Object Oriented Techniques for an Intelligent Multi-purpose English Language Dictionary System

Samia Yousif[1] and Mansoor Al-A'ali[2]

[1] Faculty of Art, Design and Computing Science, Royal University for Women
Kingdom of Bahrain

2 Department of Computer Science, College of Information Technology
University of Bahrain, Kingdom of Bahrain

Abstract. This research utilizes the features of the Object Orientation (OO) to develop TOOT, a dictionary system containing English words, their Arabic meanings, associated actions, semantic relationships, inherited actions and attributes, exceptional relationships and semantics as well as other characteristics. TOOT utilizes OO major notions such as objects, classes, aggregation, inheritances, encapsulation and polymorphism. Each word in this dictionary system belongs to a class and may have one or more subclasses. Subclasses inherit all the public attributes and operations of their super class and this concept is utilized in all types of processing on the TOOT dictionary system. TOOT is a knowledge base and can be thought of as an intelligent language model which can be used for many purposes. This research shows how simple phrases can be generated or validated to be semantically correct. In the process of using OO UML to represent semantic knowledge, we have made enhancements and additions to UML itself.

1 Introduction

Research into natural language processing has received great attention and will continue to do so because of the importance of such topic to humanity [5], [2], [16]. Natural language processing presents two main challenges: representing semantics and querying these semantics. Until now, no one claims to have the solution to store semantics of complex sentences and link all its words together with the meaning of words, phrases of other sentences. The research we present in this paper presents a novel approach by storing the semantics of English simple phrases using the well known Object Oriented technique. Object oriented technique is a method of looking at knowledge at different levels of abstraction. We identified a simple way of storing simple semantics using object orientation and we demonstrate that this method can be used to teach simple semantics to young learners of the English language.

Students of the English language need to feel more of the knowledge relating to the words to improve their skills in English. Students studying English as second language (ESL) need to use dictionaries in order to know what words mean and what they can do with them in order to understand and communicate in English [8].

Yousif S. and Al-A'ali M. (2007).
Object Oriented Techniques for an Intelligent Multi-purpose English Language Dictionary System.
In *Proceedings of the 4th International Workshop on Natural Language Processing and Cognitive Science*, pages 159-168
DOI: 10.5220/0002423801590168

Moreover, Learners' Dictionaries are used to practice English correctly, effectively and appropriately [15]. Consequently, dictionaries must have maximum semantic information for words related to semantic classification that will assist the students to understand English language very well, besides, it will improve and enhance their English.

Writing fluently in English language requires knowledge of the conventional contexts, strong vocabulary, and collocations surrounding the word. While this information may be presented implicitly in dictionaries, many students do not have the dictionary savvy to extract it and they often have difficulties finding the right meaning for unfamiliar words or phrases in their dictionaries. In response to these difficulties, some researchers have proposed new ways to access the English lexicon, such as an electronic dictionary and presenting typical phraseology rather than words in isolation [8]. In addition, some researchers have suggested a representation, with a tree-based model of runtime dictionaries, which allows a very fast morphological analysis with semantic recognition [11], for example, the path <SPORTS/Water Sports/Swimming> describes the document field <Swimming> as a sub-field of <Water Sports>, which is a sub-field of super-field <SPORTS> [6].

Many researchers continue to study a variety of new approaches in teaching English [13], [9], [3], [12], [14], [18], [1] and [17]. Communicative language teaching (CLT) approach advocates the development of communicative competence as a primary goal through the extensive use of the second language as a means of communication during classroom lessons [7]. The use of technology in the classroom has increasingly been the object of study in recent years. Computers have been used for language teaching since the 1960s [4] and [10].

Some researchers applied the ideas from CAL - Computer Assisted Language and language learning, emphasizing context for language learning, especially for abstract word learning. [18] Developed a multimedia web-based CAL system which includes 13 abstract words and five main modules: Learning Material, Testing, Communication, Help, and Extensive English Learning Web Sites. Their research indicated that pupils who worked with the more communicative methods for teaching English made greater progress in vocabulary acquisition than those who were taught in the more traditional context.

Our research presents a model for representing some aspects of semantics for simple sentences or phrases by utilizing the OO concept, for example, the sentence (boy reads book) is semantically and syntactically correct, the simple sentence (boy eats book) is grammatically correct but semantically wrong, but the simple sentence (book reads boy) is syntactically correct and semantically wrong. In turn, TOOT dictionary system rejects semantically wrong phrases or simple sentences and displays the reasons as well provides a correct example such as (boy eats apple) to help the learner understand his/her shortcomings. This is made possible because of the intelligent features of the OO approach used to model the words and phrases semantically which is embedded in the form of associated actions, relationships and inheritance features with each word (noun).

In this research, administrators create relationships between actions (verbs) and classes (nouns) to provide important information about the meanings of nouns. For instance, the action ‘eat’ is the relationship between the classes ‘human’ and ‘food’ which means ‘human eat food’ and hence any object under food (subclass of food) would automatically be related with any object (subclass) under human. These relationships would be inherited downwards to any depth of the subclasses unless exceptional relationships are stated.

TOOT is a web-based dictionary system. Each word in this dictionary belongs to a class and may have one or more subclasses. Subclasses inherit all the public attributes and operations of their super class and this concept is utilized in all types of processing on the TOOT dictionary system. For example, the super class 'human' inherit the public operation 'eat' to its subclass 'boy' and therefore we can generate or check the simple sentence ‘boy eats food’ or 'boys eat food' with semantic meaning. This research signifies that TOOT dictionary system assists teachers to make the English subject more interesting and raise student's motivation as well as it enables students to understand the meaning of words and helps them to organize sentences and build them correctly with semantic meaning.

We have conducted a survey of online dictionaries and elicited that almost all dictionaries come under three categories: leaner, translator and search. Each dictionary has a number of features, for example some dictionaries provide words definition, examples, pronunciation, words translation, text translation, pictures, rhymes, synonyms, and words functions like noun, adjective, etc. Many companies have cashed in on the concept and have developed dictionaries for a variety of purposes such as: Cambridge Learner’s Dictionary, Word Central, Ultralingua Language Software, RhymeZone, Wordsmyth Children's Dictionary, McGraw-Hill Children's Dictionary, OneLook Dictionary Search, Freesearch, Sakhr, Tarjim, WordNet & ALWafi.

2 Applying the OO Techniques to Incorporate Semantics in TOOT

Object Orientation (OO) is a paradigm for creating software systems using objects. Objects are tangible and conceptual things we find in the real world. Using OO makes TOOT Dictionary Objects more semantically related and hence better intelligence can be incorporated and implemented. OO has major notions such as objects, classes, inheritances, encapsulation and polymorphism. The OO Unified Modeling Language (UML) is the industry-standard language for specifying, visualizing, constructing, and documenting the artifacts of software systems. It simplifies the complex process of software design, making a "blueprint" for construction. These OO concepts were used as a model for representing TOOT knowledge base rather than TOOT development.

2.1 Slightly Modified UML for TOOT Dictionary System

Fig. 1 shows the UML design of TOOT Dictionary system and is continued in fig. 2. Fig. 1 shows that the super class “Thing” has a public attribute “gender” and it inherits this attribute to its subclasses as well. It presents the value of “gender” for each class and object. The super class “Information Media” has public operations (Action On; the action that can be carried out on the second class or object: browsed, bought, read, sold and written). These operations are inherited to the subclass “Reading” and consequently to its subclasses (Book, Magazine and Newspaper). In addition, it demonstrates that the super class "Human" inherits the public operations (Action By; the action that the first class or object can do: browses, buys, cooks, cuts, eats, feeds, holds, reads, sells and writes) and (Action On: fed) also the operations done by the subject without an object "Subject-Only" operations (Action By: grows, sits, sleeps, swims and weeps) to its subclasses "Family" and "People". All these operations are inherited to the objects (Father, Mother, Son, Daughter, Man, Woman, Boy and Girl). Moreover, the private operation (Action By: engages) is added to the object (Man) and the private operation (Action On: engaged) is added to the objects (Woman) and we assumed that the objects (young: Boy and Girl) can't cook, then the inherited operation (Action By: cooks) is excluded from these objects which means that the public operation (cook) is an exceptional operation (Action By) for these objects.

The super class "Vegetarian Food" has public operations (Action On: eaten, cooked, bought, held, sold and cut) and "Subject-Only" operations (Action By: grows). It inherits them to its subclasses (Fruit, Grain and vegetable) as well to its instances (objects) as shown in fig. 1 and fig. 2. We suppose that the inherited public operation (Action On: cooked) cannot be carried out on the objects (Fig, Melon, Guava, Quince, Lettuce and Radish), therefore it is excluded from these objects. In turn, this public operation (cooked) is exceptional operation (Action On) for these objects.

As shown in fig. 1 and fig. 2, the public attributes and operations are expressed with a plus sign (+). The private attributes and operations are expressed with a minus sign (-). The exceptional operations are expressed with a less-than sign (<). The "Subject-Only" operations are expressed with an ampersand (&). The letter "b" is added to these signs (+, -, < or &) when the type of the operation is "Action By" while the letter "o" is added to them when the type of the operation is "Action On". Notice that both "Exceptional" and "Subject-Only" operations are not a part of UML class diagram, but we added them, which means that the signs of them are according to our view as well as the addition of the letters "b" and "o" to the signs. The signs and the letters are utilized to distinguish between the actions.

2.2 Enhancement UML

Fig. 1 and fig. 2 were utilized UML Class Diagram to show the inheritance and the aggregation of the classes of the TOOT dictionary system. However, the relationships between classes are not obvious in these figures and can't utilize UML Class Diagram to present them. Consequently we have attempted to enhance UML diagram to demonstrate inheritance, aggregation and relationships as shown in fig. 3. Fig. 3 presents Meta Diagram of enhancement UML. In addition, is easy to draw the relationships between the classes and more clear to understand it. Fig. 3 demonstrates two super classes with their subclasses, instances and relationships. The dotted lines mean that the diagram can be extended depending on the system. The first super class is called Super class A and it has Subclass Aa. This subclass contains three instances: Instance Aa1, Instance Aa2 and Instance Aa3. The second super class is called Super class B and it has Subclass Ba. This subclass consists of three instances: Instance Ba1, Instance Ba2 and Instance Ba3. The super classes inherit the public operations (Actions) to its subclasses. Accordingly, these subclasses inherit its public Actions to its subclasses or instances. The following are the relationship between classes as shown in fig. 3:

1. **Normal Association** (Public Operations): It is a connection between classes, a semantic connection between objects (instances) of the classes involved in the association. It is normally bidirectional, which means that if an object is associated with another object, both objects are aware of each other. First direction is Action By (+b Action), and second direction is Action On (+o Action). The multiplicity (0..*) and (1..*) are used to present how many objects are linked. This association is inherited to the subclasses.
2. **Recursive Association** (Public Operations): It is a connection between a class and itself. The association still presents a semantic connection between objects, but the connected objects are of the same class. This is also bidirectional and it has multiplicity like normal association. This association is inherited to the subclasses.
3. **Private Association** (Private Operations): It is a connection between objects (instances). It is bidirectional; First direction is Action By (-b Action), and second direction is Action On (-o Action). This association can't be inherited.
4. **Exceptional Association** (Exceptional Operations): It is a connection between some objects and classes. Exceptional Action By (<b Action) means that the inherited action that the object can't do and Exceptional Action On (<o Action) means that the inherited action can't be carried out on the object. Consequently these actions are excluded from this object.
5. **Subject-Only Operations**: It means that the action is done by the subject (class) without an object. It is always Action By (&b Action) as well as is inherited to the subclasses.

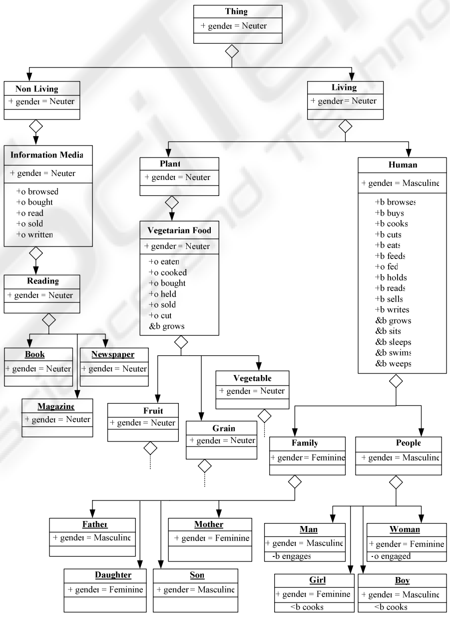

Fig. 1. UML of TOOT Dictionary System.

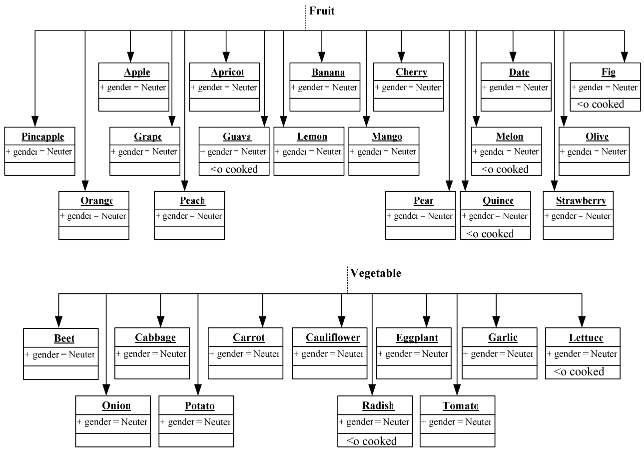

Fig. 2. Continuing UML of TOOT Dictionary System.

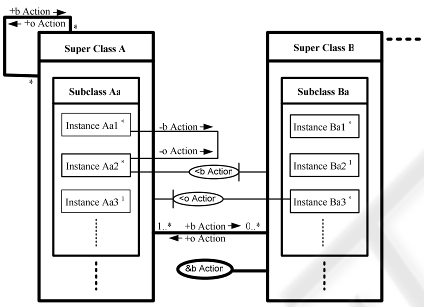

Fig. 3. Enhancement UML (Meta Diagram).

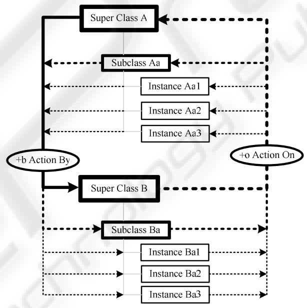

Fig. 4. Classes Relationships.

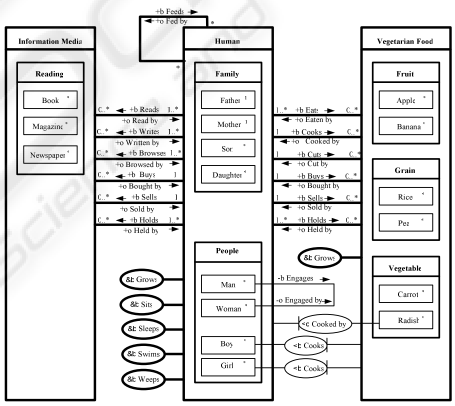

Fig. 5. A different way of representing OO for TOOT word-phrase language dictionary system.

In TOOT dictionary system, the relationships (Action By) are created between two super classes manually; subsequently the operations (Action On) are generated between them automatically. The normal line is sketched between two super classes (Super Class A and Super Class B) indicates the public relationship (Action By) which is created manually between them while the dotted lines are sketched between them and its classes signify the public relationship (Action On) which is generated automatically as well as it shows the inherence public operations (Action By and Action On) as illustrated in fig. 4. Notice that the sketched lines of relationships between classes in fig. 4 are hidden in TOOT dictionary system. The other relationships that were applied in TOOT dictionary system can be exhibited in the same manner that is presented in fig. 4. Fig. 5 shows a Meta Diagram of the enhancement UML to demonstrate the inheritance, aggregation and relationships of the classes in TOOT dictionary system.

3 TOOT Applications

This research utilized the Object Oriented approach to develop an Intelligent Object Oriented Dictionary System called TOOT to contain a number of useful and practical applications. TOOT dictionary system consists of five major educational applications: TOOT dictionary, Semantic Checker, Lesson, Quiz and History.

The classification of the word is represented in a high fashion way by a Tree structure. It provides a full path of the super classes and the subclasses in one line structure rather than separate fields. The tree structure makes it easier to perform forward and backward navigation to different levels. The tree structure tool is used for searching and displaying words.

When the student selects a word such as “Boy” as shown in fig. 6, TOOT dictionary displays the classification of the word “Boy” which is **Thing > Living > Human > People > Boy**. This helps the student understand and recognize the semantic meaning of the word “Boy”. TOOT dictionary also shows the associated diagram and images to provide further illustration on the word to expand the student knowledge. Associated diagrams include information on the word in Arabic and English. The information can be the plural of the word, gender, and description. Read is an action that Boy can do (Action By), and fed is the action that can be carried out on him (Action On). When an action is selected, TOOT dictionary system displays the Arabic translation, the corresponding picture and generates simple sentences using the selected word and the action. For example, “Boy reads” and “Boy is fed”.

As demonstrated in fig. 7, the dotted lines indicate that the object “Boy” inherited all public operations (actions) from its super class “Human”, (Action By: browses, buys, cuts, eats, feeds, holds, reads, sells and writes) and (Action On: fed). But we assume that the object (very young: Boy) can’t cook, then the inherited operation (Action By: cooks) is excluded from this object which means that the public operation (cook) is exceptional operation (Action By). Also it inherited “Subject only” operations (Action By: grows, sits, sleeps, swims and weeps).

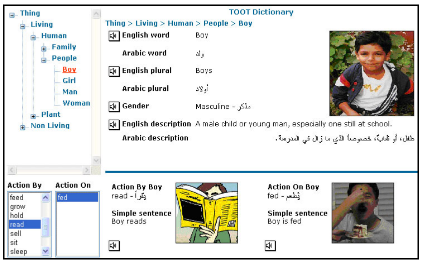

Fig. 6. TOOT Dictionary System – after selecting a word like "Boy".

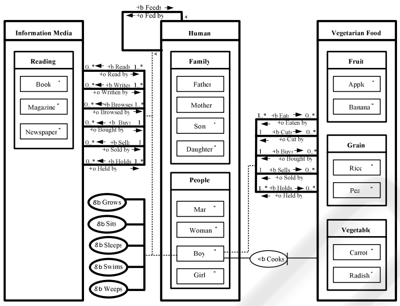

Fig. 7. The object "Boy" inherited all public actions from its Super Class "Human".

The Semantic Checker tool checks the entered simple phrases depending on their classes and its relationships that were created and stored previously. TOOT can check the syntax and semantics of simple phrases (type 1: Subject + Verb and type 2: Subject + Verb + Object) entered by the student. This is made possible because of the intelligent features of the OO technique and the actions associated with each word in TOOT dictionary system.

Checker tool represents the intelligent aspect of TOOT dictionary system such as its checking of the spelling of the entered words, listing the possible correct words of any misspelled words, accepting or rejecting the entered phrases depending on its semantic meaning, correcting the wrong phrases and generating correct phrases randomly. In addition, it provides the explanation for accepting or rejecting any phrase and determines the relationship type that is associated between the entered classes and action.

The following examples show the acceptance or rejection of the entered simple phrases depending on its semantics correctness after checking its syntax, correcting the wrong phrases and generating semantically correct phrases randomly.

Public Relationship. Consider the entered words: boy, eat and book. The simple phrase "boy eat book" is grammatically correct but semantically wrong. TOOT dictionary system rejects this phrase and displays the reasons depending on the classes, its relationships and inheritance. Fig. 5 and fig. 7 show that Boy inherits the public relationships from its super class Human, as well Book inherits the public relationships from its super class Information Media but human can't eat information media, in turn, boy can't eat information media while human can eat vegetarian food and consequently, boy can eat vegetarian food as shown in fig. 7. Hence, TOOT creates a possible class string which is human can eat vegetarian food, and selects an object from this class randomly such as Apple and therefore it corrects the semantics of the rejected phrase to become "Boy eats Apple".

Special Relationship. Consider the entered words: daughter, engage and father. TOOT dictionary system rejects this simple phrase "daughter engage father" because it is grammatically true but semantically wrong. Fig. 5 exposes that the action Engage is a special relationship between Man (Action By) and Woman (Action On) which means that Daughter can't engage Father but Man can engage Woman. Consequently, TOOT creates a possible class string and generates correct phrase "Man engages Woman".

Exceptional Relationship (Action By). Consider the entered words: boy, cook and rice. TOOT dictionary system rejects this simple phrase "boy cook rice" because Human can cook Vegetarian Food and the super class Human inherits the public relationship (Action By) Cook to its object Boy but we assume that the young Boy can not cook. In turn, Cook is an exceptional relationship (Action By) for Boy. Therefore TOOT modifies the entered phrase randomly like "Woman cooks Rice" with correct grammar and semantics.

Exceptional Relationship (Action On). Consider the entered words: woman, cook and radish. TOOT dictionary system rejects this simple phrase "woman cook radish" because Human can cook vegetarian Food and the super class Vegetarian Food inherits the public relationship (Action On) Cooked to its object Radish but we assume that this action can't be carried out on the Radish as illustrated in fig. 5. Hence TOOT amends the entered phrase randomly such as "Woman cooks Rice" with correct grammar and semantics.

Subject-Only Relationship. Consider the entered words: carrot, grow and pea. TOOT dictionary system rejects this simple phrase "carrot grow pea" and Fig. 5 demonstrates that Carrot is a Vegetarian Food and the action Grow is a Subject-Only relationship which means that the Vegetarian Food can grow and the object is not required. For these reasons, TOOT removes the object and the phrase become "Carrot Grows".

4 Conclusions

This paper presented a new approach for the development and evaluation of an intelligent English-Arabic Object Oriented dictionary system called TOOT. This research is the first to effectively use the OO concept and enhance UML to deal with English language semantics for simple phrases. TOOT dictionary system was tested on the 6^{th} class of an elementary girls school in kingdom of Bahrain to test its effectiveness in teaching some aspects of the English language and to test its level of contribution to students learning and the tests are very promising. The results of such testing are outside the scope of this paper.

References

1. Al-A'ali Mansoor: An Approach for Automatically Extracting Translation Rules For a Better English-to-Arabic Machine Translation, Journal of computer science, (2007), Science Publishers, USA
2. Al-A'ali Mansoor: An Efficient Technique for Identifying and Storing Syntactic and Semantic English Natural Language Sentence Patterns, WSEAS Transactions On Systems, Issue 1, Volume 6, January (2007), ISSN 1109-2777
3. Chuang, Y.: A brand new world for English instruction with interactive multimedia and the World Wide Web, SIGCUE Outlook, Volume 26, Issue 3, July (1998) 2-7
4. Cla´ udia B.M.J. Andrea V. Steilb, Jose´ L. Todescoc. Martinsa: Factors influencing the adoption of the Internet as a teaching tool at foreign language schools. Computers & Education Volume 42, Issue 4, May (2004) 353–374
5. Cowie J., Wilks Y. In R. Dale, H. Moisl and H. Somers (eds.): Handbook of Natural Language Processing. New York: Marcel Dekker (2000)
6. El-Sayed Atlam, K. Morita, M. Fuketa and Jun-ichi Aoe: A new method for selecting English field association terms of compound words and its knowledge representation, Information Processing & Management, Volume 38, Issue 6, November (2002) 807-821
7. Francis Mangubhai, Perce Marland, Ann Dashwood and Jeong-Bae Son.: Teaching a foreign language: one teacher's practical theory. Teaching and Teacher Education, Volume 20, Issue 3, April (2004) 291-311
8. Janice McAlpine and Johanne Myles.: Capturing phraseology in an online dictionary for advanced users of English as a second language: a response to user needs, System, Volume 31, Issue 1, March (2003) 71-84
9. Jaspers, J.G.M., Kanselaar, G.; Kok, W.A.M.: Productive language use with IT'S ENGLISH, Computers & Education, Volume 21, Issue 1-2, July-September (1993) 67-75
10. Kenneth Ruthven Sara Hennessy & Sue Brindley. T: Teacher representation of successful use of computer-based tools and resources in secondary-school English Mathematics and Science. Teaching and teacher education, Volume 20, Issue 3, April (2004) 259-275
11. Lovisa, C. R. Baudb, A. M. Rassinouxc, P. A. Michelb and J. R. Scherrer: Medical dictionaries for patient encoding systems: a methodology, Artificial Intelligence in Medicine, Volume 14, Issues 1-2, September (1998) 201-214
12. Mary Goretti Nakabugo, Rob Siebörgerb: Curriculum reform and teaching in South Africa: making a ‘paradigm shift', International Journal of Educational Development, Volume 21, Issue 1, January (2001) 53-60
13. O'Brien, P. eL: Using AI in CALL, Intelligent Tutoring Media, Volume 3, Issue 1. February (1992) 3-21
14. Shireen Motala.: Quality and indicators of quality in South African education: a critical appraisal, International Journal of Educational Development, Volume 21, (2001) 61-78
15. Stefania Nuccorini: Pragmatics in learners' dictionaries Journal of Pragmatics, Volume 19, Issue 3, March (1993) 215-237
16. Taher S.K. Homeed & Mansoor Al-A'ali, Analyzing Statistical and Syntactical English Text For Word prediction and text generation, journal of Information Technology, ANSI, (2007)
17. Teresa Chen.: Reticence in class and on-line: two ESL students' experiences with communicative language teaching, System, Volume 31, Issue 2, June)2003) 259-281
18. Wenli Tsou, Weichung Wang and Hung-yi Li: How computers facilitate English foreign language learners acquire English abstract words, Computers & Education, Volume 39, Issue 4, December (2002) 415-428
19. Yousef El-Halees.: The teaching-learning situation of English pronunciation in Jordan, System, Volume 14, Issue 1, (1986) 59-70