ABSTRACT

Language Choice and Code-Switching

Dr. Muhammad Ali Alkhuli
King Saud University/Riyad

This paper aims at defining Language Choice (LC), illustrating factors that determine LC, and showing the relationship between LC and style choice and dialect choice. The paper has also explicated the elements of the LC process and given some LC instances. The paper has especially emphasized these LC factors : the speaker's and listener's abilities, conversational habits, age, social status, conversers' degree and quality of relation, social pressure, usage distribution, topic, prestige gain, social-distance increase, and exclusion.

The paper also aims at defining Code-Switching (CS) and illustrating the difference between CS and each of iterference and LC. Further, it aims at showing the types, system, and objectives of CS. The paper has especially emphasized these objectives : influence, need, continuity, quotation, selecting the addressee, belonging, special emotion, confidentiality, exclusion, social-distance increase, and emphasis.

Newell, A. and H. Simon :
Human Problem Solving Englowed pentick-Hall 1972.

Waterman, D.A. and F. Hayes Roth :
Pattern Directed Inference System, Academic Press, 1978.

Feigenbaum, E.A. :
«The Art of Artificial Intelligence Themes and Case Studies of Knowledge Engineering», AFIPS Conf. Proc., vol. 47, June 1978, pp. 227-240.

Shortliffe, E.H. :
Computer Based Medical Consultation MYCIN, New York, Elsevier 1976.

Feigenbaum, E.A. et al. :
«On Generality and Problem Solving : A Case Study Using the DENDRAL Program», Machine Intelligence 6, B. Meltzer and D. Michie (Editors), N.Y., Elsevier, 1971, pp. 165-190.

M.C. Dermot :
«RI : An Expert in Computer System», Proc. of the 1st National Conference on AI, August 1971.

People, H.E. et al. :
«DIALOG : A Model of Diagnostic Logic for Internal Medicine», Proc. of the 4 th Int. Joint Conf. on AI, Sept. 1975, pp. 848-855.

Stallman, R.M. an G.J. Sussman :
«Forward Reasoning and Dependency-Directed Back-Tracking in a system for Computer-Aided Circuit Analysis», Artificial Intelligence, Vol. 9, 1977, pp. 135-196.

Clarcey, W.J. et. al :
«Intelligent Computer Aided Instruction for Medical Diagnosis», Proc. of the 3rd Symposium on Computer Application in Medical Care, 1979, pp. 175-183.

Ramadan Abdel Towab :
Introduction to linguistic and Linguistic research topics, Dar Al Refaie, Riyadh.

Bryle, Mousa :
«Dictionary of Arabic Newspapers», Al-Quds, 1940.

Baily, E.M. :
A list of Modern Arabic Words, Cairo, Niel Mission Press.

Akel Fakhr :
«Basic Vocabulary for Elementary Education». Damascus, Syria 1953.

Landan Jacob M :
A Word Count of Modern Arabic Phrase, New York, American Council of Learned Societies, 1959.

Lotfi, M., K. :
Changes Needed in Egyptian Readers to Increase their value (Ph. D. Dissertation). Chicago : Chicago Univ., 1948.

Abdu, D. :
«Common Vocabulary of Arabic Language», Riyadh University 1979.

Al-Khuli, Dr. M.A. :
«Statistical Analysis of Arabic Sounds», Al-Ribat, (to be published).

Al-Khuli, M.A. :
«Common Structures in the Arabic Language», Riyadh, Dar Alolum Press.

Al-khuli, Dr. M.A.,
«Analytic Review Study for the Arabic Vocabulary», Riyadh, Drasat Magazine, King Saud University, second Issue. pp. 27-66, 1973.

Ibn Manzoor,
Dictionary of «Lisan ALARAB». Cairo Ameriya Press.

Mousa, Dr. Ali,
«Statistical Analysis of the roots of Lisan «Arab Dictionary». Kuwait, University of Kuwait, 1972.

Mousa, Dr. Ali,
«Statistical Study for the Triliteral Roots of Arabic Vocabulary». Kuwait, University of Kuwait, 1971.

Mousa, Dr. Ali,
«Statistical Study for non-triliteral Roots of Arabic Vocabulary» Kuwait, University of Kuwait 1972.

REFERENCES

IEEE Proceedings :
Special issue on natural language processing, Vol. 74, n° 7, pp. 897-1056, July, 1986.

Neils, J. Nilson :
Symbolic Computation : «Principles of artificial Intelligence».
Springer-Verlag Berlin, New York, 1982.

Gevarter, W.B. :
«An Overview of Computer-Based Natural Language Processing», NBSIR 83-2687, National Bureau of Standards, Washington, DC, APR. 1983.

Hendrix, G.G. Sacerdoti, FD, Saglow icz, D :
«Developing a Natural Language Interfaces to Complex Data, ACM Trans on Database System», Vol. 3, n° 2, June 1978.

Hendrix, G.G., and Sacerdod, E.D. :
«Natural Language Processing : The Field in Perspective», Byte, Sept 1981, pp. 304-352.

Berwick, R :
«Locality Principles and Acquisition of Syntactic Knowledge» Ph. D Thesis, MIT, Cambridge, MA, 1982.

Marcus, M.P :
«A Theory of Syntactic Recognition For Natural Languages», MIT Press, Cambridge, MA, 1980.

Kay, M. :
«Algorithm Schemata and Data Structures in Syntactic Processing, Processing of Symposium on Text Processing», 1980.

Burton, R.R. :
«Semantic Grammar : An Engineering Techniquest for Constructing Natural Language Understanding Systems», BBN Rep 3453, Bolt, Beranek and Newman, Cambridge MA, 1976.

Shalom M. Weiss & Casimir A. Kuiikawski :
«A Practical Guide to Designing Expert Systems», Rowman & Allanheld, Publishers, 1984.

Slocum, Jonathan 1984 :
«Machine Translation its History and Current Status and Future Prospects», Austin : LRC working paper LRC-84-3.

Winfred P. Lehmann :
«Computer Aided Translation, State of the Art in the U.S.A. and Canada» Proceedings of SANCST Saudi International Work shop on MAT, Riyadh, March 1985.

Jonathan Slocum & Aristor :
«The Treatment of Grammatical Categories and Word Order in Machine Translation», Proceeding of SANCST Work shop on MAT, Riyadh, March 1985.

K.S. Fu :
«Applications of Pattern Recognition», CRC Press, Boca Raton, Fab. 1982.

K.S. Fu :
«Pattern Recognition and Computer Vision», Computer Magazine, October 1984.

IEEE Computer Magazine :
Special issue on knowledge representation, October 1983.

Duda, R.O. and Gasching, J.G. :
«Knowledge-Based Expert System Come of Age», Byte, sep. 1981, pp. 238-280.

Barstow, D. R. :
«An Experiment in Knowledge-based Automatic Programming» Artificial intelligence, Vol. 12, August 1979, pp. 7-119-

Davis, R. :
«Interactive Transfer of Expertise : Acquisition of New Inference Rules» Artificial Intelligence, Vol. 12. August 1979, pp. 121-157.

Duda, R. O. and J. G. Gaschning and P. E. Hart, :
«Model Design in the Prospector Consultant System», Edinburg University Press, pp. 153-167.

b) defining what lexical items are potentially possible within the language/range.
c) defining what forms are impossible to be introduced into the language.
— The KBS can be helpful in developing children's literature as well as simplified science books.
— Determining the most recurrent Arabic syntactic structures.

Conclusion

This paper incorporates a proposal for a knowledge base system for Arabic as a natural language. It has been an attempt to draw the attention of those interested in the area of Arabic linguistics, as well as related studies and research, to the vast potential of recent advances in computer science technology. This technological potential, we have seen can now accomplish once incredible analytic feats when driven by the current theoretical advance in the field of linguistic science.

We believe that serious computational linguistic research is long overdue if Arabic is to be analyzed and understood thoroughly. Furthermore, this research should be done in the Arab World in order to have the proper and appropriate grounding, and it should be distinct from any commercial interests or considerations.

The goal of this paper was to have opened wide the door for discussion of problems. It was also intended to initiate a cooperative effort towards realizing the ultimate aim of understanding and analysing the language of the Holy Quran.

The debate that this proposal is expected to engender and encourage is presented in the hope that it will sharpen the central idea put forward. Finally, this paper has been presented in order to crystalize numerous details.

* * *

0 Frequency of the various consonants and vowels (e.g. tha, wa, on, oo... etc.) ;
0 Sequence of occurence of various sounds ;
0 Relation to various morphological structures ;
0 Permissible sound clusters ;
0 Syllable structures ;
0 The relation between vowels and consonants.
4) Statistical information about syntactic structures :
0 Marked VS unmarked forms ;
0 Percentage of grammatical words.
0 Frequency of various Arabic Syllables (e.g. fi, ala, ln ; ila etc.)
0 Determining syntactic exceptions and deducing the constraints preventing their generation ;
0 Word order (sentence initial position).

The above are just a few examples of possible statistical results that may be collected during this phase of project. Other studies may be developed and considered as required by the research.

Third Phase : In the third phase, different analyses are to be made of the statistical information collected from the previous phase. These include :

1) Analysis to deduce various linguistic theorems, such as :
0 Finding analogical rules for irregular morphological structures. (e.g. triliteral verbs, and their gerunds) ;
0 Laws for phonemic arrangement ;
2) Analysis of common vocabulary, common sounds, etc., in order to determine reasons for common use of certain words, or sounds, and rare use of others ;
3) A comparative analysis of the project's results with the statistical results obtained from previous studies.
4) A comprehensive review of various results in order to determine the degree of relevance to the representation of Arabic knowledge base. As was explained in the literature section, the acquiring of a base from various sources such as experts, books, or empirical data-analysis, undergoes several stages. These stages can be characterized as (1) problem-identification, (2) conceptualization, (3) formalization, (4) implementation, and (5) testing (as shown in figure 1). However, the process is not as neat and well-defined as the figure might suggest. These stages are simply a rough characterization of complex and ill-structured activity that takes place during knowledge acquisition. Generally, the first step in acquiring knowledge for an expert-system is to characterize the important aspects of the problem. This involves identifying problem characteristics, resources and goals.
During conceptualization phase, the key concepts and relations are made explicit.

The following questions need to be answered before processing with conceptualization process :

- What types of data are available ?
- What is given and what is inferred ?
- Do the strategies have names ?
- Are there identifiable partial hypotheses that are commonly used ? What are they ?
- How is the object in the domain related ?
- Is it possible to label causal relations, set inclusion, part whole relations, etc ?
- What does it look like ?
- What is the information flow ?

The formalization process involves mapping the key concepts and data-analysis of characteristics into more formal representation based on various knowledge-engineering tools or framework.

Understanding the nature of data in the problem domain is important in formalizing knowledge. If the data can be explained directly in terms of certain hypotheses, it is useful to know if this relationship is causal, definitional or merely correlational. Such knowledge may help explain how hypotheses that directly explain data can be related to other high-level hypotheses and how these hypotheses relate to the structure of goals in problem-solving process itself. The results of formalizing the conceptual information-flow and sub-problem elements will be basic specifications for knowledge-base systems.
The utilization of empirical analysis of the Arabic language is necessary through this stage of the development of the knowledge-base system. First, three stages of development of the KBS for Arabic (problem identification, conceptualization, formalization) will be attempted during this phase of the project.

5) The experience gained through the work in this phase will provide insight into the development of several applications. This in turn allows the formulation of a plan for various uses of the Arabic knowledge base systems.

Possible applications to the Arabic language to be investigated may be selected from the following :
— the computer-aided Instruction of Arabic ;
— a system for machine translation to/from Arabic ;
— a searching for new linguistic theorems and rules ;
— developing an Arabic speech synthesis system ;
— a question-Answering System ;
— Natural Language Interfaces
— Information retrieval
— Machine Translation (MT) ;
— data base interfaces
— Arabic Computer Design and Architecture ;
— the vocalization of Arabic text by computer ;
— providing a solid basis for tackling problematic areas in Arabic grammar, such as marked VS unmarked forms, irregular and other peculiar forms in Arabic (studying broken plurals will be one example of this).
— Creating a network for the derived verbal and nominal forms in the language with the following aspects in view :
a) defining the inventory of the language (i.e. what lexical items actually exist).

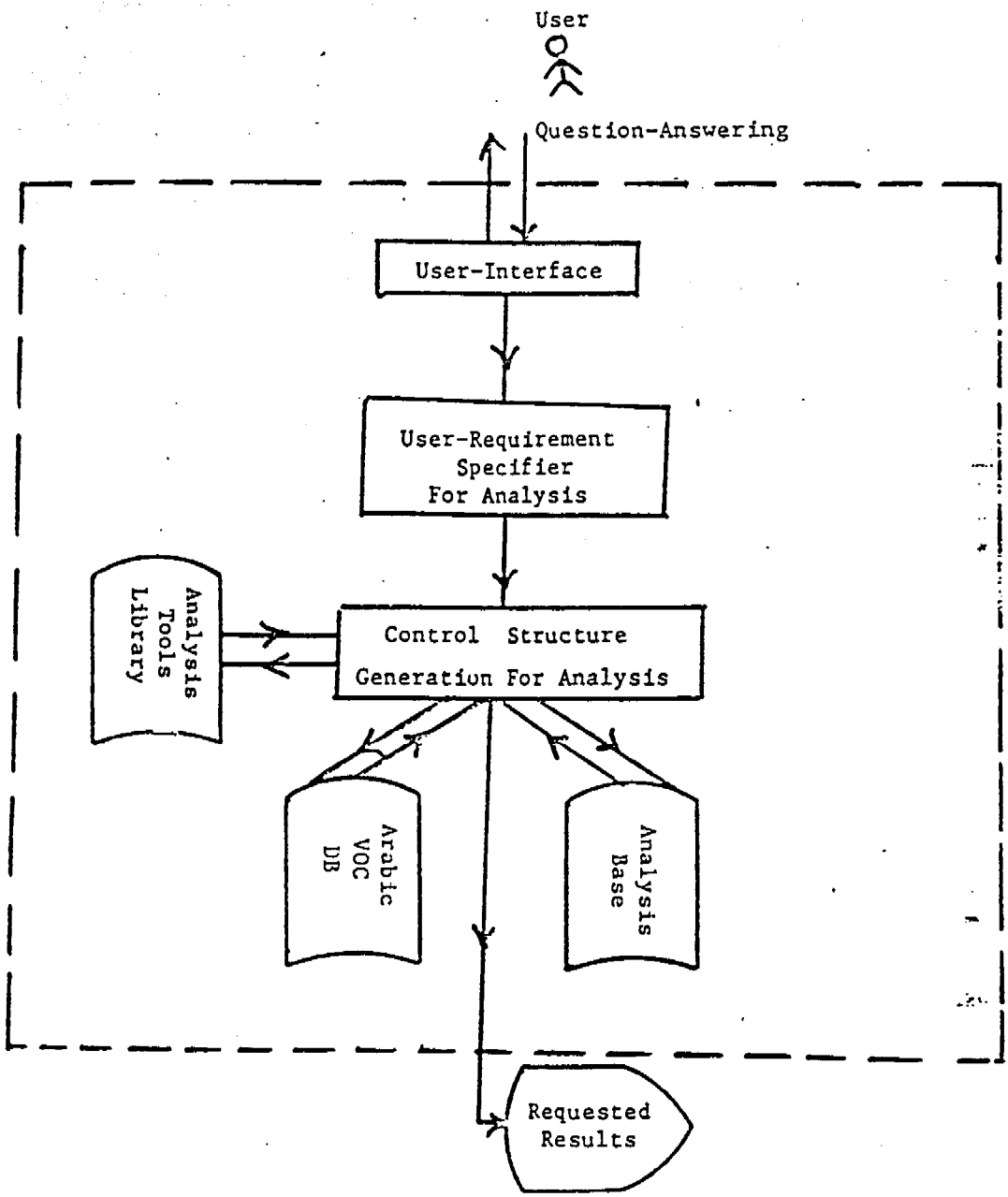

Figure 2 : Block-Structured Representation of SATA.

Applications and Utilization

The construction of empirical & statistical information for the Arabic language by the proposed research will have several direct applications which can be enumerated as follows :

1) Once the 2000 most frequent words have been selected the following can be achieved :
   a) Pocket dictionaries can be compiled for quick reference. This item is currently missing in the Arabic library.
   b) Writing textbooks for Arabic primary education. The majority of the current texts do not concentrate on the most commonly used forms.
   c) Teaching Arabic to non-native speakers. This project will make easy the learning of the most recurrent 2000 words by foreign learners.
2) Development of Arabic Educational books for elementary schools students and for non-Arabs that include commonly used vocabularies and structures suitable for various educational level.
3) Evaluation of educational books regarding their use of rare vocabulary, or complex and uncommon syntactic or morphological structures.
4) Design of various language examinations linguistic tests in particular.
5) The results of the project also constitute the first necessary step in a multi-stage plan to develop a KBS for the Arabic language. This system, upon completion, will assist in developing very important applications.
6) The application plan to be formulated will estimate the uses of the Arabic knowledge system for Arabic linguistic research. The plan will list numerous possible applications of the Arabic knowledge base that may be utilized by Arabic scientists.

Research project design

In order to achieve the planned objectives, this proposed research project should go through three main phases :

First Phase : In the first phase, hundreds of samples of modern Arabic texts are to be gathered from various Arabic sources (e.g. newspapers, scientific and literary articles, educational books, engineering, organization, computer installations, Data Processing (DP) manuals, government offices, industrial environments, texts... etc). These texts should be examined carefully in order to select the most appropriate ones for statistical analysis. The total word volume of the texts have to exceed one million words in order to fully represent the language. A team of Arabic specialists, should then work on vocalizing these texts. This is done by adding the vowel diacritics to the characters of each word, according to Arabic syntactic and morphological rules. In the meanwhile, another effort is to be made to prepare a super microcomputer or a microcomputer system to accept and process Arabic vocalized writing. Several data-base programs are to be developed to store and organize them in computer files.

By the end of the first phase, all vocalized texts will have been entered into the computer storage and will then be ready for the statistical analysis. The first goal of this project (the formation of Arabic vocabulary data base) will then be realized.

Second Phase : During this phase, an automatic analyzer SATA consisting of several tools should be developed. [Fig. 2 shows a block structure representation of SATA]. This software system should utilize state-of-the-art techniques both in software engineering and AI. Formal techniques and data-structuring from natural language processing are to be employed. [Ref. 1-10].

Heuristic Search [1] strategies must be implemented to ensure controlled processing time with a gradual extension of the Arabic vocabulary data base. This is expected to provide greater flexibility for continuous experimentation (which is the basic characteristic of knowledge base systems) for several researches in this area. This analyzer should provide a user-oriented, simple interface for management of an Arabic Vocabulary Data Base (AVDB) to perform empirical analyses on raw data. SATA may act as a front-end for development of a knowledge-base system in several application areas. The design of SATA should allow easy implementation of any further analysis. The investigators may program (SATA) by involving several tasks from the SATA library to perform analyses of desired ends. This is anticipated to include statistical information required for educational applications as well as the Arabic knowledge base. Examples of the empirical data to be extracted by using 'SATA' are :

1) The most commonly used Arabic words. These are the words that have the highest frequency among all words stored in the AVDB. This vocabulary will then be organized in different ways to obtain various statistical results :
   - Word-class (e.g. verb, noun, articles, adverbial,... etc.) ;
   - Roots (triliteral, quadriliteral...) ;
   - Kind of sociolinguistic register (e.g. scientific article, poetry, ... etc) ;
   - Semantic features ;
   - Other classifications ;
2) Various morphological statistical data such as :
   - Frequency of nouns, active participles, diminutive aspect, etc. ;
   - Frequency of ( ) article
   - Probability that a word with a given number of characters is a verb, is a noun, etc.
   - Percentage of triliteral verbs to all words and to all verbs.
3) Statistical information about Arabic sounds may include :

textbooks used by third grade students in 1948. [35].

- Dr. Douad Abdo, in 1979, prepared a list based on several previous studies of the most commonly used Arabic vocabulary. [36].
- Other studies made on various Arabic structures include :
  - A study by Dr. Mohamad Ali Al-Khuli, in 1978, on a text of a 1000 words for the frequency of the various morphological structures and phrases types. [37-39]. In 1982, he also wrote a book (18) about the statistical results gathered from a 5000-word text.

For the frequency of the Arabic sounds, the following studies were conducted.

- Ibn Manzoor, seven hundred years ago, in his famous book *Lisan al Arab*, classified Arabic sounds into three main categories based on their frequencies. [40].
- Dr. Ali Mousa, in 1971 and 1972, counted the sounds of Arabic roots from the *ALSIHAH* dictionary compiled by Al-Jwahiri a thousand years ago [41-43].

Most of the previous studies had the following shortcoming :

1) The samples chosen for analysis were relatively small, and not accurate enough for serious application.
2) The samples of the previous studies were all confined to particular linguistic use (e. g. literature, poetry, and so on). Thus, their results were not representative of the general functions of the language.
3) Most of the past studies were also limited to certain sources, such as newspapers, elementary school books, etc. Consequently, the statistical results obtained were not at all sufficient to fully represent the statistical distribution of the Arabic Language.
4) The objectives of those studies were far from being well-defined ; for this reason there was a wide diversity among them, and there was no common link between the various results obtained from those studies ; besides, much important statistical data were overlooked.

In conclusion, because of the several deficiencies described, previous statistical studies on the Arabic language were inadequate to the task of building on Arabic knowledge base. In order to build a comprehensive statistical-body for this base, the proposed study should have the following characteristics :

1) Large samples, with a total volume of about 1,000,000 words.
2) Text samples should be chosen carefully from various sources, covering different areas. This makes it a truly representative of a wide cross-section of the language.
3) Among other things, this study aims at collecting the statistical-data required for an Arabic knowledge base system. A careful analysis is necessary to direct the study to achieve its objectives.

Research objectives

The proposed research aims to provide insight to the knowledge base of the Arabic KBS. For this, it should concentrate on developing the statistical and empirical information required for the knowledge base and the numerous applications of the system to the Arabic language. The objectives can, therefore, be summarized succinctly as follows :

1) The development of a comprehensively distribution data base of Arabic vocabulary. The base will have a total volume of about one million words selected after careful evaluation of different modern Arabic written texts. The base will form the raw material necessary to conduct various studies on the Arabic language.
2) The development of a System for Arabic Text Analysis (SATA), consisting of several state-of-the-art, software tools that will provide user-oriented, simple interfaces for management of the Arabic data-base.
3) The use of SATA for analysis of the vocabulary data base in order to :

— determine and select the most commonly used Arabic lexical items, their frequency, and other related statistical information.
— The provision of comprehensive statistical data for various Arabic syntactic structures.
— the gathering of various statistical information concerning Arabic sounds, such as their phonemic arrangement, etc.
— the facilitat' ON of prospective morphological studies of Arabic.

4) A comparison between the results obtained from this study and those from previous studies.
5) The selection of the data required for the Arabic knowledge base from the large collection of statistical information obtained through this work.
6) The use of the experience gained from the previous objectives to develop a detailed plan for potential applications of this proposed knowledge base system for in Arabic linguistic research. This plan will include :

— Potential applications of the proposed knowledge base system.
— Actions required to implement, absorb, adapt, maintain and create Arabic knowledge-base and its system applications.

Formalization involves mapping the key concepts and relations into a formal representation suggested by some expert system tool or language. The language should be selected to represent basic concepts and relations within the language framework.

During implementation, the formalized knowledge is combined and reorganized to make it compatible with the information flow characteristics of the problem. The resulting set of rules and associated control structure define a prototype software capable of being executed and tested.

Finally, testing involves evaluating the performance of the prototype software and revising it to conform to standards of excellence defined by experts in the problem domain.

A Knowledge-Base System For Arabic Language

Despite several advantages offered by KBSS and the great progress made in their development, no such system has ever been developed for the Arabic language. This has mainly been due to the lack of knowledge about new computer scientist with little linguistic background and expertise far. It is, therefore, a necessity to have a common effort by experts in both subjects (linguistic topics and computer systems) in order to develop a KBS for the Arabic Language. Such a system will be a very powerful tool for analysing and solving major problems that limit progress in the various fields of the modern Arabic research. It will also serve as the base for numerous very important computer applications in the Arabic Language (e. g. machine translation, Arabic speech synthesis, CAI... etc.) These are among the many greatly needed applications of a KBS for the Arab world in general, and for Saudi Arabia in particular. The important goal here is to implement and locally these latest developments in various fields. In order to appreciate the usefulness and importance of an Arabic KBS consider the following current research topics in the Arabic language : [30].

1) Research in Arabic phonology to analyze Arabic sounds, their laws, characteristic, outputs... etc. ;
2) The study of Arabic syntaxe and the analysis of the roots of Arabic vocabulary ;
3) The study of Arabic morphology and affixation.
4) The etymology of Arabic vocabulary and the history of the Arabic language.
5) The comparative linguistic research with reference to Arabic and other world languages, exhibiting similarities and differences.
6) The lexicography and dictionary compilation.
7) The computer aided translation.

These are just examples of modern Arabic linguistic research. Most of the above mentioned areas, if not all of them, can use the Arabic KBS to reach various results. For example, in the area of phonotactics, a rule-interpreter program can analyze a large number of phonological sequences. A rule-interpreter program can process a given Arabic text, to determine its structure etc. This can be performed by comparison of the input (text in the Globe Base) with Arabic syntactic and morphological rules (in the knowledge base). After the structure of the text is completely analysed, an easy translation can be made into any foreign language.

To build a knowledge base system for Arabic is a very ambitious goal. Intensive research would be needed in order to define the knowledge base of the system, to create a method for converting it into a set of rules, and to develop intelligent rule interpreter programs for various applications. The knowledge base would need to be large enough to include the following :

1) Large set of Arabic syntactic and morphological statistical data.
2) Sizable corpus of examples representing Arabic phonology and phonotactics.
3) Vast number of Arabic lexical items with their syntactic functions and semantic features, that is, a rule-interpreter dictionary ;
4) Large corpus of analogous and anomalous data concerning Arabic morphology ;
5) Linguistic signs and their semantic reference ;
6) Other interesting linguistic theorems.

A knowledge base that contained all the above mentioned items would be enormous, indeed, and it could only be built gradually. For this reason, the development of a knowledge system for Arabic should first concentrate on developing the empirical part of the Arabic knowledge base.

Review of Previous Work

Several studies were performed in the past to determine the frequency of Arabic vocabulary, sounds, and various syntactic and morphological structures (31-42). Frequency is the number of times a given word, sound or structure appears in a text of a known volume. The studies of vocabulary frequency may be summarized as follows :

- In 1940, Mousa Bryel (31) conducted a study of the frequency of various words, collected from «Al-Ahram» and «Palestien» newspapers. The texts he chose were about 136,000 words mainly consisting of news articles.
- In 1950, E.M., Baily studied texts from Arabic Egyptian newspapers, about 200,000 words, to determine those most commonly used. [32].
- Dr. Fakhar Akel, in 1953, searched various books of elementary schools for the most common words. He counted a total of 188,000 words. [33].
- Landen Jacob, in 1959, counted 136,000 words from books of different subjects. [34].
- Dr. Mohammad Lutfi counted about 35,000 words in

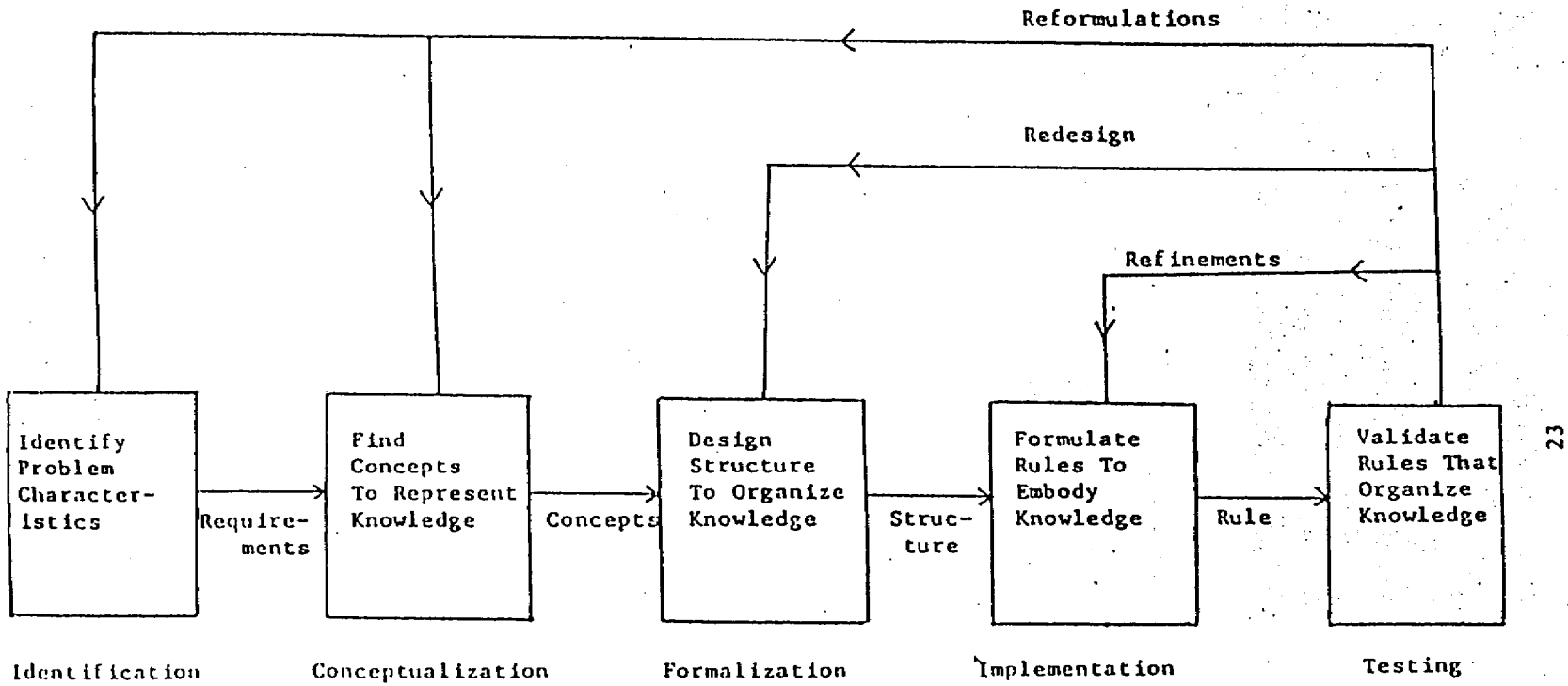

Fig. 1. Methodology for Building expert systems

as certainty C). This value indicates the degree of certainty by which the conclusion follows from the evidence. As an example, consider the following rule, from a medical knowledge base-system (17, 20) for diagnosis :

IF :(1) The site of the culture is blood, and
(2) The identity of the organism is not known with certainty, and
(3) Morphologically the organism is a rod, and
(4) The patient has been seriously burned.

THEN :There is weakly suggestive evidence (4) that the identity of the organism is pseudomonas.

The value of 4 (on a 0 to 1 scale) means the conclusion is weakly suggested from the evidence.

There are also other ways, to represent the certainty of given evidence. In some systems, the evidence may be combined by logical operators such as AND, OR, and NOT. Others may use plausible-inference rules [17, 19] indicated by the likelihood ratios (rule-strength values). These ratios are represented by either a plot or by two decimal numbers. One indicates how sufficient the evidence is for establishing the hypothesis, if the evidence is present. The second indicates how necessary the evidence is for the hypothesis. That is the degree to which the absence of the evidence will rule out the hypothesis.

From the above discussion the importance and necessity of an accurate statistical study for every KBS should be clear. This is because, most of the empirical rules are defined by an expert, usually they do not support a conclusion with a hundred percent certainty.

A KBS may contain several programs called rule-interpreter programs, or inference systems. These programs can scan through the rule-base to produce an internal representation that makes the expert-system an expert about the task's domain. The rule-interpreter program is thus a reasoning mechanism, that inputs the knowledge-base (i. e. interprets the rules), and applies it to solve the problems under investigation.

The rule-interpreter program can vary, as the rule base may be used for a variety of applications. For example, the rule base of a medical diagnosis system can be interpreted by :

- 0 A program for medical students' education.
- 0 A program for diagnosis of various diseases.
- 0 A program for medical research.

The nature of the rule-interpreter programs illustrate an important distinction between a knowledge-base system and the more conventional computer programs ; there is a clear separation of general knowledge (knowledge-base) from the methods for applying it to a particular problem (the rule-interpreter).

Various strategies (called control strategies) are used by interpreter programs to decide which rule of the knowledge to apply. References (21.22.23) describe some of these.

The Global Data Base (sometimes called the working memory) contains information about a current problem needed to be solved (e.g. manifestations noticed about a particular patient). The Global Data Base can be thought of as a file of needed facts concerning a problem. These facts are used as input data to the rule-interpreter program.

Knowledge-base Systems Characteristics and Advantages

As mentioned earlier, one major characteristic of a knowledge system is the separation between the rules forming the knowledge-base, and the rule interpreter program (16, 17). This partitioning, together with the further division of general knowlege into many separate rules, and the nature of the knowledge base itself, offer several advantages :

1) The rule-base represents the knowledge and judgement of many scientists and experts. As a result, the knowledge system can effectively substitute the knowledge of the experts for the unknown function. In effect, knowledge base expert systems offer the promise of putting experts at everyone's disposal.
2) Incremental development of the knowledge base over an extended time by letting the developers refine old rules and new ones.
3) Same general system can be used for a variety of applications, and the same knowledge can be put to use in different ways, by changing the rule interpreter.
4) Possibility of developing systems that are introspective and evolutionary (e. g. can check the consistency of their own rules and learn new ones).

Because of the above advantages, several impressive knowledge base systems were created, to solve difficult problems, in specialized areas. [Ref. 24-29]

Overview of Methodology for Building Expert Systems :

Generally expert systems emerge from experimentation, therefore, it evolves gradually. The major stages in the evolution of an expert system are shown in Fig. 1. During identification, the problem area and its scope is estimated and decided upon the goals or objectives of building expert system.

During conceptualization, the key concepts are established, relation and information-flow characteristics needed to describe the problem solving process in given domain. The subtasks, strategies, and constraints related to problem-solving activity are specified.

linguistic problems were involved. As a result, most Machine Translation (MT) research activities diminished world-wide. Since then, the tremendous advances in AI and computer hardware have brought a resurgence of interest in MT. There have been a growing number of useful though imperfect systems deployed in government and industry.

The systems developed in America and the extent of their development have been described in detail in numerous recent articles (11,12). Reference (11) sketches the status of well-known systems like SYSTRAN, LOGOS, METEO, WEIDNER, ALPS and SPANAM. Other systems outside America are also described. The state-of-the-art of MAT and MT is greatly affected by the status of research in general linguistics, on the one hand,and in applied computational linguistics on the other. This is true because the interpretation of natural languages requires simultaneous application of language-specific knowledge (word use, word order, phrase structure) and real-world knowledge about typical situations, events, roles and contexts.

Concerning Arabic, two MT systems from and to Arabic already exist : Systran and Weidner. However, they are rarely used and their performance is not entirely satisfactory. In part, this is due to shortcomings found in other language MT Systems but partially acute in Arabic ; lack of progress in Computational and theoretical linguistics. Another reason lies in the fact that too little is known regarding contrastive issues between Arabic and other languages. The production of an effective Arabic MT system depends mainly on the selection of proper linguistic and computational techniques, techniques up on which system-development should be based. This, in turn, requires a large-scale effort directed toward the formal description of Arabic. In addition, substantial effort must be concentrated on Arabic terminology, statistical distribution, and contrasts with other world languages. [12] Specific problems occasioned by Arabic should also be solved. Arabic morphology, for instance, represents a special problem that cannot be handled even by the most promising currently available techniques [13]. Topicalization, genitive constructions, and passivization are particular Arabic attributes that do not seem to exhibit important roles in other world languages. It is, therefore, obvious that the understanding of the language itself is the primary prerequisite for development of capable and effective computerized language control.

Pattern Recognition is the description and analysis of measurements taken from physical or mental objects. The general goal of pattern-recognition computer systems is to analyze images of a given scene and recognize the contents. A major application of this analysis is in document processing (Alphabetic and numerical characters recognition) [14]. Naturally, to perform this analysis easily the computer system needs a sizable set of measurements for language character recognition. Knowledge of the statistical and structural characteristic of the pattern (character) is therefore vital to achieve this goal.

For Arabic, studies especially oriented to serve Arabic pattern-recognition application are still in their infancy. Serious computational linguistic research is greatly needed to develop a practical Arabic character-recognition system. As will be noted later, an expert system that is especially designed for Arabic pattern-recognition will be of great assistance.

Knowledge Base System

Knowledge Base Systems may be defined as computer algorithms that exploit special knowledge to solve difficult problems in specialized fields [16]. A KBS consists of the following three major components ; A Knowledge Base (KB) and a Rule Interpreter Program and a Global Data Base :

The KB represents a «collection» of knowledge about a particular field, knowledge usually obtained from experts in that field, e.g. professors of medicine for diagnosis systems, geologists for mineral exploration Knowledge systems etc [17]. The Knowledge collected is represented in various forms. One popular approach is to use IF-THEN rules, (usually called KB rules) [17, 18, 19]. These rules say that if a certain kind of action can be taken. They are typically used to capture the kind of «semi-logical» responses to familiar patterns that characterize much of everyday human thinking. The rules for a particular task are usually written in a specialized language, which is then put in by a particular application program, as will be explained later. The rules have usually the following general form : [17]

IF : antecedent $_1$
0
0
antecedent $_n$

THEN :

Consequent $_1$
with certainty C_1
0
0
Consequent $_n$
with certainty C_n

The antecedents can be thought of as evidence of assumptions or patterns that can be matched against entries of the particular case to be investigated. The consequents are actions that can be performed (or conclusion that can be deduced), if all antecedents match.

An expert usually has many varying judgements. Evidence either supports or does not support a conclusion or hypothesis, some judgements are more certain than others. There are degrees of certitude. In this case, a numerical value is associated with each rule (referred to