Linguistic Research Institute

Report No. 22

FORMALIZATION OF REAL WORLD REPRESENTATION

H. Joel Jeffrey

1979

Linguistic Research Institute
Boulder, Colorado 80306

The intention of this paper is to present a mathematical formalism that can be used to state anything that can be stated in natural language. The formalism is a formalization of a conceptual analysis due to P. G. Ossorio[1]; there is nothing new here, conceptually. The contribution of this paper is to make clear the degree to which Ossorio's analysis is distinct from ordinary discursive English, a point which is very often missed, and to lay a foundation in mathematical formalism, for a mathematical theory of the real world.

SECTION I

The knowledge formats in this section are designed to provide a formalism for stating real world knowledge. By giving the information noted in each format, a particular process (or object, event, or state of affairs or type of process) is specified - i.e., distinguished from other processes.

A. Processes

1. A process is specified by giving a pair (PROC,DESCR). PROC is a member of a set PROCNAME = $\{NAME_i \mid i \leq 1\}$. DESCR is given as follows:

---

[1] P. G. Ossorio, What Actually Happens, Linguistic Research Institute, Boulder, Colorado, 1973.

2. A process divides into subprocesses; this information is specified by giving

Stages: $PROC_1, \ldots, PROC_K$ where

$PROC_i \in PROCNAME$, $1 \leq i \leq K$.

(Subscripts carry no implication of time ordering.)

3. Each stage may happen in various ways, in general. Formally, for $1 \leq i \leq K$,
$PROC_i = \{PROC_i^j \mid 1 \leq j \leq L_i,\ L_i \geq 1\}$ where each $PROC_i \in PROCNAME$.
(To be perfectly proper, since we have already stated that $PROC_i \in PROCNAME$, it is necessary to say that there is a 1-1 map from $PROC_i$ to $\overline{PROC}_i = \{PROC_i^j\}$. This distinction does not appear to be worth the added complexity.)

Pictorially, so far we have

$$\begin{array}{ccccc} PROC_1^1 & & \ldots & & PROC_K^1 \\ \vdots & & & & \vdots \\ \underbrace{PROC_1^{L_1}}_{=PROC_1} & & \ldots & & \underbrace{PROC_K^{L_K}}_{=PROC_K} \end{array}$$

, with no restrictions on $PROC_i^j$. In particular, a recursive process is specified by letting $PROC_j^i = PROC$, for some i,

4. Specifying an actual instance of a process requires a way to specify particular individuals (e.g., Sir Lawerence Olivier played Hamlet in Hamlet):

$$IND = \{I_1, \ldots, I_N\}$$

5. The individuals comprising IND may (or must) take certain parts, and not others, in the process to be specified. These parts, or roles, we call elements:

$$ELEM = \{E_1, \ldots, E_M\}$$

6. As noted in 5 above, there is restriction infor- specified. For example if PROC = "drink a cup of coffee," the same individual cannot play the role of cup and of coffee, or else the process thereby described is not drinking a cup of coffee. On the other hand, it is a legitimate production of Hamlet (i.e., qualifies as one) for Olivier to play both Hamlet and Ophelia.

This information, which we turn eligibilities, is stated formally as

$$PROCELIG_{PROC}: IND \chi ELEM \quad [0,1], \text{ where}$$

PROCELIG(I, E) is the eligibility of I to be E in process PROC. ([0,1] is the closed interval from 0 to 1 on the real line.) The range of [0,1] allows formal specification of degree.

7. Finally, it must be possible to specify what can occur and still be a case of PROC. This constraint information is of two types:

   a. Attributional Let ATTR = $[A_1, \ldots, A_Q]$ be a set of attributes, and let P = $\left\{PROC_i^j\right\}$ for PROC. Then $PROCATTR_{PROC}$: ELEM χ P χ ATTR [0,1]. $PROCATTR_{PROC}(E, PROC_i^j, A)$ is the degree to which the occurrence of $PROC_i^j$ is contingent on element E having Attribute A.

Example: If the quarterback is not human, the process is not a football game, although everything else that occurs would fit that description exactly.

b. Co-occurrence

$$PROCCO_{PROC} : P \chi P \rightarrow [0,1]$$

Here, $PROCCO_{PROC}$ $(PROC_i^j, PROC_i^{j'})$ is the degree to which $PROC_i^{j'}$ is contingent on occurrence of $PROC_i^j$.

Example: If a checking account balance is negative, the account is flagged for a message to the customer.

Both PROCATTR and PROCCO may be used to specify statistical information on the elements and options, as well as logical constraints.

Example: Ninety-five percent of the forwards in professional basketball have the attribute of being over 6 feet 5 inches tall.

B. Objects

1. The basic object description is given by a pair (OBJ, DESCR), where OBJ ε OBJNAME = $\{NAME_i \mid i \geq 1\}$, and DESCR is as follows:

Objects divide into subobjects, which are related in certain ways. This information is given by immediate constituents and their relationships:

2. Constituents $OBJ_1, \ldots, OBJ_K$

3. Relationships $R^{n_1}_1, \ldots, R^{n_m}_m$, where $R^{n_i}_i$ is an n_i - ary relation.

To specify each relation, the following must be stated:

A. The name OBJREL ε RELNAME = $\{NAME_i \mid i \geq 1\}$

B. Elements OBJELEM = $\{E_1, \ldots, E_m\}$. OBJREL is a relation between these elements - i.e., $R^{n_i}_i$ is OBJREL($E_1, \ldots, E_{n_i}$).

Example: Let the object be a room, with constituents of FLOOR, CEILING, and $WALL_i$, $1 \leq i \leq 4$. A room's ceiling is above its floor: ABOVE(FLOOR, CEILING).

C. Individuals RELIND = $\{I_1, .., I_N\}$.

As with processes, these are the actual historical individuals that take the roles $E_1, \ldots, E_M$.

D. Eligibilities $\underset{OBJ}{RELELIG} : \underset{OBJ}{RELIND} \times \underset{OBJ}{RELELEM} \rightarrow [0,1]$,

where $\underset{OBJ}{RELELIG}(I, E)$ = the eligibility of I to be E.

Example: Let OBJ = a box of pencils (the single object that is the box with pencils) Constituents are BOX, $P_0, \ldots, P_{10}$; there are 11 individuals $I_0, \ldots, I_{10}$. Let REL = WITHIN. Then

RELELIG $(I_0, BOX) = 1$

RELELIG $(I_j, BOX) = 0$, $1 \leq j \leq 10$

RELELIG $(I_j, P_k) = 1$, $1 \leq j,k \leq 10$

Colloquially, any of the 10 individuals (that are long, thin, and yellow, though we haven't, and don't need to, say that formally above) will serve as any of elements called pencils, but none may be the box. Similarly for the (square, hollow) individual, which cannot be a pencil. (The question of

what qualities must be present in order to be able to use elements differently is a matter of attributional constraints in the process and objects involved).

E. Contingencies Let ATTR = $\{A_1,\ldots,A_Q\}$ be a set of attributes.

i) $RELATTR_{OBJ}$:RELELEM χ ATTR [0, 1], where $RELATTR_{OBJ}(E,A)$ = the degree to which E must have Attribute A, (for the relation RELNAME).

Example Let A be "transparent," E be a window. Then RELATTR(E,A) = 1 states that in order to be considered a window, E must be transparent, while RELATTR(E,A) = 0 states that transparency is irrelevant to bring a window, (and thus this description would not specify a window).

ii) Co-occurrence

$RELCO_{OBJ}$:RELELEM χ RELELEM $\rightarrow$ [0, 1], where $RELCO_{OBJ}(E,F)$ is the degree to which F must be present, if E is. Example: Let

OBJ = a table, REL = attached, E_1 = top, E_2 = a leg. Then RELCO(E_1,E_2) = 1 indicates that if E_1 is present, then E_2 must be. In ordinary language, if the top is to be attached to a leg, the leg must be present. As with processes, RELCO and RELATTR may specify statistical data as well as logical constraints.

What we have so far is a formalism for completely specifying the parts of any object, and their relationships. This can be thought of as basic object information. However, the point of all this is to provide a formalism for distinguishing any part of the real world from things that are not that. It is sometimes the case that "defining" characteristics can be stated only in terms of the objects, processes, events, and states of affairs of which the object is a part. For example, what distinguishes a carburetor from other objects is that it mixes fuel and air for an internal combustion engine, rather than any particular collection or arrangement of parts. While the parts and relationships of a particular carburetor, or a kind of carburetor, may be specified by the formalism

given so far, it is not sufficient. In a similar vein, point, line, and plane are only defined in terms of the relations they may have among themselves; points have no immediate constituents.

4. Attributes of OBJ: OBJATTR:OBJNAME χ ATTR $\rightarrow$ [0, 1], the degree to which OBJ must have Attribute A.

5. Contingencies

A. Attributes a constituent must have $CONSTATTR_{OBJ}$:OBJNAME χ ATTR $\rightarrow$ [0, 1].

B. The information on an object's immediate constituents and their relations is in general only one of several possible decompositions of OBJ. Let us term such a decomposition a paradigm, and denote by $PARA_{OBJ}$ the set of paradigms of OBJ (not all possible paradigms, in general, merely the set being described.) In general, the applicability of a given name will depend on the paradigm being used. For example, if OBJ is a room with four people present,

referring to "John" is only possible in a paradigm in which the person John is an immediate constituent. If the paradigm is, e.g., top-half and bottom-half, the name John is not applicable. Let Π_{OBJ} = {paradigms of OBJ}. APPLPARA: OBJNAME χ $\Pi_{OBJ} \rightarrow$ [0, 1]; the degree to which $NAME_i$ is applicable only when $PARADIGM_j$ is involved.

C. Applicability Within a Unit

Often a given name is only used when the object is part of a larger unit (e.g., process or state of affairs). For example the term "first base" can only be applied to a square bag of sand when the object is part of the processes and states of affairs comprising baseball. Let S = the set of states of affairs. Then APPLUNIT:OBJNAME χ $S \rightarrow$ [0, 1]; the degree to which $NAME_i$ applies only within S.

D. There are cases in which having a given attribute, or being an element in a configuration, is contingent on the

specification of a particular paradigm. For example, describing a program as well structured depends on using the paradigm of a program as a set of modules. Formally, ATTRPARA:ATTR $\cup$ S χ Π_{OBJ} [0, 1], the degree to which having Attribute A (or being an element of configuration S) is contingent on specifying $PARADIGM_j$.

E. The final type of contingency is that there are cases in which OBJ has Attribute A due to some constituent having A. Example: an automobile is an internal combustion machine because its engine is an internal combustion machine
ATTRCONST:OBJNAME χ ATTR χ OBJNAME $\rightarrow$ [0, 1], where ATTRCONST(OBJ, A, OBJ_j) is the degree to which OBJ's having A is contingent on OBJ_j's having A.

6. Finally, note that

1) OBJ_j is an object name, and thus subject to decomposition (in general), and

2) There are cases in which it is necessary to specify relations such as in 1 through 3

above, but not restricted to having elements that are immediate constituents. Example: It must be possible to state such cases as "the handle of the cup [a decomposition of cup] in Room 321 of Building A [a decomposition of Building A] is broken. Thus, we remove the restriction that the E_j be the immediate constituents of OBJ.

A. States of affairs

1. A state of affairs description is given by a pair (SA, DESCR), where SA ε SANAME = $\{NAME_i \mid i \geq 1\}$. SA may be any reference that identifies the state of affairs to be described. Examples are "the man shot the bear," clauses such as "the shooting," or a formal symbol such as SA_1. DESCR is given as follows:

2. A state of affairs is a set of related objects, processes, events, and states of affairs. Thus, it is necessary,

first, to state what the relation is, and what each of the elements in the relation is - i.e., object, process, event, or state of affairs. Formally, this amounts to specifying

a. The name of the N place relation R^N : SAREL ϵ RELNAME = $\{NAME_i \mid i \geq 1\}$.

b. Elements the logical roles in the relation: SAELEM = $\{E_1, \ldots, E_N\}$.

c. Eligibilities It must be specified whether each element must or may be an object, process, event, state of affairs, attribute (a 1 place relation), or concept. (A concept is specified by an object, process, or state of affairs description. The difference between X and the concept of X is that the concept of X is an abbreviated form of "P acts on concept X," which means that in the description of P's behavior

the state of affairs description X appears in the K parameter. In other words, formally, noting that X is a concept is nothing more than noting that the description for X will be used differently i.e., will appear in different places in other descriptions, than if X is the object, process, event, or state of affairs itself. For example, we can say, "P hit the ball," but not, "P hit the concept of ball." Finally,

SAELIG:SAELEM χ {0, P, E, S, A, C}$\rightarrow$[0, 1

d. Individuals The N individuals that take the parts of SAELEM:SAIND = {I_1 ..., I_N}. ("Individual" does not mean "object").

e. Classification As with elements, the individuals must be classified as object, process, etc.

SACLASS:SAIND χ {0, P, E, S, A, C}$\rightarrow$[0, 1

f. Assignments The individuals must be related to the elements - i.e., an I_j is identified as the exemplar of E_j in the state of affairs SA:
SASSN: SAELEM χ SAIND $\rightarrow$ [0, 1]

The formalism so far allows formal statement only of such statements (SA names) as "Object A took part in Process P" ("took part in" is the relation name) or "Process P is a part of Process Q, etc. (This is not trivial; it includes such description as "The pencil and pen are inside the case," "The cat chased the rat," etc.). In general, we will need more:

g. Expansions Individuals so far are classified as objects, processes, etc. This classification may be expanded by giving an object, process, event, or state of affairs description of it, using the formalism of parts A, B, or C.

Such an expansion formally is EXP:SAIND $\rightarrow$ {(N, D)}, where (N, D) = (NAME, DESCR) = an object, process or event description Attributes are handled slightly differently: an expansion is state of affairs description of the SA in which the attribute is relation.

3. Finally, certain contingency information must be formally specifiable:

   a. Since contingencies appear in object and process descriptio these are contingencies within the state of affairs description

   b. It is sometimes the case that particular names are used for an element only when certain other names are in use for other elements. For example, "The fullback is 20 yards behind

and 5 yards to the left of the tackle" is a SA description in which the names are technical football terms; "John is 5 yards to the right of the quarterback" makes sense only in terms of a "wider" SA - i.e., one which includes the "ordinary" human object name "John" and the football term "quarterback." Formally, SANAMECON: SANAME χ SANAME$\rightarrow$[0, 1].

c. The final type of constraint that we used is a formal means to specify cases in which using a particular element is continge on its being an element of the particular SA within which it is an element. For example, a pawn is only a pawn within the SA that is the game of chess (not a particular game); if there were no such thing (SA) as chess there would be no pawns (although there might be oddly shaped pieces of wood, ivory, et SAELEMSA: SAELEM $\rightarrow$ [0, 1].