REPRINTED FROM
COMPUTER AND INFORMATION SCIENCES — II

ACADEMIC PRESS INC., NEW YORK

The Nature of Syntactic Redundancy

D. Terence Langendoen
OHIO STATE UNIVERSITY
COLUMBUS, OHIO

It has been clearly demonstrated that the syntactic structure of language is far richer than that of any "transitional probability" or stochastic model. The transformational-generative theory of Chomsky probably is the closest approximation to the "truth" that linguists have yet come up with.

Within the Chomskyan framework [particularly that of his "Aspects of the Theory of Syntax," Cambridge, 1965] we may ask how redundancies can be talked about at each of the three levels of description: phrase structure, lexicon, and transformation. At the lexical level things look fairly familiar; one can talk in terms of lexical redundancy rules which interrelate the various kinds of lexical features, e.g.,

$$[+\text{human}] \rightarrow \begin{bmatrix} +\text{animal} \\ +\text{animate} \\ +\text{physical object} \\ -\text{abstract} \end{bmatrix}$$

However it is still an open question as to the relationship between the kinds of features illustrated in the preceding rule and features which govern the substitutability of lexical items in particular grammatical contexts. Two separate problems relating to this open question are taken up: the choice of particular classes of nouns as subject or object of a particular class of verbs, and the problem of the so-called cognate object.

I. Introduction

The detailed study of the nature of redundancy to be found within the sentences of human languages has had to await the development of certain theoretical and analytical tools, which have only recently come into existence. This is not to say that certain aspects of redundancy in language have not been known and well studied for many years, in fact millennia, for example the phenomena of concord and agreement. Restrictions on the freedom of occurrence of particular words or phrases in particular environments have also been well-known for a long time, although the systematic study of this topic is quite recent. One important result of this kind of study has been to show that while it is possible to approximate closer and closer the structure

of actual sentences of a language such as English by systematically increasing the integer n indicating the number of words over which one obeys such restrictions, the approximation is asymptotic. For any fixed n it is always possible to construct a sentence of English (or of any natural language) in which there is a linguistic constraint between two words separated by more than n words of that sentence.

II. Phonological Redundancy

Modern studies of the phonological systems of natural language have also shed considerable light on the nature of redundancy in language namely, those restrictions that hold among the elements of speech or pronunciation. While this is not the topic of this paper, consideration of what this work has revealed may afford us some significant clues as to what to expect in syntax, for there are enough parallels between phonology and syntax to make a somewhat detailed survey of the former worthwhile for our purposes.

Within modern structural phonology, dating in America from the work of Leonard Bloomfield (1887–1949), the primary theoretical concern has been to develop a kind of redundancy-free notation, or transcription, in which the elements of the transcription, called phonemes, represent classes of actually occurring sounds, or phones, which make up the elements of a detailed phonetic transcription. In this view, the individual sounds making up the phoneme are seen as being environmentally conditioned (by neighboring sounds); the symbol chosen for the transcription of the phoneme was usually the symbol of the most commonly occurring of the phones, or as close to it as possible subject to extralinguistic (usually typographic) considerations. The special phonetic details of phonetic transcription was considered redundant from a structural (i.e., phonemic) viewpoint and could therefore be ignored in a phonemic transcription.

Due in part to specific limitations of the view that phonemes were non-intersecting classes of phones,[1] a redundancy-free notation was by no means achieved. It was immediately recognized that the distribution of phonemes within, say, words is still highly constrained. Certain phonemes were found to be defectively distributed vis-à-vis other phonemes. The most original attempt to handle this problem was Zellig Harris' scheme of long component analysis. If, say, a particular combination of phonemes did not occur in a language, Harris sought to extract a component from the phoneme combinations of the sequences that did occur so that the segments which remain have complete freedom of distribution and hence are redundancy-free. Of course

[1] Or more liberally, in some linguists' views, two phones are permitted to belong to different phonemes provided that the environments in which they occur as members of those phonemes do not intersect.

Harris succeeded only in pushing the problem of redundancy among the phonemic segments into his long components; the same questions of distribution can be asked about the long components, and the whole problem repeats itself. Anyway, for a number of reasons which need not concern us here, Harris' long component analysis has been all but completely abandoned by the succeeding generation of linguists, including Harris himself.

Since Harris' concept of the long component was introduced, the most significant development in phonological theory has been the elaboration of the theory of generative phonology by Morris Halle, Noam Chomsky, and others. One of the most important features of this theory is the notion that the units of phonology are not segments as such but universal phonological features having a basis in the articulatory and acoustic character of speech. The idea of a universal set of phonological features forming the basis of phonology is due to Roman Jakobson.[2] Individual segments are representable in terms of Jakobson's theory as bundles, or more properly one-column matrices, having as rows values for the individual features, such as voicing, interruptedness, nasality, lip-rounding, and so forth. In the phonetic representation of speech, specific values (usually + or −, but not necessarily) are given for all the features for all the segments, whereas in a phonemic representation, not all values are specified, but some positions may be left blank, while those that are not contain either a plus or a minus. Immediately we can draw what appears to be an obvious conclusion: those phonetic features specifications corresponding to a blank in the phonemic representation are redundant; their phonetic values may be provided by means of phonological redundancy rules.

Actually the picture is not quite as simple as that, for various reasons not all of which can be given here. According to generative phonological theory, phonological rules may change feature specifications, as well as introduce them, and rules may also delete, insert, and transpose entire segments, so that it is not always obvious what is and what is not redundant in a phonetic specification. Furthermore, certain obvious redundancies, notably those having to do with the patterning of consonants and vowels into so-called syllable structure have been notoriously difficult to describe by rule, so much so that whole aspects of the theory as originally formulated have been called into question. Finally certain of the feature specifications are redundant because the features themselves are not entirely independent of one another. For example, non-continuants are always non-stressed; nasals are never obstruents; obstruents are always consonants. There is no point in attempting to eliminate these interdependencies, for a system without them would have no significance as a set of phonological features for human languages. Thus,

[2] The influence of Jakobson's ideas can also be quite clearly seen in the work of Harris; cf. "Structural Linguistics," pp. 146–149. Univ. of Chicago Press, Chicago, Illinois, 1951.

in the feature representation of a particular phonological segment, it is possible to consider many feature specifications to be redundant simply by virtue of the nature of the features themselves. Consider for example the English segment [s], which may be represented phonetically as:

(1)
$$\begin{bmatrix} /-/ \text{ Vocalic} \\ /+/ \text{ Consonantal} \\ (+) \text{ Continuant} \\ /+/ \text{ Obstruent} \\ /-/ \text{ Nasal} \\ + \text{ Sibilant} \\ - \text{ Voiced} \\ /-/ \text{ Grave} \\ - \text{ Compact} \\ (-) \text{ Tense} \\ (-) \text{ Retroflex} \\ (-) \text{ Lip-Rounded} \\ (-) \text{ Checked} \end{bmatrix}$$

where we have put in parentheses those features specifications which are redundant as a consequence of the phonological structure of English[3] and put in solidi those that are redundant because of the interrelationships among the features themselves. In this case, we have made use of the fact that all sibilants are non-vocalic, consonantal, non-grave obstruents. Thus it appears to follow that a systematic phonemic representation of the English segment *s* should be:

(2)
$$\begin{bmatrix} + \text{ Sibilant} \\ - \text{ Voiced} \\ - \text{ Compact} \end{bmatrix}$$

where we have simply omitted the redundant features.

But there are serious problems with viewing (2) as a phonemic representation of *s*. The specification [+Sibilant] for *s* in English is not itself redundant because English does have a non-sibilant, non-grave, non-compact continuant obstruent, namely *θ*. Suppose *θ* were not a phonological element in English. Then the specification [+Sibilant] for *s* would itself be redundant, and we would then have to consider those features which were redundant because of the specification [+Sibilant] non-redundant. This includes the specification [+Continuant], a redundancy peculiar to English. The segment would have to be represented as:

(3)
$$\begin{bmatrix} + \text{ Obstruent} \\ + \text{ Continuant} \\ - \text{ Voiced} \\ - \text{ Grave} \\ - \text{ Compact} \end{bmatrix}$$

[3] Not all varieties of English *s*, needless to say, have this exact phonetic character. For example the initial *s* of *sweet* is usually articulated with lip-rounding.

(the specifications [−Vocalic], [−Nasal], and [+Consonantal] are redundant given the specification [+Obstruent]). Thus the elimination of the segment θ from English would seem to have the curious consequence of rendering the redundancy-free feature representation of *s* more complex!

This is, of course, a hypothetical difficulty. There are, however, also real ones. Consider the well-known fact that in word-initial position in English, the only obstruent that can occur before an other obstruent is *s*. Thus we have words beginning with *sp-*, *st-*, *sk-*, *sf-*, but not for example with *zb-*, *šk-*, *ft-*, *ks-*, etc. What we would like to say about English is that given an initial cluster of obstruents, the first is necessarily, that is redundantly, a voiceless, non-compact sibilant. But this means that we are considering this *s* to be phonemically represented as [+Obstruent] and redundantly [+Sibilant], whereas other *s*'s in the language are phonemically [+Sibilant], and redundantly [+Obstruent]. This is an extremely unsatisfactory state of affairs.

The difficulty can be resolved, and as I understand the situation can only be resolved, if we take the position that a systematic phonemic, or what I would call a phonologically linguistically significant representation of a language, is not the same thing as a phonologically redundancy-free representation. Feature specifications which are redundant in all languages by virtue of the interdependencies among the universal phonological features should be included in a systematic phonemic representation. Only those redundancies which are language-specific should be left out of such a representation. This means, in particular, the systematic phonemic representation of the English *s* should be:

(4)
$$\begin{bmatrix} -\text{Vocalic} \\ +\text{Consonantal} \\ +\text{Obstruent} \\ -\text{Nasal} \\ +\text{Sibilant} \\ -\text{Voiced} \\ -\text{Grave} \\ -\text{Compact} \end{bmatrix}$$

The feature specifications of (4) are all those of (1) which are not in parentheses. The specification of *s* in the event that θ were not present in the language would turn out to be:

(5)
$$\begin{bmatrix} -\text{Vocalic} \\ +\text{Consonantal} \\ +\text{Obstruent} \\ -\text{Nasal} \\ +\text{Continuant} \\ -\text{Voiced} \\ -\text{Grave} \\ -\text{Compact} \end{bmatrix}$$

which is at least not more complicated than (4). The fact that the same number of specifications have to be given in (5) as in (4) is a consequence of the fact that [+Continuant] is needed to distinguish *s* from *t* if [+Strident] is not used.

The specification of word-initial *s* preceding another obstruent would be:

(6)
$$\begin{bmatrix} -\text{ Vocalic} \\ +\text{ Consonantal} \\ +\text{ Obstruent} \\ -\text{ Nasal} \end{bmatrix}$$

which satisfies our intuitions that the specification of this *s* should be simpler than that of other *s*'s in the language.

III. Problems of Syntactic Redundancy

This concludes my discussion of the nature of phonological redundancy. I have attempted to show that a linguistically significant phonological representation is not the same thing as a redundancy-free one. In a moment I shall attempt a similar argument for the syntactic representation of natural language. Much more could be said about the lack of exact correspondence between a systematic phonemic and a redundancy-free phonological representation, but the point has been served.

In order to provide a coherent framework for our discussion of the nature of syntactic redundancy, I shall assume the so-called transformational-generative model of syntactic structure, of the type outlined in "Aspects of the Theory of Syntax." I shall restrict my attention to what may be called lexical redundancies, and not consider aspects of syntactic redundancy having to do only with the phrase structure or the transformational components. I shall also only deal with specific problems and not attempt anything like overall coverage. Quite often, the considerations I shall present will resemble very little the traditionally defined notion of redundancy—I keep the term to cover all these cases because I do not think it to be an unreasonable extension of the referential capacity of the term to make it do service in this way.

IV. Lexical Subcategorization

Following Chomsky, "Aspects of the Theory of Syntax," we assume that the individual lexical items of a language such as English are represented as a bundle of syntactic features. For example, the noun *boy* may be considered to have in its representation:

(7)
$$\begin{bmatrix} +\text{Noun} \\ +\text{Physical-Object} \\ +\text{Animate} \\ +\text{Animal} \\ +\text{Higher-Animal} \\ -\text{Adult} \\ +\text{Masculine} \\ +\text{Count} \end{bmatrix}$$

Clearly, certain of these specifications are redundant; all English nouns which are subcategorized as [+Human] are furthermore [+Higher-Animal, +Animal, +Animate, +Physical-Object] and also [+Count]. Intuitively, the specifications [+Human, +Higher-Animal, +Animal, +Animate, +Physical-Object] form a hierarchy, and following Aristotle,[4] we need only specify the value of the most specific feature of the hierarchy. In fact, the specification [+Noun] can be said to follow from [+Human], since there are no "human" verbs, adjectives, or prepositions in English (or presumably in any natural language), so that a redundancy-free lexical representation of *boy* would be:

(8)
$$\begin{bmatrix} +\text{Human} \\ -\text{Adult} \\ +\text{Masculine} \end{bmatrix}$$

Again, as in the case of phonological representation, we may ask whether the redundancy-free lexical representation (8) is linguistically well-motivated as a systematic representation. I think the answer here, too, is negative, for it is unnatural to leave out of the lexical representation of items like *boy* redundant features such as [+Noun, +Animate, etc.] which follow by universal conventions on the nature of lexical features.

Lexical features of the sort just considered clearly have to do with the meaning of the lexical items in question, but it is equally clear that some of them, at least, play important syntactic roles as well. The question is frequently raised as to the place of the "line" between syntax and semantics, but no answer to that question can be given until a clear and significant formulation of the theories of syntax and semantics is provided, and we presently lack both. Let us then call features of this sort *inherent features* of the lexical items in question, leaving aside the question of whether they are inherently syntactic or semantic.

According to Chomsky, there are two other sorts of lexical features which belong in the representation of lexical items. These features have to do (1)

[4] "Metaphysics" 1040b; "De Partibus Animalum" 642b. Chomsky, ("Aspects of the Theory of Syntax," p. 231. M.I.T. Press, Cambridge, Massachusetts, 1965) attributes the idea in its present form to Paul Postal.

with the categorization of constituents with which the item in question can co-occur and (2) with the inherent features of the consistuents with which the item can co-occur. The first type is called a *strict subcategorizational feature* and the second a *selectional restriction*. An example of the first type is the familiar subcategorization of verbs into transitive and intransitive, depending upon whether the verb co-occurs with a direct object noun phrase or not, while an example of the second type would be the designation of such verbs as *kill* and *frighten* as selecting animate objects.

It should be apparent that there is a strong connection between the inherent, strict subcategorizational and selectional feature specifications of a given lexical item.[5] The strongest conceivable connection would be for subcategorizational and selectional features to be completely predictable from inherent ones, in other words for them to be wholly redundant. As it stands, this claim appears to be much to strong; what inherent feature of an intransitive verb such as *vanish* can lead us to a prediction that it is in fact intransitive, or of a verb such as *like* that it is transitive? Similarly what inherent feature of *gallop* may be used to predict that it is selected by equine subject nouns or that the subject and object of *marry* (in its non-causative sense) are of opposite sexes? The problem in these cases is that we are not used to thinking of the inherent features of verbs to the extent that we are of nouns; we are too readily struck by the obvious features of selection and subcategorization. It seems to me that if we do postulate inherent features for verbs and attempt to predict their selectional and subcategorizational features from these, we come to be able to offer possible solutions to some puzzles about selectional and subcategorizational features that Chomsky and others have noted for English.

One such case has to do with a subclass of transitive verbs in English which express mental or emotional states or events. In many cases, the direct object is the "possessor" of the feeling involved and the subject can be construed as the cause, as in:

(9) You amuse me.
(10) All that noise outside is distracting your father.
(11) John's newest novel really pleases the editor.

However there are a number of similar verbs in which the roles of the subject and the object are reversed:

(12) The editor really likes John's newest novel.
(13) Your father doesn't enjoy all the noise outside.

[5] As Chomsky has argued, only the categories verb and adjective take on selectional features to any degree, and most of our examples from this point will be taken from these categories.

The problem is to account for the reversal of subject and object in (12) and (13) as compared with (9)–(11); cf. Chomsky, "Aspects of the Theory of Syntax," pp. 162–163. We may do so, perhaps, in the following way, making judicious use of inherent lexical features for the verbs *amuse*, *distract*, *please*, *like*, and *enjoy*. We may say that these verbs all have to do with the mind, that they are inherently [+Psychological]. Then we may say that the mental state or event is externally induced or comes from within; provisionally let us say that *like* and *enjoy* are [+Inner] along with *fear*, *believe*, *know*, *think*, while *amuse*, *distract*, *please* are [−Inner], along with *annoy*, *frighten*, *concern*, *surprise*. Given this classification, we may state as lexical redundancy rules:

(14) $\begin{bmatrix} +\text{Psychological} \\ +\text{Inner} \end{bmatrix} \rightarrow [+[+\text{Higher-Animal}] - \text{Subject}]$

(15) $\begin{bmatrix} +\text{Psychological} \\ -\text{Inner} \end{bmatrix} \rightarrow [+[+\text{Higher-Animal}] - \text{Object}]$

Notice that there are verbs classifiable as both [+Inner] and [−Inner] for example *worry*, *thrill*. Consequently we have such sentences as:

(16) The present situation worries me.
(17) I worry about the present situation.
(18) The hum of the engines thrilled the mechanic.
(19) The mechanic thrilled to/at the hum of the engines.

Notice that the difference in meaning between the sentence pairs is very hard to characterize; it seems to be very much like the difference between the following pair, in each of which in addition different verbs are used:

(20) A bright idea just occurred to me.
(21) I just thought of a bright idea.

Basically in (17), (19), (21), as opposed to the even numbered examples, the agency of the individual seems to be involved.

Much more work obviously needs to be done on this and similar problems before it can be maintained or denied conclusively that selectional and subcategorizational features are predictable entirely from inherent ones. To me, this is an attractive thesis and deserves further serious consideration.

5. The Problem of the Cognate Object

I now want to take up one other problem of syntactic redundancy which crucially involves the nature of lexical representation, particularly that of verbs: the problem of the so-called cognate object. The English language is not perhaps the best choice for illustrative purposes, but it is still not hard to find interesting illustrative material in it. Consider the sentence:

(22) Cassius fought a good fight last night.

The occurrence of the noun *fight* as direct object of the verb *fight* in (22) can be said to be an instance of a cognate object. Another example which may be said to be illustrative is:

(23) Frankie sang a good song last night.

Upon closer examination, however, (22) and (23) are seen to have rather distinct semantic and syntactic properties. Notice that (22) can be closely paraphrased by

(24) Cassius fought well last night.

whereas (25) below is not a paraphrase of (23) at all:

(25) Frankie sang well last night.

Let us say then that *song* is not a cognate object to the verb *sing* in (23); we will say tentatively that a cognate object is a noun of similar or identical phonological shape to its verb, and may be eliminated without significant overall change in meaning. Thus the noun *fight* in (22) may be considered a cognate object.

Not every occurrence of *fight* as object of the verb *fight* is a cognate object, however. Consider:

(26) The sergeant fought several fights last night.

which has no paraphrase without the direct object. Note also that this occurrence of the object *fights* is understood analogously to *song* in (23), and if we look again at example (22) in light of (26), we observe that in fact (22) is ambiguous; in one sense *fight* is a cognate object and in the other sense it is not.

The problem of the cognate object in English has, then, at least the following aspects. Its occurrence is idiosyncratic, depending on the verb in question. It appears to have no special phonological or other properties which distinguish it from non-cognate objects, although its phonological shape is constrained by the verb. When it does occur, it occurs most naturally with an evaluative adjective, such as *good*, *terrible*, etc. and with an indefinite determiner. Finally, inasmuch as we would like to say that it is the cognate object which is determined by the verb, and not vice versa, we are faced with a puzzle, since in general it is the case that verbs are selected by features of their objects and not the other way around.

This last observation, however, gives us an interesting and important clue toward a possible solution to the problem. Suppose we say, in rather too colorful language, that the cognate object represents "the verb's revenge" on the noun for determining it (by selectional restrictions) in the first place. More precisely let us say that particular verbs, such as *fight*, may be selected

by object nouns having certain features, but that a particular lexical item need not be chosen for it (in Chomsky's terminology, the item is represented by a dummy symbol of some sort). If a lexical item is not chosen, then a nominalization of the verb is inserted for the dummy by a rule which we may call the cognate object insertion rule. The nominalization has in particular the feature [−Definite], to account for the occurrence of the indefinite determiner, and we may stipulate also, if we wish, that the dummy must be modified by an evaluative adjective. In this way we can account for the syntactic ambiguity of (22) and non-ambiguity of (23). Under one interpretation of (22) and in (23) a direct object has been supplied by the cognate object insertion rule. It is a property of the lexical item *sing* that it never takes a cognate object, just as it is a property of *fight* that it may.

At least three interesting questions remain. Does a cognate object have to be phonologically similar to its verb? What is the reason for the widespread character of the phenomenon across languages? Are there such things as cognate subjects or cognate instruments?

There is evidence, I believe, for maintaining that the answer to the first question is negative. It seems to me, for example, that the noun *game* may be used as a cognate object to the verb *play*. Consider the sentence:

(27) The Reds played well last night.

which has two possible interpretations, one of them essentially synonymous with:

(28) The Reds played a good game last night.

Examples of this sort are admittedly quite rare, but this is a case in which one clear counterexample is sufficient.

The second question seeks for exaplanation, where we have been content only with description. Quite frankly, I do not have any explanations handy, but I do think that my concept of "the verb's revenge" is a step in the direction of explanation. Chomsky's account of the subordinate status of verbs and adjectives with regard to selection leaves no room for either of these categories to play a dominant role in noun selection. Presumably this is a universal state of affairs. The mechanism of cognate object selection as outlined here allows for a kind of reversal of the selectional determination within the Chomskyan framework, permitting the verb some power in noun selection.

As far as I can tell, the answer to the third question for the English language is also negative; I know of no construction in which it could be said that a cognate subject or other constituent is found, and I have not found such cases in other languages either (admittedly I have not conducted a careful search). It if is generally true that, say, cognate subjects are not found in

sentences of natural languages, once again it would be very worthwhile to try to find out why. What is "natural" about the cognate object construction which is "unnatural" about the cognate subject one? Although obviously no answer can be given to such a question at this moment, it is nevertheless not an idle one, for answers to such questions as these provide the linguist with sharp insight into the workings of natural language and the marvelous peculiarities of its design.