

Some Thoughts on Linguistic Redundancy

by M. W. S. DE SILVA

Looking at redundancy in relation to what are known as deep structures—depending, of course, upon the definition given to the concept of depth—a transformational generative grammarian would want to derive many redundancy features through transformational rules; to a systemic linguist, statements involving redundancies would exhibit a very high degree of delicacy; anyone working on linguistic universals, dealing with marked and unmarked categories, would find it expedient to handle redundancies in terms of marking. All this is when redundancy features are examined in terms of one descriptive theory or another. In this short paper, however, I do not wish to examine the grammatical status of redundancy, nor its relevance to information theory, but propose to make a few observations on the function of redundancy in the actual use of natural languages.

Elementary notions of redundancy imply that languages have more features, more characteristics, more bits, than are absolutely necessary for communication. One can chop off every third, or fifth, or sixth, or even random ones from among the consonants occurring in a given utterance and still make oneself intelligible. Taking this to an extreme, one is able to distort sound sequences in such a way as to produce riddle-like expressions which are intelligible to those who are aware of the mechanism of distortion employed in each case. The fact that certain sequences are distortable or transferable in a phonological sequence means that the number of phonological elements involved and the order of these elements are full of redundancy. Moving over to grammar, we meet English expressions like “he goes” where singularity is stated twice or Hindi expressions like “meerii laḍkii nahīī aatīī” where feminine gender is stated three times. These are grammatical redundancies. There are other kinds of redundancy which will be introduced later in this paper.

C. F. Hockett, in his *Course in Modern Linguistics*¹ discusses redundancy under four headings: information theory and phonology, the theory of secondary associations, linguistic change and writing systems.² Referring to redundancy at the phonological level, Hockett says, “If all mathematically possible arrangements of the phonemes of a language actually occurred, then the degree of clarity in articulation which would be required to prevent misunderstanding would be humanly impossible . . . There are always stringent limitations on the arrangements in which phonemes occur in utterances, so that there is always a measurable amount of redundancy; rough computation shows that the measure of redundancy in English is approximately 50% and there is some reason to believe that this

1 C. F. Hockett, *A Course in Modern Linguistics*, Macmillan, 1964.

2 Hockett, *op. cit.*, pp. 87ff., 296ff., 421, 440, 456, 540.

figure holds for languages in general".³ In his chapter written to J. Greenberg, ed. *Universals of Language*,⁴ Hockett maintains this 50% hypothesis as one among the half a dozen universals he sets up. Hockett's mode of measurement is not explicit, but even if we accept his 50% hypothesis, that as much as 50% of the elements occurring in linguistic events is redundant is nothing to be bemoaned. Because of redundancy phenomena we are able to comprehend linguistic events without paying any great attention to every articulation that is made.

It is known that many a grammatical category is expounded (or realized) in natural languages in a multiplicity of places, as, for instance, the category of number in "he goes" versus "they go" in English. These are redundant realizations. Certain categories, on the otherhand, are expounded in single slots only, (or, to put it in structuralist terms, there is no evidence in the syntax of the language for setting up such categories). What is important, however, is that where a category is not expounded in a multiplicity of places, other redundancy devices are employed by the language in order, as it were, to redress the balance. In the English expressions "I go, we go, they go", the category of person is not expounded in more than one place and does not constitute redundancy. (In English person on the whole is redundantly realized only marginally as in "I am" as opposed to "he is"). The lack of syntactic differences here within the terms of subject-verb agreement is compensated for by the availability of a cluster of reflexive pronouns which are marked for number and person from among which an obligatory selection has to be made. Thus we have expressions like "I go *my way*", "We go *our way*", "They go *their way*" etc. This is an example of the introduction of redundancy elsewhere in order to compensate for the lack of redundancy realization with respect to a particular category in the language. Unlike in English, gender, number and person in Hindi are redundantly realized, and redundancy phenomena are adequately built into the concordial system in this way. Thus, "meerii laḍkii kaam kartii hai", "meeraa laḍkaa kaam kartaa hai" etc. Redundancy requirement having been taken care of in this way, Hindi does not find it essential to employ a cluster of reflexive pronouns out of which a selection has to be made; the one reflexive *apnaa* suffices. Thus, "meerii laḍkii *apnaa* kaam kartii hai", "meeraa laḍkaa *apnaa* kaam kartaa hai", etc. These examples from English and Hindi illustrate how a balance is maintained in redundancy phenomena: this in fact supports Hockett's view that the amount of redundancy in languages tends to remain static.

To illustrate how a balance is maintained in redundancy phenomena within the internal structure of language a further example is given below from Sinhalese and Hindi. Sinhalese incidentally is a uniquely diglossic language⁵ in which the grammar of the literary variety appears at the surface to be very

3 *op. cit.*, p. 89.

4 J. Greenberg, ed., *Universals of Language*, MIT Press, 1963, pp. 1-22.

5 On Sinhalese Diglossia, see my "Effects of Purism on the Evolution of the Written Language", *Linguistics* 36, 1967;

"Sinhala sāhitya bhāṣāvē anāgataya", *Taranga* 7, 1963;

"Nūtana sinhala gadya sāhityayē bhāṣāva", *Sāhitya Dinaya*, 1963;

and J. W. Gair, "Sinhalese Diglossia", *Anthropological Linguistics*, 10. 8. 1968.

different from that of the colloquial variety. Like Hindi, Literary Sinhalese maintains a redundant realization of such categories as number and person. In Colloquial Sinhalese these categories are not realized in a redundant manner. In other words, Literary Sinhalese is more like Hindi and Colloquial Sinhalese more like English (although this comparison holds good only for the present exemplification) In Literary Sinhalese and in Hindi the concordial relationship between subject and verb is inflectionally indicated in both places, and exhibits redundancy. Like Hindi, (and as to be expected according to the previous paragraphs), Literary Sinhalese employs one reflexive pronoun which is not marked for any particular category. In both languages, then, the use of redundancy in concord is coupled with the use of a neutral reflexive pronoun. The Literary Sinhalese equivalent of Hindi "apnaa" is "tamaagee". (There are other lexical variations of "tamaagee" like "svakiiya" but these have no separate grammatical status). "tamaagee" (or "svakiiya") does not change for gender, number, person, etc. In Colloquial Sinhalese, however, concordial markers are not used in a redundant manner. In this variety the equivalents of "I go, we go, you go, she goes, he goes, they go" etc. have the same verb form. The "loss" or "absence" of redundancy here is compensated for by the introduction of a grammatically varying cluster of reflexive pronouns: this phenomenon is identical with the equivalent phenomenon in English. In Colloquial Sinhalese, the first person, sg. reflexive is "magee" first person, pl. is "apee", second person, sg. is "oyaage", etc. etc. These cases tend to show that in the historical evolution of languages the loss of redundancy in one place or category with reference to a parent language or one language in contact is likely to be reintroduced in another area as a matter of compensation.

The compensatory devices for loss or absence of a particular type of grammatical redundancy are not necessarily brought to bear through other grammatical redundancies alone as the above examples might convey. Such reintroductions may at times be through lexical load. Pali, for instance, has lost a number of grammatical markers that were available in Sanskrit, and from this point of view Pali shows a great deal of grammatical "simplification". This simplification in Pali is, however, compensated for in lexical repetitiveness: notice such expressions as "sandassesī samaadapesī samuttejesī sampahansesī" for "instructed" or "devataa . . . ekamantam aṭṭhaasi ekamantam ṭhitaa kho saa devataa" and so on and so forth. Redundancy burdens may in this way be shifted from grammatical complexes to lexical complexes. It may be mentioned in passing that this phenomenon which may be called *redundancy shifting* is a necessary operation in linguistic history, and that no account of the history of a language is complete which does not provide adequate information on such compensatory shifting. The importance of this phenomenon for the study of universal features in language is obvious.

The foregoing observations pertain to what might be referred to as *structural redundancy* in both descriptive and historical perspectives: although redundancy is a necessity in languages and languages tend to reintroduce lost

redundancies in one way or another, languages also have a tendency to disallow undue redundancy burdens. The factors behind what grammarians call "right hand deletion" are a case in point. The avoidance of the repetition of the word "I" in joining the three sentences "I went home" "I had my supper" "I went to bed" to produce "I went home, had my supper and went to bed" is a simple example of this process of right hand deletion. That the word "I" which is a necessary constituent of each of the three sentences joined together is not repeated in the compounded form is a kind of avoidance of undue redundancy.

Although these cases all go to support the hypothesis that the proportion of redundancy in languages tend to remain static (by means of one device or another), it must be emphasised that this more or less static degree of redundancy is not to be measured with reference to the internal structure of languages alone: Hockett's hypothesis seems to advocate that the static proportion of redundancy operates in terms of the internal structure alone. It seem to me that the redundancy factors in the internal structures of languages can be markedly reduced without violating the efficiency of communication if and when the non-linguistic correlations of the utterance such as the situational context are clearly indicated. In this way, redundancy seems to fluctuate not only within the internal layers of language but also between internal and external layers relevant to the speech situation or speech act in question. Examples to illustrate this view point will follow.

Creole languages are good examples to illustrate the warfare between two opposing tendencies in languages, one to achieve efficiency and the other to achieve understandability. As J. Lyons has pointed out in his *Introduction to Theoretical Linguistics*, "Languages, as they develop through time and "evolve" to meet the changing needs of the societies that use them for communication, can be regarded as homeostatic (or self-regulating) systems, the state of a language at any time being "regulated" by two opposing principles. The first of these (which is sometimes referred to as the principle of "least effort") is the tendency to maximize the efficiency of the system . . . : its effect is to bring the syntagmatic length of words and utterances closer to the theoretical ideal. The other principle is "the desire to be understood": it inhibits the shortening effect of the principle of "least effort" by introducing redundancy at various levels. As a consequence, it is to be expected that the changing conditions of communication will tend to keep the two tendencies in balance".⁶

I now wish to move on for a moment to newspaper Headlines as a particular variety of language. Newspaper headlines are good examples to illustrate the zero-information content of a number of grammatical categories or terms in grammatical categories in certain situations. The non-use of the past tense form or the progressive form of the verb in English headlines is no

6 L. Lyons, *Introduction to Theoretical Linguistics*, CUP 1968, p. 90

impedement to the understanding of the substances of a given line. When, in dealing with the topic of Britain's entry into the European Common Market, headlines often say "De Gaulle says Non", we refrain from interpreting this sentence as a habitual present tense form (notwithstanding the habitual nature of "non" in the case of this particular gentleman!). It is true that all newspaper headlines are not unambiguous in isolation even within the context of Headlines. The paragraphs that follow the headline act as situational constructs which enable the reader to disambiguate the purport of the headline. "Flying plane kills doctor" (Times, U. K., Nov. 27, 1968) is an example of ambiguous headlines. Such headlines are, however, disambiguated by the paragraphs that follow. I must hasten to add that only a small proportion of headlines are actually ambiguous in this way. The point I am trying to make is that when situational constructs are designed or the situation is made clear, grammatical markers of natural languages, especially those with discontinuous realizations (and others), become less significant for the information out-put, and thereby take the shape of redundancies of yet another type.

In British Universities, applicants for admission as undergraduates are interviewed by the respective departments in order to assess their suitability to the course involved. During the past five years I have had the occasion to interview several hundred such candidates who wished to major in linguistics. I have a stock question which I ask every one mainly to test their wits (though not to test their suitability to read linguistics, which I test by saner means). I tell them, an Eskimo who knows very little English comes to the University Campus and sees me in front of my department building. He sees me on three occasions and speaks to me each time. The first time he walks to me and says "me town going bus where stop tell"; the second time he comes to me and says "me brother rat kill"; the third time he walks to me, shows me a dead rat and says "me brother rat kill". These are three different occasions, and each time I am a stranger to the Eskimo. I then ask the candidate "what did he mean by "me town going bus where stop tell"? Every candidate says that I must have shown him to the bus stop. Then I ask the meaning of the second utterance, "me brother rat kill". Most candidates say that the Eskimo's brother has killed a rat; very few would say that it is ambiguous and might also mean that a rat killed his brother. Then I give the third utterance, "me brother rat kill", made to the accompaniment of the holding out of a dead rat. They all agree that Eskimo's brother has killed this rat that is being exhibited.

There are many factors which make the "but-stop" utterance comprehensible. Likewise the "rat kill" utterance made with a dead rat in hand is more comprehensible than the same without a dead rat in hand. It may not be as clear as the "but-stop" utterance but that everyone agreed on its meaning is significant. One factor that makes those two sentences comprehensible (one way or another) is the presence of situational information. By "situational information" I mean clues that allow one to make one and only one situational construct. In sentence 2, which is "me brother rat kill" which allows no situational information, one is at a loss to comprehend the information that the Eskimo is attempting to convey. Did his brother kill a rat? (If so, what?). Or, was his

brother killed by a rat? (a story worth listening to!). Maybe it is this latter he was attempting to say. But, none of my candidates thought of it as a possible interpretation. The moral of this exercise is as follows: in interpreting ungrammatical utterances which are low in situational information, people tend to map them with grammatical ones so as to allow the ungrammatical aspects to emerge as the smallest possible distortion of the grammatical. Notwithstanding the use or non-use of deictics, the nearest grammatical order of "me brother rat kill" is "me brother kill rat" but not "rat kill me brother" etc. etc. A reconstruction of an ungrammatical sentence in this way so as to map it with the nearest possible grammatical does not always yield the desired information. It is a truism that understanding depends on how the listener constructs his own sentences corresponding to the speaker's sentences; but, where the utterance is grammatically deviant, such reconstructing and mapping is not sufficient, and the understanding depends largely on unambiguous situational information. When the situational information is sufficiently provided, the mapping of the ungrammatical with the listener's grammatical becomes possible and accurate. The corollary is equally true: from the speaker's point of view, when the situation is loaded, most grammatical markers become redundant with reference to situation. Where the situation is not clear, most, if not all, of the grammar is imperative for the conveyance of the correct information. Although, with a loaded situational information content, grammatical devices become redundant, the reverse is, however, not true, for many a perfectly grammatical sentence can be ambiguous if the situational information is nil. From this we may generalize that extra-linguistic correlation take precedence over intra-linguistic relations in the distribution and determination of redundancy phenomena. If one sets up a hierarchy as situation, grammar, phonology, one may even generalize that once a higher level in the hierarchy is loaded the markers of the level below tend to be redundant. It is because of loaded situational criteria that baby-talk, foreigner-talk, and indeed all varieties of pidgins and creoles have become intelligible language types. This phenomena may be named *inter-level redundancy*.

In the foregoing paragraphs I have discussed a number of aspects of redundancy and its function: *structural redundancy obtaining at the phonological and grammatical levels, the relevance of redundancy to communication, the nature (and significance) of redundancy shifting within each level as well as between levels in order to maintain a more or less static redundancy load, and the implications of redundancy in terms of the hierarchy of levels of linguistic description*. In conclusion, I propose to make some observations on *the function of redundancy as a mark of varieties differentiation*. I shall examine this with reference to diglossic behaviour, of which Sinhalese is the best example. It is widely known that Literary Sinhalese is different from Colloquial Sinhalese in both phonology and grammar, although these differences may be handled as belonging to the surface only. The general tendency is not to use the grammar of the colloquial language in serious writings. I shall attempt to show how the special characteristics of the literary language may be interpreted as redundancy phenomena with reference to the spoken variety of the language.

Viewed in its own rights Literary Sinhalese has a case for being treated as having the categories of gender, number and person as syntactically relevant categories. From a structuralist standpoint, an independent analysis of the spoken language does not allow for stating these categories in that way. These are in fact only some of the categorial differences between the two varieties. But there is an interesting point to remember which is very relevant to the present thesis: although Literary Sinhalese is never spoken except when reading, even the most illiterate speakers of Sinhalese are able to listen to and understand the language of, say, newspapers, without being impeded by the grammatical peculiarities of the type mentioned above. It is assumed that this intercomprehensibility is brought about by the redundant nature of the apparently extra categories in the literary variety. If the two varieties are viewed as varieties of the same language, then these extra characteristics in the literary variety may be interpreted not as extra categories but as extra exponents of the same categories: from this point of view, these extra exponents of categories, or extra markers, may be interpreted as redundancy factors.

Where two varieties are socially accepted as varieties of the same language, with the intercomprehensibility factor mentioned above, one begins by identifying both varieties as varieties of the same language. Instead of setting up two grammatical systems, then, the same grammatical categories will naturally be set up for both varieties: these categories are, then, realized (or expounded) in a smaller number of markers in one variety (Colloquial, in this case) and in a greater number of markers in the other (Literary, in this case). The factor of redundancy operates in the variety in which the categories have a greater number of exponential correspondences. As far as Sinhalese is concerned, most categories are uni-exponential in colloquial and multi-exponential in literary. It is because the multi-exponential realization of categories in the literary language involves redundancy that the grammatical shape does not impede the understanding of literary texts.

In treating varieties which are grammatically different on the surface as varieties of the same language, then, the factor of redundancy is stated as a demarcative characteristic. This introduces a new dimension of redundancy. The notion that redundancy operates at a given level is an old one and has been applied to the study of particular languages, particular varieties, at particular times. Here I am introducing redundancy as a mark of varieties differentiation. In diglossic situations, the apparent access in number of the membership of categories or in the exponential criteria, obtaining in one variety as opposed to the other which is more restricted on those lines are, then, redundancy criteria from a communication point of view.

In treating Literary Sinhalese as characterised by redundancy in exponential criteria I have assumed that the spoken language is the norm or the *grid* in terms of which the literary language is understood. This does not, however, mean that the "shorter" grammar is always the norm. If the understanding of the variety with a shorter grammar depends on an added supply of situational criteria that variety is not the grid. From the point of view of many other

varieties of English the grammar of the language of newspaper headlines is shorter, but the meaning of the headline rests largely on the exposition that follows which contains situational constructs for the headlines. Headlines is therefore not the grid variety in English.

The concept of a grid variety in terms of which other varieties are to be comprehended grammatically is useful if such varieties are viewed as constituting the same language. I view literary and colloquial Sinhalese as varieties of the same language: these two varieties are socially accepted as belonging to the same language and they are mutually comprehensible; even the illiterate people who cannot perform in the literary language are able to listen to and understand it without being impeded by the grammatical peculiarities involved. I admit it is not an easy task to measure understanding: I believe understanding is negative in so far as no positive measurement of accuracy in understanding is possible. Positive measurement can only be made of misunderstanding, for areas of misunderstanding are more easily accessible, and understanding has to be diagnosed through misunderstanding. In some one thousand cases of tests conducted with a view to observing if illiterate people misunderstood literary texts (from newspapers and novels) because of the grammatical peculiarities involved I had no clue whatsoever to prove that these peculiarities impeded understanding. Notice that I leave lexis untouched: lexical differences are many between any two varieties. Any lack of understanding in my test cases was due to lexical difficulties, and once more common lexical items were used in place of the more Sanskritic or unknown ones that purport of the text passages was understood by every person tested: they were able to answer questions based on the passages read to them not more than two, or at most three, times. Newspaper headlines, on the other hand, consist of a grammatically reduced variety in relation to the grid variety. Where a variety is grammatically reduced, extra situational criteria are brought in for communication. This is not so in the case of Colloquial Sinhalese. The redundancies in the literary variety are *not* created by the need for constructing deliberate situations on paper.

It seems to me to be correct to assume that in comprehending the literary form in languages like Sinhalese, people ignore the "added" redundancy phenomena in that form of the language. From this point of view and from the standpoint of the spoken language the extra grammatical criteria in the literary variety may be regarded as superimposed characteristics. These theoretically superimposed and thereby really redundant phenomena in the literary language constitutes the "second dimension of redundancy", namely, its demarcative function. They constitute a second dimension because they are over and above the normal characteristics of redundancy which obtain in any variety of language: in terms of the "grid" they are in effect "created" redundancies.

The increased redundancy factor in Literary Sinhalese adds little to the grammatical mechanism which the Sinhalese person employs in understanding his language. It is my conjecture that in diglossia, the high variety as a rule presents an increased grammatical (systemic) complex while the low variety

presents an increased lexical (set) complex. I do not however propose to go into this notion in the present paper. The concept of redundancy is used here to refer to the increased systemic complexes and not to the set complexes.

Redundancy factor between varieties creates hybridisms. Just as two languages coming into contact with one another can result in the creation of pidgins and later creoles, even so two distinct forms of diglossia can invariably create a third variety, or more exactly, a number of third varieties, which constitute pidgin-like or creole-like forms, the properties of which are not systematically statable. This kind of hybridism created by people in their attempts to perform in the high variety without having gained a full mastery of it is known in Arabic as well as in Sinhalese. It is neither the low variety nor the orthodox high, and has a tendency to fluctuate so heavily that no systematic statement can be made of it. Hybridism both in diglossia as well as in creole formation is very much linked with the redundancy factor. It is the result of attempts to perform in varieties which have redundancy phenomena in terms of the grid variety of the individual or the community. The process involved is that the redundancy phenomena are partially observed and partially not, now observed here and now observed there, now coinciding with orthodox high and now violating it. Here again I refer to the grammatical and phonological redundancies between varieties, and not to the lexical distinctions.