

Embedded Pronoun Reference

VERN POYTHRESS

Westminster Theological Seminary

This paper gives a formal rule for calculating the referents of English personal pronouns when such pronouns occur in certain fixed ways in direct quotations, direct quotations of direct quotations, direct quotations of direct quotations of direct quotations, etc. The rule of reference can be expressed in the form of an informal computations method, an automaton, or an algebraic formalism, all of which result in the same output. The automaton and the algebraic formalism can be so constructed that they reject ungrammatical quotations.

1. INTRODUCTION

Can a formal rule be given for determining the referents of pronouns embedded in a series of quotations such as

I said to them, "You said to us, "He said, "I am going." " " " ?

In the present paper, after defining the type of embedded quotation amenable to my method (Section 2), I discuss an informal computation method (Section 3), a nondeterministic automaton (Section 5), and an algebraic formalism (Section 6), any one of which calculates referents of each pronoun.

Pike and Lowe (1969) have already applied group theory to give an answer to the embedded quotation problem, subject to the three major restrictions that (1) only three participants A, B, and C are involved, (2) only singular nonreflexive pronouns occur in the quotations, and (3) at each stage of the embedding a formula occurs (such as 'you said to him') specifying both a single speaker ('you') and a single addressee ('him'). Lowe's article (1969) also contains some brief suggestions about extensions to include a first-person-plural 'we' and involvement of more participants.

In this paper, I will remove all three restrictions in Pike and Lowe's work. I retain restrictions such as the following:¹

¹ I am indebted to Dr. Pike for bringing to my attention especially restrictions 2, 5, and 6.

1. At each stage of embedding there occurs a formula

X said to Y

of a certain type (cf. Section 2).

2. No editorial *we*'s occur.
3. No extragrammatical information such as intonation or pointing is available to indicate referents of pronouns.
4. The discourse involves a fixed number m of participants.
5. Any compound names such as 'President John Brown' function as unanalyzed wholes.
6. Any participant in the discourse has only one name.
7. Different participants have distinct names.

Even so, the price of generality is a less close match with actual habits of English speakers. My formal rules are intended only as a first approximation to usage.

2. DESCRIPTION OF QUOTATION TYPE

Now I describe the type of embedded quotation amenable to treatment by my method.

Assume that a given discourse involves a total of four participants: Abe, Bill, Charlie, and Don. Then we accept embedded quotations such as the following:

- (1) Abe, Bill, and I said, "He said to Charlie, "Don is going.""
- (2) ^o They said to Bill, "You and Charlie said to yourselves and us, "You said to him, "We hit him.""
- (3) He said to Bill and me, "They said to you, "Charlie is rich.""

etc.

² "o" indicates an improbable quotation, but one not excluded by our rules. "*", on the other hand, indicates an ungrammatical or unacceptable quotation.

The sequences

Abe, Bill, and I said
 he said to Charlie
 they said to Bill
 etc.

are called *formulas*. The n th formula of an embedded quotation is called the embedded formula at the n th stage. Thus in (2) above 'They said to Bill' is at the first stage and 'You said to him' at the third stage in the embedded quotation. The sentences such as 'Don is going.' or 'We hit him.' at the final stage do not really concern us; we are simply interested in determining the referents of the pronouns at the final stage. We will assume that the final stage is simple enough not to pose *further* problems of reference, beyond those caused by the previous stages.

In general, an n th stage formula consists of a subject Sub and an object Obj (which may be empty, cf. (1), stage 1), connected by 'said to'. The Sub and Obj each consist of strings of proper names and pronouns, connected by commas and 'and'.³ More precisely, assume that the discourse involves a fixed set of m people identified by distinct names N_1, \dots, N_m . Let $N = \{N_1, \dots, N_m\}$ be the set of names (in (1)–(3) above; we can take $N = \{\text{Abe, Bill, Charlie, Don}\}$). Let $P_s = \{\text{I, you, he, she, we, they}\}$ be the set of personal subject pronouns, let $P_o = \{\text{me, you, him, her, us, them}\}$ be the set of nonreflexive personal object pronouns, and let $P_r = \{\text{myself, yourself, himself, herself, ourselves, yourselves, themselves}\}$ be the set of reflexive personal object pronouns. Then $F_s = P_s \cup N$ is the set of entries which can occur in Sub, and $F_o = P_o \cup P_r \cup N$ is the set of entries which can occur in Obj. If, for convenience, we neglect commas and 'and', a complete Sub consists of a nonempty string of entries from F_s . The set of all such strings we denote by $F_s^* - \{\emptyset\}$. Likewise, a complete Obj consists of a (possibly empty) string from F_o ; the set of all such strings is F_o^* . A formula such as 'you, Bill, and he said to Charlie, yourselves, and me' can then be represented symbolically by a pair (Q, R) , where $Q = \text{you Bill he}$ ($\in F_s^* - \{\emptyset\}$) and $R = \text{Charlie yourselves me}$ ($\in F_o^*$). We allow also special formulas like 'Bill and you said', $Q = \text{Bill you}$, $R = \emptyset$, where R is empty.

Furthermore, an embedded quotation of *depth* n can be represented by a sequence $(Q_1, R_1)(Q_2, R_2) \cdots (Q_n, R_n)E$ of pairs, where each $Q_i \in F_s^* - \{\emptyset\}$, each $R_i \in F_o^*$, and the end of the sequence is marked by 'E'. (The final stage,

³ Noun phrases like 'the boy' and 'the second man on the right' are thus not included in the scope of our method.

containing a statement such as 'Don is going' or 'we hit him', is not included in the sequence of Q 's and R 's.) For example, (1)–(3) above have the representations

(4) (Abe Bill I),(he, Charlie) E ,

(5) (they, Bill)(you Charlie, yourselves us)(you, him) E ,

(6) (he, Bill me)(they, you) E ,

respectively.

Finally, we must impose grammatical restrictions to avoid formulas like '*²I said to me', '*he said to myself', and '*John and John said'. In Sections 4 and 5 we will construct formal rules that distinguish grammatically acceptable and unacceptable embedded quotations. For the moment, we leave the question of grammaticalness in the background.

3. INFORMAL COMPUTATION METHOD

We assume that a set of initial speakers S_0 (an original 'I' or 'we') and a set of initial addressees A_0 (an original 'you') are known from the context of the embedded quotation. If not, the initial speaker(s) and addressee(s) are to be chosen at random. Let $(Q_1, R_1)(Q_2, R_2) \cdots (Q_p, R_p)E$ represent an embedded quotation. Then the set S_1 of speakers at stage 1 is simply the set of referents of Q_1 , and the set A_1 of addressees at stage 1 is the set of referents of R_1 . For example, in 'Abe and I said to Bill,...', we would calculate $S_1 = \{\text{Abe}\} \cup S_0$ and $A_1 = \{\text{Bill}\}$. S_0 is included in S_1 because 'I' must refer to the set S_0 of speakers at stage 0.

In general, when we have completed the n th stage of analyzing an embedded quotation, we expect to have a new set of speakers S_n (representing first person referents of the $(n+1)$ th stage) and a new set of addressees A_n (representing second-person referents of the $(n+1)$ th stage). The third-person referents are $N - S_n - A_n$ (all the names N not in S_n or A_n). To compute S_{n+1} and A_{n+1} , we use the information in the $(n+1)$ th stage formula (Q_{n+1}, R_{n+1}) : S_{n+1} is constructed as the set of referents of Q_{n+1} , and A_{n+1} as the set of referents of R_{n+1} .

The above method of computation is best explained in terms of examples.

EXAMPLE 1. Suppose we have a discourse with four people: $N = \{\text{Abe}$,

Bill, Charlie, Don}. Suppose $S_0 = \{\text{Abe}\}$ (Abe is the initial speaker) and $A_0 = \{\text{Don}\}$ (Don is the initial addressee). Take the embedded quotation

He said to Bill, "you and Abe said to me, "We hit you." "

$S_0 = \{\text{Abe}\}, A_0 = \{\text{Don}\}.$

$Q_1 = \text{he}, R_1 = \text{Bill}, Q_2 = \text{you Abe}, R_2 = \text{me}.$

Stage 1: 'He said to Bill'.

Part A: Examine $Q_1 = \text{he}.$

'He' must have a single male referent from the set $N - S_0 - A_0 = \{\text{Bill}, \text{Charlie}\}$ of third-person referents. Choose one such at random. (This splits the calculation into two separate calculations, one for each possible choice.)

Alternative 1: 'He' refers to $\{\text{Bill}\}.$

Part B: Determine $S_1.$

$S_1 = \{\text{Bill}\}$ (S_1 is the set of all referents of Q_1 ; $S_1 = \{\text{Bill}\}$ means that he = Bill is the new speaker.)

Part C: Examine $R_1 = \text{Bill}.$

'Bill' refers to $\{\text{Bill}\}.$

Part D: Determine $A_1.$

(A_1 is the set of all referents of R_1 .)

$A_1 = \{\text{Bill}\}.$

REJECT alternative 1.

Alternative 2: 'He' refers to $\{\text{Charlie}\}.$

$S_1 = \{\text{Charlie}\}.$

(This means that he = Charlie is the new speaker.)

'Bill' refers to $\{\text{Bill}\}.$

$A_1 = \{\text{Bill}\}.$

The alternative 1 now proves ungrammatical because, if 'He said to Bill' meant '*Bill said to Bill', it would have to be put, 'he said to himself' or 'Bill said to himself'. Hence, only alternative 2,

$S_1 = \{\text{Charlie}\}, \quad A_1 = \{\text{Bill}\},$

is left.

Stage 2: 'you and Abe said to me'.

Part A: Examine $Q_2 = \text{you Abe}.$

'you' refers to $A_1 = \{\text{Bill}\}$; 'Abe' refers to $\{\text{Abe}\}.$

Part B: $S_2 = \{\text{Abe}, \text{Bill}\}$ (S_2 is the set of all referents of Q_2 .)

Part C: Examine $R_2 = \text{me}$.
 'me' refers to $S_1 = \{\text{Charlie}\}$.

Part D: $A_2 = \{\text{Charlie}\}$.

Stage 3: 'We hit you'. 'We' refers to $S_2 = \{\text{Abe, Bill}\}$ (exclusive 'We') or to a superset of $S_2 : \{\text{Abe, Bill, Charlie}\}$ or $\{\text{Abe, Bill, Don}\}$ or $\{\text{Abe, Bill, Charlie, Don}\}$ (inclusive 'We'). 'you' refers to $A_2 = \{\text{Charlie}\}$.

The quotation would have been rejected if

(i) 'us' had replaced 'me' in stage 2 ('us' cannot refer to a single referent $S_1 = \{\text{Charlie}\}$);

(ii) 'myself' had been inserted with Bill in R_1 , forming '*He said to Bill and myself' ('myself' must refer to the subject 'He', but 'He' is not first person). The causes for rejecting certain choices of referents can be still more complicated, as the following example illustrates.

EXAMPLE 2. $N = \{\text{Abe, Bill, Charlie, Don, Ed}\}$. $S_0 = \{\text{Abe}\}$, $A_0 = \{\text{Bill}\}$. Consider the quotation

°I said to him, "Charlie said to you and me, "You said to him,
 "Don and we said to him, "You are going with Bill." " " " "

Stage 1: 'I said to him'.

Part A: 'I' refers to $S_0 = \{\text{Abe}\}$.
 ('We' would cause immediate REJECT.)

Part B: $S_1 = \{\text{Abe}\}$.

Part C: 'him' refers to one of $N - S_0 - A_0 = \{\text{Charlie, Don, Ed}\}$.
 Choose.

Part D: Alternative 1 ||| Alternative 2 ||| Alternative 3
 $A_1 = \{\text{Charlie}\}$. ||| $A_1 = \{\text{Don}\}$. ||| $A_1 = \{\text{Ed}\}$.

Each alternative must be developed at length.

Stage 2: 'Charlie said to you and me'.

Part A: 'Charlie' refers to $\{\text{Charlie}\}$ (no matter which alternative we are in).

REJECT alternative 1, $A_1 = \{\text{Charlie}\}$, because then 'You' rather than 'Charlie' would be used at stage 2 to refer to $\{\text{Charlie}\}$.

Alternative 2

Part B: $S_2 = \{\text{Charlie}\}$.Part C: Examine R_2 .'you' refers to $A_1 = \{\text{Don}\}$.'me' refers to $S_1 = \{\text{Abe}\}$.Part D: $A_2 = \{\text{Abe, Don}\}$.

Stage 3: 'You said to him'.

Part A: 'You' refers to
 $A_2 = \{\text{Abe, Don}\}$.Part B: $S_3 = \{\text{Abe, Don}\}$.Part C: 'him' refers to one of
 $N - S_2 - A_2 = \{\text{Bill, Ed}\}$.
Choose one.Part D: (2a) $A_3 = \{\text{Bill}\}$ //
(2b) $A_3 = \{\text{Ed}\}$.

Alternative 3

 $S_2 = \{\text{Charlie}\}$.'you' refers to $A_1 = \{\text{Ed}\}$.'me' refers to $S_1 = \{\text{Abe}\}$. $A_2 = \{\text{Abe, Ed}\}$.'You' refers to
 $A_2 = \{\text{Abe, Ed}\}$. $S_3 = \{\text{Abe, Ed}\}$. $N - S_2 - A_2 = \{\text{Bill, Don}\}$.(3a) $A_3 = \{\text{Bill}\}$ //
(3b) $A_3 = \{\text{Don}\}$.

Now we have four alternatives:

| | | | |
|-----------------------------|-----------------------------|----------------------------|----------------------------|
| (2a) | (2b) | (3a) | (3b) |
| $S_3 = \{\text{Abe, Don}\}$ | $S_3 = \{\text{Abe, Don}\}$ | $S_3 = \{\text{Abe, Ed}\}$ | $S_3 = \{\text{Abe, Ed}\}$ |
| $A_3 = \{\text{Bill}\}$ | $A_3 = \{\text{Ed}\}$ | $A_3 = \{\text{Bill}\}$ | $A_3 = \{\text{Don}\}$ |

Stage 4: 'Don and we said to him'.

(2a, 2b):

Part A: 'Don' refers to $\{\text{Don}\}$.
'we' refers to $S_3 = \{\text{Abe, Don}\}$.
REJECT alternatives (2a) and (2b).
(Don cannot be mentioned twice in
 Q_4 .)

(3a, 3b):

'Don' refers to $\{\text{Don}\}$.
'we' refers to $S_3 = \{\text{Abe, Ed}\}$.

We are left with

(3a) 'we' refers to $\{\text{Abe, Ed}\}$.Part B: $S_4 = \{\text{Abe, Don, Ed}\}$.(3b) 'we' refers to $\{\text{Abe, Ed}\}$. $S_4 = \{\text{Abe, Don, Ed}\}$.

| | | | | | | | | |
|---|--|---|--------|--|--------|-----------------------------|--|--------------------------|
| Part C: 'him' refers to one of $N - S_3 - A_3 - S_4 = \{\text{Charlie}\}.$ | | $N - S_3 - A_3 - S_4$ $= \{\text{Charlie, Bill}\}.$ Choose either Charlie (3b1) or Bill (3b2). | | | | | | |
| Part D: $A_4 = \{\text{Charlie}\}.$ | | <table style="border-collapse: collapse; width: 100%;"> <tr> <td style="width: 50%; padding-right: 10px;">(3b1):</td> <td style="width: 5%;"></td> <td style="width: 45%;">(3b2):</td> </tr> <tr> <td style="padding-right: 10px;">$A_4 = \{\text{Charlie}\}.$</td> <td style="border-left: 1px solid black; border-right: 1px solid black; padding: 0 5px;"></td> <td>$A_4 = \{\text{Bill}\}.$</td> </tr> </table> | (3b1): | | (3b2): | $A_4 = \{\text{Charlie}\}.$ | | $A_4 = \{\text{Bill}\}.$ |
| (3b1): | | (3b2): | | | | | | |
| $A_4 = \{\text{Charlie}\}.$ | | $A_4 = \{\text{Bill}\}.$ | | | | | | |

Two alternatives (3a), (3b1) have merged in their interpretation of 'him'.

Stage 5: 'You are going with Bill'.

| | | |
|---|--|--|
| (3a, 3b1): 'You' refers to $A_4 = \{\text{Charlie}\}.$ | | (3b2): 'You' refers to $A_4 = \{\text{Bill}\}.$ REJECT. (This would be written 'You are going with yourself'.) |
|---|--|--|

Finally, 'You' must refer to {Charlie}.

Thus, although at intermediate stages several alternative interpretations of referents may be available, they may all but one be eliminated (or even *all* eliminated!) by later constraints.

4. ACCEPTABLE EMBEDDED QUOTATIONS

This is a good point for us to say more precisely just when possible alternatives for assigning referents to pronouns are accepted or rejected. First, some precise rules for pronoun reference:

(1) Any reflexive pronoun in R_n must refer to all of S_n , that is, to the set of referents of Q_n .

For example, in 'Bill and we said to ourselves and Abe', 'ourselves' must refer to the same set of participants as 'Bill and we'; in 'Bill and they said to themselves', 'themselves' refers to the same set as 'Bill and they'; '*Bill and they said to himself' is unacceptable by rule (2) below.

Problem: We have a plausible exception to the rule in '*Abe and I said to myself'. We assume a dialect in which, if such a thing is to be expressed at all, it is said as 'oAbe and I said to me'. If the reader prefers to accept only the phrasing '*Abe and I said to myself' or to accept both alternatives, the rules can be changed accordingly.

(2) Singular pronouns must have single referents and plural pronouns multiple referents. 'You' can have either single or multiple referent(s). (In languages with duals or with morphologically distinct singular and plural 'you', the rules would have to be revised.)

(3) 'I' and 'me' must refer to S_{n-1} (the speaker at the previous stage). Note that, in such a case, rule (2) above requires that S_{n-1} be only a *single* speaker.

(4) 'We' and 'us' must refer to a superset of S_{n-1} . (A 'We' referring to S_{n-1} itself is an "exclusive" 'We'.)

For example, in 'Abe and Bill said to Charlie, "We are rich"', the 'We' may refer to {Abe, Bill} (exclusive) or to {Abe, Bill, Charlie}, or even (somewhat less common) to {Abe, Bill, Don} or {Abe, Bill, Charlie, Don} (all three inclusive). 'We' may not, however, refer to {Bill, Charlie} under normal circumstances.

Since 'We' and 'us' are plural, rule (2) requires in addition that the set of referents must contain more than one element.

(5) 'You' must refer to A_{n-1} (the set of addressees at the previous stage).

Problem: Under some circumstances, the quotation (I said to Bill, "You are rich") conceivably might mean '*I said to Bill, "Bill and Abe are rich"', though this is more typically expressed by 'Abe and you are rich' or 'You all are rich'. 'You' would certainly not be so broadly interpreted without a contextual indication. 'We' is much more likely to be interpreted inclusively (cf. rule (4)) than is 'You'.

(5') 'Yourself' and 'yourselves' must refer to a superset of A_{n-1} .

For example, 'Abe and you said to yourselves'.

(6) 'He' must refer to a singleton set of a male name, 'she' to a singleton set of a female name. In languages with gender distinctions in other pronouns, this rule would have to be extended.

(7) Third-person pronouns must refer to a subset of $N - S_{n-1} - A_{n-1}$.

For example, in 'Abe and Bill said to Charlie, "He is going"', 'He' may not refer to {Abe}, {Bill}, or {Charlie}. {Charlie} would be referred to by 'you' and {Abe} or {Bill} by their names (a first-person pronoun would not do since it must refer to the whole of $S_{n-1} = \{Abe, Bill\}$).

Let RP_n be the referent(s) of reflexive pronouns of R_n . By rules (1) and

(later) (11), at most one reflexive pronoun can occur in R_n and it must refer to S_n . Thus, for any n , RP_n is either \emptyset (the empty set) or S_n .

(8) If reflexives occur, they must be (a) first, (b) second, or (c) third person according as the set $RP_n = S_n$ of referents of the reflexive is such that (a) $RP_n \supseteq S_{n-1}$, (b) $RP_n \not\supseteq S_{n-1}$ and $RP_n \supseteq A_{n-1}$, or (c) $RP_n \not\supseteq S_{n-1}$ and $RP_n \not\supseteq A_{n-1}$. Thus,

Abe and you said to *yourselves*

Abe and I said to *ourselves*

You and I said to *ourselves*

Abe and they said to *themselves*.

In combination with the other rules (1–7, 11–17), this is equivalent to saying that a first-person reflexive occurs when a first-person subject pronoun occurs in the Sub, a second-person reflexive occurs when a second-person (but not also a first-person) subject pronoun occurs in Sub, and a third-person reflexive pronoun occurs otherwise.

In addition to the above rules of reference, we must postulate further restrictions on the co-occurrence of different nouns and pronouns:

(11) Any one participant cannot be referred to twice in Q_n or twice in R_n .

This eliminates ‘*Abe and Abe said’ and implies that in ‘Abe and we said’, ‘we’ cannot refer to a set including ‘Abe’.

(12) If $S_n \subseteq A_n$, R_n must contain a reflexive pronoun.

For example, ‘Abe and Bill said to Abe’ is acceptable but ‘*Abe and Bill said to Abe and Bill and Charlie’ is not; the latter should be ‘ \circ Abe and Bill said to themselves and Charlie’. The type of reflexive pronoun (singular or plural, first, second, or third person) is then completely determined by rules (2) and (8).

Problem: How does one express the idea ‘*Abe and Bill said, “Abe and Charlie said to Abe and Bill and Charlie,...”’? We assume that such a thing would either be expressed by a circumlocution like ‘Abe and Bill said, “Abe and Charlie said to themselves,..., and they also said to Bill the same thing.”’ (outside the scope of our method) or by

‘ \circ Abe and Bill said, “Abe and Charlie said to themselves and Bill,...”’.

The reader who feels that one can also say

*Abe and Bill said, “Abe and Charlie said to Charlie and us,...”

or

*Abe and Bill said, "Abe and Charlie said to themselves and us,..."

will have to modify the rules accordingly. Admittedly, these are all cases on the borderline of acceptability.

(13) If $S_n \supseteq S_{n-1}$, Q_n must contain a first-person pronoun.

For example, '*Abe said, "Abe and Bill are going"' is unacceptable.

(14) If $A_n - RP_n \supseteq S_{n-1}$, R_n must contain a first-person pronoun.

For example, '*Abe and Bill said, "Charlie said to Abe and Bill,..."' must be replaced by 'Abe and Bill said, "Charlie said to us,..."'.

The conditions when one must use second-person pronouns are more complex because the addressee(s) may be included in a 'we', 'us', or 'ourselves'. Let SFP_n be the set of referents of any first-person pronouns in Q_n , and let AFP_n be the set of referents of any first-person pronoun in R_n . Then

(15) If $S_n - SFP_n \supseteq A_{n-1}$, Q_n must have a second-person pronoun.

For example, '*Abe said to Bill, "Bill and Charlie are going"' must be changed to 'Abe said to Bill, "You and Charlie are going"'. But '*Abe said to Bill, "Bill and Abe are going"' may be changed to either 'Abe said to Bill, "We are going"' or 'Abe said to Bill, "You and I are going"'.

(16) If $A_n - RP_n - AFP_n \supseteq A_{n-1}$, R_n must have a second-person pronoun.

(17) The referent(s) of 'him', 'her', or 'them' can form a subset of S_n only if the same referent(s) is (are) referred to by the corresponding subject pronoun 'he', 'she', or 'they' in Q_n .

For example, in 'Abe and he said to him', 'he' and 'him' may have the same referent, but 'him' cannot refer to Abe. The same rule excludes the 'him' of Section 3, Example 2, stage 4, from referring to {Abe}, {Don}, or {Ed}.

5. THE AUTOMATON

To show that the rules for calculating the referents of pronouns are purely formal, we now construct an automaton that performs the same calculations of referents that we did informally in Section 3, and at the same time rejects

those interpretations of pronouns that violate the rules of Section 4. More precisely, the automaton will do the following:

(1) It will accept indefinitely long finite sequences $(Q_1, R_1)(Q_2, R_2) \cdots (Q_n, R_n) \cdots$ of pairs (Q, R) , the n th pair (Q_n, R_n) of which represents the n th stage formula ' Q said to R '. Thus, $Q_n \in F_s^* - \{\emptyset\}$ and $R_n \in F_o^*$ for each n .

(2) Depending on the content of (Q_i, R_i) , the automaton may change its state in determinate or indeterminate fashion, after which it is ready to accept a pair (Q_{i+1}, R_{i+1}) from the next stage of embedding.

(3) If pronoun references are ambiguous at stage n , the automaton becomes indeterministic at that point and simply picks at random one of the possible interpretations of each ambiguous pronoun.

(4) If, because of the structure of the embedded quotation and the

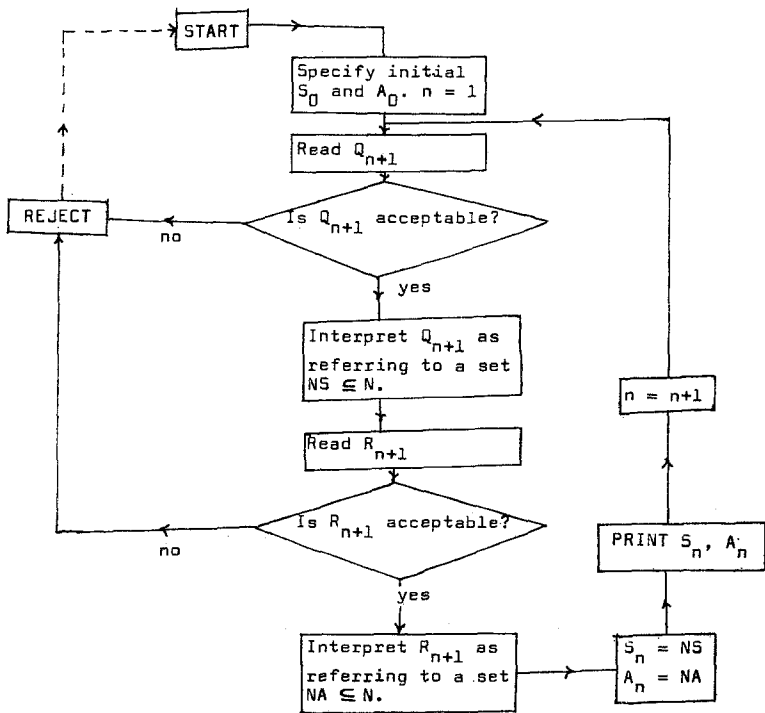


FIG. 1. Flow chart for the automaton. In actual operation the processes of accepting ('yes' or 'no') and of interpreting Q_{n+1} or R_{n+1} are intermingled.

previous choices of the automaton, it encounters a violation of any rule (1)–(17) of Section 4, it switches to a REJECT state from which it can be moved only by returning it to starting state (stage 0).

(5) The key to the automaton's operation is a collection of four "boxes" labeled S , A , NS , NA , each of which can hold subsets of N (the set of names of the participants of the discourse). While retaining in the boxes S ($=S_{n-1}$ when the automaton is at stage n) and A ($=A_{n-1}$ when the automaton is at stage n), the information about who the speaker(s) and addressee(s) at stage $n-1$ are, it gradually fills NS and NA with a set of "new speakers" and "new addressees". When NS and NA are complete, it sets $S = S_n = NS$ and $A = A_n = NA$, after which it is ready to begin stage $n+1$.

The operation of the automaton is summed up in the flow chart of Fig. 1.

We now exhibit in quasi-machine language a complete set of instructions for the automaton. Instructions are in caps and interpretations of each instruction in lower-case letters.

Let $M \subseteq N$ be the class of masculine names and $F \subseteq N$ the class of feminine names. $N = M \cup F$; $M \cap F = \emptyset$.

10 START

20 $n = 1$

n indicates the stage (in the sense of Section 3) at which the automaton has arrived.

30 $S = S_0$

This instruction is to be read, "Let S be equal to S_0 ." The initial speaker(s) S_0 is (are) specified.

40 $A = A_0$

The initial addressee(s) is (are) specified.

50 $NS = \emptyset$

Empty the new-speaker box. This box will be filled with the referents of Q_n .

60 $NA = \emptyset$

Empty the new-addressee box. This box will be filled with the referents of R_n .

70 $SFP = \emptyset$

SFP will keep track of the set of referents of first-person pronouns in Q_n , in order to check rule (13). We begin with SFP empty.

80 $B = 0$

B will keep track of whether a second-person pronoun occurs in Q_n , in order to check rule (15).

85 $STP = \emptyset$

STP will keep track of referents of third-person pronouns in Q_n , in order to check rule (17).

90 READ TAPE (X)

Let X be the first symbol on the tape.

100 IF $X = E$, STOP

' E ' signals the end of the tape.

110 IF $X \neq ($, REJECT

The first symbol on the tape should be a '(', preceding the first symbol of Q_n .

120 READ TAPE (X)

Let X be the first symbol ($\in F_s = P_s \cup N$) in the string Q_n .

130 IF $X \notin F_s$, REJECT

Recall that $F_s = P_s \cup N$ is the set of names and personal subject pronouns. X must be a name or a personal subject pronoun (by Section 2).

140 IF $X \neq I$, GO TO 180

Different procedures must be followed depending on what pronoun or name X is.

150 IF Card $S > 1$, REJECT

(Card S denotes the number of elements in the set S .) ' I ' can represent only a single referent, cf. rule (2).

160 $X = S$

' I ' must refer to $S = S_{n-1}$, cf. rule (3). This is the first of several alternative instructions (210, 240, etc.) for changing X from being a pronoun or proper name to being the set of referents of that pronoun or proper name. Instruction (160) tells how to convert from $X = I$ (first-person subject pronoun) to $X =$ referent of ' I '.

170 $SFP = X$

SFP is changed to indicate the referent of ' I '.

180 IF $X \neq$ we, GO TO 230

190 PICK $C = S$ OR C SUCH THAT $C \supset S$, $C \subseteq N$.

C is either the set S of speakers at stage $n - 1$ (exclusive 'we') or a superset of S (inclusive 'we'). Confer rule (4).

200 IF $\text{Card } C \leq 1$, REJECT

'we' can represent only multiple referents, cf. rule (2).

210 $X = C$

'we' refers to the set C . Now X represents the set of referents of 'we' (cf. remarks on instruction 160.)

220 $SFP = X$

Confer instruction 170 above.

230 IF $X \neq \text{you}$, GO TO 260

240 $X = A$

'you' refers to the set A of addressees, cf. rule 5.

250 $B = 1$

B is changed to indicate that a 'you' has occurred in Q_n .

255 IF $X \neq \text{he, she, or they}$, GO TO 310

260 IF $X = \text{he}$, PICK $X \in (N - S - A) \cap M$

Here the automaton indeterministically picks a possible referent of 'he' (cf. rules (6) and (7)). If $(N - S - A) \cap M = \emptyset$ (no possible referent), the automaton goes to REJECT.

270 IF $X = \text{she}$, PICK $X \in (N - S - A) \cap F$

280 IF $X \neq \text{they}$, GO TO 310

290 PICK $X \subseteq N - S - A$

A subset X of $N - S - A$ is chosen as the set of referents of 'they'.

300 IF $\text{Card } X \leq 1$, REJECT

'they' must have multiple referents, cf. rule 2.

305 $STP = STP \cup \{X\}$

X is added to sets of referents of third-person pronouns in Q_n .

310 IF $X \in N$, $X = \{X\}$

If X is a name, change X to be the singleton set with the name as its sole element. (For example, if $X_{(\text{old})} = \text{Abe}$, $X_{(\text{new})} = \{\text{Abe}\}$.) This is done in order that, whatever the symbol of Q_n that we started with, X will end up

being the set of the referents of the symbol of Q_n read off at instruction 120 or 340.

320 IF $X \cap NS \neq \emptyset$, REJECT

A given participant cannot be mentioned twice in Q_n (cf. rule (11)). X would here represent the second mention, the first mention already having been included in NS .

330 $NS = NS \cup X$

Add the referent(s) X to the new-speaker box.

340 READ TAPE (X)

Let X be the next symbol on the tape.

350 IF $X \in P_s \cup N$, GO TO 140

360 IF $X \neq ,$, REJECT

The only symbol besides a $F_s = P_s \cup N$ which is allowed to occur at this point is the comma ',' separating Q_n from R_n .

370 IF $NS \supseteq S$ AND $SFP = \emptyset$, REJECT

If a first-person pronoun does not occur ($SFP = \emptyset$) when it can occur ($NS \supseteq S$), rule (13) is violated.

380 IF $NS - SFP \supseteq A$ AND $B = 0$, REJECT

If a second-person pronoun does not occur ($B = 0$) when it can occur ($NS - SFP \supseteq A$), rule (15) is violated.

Now the string R_n is handled:

390 IF $X \neq)$, GO TO 420

The case where $X =)$, i.e., where R_n is empty, is treated separately.

400 PICK $NA \subseteq N$ SUCH THAT $NA \neq \emptyset$

If $R_n = \emptyset$, choose any nonempty NA as the set of new addressees.

410 GO TO 830

420 $RP = \emptyset$

RP will be the set of referents (if any) of reflexive pronouns.

422 $OFF = \emptyset$

OFF keeps track of the referents of an 'us' or 'me' in R_n , in order to check that rules (14) and (16) are satisfied.

424 $B' = 0$

B' keeps track of whether a 'you' occurs in R_n , in order to check that rule (16) is satisfied.

430 IF $X \notin P_r$, GO TO 620

Reflexives P_r are treated separately.

440 IF X IS SINGULAR, GO TO 470

450 IF $\text{Card } NS \leq 1$, REJECT

A plural reflexive can occur only with a plural subject (by rules (1) and (2)).

460 GO TO 480

470 IF $\text{Card } NS > 1$, REJECT

A singular reflexive can occur only with a singular subject (by rules (1) and (2)).

480 IF X IS FIRST PERSON, GO TO 510

490 IF X IS SECOND PERSON, GO TO 530

500 IF X IS THIRD PERSON, GO TO 560

510 IF $SFP = \emptyset$, REJECT

'Myself' or 'ourselves' can occur only if 'I' or 'we' occurs in Q_n (cf. rule (8)).

520 GO TO 580

530 IF $SFP \neq \emptyset$, REJECT

540 IF $B \neq 1$, REJECT

'Yourself' or 'yourselves' can occur only if 'you' and not 'I' or 'we' occurs in Q_n (cf. rule (8)).

550 GO TO 580

560 IF $SFP \neq \emptyset$, REJECT

570 IF $B \neq 0$, REJECT

If either 'you', 'we', or 'I' occurs in Q_n , 'himself', 'herself', and 'themselves' cannot occur in R_n (cf. rule (8)).

580 $X = NS$

X , the set of referents of a reflexive, is NS , the set of referents of Q_n (cf. rule (1)).

590 $RP = X$

RP keeps track of the referents of reflexives.

620 IF $X \neq \text{me}$, GO TO 650

630 IF Card $X > 1$, REJECT

'Me' can represent only a single referent (rule (2)).

640 $X = S$

Let X be the set of referents of 'me'.

645 $OFP = X$

OFP keeps track of the referents of first-person pronouns in R_n .

650 IF $X \neq$ us, GO TO 680

660 PICK X SUCH THAT $X \supseteq S$, $X \subseteq N$, Card $X > 1$, $X \neq NS$

X is a set of possible referents of 'us'.

670 $OFP = X$

Confer instruction 645.

680 IF $X \neq$ you, GO TO 710

690 $X = A$

Let X be the set of referents of 'you'.

700 $B' = 1$

B' is changed to indicate that a 'you' has occurred in R_n .

710 IF $X =$ him, PICK $X \in (N - S - A) \cap M$ SUCH THAT $\{X\} \in STP$

The condition $\{X\} \in STP$ checks rule (17).

720 IF $X =$ her, PICK $X \in (N - S - A) \cap F$ SUCH THAT $\{X\} \in STP$

730 IF $X =$ them, PICK $X \subseteq N - S - A$ SUCH THAT $X \in STP$ AND
Card $X > 1$

740 IF $X \in N$, $X = \{X\}$

Change X from denoting a name to the singleton set with the name as element.

750 IF $X \cap NA \neq \emptyset$, REJECT

A referent cannot be mentioned twice in R_n (rule (11)).

760 $NA = NA \cup X$

Add the referents of X to the new-addressee box.

770 READ TAPE (X)

Let X be the next symbol of R_n .

780 IF $X \in F_0$, GO TO 430

Begin cycling the next symbol of R_n .

790 IF $X \neq \text{)}$, REJECT

The last symbol of R_n must be followed by a closed parenthesis.

800 IF $NS \subseteq NA$ AND $RP = \emptyset$, REJECT

If $NS \subseteq NA$, R_n should have a reflexive (rule (12)).

810 IF $S \subseteq NA - RP$ AND $OFFP = \emptyset$, REJECT

If R_n can ($S \subseteq NA - RP$) contain a first-person nonreflexive pronoun, it must ($OFFP \neq \emptyset$) (cf. rule (14)).

820 IF $A \subseteq NA - RP - OFFP$ AND $B' = 0$, REJECT

If R_n can ($A \subseteq NA - RP - OFFP$) contain a 'you', it must ($B' \neq 0$).

830 $S = NS$

Information in the new-speaker box (NS) is shifted to the speaker box (S).

840 $A = NA$

Information in the new-addressee box (NA) is shifted to the addressee box (A).

850 PRINT (n, S, A)

The contents of S and A are printed out for a permanent record, along with the number n indicating the stage of embedding.

860 $n = n + 1$

Advance to stage $n + 1$.

870 GO TO 50

Begin processing the next stage.

880 END

With the description of the automaton completed, we can say something more about the indeterminacy of its decisions. We say that a sequence $(Q_1, R_1)(Q_2, R_2)...$ representing an embedded quotation is *indeterminate of order k at stage n* if there are exactly k possible printouts (n, S, A) resulting from choices made by the automaton at stages up to the n th stage. Each printout represents a possible interpretation of who the speaker(s) S and the addressee(s) A are after the n th stage. A sequence is *indeterminate* (or *ambiguous*) at stage n if $k \geq 2$, *determinate* (or *unambiguous*) at stage n if $k = 1$; *overdeterminate* (or *unacceptable*) if $k = 0$ (in this case, every possible choice sequence by the automaton eventually leads to REJECT). A sequence is *totally determinate* at stage n if it is determinate at all stages n' for $n' \leq n$. Sequences involving only names are either overdeterminate (e.g., 'John and John said'; 'John said to John') or totally determinate.

In case the number of participants is $m = 3$, and we restrict ourselves to sequences $(Q_1, R_1)(Q_2, R_2)\dots$ of the type in Pike and Lowe's paper (1969),⁴ the sequences are all totally determinate. In the printout (n, S, A) , S then represents the referent of 'I' and A the referent of 'you' after the n th stage. The referent of 'he' is $N - S - A$. Thus, in the restricted case, our results are the same as Pike and Lowe's.

6. ALGEBRAIC FORMALISM

We can bring our formal method into closer alignment with Pike and Lowe's (1969) group theoretic approach by speaking of multivalued operators $[Q_i, R_i]$ instead of an automaton. Let pN be the set of all subsets of N , and let $(S_0, A_0) \in pN \times pN$, i.e., $S \in pN$, $A \in pN$. Let $(S^{(1)}, A^{(1)}), \dots, (S^{(k)}, A^{(k)})$ (k possibly 0) be a list of all the possible printouts at stage 1 when the automaton starts at stage 0 in the stage $S = S_0$, $A = A_0$ and processes the string $(Q, R)E$.

Then denote by $[Q, R]$ the multivalued operator such that

- (i) $\{(S_0, A_0)\}[Q, R] = \{(S^{(1)}, A^{(1)}), \dots, (S^{(k)}, A^{(k)})\}$ and
- (ii) $\{b_1, \dots, b_i\}[Q, R] = \bigcup_{1 \leq j \leq i} \{b_j\}[Q, R]$

for any finite set consisting of pairs

$$b_j = (S, A) \in pN \times pN.$$

Given a set \mathcal{S} of pairs (S, A) , $\mathcal{S}[Q, R]$ thus represents the set of all new-speaker/new-addressee combinations obtainable from an automaton that starts at any *one* of the elements of \mathcal{S} and processes $(Q, R)E$. Denote by $[Q_1, R_1] \circ [Q_2, R_2]$ the composite operator \mathcal{O} such that

$$\{(S_0, A_0)\}\mathcal{O} = \{(S_0, A_0)\}[Q_1, R_1][Q_2, R_2].$$

The following theorem follows immediately from the sequential nature of the automaton:

THEOREM. *When the automaton starts at stage 0 in $S = S_0$, $A = A_0$*

⁴ This restriction amounts to (i) $\text{Card } N = 3$; (ii) $M = N$ (all males); (iii) $Q_i \in \{\text{I, you, he}\}$, $R_i \in \{\text{me, you, him}\}$; (iv) Q_i and R_i not the same person for any i .

and processes the sequence $(Q_1, R_1) \cdots (Q_n, R_n)E$, the set of all possible results at stage n is

$$(S_0, A_0)([Q_1, R_1] \circ [Q_2, R_2] \circ \cdots \circ [Q_n, R_n]) \\ = (\cdots (\{(S_0, A_0)\}[Q_1, R_1][Q_2, R_2]) \cdots)[Q_n, R_n].$$

Thus, in order to do arbitrary calculations of pronoun referents, it suffices to compute once and for all the value of each $[Q, R]$ on each (S, A) separately.

If each $[Q, R]$ is regarded as operating on sets \mathcal{S} of (S, A) 's, $\mathcal{S} \subseteq pN \times pN$, then the whole set of $[Q, R]$'s forms a monoid on endomorphisms of $p(pN \times pN)$, and the pair $[I, \text{you}]$ corresponds to the identity map of $p(pN \times pN)$ into itself. If $\text{Card } N = 3$, Pike and Powe's judicious limitations single out of this monoid the subgroup S_3 .

We prefer to start with the automaton rather than with the $[Q, R]$ directly, because a direct definition of all operators $[Q, R]$ in complete generality would be complicated and unilluminating. However, the algebraic formalism is useful in dealing with situations where one may wish to change the rules of Section 4 governing which embedded quotations are acceptable. Rather than having to construct a completely new automaton for each choice of rules, one simply has to do with (for the most part slight) differences concerning which operators $[Q, R]$ can occur and concerning the relation between pairs $(Q, R) \in (F_s^* - \{\emptyset\}) \times F_o^*$ and their corresponding operators $[Q, R]$.

For example, suppose we wish to add a rule that 'we' and 'us' *must* be exclusive, that is, they must refer to the set of speakers at the previous stage. Then for any (Q, R) containing no 'we' or 'us', the operator $[Q, R]$ remains the same; while for (Q, R) containing 'we' or 'us', the $[Q, R]$ in the new system of rules, call it $[Q, R]'$, is such that $\mathcal{S}[Q, R]' \subseteq \mathcal{S}[Q, R]$ for any set \mathcal{S} of initial pairs (S_0, A_0) . $\mathcal{S}[Q, R]'$ is to be interpreted as no longer including those pairs (S_1, A_1) which result from an inclusive interpretation of 'we' and 'us'. More precisely, $(S_1, A_1) \in \mathcal{S}[Q, R]'$ if and only if $(S_1, A_1) \in \mathcal{S}[Q, R]$ and (S_1, A_1) results from some $(S_0, A_0) \in \mathcal{S}$ using exclusive interpretations of 'we' and 'us'.

As a second example, suppose we imagine a more formal style of embedded quotation where, within a given Sub or Obj, pronouns must follow nouns and first-person pronouns must follow other types of pronouns. Then 'Abe and I said' and 'Abe and he said' are still acceptable, but 'I and Abe said' and 'he and Abe said' are not. This is equivalent to barring the occurrence of those operators $[Q, R]$ such that (Q, R) represents a prohibited order of nouns and pronouns.

RECEIVED: October 8, 1971

REFERENCES

- LOWE, I. (1969), An algebraic theory of English pronominal reference (part I), *Semiotica* 1, 397-421.
- PIKE, K. (prepublished article), Sociolinguistic evaluation of alternative mathematical models—English pronouns.
- PIKE, K., AND LOWE, I. (1969), Pronominal reference in English conversation and discourse—A group theoretical treatment, *Folia Linguistica* 3, 68-106.