Augmented semantic networks for an enterprise knowledge base

Paltin Sturdza discusses the use of knowledge processing in database research

Knowledge processing has spread from its original applications in artificial intelligence to other fields such as systems analysis, data analysis and database modelling. Such techniques could also be used in many business applications. This paper describes an augmented semantic network model which has been used for building, storing and processing an enterprise knowledge base. It describes how the model represents the world in terms of an open set of entities, events and properties and a closed set of relationships between them. The application of this model to systems analysis, problem solving, data definitions and database design is also discussed.

Keywords: artificial intelligence, knowledge processing, systems analysis, data analysis, database modelling.

Knowledge representation schemes such as semantic networks, predicate calculus, and production rules, have been extensively used in artificial intelligence (AI) systems which understand natural language, give expert advice, or recognize scenes. The preoccupation with knowledge processing has passed from AI to other areas in computer science, and it seemed natural that research in systems analysis, data analysis and database modelling would take a cognitive/semantic orientation (Brodie and Mylopoulos, 1983).

Within the database research field, this cognitive orientation has led to the experimental implementation of semantic interfaces for conventional database systems such as SDM (Hammer and McLEod, 1981) and to the development of a new data model, the entity–relationship (E–R) model (Chen, 1980). The E–R model represents the world in terms of entities and named relationships. The model has recently been implemented both experimentally (Muntz, 1980) and commercially (Mobsiii).

The trend toward semantic modelling is also present within the field of systems analysis. A recent model for business systems planning implemented with an IBM data dictionary closely resembling the E–R model has been developed for requirements analysis and systems design (Wilson, 1979; Sakamoto, 1980). In the area of process specification languages, designers are incorporating modelling features from both programming and natural language semantics (Goldman, 1980). The growing interest in the modelling power of natural languages is also evident from current research on the E–R model. English statements are to be directly translated into E–R diagrams. Nouns will be treated as entities, verbs as relationships, adjectives and adverbs as attributes (Chen, 1980).

Within the data analysis area, the cognitive/semantic trend appears as a significant effort to build subject data models which are independent of software and hardware considerations. In this approach, data elements represent particular facts about the enterprise while fields represent manifestations of these elements as named locations in programs or database records. Although the relationship between data analysis and the semantics of English has been recognized by the industrial analysts (Symons and Tijsma, 1982), defining data still remains a mere documentation exercise. Database models, in their turn, still do not use semantic cases and in systems analysis, the E–R model and the structured analysis method do not allow temporal and causal chains of events to be represented.

The Medical College of Pennsylvania, 3300 Henry Avenue, Philadelphia, PA 19129, USA
Received 20 May 1983, revised 18 July 1983

Accelerating the import of ideas from AI to business data processing could reduce the limitations described above and further expand the semantic/cognitive trend present in these areas. The AI field has produced some spectacular achievements in game playing, expert systems, and natural language understanding. Long-term research projects in AI, such as the Japanese fifth generation computer project, have received world wide attention. However, the practical impact of the AI field on business has been almost negligible. From 84 on-going AI research projects throughout the world, there are only a few systems which have become products destined to solve business problems. This slow penetration of AI systems in the business world can be explained in part by the extreme dependence of AI methodology on programming languages such as LISP or PROLOG which are seldom used in business applications. Representing world knowledge in terms of LISP or PROLOG structures removes AI methodology even further from solving business problems.

It is possible that in the long run business applications will be written in LISP or PROLOG. Yet, for the time being, business data processing could benefit from the AI methodology by removing it from specialized languages and implementing it with available software packages. If the productivity of these commercial systems increases or if new ways of using these systems are discovered, AI methodology should be adopted to solve business problems in a more efficient way.

The present paper describes an AI methodology which has been used for building, storing, and processing an enterprise knowledge base. The knowledge base, stored and manipulated with an IBM data dictionary product, has been used for data analysis, systems analysis, and database design within a tuition reimbursement application at the Medical College of Pennsylvania. In the process, the IBM data dictionary has acquired a universal model which significantly facilitates both its population with extensibility subjects and its ability to control data definitions.

AUGMENTED SEMANTIC NETWORK MODEL

The basic semantic model represents the world in terms of an open set of entities, events, properties, and a closed set of relationships between them. In addition to the basic model discussed in more detail in Sturdza (1983) we will extend the model by orienting the flow of events in the network through production rules (Winston, 1979).

Entities

We define an entity as an object, person, location that has individual existence and corresponds to a concrete or abstract noun in English. Entities are classified using a 'VARiety' relationship between the superclass and the subclass and are instantiated using the 'INStance' link between the class and its members. For example, 'machines' have 'computers' as a 'variety' and 'computers' have 'IBM-4331-SN12345' as 'instance'.

Events

An event is a happening, an action, a process in which entities are participating with different functions. Events correspond to verbs seen as 'miniature dramas' (Tesniere, 1959) in which actors are involved according to a limited number of roles (Greimas, 196). It is difficult to distinguish linguistically between generic events and individual events: in the sentences 'computers calculate' and 'Smith calculates' the two classes of events have the same verbal form. To solve this problem when semantic networks are implemented with he IBM data dictionary, each event class or instance will be subscripted and will thus be uniquely identified in the same way entities are. It is generally agreed that verbs (events) carry more information than nouns and thus all nominalizations of verbs such as 'marriage', 'sale', will be returned to their original verbal form: to sell, to marry, etc. This reconversion to original verbal forms is also dictated by practical reasons: temporal and locative information can be related directly to verbs (events) and only indirectly to 'nouns' (entities).

Properties

We refer to properties as characteristic traits or qualities (fast, slow, red). If the property refers to an entity it appears as an adjective. If it refers to an event it will appear as an adverb. Like entities and events, properties can also be classified with 'variety' relationships. Properties are connected to the entities and events they characterize by a 'QUALifies' relationship which can be further described using the relationship data feature of the data dictionary. For example, if 'color' qualifies 'car', we can give 'red' as value for this relationship.

Event modelling

Classification schemes (taxonomies) are not sufficient to build knowledge about the world. It is not enough to know what a thing is, we must also know what it does. We must define the roles various entities will play in an event. Most questions humans ask are structured around an event and refer to a role an entity might play in the event. In cognitive theory (Hays, 1971; Phillips, 1978) a limited number of such role relations can exist between an entity and an event: AGENT, EXPERIENCER, INSTRUMENT, OBJECT, LOCATION, PART-WHOLE, and OBJECT-RESULT.

Events are related to other events through a set of inter-event relationships: CAUSE, PURPOSE, SIMULTANEITY, SEQUENCE, ENABLEMENT, EVENT-OBJECT, RESULT, and METALINGUAL DEFINITION. In cognitive/semantic modelling, events, together with their participants, are the basic units of meaning. They correspond to propositions in logic.

Metalingual definitions (MTL)

Entities, events, and properties can be defined in terms of other events. For example, a simple metalingual definition for 'computer' could be: 'an electronic machine which performs calculations, correlates and selects data.' The entity 'computer' has been defined using three events and their participants. Like entities, events also can be MTL defined by still other events. For instance, 'programming' can be MTL defined as 'writing instructions for a computer and having these instructions executed'. The defining events can also be MTL defined within a complex architecture of events: 'writing instructions' can be further defined as 'key-punching' and 'coding'. MTL definitions of events can be used for process specification and business function analysis. Properties can also be expressed in terms of events using MTL definitions: the property 'efficient' can be viewed as the event 'producing results quickly'.

Production rules

Encoding isolated sentences into semantic nets is fairly simple. The sentence 'The Medical college reimburses students for tuition' has the semantic formula: ((AGENT: Medical College) (Event: reimbursing) (EXPERIENCER: employee) (OBJECT: tuition)). The encoding of larger text using the basic semantic networks is a more complex operation. The reason is that natural language sentences making the text have rarely the necessary organization which would reflect the flow of events in a business function. To model the flow of events and their conditional execution, we need to augment the semantic nets with production rules.

A production rule consists of a condition–action pair: if condition X holds then action Z is performed. A production whose condition part is satisfied has its action part executed, it 'fires' in AI parlance. A production which fires might change its status to a holding condition which might make the next production fire, and so on. Complex patterns of events in a semantic network can be viewed as productions and the order in which they fire can indicate the flow of events in a system.

ENTERPRISE KNOWLEDGE BASE IN SYSTEMS ANALYSIS

Using the semantic network model and the IBM data dictionary, we have built a knowledge base which can function as a policy advisor for the personnel department of the Medical College. The system is able to answer questions about the policy of tuition reimbursement followed at the college. The main source for the knowledge base has been the policy manual which guides administrators in authorizing tuition reimbursements. The natural language text of the policy has been manually transposed into cognitive/semantic networks, paragraph by paragraph. In the process, inconsistencies, redundancies, and ambiguities have

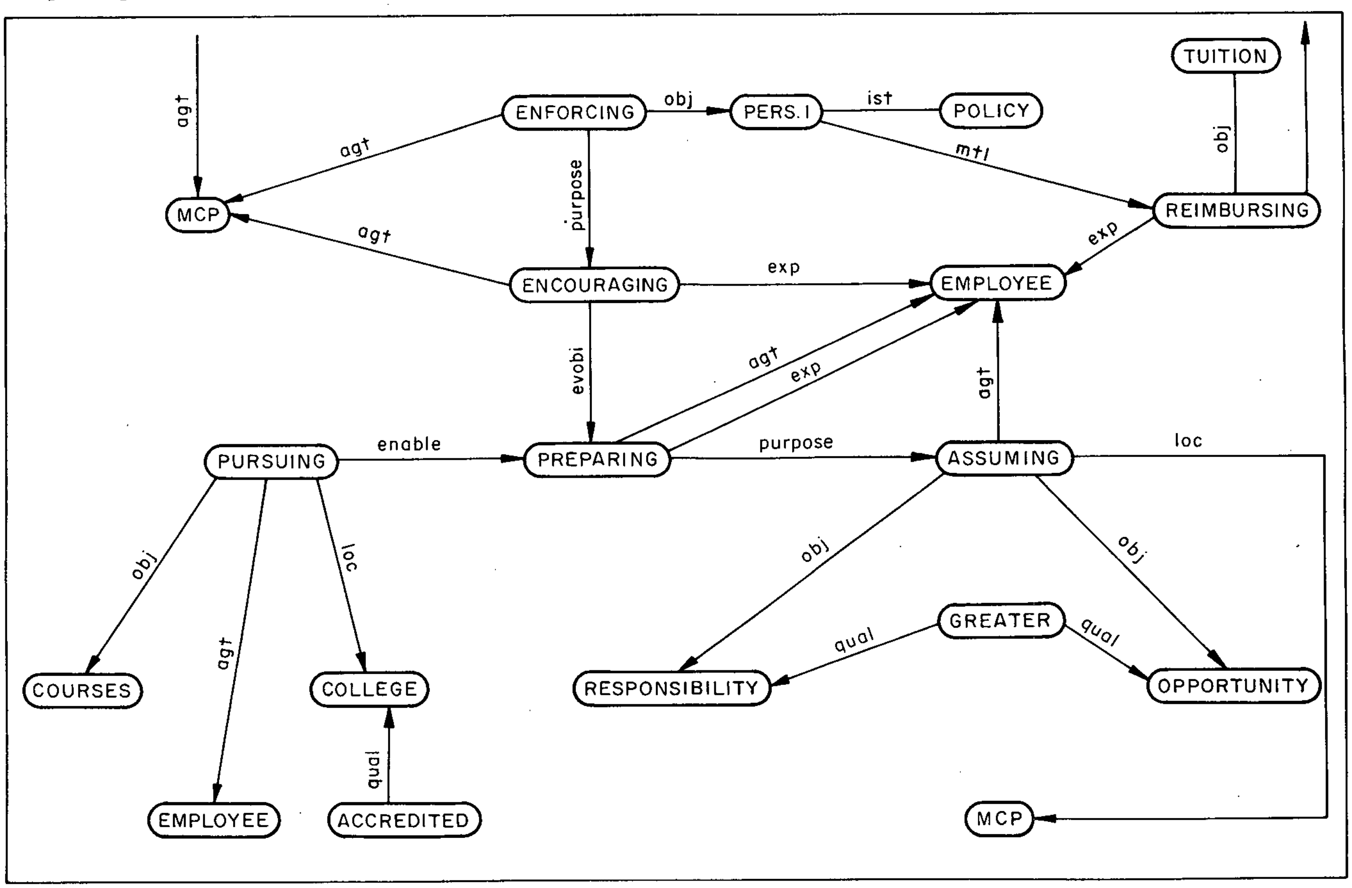

Figure 1. Semantic networks for reimbursement policy (paragraph 1)

been eliminated and the subnetworks have been combined into a final semantic network organized according to production rules. Some of the tuition reimbursement policy paragraphs and their cognitive representations appear below.

Paragraph 1

'The purpose of the PERS-1 Policy is to encourage employees to prepare for greater vocational responsibilities and opportunities within MCP (Medical College of Pennsylvania) by taking courses at an accredited college'.

Analysis

The text omits some facts about the personnel policy: policies are enforced and responsibilities are assumed. The verb 'prepare' should be in its reflexive form. Otherwise we would not know to whom the verb refers. This ambiguity is removed in the semantic net by making 'employee' both AGT and EXP in the event shown in Figure 1. The semantic net can be stored in the IBM data dictionary extensibility database using interactive screens. The concept nodes in the network are entered as dictionary subjects under a universal category *X*. Semantic relations are stored as relationships between these subjects. Details about the implementation are given in Sturdza (1983). The semantic net in Figure 1 can be entered using the following screens:

SUBJECT_ RELATIONSHIP I	X CODE	: ENFORCING : OCC: 0			STAT: P	
RELATIONSHIP	CATEGORY	SUBJECT-2 NAME	OCC	STAT	SEQUENCE	MK
AGT	X	MCP	0	P		
OBJ	X	PERS_ 1	0	P		
PURPOSE	X	ENCOURAGING	0	P		★

Figure 2. Dictionary screens for semantic network of paragraph 1

SUBJECT_ RELATIONSHIP I	X CODE	: ENCOURAGING : OCC: 0			STAT: P	
RELATIONSHIP	CATEGORY	SUBJECT-2 NAME	OCC	STAT	SEQUENCE	MK
AGT	X	MCP	0	P		
EXP	X	EMPLOYEE	0	P		
PURPOSE	X	PREPARING	0	P		★

Figure 3. Semantic network of event 'encouraging'

SUBJECT_ RELATIONSHIP I	X CODE	: PREPARING : OCC: 0			STAT: P	
RELATIONSHIP	CATEGORY	SUBJECT-2 NAME	OCC	STAT	SEQUENCE	MK
AGT	X	EMPLOYEE	0	P		
EXP	X	EMPLOYEE	0	P		
REVUBJ	X	ENCOURAGING	0	P		
RENABLE	X	PURSUING	0	P		
PURPOSE	X	ASSUMING	0	P		

Figure 4. Event 'preparing and its semantic links'

Paragraph 2

'Tuition reimbursement will be approved only for courses beginning after the employee has worked for six months at MCP. The courses must relate to the employee's job.

Analysis

A direct transposal of the text into semantic nets would miss important facts buried in the ambiguities of natural language when it expresses casual chains of events. The paragraph attempts to state the conditions under which reimbursement will be done. However, one of the conditions (antecedent event) appears at the end of the text, after the effect (consequent event). The events in the network must be arranged in such a way as to reflect the flow of events in a manner similar to the 'firing' of production rules. If the course relates to the employee's job, then the truth condition must propagate to the next event which is 'approving' and not 'reimbursing' as the text seems to indicate. The event 'approving' appears to have been skipped in the causal chain. The approval for enrolment in a course must be one of the preconditions for tuition reimbursement. Otherwise, employees could end up taking courses for which they might not be reimbursed. The antecedent events are connected to consequent events by CAUSE arcs:

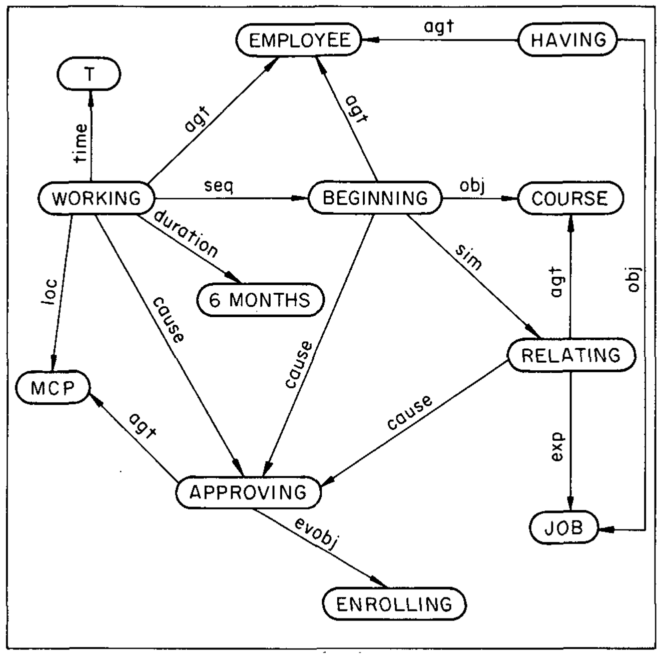

Figure 5. Antecedent events for 'approving enrolment'

Using the dictionary screens we can traverse the CAUSE arc (forward inferencing):

SUBJECT_ RELATIONSHIP I	X CODE	: WORKING : OCC: 0			STAT: P	
RELATIONSHIP	CATEGORY	SUBJECT-2 NAME	OCC	STAT	SEQUENCE	MK
AGT	X	EMPLOYEE	0	P		
LOC	X	MCP	0	P		
DURATION	X	6_ MONTHS	0	P		
START_ DATE	X	XX/XX/XX	0	P		
SEQ	X	BEGINNING	0	P		
CAUSE	X	APPROVING	0	P		★

Figure 6. Event 'working' and its semantic associations

Traversing the RCAUSE arc (reverse cause) results in reaching the antecedent event starting from its consequent (backward inferencing):

SUBJECT_ RELATIONSHIP I	X CODE	: APPROVING : OCC: 0			STAT: P	
RELATIONSHIP	CATEGORY	SUBJECT-2 NAME	OCC	STAT	SEQUENCE	MK
AGT	X	MCP	0	P		
OBJ	X	COURSE	0	P		
RCAUSE	X	WORKING	0	P		
RCAUSE	X	RELATING	0	P		
RCAUSE	X	BEGINNING	0	P		
EVO6J	X	ENROLLING	0	P		

Figure 7. Antecedents for 'approving'

Paragraph 3

'MCP reimburses employees after satisfactory completion of courses. Satisfactory completion of courses requires a final grade of A, B, C or Pass'.

Analysis

'Completion' is the nominalized form of the verb and will thus be considered as an event in which an employee participates as an AGT and 'course' as OBJ. Since the noun has been returned to its original verbal form the adjective 'satisfactory' will take an adverbial form and will qualify the verb (event) 'completing'. The text further defines the meaning of 'satisfactory completion' as being 'obtaining a certain grade.' The first event will be thus defined by the latter and we can use a MTL arc to connect the two:

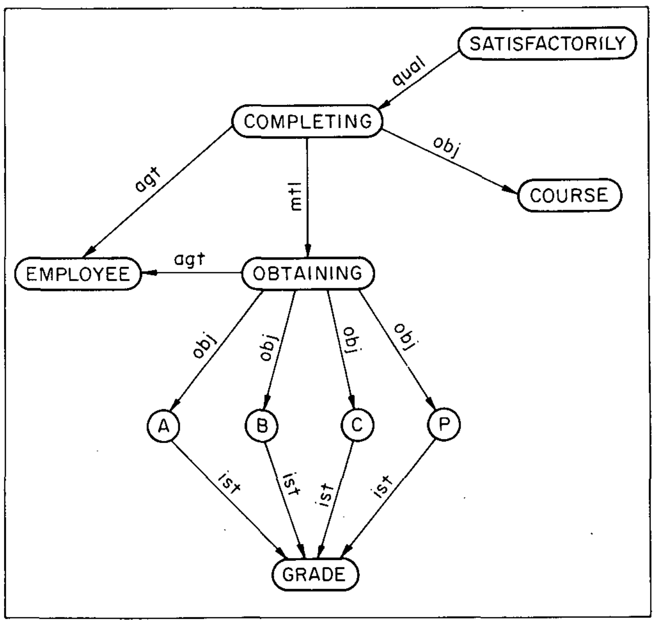

Figure 8. Metalingual definiton for 'completing satisfactorily'

It has to be noted that 'employee' maintains its AGT arc to both events connected by the MTL relationship. The IST (instance) arcs leaving the nodes A, B, C, P, indicate that the OBJects obtained are instances of the general concept 'grade'.

Paragraph 4

'Procedure: the employees must submit a tuition receipt and a grade slip as evidence to the department head which will send them along with a requisition form to the personnel department. The personnel department sends a check to the employee'

Analysis

The paragraph describes the procedures which make up the tuition reimbursement policy (event). The events enumerated are defining the main event 'reimbursing' and are also connected through MTL arcs (Figure 9). The MTL definitions provide an architecture of events where procedural knowledge is organized hierarchically. The highest node in the definition resumes all the information in the lowest nodes.

The following screens store a fragment of the network which appears in Figure 9:

SUBJECT_ RELATIONSHIP I	X CODE	: REIMBURSING : OCC: 0		STAT: P		
RELATIONSHIP	CATEGORY	SUBJECT-2 NAME	OCC	STAT	SEQUENCE	MK
AGT	X	MCP	0	P		
OBJ	X	TUITION	0	P		
EXP	X	EMPLOYEE	0	P		
MTL	X	SUBMITTING	0	P		*
MTL	X	PREPARING	0	P		
MTL	X	SENDING	0	P		

Figure 10. Screens implementing MTL definition for 'reimbursing'

SUBJECT_ RELATIONSHIP I	X CODE	: SUBMITTING : OCC: 0		STAT: P		
RELATIONSHIP	CATEGORY	SUBJECT-2 NAME	OCC	STAT	SEQUENCE	MK
AGT	X	EMPLOYEE	0	P		
OBJ	X	EVIDENCE	0	P		
EXP	X	DEPT_ HEAD	0	P		

Figure 11. Semantic network for event 'submitting'

ANSWERING QUESTIONS WITH ENTERPRISE KNOWLEDGE BASE

The interesting point about expert systems is that the knowledge they contain can be accessed using natural language queries which are translated into a formal language which manipulates the knowledge. Since our Enterprise Knowledge Base does not yet have natural language capabilities, the user must perform this translation himself. He must find what arcs should be traversed in the network in order to answer his questions. The questions most people ask fall into five categories: why, how, yes/no, occurrence and component Lehnert (1977).

'Why' questions embody a sense of causality. In terms of antecedent–consequent rules, it means that the consequent is given and the user must find the antecedent by traversing a RCAUSE arc. Sometimes 'why' questions refer to the purpose of an event and the arc to be traversed is PURPOSE. To answer the question 'why is MCP enforcing the Pers-1 policy?' we access the event 'enforcing' and follow the PURPOSE arc which will give 'encouraging' as an answer (Figure 1).

'Yes/no' questions establish the truth of a statement. In order to answer them we must retrieve the corresponding concepts in the semantic net.

'Occurrence' questions imply that the antecedent is known and the consequent must be found. For example, to answer 'what happens when an employee completes satisfactorily a course' the user must follow a CAUSE arc from 'completing' to 'reimbursing' (see Figure 9).

'Component' questions occur when a concept node in the network corresponding to the question is not known: 'to whom is an employee submitting a request form?'. 'To whom' indicates that an EXPeriencer arc must be followed starting from the event 'submitting'. The arc leads to the concept node 'department head' (Figure 9).

Using the dictionary REPORT command we can specify the starting node and the arc to be traversed. The

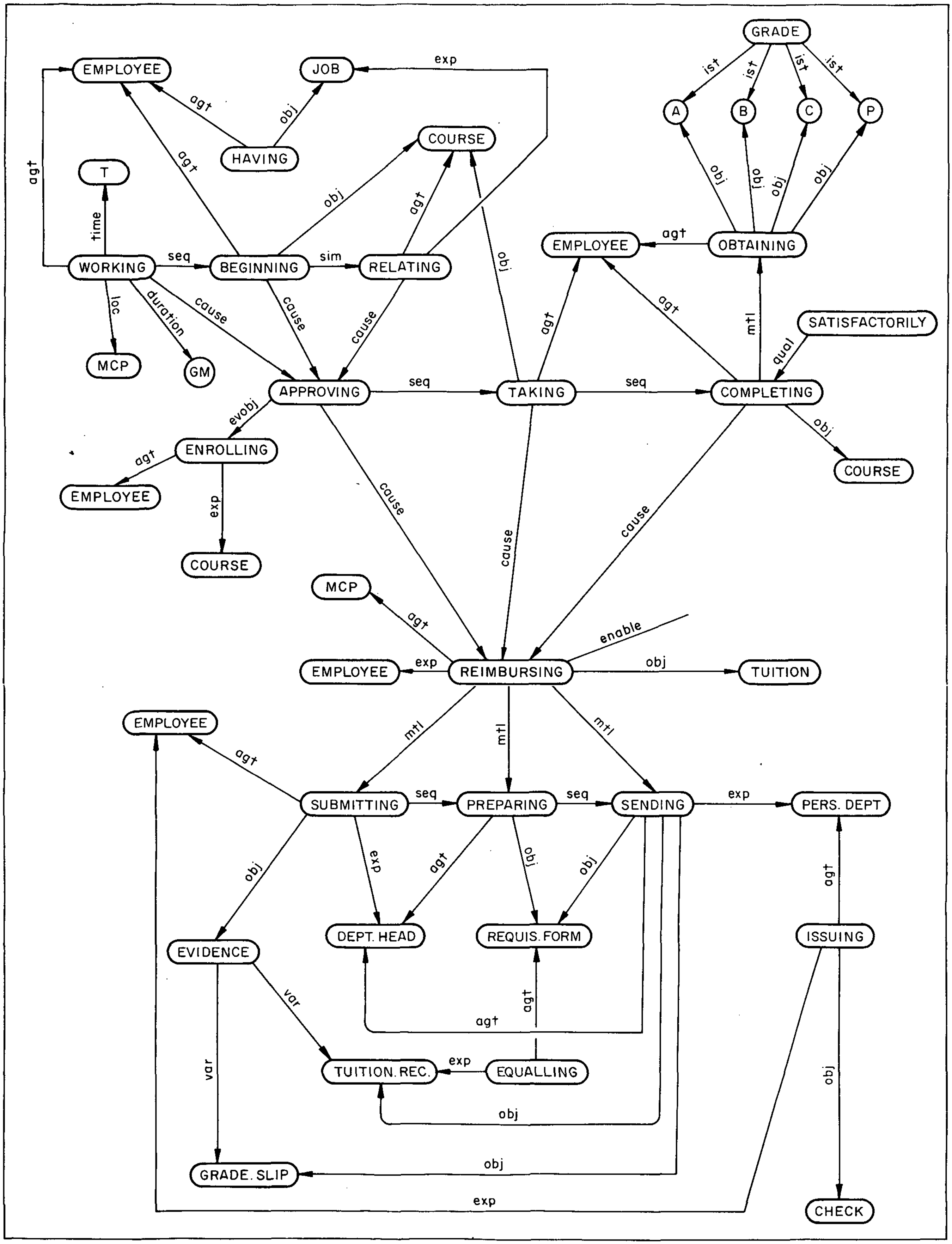

Figure 9. Metalingual definition for 'reimbursing'

dictionary will return only the related concept node since the command focuses on a single relationship ($E \times P$):

DB/DC DICTIONARY COMMAND FORM I
REPORT X (P,,SUBMITTING,0) RELATION = (EXP,X);

Figure 12. 'Report' command focusing on an 'experiencer' relationship

★ ★ ★ ★ ★ ★ ★ DB/CC DATA DICTIONARY REPORT ★ ★ ★ ★ ★ ★ ★				
DICTIONARY DATA BASE = EXT			CATEGORY = X	
NAME:	P	REIMBURSING	0	
DESCRIPTION:				
ATTRIBUTES:	NONE			
RELATIONSHIPS:				
EXP	X	P	PERSONNEL_DEPT	0

Figure 13. Answering a question with 'report' command

'How' questions refer to events which are connected to other events by an ENABLEMENT arc. To answer the question 'how does an employee prepare himself to assume greater responsibilities at MCP' can be answered by following the arc RENABLE from the event 'preparing' to the event 'pursuing':

SUBJECT_RELATIONSHIP I	X CODE	: PREPARING : OCC: 0		STAT: P		
RELATIONSHIP	CATEGORY	SUBJECT-2 NAME	OCC	STAT	SEQUENCE	MK
AGT	X	EMPLOYEE	0	P		
OBJ	X	EMPLOYEE	0	P		
EVOB	X	ASSUMING	0	P		
RENABLE	X	PURSUING	0	P		★
RPURPOSE	X	ENCOURAGING	0	P		

Figure 14. 'How' type of question

SUBJECT_RELATIONSHIP I	X CODE	: PURSUING : OCC: 0		STAT: P		
RELATIONSHIP	CATEGORY	SUBJECT-2 NAME	OCC	STAT	SEQUENCE	MK
AGT	X	EMPLOYEE	0	P		
OBJ	X	COURSES	0	P		
LOCATION	X	COLLEGE	0	P		
ENABLE	X	PREPARING	0	P		

Figure 15. Semantic links for 'pursuing'

USING THE ENTERPRISE KNOWLEDGE BASE FOR DATA DEFINITIONS

In data analysis the concept nodes of the semantic network are data elements. It follows that building an Enterprise Knowledge Base coincides with formally defining these data elements. The definition of a concept viewed now as a data element results from its position in the overall semantic net, a position determined by its relationships to other concepts.

In our application, 'course' is defined as being taken (pursued) by an employee, as relating to his job, and as beginning after the employee has worked for 6 months. This definition reflects the point of view of an institution which wants to reimburse his employees only for courses which improve job skills. The definition can be expanded by adding other events in which a course participates: 'teaching', 'offering', etc. The data element in question could also be classified by adding the arc VARiety to the concept 'activity'. We can say that 'course' has been defined by classification (what a thing is) and by event participation (what a thing does). To the latter we can add the MTL definitions which can be expanded at any desired depth using still other events.

Defining a data element by its position in the semantic net has a significant practical advantage: it eliminates the need to store separately the data element from its definition. In our application example, the semantic net from Figures 9 and 12 can be said to completely define the data element (concept) 'reimbursing'. However, for practical reasons, a smaller portion of the network would suffice as a definition. This definition need not be stored; it should be extracted when needed by a user. Finding a definition for a data element amounts to answering questions about them in the manner described in the preceding section. The user can thus consult the knowledge base with interactive screens and find the meaning of the data elements of interest.

Another advantage for defining data elements with semantic networks is that English statements corresponding to these definitions could be directly generated from the networks according to procedures already used in AI. We are currently working on generating natural language sentences from the semantic nets stored in the data dictionary.

Using semantic nets for data definitions presents still another advantage when compared with the 'of language' or the 'list of permitted aspects' (SYmons and Tijsma, 1982): the analyst is not constrained by a fixed syntax or a fixed number of 'aspects' in defining a data element. A concept (data element), be it entity or event, can be defined using any number of concepts. There are no universal attributes which will characterize in the same meaningful way all entities and events in an enterprise. There is, however, a limited set of universal relationship types between concepts (data elements) such as the semantic cases and inter-event arcs discussed in the present paper.

Having a set of universals as semantic relations produces significant positive consequences for the computer implementation of data definitions. Once this set of relations has been defined and installed in the dictionary, any data element can be entered and its position in the net established. The dictionary user need not install additional categories and relationships for each new application: the same initial set of relationship types and the unique dictionary category X will suffice.

In the data dictionary implementation of semantic networks, each concept (data element) is directly retrievable within a dictionary category X. The classification of data items is made entirely through semantic relationships and not through dictionary categories in which these data elements would be entered. The practical consequence of this approach is that the retrieval of an element from the extensibility database of the data dictionary is made independently of the classification used. One can access 'employee' within the universal dictionary category X, without having to know that the data element has been previously classified as 'person'. By making retrieval of a data item independent of its classification, the dictionary user is allowed great flexibility in browsing through the dictionary: the semantic network can be entered at any node (data element).

The data elements (concepts) within the Enterprise Knowledge Base must be related to the corresponding fields which are simply named locations in programs or database records. A relationship type MANifestation has been established between the data elements which are stored in the extensibility database of the data dictionary and the corresponding field names stored in the standard databases of the dictionary. In our application example, the field which corresponds to the concept 'employee' can be retrieved by entering MAN in the relationship slot no. 5 on the data element screen:

```
X    |        NAME: EMPLOYEE
                OCC:                         STAT:
DESC

ACT: _  1-PROC   2-PROC   3-DESC   4-ALIAS   5-RELATIONSHIPS: MAN
```

Figure 16. Retrieving field corresponding to concepts

In the next screen, the corresponding field is displayed:

```
FIELD |     NAME: EMP. NAME
            CODE:             OCC.              STAT
DESC:
DATE:      BYTES:        TYPE:        DIGITS:        DECIMALS: |
(OBOL & ASSEMBLER DATA) —
 USAGE:        SIGN:          BLANK:         JUST:          SYNCH:
PICTURE:
OCCURS:
 VALUE:
ACT:___   1-PROC   2-REGEB   3-DESC   4-ALIAS   5-SEGS   7-7PL/I   10-HRD.
```

Figure 17. Screen for EMP-name field

The concept/data element which corresponds to a given field can also be retrieved by following the same procedure but in reverse. On the field screen we enter RMAN in the relationship slot and the next screen returns the corresponding concept:

```
X    |        NAME: EMP_NAME
                OCC:                         STAT:
DESC

ACT: _  1-PROC   2-PROC   3-DESC   4-ALIAS   5-RELATIONSHIPS: MAN
```

Figure 18. From field to concept

SUBJECT_RELATIONSHIP |		X CODE	: EMPLOYEE : OCC: 0		STAT: P	
RELATIONSHIP	CATEGORY	SUBJECT-2 NAME	OCC	STAT	SEQUENCE	MK
AGT	x	WORKING	0	P		
EXP	x	REIMBURSING	0	P		

Figure 19. Semantic network for concept 'employee'

USING ENTERPRISE KNOWLEDGE BASE FOR DATABASE DESIGN

The Enterprise Knowledge Base not only contains data definitions and systems analysis information but also can function as a conceptual database schema. There is unanimous agreement that the conceptual schema should represent the inherent properties of data independently of the computer implementation. However, there is considerable disagreement concerning the content of such a schema. For us, the conceptual schema of a database is a portion of the semantic network which represents the enterprise knowledge necessary for a database application.

Since commercially available DBMS do not support a conceptual schema, a conventional schema must be derived. In order to map a conceptual schema into a conventional schema, a canonical schema is a nonredundant network of data elements and their associations used to deduce a logical model of keys, attributes, segments and hierarchical paths (Raver and Hubbard, 1977).

The canonical schema contains groupings of data elements into canonical records (segments) and essential associations between these data elements. An essential association is a type 1 association that provides the only path between two data items. A type 1 association is one in which every occurrence of the 'from' element identifies one occurrence of the 'to' element (Raver and Hubbard 1977). Data elements from which essential associations emanate are candidates for keys.

The semantic net in Figure 20 represents a portion from the knowledge base about the tuition reimbursement policy. The database designed from the net must answer queries concerning employees reimbursed, amount of reimbursements and for what courses.

We will introduce the convention that entities and events to be incorporated in the database design will be represented in the semantic net by their type name and ID number if one is assigned. Thus, 'employee-no' represents the entity 'employee' and will manifest itself as a field in the database. The employee name is a property and qualifies the 'employee-no' data element. In this way, all nodes in the semantic net are representable as fields in the database records if necessary.

The semantic nets can be used to derive the structural model (Raver and Hubbard, 1977) in which type 1 associations are identified. The QUALIFIED_BY semantic relation from 'employee-no' to 'employee-name' is a type 1 association and is marked with a dotted line (Figure 21). The semantic relationship between 'job-no' and 'employee-no' appears as an event node (working) plus the two case relationships which will map into a single type 1 association. A similar situation exists between 'course-no' and 'employee-no' (Figure 21). In the next step, the transitive and the redundant type 1 association are removed from the structural model and the canonical schema is designed (Figure 22).

'Employee-no' is the root key and 'reimbursing-no' is the initial key. The forward type 1 associations are of type M backward and establish thus $M:1$ mappings. These mappings define a parent–child relationship between the pair of related keys which can be used to implement the above canonical schema as an IMS hierarchy of segments (Figure 23).

CONCLUSIONS

An augmented semantic network model has been used to

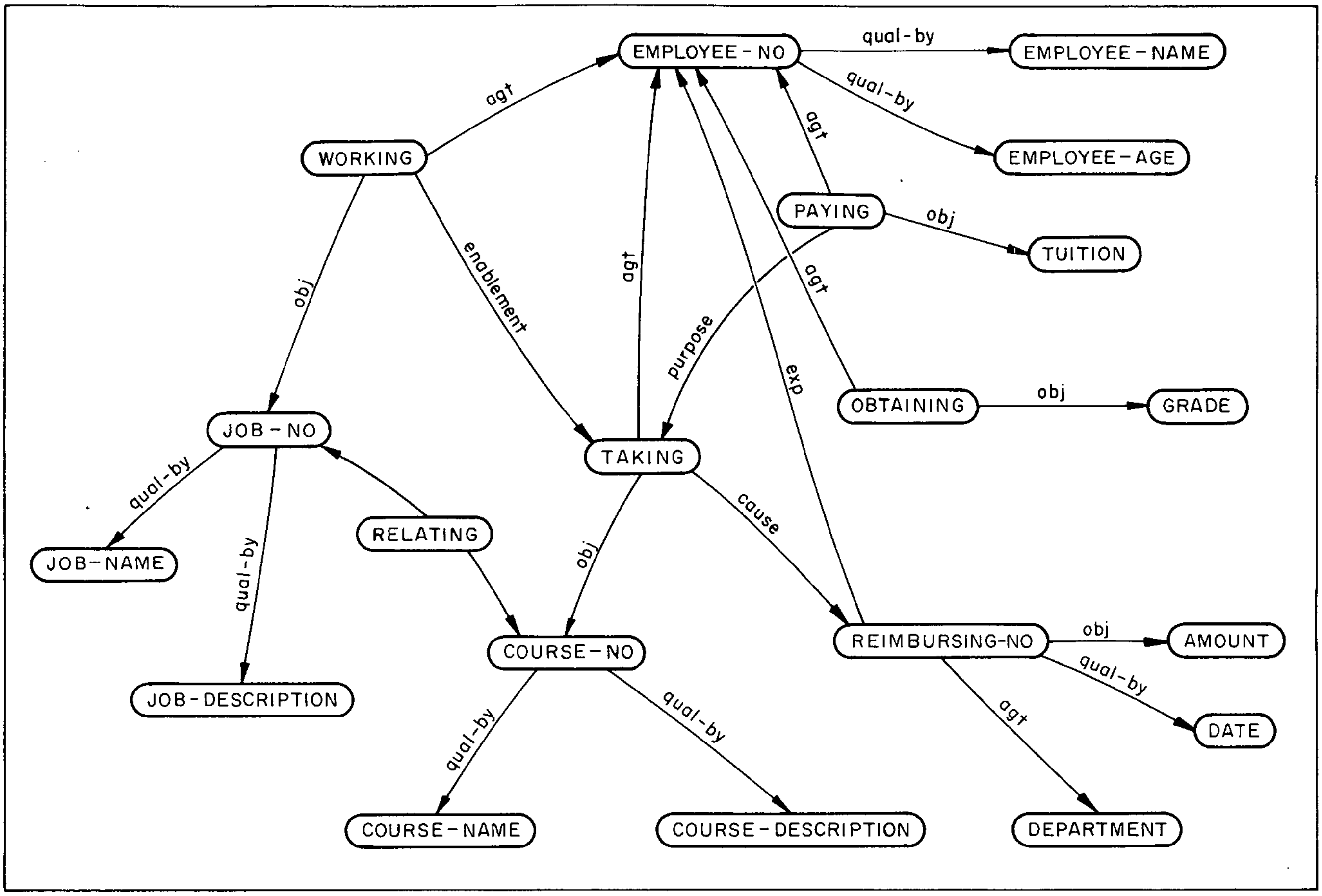

Figure 20. Portion of semantic network for reimbursement policy

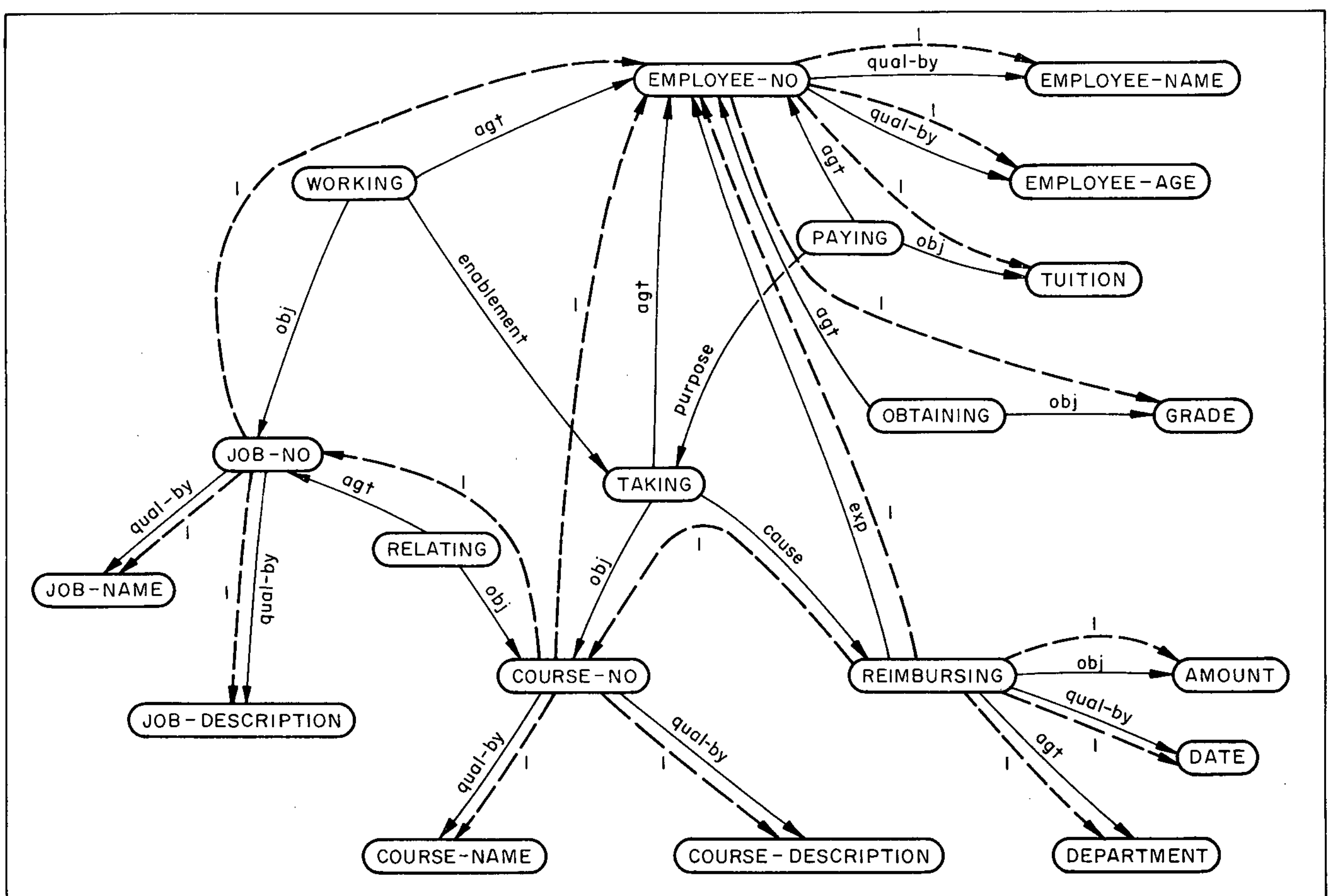

Figure 21. Marking type 1 association of semantic network

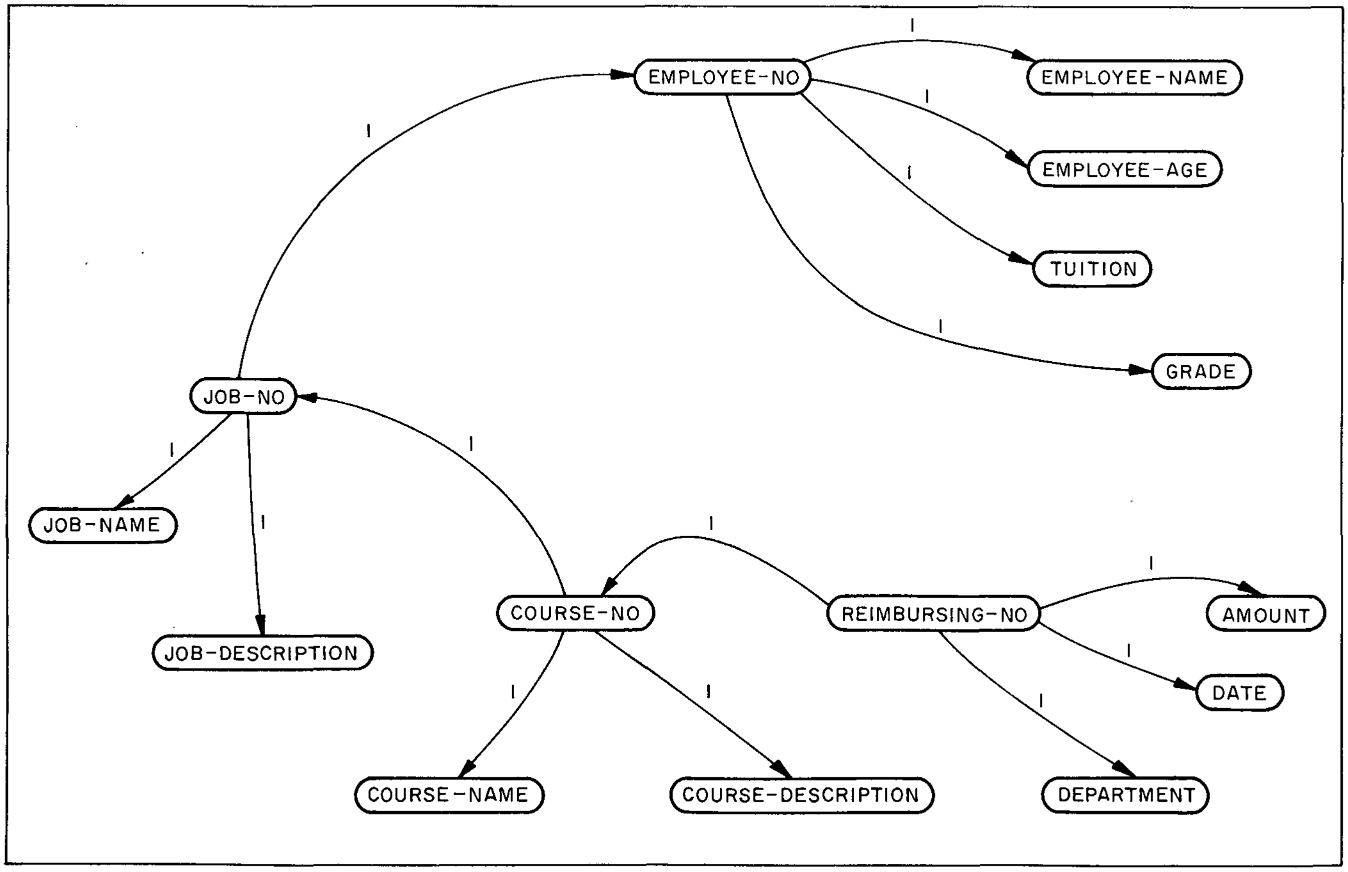

Figure 22. A canonical schema derived from a semantic net

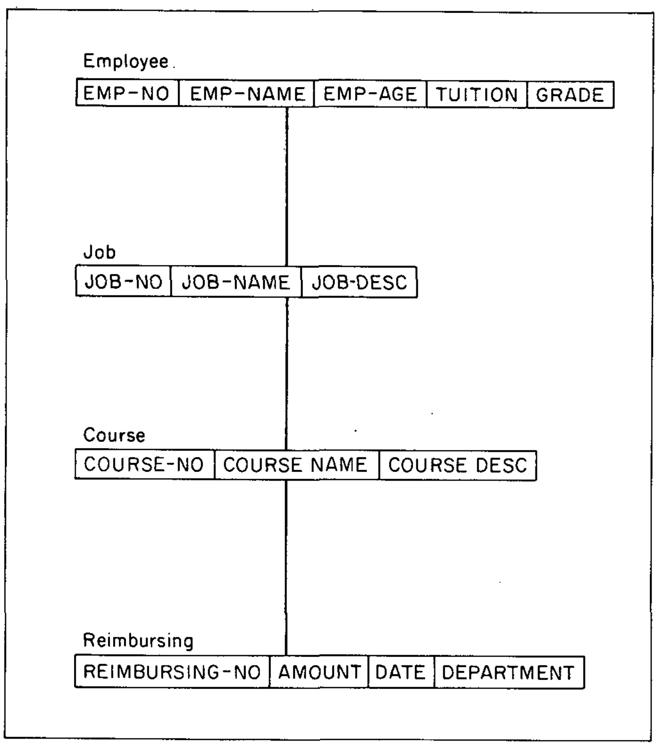

Figure 23. An IMS tree derived from canonical schema

build an Enterprise Knowledge Base for systems analysis, data analysis and database design. The Enterprise Knowledge Base has been implemented with a commercially available software package, the IBM data dictionary. In the process, the dictionary has acquired new functional capabilities which considerably increased its productivity. It is hoped that the application described here will contribute to a more useful exchange of ideas between AI and business data processing.

REFERENCES

Brodie, M and Mylopoulos, J *On Conceptual Modelling Perspectives from Artificial Intelligence, Databases, and Programming Languages* Springer-Verlag (1983)

Chen, P *The Entity-Relationship Approach to Systems Analysis and Design* North-Holland (1980)

Goldman, N 'A relational database foundation for process specification' *The Entity–Relationship Approach to Systems Analysis and Design* (P Chen, ed) North-Holland (1980)

Greimas, A *Semantique Structurale* Libraire Larousse (196?)

Hays, D 'Dependency theory: A formalism and some observations' *Language* (1971)

Lehnert, W 'Human and computational question answering' *Cognitive Science* (1977) pp 47–73

Hammer, M and McCleod, D 'Database description with SDM: A semantic database model' *ACM Transactions on Database Systems* (1981)

Mobsiii — Product Description ISE-USA
Muntz, R 'Design of the well system' *The Entity–Relationship Approach to Systems Analysis and Design* (P Chen, ed) North-Holland (1980)
Phillips, B *Topic analysis* PhD Thesis Department of Linguistics, Suny at Buffalo (1975)
Symons, C R and Tijsma, P 'A systematic and practical approach to the definition of data' *The Computer Journal* (1982)
Sakamoto, J *Use of DB/DC data dictionary to support business systems planning: An approach* IBM Scientific Center (G320–2705) Los Angeles (1980)
Raver, N and Hubbard, G 'Automated logical database design: Concepts and applications' *IBM Systems Journal* (1977)
Sturdza, P 'Data dictionary design with an artificial intelligence model' *Proceedings of the 15th International Conference on System Sciences* pp 406–4 University of Hawaii (1975)
Tesniere, L *Elements de Syntaxe Structurale* Librairie Larousse (1959)
Wilson, M *A semantics-based requirements and design method* IBM Santa Teresa Lab (TR 03.072) (1979)
Winston, P *Artificial Intelligence* Addison–Wesley (1979)
IBM DB/DC Data Dictionary Applications Guide SH2 0–9173