presence must be due to a phonological rule. Or, in other words, the low part of the tone has become an EXTRINSIC part of the signal. Finally, in stage III, the voiced obstruent becomes devoiced and we get a phonemic opposition between a high tone and a low-high rising tone. The transition from stage I to stage II I term 'phonologization': an intrinsic property of the speech signal becomes part of the language-specific phonology. The transition from stage II to stage III I term 'phonemicization': a predictable phonological property introduced by rule becomes unpredictable, i.e. distinctive. The same stages are observable in the historical development of distinctive vowel nasalization, although in Hyman (1975) I argued that there may be as many as five distinguishable stages in such a development.

At this point it may not be clear why I introduced these facts as a puzzle, or why they are intriguing or even interesting to me. After all, I have simply demonstrated that languages acquire natural phonological rules, that is, rules which relate to natural phonetic processes. The puzzle in my opinion is the following, stated as the question in (5).

(5) Given that property P (e.g. vowel nasalization before/after a nasal consonant, tone lowering after a voiced obstruent) has an external (i.e. extragrammatical) origin and *raison d'être*, why doesn't P stay out of the grammar? I.e. why do languages acquire rule R?

In other words, why don't speakers just always nasalize vowels as suits their vocal tract? Why don't speakers just slightly lower the fundamental frequency of a vowel following a voiced obstruent? Why get carried away about it and make it a formal property of one's language?

An initial response might be that in order for such phonological phenomena to occur over and over again, they must represent some kind of 'simplification', or some kind of advantageous functional or formal mutation over the previous stage. But what is the nature of this putative advantage?

The question in (5) can be reformulated without reference to either property P or rule R. In more general terms I am asking why languages must have grammars, i.e. formal systems containing a syntactic component, a phonological component, etc. There has been a tendency to view

grammar as a compromise mediator between sound and meaning (see, for example, Vennemann 1972 for a statement of this traditional view). In this view, since grammar is the result of a struggle between the two components of the linguistic sign, it acquires arbitrary (i.e. noniconic) properties of its own. This view may have relevance to describing the arbitrariness of form/content relations in the bulk of lexical items in any language, but we cannot explain the recurrent properties of GRAMMAR (syntax, phonology) without assuming that grammar has 'a mind of its own'. That is, the concerns of Grammar with a capital 'G' are not derivable from extragrammatical factors. Let us, as a convenience, refer to Grammar as *form* and non-Grammar as *substance*. In the preceding discussion I have pointed out that the Grammar (i.e. the phonological component of the Grammar) has had a mind of its own: rather than leaving the phonetics to itself, the Grammar gets into its head the idea that nasalization or new tonal distinctions ought to be a part of it. What has begun as substance is now form. In other words, the true struggle is between the *laws of substance* and the *laws of form* (Grammar). Each set of laws wishes to control the raw material (substance). In our example, the vocal tract 'wishes' nasalization and pitch lowering to be intrinsic, while the Grammar wishes both to be extrinsic. Once extrinsic, the substance, which is now form, must conform with the laws of Grammar. In the example, once the phonetic substance has become phonological, it may become 'morphologized', 'lexicalized', and eventually dropped from individual grammars as new phonologization processes are introduced.

2. The problem: (ii) syntax

In the preceding section I have attempted to justify what I think is obvious to all linguists; namely, that there are independent principles or laws of Grammar which constitute a FORCE ever-present to appropriate substance for its own purposes. I hope that it is clear that one does not 'explain' the presence of rules such as in (2) by merely pointing to phonetic facts at an intrinsic level. Similarly, one does not explain the phonologization and phonemicization processes in (4) in strictly phonetic terms. The substance (phonetics) provides the input for these processes but cannot explain why formal rules come into being.

Phonologists are well aware of the so-called naturalness of phonologization processes. What is less well understood is that the same formalization of substance occurs as frequently in the syntactic component. In this particular case the substance is pragmatics, i.e. intrinsic properties of communication. When pragmatic factors become part of a grammar, the

result is syntax and morphology. Let us consider an example. It is claimed for certain languages, e.g. Samoan, that all subject noun phrasès must be definite. The case parallel to our phonological examples is one where a language has an overt definite article and either an overt indefinite article or ∅ marking indefiniteness on common nouns. In this language the grammar will have to ensure that the determiner of the subject NP not be filled by an indefinite article (or be null, if a common noun). As seen in (6), the property P is the [+definite] specification that the subject NP must receive by rule, and the language in question is like the language A seen earlier in (1).

(6) a. Language A: NP → [+definite] / [$_S$
b. Language B: [$_S$ NP 'tends' to be [+definite], statistically

Language B, on the other hand, is one where, statistically speaking, the subject NP tends to be or is almost always [+definite]. Or, statistically, the proportion of [+definite] subject NPs is greater than, say, the proportion of [+definite] direct-object NPs. The frequency counts done by Givón and others indicate that this statistical bias holds across languages, wherever it can be measured. So, definite marking on subject NPs is an intrinsic byproduct of being 'subject' in the exact same way as nasalization in language B is an intrinsic byproduct of being adjacent to a nasal consonant. In this latter case we are still in the realm of substance as far as the distribution of definite markers is concerned.

The question of why subjects should tend to be definite arises, just as the question arises as to why pitch should be depressed after a voiced obstruent. I am sure there is a natural external explanation as to why speakers organize discourse so that the subject position receives a greater preponderance of more identifiable or determined referents than certain other positions. However, it is not any more important to our understanding of the grammar to know at this time WHY subjects tend to be definite (or more definite) than it is to know why voiced obstruents lower a following pitch. So I need not speculate here as to what the communicative function or 'meaning' of being subject might be in a given language or languages. The Grammar need only detect that there is a clustering of definiteness and subjecthood, which someone, perhaps a linguist, sociologist, or ethnologist, like our earlier phonetician for the nasalization and tone problems, may wish to take to an external level of explanation and account for why the substance is the way it is.

From this example and many like it we can conclude that pragmatics feeds into syntax exactly as phonetics feeds into phonology. I shall stay with the interaction between pragmatics, on the one hand, and syntax/

morphology, on the other, recognizing both the sometimes-vague line between semantics and pragmatics as well as the possibility of considering other interactions. If we have used the term 'phonologization' for the earlier examples, the term 'grammaticalization' seems appropriate to describe the harnessing of pragmatics by a grammar. It would be convenient to have a term to cover both phonologization and grammaticalization. 'Grammaticalization' could apply to both situations, but one interpretation of the term would overlook its phonological instantiation. On the other hand, a term like 'codification' (suggested to me by Henning Andersen), used when substance becomes part of the linguistic code, does not sound linguistic enough for my taste. So I will change off between phonologization and grammaticalization, according to the case, but wish to emphasize that these are two instances of the same phenomenon.

I should like to return to the notion of basic conflict and of conflicting interests. The property P, being part of substance, belongs to the extragrammatical world from whence it came. However, grammars, having P as a universally available formal feature, struggle to appropriate it from the world of substance and subject it to the laws of Grammar. The struggle between universal phonetics, i.e. the physiological properties of the vocal tract, and the laws of Grammar is obvious to any phonologist, but it is no less evident that the struggle between universal grammar ('Grammar' with a capital 'G') and pragmatics. We can differentiate three distinct situations or stand-offs which may obtain with respect to a given P in this struggle. These are listed in (7).

(7) a. Grammar takes care of its interests [linguistic *form*]; pragmatics takes care of its interests [nonlinguistic *substance*];
b. The interests of pragmatics encroach on the interests of Grammar;
c. The interests of Grammar encroach on the interests of pragmatics.

The situation in (7a) is the one where everything stays in its place: there is, with respect to some property P, no overlap between interests. An example might be the presence of conjugational or declension classes, as in Latin and other languages. This seems to be an area of grammar not of interest to pragmatics. The pragmatics, on the other hand, might like to have greater amplitude (loudness) on an imperative or negative form, but few grammars have this as a requirement (see below, however, for the interaction between such forms and focus marking).

The situations in (7b) and (7c) are more interesting and are dealt with in the following two sections.

3. Pragmatics encroaching on the interests of Grammar

In this situation the grammar of a particular language has a construction which is used for a certain grammatical purpose, and the pragmatics meddles and OVERRIDES the use of this grammatical construction for that purpose. I shall draw my examples from the syntax of body parts, an area I have investigated in several European and African languages. Consider, then, the French sentences in (8).

(8) a. j'ai lavé la chemise de l'enfant
'I washed the child's shirt.'
b. j'ai cassé le bâton de l'enfant
'I broke the child's stick.'

As seen in (8), the normal possessive or genitive construction involves the preposition *de* 'of' plus a noun phrase. The sentences in (8) thus have an 'NP of NP' in direct-object position. In (9), however,

(9) a. j'ai lavé les mains à l'enfant
'I washed the child's hands.'
b. j'ai cassé le bras à l'enfant
'I broke the child's arm.'

we see that when the possessed object NP is a body part, the expected construction is the preposition *à* 'to' plus an NP, i.e. identical in form to the indirect object in French. The sentences in (9) thus mean, literally, 'I washed the hands to the child' and 'I broke the arm to the child'. If we use the 'normal' possessive construction, as in (10),

(10) a. ?j'ai lavé les mains *de* l'enfant
b. ?j'ai cassé le bras *de* l'enfant

the impression gotten is that the 'hands' in (10a) and the 'arm' in (10b) are not part of the child's body, but rather some loose objects he may have found lying around somewhere. Many languages disprefer or ban the use of a possessive construction in such cases. The details may vary and depend upon the nature of the verb, the object NP, and the possessor NP, as argued in Hyman (1977b) for Haya, a Bantu language spoken in Tanzania. Details aside, the problem is always the same. The direct object of a verb is in semantic-case terms the 'patient' of that verb. If there is any semantic unity to this notion, it is that the direct object undergoes or 'is affected by' the action of the verb. In cases where the direct object NP is a body part belonging to a human NP, a potential confict arises because the possessor NP is necessarily and perhaps more critically affected by the action than the body part. This is not usually the case with the detached

object (body part or not) in (8) and (10). Although an idealized grammar (i.e. one without interference from the external world) would like to produce the sentences in (10), the pragmatics is not happy with this representation of the actions involved and their effect on the two NPs. In English, the concerns of Grammar win out in the sense that the possessive construction is used freely whenever possession is involved. (Remnant sentences such as *look me in the eye!* and *don't look a gift horse in the mouth!* reveal that English once had an alternative construction.) In this particular sense of 'concerns of Grammar' we mean, following Fodor (this volume), that an idealized Grammar wishes maximal generality, simplicity, and 'tidiness'. Fodor hypothesizes, further, that the Grammar wants the fewest statements possible, which is what the grammar of English gets with respect to body-part syntax.

There is another sense of 'concerns of Grammar', however, which is put into effect in the grammar of French and other languages which accommodate the pragmatics of body-part syntax in at least the two ways indicated in (11).

(11) a. The 'affected possessor NP' is expressed in the direct-object relation, with the semantic patient being expressed in the *chômeur* relation (e.g. in most Bantu languages).
b. The 'affected possessor NP' is expressed in the indirect-object relation, with the semantic patient left expressed in the direct-object relation (e.g. in many European languages).

Since the situations described in (11) represent an encroachment of pragmatics into individual grammars, we assume that a 'compromise' has been reached. The concerns of pragmatics are obvious: optimal expression of what an individual, society, culture wants to express. This corresponds to Fodor's notion of an 'expressor'. The concerns of Grammar in this process are different from those isolated by Fodor. In the grammaticalization process what we discover is the desire of the Grammar, and therefore individual grammars, to control anything and everything they can. In order to see that this is so, let us imagine French to be slightly different from what it is. In this slightly different French the sentences in (10) are used when the speaker is concerned with the effect of washing on hands and breaking on an arm; the sentences in (9) are used when the speaker is concerned with the effect of washing and breaking on the child. In this case the use of the possessive vs. indirect-object construction would be an intrinsic byproduct of the pragmatics. This, then, is the only situation where the pragmatics can be said to have 'won out'. The real French as we know it, however, virtually requires the sentences in (9) when the body parts are attached to a live possessor. This,

then, represents the corresponding extrinsic stage brought about by the grammaticalization of the pragmatic substance. We can look at the stages in (4) in the following way: the intrinsic stage represents the concerns of substance winning out over the concerns of Grammar; the 'emic' stage represents the concerns of Grammar winning out over the concerns of substance. In between the extremes is the extrinsic stage: this is the true compromise achieved because the substance is receiving formal recognition in a natural, nonarbitrary way. Pragmatics and Grammar reach a happy medium.

One of the interesting features about the encroachment of pragmatics into the realm of Grammar is that the loosening of the grammatical grip often extends beyond a single subarea of the individual grammar. When in the struggle between Grammar and pragmatics the latter finally breaks through, it is not just for one instance, but more generally. We see this particular clearly in the French *à* construction which marks an NP affected by an action. Consider the slight meaning difference in the following two sentences involving the *faire* causative construction (cf. Hyman and Zimmer 1976):

(12) a. j'ai fait laver la vaisselle par la bonne
'I had the dishes washed by the maid.'
b. j'ai fait laver la vaisselle à la bonne
'I had/made the maid wash the dishes.'

In (12a) the normal causative construction is used: the subject/agent of the lower clause becomes a *par* 'by' phrase as in the English translation. However, note in (12b) that an alternative construction is available with the agent expressed as an indirect object, i.e. the preposition *à* 'to' plus an NP. The reason for this is essentially the same as before. In (12a) the speaker is interested in the effect of the verb 'wash' on the dishes; in this case, the patient *la vaiselle* 'the dishes' gets no competition from the other NP, *la bonne* 'the maid', since the latter is expressed as a nonterm, i.e. by a 'by' phrase. The implication is that the agent in (12a) is only secondary, i.e. that I wanted to get the dishes washed and it happened to be the maid whom I found to do it. In (12b), however, a different set of circumstances obtains. In this case, as seen in the gloss, I am interested in the effect of the dishwashing on the maid in addition to the effect of the washing on the dishes. (12a) does not sufficiently express this, since the *par* 'by' relation is too low in the grammatical-relation hierarchy, just as it was in the *de* + NP possessive construction in (10).

How does one test this hypothesis? Consider the sentences in (13).

(13) a. ?j'ai fait laver Pierrot à la bonne
'I had the maid wash Pierrot.'
b. ?*je t'ai fait laver à la bonne
'I had the maid wash you.'
c. ?*ils m'ont fait élever à ma pauvre grand'mère
'they had my poor grandmother raise me.'

These sentences have human-patient direct objects. As indicated, it is hard to get French speakers to accept these sentences, even if providing precise and appropriate contexts. The reason is that with a human direct object the competition between the two NP referents ('maid' vs. 'me', 'you', 'Pierrot') is keener. That is, the two referents in each sentence are (intrinsically) equally capable of being affected by the action. In the Hyman and Zimmer study it was demonstrated that the acceptability of *à* + NP to express the causative agent depended upon the following person–animacy hierarchy:

(14)

1st person		(sg. > pl.)
2nd person		(sg. > pl.)
3rd person human	(definite > indefinite)	(sg. > pl.)
3rd person animate	(definite > indefinite)	(sg. > pl.)
3rd person inanimate	(definite > indefinite)	(sg. > pl.)

This hierarchy has been studied by a number of linguists besides myself, including Kuno (1976), Silverstein (1976), Hawkinson and Hyman (1974), Duranti (1979), Hopper and Thompson (1980), and others. Languages either do or do not allow it to influence their grammars — making cuts in different places — and perhaps making finer distinctions as in the Navaho 'great chain of being' — but they do not reverse the hierarchy.

As a final note in the French causative construction, consider the following pair of sentences (cf. Pinkham 1974):

(15) a. j'ai fait voir le film à l'enfant
'I had the child see the film.'
b. ?j'ai fait voir le film par l'enfant
'I had the film seen by the child.'

Since the direct object 'film' is inanimate, both constructions should be possible. However, with such objects, perceptual verbs such as *voir* 'see' do not readily allow the causative agent to be expressed with a 'by' phrase, as seen in (15b). Given what has been said, it is not hard to explain this fact. The film in (15) is not affected by the action of the verb; instead, only the 'experiencer' *l'enfant* 'the child' can be said to be affected. There is, however, a reading of (15b) which makes that sentence acceptable. This

is the sense where the child is acting as agent getting the film seen by others. The translation might be 'I had the film shown by the child'.

Returning to the hierarchy in (14), one might ask why this or any other such array of pragmatic features should become so involved with the giving out of grammatical relations and other grammatical properties. The idea that I have been developing here is one of constant conflict. Real speaker/hearers want to have the communicative freedom to exploit linguistic material at will. Thus, they want to be able to nasalize when it is convenient to them and use *à* vs. *de/par* according to the intended message. Let us refer to such freedom in the area of syntax as *pragmatic control.* To this we oppose the ever-present pressure by grammars on substance. An idealized Grammar existing without any real-world constraints would like all choices to be controlled by individual grammars, not by speakers. This means choices such as whether to nasalize or not, whether to use *à* vs. *de* or *par*, etc. Let us refer to this force as *grammatical control.* The result is compromise: what speakers wish to exploit is encoded, but with resultant grammatical control. What is achieved, in effect, is an ICONICITY between degrees of speaker concern or 'empathy', in Kuno's terminology, and the feature hierarchies in (14). This iconicity is extended as well to a hierarchy of semantic roles (or the grammatical cases and grammatical relations the semantic roles tend to receive). Thus, languages show a tendency to associate the higher feature values in (14) with the higher semantic roles in the hierarchy: agent > recipient/benefactive > patient > instrument, etc.

4. Grammar encroaching on the interests of pragmatics

In the preceding section we saw that French speakers entered into a compromise arrangement with the idealized Grammar. In both the possessive and the causative constructions the result was that they yielded to the grammar of French some control over whether one vs. another form would be used in a given context. The trade-off was the encodability of affectedness in the grammar of French. Rather than remaining apart, the interests of pragmatics encroached upon the grammar of French and the resulting compromise was struck. What started therefore as a grammatical opposition (e.g. 'possessive' vs. 'indirect-object' construction) now has some pragmatics built into it. In other languages what starts as a pragmatic opposition comes to have grammar built into it. This is the third possibility mentioned in (7c). I would like to illustrate this reverse situation with respect to cases where the choice of focus marking, which should be pragmatically controlled, is instead grammatically controlled.

Consider the sentences in (16).

(16) a. *John* ate rice (ans. to Q: who ate rice?)
b. John *ate* rice (ans. to Q: what did John do to rice?)
c. John *did* eat rice (ans. to Q: didn't John eat rice?)
d. John ate *rice* (ans. to Q: what did John eat?)

The focus in unmarked declarative English sentences is realized by putting the high pitch of the high-low intonation contour on the main stress of the focused material. As seen, the focus may be placed on the subject NP (16a), the verb (16b), the auxiliary (16c), or the direct object (16d) in a simple transitive sentence. The placement of focus marking in one vs. another place depends on two considerations: (a) the context; and (b) the intention of the speaker. If the context is one of the WH questions given, the only appropriate answer is as shown. Since it is the speaker who decides where focus marking will be, English has pragmatic control of focus marking. This is not to say that rules of grammar do not place the focus; only that the choice of one place rather than another is not decided by the grammar.

Now let us modify English only slightly. Assume that there is a dialect of English which differs from the standard in only one respect: if the speaker of this dialect chooses to use an imperative construction, he must put the focus on the imperative verb. Thus, we would have an exchange such as in (17).

(17) Speaker 1: What should I eat?
Speaker 2: *Eat* rice!

Clearly the utterance '*Eat* rice!', with stress on the verb, does not sound right to us in this context. Yet parallels to just this development are found in a number of African languages, e.g. Aghem and Somali, which treat the imperative as inherently [+focus] (Hyman and Watters 1980). Aghem also requires that [+focus] be marked on a negative auxiliary. Thus we would have the exchange in (18) in our imaginary modified English dialect:

(18) Speaker 1: Why isn't Professor Hawkins here?
Speaker 2: Because he *isn't* in town.

Again the exchange does not seem natural, but it reflects what Aghem does in its morphological focus marking (see the studies in Hyman 1979). It is not possible for me to go into what a grammatical account of Aghem or this hypothetical English dialect might look like. Let us assume for the purpose of discussion that rules such as in (19) will become part of their grammar:

(19) a. [+imperative] → [+focus]
b. [+negative] → [+focus], etc.

I have put 'etc.' because it is not only these two features which acquire [+focus] marking, although they are the most common. A more complete picture would be as in (20).

(20) a. 'marked' polarity = negative
b. 'marked' mood = imperative (possibly subjunctive too)
c. 'marked' aspect = progressive
d. 'marked' tense = perfect

Any of these may attract focus marking, and since there can be only one [+focus] per clause in these languages, the [+focus] is robbed from where it might have otherwise been according to the context and the intentions of the speaker. Why should this be?

The answer lies in the question, what is focus? We are accustomed to seeking sophisticated responses to this question, relying on presupposition, scope of assertion, exhaustive listing vs. counterassertive focus, etc. But I think the answer is much simpler: grammatical focus is the assignment of [+focus] by a grammar, and this [+focus] is a mark of salience within the grammar. Given that not all information communicated is of equal salience, languages could conceivably choose to deal with this in a number of ways, as indicated in (21).

(21) a. languages could ignore salience altogether;
b. languages could allow gradated and unlimited marking of salience;
c. languages, through their grammars, could 'harness' the pragmatics and create a formal system for focus.

(21a) seems unreasonable since it would completely ignore the needs of speakers: there simply have to be means of highlighting information at the expense of other information, if for no other reason than to get the attention of listeners who might be less interested in monotone and an utterance devoid of affect. (21b) is more reasonable and represents the complete pragmatic control of salience. Imagine a language which places a [+focus] marker *baa* after any and as many items as the speaker chooses. This is a situation which the speakers might like, just as their vocal tract might like to nasalize only when it's functional to do so. But an idealized Grammar would not like it. The driving force of Grammar is to get control of whatever it can, as in (21c). This is normally done in a simple way by reference to focus within the *propositional* content of an utterance: e.g. a WH element may receive [+focus] marking, as will its answer. Otherwise, put [+focus] on a salient part of the proposition.

What Aghem and many other African languages do is consider relative salience within the OPERATOR system, that is, within the auxiliary. It is intuitive than when one tells someone 'no', it is more salient than when one tells them 'yes'. The operator is more salient when you order someone via an imperative than when you make an indicative declarative statement. Progressive aspect focuses attention on the ongoingness of some action (e.g., 'John is playing cards'), while the nonprogressive, being unmarked, does not focus attention on the non-ongoingness of the action ('John plays cards'). Instead, the nonprogressive competes less with the propositional content of the utterance, thereby allowing [+focus] more ready access to one of its constituents outside the auxiliary. Finally, the perfect tense (or aspect, as you will) focuses attention on the effect of some prior action on some later state. The absence of a perfect does not do the reverse, but again, as in the nonprogressive, simply allows the proposition more relative salience.

What the above indicates is that there is a competition between salience within a proposition and salience within a system of operators. As a confirmation of this competitive view of salience, with different kinds of saliences vying for one [+focus] marking, let us consider constructional focus. Some clause types are inherently more salient than others. In particular, main clauses are more salient than nonmain clauses. Let us return briefly to the hypothetical English dialect. Imagine that this dialect, like many African languages, starts to forbid any [+focus] assignment within a relative clause. We now have an exchange as in (22).

(22) Speaker 1: which book did he read?
Speaker 2: [he read] the *book* that you gave him.

Again there is an unnatural focus marking, but one which we would have to have in some languages. In (23) I have indicated the kinds of clauses which do not allow [+focus] marking in certain African languages:

(23) a. cleft clauses (it's the *child* that I saw)
b. relative clauses (the child that I saw...)
c. adverbial clauses (when I saw the child...)
d. if-clause (if I saw the child...)
e. consecutive clauses (he came and — I saw him)

(23) represents a proposed hierarchy. Even in English it is rather odd to get a [+focus] within the cleft clause. The above clauses typically create islands and essentially constitute the class of nonroot Ss in Emond's (1976) framework. Why should it be these that repulse the [+focus]?

A moment's reflection will reveal that (23) represents a class of 'backgrounded' clauses (cf. Schachter's [1973] discussion of focus and

relativization). They are typically used not to make new assertions but rather to provide circumstances under which the assertions of main clauses hold true. Thus, to look only at these embedded clauses would not tell us much about the story line of a narrative but only about the periphery of the tale. We see some indication of this in the English sentences in (24).

(24) a. he *did too* hit me! (~did so)
b. *it's John who did too hit me
c. *the child who did too hit me
d. *when he did too hit me...
e. ?*if he did too hit me...

The counterassertive *too/so* construction goes only in assertive or main clauses. Aghem and other languages simply extend this so that relatively less salient clauses are exempted from any [+focus] marking. Thus, the [+focus], which should have been placed on the basis of speakers' wishes relative to a given context is largely controlled by the grammars of these languages.

This does not mean, necessarily, that differences in focus cannot be indicated in the context of negation or backgrounded clauses. Besides the morphological marking which is sensitive to [+focus] specifications, these languages allow SYNTACTIC devices, e.g. word-order variations, to express differences in relative salience. Note, finally, that if you choose in our hypothetical English dialect a main clause, affirmative, indicative, nonprogressive, nonperfect, one still, as a speaker, has some choice in placing the [+focus].

5. Conclusion

At the beginning of this paper I distinguished internal vs. external explanations for language universals. Given the preceding discussion, it should be clear that the totality of language will be accounted for only by a combination of 'explanations'. If we take as a major goal the explanation of grammatical properties recurrent in languages, we shall have to postulate, on the one hand, abstract principles of Grammar, as Chomsky has argued most forcefully, which constrain what individual grammars can do, and, on the other hand, principles, abstract or not, of communication, cognition, etc., to predict the kinds of substance which may become grammaticalized — and in what ways, order, etc. The discussion of focus in section 4 was, admittedly, sketchy, but some very basic predictions can be derived from it. First, the notion of marked (=salient) operators in

(20) predicts that negatives, imperatives, progressives, and perfects may 'associate with focus' (Jackendoff 1972) but that no language will have such an association only in the case of the unmarked values (affirmative, indicative, nonprogressive, nonperfect). This association may result in a [+neg] attracting the one [+focus] marking of a clause. Second, the notion of background clauses in (23) predicts that main clauses will always make as many focus distinctions as, or more than, cleft, relative, adverbial, 'if', and consecutive clauses.

External explanations thus relate properties of Grammar to substance (phonetics, psychology, sociology, etc.). Since substance is taken to be 'real', e.g. often physically measurable or capable of statistical analysis, external explanations are in a sense more concrete and accessible than internal explanations. In this paper I have tried to place the two kinds of explanations in perspective. Phonetics provides much of the substance of phonology, and pragmatics provides much of the substance of syntax. However, the ever-present phenomena of phonologization and grammaticalization cannot be explained by reference to the origin of the substance. Grammar has its own laws which, whether innate or learned, are species dependent. While it is universally asserted that LANGUAGE is present for communicative purposes, it is harder to demonstrate that GRAMMAR exists for the same reason. Or, restated as a problem in language acquisition, it is as if the need of the child to communicate is subordinated to his need to develop a formal system; that is, to grammaticalize as much substance as possible. Phonologization and grammaticalization become, then, an overformalization of substance by the child which gradually works its way into the adult language. We can thus reformulate Chomsky's (1965) distinction between substantive and formal language universals as follows. Substantive universals provide statements, hierarchies, implicational universals, etc., on how substance comes to be grammaticalized. Formal universals provide statements, hierarchies, implicational universals, 'parameters' concerning the internal formal operations of grammars. All of the generalizations reported in this paper fall, then, in the category of substantive universals. It is hoped that the years ahead will see the convergence of substantive and formal universals into a unified theory of language structure.

Department of Linguistics
University of Southern California
Los Angeles, Calif. 90089
USA

Note

* This paper represents a compromise between the presentation made at the Workshop, which I entitled 'Universals of focus', and a Professorial Address made January 20, 1982, at the University of Southern California, entitled 'Language [read: Grammar] has a mind of its own'. I would like to thank participants at both events for their helpful comments and discussion, especially Bernard Comrie, Edward Finegan, Janet Fodor, Jack Hawkins, Osvaldo Jaeggli, Stephen Krashen, and Elinor Ochs.

References

Chomsky, Noam (1965). *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.

Duranti, Alessandro (1979). Object clitic pronouns in Bantu and the topicality hierarchy. *Studies in African Linguistics* 10, 31–45.

Emonds, Joseph E. (1976). *A Transformational Approach to English Syntax*. New York: Academic Press.

Hawkinson, Anne K., and Hyman, Larry M. (1974). Hierarchies of natural topic in Shona. *Studies in African Linguistics* 5, 147–170.

Hombert, Jean-Marie (1978). Consonant types, vowel quality, and tone. In Victoria A. Fromkin (ed.), *Tone: A Linguistic Survey*, 77–112. New York: Academic Press.

Hopper, Paul, and Thompson, Sandra (1980). Transitivity in grammar and discourse. *Language* 56, 251–299.

Hyman, Larry M. (1975). Nasal states and nasal processes. In Charles A. Ferguson et al. (eds), *Nasalfest: Papers from a Symposium on Nasals and Nasalization*, 249–264. Stanford: Stanford University Press.

—(1977a). Phonologization. In Alphonse Julliand (ed.), *Linguistic Studies Presented to Joseph H. Greenberg*, 407–418. Saratoga, Calif.: Anima Libri.

—(1977b). The syntax of body parts. In Ernest Rugwa Byarushengo et al. (eds), *Haya Grammatical Structure*, 99–117. Southern California Occasional Papers in Linguistics No. 6. Los Angeles: University of Southern California.

—(ed.) (1979). *Aghem Grammatical Structure*. Southern California Occasional Papers in Linguistics No. 7. Los Angeles: University of Southern California.

—, and Watters, John Robert (1980). Auxiliary focus. Paper presented at Conference on Auxiliaries, Fourth Annual Round Table in Linguistics, University of Groningen, Netherlands, July 4–8, 1980.

—, and Zimmer, Karl E. (1976). Embedded topic in French. In Charles N. Li (ed.), *Subject and Topic*, 189–211. New York: Academic Press.

Jackendoff, Ray S. (1972). *Semantic Interpretation of Generative Grammar*. Cambridge, Mass.: MIT Press.

Kuno, Susumo (1976). Subject, theme, and the speaker's empathy — a reexamination of relativization phenomena. In Charles N. Li (ed.), *Subject and Topic*, 417–444. New York: Academic Press.

Mohr, Burckhard (1971). Intrinsic variations in the speech signal. *Phonetica* 23, 65–93.

Ohala, John J. (1973). The physiology of tone. In Larry M. Hyman (ed.), *Consonant Types and Tone*, 1–14. Southern California Occasional Papers in Linguistics No. 1. Los Angeles: University of Southern California.

—(1978). Production of tone. In Victoria A. Fromkin (ed.), *Tone: A Linguistic Survey*, 5–39. New York: Academic Press.

Pinkham, Jessie (1974). Passive and *faire-par* causative construction in French. Senior Essay, Harvard University.

Schachter, Paul (1973). Focus and relativization. *Language* 49, 19–46.

Silverstein, Michael (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (ed.), *Grammatical Categories in Australian Languages*, 112–171. New Jersey: Humanities Press.

Vennemann, Theo (1972). Phonetic analogy and conceptual analogy. In Theo Vennemann and Terence H. Wilbur (eds), *Schuchart, the Neogrammarians, and the Transformational Theory of Phonological Change*, 181–204. Frankfurt am Main: Athenäum.