Zellig Harris's theory of syntax and linguistic reflexivity

Philippe De Brabanter

Zellig Harris's theory of syntax and linguistic reflexivity

Though Zellig Harris was originally one of the leading names in the transformational 'revolution' that rocked the linguistic world in the 1950s, his theories did not meet with the same success as Chomsky's. In particular, Harris did not attract a large following in the later years of his activity as a linguist. To my knowledge, introductions into Harris's personal conception of grammar are few and far between. Hiz (1994) and Matthews (1999), both of which are obituaries, provide a useful overview of his biography and theories.[1] Harris (1988) is the closest one has to an introductory exposition by the linguist himself, but that book already makes for a taxing reading experience.

My interest in Harris stems from a concern with the analysis of the metalinguistic dimension of discourse. A fairly large body of literature has been devoted to metalanguage, especially in logic and the philosophy of language. Most of it is to do with artificial metalanguages for the description of formalised symbolic systems (e.g. Carnap 1937; Tarski 1944), or with the use/mention distinction (see Saka 1998 for a bibliography). Very little has been written about the syntax of metalinguistic use in natural language; even linguists seem not to have been too taken with the subject. In this respect, Harris constitutes a welcome exception. Furthermore, once you have taken the trouble to immerse yourself in his system, reading Harris becomes a rewarding experience. Here is a linguist who, despite appearances, strives for simplicity — his grammar depends on a handful of fundamental principles — and who methodically justifies every step of his reflections. In a nutshell, here is a linguist from whom a lot can be learned.

In the next few pages, we shall see how Harris builds a full-fledged grammar of a natural language (henceforth often 'L_n'). First, he provides means of identifying what are the basic units of L_n, i.e. its phonemes and morphemes. Then, he shows how these units can combine to form larger units, i.e. phrases and sentences, all of which is achieved on the basis of a single principle. When that has been done, it is useful to put the scheme to the test. To that end, I have chosen to examine to what extent it is capable of accounting for some of the notorious difficulties that stem from metalinguistic use.

---

[1] See also Gross's introduction to Harris (1976) and the entry in *Encyclopedia Universalis, Thesaurus Index*, vol. 2 (1996: 1649).

<u>Identifying the units of L$_n$</u>

Harris's whole method of doing syntax is rooted in the observation that natural languages have no external metalanguage (1968: 17; 1988: 3; 1991: 31-32). This simple thought has far-reaching consequences. In logic and mathematics, "[t]he statements that describe a system are in a different system, called the metalanguage [...] which is richer in certain respects than the system it is describing" (1991: 32). The elements of logic and mathematics are determined which such precision that one can readily distinguish statements about the field ('meta-statements') from statements belonging to the field (e.g. mathematical formulas). By contrast, with respect to natural languages, "we have no different system in which the elements and combinations of language can be identified and described" (1991: 32). As a result, any system that we choose to describe the elements and structures of a natural language must make use of elements and combinations that are essentially similar to the language described (1988: 3; 1991: 32).

This means that Harris is going to be looking for a characterisation of the language that is internal to it: the grammar, then, is just a subset of the set of sentences that make up a full-fledged natural language. Harris observes that the linguist's task would be impossible were it not for one characteristic of natural languages, namely that they exhibit 'departures from equiprobability', i.e. from randomness. By this, Harris means that not all combinations of discrete elements are equally likely to occur, and indeed some are downright impossible. For instance, such combinations of phonemes as [ktfnp] do not occur in English (or, presumably, in any other language). By the same token, English rules out such combinations of morphemes as *no the here yes go*.[2] What a decent grammar must do is to bring out the rules or regularities that govern departures from equiprobability.

This still begs the question as to how one can identify the basic building blocks that are phonemes and morphemes. Given that there is no external language in which these elements could be catalogued, what are the internal procedures yielding reliable lists? Harris argues that the list of the phonemes of a natural language can be established for instance by means of the so-called 'pair test' (cf. 1968: 21-23), a type of experiment involving two members of a single language community. The first one, the speaker, repeats at random each of two sequences that are felt to be similar, e.g. *roll* and *role*, or *cart* and *card*. The second, the hearer, is requested to

---

[2] Note, however, that tokens of these impossible sequences *have* just been produced. But such tokens occur only in metalinguistic discourse.

guess, utterance after utterance, which sequence was being pronounced. Harris notes that for the first pair, about 50% of the guesses will turn out to be correct — which indicates that the hearer could make out no pronunciation difference between the two words so that he or she responded at random — whereas, for the second, there will be close to 100% of accurate responses — indicating that a 'phonemic' difference is detected by the hearer.

Now that phonemes have been identified, so can morphemes. This is done on the basis of a stochastic method (cf. 1968: 24-28) brought to bear on a fairly large corpus of sentences. This process is used to evaluate the number and range of phonemes that can follow after a particular phonemic sequence. It turns out that the peaks — the points at which the number of possible successors is high, i.e. does not significantly depart from randomness — signal morphemic boundaries: at the end of (what can therefore be recognised as) a morpheme, there is a much greater freedom for the selection of the successor phoneme than, say, in the middle of a morpheme. Given the British English sequence [rɪ»memb´], the choice of the next phoneme is almost entirely free. By contrast, after only [rɪ»memb], the range of possible successors is much more limited (see also Hiz 1994: 521; Matthews 1999: 117). This process can also be stated in terms of the 'predictability' of the following phoneme: the points at which predictability is lowest are likely to be morpheme boundaries. Note that the recourse to stochastic processes is motivated by Harris's reluctance to use a meaning-based criterion. In any case, stochastic processes have a wider applicability, since they lend themselves also to the study of languages whose semantics is not well-known to the linguist. Besides, Harris claims that their validity is confirmed by the fact that they produce results that match native speakers' experiential knowledge of what is a word in their own language (cf. 1988: 6).

Combining the units of Lₙ

At this stage, Harris has provided means of identifying the discrete elements that can be combined to form larger units in the language. But what principle governs these combinations, and allows speakers to distinguish between acceptable and unacceptable sequences? Harris's answer is fairly straightforward: syntax depends entirely on one <u>relation</u> whose presence can be made out in every complex arrangement of the language, in particular in sentences.

But, before I attempt to characterise this relation, a distinction needs to be drawn between two sets of strings in any Lₙ, the so-called 'base' sentences and 'reduced' sentences. In a nutshell,

base sentences (or 'kernel' sentences in earlier writings) are those from which all the actual or possible sentences of a given L_n can be obtained through transformations.[3] This latter set is that of reduced sentences, so called because, especially in Harris (1988) and (1991), transformations are essentially 'reductions' (including 'zeroings' of certain elements). In Harris's scheme, reductions, which are defined as "changes [i.e. alterations of the sound shape] in word-occurrences, not recastings of the whole sentence" (1991: 109), affect word forms, not abstract structures.

Now, as I announced above, Harris claims that there is a universal principle regulating sentence formation, a principle which is another variation on the departure-from-randomness motif. Harris describes it as a 'dependence on dependence' between 'operators' and 'arguments', that is to say, between the two primary types of words that he distinguishes in the lexicon of any L_n. Every grammatical sentence of a natural language conforms to this principle, either overtly or implicitly. It is on this basis that the division between the base set and the reduced set relies. Base sentences are those whose operators (there may be just one) are <u>overtly</u> accompanied by their requisite arguments (cf. 1991: 54). Reduced sentences are those where the combinations between operators and arguments are not entirely explicit. This means that *Kenneth eats rubbish* is a base sentence, whereas *Maud was eating* is not: its Object-NP, though reconstructible, is not present.[4]

As a result of the universal prevalence of the dependence-on-dependence principle, any sequence that aspires to the status of grammatical sentence must have all its operators satisfied by the right kind of arguments. This is a necessary condition for all sentences of a natural language. Assessment is straightforward in the base set, since satisfaction must be explicit. But it is not in the reduced set. There, the criterion for grammaticality is the existence of a path (an ordered sequence of transformations) that leads from the reduced sentence back to a grammatical source-sentence in the base. Such a path *ipso facto* demonstrates that the dependences on dependences in the reduced sentence were indeed satisfied.

Let us now take a very simple example:

(1) John walks.

---

[3] Though Harris's grammar is a 'transformational theory', it is essentially different from Chomsky's in that it does not distinguish between a surface and a deep level: transformations alter word forms as they appear in sentences: there are no 'abstract' modules through which forms are processed before they finally surface.

[4] In the base set one finds only affirmative sentences in the present tense of the indicative mood. Moreover, most of the words used in them are monomorphemic and affixless. That is because Harris analyses most affixes as reductions from free morphemes appearing in base sentences. This holds, notably, for the past tense, the perfect, and the plural.

This sentence is made up of an operator, *walks*, which takes only one argument to be satisfied. This argument, *John*, is itself already satisfied, in the sense that it requires nothing (1988, 1991: *passim*; Harris uses the term 'null'). This way, the two dependences — one of which is a zero-dependence — are satisfied, and the sentence is grammatical. Two more examples:

(2) Fred wears a coat

(3) That Joan hates John is unlikely

In (2), *wears* is combined with the two arguments it requires. These, being like *John* in (1), are also satisfied. In (3), the situation is slightly different: *is unlikely* is satisfied by the presence of the argument *hates*. However, this argument, being an operator, does not require null: it needs in turn to be satisfied by two words like *Joan* and *John*, which, being zero-level arguments, need no further satisfying.[5]

These three examples suggest that the lexicon subdivides into three subsets rather than two: First comes the set of 'zero-level' arguments. Central in this *N* set — with *N* for *null* — are basic nouns like *John*, *table*, or *frog*. Next are the first-level elements, namely operators which, in a sentential context, require the presence of one or more zero-level elements. This set subdivides into O_n (*walk, arrival, tall*), O_{nn} (*wear, father, identical*), and O_{nnn} (*give*), according to the number of zero-level arguments they need to combine with. Third, the second-level elements, i.e. operators which, in a sentential context, require the presence of at least one first-level element (with, as the case may be, one or more zero-level arguments). This set can be split up into O_o (*likely*), O_{oo} (*entail*), O_{no} (*assert*), and O_{nno} (*tell*).[6] Note in passing that this type of categorisation is in principle applicable universally and therefore supersedes the division into parts of speech recognised by traditional grammar.

On the basis of this classification of lexemes, one can characterise an essential difference between the following pairs of sentences: *John plays violin and Mary plays piano* vs. *John plays violin and Mary piano*; *Sam was eating something* vs. *Sam was eating*, *I'm expecting John to come* vs. *I'm expecting John*; *John can for John to swim* vs. *John can swim*.[7] The first of each pair is a base sentence, while the second is reduced. In the first example, the base sentence needs to include the repetition of *plays* for otherwise the two *N* arguments *Mary* and *piano* would be

---

[5] The reader is asked to forgive me for glossing over the presence of the definite article and the subordinator *that*.
[6] The various pairs and triplets of indices are ordered.
[7] Matthews rightly emphasises that, in Harris's scheme, "the forms of words which speakers use are ordinarily reduced from other forms of words that, to varying degrees, *they do not use*" (1999: 115; emphasis mine).

deprived of an operator. In the second, the base sentence must include a second *N* for the operator *eat*, which is an O_{nn}. In the third, the base sentence must include a second argument for *expect* which is itself an operator, given that *expect* is an O_{no}. In the fourth, the base sentence must include an *N* for *swim*, failing which this operator is not overtly satisfied.

Each second member in the pairs above is obtained from the first through one or more transformations, essentially reductions. It is easy to accept that these reductions do not affect the information conveyed by the sentences, provided one understands information in a sense similar to Harris, i.e. essentially as something like the propositional content of sentences (cf. Lyons 1995). Harris (1991: 130-31) briefly shows how *A small boy disappeared* can be obtained from two base sentences, one interrupting the other:[8]

A boy — a boy is small — disappeared.
A boy who is small disappeared. (the immediate repetition of the NP allows the application of
    a *wh*-operator)
A small boy disappeared. (the *wh*-pronoun and the copula are zeroed)[9]

The information is not affected by the derivation. Neither, for that matter, is acceptability. If the reduced sentence had been the barely acceptable *A liquid table laughed*, its base and the various transforms would have exhibited the same degree of near-unacceptability too.

Finally, a word is in order regarding the conditions under which reductions, especially zeroings, can be carried out. In the pairs of sentences above, only easily predictable items have been reduced or zeroed. High predictability generally goes together with low information. Partial evidence of this is provided by the fact that English speakers who were given the second members of the pairs above would be in no trouble to restore the zeroed items. If these had carried much information, that task would have proved impossible (cf. 1991: 83ff; 94f; and *passim*). This remark concludes my rapid overview of Harris's syntax. We are now more or less equipped with the notions required for a meaningful appraisal of Harris's discussion of metalanguage.

---

[8] Clearly, several steps are skipped, and some difficulties sidestepped; but this is meant as a mere illustration.
[9] The zeroing of *wh- is* is described in Harris (1991: 89-90).

Harris on metalanguage

Metalinguistic discourse provides an excellent testing ground for evaluating the merits of Harris's syntax. In particular, sentences containing mentioned or quoted items present linguists with challenging peculiarities. Thus, quotable items extend far beyond the standard lexicon used in non-metalinguistic use. Furthermore, quoted sequences exhibit a quite idiosyncratic morphosyntactic behaviour, as they tend to be invariable in number (and gender and case where these are marked). A case in point is this logician's pseudo-paradox, highlighted by Josette Rey-Debove (1978: 67): "Dans /*Table* est un nom féminin/, /*table*/ est un nom masculin". Is Harris's theory capable of handling this? Can it also deal with what some linguists have called 'mixed uses', i.e. sentences which are simultaneously about a state of affairs in extralinguistic reality *and* about language?[10]

In the next few pages , I shall try to assess how well Harris's syntax takes care of these tricky issues. But the first question to ask is: does it leave any room for metalinguistic sentences in general? In other words, can it generate the subset of sentences which make up the metalanguage of L_n? Remember that this is an elementary requirement, given that L_n's metalanguage is part of L_n.

Harris has no trouble handling this. In English, he writes,

all metalinguistic sentences contain transforms of the sentence form *X is a sentence*, *X is a word*, *X is a linguistic form of English*, etc., also *'X' is a sentence*, etc. (1968: 125)

This postulate, whose validity will be partially assessed below, implies that all the metalinguistic sentences of L_n are derived from a source sentence in the base which is (or includes, if it is complex) a sentence modelled on one of the patterns illustrated in the citation. The predicates in these examples belong to the class of predicates labelled *is N,,*, which is a subclass of *is N*, in which one finds classifying predicates such as *is a mammal*, *is a book*, etc. Note that this characterisation provides Harris with a formal criterion for the so-called 'mention' of words (cf. Quine 1940: 23-26): an *X* is mentioned if and only if it occurs as the argument of an *is N,,* operator (in the case of base sentences), or if the derivation that yielded the reduced sentence in which it occurs contains a base sentence on the pattern *X is N,,*.

---

[10] Their commonness has been pointed out by several language philosophers and they have been the focus of close scrutiny from such writers as Recanati (1979) and Rey-Debove (1978).

On this simple basis, Harris appears to be able to account for various peculiarities of metalinguistic sentences.

— First, Harris recognises the wide variety of objects that can be mentioned or quoted.[11] Basically any sound (or any sequence of letters) can be used autonymously. This stands in contrast with most of the non-metalinguistic sentences of English, whose subject must be nominal or nominalised. This observation prompts Harris to define more precisely the set of metalinguistic predicates. Indeed, given that any sound can be mentioned or quoted, such predicates as *is a sound* and *is a noise* must also be included in *is $N_{,,}$*, although they are not strictly metalinguistic. Therefore, metalinguistic predicates proper will be said to make up *is N_{meta}*, a subset of *is $N_{,,}$*. Also, metalinguistic classifiers will be said to belong to N_{meta}, which is a subclass of $N_{,,}$. Non-metalinguistic quoted items, such as the song of a bird, will be dealt with in terms of the non-metalinguistic predicates in *is $N_{,,}$*.

— To proceed with a rather simple point, some philosophers (e.g. Saka 1998) have made much of a difference between the straightforward (unmarked) mention of a graphemic or phonemic sequence and its quotation with the help of quotation marks, italics, or similar means. Though it is unlikely that the distinction is systematically observed, it needs to be taken into consideration, failing which no role, morphosyntactic or semantic, can be ascribed to markers of quotation. In Harris's scheme, mention becomes quotation through the application of a quoting operator[12] which turns the first of the next two pairs of sentences into the second:

He went is a sentence/Mary is a word.
'He went' is a sentence/'Mary' is a word. (1968: 125)

— I pointed out above some of the morphosyntactic peculiarities of autonyms. This is something that has not escaped Harris's attention. He claims to be able to give reasons for the usually singular number of autonyms. To account for the contrastive pair of sentences *Bookworms is on p. 137 in this dictionary* and *Bookworms are all over in this dictionary*, one need only postulate

---

[11] In the rest of this paper, I shall often use the word *autonym* (and its derivatives) as an umbrella term for *mention* and *quotation*. The term is borrowed from Carnap's *Logical Theory of Syntax*. It has been put to all sorts of interesting uses in Rey-Debove (1978).

[12] This 'quoting operator' — the label is mine — is used in order to reflect what Harris recognises as a "sentential or other intonation", which generates only "morphophonemic variation" (1968: 125).

the zeroing of the singular presenter *The word* or *The phoneme-sequence* in the first sentence, as opposed to the plural *The objects referred to by the word ...* in the second (cf. 1991: 136).[13]

The deviant morphosyntactic behaviour of autonyms comes out most clearly in the case of nouns, where not only number but also gender and/or case may be affected. Harris, all of whose examples are in English, does not discuss gender and case. However, it is not too difficult to understand how his theory would handle them. For instance, in Rey-Debove's "Dans /*Table* est un nom féminin/, /*table*/ est un nom masculin", the autonym *table* could be said to be reduced from *le nom table*, and would therefore receive its masculine gender from the masculine head of the NP, i.e. *nom*.[14] A similar reasoning could be applied to case.

— Many linguists assume that there may be metalinguistic sentences containing no autonyms (e.g. *The word has four letters*). We saw above that Harris postulates that every autonymous use can be traced back to an X *is* N_{meta} sentence. As a matter of fact, Harris extends this criterion to every single metalinguistic sentence. Now, does that mean that his grammar does not allow for 'autonym-less' metalinguistic sentences? The answer is *No*. Here again, the distinction between base and reduced sentences is relevant. In the base, every metalinguistic sentence contains at least one argument which is the 'name' of a phoneme-sequence (i.e. at least one autonym) just as it contains an operator of the *is* N_{meta} set. In the set of reduced sentences, either one, the autonymous sequence or the meta-operator, may have disappeared. In the next paragraph, we shall consider an instance of zeroing of the meta-operator. What happens in the case of metalinguistic sentences without autonyms is just the mirror image of that process: these sentences are reduced from base sentences including mention, with zeroing of the autonymous sequence.

This way, it is possible to account for *The word has four letters* or *English sentences contain verbs* as metalinguistic sentences. Note, however, that zeroing applies more straightforwardly to general sentences like the second in the pair: what is zeroed here is basically the disjunction of all the sentences and all the verbs in English (cf. 1968: 126), which is clearly no more informative than quantifiers like *any* or *some*, i.e. hardly at all. On the other hand, the particular word to be zeroed in the former sentence cannot be said to be minimally informative. Thus, on the sole basis of that sentence, it would be impossible for any English speaker to restore the zeroed item, except

---

[13] In his introduction to Harris (1976), Maurice Gross assumes that there is a general agreement that such sentences as *Dieu a quatre lettres* are obtained from *Le mot Dieu a quatre lettres* through reduction, or the application of an equivalent operator (cf. Harris 1976: 9).

[14] But perhaps things are not that uncomplicated. Harris has a soft spot for the presenter *The phonemic sequence*. Yet, in French, *La séquence phonémique* is feminine, whereas French autonyms are always masculine.

by pure chance. This seeming problem for the theory, however, disappears as soon as one realises that a sentence like *The word has four letters* is most likely to occur in conjunction with a sentence containing the zeroed word. That is presumably what Harris has in mind, even though his treatment of this point is very succinct and provides no definite answer.

— Let us now see if Harris's syntax has anything convincing to say about 'mixed uses'. Sentences of this kind contain at least one sequence which, though it performs its ordinary role, is simultaneously contemplated as a piece of language. Such is *'intelligentsia'* in *He is of the 'intelligentsia'*: As the head of an NP within a PP-complement of *is*, it has its run-of-the-mill reference to those members of society who are most educated and think up new ideas. At the same time, however, the sequence between quotation marks also makes a comment about the word *intelligentsia* or about its utterance, a comment whose interpretation may vary with the context of utterance and the speaker's intentions (e.g. "this is not the right word"; "this is the word others use"; "this is an overstatement"; etc.).

Harris (1968: 125fn) briefly discusses the sentence *He is of the 'intelligentsia'*, and in so doing provides us with a basic insight into how such a sentence could be derived:

He is of X. X is called the intelligentsia.
He is of X. X is called the 'intelligentsia'. (through application of the quoting operator)
He is of what is called the 'intelligentsia'. (through application of the *wh*-operator)
He is of the 'intelligentsia'. (through zeroing)

Remember that zeroing applies only to highly predictable, minimally informative items. That condition is met here. The quoting operator can only be applied to sequences which are arguments of an *is N_{meta}* operator. This means that, whenever there is a quoting operator, the *is N_{meta}* operator is weakly informative (since the quoting operator necessarily indicates the presence of an *is N_{meta}* operator in the base sentence). Therefore this operator is more or less redundant and can be zeroed after the quoting operator has acted (cf. 1968: 126). Now, I have not listed all the transformations performed, and Harris's own account is even more embryonic. Nowhere is there any justification of why *intelligentsia* rather than *the intelligentsia* comes to be put within quotes. A solution, however, emerges if one considers that the quoting transformation must presumably be applied very early in a derivation. If the quoting operator is made to act on sentence-patterns like *X is a name* or *X is a word*, then perhaps we can justify the scope of the quotes in the above example. The justification might take the form of the following derivation:

We call X by a name. The name is intelligentsia.

We call X by a name. The name is 'intelligentsia'. (the quoting operator would apply here)

We call X by a name which is 'intelligentsia'. (application of the *wh*-operator)

We call X by the name 'intelligentsia'. (zeroing)

We call X 'intelligentsia'. (zeroing)

Assuming that the three sentences *He is of X*, *We call X by a name*, *The name is intelligentsia* are compounded in the base — yielding a complex base sentence —, and that the above transformations (plus passivisation) are performed, one should be able to arrive at the reduced form *He is of the 'intelligentsia'*.

The metalinguistic apparatus of language

Harris (1968) identifies different types of metalinguistic sentences, in increasing order of complexity: 'metatype sentences', 'metatoken sentences', and 'metasentences'. Metatype sentences are built on the simple pattern, *'X' is N_{meta}*. Metatoken sentences are of the form, *a, 'q' in 'X' is N_{meta}*, where *a* indicates *q*'s position within *X*. An example is *The word 'book' in word-position 2 of 'the book' is a noun, but 'book' in word-position 3 of 'They will book him' is a verb* (1968: 127). Harris states that these patterns must necessarily be used whenever we want to talk about linguistic types and tokens, respectively. In so doing, Harris shows his awareness of a fact that is often overlooked, i.e. that the widespread view according to which autonyms 'refer to themselves' is not strictly correct: an autonymous token occurring in a sentence seldom actually refers to itself. Mostly, it does to one or more other tokens, or to its type. But not only is Harris aware of this complication, he also supplies a criterion for deciding if a given utterance talks about a type or a token. This is more than welcome, as experience demonstrates that, failing such a test, it is often very difficult to decide one way or the other. Unfortunately, Harris's test cannot itself be put to the test within the confines of this paper.

Harris identifies a third variety of metalinguistic sentence, the so-called 'metasentence', which is "a metatoken sentence about S1 which is adjoined to S1" (1968: 128). Contrary to metatoken and metatype sentences, metasentences are not usually encountered in actual productions by native speakers. Rather, they are a theoretical construct of Harris's, one which plays an essential role in his syntactic theory, in conformity with the observation that a natural language contains its metalanguage. The idea is to avoid devising a complex metasystem for grammatical description:

with the help of the operator for coordination, one can adjoin metasentences to any empirically observed or observable sentence of an L_n:

> [...] all sentences can be thought of as originally carrying metalinguistic adjunctions which state all the structural relations and word meanings necessary for understanding the sentence, these being zeroed if presumed known to the hearer. [...] We can thus append to a sentence in a language all the metalinguistic statements necessary for accepting and understanding it, with the whole still being a sentence of that language (1991: 127; also 1988- 70-72)

The metalinguistic descriptions provided by the metasentences of S_1 also stipulate all the transformations that have been necessary to arrive at S_1, or, if S_1 is a base sentence, they indicate which elements are reducible or zeroable. Harris's point is that the complete set of metasentences constitutes a grammar of L_n, stating as it does all the elements that can occur in sentences of L_n and all the grammatical operations that can be performed in L_n. This means that, given the base set of sentences of L_n and a metasentence-based grammar, it should theoretically be possible to generate all its reduced sentences, that is, the whole variety of sentences that are actual or potential productions by speakers. In the two-tiered terminology of transformational generative syntax they would be called 'surface' structures. But Harris's reliance on metasentences means that he does not have to postulate the existence of distinct levels in the grammar.

The metasentences are mostly only implicitly present. Such implicitness reflects the generally non-reflexive knowledge that speakers have of the grammatical characteristics of the sentences they utter. Most speakers would be at a loss to enumerate the putative operations they have carried out to generate a reduced sentence from its source in the base. Furthermore, they usually also could not define most of the terms that occur in the appended metasentences. Let us illustrate this by means of the metasentences that would be adjoined to *John likes to read*:

> John likes to read; in this sentence 'John' is a word in *N*, 'like' is a word in O_{no} with ordered arguments 'John' and 'read', 'John' is a word in O_{nn} with ordered arguments 'John' and 'things', and '−s' is the operator-indicator on 'like', and 'for … to' is the

argument-indicator on 'read', and the words 'for John' are zeroed *qua* repetition while the word 'things' is zeroed *qua* indefinite. (adapted from 1991: 275))[15]

There is no need to go into the details. The very complexity of the example leaves no doubt that metasentences are essentially theoretical reflections of putative mental operations performed by a speaker, to the same extent that a Chomskyan phrase-marker would be, or lexical rules plus constituent structure plus functional structure in Lexical Functional Grammar.

I cannot go into the actual ability of the scheme to account for all the sentences of a language, but some areas of the grammar offer encouraging applications. For example, Harris's original depiction of pronominal reference and co-reference (1968: 139ff; 1991: 128-35) gives a convinving illustration of the effectiveness of his method. More generally, one of the model's obvious advantages is its universal applicability, owing to the simple fact that every L_n contains its own metalanguage.

Conclusion

I hope this paper may have given an intimation of the significance of Harris's theory of syntax. Once its basic principles are mastered, the method proves particularly flexible: the metasentences can always be amplified to accommodate newly observed phenomena. The risk, of course, is ad-hoccery. But it can be avoided provided one makes sure, as Harris is careful to do, that transformations are made to act as widely and in as many diverse contexts as possible.

Arguably, that is just what Harris has achieved with metalanguage: the syntax of metalinguistic sentences does not appeal to anything that is not found in the syntax of L_n at large. Metalinguistic sentences are a subset of the *X is N* set, and even such an apparently specific transformation as the addition of quotation-marks actually applies beyond the metalanguage, with such predicates as *is a noise*. Remember too that a vast number of metalinguistic sentences no longer exhibit both an autonym and a metalinguistic classifier. Either or both may have been zeroed in the process of generating the reduced sentence. Here again, these zeroings are exactly of the same nature as those performed on non metalinguistic sentences.

It is unfortunate that Harris's theorising had little impact on the community of linguists at large. This means that there are few extensive discussions of the validity of Harris's scheme and,

---

[15] The terms *operator-indicator* and *argument-indicator* designate elements that are found in base sentences, prior to the application of transformations.

in particular, of its ability to account for a natural metalanguage. Rey-Debove is one linguist who has ventured an assessment, albeit a lukewarm one. Although acknowledging the significance of Harris's contribution — she calls him "one of the few linguists to have assigned to the metalanguage a place of its own within the language" (1978: 41) — she judges that, like Jakobson, he was content with setting the scene for a description of metalanguage without developing it in detail.[16] Though it is true that Harris did not supply an exhaustive account of the workings of natural metalanguage — after all, he was trying to address the whole range of problems that are of interest to the grammarian — I believe he provided linguists with useful insights as to how to build such an account. One point is especially noteworthy: the possibility of giving a Harrissian analysis of the notoriously tricky 'mixed uses'. This in itself is a strong indication in favour of the fecundity of Harris's scheme.

Bibliography

CARNAP, Rudolf (1937), *The Logical Syntax of Language*, translated from the German by Amethe SMEATON, London, Routledge & Kegan Paul.

HARRIS, Zellig (1968), *Mathematical Structures of Language*, New York, London, etc., Interscience Publishers.

HARRIS, Zellig (1976), *Notes du cours de syntaxe*, translated by Maurice GROSS, Paris, Seuil.

HARRIS, Zellig (1988), *Language and Information*, New York, Columbia University Press.

HARRIS, Zellig (1991), *A Theory of Language and Information*, Oxford, Clarendon Press.

HIZ, Henry (1994), "Zellig S. Harris", *Proceedings of the American Philosophical Society*, 138-4, 519-27.

LYONS, John (1995), *Linguistic Semantics: An introduction*, Cambridge etc., Cambridge University Press.

MATTHEWS, Peter (1999), "Zellig Sabbettai Harris", *Language*, 75-1, 112-19.

QUINE, Willard Van Orman (1940), *Mathematical Logic*, Cambridge, Mass., Harvard University Press.

RECANATI, François (1979), *La transparence et l'énonciation. Pour introduire à la pragmatique*, Paris, Seuil, coll. L'ordre philosophique.

REY-DEBOVE, Josette (1978), *Le Métalangage. Etude linguistique du discours sur le langage*, Paris, Armand Colin.

SAKA, Paul (1998), "Quotation and the Use-Mention Distinction", *Mind*, 107, 113-35.

---

[16] Rey-Debove (1978) did not rely on publications of Harris's post 1968. A little disappointedly, that situation remained unchanged in the 1997 second edition of *Le métalangage*: no account was taken of Harris (1988 1991).

TARSKI, Alfred (1944), "The Semantic Conception of Truth and the Foundations of Semantics",
Journal of Philosophy and Phenomenological Research, 4, 341-375.