608
Notre Dame Journal of Formal Logic
Volume XV, Number 4, October 1974
NDJFAM

SELF REFERENCE IN FORMAL LANGUAGES

PETER SWIGGART

A syntactical ambiguity can be shown to characterize R. M. Smullyan's S_0 language in which self reference is said to be formally possible.[1] The ambiguity in question is linked to the inclusion of devices for transforming concatenations of symbols into designators or names of the symbol concatenations. As a result of the use of formal quotation marks, a situation is created in which the logical status of symbols that are part of designators can become systematically uncertain. In view of this syntactical ambiguity it is no longer possible to say unequivocally of S_0 that its well formed sentences constitute a subset of the expressions of the language, where an expression is defined as a concatenation of the symbols of the language. If a given string of the symbols of the language proves to be ambiguous concerning its syntactical structure, our best recourse is to define a sentence of the language not as a concatenation of symbols but as a juxtaposition of syntactical units. But if we insist upon the systematic application of such a distinction, then it can be shown that Smullyan's language fails to be self-referring.

Smullyan's S_0 language contains three primitive signs, 'φ', '*', and 'N', and is based upon the following definitions:

D1. By an *expression* of S_0 is meant any string built from the three signs of S_0.
D2. By the (formal) quotation of an expression is meant the expression surrounded by stars.
D3. By the *norm* of an expression is meant the expression followed by its own formal quotation.
D4. A *sentence* of S_0 is an expression consisting of 'φ' followed by a designator.

1. "Languages in which self reference is possible," *The Journal of Symbolic Logic*, vol. 22 (1957), pp. 55-67.

Received March 13, 1972

A designator in S_0 is either the quotation of an expression or such a quotation preceded by one or more 'N's. To the above definitions two rules of designation may be added:

R1. The quotation of an expression E designates E.
R2. If E_1 designates E_2, then $\ulcorner NE_1 \urcorner$ designates the *norm* of E_2.

The difficulty in question involves the status of designators that quote complex expressions that can themselves be construed to be designators or to contain designators. This creates a situation in which the logical interpretation of a string of concatenated symbols can depend upon arbitrary conceptualizations of internal structure. For example:

(1) *N*N*
(2) φ*N**N*

Definitions D3 and D4 suggest that the logical status of any complex expression of S_0 (for example, its status as a designator or as a meaningless concatenation of signs) will depend upon analysis of what expression or expression complexes follow what expressions or expression complexes. But such an analysis is indeterminant in the case of (1) and (2). Assuming that (1) is an expression of S_0, or a string built from its primitive symbols, we can construe (1) either as the designator of 'N*N' or as an ill formed expression consisting of the designator '*N*' followed by the expression 'N*' (or preceded by '*N'). Correspondingly, we may interpret (2) either as a well formed sentence, a predicate followed by a designator, or as an ill formed expression consisting of a predicate followed by two juxtaposed designators. Of course the ambiguity can be resolved by the addition to S_0 of a rule to the effect that in the case of syntactical ambiguity precedence must be given to that syntactical interpretation which yields either a well formed syntactical complex or a well formed sentence. Yet the necessity for such a rule suggests the need to discriminate the interpreted expressions of S_0 from syntactically uninterpreted concatenations of signs, a distinction which can be seen to undermine Smullyan's formal proof of self reference in S_0.

It is worthy of note that if we create a language S_c which consists of S_0 with the addition of sentence connectives, we can generate within S_c an ambiguity that no simple interpretation rule can dissipate. Adding the connective '$\supset$', we can obtain such expressions as the following:

(3) φ*N* $\supset$ φ*N*

(3) can be taken as either of two quite different sentences, a sentence composed of 'φ' followed by the designator '*N* $\supset$ φ*N*' or a sentence composed of the sentences 'φ*N*' and 'φ*N*' joined by the connective '$\supset$'.

The source of syntactical ambiguity in S_0 and related languages is the fact that we can no longer deduce from the formation rules of S_0 the syntactical role that individual symbols will perform in well formed expressions and sentences. For example, although 'N' is defined as the 'norm of' function, a means of converting certain expressions into

designators, 'N' does not always perform this function in well formed sentences. In effect S_0 possesses the capability of generating sentences in which the expressions of S_0 are mentioned rather than used.

The need to distinguish the syntactically interpreted from the uninterpreted expressions of S_0 amounts to the need to distinguish mere concatenations of the symbols of S_0 from those symbol strings where it has been decided which symbols are merely mentioned—are component parts of designator expressions—and which symbols are provided with a syntactical use or function. Once such a necessary distinction is enforced it becomes possible to speak of a given concatenation of symbols as expressing a well formed sentence, but only if it is acknowledged that the same concatenation could be said also to express an entity that is not a well formed sentence. For example, it is possible to say that the string of symbols 'φ*N*' expresses a sentence in which the predicate 'φ' is followed by the designator '*N*'. However, we may also supply a logical interpretation of the string of symbols from which it follows that what the string expresses is not a sentence in which 'N' is mentioned rather than used, but a meaningless combination of uses of the four symbols. In short, the claim that a given concatenation of symbols expresses a given sentence must now be understood to imply that the string of symbols has been subjected to an appropriate syntactical interpretation.

Because of the necessity for a distinction between uninterpreted and interpreted symbol concatenations, Smullyan's S_0 language can be shown to fail to permit linguistic self reference. According to Smullyan's argument, the expression 'N*N*' functions in S_0 as a self referring designator. We are told that since ⌜N*E*⌝ designates the expression ⌜E*E*⌝, the expression 'N*N*' must designate the expression 'N*N*', namely itself. But the conclusion that 'N*N*' is self referring presupposes an assimilation of uninterpreted to syntactically interpreted expressions of the language. We are initially told that 'N*N' is to be taken as a designator, and we thus know that 'N*N*' in its function as a designator must designate the norm of 'N'. Moreover, we know by D2 that the formal quotation of an expression is the expression surrounded by stars. Therefore, from the fact that the designator 'N*N*' designates the norm of 'N' it follows that the designator 'N*N*' designates the expression 'N*N*'. What we do not know, and cannot know, is that the expression 'N*N*' is identical in syntactical status with the designator 'N*N*'. For the knowledge that 'N*N*' consists of the expression 'N' followed by its own quotation is not the same as the knowledge that 'N*N*' consists of 'N' followed by a designator. For not every expression of S_0 that consists of an expression surrounded by stars can be said to function as a designator, even in well formed sentences. As an example, take the following expression:

(4) φ*N**

(4) can be taken as expressing a sentence composed of the predicate 'φ'

followed by a designator. Moreover, (4) contains the three concatenated signs '*N*' which by D2 must be said to constitute the formal quotation of 'N'. But the designator in sentence (4) is not '*N*' but '*N**'. If a designator is defined as a syntactical unit, then we must say that (4) contains a quotation of 'N' that defies classification as a designator. From such a demonstration it follows that Smullyan's R1 rule admits a multiplicity of exceptions. From the fact that ⌜*E*⌝ is the quotation of E it does not follow that ⌜*E*⌝ is a designator that designates E. From the fact that the designator 'N*N*' designates an expression consisting of 'N' followed by the quotation of 'N' it does not follow that this designated expression must be construed to be 'N' followed by a designator. But if '*N*' in the expression 'N*N*' is not taken to be a designator, then R2 fails to apply and 'N*N*' cannot itself be classified as a designator. Consequently we cannot say that the designator 'N*N*' necessarily designates an expression that is identical with itself.

To some extent the definition of a designator as a syntactical unit, and not as a concatenation of certain symbols, is implied by Smullyan's account of the formation rules of S_0. But Smullyan does not investigate the need for distinguishing complex syntactical units from the symbol strings that express them. Thus we may accept his definition of a sentence as implying that a sentence is formed not by concatenating symbols but by juxtaposing syntactical units. However, such a reading merely emphasizes the defective nature of R1, which identifies the quotation of an expression as constituting a syntactical unit, regardless of the context in which the quotation appears. Yet in rejecting R1 as a valid formation rule we reject any hope of proving formally that the designator 'N*N*' designates itself. This conclusion can be easily extended to Smullyan's example of a self referring sentence, namely 'φN*φN*'. The discrediting of R1 removes the possibility of proving that 'φN*φN*' contains a designator that refers to 'φN*φN*' in its function as a sentence. We must say instead that the reference is to a linguistic entity that may be a syntactically uninterpreted concatenation of symbols.

An additional point can be made concerning the similarity between the designators of Smullyan's S_0 language and the quoted or otherwise referentially opaque contexts of ordinary language. It was argued above that the sentences of S_0 are composed of syntactical units, and that in S_0 a syntactical unit cannot be identified with a concatenation of the individual symbols of which it is composed. By enforcing this implicit distinction, Smullyan has created a formal language that is like an ordinary language in its ability to discriminate between the use and the mention of its own symbols. But S_0 differs markedly from an ordinary language in its failure to create the equivalent of quotation contexts that, at least in principle, are invulnerable to quantification. Of course S_0 has no quantifiers; nor does it have the usual sort of individual constants or variables. My point is that since the designator contexts of S_0 are less opaque than the quotation

contexts of ordinary language, there is no possibility in S_0 of approximating the types of limited quantification that we find in ordinary language. This is because the symbols that are included within designator expressions retain at least a residual logical function, which can be formally described as their vulnerability to quantification, were quantifiers possible in the language.

This potential difficulty can be elucidated by the following example. Let us create the language S_q by adding the individual variables 'x' and 'y' to S_0, and by including existential quantification. Such sentences as (5) thus become well formed in S_q:

(5) $(\exists x)[(\exists y)(\varphi *x*)]$

Upon the assumption that 'x' ranges over any object to which reference is formally possible, we can substitute for 'x' the expression '$*y*$', which names the symbol 'y'. But if this is done, the 'y' that is part of the name of itself is now subject to quantification:

(6) $(\exists y)(\varphi **y**)$

We cannot block such a result by appealing to the notational rule that no free variable can be captured by a quantifier; for 'y' does not occur in '$*y*$' as a free variable. The expression '$*y*$' contains mention rather than a syntactical use of the variable 'y', and in this respect the expression can be said to contain a symbol that is neither free nor bound.

This difficulty which would apply to a richer language than S_0 is a heightened version of the syntactical ambiguity basic to the use of designators in S_0. The device of adding quotation marks to a formal language permits the generation of logical environments in which symbols can be mentioned rather than used. Yet the mention-use distinction, so crucial to ordinary languages, seems alien to those notational conventions by which in contemporary times we construct formal languages. That is, our formal languages are generally constructed upon the principle that no distinction is possible between the appearance of a symbol in a well formed sentence and its performance of a syntactical function that is defined by the formation rules of the language. Smullyan's language designed for self reference achieves a modification of this basic convention, so that in quoting an expression it becomes possible to deprive the symbols that constitute the expression of the major part of their customary syntactical function. But there is not, nor can there be any definitive procedure in S_0 for distinguishing opaque from open contexts—not as long as Smullyan's quotation marks function as symbols that can themselves be quoted. As a consequence, the mention-use distinction in S_0 remains radically unstable. There is always the formal possibility of interpreting an S_0 symbol according to the syntactical function assigned to it by formation rules, even in contexts where the presence of quotation marks suggests that the symbol is being mentioned rather than used.

Brandeis University
Waltham, Massachusetts