INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

LEXICAL-SEMANTIC RELATIONS IN ARABIC TEXT

BY

MOHAMMED ALSHENIFEY

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Computer Science
in the Graduate College of the
Illinois Institute of Technology

Approved Martha Evens
Adviser

Chicago, Illinois
May 1998

ACKNOWLEDGMENT

First of all, I would like to thank my God, Alla the Almighty, Who gave me the strength to finish this thesis. My sincerest thanks go to my advisor, Professor Martha Evens for her support and guidance during the time I have been carrying this research. I cannot remember how many of my drafts she went through. Without the help of this remarkable scientist, this work might never been completed.

I am also thankful to the member of my committee, Professor Teymour, Professor Burnstein, Professor Elrad, and Professor Hood. I am thankful to my sister Hya and her husband Abraham and all the rest of the family. I would like to thank all my friends and my colleagues at IIT, who helped me, carry out my experiment, especially Saleem Abuleil, who has carefully read all my data.

I am especially grateful to Michael Glass who has taught me about many aspects of Computers Science for the last four years.

M.A.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 The Contribution of This Research

As natural language processing (NLP) systems grow more sophisticated, they need larger and more detailed lexicons. This seems to become all the more true as NLP systems get bigger and move out of the laboratory, into applications that require more information in each lexical entry. This research involves one of the most important areas of the attempt to make computer systems available in Arabic. My focus is on an investigation of lexical-semantic relations that can be used for not only building a thesaurus for information retrieval but also for parsing and text generation applications. A lexical-semantic relation is a rule that links words together.

Natural language processing (NLP) systems need a large amount of explicit information for each vocabulary item known to the system. Different applications, in turn, require different information in each lexical entry. As far as we know there is no Arabic Thesaurus for computers available at this time, so I am developing a Thesaurus for Arabic that can be used in information retrieval, text generation and other applications in natural language.

To find the relationships between words, I looked in many Arabic dictionaries and discussed the problem with some Arabic linguists and they suggested that Synonyms, near-Synonyms, Antonyms, and indications of Taxonomy, Part-Whole, and Grading relations can be found in Arabic language dictionaries. They also suggested that many other relations exist in Arabic Lexicography and can be used for developing a thesaurus.

The goal of this research is to develop a list of lexical-semantic relations for Arabic, and to use them in building a thesaurus for the vocabulary extracted from 258 Arabic abstracts of papers in computer science in machine readable form. This thesaurus is designed to support several natural language applications including information retrieval. For many years, researchers have been trying to improve retrieval performance. One direction of these studies is to test the usefulness of a relational thesaurus in improving retrieval performance in Arabic.

Several experiments have been done in the past using a relational thesaurus to improve the effectiveness of information retrieval systems containing English documents. Fox [1981] performed a series of experiments on the SMART system and found that using a relational thesaurus improved the performance of the system. Evens et al. [1985] and Wang et al. [1985] demonstrated the efficacy of the thesaurus as well. Abu-Salem [1992] used a thesaurus built by Al-Khrisat [1992a,b] to show that relations can also be extremely effective in Arabic information retrieval. The relational thesaurus is proposed as a new methodology to enhance the effectiveness of Arabic-IRS and to solve the problems created by the suffix, infix, and/or prefix chopping method. The advantage of this methodology is that the construction of the relational thesaurus is independent of the document collection and the user queries.

Relational models have been used widely in computer science, anthropology, linguistics, and psychology [Evens et al., 1983]. A wide study of lexical semantic relations was launched in the former U.S.S.R in connection with the development of the Explanatory Combinatory Dictionary (ECD) [Apresyan et al., 1970]. Lexical semantic relations have been studied by considering what knowledge would be needed in a lexicon

to support question-answering systems [Evens and Smith, 1978]. The resulting set of relations motivated an empirical study of the usefulness of lexical-semantic relationships in information retrieval [Evens et al. 1985; Fox, 1981; Nutter et al., 1990]. In a combined project, an Illinois Institute of Technology group and a Virginia Polytechnic Institute group have built a large relational lexicon from machine-readable dictionaries for use with information retrieval systems, using the Collins Dictionary of the English Language and Webster's Seventh Collegiate Dictionary [Fox et al., 1988, Nutter et al., 1990].

Evens et al. [1983] constructed a relational thesaurus, which was used to automatically enhance queries with related index terms in a batch information retrieval system. The improvement in performance was especially noticeable with a set of ill-formed queries that contained few index terms initially. The best results were obtained in their batch environment, when all the relations except the antonymy relation were combined together.

The classical relations of synonymy and antonymy are marked explicitly in many English dictionaries, such as Webster's Seventh New Collegiate Dictionary. The entry for 'feminine', for example, includes 'Syn. female' and 'Ant. masculine'. Many noun definitions involve taxonomy, the species-genus relation, e.g., a ferret is defined as 'An animal (Mustela fero) of the weasel family.' Other definitions involve the part-whole relation, the foot is defined as 'the tournament part of the leg.'

Lexical-semantic relations provide a way to encode information in a form that is compact and easy to retrieve. Suppose, for example, that a question-answering system is processing a story that begins:

> Nancy was picking heather when she heard a baa. She turned and saw a little black lamb in the shade of a large rock. It rose shakily to its feet and started to follow her.

To answer the question, 'What did the sheep do?' the system must use the information that a lamb is a young sheep. This fact could be stored as a predicate calculus formula, for example, but the same kind of information must be repeated in lexical entries for kitten, calf, and puppy. This information can be expressed more compactly, instead, by the lexical relation, CHILD. The lexical entry for lamb need only contain "CHILD sheep" while the lexical entry for kitten contains "CHILD cat."

1.2 Organization of the Lexicon

There are also differing views about how the lexicon ought to be structured. For one thing, should syntactic, semantic and pragmatic information be separated, and if so, where do we draw the line? Intended use is only one factor in the design of the lexicon. Although much successful work has been done using a separate dictionary and encyclopedia it is difficult to decide what goes where. One of the basic operations is finding words related to a given word.

We prefer to view the lexicon as a large directed graph; the nodes correspond to lexical items; the arcs are labeled with relation names. Each relation has an inverse [Ahlswede and Evens, 1988a].

1.3 Historical Background of the Arabic Languages

The Arabic language is a Semitic language with special characteristics that make it a difficult language to deal with on a computer. The Arabic alphabet consists of twenty-eight characters. Texts are written in three different forms, non-diacritized script without vowels, fully-diacritized script with vowels and partially-diacritized. In non-diacritized text, like the text we deal with, it is the responsibility of the reader to determine the intended meaning of the word in its context. Arabic script is written from right to left.

Arabic is one of the official languages of the world. There are twenty-two countries where Arabic is spoken. It is the language of the Qur'an and is the religious language of all Muslim peoples. Arab linguists have been producing studies of the language since 656 A.D [Saliba and Al-Dannan in 1989]. Therefore, the Arabic language was well described and investigated, but some works on the Arabic language are interpretations of the previous ones, whose style is difficult for students of today.

Mohamed Esmail Sieny [1985, p. 195] working at the King Saud University teaching the Arabic language wrote that:

> Arabs were known to be leaders in the field of lexicography for many centuries. Probably before the Middle Ages, they were the most prolific producers of lexicographic works and dictionaries. If we count the ways in which they presented their works we will find that there are different types of dictionaries produced by Arabs, basically three we may say. The first was the general type arranged according to words, the second type was the thesaurus, and the third type was terminology dictionaries. On the thesaurus side, they started with 'special topic' dictionaries, such as dictionaries on horses, dictionaries on man, on insects, on arms etc., and many of those were common in the tenth century A.D. More comprehensive thesauri were compiled in the twelfth century. For example, we

> have the best known thesaurus Al-Mukhassas by Iben Sidah. It was compiled in seventeen volumes. The author was an Arab born in Andalusia, and he was a blind man. Other ones include Figah Al-Lugha by Al-Tha'alibi, adab Al-kaatib by Ibn Qutaybah, etc.

Some other languages use the Arabic alphabet, like Urdu, the language of Pakistan and some parts of India, and Farsi, the language of Iran. The extensive use of Arabic script dates back to the emergence of the Islamic faith.

1.4 Significance of this Research

We have read much previous and current research on lexicon-thesaurus design and construction peculiar to other languages. We wondered whether relation based approaches would work for the Arabic language. Various theoretical and methodological opinions have evolved over the research period, by no means always in a straight line. We are interested in what information we need in the lexicon-thesaurus, concentrating on the list of lexical-semantic relations. Our interest is in morphological, syntactic and semantic feature-like properties of words attached to nodes in a lexical network. Lexical semantic relations have been shown to be valuable in language processing in languages as different as English, Russian, Papago, and Chinese [Werner and Schoepfle, 1987]. Earlier experiments using a relational thesaurus in Arabic information retrieval by Abu-Salem [1992] suggest that this approach can be of use. I propose to provide enough relational data so that the Arabic Language Processing Group can experiment with the use of relations in parsing and text generation as well as information retrieval.

1.5 Review by Chapter

Chapter II is a review of the literature. Chapter III describes my synonymy experiment. Chapter IV gives definitions of taxonomy, antonymy, grading, attribute, collocational relations, paradigmatic relations and part-whole. Chapter V describes the data and the database. Chapter VI gives a description of the thesaurus. Chapter VII discusses the morphology and some systematic relations expressed in Arabic by standard morphology. Chapter VIII includes a summary and a discussion of possible future research.

CHAPTER II

REVIEW OF THE LITERATURE

This chapter discusses some related research in the area of lexical-semantic relations. Researchers who contributed to more than one area may appear in more than one section as seems appropriate. The work described in this thesis is inspired and guided by much previous work with lexical-semantic relations.

2.1 Applications of Relational Models of the Lexicon

Applications fall into two fundamental categories: computer implementation of relational models constructed by anthropologists, linguists, and psychologists in order to investigate the implications of their theories or simply to store and manipulate data effectively; and lexicons and knowledge bases built by computer scientists, borrowing ideas from everywhere, for information retrieval, natural language interfaces, or other natural language processing projects.

Apresyan, et al. [1970] designed a dictionary containing explicit information about all entries. They named it the Explanatory-Combinatorial Dictionary (ECD). Their intention was to provide the non-native speaker of the Russian language with enough details about vocabulary to construct correct idiomatic Russian texts. A major feature of the ECD was the Lexical Functions (or lexical-semantic relations) that describe the relationship of the entry word to other words in its semantic field. They identified 47 lexical functions (later expanded by Mel'cuk and Zholkovsky [1988] to 53). The lexical functions include such traditional semantic relations as synonymy, antonymy and

taxonomy as well as more specifically lexical relations like So which relates verbal nouns (*motion*) to verbs (*move*) or Ao which relates denominal adjectives (*national*) to nouns (*nation*).

Casagrande and Hale [1967] studied 800 definitions in Pima and Papago and extracted thirteen relations. This set of relations has been used in an information retrieval experiment in Chinese [Wan, 1995]. They are listed in Figure 1 in order of frequency in their corpus of definitions.

Williams [1966] organized a large semantic network with relations KIND, PART, and NAME to describe Japanese food, plant, season, and weather categories.

Evens et al. [1983] studied lexical-semantic relations in anthropology, linguistics, computer science and psychology. Work in these fields deals with lexical-semantic relations as a model of semantic organization. Evens [1979, 1981] explored the use of relations to make inferences in answering questions: e.g., given the information "Ted has a puppy. His name is Happy," we cannot answer the question "the pet is a: dog boy toy?" without the information that a puppy is a young dog (puppy CHILD dog) and that a dog may be a pet (dog TAX pet). Evens has developed an extensive list of relations for question answering [Evens and Smith, 1978], and text generation [Ahlswede and Evens, 1988a, b], and information retrieval [Nutter et al., 1990] and has investigated the relations implicit in dictionary definitions [Evens et al., 1987].

Oswald Werner's work in ethnography began with the Anatomical Atlas of the Navajo, in wide use by the U.S. Public Health Service [1969/1981]. It is organized in terms of the part-whole relation and the taxonomy relation. He has decided that the ideal set of primitive relations consists of just three: Modification, Taxonomy, and Queuing.

1. Attribution: X has defining attribute(s) Y

   burrowing owl: 'but they are small; and they act like mice and they live in holes.'

2. Contingency: X is normally or necessarily associated with Y.
   to lightning: 'when it rains, it lightnings.'

3. Function: X is the means of effecting Y
   shovel: 'with which earth is scooped up.'

4. Spatial: X has a characteristic spatial orientation to Y
   nose: 'it stands below our eyes.'

5. Operational: X is a characteristic goal or recipient of Y
   bread: 'which we eat.'

6. Comparison: X is like or unlike Y
   wolf: 'they are rather like coyotes, but they are big.'

7. Exemplification: X has typical co-occurrent Y
   sweet: 'as sugar.'

8. Class inclusion: X belongs to the class Y
   crane: 'a bird.'

9. Synonymy: X is the same as Y
   amusing: 'funny.'

10. Antonymy: X is the negation or opposite of Y
    low: 'not high.'

11. Provenience: X has source Y
    milk: 'we get it from a cow.'

12. Grading: X is part of a series containing Y
    Monday: 'the one following Sunday.'

13. Circularity: X is X
    to teach: "if someone teaches us we call it 'to teach.'"

Figure 1. Relations Identified by Casagrande and Hale [1967]

While Werner began with relational models of the lexicon and later applied some of the same strategies to texts, the Cornell anthropologist Joseph Grimes [1996] has moved in the opposite direction. He is probably best known for his work on narrative structures in different cultures, but he is now working at the level of the individual word.

The problem of designing and creating a lexicon containing all the syntactic and semantic information to support natural language processing over a broad range of subject matter is a center of concern for information retrieval.

Ahlswede's MS thesis [1981] made a systematic effort to identify relations in W7 adjective definitions, on the hypothesis that defining formulas expressed relations. In his Ph.D. Thesis [Ahlswede, 1988], he analyzed W7 definitions to build a lexicon using relational primitives. He found about two hundred lexical-semantic relations ("generous JCHAR ample," "nervous JCAUSE2 irritated").

In [1992b] Al-Khrisat designed an Arabic lexicon-thesaurus using lexical-semantic relations for the computer science sublanguage. He identified some of the morphological, syntactic, and semantic relations that can be used for language understanding and text generation. The Arabic lexicon was based on the list of keywords identified by Abu-Salem in 120 abstracts. He used the set of lexical-semantic relations developed by Ahlswede and Evens [1988a] and created a thesaurus containing word-relation-word triples.

Hammouri [1994] built an Arabic Lexical Database (ALDB) from 120 Arabic abstracts of papers in computer science, to support parsing, information retrieval, and text generation. The Arabic Lexical Database contains words and phrases in the computer

sublanguage. Each contains information about the entry part of speech, and morphological information.

Al-Shalabi [1996] designed and implemented an Arabic morphological system to support natural language processing applications. This system detects the root and the pattern of Arabic words with verbal roots.

Alsamara [1996] carried out further lexicographic research at IIT. He built an Arabic lexical database for the computer sublanguage for 242 abstracts from the Proceedings of the Saudi Arabian National Conferences. Then he converted the corpus of 242 abstracts into a machine readable form in an Arabic Windows environment. He created a relational database to store lexical entries for nouns, verbs, particles, and adjectives in separate tables. The research described in this dissertation makes use of the fundamental effort by Alsamara.

2.2 The Nature of Relations

There are many questions about the nature of lexical semantic relations themselves. The possible range of answers depends on whether the relations in question are relations between concepts (purely semantic) or relations between words (purely lexical) or somewhere in between.

Anthropologists, psychologists, and linguists have all been concerned about the fundamental nature of the tools that they employ in model building. Among anthropologists Werner, in particular, has been concerned to establish the universal nature of his system of relations, which consists of Modification, Taxonomy, and Queuing (MTQ) and certain Boolean relations from the prepositional calculus. He has used this

system in the analysis of texts in languages as far apart as Navajo, English, Hungarian, and Rapa Nui (Easter Island) and his students have used these relations to study texts from a number of African languages as well [Werner and Schoepfle, 1987].

Riegel [1970] was one of the first psychologists to use relational models to describe the organization of memory. He was revolutionary in another way as well; he looked at language and memory not in children or college students but in mature adults. He also studied the dissolution of language in the aging and in aphasics. Riegel divided relations into two fundamental categories depending on their "nature." Logical relations are derived by abstraction from the words themselves. Examples of logical relations are superordination (as in table-furniture) and coordination (as in table-chair). Infralogical relations or physical relations are based on the denoted objects, events, or qualities, and are a product of abstracting physical features from items. Examples of Riegel's infralogical relations are: parts (as in table-leg), locations (as in zebra-Africa), and substance (as in table-wood). Mel'cuk seems to make the same distinctions, at least tacitly, with a distinct preference for the logical as opposed to the infralogical relations. When asked why he does not include the part-whole relation among his collection of lexical functions, he explained [personal communication to M. Evens] that part-whole is "too semantic" and "too vague."

Becker [1975] has made a strong case for the phrasal lexicon, with the argument that when we produce language we seem to use units much larger than single words, sometimes even whole sentences. Professional lexicographers seem to agree; more than sixteen percent of the entries in Webster's Seventh New Collegiate Dictionary are headed by multi-word expressions. People who work on machine translation are also already

convinced that they need to store many idiomatic phrases. Admission of multi-word lexical items to a relational lexicon is likely to require a richer stock of relations to describe the organizational structure of the lexicon.

Susanna Cumming [1986 a, b; Cumming and Albano, 1986] contributed to the field of lexicon design through the Master Lexicon of the JANUS Project. This lexicon is intended to handle all the lexical needs for the grammars associated with the JANUS system: RUS, an ATN parsing grammar, and Nigel, a systemic generation grammar. The Master Lexicon contains morphological information, syntactic information, semantic information, and phrasal entries as well as word entries.

Ahlswede [1988] built a lexical database containing syntactic and semantic data from information in Webster's Seventh Collegiate Dictionary (W7). He investigated the role of relations other than taxonomy in the dictionary, and he studied the relational links between the entries for the words. Evens has been studying lexical-semantic relations for many years [1983]. One focus of her work is automatic thesaurus construction using this analysis of W7's definitions with Ahlswede. She argues that it is impossible to separate lexical and encyclopedic knowledge due to the fact that NLP systems need both knowledge about words and knowledge about the world [1981].

Ahlswede and Evens [1988b] describe the design of a relational lexicon for the use of the Michael Reese Hospital Stroke Consultant. They structured their lexicon as a semantic network with lexical-semantic relations. The lexicon contains semantic, syntactic, and pragmatic information to support parsing and text generation. Their relational lexicon consists of a set of nodes, one for each entry in the vocabulary. Some of

the lexical-semantic relations that they present are synonymy, taxonomy (the "is-a-kind-of" relation), and the part-whole relation. Here are some examples of their work:

artery	TAX	blood-vessel
aphasia	DYSFUN	speech
cerebellum	PART	brain
temporal-lobe	ABOVE	third-nerve
frontal-lobe	BEFORE	parietal-lobe
reactions	PLURAL	reaction
weakness	STATE	weak

Pin-Ngern [1990] built a Lexical Database for English that contains syntactic and semantic information for 50,000 words separated by parts of speech. There is information about word classification, selection restrictions, lexical semantic relationships, and syntax. She used the Collins English Dictionary (CED) as the major source. This lexicon was designed to support the needs of most NLP applications, including text generation applications, information retrieval and natural language understanding. The lexical database has been used in several experiments in information retrieval [Ahlswede and Evens, 1988 a, b; Fox et al., 1988; Nutter et al., 1990]. This lexicon is stored in an Oracle Relational Database Management System.

Ingria [1987] made an extensive study of verb complements and a compendium of verb feature information in natural language processing systems.

Boguraev [Boguraev et al., 1987, Boguraev and Briscoe, 1987, Boguraev, 1987] built the British National Lexicon on the basis of automatic analysis of the Longman Dictionary of Contemporary English at the University of Cambridge. Boguraev then moved to IBM, and he is now at Apple, but still involved in problems of Lexicography.

2.3 Explanatory Combinatorial Dictionary

An Explanatory Combinatorial Dictionary (ECD) is an essential component of any full-fledged linguistic description within the Meaning-Text Model (MTM). This model describes a natural language as a kind of logical device that associates with any given meaning M the set of all the texts in this language that are expressions of M (and which are consequently synonymous with one another), and with any text T, the set of all the meanings that are expressed by T (and that are, so to speak, homonymous with one another) [Zholkovskii and Mel'cuk, 1967].

Establishing correspondences between meaning and texts is conceived of as a multi-stage process: translating a given meaning, that is, a semantic representation from one level to another, until one of the corresponding texts is reached (or vice versa: 'translation' of a given text, that is a phonetic representation, from one level to another, until one of the corresponding meanings is reached).

Until recently, the predominant type of monolingual dictionary was the comprehensive dictionary, to which the user would refer on encountering an unfamiliar word or phrase in a text. In other words, such dictionaries were oriented toward making texts comprehensible (i.e., providing for the transition from a text to the meaning expressed by it). Using the well known opposition between passive grammar (= text

understanding) and active grammar (= text production), the ECD tries to provide help in moving from meaning to text.

The ECD [Zholkovsky and Mel'cuk, 1988] allows for the representation of three basic type of relations between words. The first type is semantic (paradigmatic) relationships between words: e.g., synonymy, semantic proximity, etc.

The second type is syntactic (syntagmatic) relationships between the entry word, which is semantically a predicate, and other words or phrases, which may be syntactically dependent on it in a sentence and which are the expression of its semantic actants. These sentence elements are used to fill in the slots of the active syntactic valence of the entry word and are called its actants. The active syntactic valence is specified by means of a table called a government pattern.

The pattern includes three parts. For each semantic actant of the entry word, the corresponding syntactic actant; the form that each syntactic actant takes on the surface; and which of the syntactic actants are incompatible (or, conversely, are inseparable, i.e., invariably used together), and under what conditions. The third part of the pattern includes the lexical (both paradigmatic and syntagmatic) relationships between the entry word and those other words that can either replace it in a text (under specific circumstances), or be joined to it in more or less fixed word combinations (also known as phraseological combinations).

For several years, Nicoletta Calzolari and her team at the University of Pisa have been working on the development of a large lexical database designed to be a repository of the entire vocabulary of Italian almost as detailed as the ECD. They have collected approximately 106,000 lemmas or root forms, more than one million word forms, and

about 186,000 definitions. Their database contains morphological syntactic, semantic, and thesaurus information [1988]. The semantic relations included in their database are shown in Figure 2.

hierarchical relations; synonymy relations; derivational relations; other taxonomies not organized on the IS-A relation; co-occurrence or collocation relations; terminological sublexicons; case-type or argument relations; lexical fields;

Figure 2. Components of Calzolari's Lexical Date Base

2.4 Research in Arabic Language Processing Elsewhere

Natural Language Processing in the Arabic language still needs a lot of work if we compare it with European languages. The first Arabic institution had computers in 1962. This was the National Planning Institute in Egypt. Most work there focused on Arabization, printing Arabic characters like names and addresses [Ali, 1988a, 1988b]. The first application dealing with Arabic computational linguistics was developed by Said Haidar in 1973. He built a computer system to deal with Arabic script.

In 1985 the Al-Lamieh company built a morphological system to deal with Arabic words. This system can handle vowelized or unvowelized Arabic words. This step was very important for Arabization.

Hegazi and Elsharkawi [1986] described a computer aided morphological system for vowelized Arabic words. The system was used to support a lexical analyzer for

diacriticized Arabic text and used to derive the root of the word based on the morphological and phonetic rules of the Arabic language.

El-Dessouki et al. [1988] built an expert system for Arabic sentences by using Prolog on the IBM-PC. They implemented a syntactic analyzer. This analyzer can extract the triliteral root and generate the morphological patterns for each root and its meaning.

El-Sadany and Hashish [1989] described a morphological system that was able to find the root and the morphological stricture of Arabic words that have triliteral roots; they tested their system on vowelized words.

Khayat and Al-Muhtaseb [1988] presented a method of knowledge representation for natural language. This method divided knowledge into subject and action. Subjects were represented in a manner similar to semantic networks; relations among subjects were represented by arcs. The system was used in understanding Arabic in a subject domain involving human beings, plants, and the interactions between them.

Gheith and Aboul-Ela [1989] presented a Computer Based Arabic Syntax Analyzer based on the concept of separation between processing algorithms and linguistic information. The aim of their research was to develop a system that can understand Arabic and to complete the modeling of Arabic grammar.

Wahba et al. [1990] designed a system used to provide a phonetic transcription for an experimental Arabic text-to-speech system. The system could also be used to find word base forms for an Arabic speech recognition system. Farghaly [1989] developed a natural language understanding system for Arabic. This system (NLUSA) has two main components: one is a query language that works on semantic representations of text; the other is a subsystem that takes Arabic input and produces a semantic representation

passing through a level of syntactic analysis. Hanash [1989] proposed to build an Arabic dictionary for Arabic verbs, since Arabic verb morphology is so complex. He used the Almoheet Dictionary by Farozbady, extracted all the verbs, then sorted them alphabetically and studied the characteristics of the verbs in addition to their selection restrictions.

2.5 Research in Arabic Language Processing at IIT

Work on Arabic Language Processing at Illinois Institute of Technology is led by Martha Evens. She established a Laboratory for Research in Arabic Language Processing several years ago. During the last seven years eleven students have completed Ph.D. dissertations in Arabic Language Processing. Work in Information Retrieval was begun by Al-Kharashi [1991]. He developed a system called MICRO-AIRS and used it to experiment with alternative choices of index terms, words, stems and roots. Jambi [1991] designed a system for recognizing Arabic characters. Al-Khrisat [1992a, 1992b] designed an Arabic lexicon-thesaurus using lexical-semantic relations to support information retrieval. In 1992 Abu-Salem developed a microcomputer-based bibliographic information retrieval system for Arabic documents (Arabic IRS), that interprets queries and retrieves relevant abstracts using the thesaurus developed by Al-Khrisat. Hammouri [1994] built an Arabic lexical database to support natural language processing. Abu-Arafah [1995] developed a partial grammar for the Arabic language suitable for machine parsing and automatic text generation. In [1995] Hmeidi designed and implemented an automatic word and phrase indexing system for information retrieval. Arif [1995] and

Armouti [1995] designed and implemented Al-Risalah, a pure object-oriented programming language with an Arabic language interface.

Al-Shalabi [1996] designed and implemented an Arabic morphological program to support language processing applications. Alsamara [1996] built a much larger, more detailed Arabic lexical database for the computer sublanguage to support information retrieval, text generation, and parsing. I have used Alsamara's lexicon to develop a list of lexical-semantic relations for Arabic, and to use them in building a thesaurus for the vocabulary extracted from Arabic abstracts of papers in computer science.

CHAPTER III

SYNONYMY EXPERIMENT

Apresyan et al. [1970, p.5] define synonymy as "two words should be semantically substitutable for each other, and the meaning of one should be expressible through the other in any context." Synonymy is said to be reflexive, symmetric, and transitive [Evens et al., 1983].

The synonymy relation is very important in thesauri and dictionaries. It has also played a central role in linguistic theory. It does not seem to be as central in folk definitions as taxonomy (Casagrande and Hale, 1967).

Many American linguists have argued that synonymy in the sense defined by Apresyan et al. does not exist. Instead, they argue that, if the sounds and the shapes of the words are different, there is no synonymy. Bloomfield was the most influential American linguist before Chomsky and his book is still required reading in most American Linguistics Departments. He states that two different words never have exactly the same meaning. "Each one of a set of forms like *quick, fast, swift, rapid, speed*, differs from all the others in some constant and conventional feature of meaning." [1933/1965, p.145]. Goodman in his paper on Likeness of Meaning [1966] insists that there are no two substitutable words without semantic differences.

I went to ask our Illinois Institute of Technology expert in Philosophy of Language, John W. Snapper, Professor of Philosophy, what he thinks about this issue. Snapper believes that "there is synonymy if X can be substituted for Y in context C (sub

X/Y in C) without changing the significance of C, but in this case synonymy is related to context C." He states that no words are substitutable in all contexts.

Khalil Ahmad Khalifeh, a linguist from King Saud University told me that he believes that synonymy exists in Arabic Language. He cited their standard text "Science of Semantics" by Ahmed Omar [1982].

Lyons [1968] developed the idea of mutual substitutability into an elaborate linguistic methodology. It requires five steps (adapted from Lyons. p.150):

- Take two candidates, both differing in one element.
- Make sure the contexts of both are identical.
- Get informant judgment of sameness of meaning of the items by which the target sentences differ.
- Place these two items in a variety of sentences.
- Get informant judgments on sameness of meaning in all sentences.

I decided to use this methodology to find out for myself whether synonymy exists in our corpus of abstracts.

I took seven abstracts from the Computer Science corpus and identified several synonyms. These seven abstracts were chosen because they were successfully parsed by Abu-Arafah [1995]. I also wrote a program to substitute one for another in the text with the goal of getting a better understanding of the synonymy relation.

I displayed the seven abstracts with the synonyms substituted for the original words at one of our regular Wednesday meetings of the Arabic Language Processing Laboratory. I gave each member of the group a form to fill out with his judgments about the substitution. For each substitution I asked people to state whether the resulting

sentence made perfect sense (100%) or only partial sense (50%) or made no sense at all (0%).

To our surprise everybody agreed that the sentences with substitutions were perfectly good except for one particular substitution of [صممنا persist] for [اصررنا insist] in thirty abstracts.

Again we had perfect agreement that the result of this substitution was terrible. Let me emphasize that all these abstracts belong to the computer sublanguage.

If we had made this substitution in psychology texts we might have obtained different results. The definition of synonymy given by Apresyan is not so outrageous; as eventually some thirty substitutions were judged to be workable by the group. Note that the translations of the abstracts also come from the Proceedings.

Examples of the synonyms in the abstracts:

SYN [Q(u)ran قران] = [m(u)sh(a)f مصحف Holy Book]
SYN [alm(9)lomat المعلومات information] = [albynat البيانات data]
SYN [alhdeth الحديثة recent] = [aljdedh الجديدة new]
SYN [mazam معظم most] = [glabeh غالبية majority]
SYN [almstamlet المستعملة used] = [almsthtmh المستخدمة decided]
SYN [sammna صممنا persist] = [asrana اصررنا insist]
SYN [almgdarht المقدرة ability] = [alstt(a)(9)h الاستطاعة capability]
SYN [tgleale تقليل decrease] = [tanges تنقيص reduce]
SYN [altklfh التكلفة cost] = [algemh القيمة value]
SYN [yast(9)reiz يستعرض show] = [ygadem يقدم introduce]
SYN [almshbh المشبهة like] = [almm(a)thlah المماثلة similar]
SYN [sahlht سهلة easy] = [basedh بسيطة simple]
SYN [almrad المراد wanted] = [almtlwb المطلوب required]
SYN [almjawrh المجاورة adjacent] = [almhydeh الحادية close]
SYN [almkhtalefh المختلفة different] = [almtnaw(9)h المتنوعة variant]
SYN [almhddt المحددة specified] = [alm(9)ynah المعينة designated]

SYN [almshakel المشكل problems] = [als(9)wbaht الصعوبة difficulty]
SYN [lttheh لتوضيح to show] = [ltbsedh لتبسيـــــــط to simplify]

Abstract Number 7

يستعرض هذا البحث طرقا متوازية حديثة للتسريع الفائق لمعالجة معلومات الصور الرقمية متعددة الأطياف مع إمكانية مضاعفة استيعاب المعالجات لهذه المعلومات بتحميل المقدار التخيلي لأرقام تحويل فوربير بالنصف الأخر من هذه المعلومات و تستخدم هذه الطرق معالجات الترانسبيوتر الحديثة مع لغة اوكام للبرمجة.

English Translation of Abstract Number 7

Parallel processing techniques are used to increase the speed and efficiency of handling remote sensing data. This technique with inter leaving of data from images are proposed in order to expedite the performance of Fast Fourier Transfs. The algorithm uses the fast transputer processors of simultaneous handling of two or more data streams are demonstrated.

Abstract Number 7 (underlined words are inserted synonyms)

يقدم هذا البحث طرقا متوازية جديدة للتسريع الفائق لمعالجة بيانات الصور الرقمية متعددة الأطياف مع إمكانية مضاعفة استيعاب المعالجات لهذه البيانات بتحميل المقدار التخيلي لأرقام تحويل فوربير بالنصف الآخر من هذه المعلومات و تستخدم هذه الطرق معالجات الترانسبيوتر الجديدة مع لغة اوكام للبرمجة

In this abstract (#7) each underlined word is a synonym replacing the original word. The words involved were [يستعرض show ,يقدم introduce], [الخديثة حديثة recent, الجديدة new] and [معلومات المعلوماتinformation بيانات data]. If we look at the meaning of the Arabic text, we find it reasonable. The sentences with substitutions were judged to be perfectly reasonable and

to have unaltered meaning. We performed the same experiment with Abstract Number 17.

Abstract Number 17

يقدم هذا البحث طريقة سهلة و سريعة لتحديد شكل و حجم الحرف في النص العربي. تستخدم هذه الطريقة ثلاث جداول لأشكال و حجوم و خصائص الحروف العربية و اللاتينية, وبمعالجة سهلة لخصائص الحروف المجاورة يمكن تحديد شكل و حجم الحرف المراد.

English Translation of Abstract Number 17

The paper presents a simple and fast algorithm to determine the shape and width of a given Arabic character within Arabic text. The same algorithm is currently implemented in a bilingual information retrieval system utilizing the alphanumeric mode of the IBM PC/AT. The working and characteristic of the text editor are discussed.

Abstract Number 17 (underlined words are inserted synonyms)

يعرض هذا البحث طريقة بسيطة و سريعة لتحديد شكل و حجم الحرف في النص العربي. تستخدم هذه الطريقة ثلاث جداول لأشكال و حجوم و خصائص الحروف العربية و اللاتينية, وبمعالجة سهلة لخصائص الحروف المحاذية يمكن تحديد شكل و حجم الحرف المطلوب

In Abstract 17 each underlined word is synonym replacing the original word. The substitutions are [يقدم introduce, يعرض show], [سهلة easy, بسيطة simple], [المجاورة adjacent, المحاذية close] and [المراد wanted, المطلوب required]. If we look at the Arabic text, we find it reasonable. The sentences with substitutions were judged to be perfectly

reasonable and to have unaltered meaning. We carried out the same experiment with Abstract 44.

Abstract Number 44

في هذا المقالة نعرض الطرق المختلفة لتخطيط و تصميم شبكات من منظور الملكة العربية السعودية, و تركز المقالة على مشكلة المهام المحددة للوصلات و ذلك لأهميتها و تركز المقالة على ثلاثة أنواع من الشبكات و و الموزعة, وتتعرض المقالة أيضا للمشكلة الهامة و الخاصة بحماية البيانات باستخدام هي :النجمية, الشجرية طرق الشيفرة.

English Translation of Abstract Number 44

Discusses the approach in the planning and designing the required computer communications network. Definition of user requirements and network design methodology constitutes the core of the planning methodology. User requirements of Saudi, universities, banks and Ministries of health and Education are outlined. Three different network topologies, namely star, tree and distributed are analyzed for the Kingdom situation. Link assignment problem in the design process is tackled. Two basic approaches for data security, namely data encryption standard and public key cryptosystem are briefly described.

Abstract Number 44 (underlined words are inserted synonyms)

في هذا المقالة نقدم الطرق المتنوعة لتخطيط و تصميم شبكات من منظور الملكة العربية السعودية وتركز المقالة على مشكلة المهام المعينة للوصلات و ذلك لأهميتها وتركز المقالة على ثلاثة أنواع من الشبكات و هي :النجمية, الشجرية و الموزعة, وتتعرض المقالة أيضا للمشكلة الهامة و الخاصة بحماية البيانات باستخدام طرق الشيفرة.

In Abstract 44 each underlined word is a synonym replacing the original . The substitutions are [نعرض show, نقدم introduce], [المختلفة different, المتنوعة variant] and [المحددة specified, المعينة designated] If we look at the meaning of the Arabic text, we find it reasonable. The sentences with substitutions were judged to be perfectly reasonable and to have unaltered meaning.

Abstract Number 83

معظم الكمبيوترات العربية المستعملة حاليا تعتمد على مقياس و.و.م. والتي معها وحدة المدخلات و المخرجات لمعالجة النص العربي. ومن جهة أخرى لقد صممنا ونفذنا و.و.م. خاصة لها المقدرة على معالجة شكل الحرف مع التشكيل, و الميزة هي تقليل التكلفة, وتحسين الفعالية والنتائج , سهولة برمجة و معالجة الحرف العربي

English Translation of Abstract Number 83

The design, simulation and implementation of SPARC- a special purpose computer for Arabic text processing are discussed. The machine includes special instruction to handle Arabic characters along with their shapes and vowels (Harakat). The system will be particularly useful for office automation and for Arabic databases where unvowelizwd and vowelized text might be mixed. The advantage of this computer are: reduced overall system cost. improved performance and throughput, ease of programming and flexibility in character manipulation.

Abstract Number 83 (underlined words are inserted synonyms)

غالبية الكمبيوترات العربية المستخدمة حاليا تعتمد على مقياس و.و.م. والتي معها وحدة المدخلات ومخرجات لمعالجة النص العربي.ومن جهة أخرى لقد أصررنا ونفذنا و.و.م خاصة لها المقدرة على معالجة شكل الحرف مع التشكيل, و الميزة هي تقليل التكلفة, وتحسين الفعالية والنتائج , سهولة برمجة و معالجة الحرف العربي

In this Abstract 83 each underlined word is a synonym substituted for the original as follows: [معظم most, غالبية majority], [المستعملة used, المستخدمة decided], [صممنا persist, أصررنا insist], [تقليل decrease, تنقيص reduce] and [التكلفة cost, القيمة value]. The sentences with substitutions were perfectly correct except for the particular substitution of [صممنا persist, أصررنا insist]. If we look at the meaning of the Arabic text we find a problem in the new text at this point.

Note that the results here suggest that synonymy does exist at least with the Computer Science sublanguage. The problem in the Abstract 83 suggest that with polysemous words, like [صممنا persist, أصررنا insist], we need to consider word senses, not words. The WordNet created at Princeton University by George Miller and a team of colleagues uses "synsets" to handle this problem [Miller et al., 1990]. In other words, they record of synonymy between word senses, not between words.

CHAPTER IV

LEXICAL-SEMANTIC RELATIONS

4.1 Definitions

A lexical-semantic relation is a rule that links words together. These links may represent semantic, syntactic, or morphological relationships. The properties of a relation are important, since they enable us to make inferences. These relations should be made a part of an Arabic lexicon.

4.2 Taxonomy

Taxonomy can be used to relate nouns or adjectives or verbs: a swan ISA bird and a bird ISA animal. Taxonomy has the transitivity property. If A ISA B and B ISA C, then A ISA C. Computer models of memory have used taxonomy extensively starting with Quillian [1968] and Raphael [1968].

Example:

Tax [Lion اســـد] = [Animal حيـــوان]

4.3 Antonymy

Antonymy is not an easy relation to describe even though this relation is familiar. Casagrande and Hale [1967, p.183] discovered that antonymy appears often in Papago folk definitions. They classified as antonymy all examples in which x is described as "the negation of y, its opposite." Researchers have identified the following kinds of oppositeness [Evens et al. 1983].

Anti [sakh(i)n ساخى hot] = [bar(i)d بارد cold]

where assertion of one implies the denial of the other, but the denial of one does not imply the assertion of the other.

Comp [mutazawij متزوج married] = [a9zab اعزب single]

The assertion of one implies the denial of the other and the other way around.

Conv [yu9ty يعطي give] = [yaakhuth ياخذ take]

The converse of give is take. We describe this as conversiveness. The core meaning is the same but the arguments are switched.

Reck [zawji زوج husband] = [zawjh زوجه wife]

Reck stands for reciprocal kinship. This relation can be used to represent relationships between people [Evens and Smith, 1978].

4.4 Grading

Grading has been used by different researchers and authors to refer to linearly ordered, or scalar continua. Casagrande and Hale [1967, p. 184] defined grading as "where x is defined with respect to its placement in a series or spectrum that includes y." Grading is not reflexive or symmetric but some types of grading are transitive; others are not.

Examples:

Queuing [elahad الأحد Sunday] = [elthnean الاثنين Monday]

Stage [jled جليد ice]=[mai ماء water]

Here, the queuing relation seems to be transitive but the stage relation is not.

4.5 Attribute

The attribute or modification relation is used to describe distinguishing aspects of nouns or verbs. Casagrande and Hale [1967, p.168] classify as attributive any definition in which "X is defined with respect to one or more distinctive or characteristic attributes Y."

Examples:

Time [h(a)r حر hot] = [S(a)yf صيف summer]
Color [sma' سماء sky] = [ezarga'a أزرق blue]
Home [jamal جمل camel] = [sahra'a صحراى desert]

4.6 Collocational Relations

Collocational relations are found by the study of repeated co-occurrences of words in consistent patterns. The following collocation relations are derived from the work of Apresyan et al. [1970].

Cont [s(a)lam سلام peace] = [y(a)hf(a)th يحفظ maintain]

The Cont relation relates a noun to the verb meaning to cause it to continue. The Fact relation also relates nouns to verbs. In this case the associated verb indicates the standard function of the noun. Sometimes these verbs are called "functional verbs."

Fact [thyab ثياب clothes] = [y(a)lb(a)s يلبس wear]

4.7 Paradigmatic Relations

Paradigmatic relations relate words that have the same core meaning. Sometimes, the relation is expressed by a kind of derivational morphology.

Examples:

Become [ahmr احـمـر red] = [ahmmar احمرار redden]
Mode [yalbas يلبس wear] = [sha^by شعبي traditional]
Terminate [juo(9) جوع hunger] = [ysh(9) يشبع satiate]

4.8 Part-Whole

In 1967, Casagrande and Hale called the part-whole relation "the constituent relation" and described it as "X is defined as being a constituent or part of Y" [p.156].

"In English the part-whole relation seems to be expressed most often with have, of, or the possessive". The part-whole relation is important for Arabic.

[ras راس head] = part [jesm جسم body]
[worayga ورقه petal] = part [zahre زهره flower]
[isba اصبع finger] = part [yad يد hand]
[yad يد hand] = part [jesm جسم body]

The part-whole relation has been proven to be very important in definition-making as demonstrated by Smith's study [1985], which found part to be the second most common noun used in noun definitions in Webster's Seventh Collegiate Dictionary

[Iris et al., 1988] analyzed the use of the part-whole concept and the word "part" in definitions of body-part words. After they analyzed dictionary data, they found the

part-whole relation to have four different senses: the relation of the functional component to its whole, the relation of the segment to the segmented whole, the membership relation, and the set inclusion relation.

CHAPTER V

DATA AND DATABASE

This chapter describes the design of the database in which I have stored the information about lexical-semantic relationships that I extract from the Arabic abstracts of papers in computer science. This database is an extension of the one developed by Alsamara and described in his thesis [1996]. I agree completely with the relational database approach used by Alsamara.

This database approach has several advantages. The data model is completely independent of how data is stored and accessed. The resulting database is a combination lexicon and thesaurus like those described by Calzolari [1988] and Evens [Evens et al., 1991]. Alsamara's lexicon database is a relational database stored in the PC Access System marketed by Microsoft. It is designed to store sixteen bit characters and it is available for use with Arabic Windows.

Abu-Salem [1992] entered 120 abstracts. Hmeidi entered 122 more to give us 242 abstracts, which he used in a series of information retrieval experiments [Hmeidi et al., 1997]. The lexicon developed by Alsamara was extracted from a corpus of 242 abstracts of papers in computer science from the Saudi Arabian National conferences.

Alsamara's lexical database contains five tables. The Main Table has entries for each word in the 242 abstracts in our corpus. The other four tables are the Verb Table, the Particle Table, the Noun Table, and the Adjective Table. (Although adjectives and nouns are usually considered to be the same part of speech in Arabic, Alsamara separated them because of the necessity of storing masculine and feminine forms for adjectives.) My

database is stored as a group of tables added to the core set of five tables defined and built by Alsamara. As a first step I entered 16 more abstracts and then added to Alasamara's database all the new words in those 16 abstracts. This gave me an opportunity to get thoroughly familiar with Alsamara's work.

I have considered two different approaches to the problem of extending the lexical database to support the thesaurus that I am building. The simplest approach is to add just one table to contain all the word-relation-word triples, sorted by the first word in the triple. This approach is ideal if the most common use is to find all words related to a given word. It will not work so well when we need to manage applications where we want to use some relations but not others.

The alternative design is one in which data about different relations is stored in separate tables. This design supports a wide variety of applications. I have implemented the thesaurus in this form first. If this design proves to create performance problems for a particular type of application, I can write a simple SQL program to produce combined tables. My design for the Thesaurus Database appears in Figure 5 below. It uses Alsamara's Main Table plus a table for each of the major relations in the thesaurus.

5.1 Thesaurus Tables

The thesaurus contains approximately 4547 different word forms (258 Arabic abstracts of papers in computer science)

Relation	No. of Entries
Antonymy	1073

Synonymy	2303
Taxonomy	210
Grading	300
Attribute	150
Collocational Relations	190
Part-Whole	180
Paradigmatic Relations	141
Total	4547

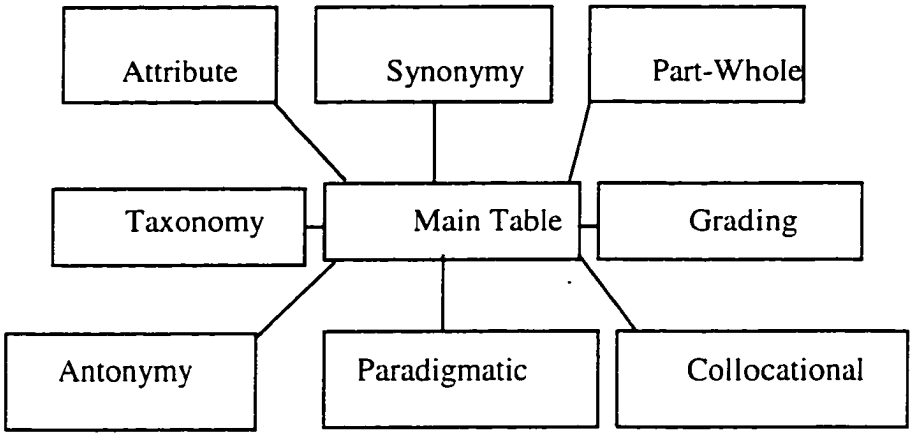


Figure 3. Organization of the Thesaurus Database

The Synonymy Table contains words and their synonyms. Each word from the first column of the Synonymy Table (Figure 4) appears in the Main Table.

Word	Synonym
المعلومات Information	البيانات Data
الحديثة Recent	الجديدة New

Figure 4. Synonymy Table for Arabic Words

In the Taxonomy Table we store all ISA relationships between words. The Taxonomy Table for Arabic words is illustrated in Figure 5.

Word	Taxonym
أسد Lion	حيـــوان Animal
Microcomputer	حاسب Computer

Figure 5. Entries in the Taxonomy Table.

In the Antonymy Table we store all examples of opposites found in the abstracts. A small piece of the Antonymy Table for Arabic words is illustrated in Figure 6.

Word	Antonym
ساخن hot	بارد cold
متزوج married	أعزب single

Figure 6. Entries in the Antonymy Table

5.2 Methodology for Finding Thesaurus Entries

Most Arabic dictionaries are organized in the traditional way; each entry consists of a root and all its derivatives. I searched entries for related words and then tried to figure out the relationships involved. The Arabic dictionaries do not include all semantic relations many of the new technology related words that appear in the Arabic Computer Science abstracts did not appear in the dictionary at all. The Computer Science abstracts do not handle vowelization. The lack of vowelization in the abstracts sometimes made it hard to ascertain the words in question. I also followed Apresyan's definitions of the

different semantic relations; with new words this seemed to be a more suitable way to find thesaurus entries. The problem of prefixes and suffixes is avoided using these definitions, and entries can be found for any word without going back to its root.

CHAPTER VI

THE THESAURUS INTERFACE

While building the Thesaurus required a great deal of the manual effort, there are several ways in which the computer provided support for this effort. I have stored those relationships in different files using a Microsoft Access Database. Also, I wrote a program in Visual Basic as an interface to read those files from the database, so that the user can enter a word and find all the related words. The system can be used to enter relationships as well. This application requires a Microsoft Arabic Windows environment to run.

Table 1. System Requirements for the Thesaurus

Component	Description/Comment
Microprocessor	80386 or higher
RAM	8MB
Hard disk	A hard disk is required
Microsoft Windows 95	Arabic Version 7
Mouse	Recommended

I designed and implemented an Arabic Thesaurus containing the vocabulary of the 258 computer science abstracts. The system was designed to help the user invoke various procedures for manipulating the system from a Graphic User Interface (GUI). It is

Windows 95 oriented and menu driven. The instructions for using this system are very simple:

- Double-click القاموس الموضوعي on the Windows icon from the Program Manager.
- Enter the word you want to search for and click on the relation you want.

The user has the opportunity to print the relationships that appear in each window. The user can exit by hitting the EXIT button.

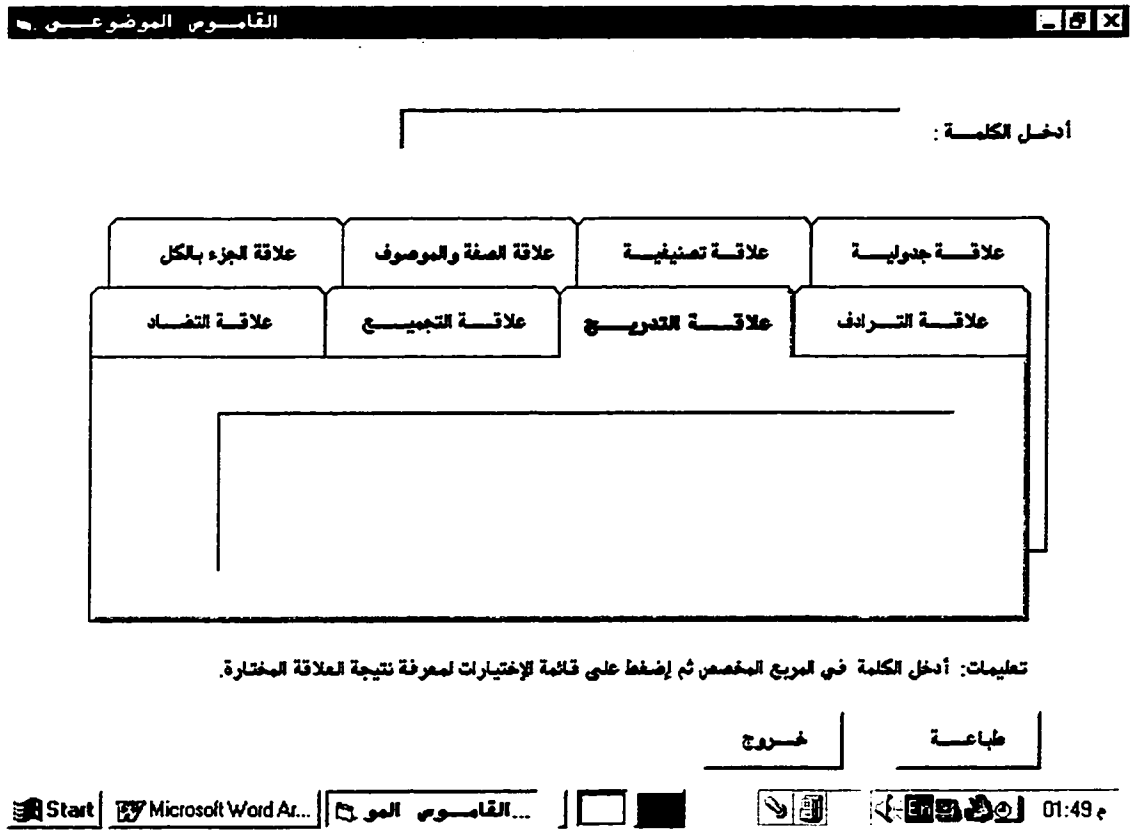


Figure 7. Basic Information Window.

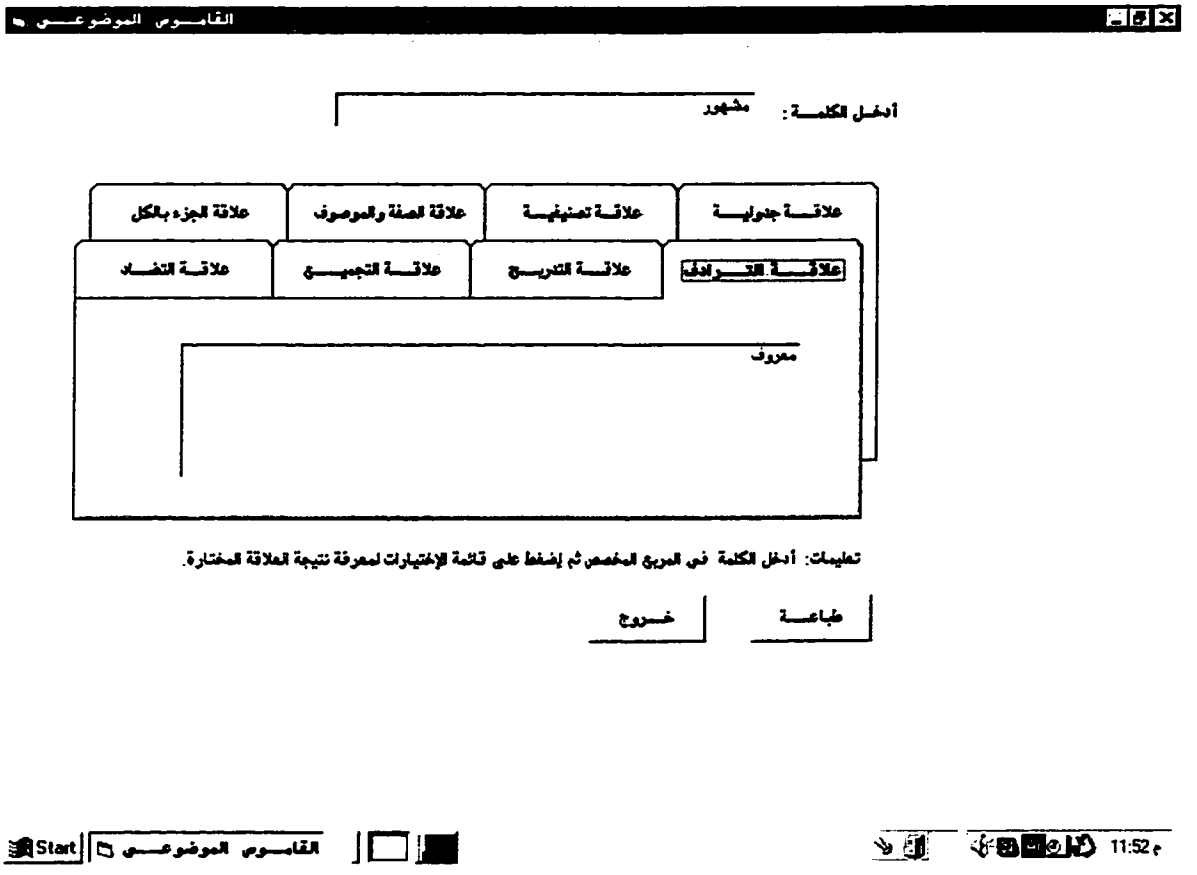


Figure 8. Synonymy Window. Figure 8 shows the Synonymy Window. The user has entered the word [مشهور] and found the Synonym[معروف] listed in the window.

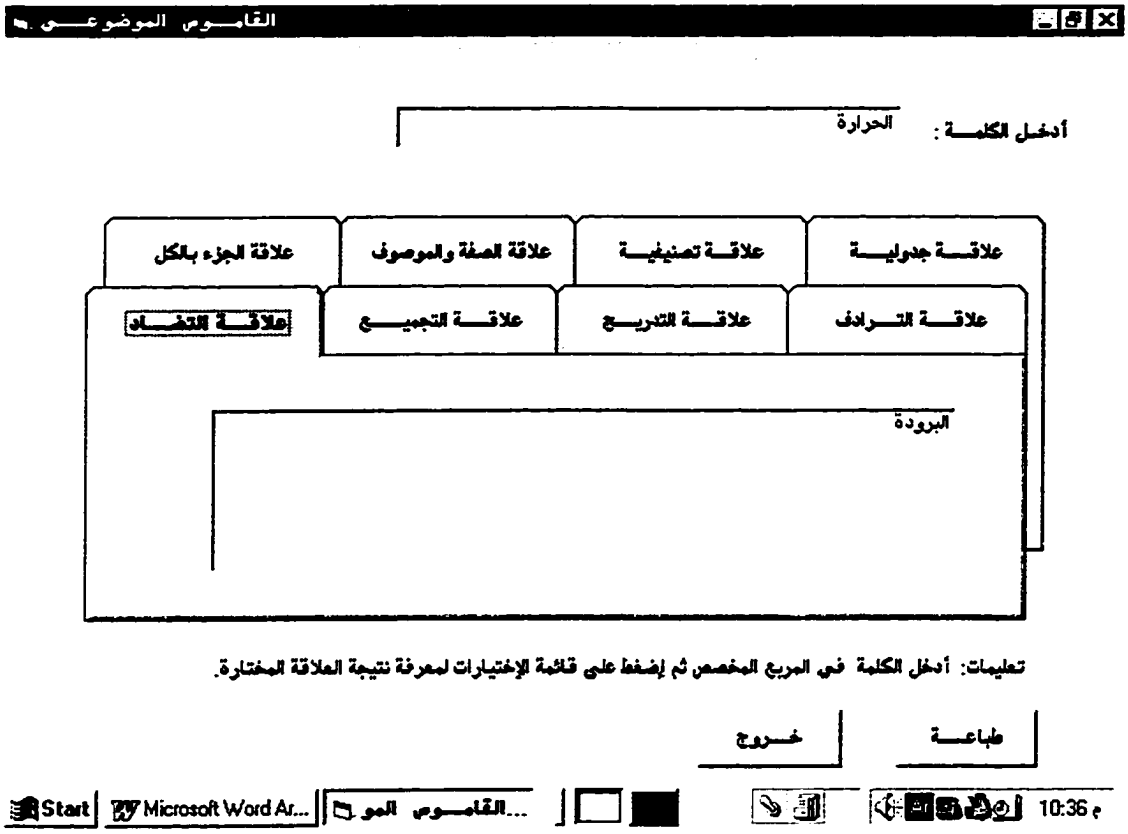


Figure 9. Antonymy Window. Figure 9 shows the Antonymy Window. The user has entered the word [heat, الحرارة], and found the Antonym [cold, البرودة], listed in the Window.

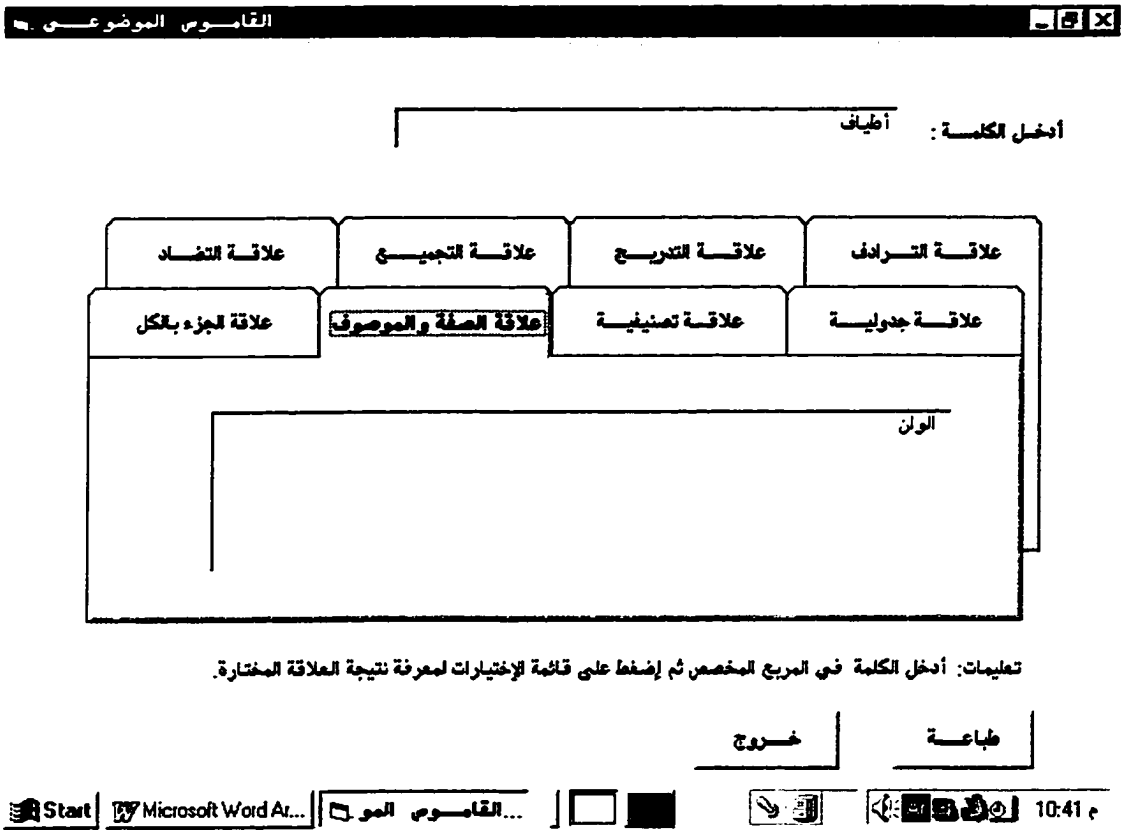


Figure 10. Attribute Window. Figure 10 shows the Attribute Window. The user has entered the word [spectrum, أطياف] and found the Attribute [colors, ألوان] listed in the window.

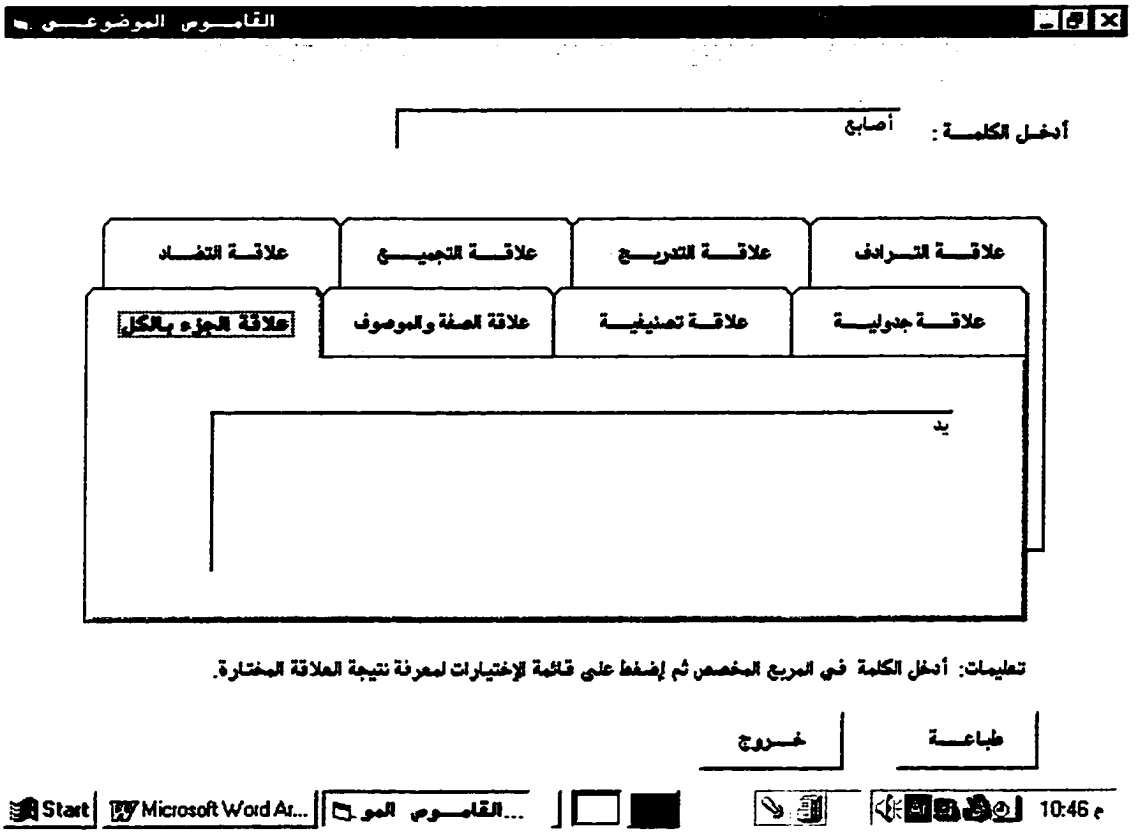


Figure 11. Part-Whole Window. Figure 11 shows the Part-Whole Window. The user has entered the word and found [finger, أصابع] Part [hand , يد] listed in the window.

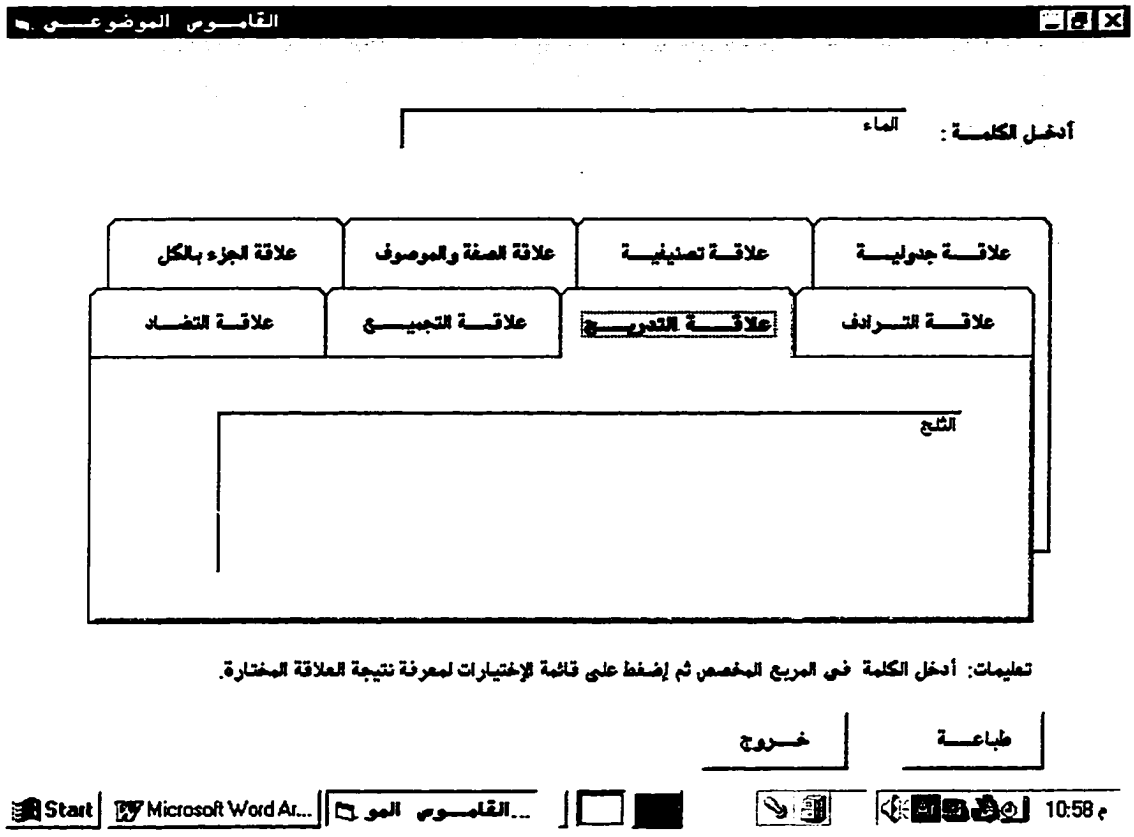


Figure 12. Grading Window. Figure 12 shows the Grading Window. The user has ientered the word and found the [water, الماء] Grad [ice, الثلج] listed in the window.

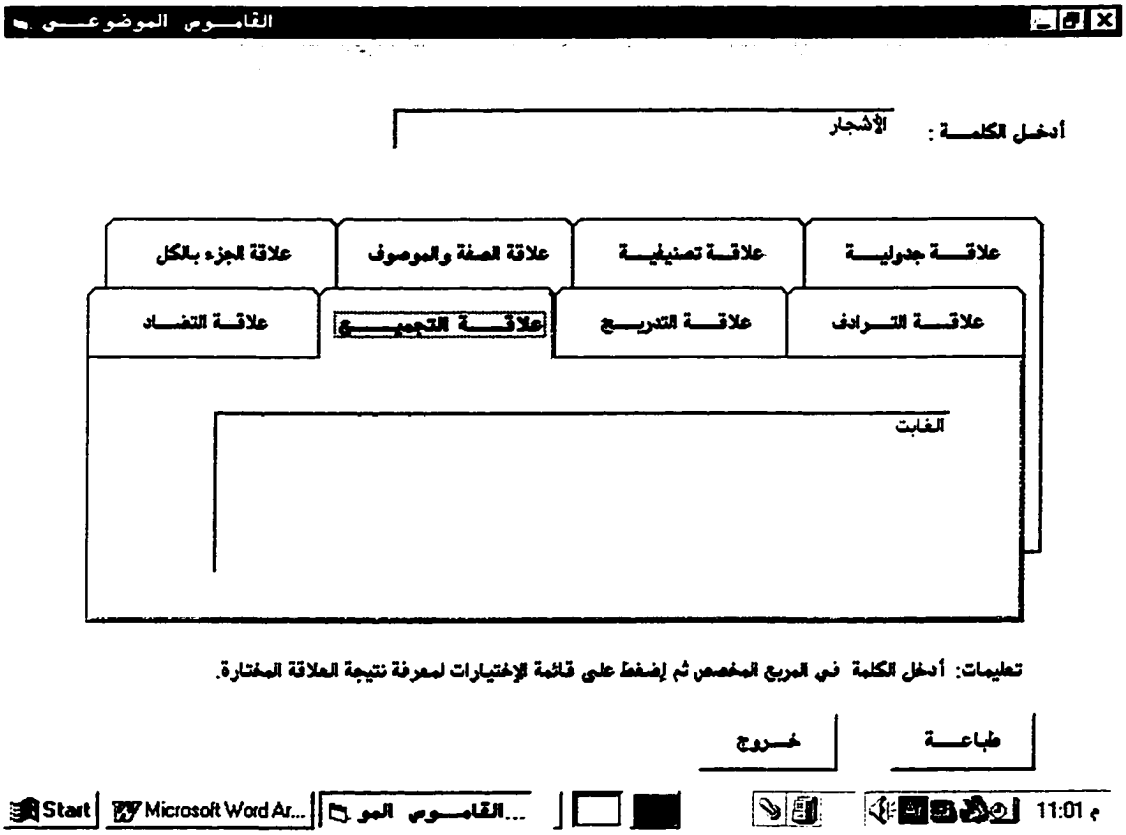


Figure 13 Collocation Relation Window. Figure 13 shows the Collocation Window. The user has entered the word [trees, الاشجار] and found the collocation [forest, الغابات] listed in the window.

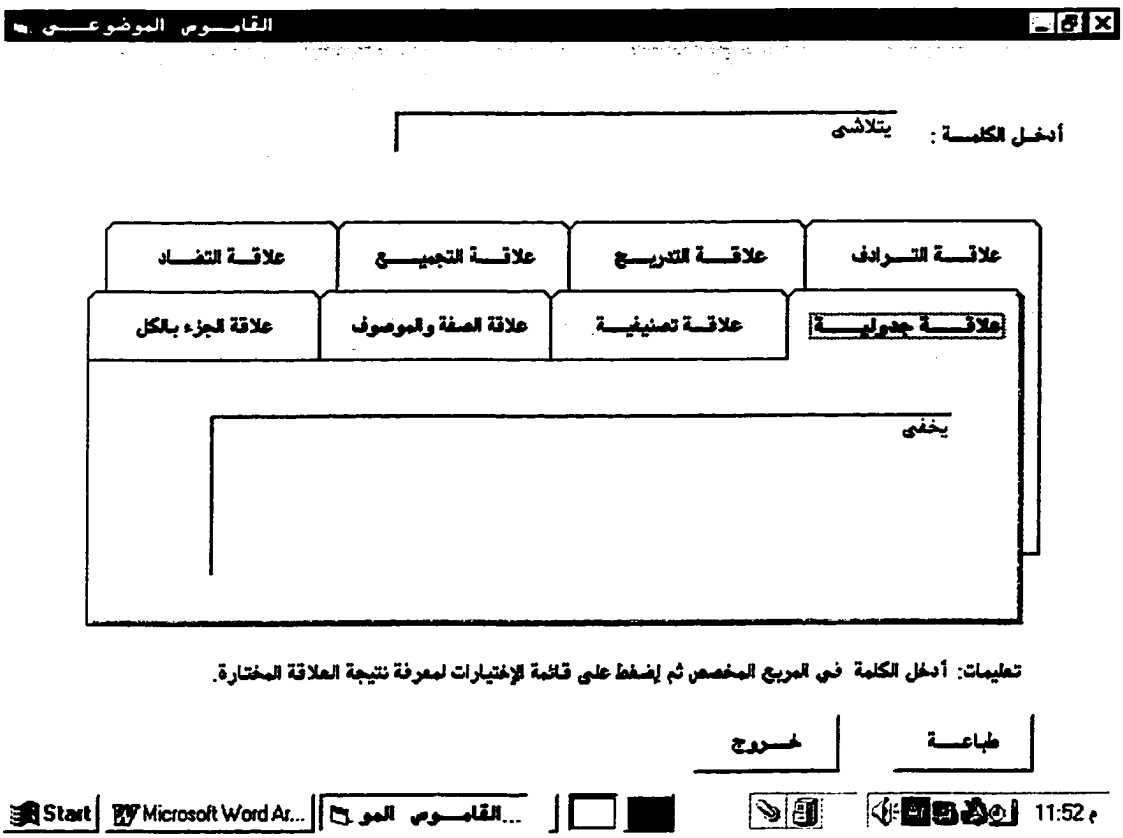


Figure 14. Paradigmatic Relation Window. Figure 14 shows the Paradigmatic Relation Window. The user has entered the word [perish, يتلاشى] and found the related word [disappear, يخفي] listed in the window.

CHAPTER VII

MORPHOLOGY

Morphology is the study of word structure [Ritchie et al., 1992], the area of Linguistics that is concerned with studying the internal structure of words. There are two kinds of morphology: derivational and inflectional morphology. Derivational morphology is the formation of new words from existing words. Inflectional morphology involves adding affixes in order to produce different forms. The best choice of morphological information for the lexicon of the Arabic language is still an issue of great controversy. Some researchers believe that the lexicon must be based on roots; others believe it has to contain all inflected forms.

In some languages, such as English, the common practice is to store all words needed, including those formed by adding prefixes and suffixes to other words [Yahya, 1989]. There are systems that store all the inflected forms of a lexical entry because of efficiency considerations [Ingria, 1987]. The strategy of storing all possible words in Arabic, which is used for some systems in the English language, does not appear to be practical.

One of the easiest ways to understand which form variants should be stored in the lexicon-thesaurus is to consider the parser. A parser is presented with many sentences; the sentence, in turn, is made up of sequences of characters delimited by white spaces. Such strings of characters (orthographic words) can be one of the following:

Root: the word is not derived from any other word.

Stem: the word is derived from a root word. Most stems denote a related entity, action, or concept.

Augmented word: the word is either a stem word or root word with bound morphemes, attached to it. The term "bound morphemes" is used to denote specifically the prefixed definite article, pronominal suffixes, and certain prepositions, which are attached to the words they modify without a word space.

The morphology of the Arabic language depends on a root-pattern structure. Most Arabic words are built up on a basic consonantal skeleton. The two concepts of root and pattern are fundamental to the structure of Arabic words. The root usually has some fundamental kernel of meaning, which is expanded or modified by the pattern. A verb root is usually a sequence of three consonants (a triliteral root). Yahya [1989] defined a root word as a word that is not derived from any other word or any combination of words. Some roots contain four letters or five letters. Al-Shalabi [1996] found that 95% of the Arabic word types in the abstracts are derived from a triliteral root.

An example of a root is the consonant sequence KTB. The core meaning of this root is the concept of "writing".

For example:

[k(a)t(a)ba كـتـب he wrote]
[iketbu اكـتبو they write]

7.1 Tasrif

The Arabic word for morphology has the root (srf). The basic idea is changing direction, averting, and flowing freely. Tasrif is the total range of morphological patterns derived from a given root [Owens, 1988]. The Tasrif of the root "ktb" (concerning writing) is shown as an example in Figure 16.

Figure 15. The Tasrif of the Root (ktb)

Word		Pattern	Relation
kataba	كتبوء	f(a)9(a)l(a)	past
yaktubu	يكبو	y(a)f9l (u)	imperfect
maktub	مـكتوب	m(a)f9(u)wl	past participle
takataba	تكتبوا	t(a)f(a)9(a)l(a)	reciprocal

Tasrif describes the changes in the word as different form variants derived from the same root.

7.2 Ishtiqaq

"Ishtiqaq" in Arabic is the process of forming one word form and meaning from another, which involves a change in both form and meaning [Owens, 1988]. Most Arabic words are derived from roots by adding affixes to the root. Ishtiqaq refers to seven nouns, all derived from verbs: the comparative noun, the active and passive participles, the

verbal nouns of time and place, the verbal noun of instrument, and the noun of exaggeration.

Examples:

[walad ولد boy]=[waladan ولدان two boys] by adding the suffix
[ktb كتب writing]=[yaktbu يكتب writes] by adding the prefix

7.3 Added Sounds and Morphemes

When we talk about tasrif, we are concerned with the root and morphological pattern. Augmented or derived words contain a root morpheme and suffix. For example:

al-rajul	الرجل	the man	(def + man)
katab-a	كتبا	they (dual)	wrote
katab-u	كتب	they (masculine plural)	wrote

Figure 16. Augumented Morphemes and Morphological Patterns

In this example, there are two obvious candidates for segmental morphemes, a root and prefix and verb + pronoun suffix.

In the structure of the lexicon-thesaurus it is necessary to distinguish various pieces of information regarding orthographic words. Such information, therefore, can be used by natural language application systems, in order to associate each word presented to the system with at least one word in the lexicon-thesaurus. In case of augumented words, it is necessary to identify the prefixed definite article and suffixed pronouns and other affixes before any other processing.

CHAPTER VIII

CONCLUSION

8.1 Summary

The major goal of the research described in this thesis was to find all the relationships between the words that occur in the Alsamara Lexicon, to investigate lexical-semantic relations for Arabic and make a list of appropriate ones, and use them in building a thesaurus. The vocabulary in the Alsamara lexicon was extracted from our corpus of 258 Arabic abstracts of papers in computer science in machine readable form. This thesaurus was designed to support several natural language applications including information retrieval parsing and text generation; it will be tested in ongoing research at the Arabic Language Processing Laboratory at Illinois Institute of Technology and in future research on the Arabic language. My own first goal was to learn a variety of things about the language, most particularly about relationships between word form and word meaning.

We realized that much of the work done for other languages such as English in the field of lexicon design for natural language applications is applicable to the Arabic language. Some relations seem to be language universal, particularly taxonomy (often called the IS-A relation by computer scientists). Others are language specific.

We started with the relational database created by Alsamara and redesigned it to store relationships between words as well as the words themselves. We created thesaurus entries. We wrote a program in Visual Basic as an interface to retrieve words and relationships from the database, so that the user can select the appropriate relation or

related word. This application requires a Microsoft Arabic Windows 95 environment and Visual Basic Version 5.0 to run.

8.2 Future Research

The first area that needs further research is the expansion of the thesaurus to cover a more extensive vocabulary. Most of the papers in the our corpus involve computer science. My next version will include the vocabulary from a newspaper corpus, the Al Raya corpus, and investigate the lexical semantic relations between those words.

The second area of planned research is to study the effectiveness of the thesaurus in Information Retrieval. I am already planning an experiment with Akkawi's system [1998].

A third area of planned research to use the thesaurus to support natural language understanding.

Finally, we want to try out the system with other natural language applications including text generation.

APPENDIX

LIST OF WORDS AND SYNONYMS

الكلمة	المرادف
أجبر	أرغم
أحد	احث
أحد	وحيد
أحد	مفرد
أحزاب	جماعة
أحزاب	تجمع
أحزاب	معشر
أحزاب	زمرة
أحزاب	فرقة
أحزاب	فريق
أحزاب	جماعات
أحزاب	طائفة
أخذ	اتخذ
أخذ	تناول
أخذ	تعاطى
أخذ	تسلم
أخيرا	ختام
أخيرا	نها
أداة	وسيلة
أداة	واسطة
أداة	سبيل
أداة	سبب
أدلة	علامة
أدلة	علم
أدلة	رمز
أدلة	سمة
أدلة	برهان
أدلة	اثبات
أدلة	شاهد
أساسية	رئيسية
أسجد	اتعبد
أسجد	اعبد
أفضل	اظل
أفضل	نفوق
أفضل	ابقى
أمتحن	اختبر
أمتحن	جرب
أمتحن	فحص
أمتحن	راجع
أمتحن	استجوب
أوزان	اثقال

أية	علامة
ابتكار	ابتداع
ابتكار	استنباط
ابتكاره	تقليده
ابتكاره	ابتداعه
ابتكاره	استنباطه
اتقان	استحكم
اجتذب	استهوى
اجتذب	استدرق
اجتذب	استقطب
اجتذب	جلب
اجتذب	استجلب
اجمالى	تحمين
احتمالات	توقعات
احتمالية	توقعية
احدث	استجد
ادرجت	ادخلت
ارشاد	توجيه
اركان	اعمدة
اساس	قاعدة
اساسى	قاعدي
اساسى	رئيسى
اسم	مسمى
اسم	تسمية
اشتراك	مساهمة
اشتراك	تشارك
اشتراك	مشاركة
اشتراك	مقاسمة
اشتراك	مشاطرة
اشتملت	تضمنت
اشتملت	احتوت
اشتراكات	مساهمات
الاحتمالا	التصورات
الادراك	الوعى
الادراكى	الوعى
الارشادات	التوجيهات
الاستخدام	الاستعمال
الاستدلال	الاستنتاج
الاستدلال	الاستنباط
الاستدلال	الاستخلاص
الاستدلال	الاستخراج
الاسترجاع	الاسترداد
الاسترجاع	الانسحاب
الاستعداد	الاهابة
أدرس	ادرس
الاستعداد	العدة

الاستعلام	الاستفهام
الاستعلام	الاستيضاح
الاستعلام	السؤال
الاستعمال	الاستخدام
الاستغلال	الاستثمار
الاستفهام	الاستفسار
الاستفهام	الاستيضاح
الاستفهام	الاستجلاء
الاستفهام	السوال
الاستفهام	الاستخبار
الاستقبال	الملاقاة
الاستقرار	الاثبوت
الاستقرار	الارسوخ
الاستقرار	التوازن
الاستقرار	الاستتباب
الاستقصاء	الفحص
الاستقصاء	المعاينه
الاستقصاء	المراجعة
الاستقصاء	البحث
الاستقصاء	استعراض
الاستقلال	التحرر
الاستنتاج	استدلال
الاستنتاج	استخلاص
الاستنتاج	استخراج
الاستهلاك	استنفاد
الاستهلاك	افناء
الاستهلاك	نهك
الاستيراد	التوريد
الاستيراد	ايراد
الاستيراد	اجتلاب
الاستيراد	استجلاب
الاسمية	اعتبارية
الانقلاب	التغير
البالغ	العاقل
البالغة	العاقلة
البالغين	العاقلين
الباهضة	المرتفعة
الباهضة	المكلفة
التطوير	النماء
التطوير	التنمية
التطويرية	النماية
التظهير	التوضيح
التمثيل	المناب
التمثيل	النوب
التنازلى	الخدلان
التنازلى	التفضل

التنازلى	التلطف
التنبؤ	التكهن
التنسيق	الترتيب
التنسيق	النظام
التنظيم	الترتيب
التنظيم	النظام
التنظيم	التمهيد
التنظيمى	الترتبى
التنظيمى	النظامى
التنظيمى	التمهيدي
التنظيمية	الترتبية
التنظيمية	التمهيدية
التنظيمية	التسوية
التنظيمية	الترتبية
التنفيذ	التطبيق
التنفيذ	الانجاز
التنفيذي	التطبيقى
التنفيذي	الانجازي
التنفيذية	التطبيقية
التنمية	التطوير
التنمية	انماء
التنويه	الثناء
التنويه	التلميح
التنويه	التعريض
التهديدات	التوعيدات
التوسيع	التعريض
التوعية	التنوير
التوعية	التبصير
التوعية	التحذير
التوعية	التنبيه
التوقع	الاحتمال
التوقع	الترجح
التوقع	التقدير
التوقع	التكهن
التوقع	تبؤة
الجر	السحب
الحفر	النقر
الحمل	الثقل
الرسوم	التصوير
الرسوم	الجباية
الرصيد	الباقى
الرطوبة	الندى
الرطوبة	التبلل
الرطوبة	البلة
السعر	القيمة
السفر	التجول

السيطرة	التحكم
الشيخ	الهرم
الشيوع	الانتشار
الشيوع	التفشى
الشيوع	السطوع
الصانعة	الخالقة
الصانعة	المبتعدة
الصانعة	المستحدثة
الصدارة	التقدم
الصدد	الخصوص
الصدد	الوجهة
الصدد	الناحية
الصدد	القصد
الصرف	العزل
الصرف	الطرد
الصرفى	العزلى
الصرفى	الطردي
الصرفية	العزلية
الصرفية	الطرية
الصريح	المصارح
الصريح	النقى
الصريح	الصافى
الصعوبات	المشقات
الصعيد	التلة
الصعيد	الهضبة
الصعيد	المرتفع
الصفا	الطابورا
الصفات	الخصلات
الصفر	الذهب
الصفر	العسجد
الصفر	النضار
الصفوف	الطوابير
الصلة	الرابطة
الصلة	العلاقة
الصم	الطرش
الصم	الصلب
الصم	الجامد
الصناعة	الحرفة
الصناعى	الحرفى
الصهر	اذابة
الصوري	البوق
الصياغة	التشكيل
الصياغة	التكوين
الصياغى	التركيب
الصياغى	التكوين
الضرورة	الزوم

الضرورة	الواجب
الضرورة	المدعاة
الضوء	النور
الطبع	الختم
الطبع	الدمغ
الطبع	المهر
الطبع	الوسم
الطبيعة	الفطرية
الطبيعة	الخلقية
الطبيعة	البسيطة
الطبيعى	الفطري
الطبيعى	الخلقى
الطبيعى	البسيط
الطيب	النيذ
الطيب	النمير
الطيران	التحليق
الظرف	الحال
الظرف	وضوع
العادي	العتيق
العادي	الدهري
العادية	العتيقة
العادية	الدهرية
العشرة	الصحبة
العشرة	الرفقة
العشرة	المخالطة
العصبية	القبلية
العصبية	العشارية
العصر	الزمن
العصور	السنين
العصور	القرون
العلاج	الدواء
العلاقة	الروابط
العلوم	المعرفة
العمل	الشغل
العمل	المهنة
العمل	الوظيفة
العملة	النقد
العمود	السارية
العميقة	السحيقة
العميقة	المقعر
العميل	الزبون
العميل	المأجور
العوانق	الصعانب
العوامل	الموثرات
الفتاة	البنة
الفصلين	الصفين

الفيصل	الحكم
الفيصل	المحكم
الفيض	التدفق
الفيض	الدفق
الفيض	الزيادة
الفيض	المزيد
الفيض	الغرط
الفيض	الغزارة
القائم	الناهض
القائم	المنتصب
القائمة	رجل
القائمة	كراع
القائمة	قدم
القائمة	عمود
القائمة	سارية
القائمة	جدول
القابل	الموافق
القابل	الراضى
القابل	المتجاور
القابلة	الموافقة
القابلة	الراضية
القابلة	المتطابقة
القادمة	المقبلة
القادمة	آتى
القادمة	الوافدة
القادمة	الواصلة
القادمين	المقبلين
القادمين	الوافدين
القادمين	الواصلين
القارنة	مطالع
القاعدة	القانون
القاعدة	الدستور
القاعدة	النظام
القاعدة	النقش
القاموس	المعجم
القانون	الشريعة
القانون	التشريع
القانون	الدستور
القانون	الناموس
القاهر	القوي
القاهر	الجبار
القاهر	العظيم
القاهر	المستبد
القاهر	الطاغية
القاهر	الغاشم
القدرة	الاستطاعة

القرآن	المصحف
القواعد	الاسس
القواميس	المعاجم
الكف	المنع
الكف	الردع
الكف	الحجز
المؤلف	الكاتب
المؤلف	الواضع
المبسط	الميسر
المبسط	المتيسر
المبسطة	اليسرة
المبسطة	المتيسرة
المتحولات	المتغيرات
المترجم	المنقول
المجرد	البحث
المجرد	الصرف
المجرد	الخالص
المجرد	الصافى
المجرد	العاري
المحايد	المتحايد
المحددة	المنحصر
المحددة	المحجوز
المحددة	المحصور
المحرر	المخلص
المحرك	الحافز
المحرك	باعث
المحرك	الدافع
المحرك	المثير
المحرك	الجالب
المحرك	المسبب
المرور	العبور
المسار	مسيرة
المساعد	النافع
المساهمة	المساعدة
المستجدات	الجديدات
المستجدات	الحديثات
المستجدات	المستحدث
المستخدم	الموظف
المستخدم	العامل
المستخدم	المشتغل
المستخدم	الشغال
المستخدم	اجير
المستخدم	المستعمل
المستخدمة	الموظفة
المستخدمة	العاملة
المستخدمة	المشتغلة

المستخدمة	الشغالة
المستخدمة	اجيرة
المستخدمة	المستعملة
المستشار	الناصح
المستشار	المرشد
المستشار	القاضى
المستعمل	المستخدم
المستعمل	المعتمد
المستعملة	المستخدمة
المستعملة	المعتمدة
المستعملى	المستخدمى
المستعملى	المعتمدي
المستفيد	المنتفع
المستفيدة	المنتفعة
المستفيدي	المنتفعى
المستقرة	الثابتة
المستقرة	الراسخة
المستوى	المقياس
المستوى	المعيار
المستوى	المحك
المستويات	ممهد
المستويات	سوى
المسح	التدليك
المسح	الدهن
المسموح	الجائز
المسموح	المقبول
المسموح	المستساغ
المسموم	مسمم
المشابهة	الماثلة
المشار	المذكور
المشاعر	الحساس
المشاعر	التأثر
المشاهد	المنظور
المشاهد	المشهود
المشتركة	المشاعة
المعتادة	المالوفة
المعتادة	المعهودة
المعتز	الفخور
المعتقدات	الديانة
المعتقدات	المذاهب
المعتمد	المفوض
المعتمد	الموكل
المعتمد	الوكيل
المعتمد	المندوب
المعتمد	الممثل
المعد	المستعد

المعد	المتأهب
المعد	الجاهز
المعد	المدبر
المعدل	المحور
المعدل	المضبوط
المعدل	المصحح
المعدل	المقوم
المعرف	الدراك
المعرف	التميز
المعرف	الوعى
المعروضة	المسرودة
المعروضة	المقدمة
المعروضة	المقترحة
المعروف	المشهود
المعروف	المشهور
المعروف	المعلوم
المعروفة	المشهورة
المعروفة	المعلومة
المعزولة	المفصولة
المعزولة	المفرق
المعزولة	المشتت
المعزولة	الشتات
المعضلة	المشكلة
المعضلة	المشكل
المعطاة	الممنوح
المعطاة	المهوب
المعطاة	المقدم
المعطاة	المخول
المعطيات	الموهوبات
المعطيات	المقدمات
المعقد	الصعب
المعقد	العويص
المعقد	الدقيق
المعقد	الحرج
المعقدة	الصعبة
المعقدة	العويصة
المعقدة	الدقيقة
المعقدة	الحرجة
المعكوس	المقلوب
المعكوس	المنقلب
المعكوس	المنكوس
المعكوس	المنكس
المعكوسة	المقلوبة
المعكوسة	المنقلبة
المعكوسة	المنكوسة
المعكوسة	المنكسة

المطابقة	المماثلة
المطالب	المسئول
المطالب	الضامن
المطالب	المواخذ
المطبقة	المشابهة
المطبقة	الممثالة
المطبوع	المختوم
المطبوع	المدموغ
المطبوع	الممهور
المطروح	المخصوم
المطروح	المحسوم
المطروح	المحذوف
المطروحة	المحصومة
المطروحة	المحسومة
المطروقة	المدقوقة
المطلوب	المرغوب
المطلوب	المبتغى
المطلوب	المنشود
المطلوب	المراد
المطلوب	المستوفى
المطلوبة	المرغوبة
المطلوبة	المنشودة
المطلوبة	المرادة
المطلوبة	المستوفة
المطهرة	المعقمة
المظاهر	المحيا
المظاهر	الطلعة
المعاجم	القواميس
المعادلة	المساوة
المعادلة	المتوازنة
المعارف	الطلاع
المعارف	الحاطة
المعارف	الوقوف
المعارف	الدرابة
المعاكس	العكسى
المعالج	الطبيب
المعالج	المطبب
المعالج	المداوي
المعالج	الشافى
المعالم	المنار
المعالم	الثوة
المعامل	المعالج
المعامل	المشغول
المعامل	المصنوع
المعامل	الصنيع
المعانى	المدلول

المعانى	المغزى
المعانى	الفحوى
المعانى	المودى
المعانى	مضمون
المعانى	محتوى
المعلومات	البيانات
المعلومات	المعطيات
المعمر	التفصيلى
المعمر	المفصل
المعمر	المستفيض
المعمر	المسهب
المعمر	الطويل
المعمل	المصنع
المعمل	المشغل
المعمل	المختبر
المعمم	المروج
المعممة	المروجة
المعمول	المفعول
المعمول	المصنوع
المعمول	المشغول
المعنى	المدلول
المعنى	المغزى
المعنى	المضمون
المعنى	المحتوى
المعنى	المقصود
المعنى	المعتمد
المعنى	المراد
المعنية	المقصودة
المعنية	المعتمدة
المعنية	المرادة
المعهودة	المعلومة
المعهودة	المعروفة
المغلوب	المنهزم
المغلوب	المهزوم
المغلوب	المقهور
المغلوب	المدحور
المغلوب	المكسور
المفتاح	المقلد
المفحوصة	المنتحنة
المفخمة	الجليلة
المفخمة	العظيمة
المفخمة	المهيبة
المفردات	الكلمات
المفردة	الكلمة
المفردة	القطة
المفروضة	المفترضة

المفصل	التفصيلى
المفصل	المستفيض
المفصل	المسهب
المفصل	المعمق
المفضلة	المختارة
المفلح	الناجح
المفلح	الموفق
المفلح	الفالح
المفلح	الضافر
المفلح	المضفر
المفهوم	المعقول
المفهوم	التصور
المفيدة	النافعة
المفيدة	الصالحة
المفيدة	المجدية
المفيدة	الناجعة
المقارنة	المقابلة
المقارنة	المقايسة
المقارنة	المفاضلة
المقام	المنزلة
المقام	المكانة
المقام	المكان
المقام	القدر
المقام	المركز
المقاومة	المناهضة
المقاومة	الصمود
المقاومة	الثبوت
المقاومة	المعارضة
المقترح	المعروض
المقترح	الرأي
المقدار	الكمية
المقدار	القسط
المقدار	الحسب
المقصود	المتوخى
المقصود	المتعمد
المقصود	المفتعل
المقصود	المراد
المقياس	المعيار
المقياس	المكيال
المقياس	المحك
المقياس	المقاس
المقياس	المدى
المقيدة	المربوطة
المقيدة	الموثقة
المقيدة	المشدودة
المكاتب	المراسل

المكتب	الديوان
المكتوب	المحرر
المكتوب	المدون
المكتوبة	المحررة
المكتوبة	المدونة
المكفوف	الضرير
المكفوف	اللاعمى
المكون	المعمول
المكون	المؤلف
المكون	المتالف
المكون	المخلوق
الملفوظ	المنطوق
الملك	السلطان
المنازعات	المخاصمات
المنازل	المساكن
المنازل	البيوت
المنازل	المثوى
المناسب	الملائم
المناسب	الموافق
المناسب	الانق
المناسبة	الملائمة
المناسبة	الموافقة
المناسبة	اللائقة
المناظرة	المناقشة
المناظرة	الحوار
المناظرة	المحاورة
المنافع	الفائدة
المناقشات	المحاورة
المناقشات	المناظرة
المناقشات	التحاور
المناهج	الطريقة
المناهج	الوتيرة
المناهج	النظام
المنبثقة	الصادرة
المنبثقة	المتصاعدة
المنتج	الصانع
المنتجة	الصانعة
المنتشرة	المشهورة
المنتشرة	المشاع
المنتشرة	المتداول
المنجى	المنقذ
المنجى	المخلص
المنزل	المسكن
المنزل	البيت
المنزل	الدارة
المنزل	المثوى

المنزل	المقر
المنزل	المستقر
المنسوخ	اللاغى
المنسوخ	الملغى
المنسوخ	المنقوض
المنسوخ	المنقول
المنسوخة	اللاغية
المنسوخة	المنقولة
المنشود	المرغوب
المنشود	المبتغى
المنشود	المطلوب
المنشود	المراد
المنشود	المتوخى
المنشودة	المرغوبة
المنشودة	المبتغة
المنشودة	المطلوبة
المنشودة	المرادة
المنشودة	المتوخة
المنطلق	المندفع
المنطلق	المسترسل
المنطوق	الملفوظ
المنطوق	النص
المنطوقة	الملفوظة
المنظور	المرئى
المنظور	المشاهد
المنظور	المشهود
المنظور	المرتقب
المنعزل	المنفرد
المنعزل	المفصول
المنعزل	المحبوس
المنعزل	المفروز
المنفذ	المنجز
المنفذ	المحقق
المنفصل	المتباعد
المنفصلة	المتباعدة
المنقوصة	الناقصة
المنقولة	المحمولة
المنقولة	المشحونة
المنقولة	المسرودة
المنهج	الطريقة
المنهج	النهج
المنهج	النسق
المنهجى	الطريقة
المنهجى	المنوال
المنهجى	النظام
المنورة	المزهرة

المنورة	المتفتحة
المنورة	المثقفة
المنورة	المشرقة
المنورة	المشعة
المهارة	البراعة
المهارة	الحدق
المهام	الوظيفة
المهام	العمل
المهام	الشغل
المهام	المامورية
المهام	اختصاص
المهمة	الوظيفة
المهمة	العمل
المهمة	المأمورية
المواد	الجوهر
المواد	الجسم
المواد	العنصر
المواد	المكون
المواد	المقوم
المواد	الفقرة
المواد	البند
الموارد	المأخذ
الموارد	المنهل
المواسم	العيد
المواضع	المكان
المواضع	المواقع
المواضع	المحل
المواضع	المطرح
المواضع	المقام
المواضع	الحيز
الموافق	القابل
الموافق	الراضى
الموافق	المطابق
الموافقة	القابلة
الموافقة	الراضية
الموافقة	المطابقة
المواقع	المكان
المواقع	الموضوع
المواقع	المطرح
المواقع	الجهة
المواقع	المقام
المواقف	الوضع
المواقف	الحالة
المواقف	الحال
الموزع	المشتت
الموزع	المفرق

الموزعة	المشتته
الموزعة	المفرقه
الموسعة	المكبره
الموضوعى	الواقعى
الموضوعى	الدقيق
الموضوعية	الدقه
الموظف	العامل
المولدة	المنتجه
الميادين	المجالأت
الميدان	الساحه
الميول	الانحياز
النابض	الحيوي
الناجح	المجتاز
الناجمة	الناتجه
الناحية	الاتجاه
الناحيتين	الاتجاهين
النسبية	الجزنيه
النسخ	التصوير
النسخة	الصوره
النشاط	الفاعلية
النشاطات	الفاعليات
النشر	الاعلان
النشرات	الاعلانات
الوجود	الكون
الولاء	اخلاص
الولاء	طاعة
الولاء	وفاء
تتجلى	تعرض
تتجلى	تظهر
تتجلى	تتكشف
تتجمع	تتحشد
تتجمع	تتجمهر
تتجمع	تتحزب
تتجمع	تتلاقى
تتجنب	تتفادى
تتجنب	تتحاشى
تتجنب	تتلافى
تتحقق	تتجسد
تتحقق	تتجسم
تتخبط	تتعثر
تتخبط	تتلعثم
تتخبط	تسقط
تتدرج	تتدرب
تتراوح	تتناوب
تتراوح	تتعاقب
تتراوح	تتوالى

تتراوح	تتبادل
تتسبب	تتقايض
تتسم	تتحمل
تتشارك	تضم
تتشارك	تندمج
تتضمن	تحتوى
تتضمن	تضم
تتضمن	تشمل
تتطرق	تعالج
تتطرق	تبحث
تتطرق	تداوي
تتطرق	تعامل
تتطلب	تأمر
تتطلب	تستلزم
تتطلبه	تأمره
تتطلبه	ترفضه
تتطلبه	تحتمه
تتطلبه	يستلزمه
تتطلبها	تامرها
تتطلع	تأمل
تتطلع	تمنى
تتطلع	ترجى
تتعدى	تتجاوز
تتعدى	تتخطى
تتعدى	تخرق
تتعرض	تعقد
تتعرض	تتعاقد
تتعرض	تتعاهد
تتعلق	تشق
تتعلق	تتهدل
تتعلق	تتدلى
تتغير	تتبدل
تتغير	تتحول
تتغير	تتعدل
تتفرع	تتشعب
تتفق	تنسجم
تتفق	تتخالص
تتفق	تتهايأ
تتفوق	تجتاز
تتكون	تتشكل
تتكون	تتصاغ
تتكون	تتصور
تتلخص	توجز
تتلخص	تحمل
تتمتع	تتظهر
تتمتع	تتجلى

تتمثل	تتوضح
تتمثل	تتبدى
تتمكن	تستطيع
تتمكن	تتحكم
تتناول	تتخذ
تتناول	تتقبل
تتناولها	تتخذها
تتناولها	تتقبلها
تتوافق	تتطابق
تتوفر	تتكاثر
تتوقف	تمتنع
تتولد	تتكون
تتولى	تخيل
تثير	تهيج
تثير	تحرك
تجارب	بروفات
تجارب	محاولة
تجاه	قبالة
تجاه	مقابل
تجاه	حيال
تجاه	ازاء
تجاهل	تناكر
تجاهل	أهمل
تجاهل	نبذ
تجاوز	عبر
تجاوز	مضى
تجاوز	فات
تجاوز	ولى
تجاوز	سلف
تجاوز	دبر
تجرؤ	جازف
تجرؤ	خاطر
تجرؤ	غامر
تجرؤ	جسر
تجرؤ	تجاسر
تجربة	بروفة
تجربة	محاولة
تجريبى	تجربة
تجريبى	اختبار
تجريبية	اختبارية
تجعل	تصير
تجعل	تخلق
تجعل	تفطر
تجعل	تصور
تجعل	تكون
تجنب	تفادى

BIBLIOGRAPHY

Abu-Arafah, A. 1995. A Grammar for the Arabic Language Suitable for Machine Parsing and Automatic Text Generation. Ph.D. Dissertation, Computer Science Department, Illinois Institute of Technology, Chicago, IL.

Abu-Salem, H. 1992. A Microcomputer Based Arabic Bibliographic Information Relational System with Relational Thesauri (Arabic-IRS). Ph.D. Dissertation, Computer Science Department, Illinois Institute of Technology, Chicago, IL.

Ahlswede, T. E. 1981. A Linguistic String Grammar for Adjective Definitions from Webester's Seventh Collegiate Dictionary. M.S. thesis, Computer Science Department, Illinois Institute of Technology, Chicago, IL.

Ahlswede, T. E. 1988. Syntactic and Semantic Analysis of Definitions in Machine Readable Dictionary. Ph.D. Dissertation, Computer Science Department Illinois Institute of Technology, Chicago, IL.

Ahlswede, T. E. and Evens, M. W. 1988a. Generating a Relational Lexicon from a Machine-Readable Dictionary. International Journal of Lexicography, Vol.1, No. 3, pp. 214-237.

Ahlswede, T. E. and Evens, M. W. 1988b. Parsing vs. Text-Processing in the Analysis of Dictionary Definitions. Proceedings of the Association for Computational Linguistics: 26th Annual Meeting. June 7-10. Buffalo, NY: ACL. pp.217-224.

Akkawi, K. 1998. The Effectiveness and Efficiency of Clustering in Arabic Information Retrieval. Ph.D. Dissertation, Computer Science, Illinois Institute of Technology, Chicago, IL.

Ali, N. 1988a. Arabic Language and Computer (research study). Al-khat Publishing Press, Ta' reep, Cairo, Egypt.

Ali, N. 1988b. Formalization and Computation of Arabic Syntax. 11th National Conference Proceedings, KFUPM, Dhahran, Saudi Arabia, Vol. I, pp. 309-320.

Ali, N., and Morcos, I. 1987. Data Compression Technique for Arabic Text. Proceedings of the First King Saud University Symposium on Computer Arabization, April, pp. 67-77.

Ali, N., Hegazi, N., and Abed, E. 1984. A Morphology-Based Data Compression Technique for Arabic Text. 1st African Conference on Computer Communications, Tunis 21-23 May, pp.241-251.

Ali, S. and Shapiro, S.C. 1983. Natural Language Processing Using a Propositional Semantic Network with Structured Variables. Technical Report 93-16. Buffalo, NY: Department of Computer Science, State University of New York at Buffalo.

Al-Kharashi, I. A., 1990a. Micro-AIRS, A Microcomputer Based Arabic Information Retrieval System, Design, Implementation and Evaluation, The 12th National Computer Conference, King Saud University, Riyadh, Saudi Arabia, October 21-24, Vol.II, pp. 515-529.

Al-Kharashi, I. A., 1990b. An Efficient Contextual Analysis Algorithm for Arabic Text Handling. The 12th National Computer Conference, King Saud University, Riyadh, Saudi Arabia, October 21-24, Vol.II, pp. 465-473.

Al-Kharashi, I. 1991. Microcomputer Based Arabic Information Retrieval System: Comparing Words, Stems, and Roots as Index Terms. Ph.D. Dissertation, Computer Science Department, Illinois Institute of Technology, Chicago, Illinois.

Al-Kharashi, I., and Evens, M. 1994. Comparing Words, Stems, and Roots as Index Terms in an Arabic Information Retrieval System. Journal of American Society for Information Science. Vol. 45, No. 8, September 1994, pp. 548 - 560.

Al-Khrisat, M. M. 1992a. Structuring the Arabic Lexicon and Thesaurus with Lexical-Semantic Relations. Proceedings of the Fourth Midwest Artificial Intelligence and Cognitive Science Society Conference, Utica, IL, May 3-4, 1992, pp.117-121.

Al-Khrisat, M. M. 1992b. Structuring the Arabic Lexicon and Thesaurus with Lexical-Semantic Relations to Support Information Retrieval. Ph.D. Dissertation, Computer Science Department, Illinois Institute of Technology, Chicago, Illinois.

Allen, J. 1987. Natural Language Understanding. The Benjamin/ Cummins Publishing Company, Inc., Menlo Park, CA.

Alsamara, K. 1996. An Arabic Lexicon to Support Information Retrieval, Text Generation, and Parsing. Ph.D. Dissertation, Computer Science Department, Illinois Institute of Technology, Chicago, IL.

Al-Shalabi, R. 1996. Design and Implementation of an Arabic Morphological System to Support Natural Language Processing. Ph.D. Dissertation, Computer Science Department, Illinois Institute of Technology, Chicago, IL.

Amsler, R. 1981. A Taxonomy for English Nouns and Verbs, Proceedings of the 19th Annual Meeting of the Association for Computational Linguistics, Stanford University, Stanford, CA, pp. 133-138.

Apresyan, Yu. D., Mel'cuk, I.A., and Zolkovsky, A. K. 1970. Semantics and Lexicography: Towards a New Type of Unilingual Dictionary. In F. Kiefer, (ed.) Studies in Syntax and Semantics. Reidel, Dordrecht, Holland, pp. 1-33.

Arif, E. 1995 Designing an Arabic Object-Oriented Programming Language and Help System for Pedagogical Purposes. Ph.D. Dissertation, Computer Science Department, Illinois Institute of Technology, Chicago, IL.

Armouti, H. 1995. An Intelligent Editor for an Arabic Object-Oriented Programming Language. Ph.D. Dissertation, Computer Science Department, Illinois Institute of Technology, Chicago, IL.

Attar, R., Choueka, Y., Dershowitz, N. and Fraenkel, A. 1987. KEDMA - Linguistic Tools for Retrieval Systems. JACM, Vol. 25, No. 1, January. pp. 52 - 66.

Becker, J. D. 1975. The Phrasal Lexicon. In R.C. Schank, and B. Nash-Webber, eds., Theoretical Issues in Natural Language Processing, ACL Annual Meeting, Cambridge, MA, June 1975. pp. 38 - 41.

Becker, J. D. 1987. Arabic Word Processing, Communications of the ACM. Vol. 30, No. 7, July 1987. pp. 600 - 610.

Beesley, K. 1988. Computer Analysis of Arabic Morphology: A Two-Level Approach with Detours. Perspectives on Arabic Linguistics III, ed. by Mushira Eid and Bernard Comrie, John Benjamins, Amsterdam. pp. 155 – 172.

Bloomfeld, L. 1933/1965. Language. Holt, Rinehart and Winston, Inc., New York.

Boguraev, B. 1987. Machine-Readable Dictionaries and Computational Linguistic Research. In D. Walker, A. Zampolli, and N. Calzolari, eds. Proceedings of the Lexicon Workshop of the 1987 Linguistics Summer Institute, July 13 – 24; Stanford CA. pp. 279 - 303.

Boguraev, B. and Briscoe, T. 1987. Large Lexicons for Natural Language Processing: Utilizing the Grammar Coding System of LDOCE. Journal of Computational Linguistics, Vol. 13, Numbers 3 - 4, July - December 1987, pp. 219-240.

Boguraev, B., Carroll, J., Pulman, S., Russell, G., Ritchie, G., and Black, A., 1987. The Lexical Component of a Natural Language Toolkit. In D. Walker, A. Zampolli, and N. Calzolari, eds. Proceedings of the Lexicon Workshop of the 1987 Linguistics Summer Institute, July 13 - 24; Stanford, CA. pp. 265 - 372.

Byrd, R. J. 1987. Dictionary Systems for Office Practice. In D. Walker, A. Zampolli, N. Calzolari, eds. Proceedings of the Lexicon Workshop of the 1987 Linguistics Summer Institute, July 13 - 24; Stanford, CA. pp. 186 - 205.

Byrd, R. J., Calzolari, N., Chodorow, M. S., Klavans, J. L., Neff, M. S., and Rizk, O. A. 1987. Tools and Methods for Computational Lexicology. Journal of Computational Linguistics, Vol. 13, Nos. 3 - 4, July - December1987, pp. 219 - 240.

Calzolari, N. 1988. The Dictionary and the Thesaurus Can Be Combined. In Evens, M., ed. Relational Models of the Lexicon, Cambridge University Press, Cambridge. pp.75-96.

Casagrande, J. and Hale, K. 1967. Semantic Relations in Papago Folk Definitions. In D. Hymes and W. E. Bittle, eds., Studies in Southwestern Ethnolinguistics. Mouton, The Hague: pp. 165-196.

Conlon, S. Pin-Ngern, Evens, M., Ahlswede, T. 1990. Generating a Lexical Database for Adverbs. Proceedings of the University of Waterloo Centre for the New Oxford English Dictionary. October 28-80, pp. 95-109.

Conlon, S., and Evens, M. 1994. A Lexical Database for Nouns to Support Parsing Generation, and Information Retrieval. In S. Hockey and N. Ide, eds. Research in Humanities Computing 3, Oxford University Press, Oxford pp. 74-78.

Conlon, S., Dardaine, J., D, Souza, A., Evens, M., Haynes, S., Kim, J.S., and Strutz, R. 1994. The Lexical Database: Dream and Reality. In Current Issues in Computational Linguistics: In Honour of Don Walker. A. Zampolli, N. Calzolari, and M. Palmer, eds. Linguistica Computazionale, Vol. IX. Giardini Editori, Pisa, distributed in the US by Kluwer Academic Publishers, Norwell, MA, pp. 201-225.

Cumming, Susanna, 1986a. Design of the Master Lexicon. Report to the National Science Foundation. In xerograph.

Cumming, Susanna, 1986b. The Lexicon in Text Generation. Presented at the Workshop on Automating the Lexicon, Grosseto, Italy, June, 1986.

Cumming, Susanna, and Robert Albano, 1986. A Guide to Lexical Acquisition in the JANUS System. Report to the National Science Foundation. In xerograph.

El-Dessouki, O., and Munir, M. 1987. A Dictionary Based Arabic Text Compression Technique. Proceedings of the First King Saud University Symposium on Computer Arabization, April, Riyadh, Saudi Arabia, pp. 79 - 87.

El-Dessouki, A., Nazif, A., El-Dessouki, O., and Ahmad, M. 1988. An Expert System for Understanding Arabic Sentences. Proceedings of the 10th National Computer Conference, Vol. II, Feb, 28 - March, 2, 1988. Computer Center, King Abdulaziz University, Jeddah, Saudi Arabia, pp. 745 - 759.

El-Dessouki, A., El-Dessouki, O., Nazif, A., and Ahmad, M. 1989. An ATN Approach for Understanding Arab Sentences. Proceedings of the 11th National Computer Conference and Exhibition. 4-7 March, 1989, Dhahran, Saudi Arabia, pp. 762-773.

El-Kareh, S., and Cote, P. 1990. Development of Computer Tools to Analyze Arabic Questions. The 12th National Computer Conference and Exhibition Proceedings, King Saud University, Riyadh, Saudi Arabia, October 21 - 24, 1990, Vol. II, pp. 693 - 705.

El-Sadany, T. A., and Hashish, M. A. 1989. An Arabic Morphological System. IBM Systems Journal. Vol. 28, No. 4, pp. 600 - 612.

Evans, D., and Zhai, C. 1996. Noun-Phrase Analysis in Unrestricted Text for Information Retrieval. Proceedings of the 34th Annual Meeting of the ACI. June 24-27, 1996. Santa Cruz, CA, pp. 17-24.

Evens, M. W. 1979. Structuring the Lexicon and the Thesaurus with Lexical-Semantic Relations. Proposal to the National Science Foundation. In xerograph.

Evens, M. W. 1981. Structuring the Lexicon and the Thesaurus with Lexical-Semantic Relations. Final Report to the National Science Foundation on Grant IST-79-18467. In xerograph.

Evens, M., ed. 1988. Relational Models of the Lexicon. Cambridge, UK: University Press.

Evens, M. 1993. Lexical and Semantic Relations for an Information Retrieval Thesaurus. Proceedings of the 56th Annual Meeting of American Society of Information Science, Vol. 30, October 24 - 28, 1993, p. 294.

Evens, M. and Smith, R. 1978. A Lexicon for a Computer Question-Answering System. American Journal of Computational Linguistics, Microfiches 81, pp.16-24, and 83, pp.1-98.

Evens, M., Litowitz, B., Markowitz, J., Smith, R. N., and Werner, O. 1983. Lexical-Semantic Relations: A Comparative Survey. Linguistic Research, Inc., Edmonton, Alberta.

Evens, M., Vandendorpe, J., and Wang, Y. 1985. Lexical-Semantic Relations in Information Retrieval. In S. Williams, ed., Humans and Machines. Ablex, Norwood, NJ. pp. 73-100.

Evens, M., Markowitz, J., Ahlswede, T., and Rossi, K. 1987. Digging in the Dictionary: Building a Relational Lexicon to Support Natural Language Processing Applications. Issues and Developments in English and Applied Linguistics, Vol. 2, pp. 33-44.

Evens, M., Dardaine, J., Huang, Y-F., Li, S.M., Markowitz, J., Rinaldo, F., Rinado, M., and Strutz, R. 1991. For the Lexicon That Has Everything. Proceedings of the Siglex Workshop, Berkeley, CA. pp. 179-187.

Farghaly, A. 1989. A Natural Language Understanding System for Arabic. Proceedings of the First Kuwait Computer Conference, Kuwait, March, 1989, pp. 595-622.

Fassi, F. A. 1993. Issues in the Structure of Arabic Clauses and Words. Kluwer Academic Publishers, Dordrecht, Holland.

Feddag, A. 1992. Arabic Morpho-Syntax and Semantic Parsing. Proceedings of the 13th National Computer Conference, King AbdulAziz City of Science and Technology, Riyadh, Saudi Arabia, November 11 - 14, Vol. 2, pp. 717 - 749.

Felber, H. 1986. Terminology and Machine Assisted Translation. Infoterm, Vienna, Austria. Proceedings of the International Workshop on Computer Aided Translation. Organized by the King Abdulaziz City for Science and Technology in Riyadh, 25 -26 Jumada II, 1405/ 16 - 17 March, 1985, pp. 117- 122.

Fox, E. 1981. Lexical Relations: Enhancing the Effectiveness of Information Retrieval Systems. ACM SIGIR Forum. pp.15-36.

Fox, E., Nutter, J. T., Ahlswede, T., Evens, M., and Markowitz, J. 1988. Building a Large Thesaurus for Information Retrieval, Proc. ACL Conference on Applied Natural Language Processing, Austin, Texas, Feb, pp. 101-108.

Gheith, M., and Aboul-Ela, M. 1989. A Computer Based Arabic Syntax Analyzer. Proceedings of the 11th National Computer Conference and Exhibition, March, Dhahran, Saudi Arabia, pp. 352 - 360.

Gheith, M., and El-Sadany, T. 1987. Arabic Morphological Analyzer on a Personal Computer. Proceedings of the First King Saud University Symposium on Computer Arabization, April, pp. 55 - 65.

Goodman, N. 1952. On Likeness of Meaning. Semantics and the Philosophy of Language. In L. Linsky, ed.: University of Illinois Press, Urbana, pp. 67-74.

Hammouri, A. 1994. An Arabic Lexical Database to Support Natural Language Processing. Ph.D. Dissertation, Department of Computer Science, Illinois Institute of Technology, Chicago, IL.

Hanash, M. 1989. The Electronic Dictionary for Arabic Language. Proceedings of the First Kuwait Computer Conference, Kuwait, March 1989, pp. 173 - 183.

Hegazi, N., and Elsharkawi, A. A. 1986. Natural Arabic Language Processing, Proceedings of the 9th National Computer Conference and Exhibition, Riyadh, Saudi Arabia, pp. 1 - 17.

Hmeidi, I. I. 1995. Design and Implementation of Automatic Word and Phrase Indexing for Information Retrieval with Arabic Documents. Ph.D. Dissertation, Department of Computer Science, Illinois Institute of Technology. Chicago, IL.

Hmeidi, I., Kanaan, G. and Evens, M. 1997.Design and Implementation of Automatic Indexing for Information Retrieval with Arabic Documents. Journal of the American Society for Information Science, Vol. 48. No. 10, pp. 867-881.

Huang, S - F. 1975. A Study of Adverbs. Mouton, the Hague.

Ingria, Robert, 1987. Survey of Verb Complement Types in English. Presented at the Workshop on the Lexicon, Stanford University, July, 1987.

Iris, M., Litowitz, B., and Evens, M. 1988. Problems of the Part-Whole Relation in the Lexicon: An Investigation of Semantic Primitives. In M. Evens, ed., Relational Models of the Lexicon. Cambridge University Press, Cambridge, England: pp.261-288.

Jambi, K. M. 1991. Design and Implementation of a System for Recognizing Arabic Hand Written Words with Learning Ability. Ph.D. Dissertation, Department of Computer Science, Illinois Institute of Technology. Chicago, IL.

Kanaan, G. 1997. Comparing Automatic Statistical and Syntactic Phrase Indexing for Arabic Information Retrieval. Unpublished Ph.D. Dissertation, Department of Computer Science, Illinois Institute of Technology. Chicago, IL.

Khayat, M. G. and Al-Muhtaseb, H. A. 1988. Knowledge Representation in Natural Language Systems. Proceedings of the 10th National Computer Conference, Vol. II March 28-2; Computer Center, King Abdulaziz University, Jeddah, Saudi Arabia. pp. 667 - 676.

Lee, W. J., and Evens, M. 1992. Generating Cohesive Text Using Lexical Functions. Proceedings of the International Workshop on Meaning-Text Theory. Darmstadt, Germany: Springer-Verlag. pp. 207-214.

Light, Marc. 1996. Morphological Cues for Lexical Semantics. Proceedings of the Association for Computational Linguistics: 34th Annual Meeting. June 24-27. Santa Cruz, CA: ACL. pp.25-31.

Litowitz, B., Evens, M., and Iris, M. 1988. Problems of the Part-Whole Relation, In M. Evens, ed., Relational Models of the Lexicon, Cambridge University Press, Cambridge. pp.261-288.

Lyons, J. 1968. Introduction to Theoretical Linguistics. Cambridge University Press, Cambridge.

Lyons, J. 1977. Semantics. Vols. I, II. Cambridge University Press, Cambridge.

Markowitz, J., Nutter, J. T., and Evens, M., 1992. Beyond Is-A and Part-Whole: More Semantic Network Links. Computers and Mathematics with Applications, Vol. 23, Nos. 6-9, 1992. pp. 337-390.

Mehdi, S. A. 1986. Arabic Language Parser, Int. J. Man-Machine Studies, Vol. 25, No. 5, pp. 592 - 611.

Mel'cuk, I. and Zholkovsky, A. 1988. The explanatory combinatorial dictionary. In M. Evens, ed. Relational Models of the Lexicon, Cambridge University Press, Cambridge: pp. 41-74.

Merriam. 1978. The Merriam-Webster Thesaurus. G&C Merriam Company. Springfield, MA.

Miller, G., Beckwit, R., Fellbaum, C., Gross, D., and Miller, K. 1993. A Revised Version of Introduction to WordNet: An On-line Lexical Database. Interalnonal Journal of Lexicography, Vol. 4, No.1. , pp.1 - 85.

Mohammed, F. A., Nasser, K., and Harb, H. M. 1993. A Knowledge Based Arabic Question Answering System. SIGART Bulletin, Vol. 4, No. 4, October 1993, pp. 21 - 33.

Nafitha 1988. User's Guide and Advanced Programmers Reference, release 3.0, Manama, Bahrain: 01 system.

Ng, T. H., and Lee, H. 1996. Integrating Multiple Knowledge Sources to Disambiguate Word Sense: An Exemplar-Based Approach. Proceedings of the Association for Computational Linguistics: 34th Annual Meeting. 24-27. Santa Cruz, CA: ACL. pp.17-24.

Nutter, J. T., Fox, E. A., and Evens, M. W. 1990. Building a Lexicon from a Machine-Readable Dictionary for Improved Information Retrieval. Literary and Linguistic Computing, Vol. 2, No. 5, pp.1 - 17.

Omar, A. 1982. Science of Semantics. Dare Aloroba lenashr waltawzee. Kuwait.

Owens, J. 1988. An Introduction to Medieval Arabic Grammatical Theory. The Foundations of Grammar. John Benjamins, Amsterdam.

Pin-Ngern, S. 1990. A Lexicon Database for English to Support Information Retrieval, Parsing, and Text Generation, Ph.D. Dissertation, Computer Science Department, Illinois Institute of Technology, Chicago, IL.

Pin-Ngern, S., Strutz, R., and M. Evens. 1991. Lexical Acquisition for Lexical Databases. Proceedings of the First Great Lakes Computer Science Conference, Kalamazoo, October, 1989. Published by Springer, New York, 1991. pp. 264-270.

Quillian, M. R., 1968. Semantic Memory. In M. Minsky, ed. Semantic Information Processing, MIT Press, Cambridge, MA: pp. 227-271.

Raphael, B. 1968. SIR: A Computer Program for Semantic Information Retrieval. In M. Minsky, ed., Semantic Information Processing, MIT Press, Cambridge, MA: pp.33-146.

Riegel, K. 1970. The Language Acquisition Process: a Reinterpretation of Related Research Findings. In L.R. Goulet and P.B. Baltes, eds., Theory and Research in Life-Span Developmental Psychology. Academic Press, New York: pp.357-399.

Ritchie, G., Black, A., Russell, G., and Pulman, S. 1992. Computational Morphology., MIT Press, Cambridge, MA.

Rumelhart, D., Lindsay, P. and Norman. D. 1972. A Process Model for Long Term Memory. In E. Tulving, W. Donaldson and G. Bower, eds. Organization of Memory. Academic Press, New York, NY: pp.198-248.

Sager, N. 1975. Computerized Discovery of Semantic Word Classes in Scientific Fields. In R. Grishman, ed. Directions in Artificial Intelligence: Natural Language Processing, Courant Institute of Mathematical Sciences, Computer Science Department, New York University, pp. 27- 48.

Saliba, B., and Al-Dannan, A. 1989. Automatic Morphological Analysis of Arabic: A Study of Content Word Analysis. Proceedings of the Kuwait Computer Conference, Kuwait, March, pp.3-5.

Salton, G. and McGill, M. J. 1983. Introduction to Modern Information Retrieval. McGraw-Hill, Inc. New York:

Sieny, M. 1985. Studies in Machine Translation, Proceedings of the International Workshop on Computer Aided Translation, Organized by the King Abdulaziz University for Science and Technology in Riyadh, 25-26 Jumada II, 1405/16-17 March, 1985. pp. 195-219.

Sinclair, J. ed. 1987. Collins Cobuild English Language Dictionary. Collins & Son Publishers, Glasgow, Scotland.

Slator, B. 1992. Sense and Preference, Computers and Mathematics with Applications, Vol. 23, Nos. 6-9, 1992. pp. 391-402.

Smith, R. N. 1985. Conceptual Primitives in the English Lexicon, Papers in Linguistics, Vol. 18, No. 1, pp. 99-137.

Sowa, J. F. 1988. Using a Lexicon of Canonical Graphs in a Semantic Interpreter. In M. Evens, ed. Relational Models of the Lexicon, Cambridge University Press, Cambridge: pp. 113-137.

Snapper, J. W. 1998. Associate Professor of Philosophy, Illinois Institute of Technology, Chicago, IL. Personal Communication.

Testen, D. 1993. On the Development of the Energic Suffixes. In M. Eid and C. Holes, eds. Perspectives on Arabic Linguistics V, John Benjamins, Amsterdam, pp. 293 - 311.

Viegas, E., Onyshkevych, B., Raskin, V., and Nirenburg, S. 1996. From Submit to Submitted via Submission: On Lexical Rules in Large-Scale Lexicon Acquisition. Proceedings of the 34th Annual Meeting of the ACL. June 24-27, 1996. Santa Cruz, CA, pp. 32-39.

Wahba, K. M., Emam, O. S., and Hashish, M. A. 1990. An Orthographic-to-Phonetic Transcription System for Arabic Language. The 12th National Computer Conference and Exhibition Conference Proceedings, October 21 - 24, 1990. King Saudi University, Riyadh, Saudi Arabia, Vol. II, pp. 439 - 448.

Wan, T-L. 1995. Experiments with Automatic Indexing and a Relational Thesaurus in a Chinese Information Retrieval System. Ph.D. Dissertation, Computer Science Department, Illinois Institute of Technology, Chicago, IL.

Wang, G. N. 1989. A Microcomputer Based Information Retrieval System with Relational Thesauri to Support a Medical Expert System. Ph.D. Dissertation Computer Science Department, Illinois Institute of Technology, Chicago, IL.

Wang, Y., Vandendorpe, J., and Evens, M. 1985. Relational Thesauri in Information Retrieval. Journal of the American Society of Information Science. Vol. 36, pp. 15 - 27.

Werner, O., and Schoepfle, G. M. 1987. Systematic Fieldwork. Sage Publication, Newbury Park, CA.

Werner, O., Austin, M., and Begishe, K. 1969/1981. The Anatomical Atlas of the Navajo. Department of Anthropology, Northwestern University and the U.S. Public Health Service.

Williams, G. 1966. Linguistic Reflections of Cultural Systems. Anthropological Linguistics, Vol. 8, No. 8, pp.13-21.

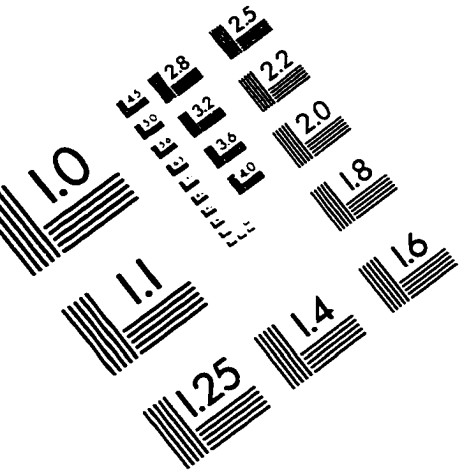
Wright, W. 1859. A Grammar of the Arabic Language. Vol. I. Covent Garden, London.

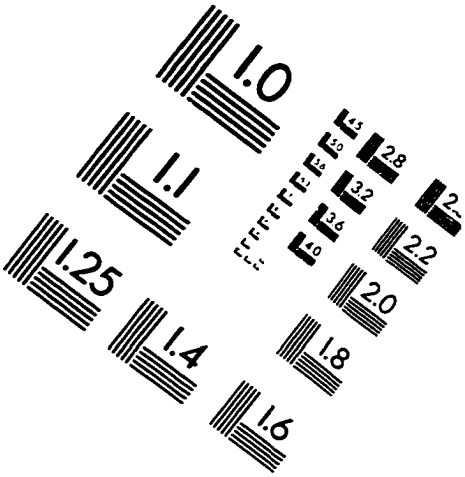
Yahya, A. H. 1989. On the Complexity of the Initial Stage of Arab Text Processing, Proceedings of the Great Lakes Computer Conference, Kalamazoo, MI, October 1989.

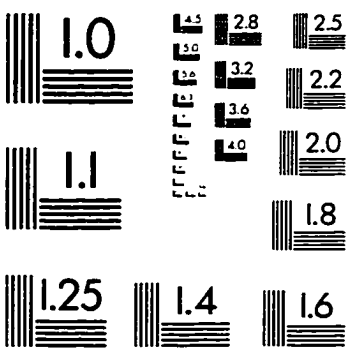
Zholkovskii, A. K. and I. A. Mel'chuk. 1967. Semantic Synthesis. Systems Theory Research – a Translation of Problemy Kibernetiki Vol. 19, pp. 179-243.

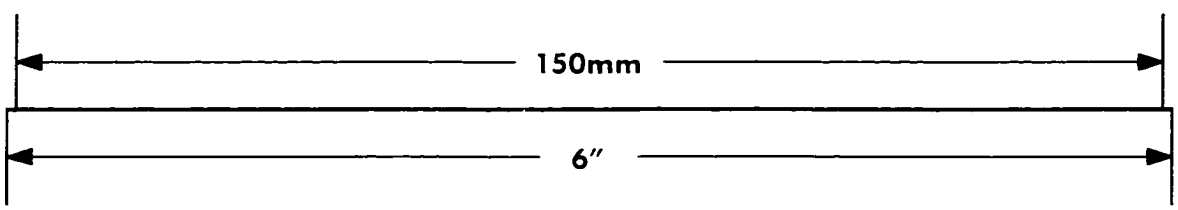
Ziadeh, F., and Winder, R. 1988. An Introduction to Modern Arabic. Princeton University Press, Princeton, NJ.

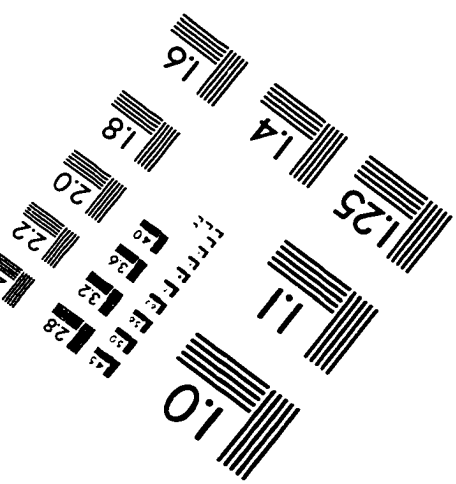
IMAGE EVALUATION TEST TARGET (QA-3)

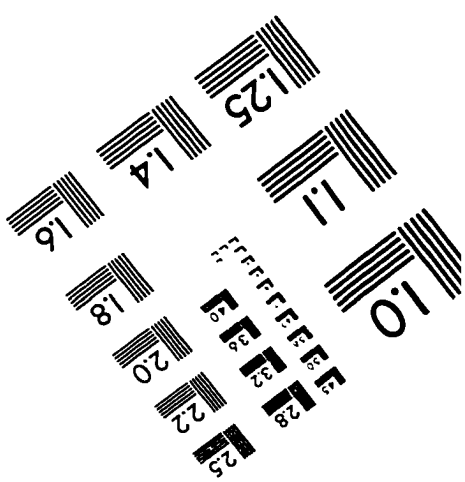












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989