A Suite of Tools for Arabic Natural Language Processing: A UNL Approach

Sameh Alansary, Magdy Nagi, Noha Adly
Bibliotheca Alexandrina
Alexandria, Egypt.
sameh.alansary@bibalex.org, magdy.nagi@bibalex.org,
noha.adly@bibalex.org

Sameh Alansary
Department of Phonetics and Linguistics, Faculty of Arts
Alexandria University
Alexandria, Egypt.
sameh.alansary@bibalex.org

Magdy Nagi, Noha Adly
Department of Computer and System Engineering, Faculty of Engineering
Alexandria University
Alexandria, Egypt.
magdy.nagi@bibalex.org, noha.adly@bibalex.org

***Abstract*—This paper introduces the UNL framework as a collaborative framework that encourages and promotes the participation of linguists and non-linguists in the development of an integral natural language processing workbench. The UNL workbench includes a multitude of user-friendly back-end and front-end applications that facilitate the process of learning the UNL basics, participating in the development of resources within the UNL framework, as well as applications that perform several NLP tasks such as machine translation and editing. This workbench claims the ability to analyze automatically natural languages into their abstract semantic meanings, with the aim of finding the common denominator between all languages.**

Keywords—automatic semantic analysis; universal networking language; language resources; Arabic analysis grammar; Arabic generation grammar; UNL; Arabic natural language processing

I. INTRODUCTION

Although 5% of the world populations speak Arabic as a native language, the research on the Arabic language does not measure up to its usage. The Arabic language lags far behind other languages with respect to language resources; very few language resources are available for Arabic. Moreover, most of the resources available are not very efficient. Because many resources are limited to the same field of study. However, more research has been done on Arabic in the last decade due to many factors including the widespread use of communication and information technology applications. For example, the NEMLAR project drew a map of the existing Arabic resources, tools, technologies and actors in the Arabic speech and language processing fields, this led to the identification of the needs and the availability of resources and tools in the BLARK "Basic Language Resources Kit" [1],[2]. Also, two institutions: the ELRA and LDC distribute useful resources for the development of Arabic speech and language processing technologies and applications [3]. However, most of these attempts are individual and there are no joint efforts to build collaborative language resources to be available for linguistic research. In addition the available resources are mostly commercial.

Most of the Arabic resources available handle the Arabic language morphologically and some syntactically, but, up to our knowledge, none are devoted to analyzing Arabic sentences semantically. Some natural language processing systems can indeed analyze the underlying syntactic structure of a sentence but no system claims the ability to understand Arabic sentences and represent them semantically. There are many acknowledged attempts for the semantic analysis of natural language such as the FrameNet1 [4] and the Propbank [5]. These approaches aimed to build semantically analyzed corpora but they did not mount up to be a collaborative framework for semantic analysis.

This paper presents a detailed description of a framework for natural language processing in general and semantic analysis in particular in a manner that guarantees the existence of a cooperative framework. This framework contains an educational subdivision, a technical subdivision, and a practical subdivision. This framework is the universal networking language. The paper also presents some of the tools that created using this environment and how these support linguistic research in general and semantic analysis in particular. In this regard, we present some models of Arabic morphological and syntactic analysis in a way that leads to the semantic analysis of Arabic within this aforementioned framework demonstrated with examples.

This paper presents the recent developments in the UNL system. The authors published many other works discussing the research and development of the UNL from its early beginning in 1996 till 2012. In addition, a number of these publications discussed and evaluated the use of UNL in machine translation to translate 1000 pages from the Encyclopedia of Life support Systems (EOLSS) into the six official languages of the UNESCO, and comparing the results of UNL with other interlinguas [20] and [21].

Section 2 of this paper, introduces the UNL system as a system that claims the ability to understand natural language sentences, analyze them morphologically, syntactically as well as semantically. Section 3 will introduce the online portal to the UNL system; the UNL web with all its component platforms and environments. Finally, section 4 will examine the Arabic component in the UNL system.

II. UINVERSAL NETWORKING LANGUAGE

The Universal Networking Language (UNL) is a formalism for representing and expressing the knowledge conveyed by natural language utterances. It is an artificial language created in order to mediate between the natural languages of the world, as it is also an intermediary language between natural and artificial (programming and markup) languages.

[1] https://framenet.icsi.berkeley.edu/fndrupal/

978-1-4673-2821-0/13/$31.00 ©2013 IEEE

The UNL is not intended to function the same as other man-made interlinguas (e.g. Esperanto), rather, it can be considered as a sort of encoding of the meaning of natural language sentences that only computers can understand [6], and [7]. The UNL depends on the semantic analysis of words and phrases in order to reach the meaning and encode it in a universal format.

The UNL program was launched in 1996 as a communication protocol, originally proposed by the Institution of Advanced Studies of the United Nations University, Tokyo, Japan under the auspices of the UNESCO to be a resource for developing a multilingual platform for information exchange. In January 2001, the United Nations University set up an autonomous non-profit organization in Geneva, Switzerland to be responsible for the development and management of the UNL Program; the Universal Networking Digital Language (UNDL) Foundation1. 16 different languages joined the project until now, with each responsible for the development and maintenance of the components of its respective language module, and this number is expected to grow. Since 1996, the UNL program has passed through many phases of development and enhancement and crossed important milestones (More information about the earlier system can be found in [7],[8], [9], [10] and [11].

A. *How UNL represents language?*

UNL uses components similar to those natural languages use. The main components are Universal Words (UWs), Relations and Attributes.

Universal Words constitute the vocabulary of UNL. They are used to express the meaning of any concept. Currently, the UNL system uses a numerical ID to refer to concepts. These IDs are extracted from the English WordNet 3.0 [12], [13] and [14].

Relations, on the other hand, constitute the syntax of UNL. Relations are three-letter symbols that signify the kind of semantic relationship that ties two UWs in a natural language utterance. Examples are the agt (agent) relation, plc (place) relation, and tim (time) relation.

Finally, attributes are additional tags that encode the contextual and/or subjective knowledge present in the original sentence into the UNL graph. They are used to further modify the semantic network and add information that is not expressed via UWs or Relations.

The UNL system also uses an extensive knowledge base that hierarchically arranges the human concepts in a way that enables the inheritance of the linguistic and non-linguistic information encoded with each concept from higher to lower levels. For more information on UNL components, cf. www.unlweb.net.

III. THE UNLWEB

The UNLweb is the UNL workplace; it was created to bring together the community of researchers and developers of UNL as well as the general people interested in UNL in particular and Natural Language Processing (NLP) in general in order to share their expertise, documentation and resources. It also aims at collecting all information and training related to the development of UNL in one place for those who are interested in joining the effort. It contains several working platforms including the UNLarium, the UNLdev, the UNLwiki, the UNL EDGES and VALERIE (VirtuAlLEaRnIng Environment) which is introduced to teach those wishing to be part of the project how to deal with natural language phenomena, especially in connection with the Universal Networking Language. Successful candidates are granted a permission to work within the UNLarium. VALERIE comprises two series of certificates; CLEA (CLEA250, CLEA450 and CLEA700) and CUP (CUP500) [2]. These certificates are required in order to be an active user and participate in the development of UNL resources.

No previous knowledge in UNL or NLP is required in order to create an account in the UNLweb environment.

The UNLweb is entirely free and all its results are released under an Attribution Share Alike (CC-BY-SA) Creative Commons license.

A. *The UNLarium*

The UNLarium is the development environment used in the production of the language resources. Members in the UNLarium are allowed to search or browse the language resources. The UNLarium comprises three main sections; Dictionary, for creating, editing and reviewing lexical resources. Grammar, for creating and editing inflectional paradigms, subcategorization frames, and other analysis and generation grammar rules. And Corpus, for adding, editing and exploring UNL documents. Users can also import language resources developed elsewhere into the UNLarium or export resources developed within the UNLarium to any other environment.

The UNLarium is considered a research workplace for exchanging information and testing the linguistic constants that have been proposed for describing and predicting natural language phenomena. Its main goal is helping the UNL system in devising a language-independent interlingua that would be as comprehensive and harmonized as required for NLP tasks.

Participants can develop resources related to any of the languages involved in the UNL system currently.

The advantages of participating in the development of UNL include promoting access to information and ideas in native languages. UNL resources are not limited to use within the UNL program, as strongly standardized resources they can be used in any field related to Natural Language Processing (NLP). Moreover, the data is exportable in several different formats (http://www.unlweb.net/unlarium/).

B. *The UNLdev*[3]

The UNLdev is the wrapper application for the development of various UNL tools and applications to achieve the aims of the Universal Networking Language. These tools implement different tasks; they perform data transformations, analysis, search and generation. Secondary tools have been developed to aid the basic ones in the completion of their tasks. These secondary tools help in building the resources (such as dictionaries and grammars) that are later used by the basic tools and engines. There are tools for professional (linguists, computational linguists) and non-professional users.

Most of these tools are shareware, in the sense that the source code is not provided and users are supposed to sign the UNL Development Set agreement in order to have access to the software. UNLdev includes three UNLization software; IAN, the UNL Editor and SEAN take a natural

[2] http://www.unlweb.net/valerie/

[3] The UNLdev is developed by the UNDL Foundation, Geneva, Switzerland in cooperation with Bibliotheca Alexandrina, Alexandria, Egypt.

language input and deliver output in UNL. On the other hand, the UNLdev also contains one NLization software, EUGENE, its input is in UNL while the output is in natural language. The UNLdev also includes other tools and applications [4].

- **IAN:** this (Interactive ANalyzer) is a tool for the semi-automatic analysis of natural language texts and the generation of UNL Documents; word sense disambiguation is still carried out by a language specialist, nevertheless, the system can filter the candidates using an optional set of disambiguation rules. Syntactic processing, on the other hand, is carried out automatically using the natural language analysis grammar, but syntactic ambiguities are referred back to the user, who may choose to backtrack to a different syntactic path. Nevertheless, human interaction is always optional, and is only used to improve the results. If no human intervention took place, the system would output the most likely alternative. It uses UNL Resources (NL-UNL Dictionaries, UNL Transformation and Disambiguation Rules) to process Natural Language texts and produce UNL Documents. At any rate, IAN would be fully automatic as soon as the disambiguation devices it employs are perfected.
- **The UNL Editor**: it is a more realistic manual analysis engine and it is supposed to take over the analysis process until the disambiguation devices employed by the system mature enough for the semi-automatic engine (IAN) to operate. The UNL editor is a graph-based UNL authoring tool that facilitates the work of developers by providing them with a functionality to load/create UNL documents and manage their contents. A detailed description about this tool can be found in [19].
- **SEAN**: is a language-independent information extraction system expected to automatically extract structured information in the format of the UNL Knowledge Base, from unstructured machine-readable natural language documents.
- **EUGENE:** Eugene (the dEep-to-sUrface natural language GENErator) is a tool for the generation of natural language texts from UNL Documents. It uses UNL Resources (UNL-NL Dictionaries, UNL Transformation and Disambiguation Rules) to process UNL Documents and to produce an equivalent text in a target natural language. The processing of the documents can be done sentence by sentence or for the whole document at once. Similar to IAN, EUGENE is language-independent, it simply uses the target language grammar rules and UNL-NL dictionary in order to decode the incoming UNL document and generate it in natural language format.

C. EDGES

Concepts are the most fundamental constructs in theories of the mind; given their importance to all aspects of cognition, it is no surprise that concepts raise so many controversies in philosophy and cognitive science [15]. The mental lexicon differs from the man-made lexicon in that it is not just a general collection of words; instead, it deals with how those words are activated, stored, processed, and retrieved by each speaker. The mental lexicon is not organized alphabetically like a dictionary; rather, it seems to be organized in a more complex manner, with links between phonologically and semantically related lexical items [16]. Many researches have been conducted to identify the exact way that words are linked and accessed in the mental lexicon.

[4] http://dev.undlfoundation.org/index.jsp

A common method to analyze these connections is through a lexical decision task, this method has been used for many years to figure out how the mental lexicon is structured. Participants in this task are required to respond as quickly and accurately as possible to a string of words and determine if the string is a non-word or a real word. Reaction times from this task indicate that certain words are more active in participants minds after related words have been presented. An example of this would be to present the word "bread" to the participant and then see an increased reaction time later to the word "butter", this increased reaction time demonstrates that related words are stored closely in the mental lexicon. By doing lexical decision tasks, researchers have been able to analyze what words are stored with what related counterparts [17]. The results of this experiment are what the psycholinguists seek to prove.

Using one of the applications based on UNL we can achieve this task. This application can draw networks of related concepts in any natural language. EDGES (Entity Discovery and Graph Exploration System) is a UNL visualization tool that displays UNL graphs in a hyperbolic mode, and allows for node expansion, collapse and navigation. It offers a visual attractive, user-friendly and intuitive way for exploring semantic networks. It can be used to show the relations between the concepts.

Moreover, Edges is capable of illustrating the distance between concepts as shown in figure 1. For example the word "كتاب" (book) is related to many words, however, not all relations are at the same distance; some are far such as "عصر" (age/era) and some are nearer such as "قصة" (story). In addition, some words are related to "كتاب" (book) but indirectly such as the word "مشكلة" (problem).

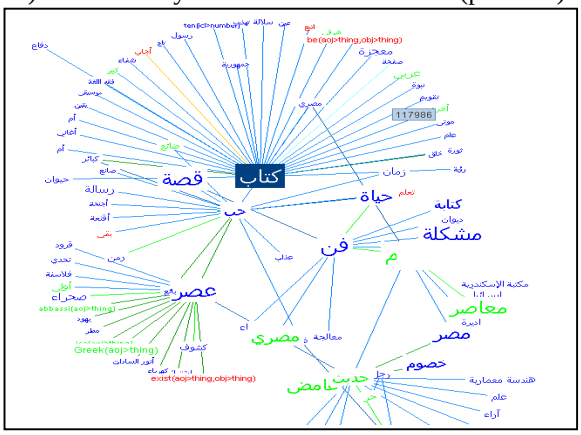

Figure 1. Example of EDGES graph

In addition to Arabic, EDGES can explore the network of the concepts in many other languages such as English, French, Spanish, Russian, Italian, etc. (http://www.undlfoundation.org/edges/)

The variation in the distances between concepts in figure 1. might be a clue that helps psycholinguists measure the distances between concepts in the human mind and how the human brain orders and responds to words.

IV. THE ARABIC UNL SYSTEM

Each of the participant 16 languages mentioned above are expected to make use of the UNL system components and resources to fit into the UNL system. The Arabic UNL center in Bibliotheca Alexandrina is responsible for developing the Arabic resources necessary to encode and

decode Arabic utterances into and from UNL. The resources required include Arabic analysis and generation dictionaries as well as Arabic analysis and generation grammars. Currently, a foundation grammar has been built for Arabic analysis and generation.

A. The Arabic dictionary

Building the Arabic dictionary within the UNL framework requires that the user acquired the first Valerie certificate; CLEA 250.

Entries in any dictionary can either be nouns, verbs, adjectives or adverbs. The user is given a gloss, and example as well as the headword, he/she is expected to determine the most appropriate Arabic lemma to translate the English headword and gloss. The interface also shows the translation of this entry to different other languages. Some basic information are added to define each part of speech such as lexical structure, part of speech, gender, number, transitivity and so on and so forth.

Morphological information responsible for generating the different word forms of entries are added. This morphological information include inflectional paradigms; 332 inflectional paradigms containing 10386 rules are created to describe the morphological behavior of Arabic nouns, verbs and adjectives. 125 inflectional paradigms generate the different forms of Arabic verbs according to their pattern. 143 inflectional paradigms generate the different dual and broken plural forms of Arabic nouns. 62 inflectional paradigms generate the different forms of Arabic adjectives

The word form generated by inflectional paradigms depends on several linguistic features assigned to the base form of the Arabic word itself, as well as the linguistic feature of the subject of the verb [18].

Other than morphological information, syntactic information are also added to entries in the generative dictionary. These are specified in subcategorization frames which define the number and the type of arguments that a base form needs. In total, 34 subcategorization frames describe the syntactic behavior of the Arabic words[18].

When adding an entry in the Arabic UNL dictionary, the user is also supposed to specify the region in which the entry is used. For example Jordon, Kuwait, Egypt, Oman, etc.

The Arabic analysis dictionary (enumerative dictionary) lists all Arabic words and enumerates all the possible word forms of each word. It currently includes about 3,500,000 entries. On the other hand, the Arabic generation dictionary (generative dictionary) only lists a base form to represent each Arabic lexical items. Along with each base form is a set of morphological rules that enable the generation of the different word forms required in different context. It currently lists about 136,000 entries.

B. The Arabic Grammar

Participating in the building of the Arabic grammar within the UNL framework requires that the user acquires the Valerie certificates; CLEA 700 and CUP500. Grammars are the formalizations required to transform Arabic sentences into the Universal Networking Language and vice versa. The Arabic grammar is composed of two branches; the analysis grammar and the generation grammar. The analysis grammar is composed of rules that analyze Arabic input sentences morphologically, syntactically and semantically in order to reach its abstract meaning and put in UNL format. It requires an accurate understanding of the human language, thus, the UNL system provides the tools and methods capable of effectively decomposing the sentence into its basic constituents, understanding and encoding the intended meaning behind each constituent and the meaning reflected by its superficial grammatical form as well as its position in the sentence. Moreover, the semantic relation between each constituent and the others should also understand and encoded. On the other hand, the generation grammar must be able to generate language accurately in order to answer questions, for example, or interact with the user for the purposes of translation, information retrieval, etc. The UNL is an efficient and robust means for generating language. The process of generation may be seen to some extent as the mirror image of the analysis process; the abstract meaning reached after the analysis process is transformed into a natural language sentence.

Moreover, both analysis and generation grammars are divided into two main types: transformation rules and disambiguation rules. Disambiguation rules are used to prevent wrong lexical choices, provoke best matches and check the consistency of the graphs, trees and lists. There are three subtypes of disambiguation rules; Network disambiguation rules, Tree disambiguation rules, and List disambiguation rules

However, the second type of rules; transformation rules are the rules responsible for transforming natural language into UNL and vice versa. Thus, they are more complex and are examined in more detail in the following.

The UNDL Foundation's generation and analysis tools assume that the transformation rules should be carried out progressively, i.e., through several stages: LL - List Processing (list-to-list), LT - Surface-Structure Formation (list-to-tree), TT - Syntactic Processing (tree-to-tree), TN - Deep-Structure Formation (tree-to-network), NN - Semantic Processing (network-to-network), NT - Deep-Structure Formation (network-to-tree),TL - Surface-Structure Formation (tree-to-list).

These seven types constitute the main body of rules that are responsible for analyzing natural language into UNL semantic networks (the UNL graph) and generating natural language out of them.

1) Analysis grammar: the process of analyzing natural language into semantic networks is carried out in a modular manner. It uses five of the seven types of transformation rules mentioned above. These five rule types represent the five stages through which input sentences pass starting from natural language sentences passing through syntactic trees and finally forming the semantic network. figure 2. illustrates the elaborate design of the analysis grammar as postulated by the UNL+3 program showing the five stages of analysis.

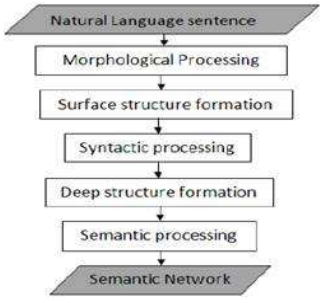

Figure 2. The elaborate design of the analysis grammar

The process of analyzing language includes first decomposing and analyzing the words in a natural language sentence; second, linking together these words to form a syntactic structure then a semantic network that reflects the meaning of the whole sentence.

Word analysis: The Arabic language is highly inflectional and is especially rich in word forms. Thus, Arabic words such as "فتجاهلوني", although a single orthographic word in Arabic, is the equivalent of a whole phrase in some other languages. Therefore, in order to understand the full meaning of such a complex word, the information communicated by the bound morphemes in it must be included into its meaning. Uncovering the bound morphemes in a word (i.e. affixes) and what they represent involves separating them from the core open-class concept by scanning the input words and matching them with the entries in the NL-UNL dictionary. However, there are usually several matches and, consequently, several potential analyses for a single input word. For example, figures 3 and 4 show two of the potential morphological analyses for the previous example word "فتجاهلوني".

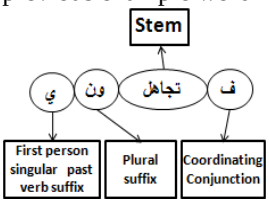

Figure 3. Morphological Analyses 1 of the Word "فتجاهلوني"

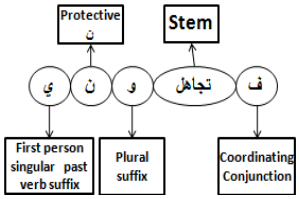

Figure 4. Morphological Analyses 2 of the Word "فتجاهلوني"

To resolve this sort of ambiguity, morphological disambiguation rules are used. Disambiguation rules assign priorities to the possible sequences of morphemes. All of the constituent morphemes are disambiguated and the wrong analyses are refuted until only the most appropriate analysis is left which is the analysis in figure 4.

Sentence Analysis: Understanding and encoding the meanings conveyed by every single morpheme in a certain sentence is far from sufficient to constitute an understanding. A simple list of concepts and tags will be hardly comprehensible even for the native speaker. Grammar rules are required to link these morphemes into a semantic network that represents the meaning of the sentence as a whole.

Deducing the pure semantic meaning directly from a simple list of concepts can be deemed impractical if not impossible; hence, the UNL system has opted for the use of an intermediary stage that maps this list onto an elaborate syntactic tree structure. The ordering of constituents in this tree and the syntactic relations that link them together can, subsequently, help point out the kind of semantic links the sentence implies.

A significant section of the UNL transformation grammar is devoted to transforming the incoming natural language list into an elaborate syntactic structure. To demonstrate this process, the Arabic sentence " منح الرئيس قلادة النيل لمجدي يعقوب" 'The president granted the Nile Medal to Magdi Yacoub' will be used as an example.

The information assigned in the word analysis stage will come into use here; transformation grammar rules use the grammatical, semantic and pragmatic information as guides to determine the syntactic position each morpheme holds in the syntactic structure of the sentence. Along with the transformation rules, disambiguation rules are at work. The result of the syntactic analysis would be the syntactic structure in figure 5.

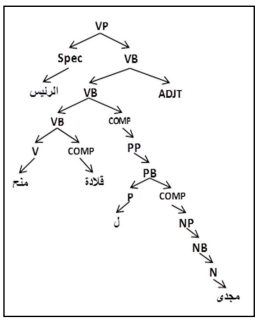

Figure 3. the syntactic structure

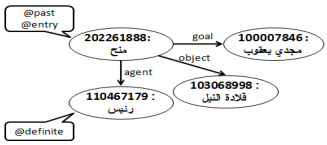

Figure 4. The Semantic Network

Finally, in order to generate the final understanding of an input sentence, different types of transformation rules apply to transform the syntactic structure in figure 5 into a semantic network. This sematic network will incorporate all of the information extracted and defined in the previous stages.

Also on the semantic level, network disambiguation rules apply over the network structure of UNL graphs to constrain the application of transformation rules. Disambiguation rules constrain the type of constituents to be a member in a binary semantic relation.
All of the previous processes work in unison to finally generate the semantic network in figure 6.

2) Generation Grammar: Similar to analysis, the process of generating natural language from a semantic network is modular. It uses five of the seven types of transformation rules mentioned above. Figure 7 illustrates the more elaborate design of the generation grammar as postulated by the UNL+3.

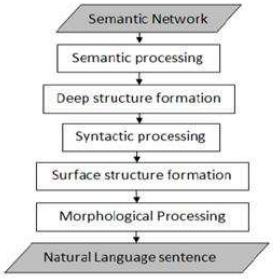

Figure 5. The design of the generation grammar

Generating well-formed sentences has to pass through five stages of transformation rules, in addition to disambiguation rules; passing from the abstract semantic network to a syntactic representation from which the final natural language sentence is generated. The syntactic structure is responsible of constituting a well-formed target language structure. Thus, the UNL framework uses a set of formal rules to convert the pure semantic links that make up the abstract meaning representation (i.e., the UNL network) into syntactic relations. There are two types of syntactic structures; the deep structure and the surface structure. The deep structure of a sentence represents its meaning but interpreted using syntactic tags rather than semantic ones. The surface structure, on the other hand, reflects the ordering of the constituents in the final natural language sentence. In the process of forming a sentence's deep

structure, grammar rules are devoted to mapping the semantic relations from the semantic network onto their equivalents in a syntactic tree. As an example, the semantic network in figure 8 requires the mapping rule in (1) to map the semantic agent onto its counterpart syntactic relation of verb specifier, the rule in (2) maps the semantic object relation onto the position of a verb complementizer relation.

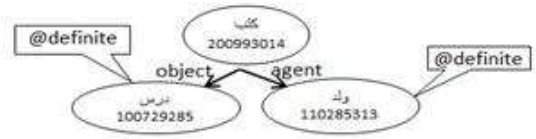

Figure 6. The Input Semantic Network

(1) agt(VER,%01;%02,NOU):=VS(%01;%02);
(2) obj(VER,%01;NOU,%02):=VC(%01;%02);
Then, a different type of grammar rules is subsequently used to determine the exact position of a constituent with regards to the others, when certain conditions are fulfilled such as the rules (3) and (4).
(3) VC(%x;%y):=VB(%x;%y);
(4) VB(%x;%y) VS(%x;%z) :=VP(VB(%x;%y);%z);

the rule in (3) states that if there is a verb complement it will have a minimal projection of a verb phrase, rule (4) states that the previous minimal projection will be combined with the verb specifier to generate the maximal verb phrase VP which include the verb, specifier, and complement. Now the Arabic natural sentence is represented in the form of syntactic tree shown in figure 9.

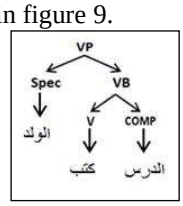

Figure 7. The Syntactic Tree

After that the rules in the tree to list phase will carried out to transform the syntactic tree to a list , when certain conditions are fulfilled such as the rules in (5) and (6).
(5) VB(%x;%y):=(%x)(“ ”)(%y);
(6) VP(%x;%y):= (%x) (“ ”) (%y);
The rules in (5) and (6) transform the syntactic representation to a linear representation. After that the phase of list processing will take place and affixation rules assume the responsibility of generating all the required affixes to fit the tags explicitly marked in the semantic network, for example the attribute @definite will be realized in the list as “ال” by applying the rules in (7),(8).

(7) (%x,@definite):= (%x,-@definite,+DEF);
(8)({NOU|ADJ},%x,DEF,^APOS,^definite):=(%x,+DFN(DEF:="ال"),definite);

This was a brief description of the processes undertaken by the Arabic grammar; whether in the analysis or generation of Arabic natural language sentences along with examples to demonstrate their methods of action.

REFERENCES

[1] B. Maegaard, M. Atiyya, K. Choukri, S. Krauwer, C. Mokbel , M. Yaseen, “MEDAR: Collaboration between European and Mediterranean Arabic Partners to Support the Development of Language Technology for Arabic”, In Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC), Marrakech, Morocco, 2008.

[2] B. Maegaard, S. Krauwer, K. Choukri, L. Damsgaard Jørgensen, “The BLARK Concept and BLARK for Arabic”, In Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC), Geneva, Italy, 2006.

[3] M. Alghamdi, C. Mokbel, M. Mrayati, “Arabic Language Resources and Tools for Speech and Natural Language”, Proceedings of the 2nd International Conference on Arabic Language Resources & Tools, Cairo, 2009.

[4] R. Christopher Johnson, J. Charles Fillmore, J. Esther Wood, J. Ruppenhofer, Ma. Urban, R. L. Miriam Petruck, F. Collin Baker, “The FrameNet Project: Tools for Lexicon Building Version 0.7”. Barker,CA: international computer science institution, 2001.

[5] M. Palmer, P. Kingsbury, D. Gildea, "The Proposition Bank: An Annotated Corpus of Semantic Roles". Computational Linguistics 31 (1): 71–106, 2005.

[6] S. Alansary, M. Nagi, N. Adly, UNL+3: “The Gateway to a Fully Operational UNL System”, in proceedings of the 10th International Conference on Language Engineering, Ain Shams University, Cairo, Egypt, 2010

[7] H. Uchida, M. Zhu and T. Della Senta, UNL: A gift for a millennium, Institute of Advanced Studies, United Nations University, Tokyo.1999.

[8] H. Uchida, UNL: Universal Networking Language – An Electronic Language for Communication, Understanding, and Collaboration. UNU/IAS/UNL Center. Tokyo, Japan, 1996.

[9] I. Boguslavsky, J. Cardeñosa, C. Gallardo, L. Iraola, “The UNL Initiative: An Overview”, Lecture Notes in Computer. Science, Volume 3406: 377– 87, 2005 .

[10] J. Cardeñosa, A. Gelbukh and E. Tovar (eds.), “Universal Networking Language: advances in theory and applications”, Mexico City, National

[11] S. Alansary, M. Nagi and N. Adly, “Machine Translation Using the Universal Networking Language (UNL)”, in proceedings of the 8th International Conference on Language Engineering, Ain Shams University, 2008.

[12] J. Bekios, I. Boguslavsky, J. Cardeñosa and C. Gallardo. “Using Wordnet for building an Interlingua Dictionary”, in proceedings of 5th International Conference on Information Research and Applications, (I.TECH). vol.1, pp. 39-46, June, 2007.

[13] S. Boudhh, P. Bhattacharyya, “Unification_ of Universal Words Dictionaries using WordNet Ontology and Similarity Measures”, in proceedings of 7th International Conference on Computer Science and_Information Technologies, (CSIT 2009), Yerevan, Armenia, 28 September – 2 October, 2009.

[14] R. Martins, and V._ Avetisyan, “Generative and Enumerative_ Lexicons in the UNL Framework”, in proceedings of 7th International Conference on Computer Science and_Information Technologies, (CSIT 2009), Yerevan, Armenia, 28 September - 2 October, 2009.

[15] S. Laurence, E. Margolis, concepts: core readings, Concepts and Cognitive Science, chapter1, pp. 3-82, Cambridge, Mass.: MIT Press, 1999

[16] J. Aitchison, Words in the Mind: An Introduction to the Mental Lexicon. Malden, MA: Blackwell, 2003.

[17] G. Altman, the Ascent of Babel: An Exploration of Language, Mind, and Understanding. Oxford University Press, ISBN 0198523777, 1999.

[18] S. Alansary, A UNL Based Approach for Building an Arabic Computational Lexicon, 8th International Conference on Informatics and Systems, Faculty of Computers and Information, Cairo, Egypt, 2012.

[19] S. Alansary, M. Nagi, N. Adly, UNL Editor: An Annotation tool for Semantic Analysis 11th International Conference on Language Engineering, Ain Shams University, Cairo, Egypt. 2011.

[20] N. Adly, S. Alansary, Evaluation of Arabic Machine Translation System based on the Universal Networking Language, 14th International Conference on Applications of Natural Language to Information Systems, 2009, Saarland University, Saarbrücken/Germany.