A Knowledge Base System For Arabic: Retrospective and Prospective

Abdulaziz I. Al Sweel * Sami S. Al Wakeel **

* Ph.D. Linguistics, Washington University of Seattle, 1983.
Professor, English Department - King Saud University.

** Ph.D. Electrical Engineering, Stanford University, U.S.A. 1983.
Assistant Professor, Science College, King Saud University.

Abstract

Recent developments in Computer Science have made possible far-reaching and linguistically relevant research into and implementation of an Arabic language Knowledge Base System (KBS). Although no Arabic KBSs yet exist, they have shown a great potential to the task of processing the multifold systems within the Arabic natural language, its morphology and syntax. Arabic language machine translation, pattern-recognition and Artificial Intelligence (AI) applications are as yet in their infancy. The purpose of this paper is to review past work in the same and related fields of investigation into AI applications for Arabic language, suggest various interesting possible roads of development, with special attention given to a KBS for the Arabic language; enumerate pressing research problems and research objectives; and, provide possible applications. This paper concludes with a concrete and detailed research project design for an Arabic language KBS. The design is made so that its implementation can be realized in distinct research phases. Explanatory tables and references are included with the text.

Introduction

Since the first computers appeared in early 1940's many scientists saw the potential of using them for processing of human (natural) language. First, computer programs have been developed to manipulate texts to produce words lists, indexes...etc. Later on, it has proposed that computers might be used to translate one natural language to another (IEEE Proceedings, 1986). With time, more research subjects and applications were identified. Currently, the domain of computational treatment of natural language is vast enough and covers a wide collection of theories, techniques, and research results. Two sets of compative software (programs) however exist. The first, are those programs that process texts without looking to their linguistic meaning, structures, or relations. Examples of these include word processing programs, programs that produce text indexes archives etc. The second set of programs, often called knowledge based systems, expert system or advisory systems are expected to "understand" texts and generate outputs with as much fluency as a native speaker would (IEEE Proceedings, 1986).

Knowledge Base Systems are the result of the research progress made in the field of Artificial Intelligence (AI) and constitute one of the key AI disciplines. For this, We review AI in what follow:

AI is the computer science that focuses on "intelligence," and is applied to solving real-world problems. It offers viable solutions for dealing with complex situations that require interpretation of information, the use of heuristic rules or problems dealing with uncertainty. AI models emulate human expert behaviour in recalling past experiences needed to judge when an exceptional case is encountered. Besides Knowledge Base Systems, major AI disciplines include symbolic processing, natural language processing machine translation and speech and pattern recognition.

Symbolic processing is a technique used to represent real world objects and properties associated with them as symbols. (Neils, 1982) It allows the manipulation of these symbols in a way analogous to human reasoning. Symbols can be linked together, using super structures such as networks or graphics, to represent such relationships as hierarchy and dependency. As conventional computer languages are not suited for symbolic processing, special AI logic languages, such as "Lisp." "Planner" and "Prolog" were developed.

Natural language processing by computers has as its goal humans communication with computers in normal everyday languages such as English or Arabic rather than in conventional programming languages. Major applications of natural-language processing are in data-base interfaces. Users of data-bases are allowed to retrieve and manipulate stored data by using ordinary natural language commands (Gevarter, April 1983.... Burton, 1976).

Using computers to assist in translation from one natural language to another is an active research area of AI. The systems developed in America for machine translation

and the extent of their development have been described in detail in numerous recent articles (11,12). Well-known systems like SYSTRAN, LOGOS, METEO, WEIDNER, ALPS and SPANAM, as well as, other systems outside America are described in Reference (Slocum, 1984).

Concerning Arabic, the development of MT systems to Arabic are rare and their performance is not entirely satisfactory. In part, this is due to lack of progress in Computational and theoretical linguistics, and the lack of understanding of contrastive issues between Arabic and other languages. The production of an effective Arabic MT system, therefore, requires a large-scale effort directed toward the formal description of Arabic, Arabic terminology, statistical distribution and contrasts with other world languages. (Winfred, March 1985). Specific problems occasioned by Arabic should also be solved. Topicalization, genitive constructions, and passivization are particular Arabic attributes that do not seem to exhibit important roles in other world languages (Jonathan, March 1985). The understanding of Arabic itself is the primary prerequisite for development of capable and effective computerized Arabic 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) (K. S. Fu, 1982.... K.S. Fu, 1984). 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 (IEEE Computer Magazine, 1983). 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 a mineral exploration knowledge systems etc (Duda and Gasching, 1981: 238-280). The knowledge collected is represented in various forms. One popular approach is to use IF-THEN rules, (usually called KB rules) (Duda and Gasching, 1981: 238-280... Davis, 1979: 121-157). The rules have usually the following general form: (Duda and Gasching, 1981: 238-280).

IF: case $_1$
0
0
case $_n$
THEN:
Result $_1$

with Probability P_1
0
0
Result $_n$
with Probability P_n

The cases are the evidences or patterns that can be matched against entries of the particular case to be investigated. The results are conclusions that can be deduced, if all cases match.

As some conclusions are more certain than others, a numerical value is associated with each rule (referred to as Probability P). This value indicates the degree of certainly by which the conclusion follow from the evidence. As an example, consider the following rule, from a medical Knowledge Base System: given by references (Duda and Gasching, 1981: 238-280):-

IF : (1) The site of the culture is blood, and
(2) Morphologically the organism is a rod, and
(3) 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 result is weakly suggested from the case.

There are also other ways, to represent the certainty or probability of a given evidence. The evidence may also be combined by logical operators such as AND, OR, and NOT, may use plausible-inference rules, indicated by the likelihood ratios (rule-strength values) (Duda and Gasching, 1981: 238-280, Davis, 1979: 121-157). These represents how sufficient the evidence is for establishing the hypothesis , and indicate how necessary the evidence is for 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 or result with a hundred percent Probability.

A KBS may contain several programs called Rule Interpreter Programs, or infer-

ence systems. These programs are 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: A program for medical students' education, a program for diagnosis of various diseases of a program for medical research.

It is clear, therefore, that for a Knowledge-Base System exists a clear separation between the general knowledge (knowledge-base) and 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 (Shalom, 1984.... Feigenbanm, 1978) 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 (Duda and Gasching, 1981: 238-280).

Knowledge-base systems Characteristics and Advantages

As mentioned earlier, one major characteristic of a Knowledge Base System is the separation between the rules forming the Knowledge-Base, and the Rule Interpreter Program (Duda and Gasching, 1981: 238-280- Barstow, 1979: 7-119). This partitioning, together with the further division of general knowledge into many separate rules, and the nature of the Knowledge Base itself, offers 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 add 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, like medical diagnosis, computer design and circuit analysis (Shortliffe, 1976.... Clarcey, 1979: 175-183).

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 systems by many Arab linguists, and, on the other side, due to the great difficulties faced by a computer scientist with little linguistic background and expertise so far. It is, therefore, a necessity to have a common effort by experts in both subjects (linguistics 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, Computer Aided Design (CAI)...etc). These are among the many greatly needed applications of a KBS for the Arab world. The important goal here is to implement and locally developed 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: (Ramadan, Riyadh).

1) Research in Arabic phonology to analyze Arabic sounds, their laws, characteristic, outputs..etc.:
2) The study of Arabic syntax 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 advantage. 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 for statistical analysis and to determine the frequency of Arabic vocabulary, sounds, and various syntactic and morphological structures (Bryle, 1940... Mousa, 1972). Frequency is the number of times a given word, sound or structure appears in a text of a known volume.

Most of the previous studies, however, 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 an 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 2,000,000 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 stydy 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 and to develop the Knowledge Base of the Arabic KBS. For this, it concentrates on developing the statistical and empirical information required for the Knowledge Base and for the numerous applications of the system to the Arabic language. The objectives can, therefore, be summarized succinctly as follows:

1) The development of a comprehensive data base of Arabic vocabulary. The data is a collection of Arabic words and word use, stored together with controlled redundancy to serve more than one applications. The base will have a total volume of about two 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 frequency, relationship to each other, their phonemic arrangement, etc.
   — the facilitation 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 development of the Arabic Knowledge Base System. Generally, the Arabic KBS will emerge from experimentation. Therefore, it will evolve gradually. The major stages in the evolution of the KB system are shown in Fig 1. During identification, the problem area and its scope are estimated and the goals or objectives of building the Arabic KBS are decided.

   During conceptualization, the key concepts are established. Relation and information-flow characteristics needed to describe the application problem solving pro-

cess in given Arabic domain are selected. The subtasks, strategies, and constraints related to problem-solving activity are specified.

Formalization involves mapping the key concepts and relations into a formal representation suggested by some expert system tool or computer 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.

7) The use of the experience gained from the previous objectives to develop various rule-interpreter programs of the developed knowledge base for Arabic.
These programs will implement:
   — Potential applications of the proposed Knowledge Base System.
   — Actions required to absorb, adapt, maintain and create Arabic Knowledge-Base System applications.

Applications and Utilization

1) The results of the project constitute the necessary steps to develop a KBS for the Arabic language. This system, upon completion, will assist in developing very important applications, for Arabic in the fields of symbolic and natural language processing, machine translation and Arabic pattern recognition. Major Arabic application of these fields will include:

   — Development of Arabic data-base interfaces that allow users of data-base to retrieve and manipulate stored computer data using ordinary Arabic commands.

   — Development of computer-based translation systems that could produce and assist high quality, high speed translation of arbitrary Arabic texts.

   — Document processing of Arabic by computer systems. This allows the pattern recognition of Arabic alphabetic and numerical characters, and the vocalization of Arabic texts.

   — Development of Arabic speech synthesis system.

   — A question - Answering computer-based system.

   — Arabic computer design and Architecture.

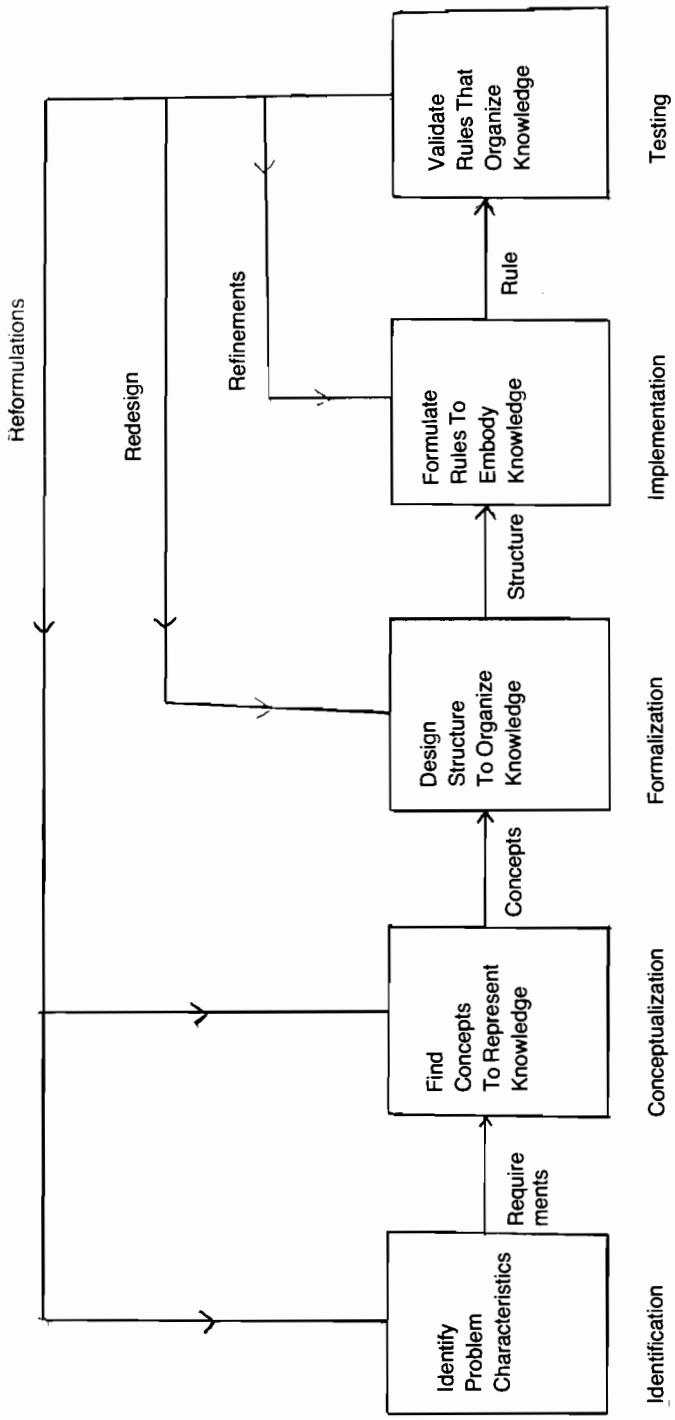

Fig.1. Methodology for Building expert systems

2) The applications to be formulated will stimulate the uses of the Arabic knowledge system for Arabic linguistic research. Numerous possible linguistic theorms and rules may be easily investigated by Arabic scientists. In addition, a solid basis will be made for tackling problematic areas in Arabic Grammar. These includes:

   — Studying the irregular and other peculiar forms in Arabic (studying broken plurals will be one example of this).

   — Studying of characteristics of marked in contrast to unmarked forms.

   — 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 lexial items actually exist).

      b) defining what lexial items are potentially possible within the language range.

      c) defining what forms are impossible to be introduced into the language.

   — Determining the most recurrent Arabic syntactic structure.

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

   — Development of pocket dictionaries compiled for quick reference, and wirting textbooks for Arabic primary education, concentrated on the most commonly used forms.

   — Development of children's literature, simplified science books, and Arabic Educational books for elementary schools students and for non-Arabs that include commonly used vocabularies and structures suitable for various educational level.

   — Evaluation of educational books regarding their use of rare vocabulary, or complex and uncommon syntactic or morphological structures.

   — Design of various language examinations linguistic tests in particular.

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. (IEEE Proceedings, 1986.... Slocum, 1984).

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:
   - o Word-class (e.g. verb, noun, articls, adverbial, ... etc.):
   - o Roots (triliteral, quadriliteral ...):
   - o Kind of sociolinguistic register (e.g. scientific article, poetry, ... etc):

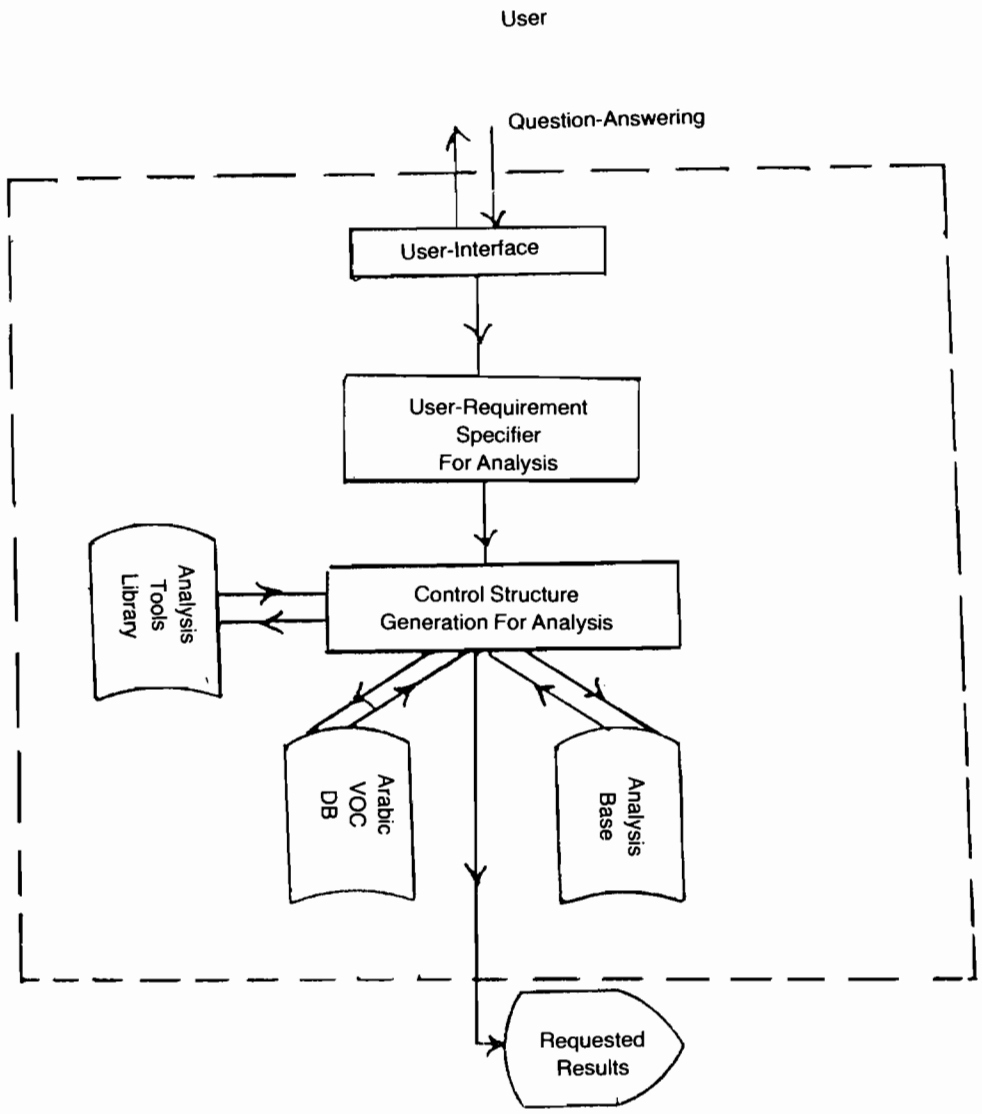

Figure 2 : Block-Structured Representation of SATA

- o Semantic features:
- o Other classifications:

2) Various morphological statistical data such as:
   - o Frequency of nouns, active participles, diminutive aspect, etc.:
   - o Frequency of (أل) article

- o Probability that a word with a given number of characters is a verb, is a noun, etc.
- o Percentage of triliteral verbs to all words and to all verbs.

3) Statistical information about Arabic sounds may include:
   - o Frequency of the various consonants and vowels (e.g. tha ث, wa و, on وو, oo.... etc.):
   - o Sequence of occurrence of various sounds:
   - o Relation to various morphological structures:
   - o Permissible sound clusters:
   - o Syllable structures:
   - o the relation between vowels and consonants.

4) Statistical information about syntatic structures:
   - o Marked Versus unmarked forms:
   - o Percentage of grammatical words.
   - o Frequency of various Arabic Syllables (e.g. fi, ala, In: ila etc).
   - o Determining syntactic exceptions and deducing the constraints preventing their generation:
   - o 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:
   - o Finding analogical rules for irregular morphological structures. (e.g. triliteral verbs, and their gerunds):
   - o 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 Kowledge 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) im-

plementation, 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 proceeding 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 hypothesis 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, and rural interpreter programs. This in turn allows achievement of various objectives of the Arabic Knowledge Base Systems.

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 advances 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.

Bibliography

Abdu, D. **Common Vocabulary of Arabic Language,** Riyadh University, Riyadh, 1979.

Akel Fakhr. **Basic Vocabulary for Elementary Education,** Damascus, Syria 1953

Baily, E.M. **A list of Modern Arabic Words,** Niel Mision Press, Cairo, 1950.

Barstow, D.R. "An Experiment in Knowledge-Based Automatic Programming" **Artificial Intelligence,** Vol. 12, August 1979, pp. 7-119.

Berwick, R. **Locality Principles and Acquisition of Syntactic knowledge,** Ph.D Thesis, MIT, Cambridge, MA, 1982.

Bryle, Mousa, **Dictionary of Arabic Newspapers,** Al-Quds, 1940.

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

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

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

Duda, R.O. and Gasching, J.G. "Knowledge-Based Expert System Come of Age." **Byte,** Sep. 1981, pp. 238-280.

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

Feigenbanm, E.A. "The Art of Artificial Intelligence Themes and Case Studies of Knowledge Engineering". **AFIPS Conference Proceeding,** Vol. 47, June 1978, pp. 227-240.

Feigenbaum, E.A., et al, "On Generality and Problem Solving: A Case Study Using the

DENDRAL Program," **Machine Intelligence,** No. 6, B. Meltzer and D. Michine (Editors), Elsevier, N.Y., 1971, po. 165-190.

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. "Developing a Natural Language Interfaces to Complex Data." **ACM Trans on Data Base System,** Vol. 3, No. 2, June 1978.

Sacerdoti, ED: Saglow, D; Hendrix, G.G., and Sacerdoti, E.D. "Natural Language Processing: The Field in Perspective," **Byte,** Sept 1981, pp. 304-352.

Ibn Manzoor. **Dictionary of Lisan Al-Arab,** Cairo, Ameriya Press.

IEEE Computer Magazine, Special issue on knowledge representation, Institute of Elecrical and Electronic Engineers, October 1983.

IEEE Proceedings Special issue on natural language processing: Institute of Electrical and Electronic Engineers (IEEE), Vol. 74, No. 7, July, 1986, pp. 897-1056.

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

Kay, M. "Algorithm Schemata and Data Structures in Syntactic Processing." **Processing of Symposium on Text Processing,** 1980.

Al-Khuli, M.A. **Common Structures in the Arabic Language,** Dar Al-Olum Press, Riyadh, 1978.

Al-Khuli, DR. M.A. "Analytic Review Study for the Arabic Vocabulary," Drasat Journal, Second Issue, King Saud University, Riyadh, 1978, pp. 27-66.

K.S. Fu **Applications of Pattern Recognition,** CRC Press, Boca Raton, Florida, 1982.

K.S. Fu "Pattern Recognition and Computer Vision." **Computer Magazine,** October 1984.

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

Lotfi, M., K. **Changes Needed in Egyptian Readers to Increase their Value,** Ph.D. Dissertation, Chicago University, Chicago, 1948.

Marcus, M.P **A Thoery of Syntactic Recognition For Natural Languages,** MIT Press, Cambridge, MA, 1980.

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

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

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

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

Neils, J. **Symbolic Computation: Principles of Artificial Intelligence,** Springer-Verlag, Berlin, New York, 1982.

Newell, A. and H. Simon, **Human Problem Solving,** Prentice-Hall, Englowed, 1972.

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

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

Shalom M. Weiss & Casimir A. Kulikawski. **A Practical Guide to Designing Expert Systems."** Rowman & Allanheld, Publishers, 1984.

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

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

Stallman, R.M. and 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.

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

Winfred P. Lehmann "Computer Aided Translation, State of the Art in the U.S.A. and Canada." **Proceedings of SANCST Saudi International Workshop on MAT,** Riyadh, March 1985.