Grammatical Relation Extraction in Arabic Language

Othman Ibrahim Hammadi and Mohd Juzaiddin Ab Aziz
School of Computer Science,
Faculty of Information Science and Technology,
University Kebangsaan Malaysia, 43600, Selangor, Malaysia

Abstract: Problem statement: Grammatical Relation (GR) can be defined as a linguistic relation established by grammar, where linguistic relation is an association among the linguistic forms or constituents. Fundamentally the GR determines grammatical behaviors such as: placement of a word in a clause, verb agreement and the passivity behavior. The GR of Arabic language is a necessary prerequisite for many natural language processing applications, such as machine translation and information retrieval. This study focuses on the GR related problems of Arabic language and addresses the issue with optimum solution. **Approach:** We had proposed a rule based production method to recognize Grammatical Relations (GRs), as the rule-based approach had been successfully used in developing many natural language processing systems. In order to eradicate the problems of sentence structure recognition, the proposed technique enhances the basic representations of Arabic language such as: Noun Phrase (NP), Verb Phrase (VP), Preposition Phrase (PP) and Adjective Phrase (AP). We had implemented and evaluated the Rule-Based approach that handles chunking and GRs of Arabic sentences. **Results:** The system was manually tested on 80 Arabic sentences, with the length of each sentence ranging from 3-20 words. The results had yielded the F-score of 83.60%. This outcome proves the viability of this approach for Arabic sentences of GRs extraction. **Conclusion:** The main achievement of this study is development of Arabic grammatical relation extractions based ob rule-based approaches.

Key words: Arabic language processing, chunking, grammars, grammatical relations

INTRODUCTION

Grammatical Relation (GR) can be defined as a linguistic relation established by grammar, where linguistic relation is an association among the linguistic forms or constituents. Fundamentally the GR determines grammatical behaviors such as: placement of a word in a clause, verb agreement and the passivity behavior. The GR of Arabic language is a necessary prerequisite for many natural language processing applications such as machine translation and information retrieval.

Every language has its own set of grammars that makes them unique Diab *et al.* (2004; 2007) and Diab (2009). But the serious case of ambiguity arises when they have to be dealt with computers, in terms of translation or information retrieval. The positions of subject, verb and object in every language are the biggest challenge in the information or translation task. If not properly dealt with, the machine enabled translation will deteriorate the whole meaning of the document. Hence it is vital to consider a lot of factors prior to developing any NLP applications. Generally the NLP applications fall under the following categories: Information Retrieval (IR), Information Extraction (IE), Question-Answering (QA), Summarization, Machine Translation (MT) and Dialogue Systems (DS). As mentioned earlier all the applications must carefully analyze the relationship between the grammar of both, the destination and target languages.

Basically the complex and ambiguous sentences and the unique positioning of verbs-subjects-objects in some languages would create problems to the executions of NLP applications, especially in translations. For instance if a sentence contains two names, it would be confusing for the machine to distinguish them as subjects and objects.

These challenges however, have drawn the attention of a lot of researchers towards the semantic analysis of natural language especially in the domains of information extraction, translation and retrieval.

Corresponding Author: Othman Ibrahim Hammadi, School of Computer Science, Faculty of Information Science and Technology, University Kebangsaan Malaysia, 43600, Selangor, Malaysia

Related work: There are many syntax analyzing software (Abney, 1996) but only a few focus on grammatical relation extraction. Most of techniques on full parsing, parser will not have specific grammatical relation extraction. There are applications such as developing an Arabic parser, Arabic parsing using grammar transforms, a rule-based approach for tagging non-vocalized Arabic words and pola grammar technique for grammatical relation extraction in Malay language.

After a long period of dominance of the statistical paradigm in NLP arena, the dawn of a improved interest has been witnessed in Rule-Based approaches to solve general problems like morph syntactic tagging (Neumann *et al.*, 2000; Hinrichs and Trushkina, 2002) and (Oliva and Petkevifc, 2002) also, partial syntactic parsing (Grover and Tobin, 2006). Much focus was given towards coupling statistical and rule based techniques (Piasecki, 2006).

The benefits of Rule-Based grammatical relations are that, the rules can be manually inscribed and easily understood. However, the drawbacks are that, the rules are linguistic and corpus dependent and consumes huge volume of work and requires lots of language expertise (Albared *et al.*, 2009; Shaalan, 2010).

According to Kinyon (2001) a rule-based grammatical relations compiler generates a solid grammatical relation extraction, which is applicable for all the texts from any field in all the language. However even devoid of the training data, he had applied a very restricted number of rules to recognize boundaries and demonstrated that the parsing is done steadily: The input is scanned meticulously from left to right, in a solo pass. He has utilized his compiler to produce grammatical relations for French to examine the linguistic features of his tool for the noun phrase grammatical of English (Penn Treebank). A preciseness of 90.8% and a recall of 91% were achieved for opening brackets, however for the closing brackets a precision of 65.7% and recall of 66.1% were achieved. For French (newspaper corpus "LeMonde") he got a recall of 94.3% with a preciseness of 95.2% for opening brackets and he got a preciseness of 92.2% and a recall of 91.4% for closing brackets.

There are several parser and NLP techniques that have been proposed and used by the applications discussed in this chapter. Ahmed (1999) has solved the parser Arabic language using the Rule-based approach and has achieved 77% as highest accuracy. Aziz *et al.* (2006) have solved the problem, which was similar to ours in grammatical relation extraction, but in other Arabic language and in Malay language the authors have achieved 87% for Adjunct, 89% for subject, 80% for post-subject, 83% for conjunction and 86% for predicate.

Loftsson (2007) has solved the parser Icelandic language with Rule-based approach and has achieved 85.43% for subject and 72.60% for predicate.

It is practically difficult to make computers to think like human beings, especially in decision making. Hence researchers, face lot of challenges while analyzing different languages. This research focuses on the challenges faced by one of the researchers. Many techniques have been proposed to tag Arabic, English and other European language corpora. One of these techniques is the rule-based technique and all other techniques are extended to it. We have employed the Rule-based technique in our system, to utilize the rules in the morphological analyzer to construct a new technique like statistical model or semantic analysis to map a given word to the corresponding TAG.

Arabic sentence: In general, the sentence is a sequential combination of words. Arabic has flexible syntax. Therefore, Arabic sentences will have different types of word orders such as: VSO, SVO and VOS. Furthermore Arabic sentence can also be constructed without verbs, such as subject + predicate. See following example:

- (الشمس مشرقة) (The sun is shining)…subject + predicate
- Furthermore, a full sentence can be constructed with just only one word, without any syntactic errors (Attia, 2008). See next example
- اعطیتمونیها (You gave it to me) …. verb + subject + Object

In Arabic language, there are two types of sentences

- Simple sentence
- Complex sentence

The simple sentence can be constructed with subject and predicate or verb with subject. Complex sentence consists of more than one subject, predicate and verb. By conjunction particle "و" "and" can be two or more sentences can be joined together:

ذهب الولد و البنت الى المدرسة

The basic grammars of Arabic sentence are as follows:

- Verb Phrase + Verb Phrase (VP +VP)

- Verb Phrase + Noun Phrase (VP +NP)
- Noun Phrase + Verb Phrase (NP + VP)
- Noun Phrase + Noun Phrase (NP + NP)
- Noun Phrase + Preposition Phrase (NP + PP)
- Noun Phrase + Adjective Phrase (NP + AP)

These clauses of sentence are used in our system to recognize grammatical functions (Subject, Predicate and Object). In our research we have split the sentences into three phrases, based on the Rule-Based approach:

Phrases in Arabic Sentence: Arabic sentences are made up of three main phrases:

Noun Phrase (NP): In Arabic sentence the noun phrase starts with noun or pronoun, nouns like proper noun, place noun, animal noun.

Verbal Phrase (VP): The verb initiates a verbal phrase in the following forms: present, past and order verbs. It is mightier than a noun phrase. The verb "الفعل" and "subject" "الفاعل" are the components of verbal sentence. It is evident that just with the subject the verb can express the meaning of a sentence. Hence the verb "الفعل اللازم", is called as "Intransitive Verb". For example "سافر الوالد", "the father traveled".

The sequence of a verbal sentence is verb "الفعل" - subject "الفاعل" and the object "المفعول به", the actions of that verb are received by the 'who' or 'what' . In this case the verb "الفعل المتعدى", call is called as "transitive verb". For example "أكل الولد التفاحة", "The boy eats an apple". Time is the key to know the tense of any verb in any language whatsoever. There are four main tenses in the Arabic language:

- Present Tense "الفعل المضارع": refers to present time actions or nearest immediate future and the action is still continuing. For example, "The student writes the lesson", "يكتب الطالب الدرس". The action is still continuing, he is still writing while the statement was made
- Past Tense "الفعل الماضى": refers to past time actions. For example "The student wrote the lesson", "كتب الطالبالدرس". The student has finished writing while the statement has been made
- Order / Imperative Tense "الفعل الأمر": refers to ordered time actions. In this case, they are orders directed from persons of high status to lower status. For example "Read the lesson", "أقرأ الدرس". An order verb form is also done if the verbal sentence is just a verb, For example "read" "أقرأ"

- Future Tense "المستقبل": is indicated in Arabic by adding the word "سوف" or the prefix "س" to the imperfect form of the verb. For example "I will read" "سأقرأ" or "سوف أقرأ"

Preposition Phrase (PP): Prepositional Phrases (PP)"الجار و المجرور" is identical in Arabic and English. The sequence of preposition is trailed by a word or phrase. The Arabic language consists 20 meaningful particles "حرف الجر" and cannot be preceded by another preposition, such as "with""مع", "from" "من", "to""إلى", "for""لـ".

MATERIALS AND METHODS

The Arab-GR system is restricted to identify the best methodology for rule-based for two parts (i) shallow parsing for Arabic language. The boundaries of the noun phrases, verb phrases and preposition phrases will be discussed by the analysis of Arabic phrases. This means that the components of each of the noun phrases, verb phrases and preposition phrases phrase will be explained with examples for each phrase and (ii) extraction of Grammatical Relation (GRs) for Arabic language. To identify how to extract grammatical relations of the Arabic text in accordance the rules-based approach with the highest accuracy. This means that the components of each of parsing such as: subject, object and predicate.

Structure of the system explains the structural design of the system. Figure 1 illustrates the structural design of the system. The input of this system is the progression of lexemic objects. The system makes reference to three optional modules Part Of Speech (POS), shallow parsing (Chunking) and Grammatical Relation extraction (GRs).

The Arab-GR system firstly makes the tokenization of the Arabic sentence and stores it in the lexical source, where the Part Of Speech (POS) is given a word token. This is followed by the second step, shallow parsing (Chunking) and finally, through (POS) and with (Chunking) based on the Rule-Based the Grammatical Relation extraction (GRs).

Pre-processing modules: The proposed approach includes triple functional pre-processing compartments, employed prior to the shallow parser. It is the input that decides the module to be used. Basically the proposed system is used to deal with unprocessed text. Nevertheless it would be useless in the case of annotated corpus. The modules are normalization, tokenize and POS certainty Albared (2009; 2010; 2011).
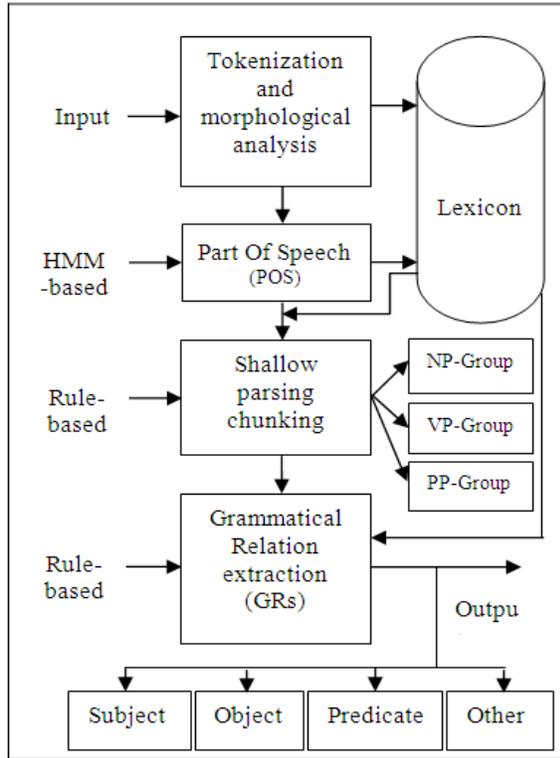
Fig. 1: Architecture of the Arab-GR system

Table 1: (POS) and Marks

POS / Marks	Meaning
PN	Proper Noun
SN	Singular Noun
PSN	Plural Noun
POSS_PRN	Possessive Pronoun
PRN	Pronoun
DEM_ PN	Demonstrative Pronoun
ADJ	Adjective
CC	Conjunction
CD	Number
PART	Particle
PRP	Preposition
P_PN	Personal Pronoun
V	Verb
VP	Verb-Phrase
NP	Noun-Phrase
PP	Preposition-Phrase
? / mark	zero or one parameter
+ / mark	one or many parameter (s)

The normalization module: Before tokenizing the Arabic text the Normalization process should be carried out. The Normalization involves in reducing noise in the data (Kholy and Habash 2010), The Normalization processes are as follows:

- Removing the diacritics "ذَهَبَ ألطّالبُ إلىٰ المدرسة"become "ذهب الطالب الى المدرسة"

- Adding deleted characters. In Arabic, sometimes, some characters of a noun or verb are deleted due to its position in a sentence or if it is preceded with a special particle "اشترا الولد خبزا"become "اشترى الولد خبزا"
- Removal of redundant and misspelled space
- Resolution of the orthographic ambiguity "ىٰ ,"أإآ"", "ئ ي" in Arabic
- Removing the stretching character "~ "

Rules implementation: A rule-based constituent for the grammatical relations is used when the input is a sequence of lexical trees with no constituent structure. The input data is prepared in a specific format and each line contains only a POS tag matching with the word in the sentence. The rule formalism has been designed specifically, for grouping sequences of categories into structures, to facilitate the dependency analysis. These rules are structured in layers that are applied on to the input sequences of sequential categories and they deal with syntactic structure and typical Arabic linguistic grammars, to recognize several major categories of words in Arabic language.

Table 1 explains the Part Of Speech (POS) and some marks that need to apply rules to extract of grammatical relations.

Subject "الفـاعـــل": Arabic language has multi-faceted representation of subject,"الفاعل" the subject describes the verb and evidence of the action in the sentence, in Arabic language the subject always, comes after verb, whereas in English language it comes before verb. If the subject comes before verb here the name is "المبتدأ" but in English it remains the same name is the subject or inchoative.

The subject in the Arabic language on takes several forms. It could come in single or plural of proper-noun, or in pronoun, the pronouns in the Arabic language comes either separate or connected or hidden:

- R1 (s) → VP + NP {(Proper-Noun) + + complement}
- R2 (s) → VP + NP {Proper-Noun + Particle + Proper-Noun}
- R3 (s) → VP {V + PRP} + NP | PP
- R4 (s) → VP {V} + NP {SN + ADJ}

Object "المفعـــول بـــه": The object is name evidence of a verb signed by the subject. There must be an object name or pronoun in the sentence for the verb to be present. There are two types of objects in Arabic language.

Table 2: Separate object pronouns

Pronouns	Meaning
اياي	Me
اياك	You
اياه	Him
اياها	Her
اياكما	You
اياهما	Them
اياكم	You
اياكن	You
اياهن	Them
اياهم	Them
ايانا	Us

Table 3: Attached object pronouns

Pronouns	Meaning
ي	Me
ـك	You
ـه	Him
ـها	Her
ـكم	You
ـكما	Them
ـنا	Us
ـكم	You
ـكن	You
ـهم	Them
ـهن	Them

Direct object: Direct object is either general name or proper-noun or clause. Direct object extracts two types: name and pronoun. Some rules to extract direct Object are explained herewith.

Name-direct object:

- R5 (s)→VP {V} + NP {PN + SN}
- R6 (s) →VP {V} + NP {PN + Particle + PN} + ADVP + NP + Complement
- R7 (s) →NP {SN} + VP {V + PN}

Pronoun-direct object: In Arabic language a set of pronouns comes as a separate object. Table 2 explains pronouns as object:

- R8 (s)→V + {some of one above Pronouns} + Complement
- R9 (s) →{some of one above Pronouns} + V

Indirect object: In Arabic language a set of pronouns comes as an indirect object. Table 3 explains pronouns as object:

- R10 (s) → one word contain to {Verb + Subject + Object}
- R11 (s) → V + {Some of one above Pronouns} + Complement

Subject and predicate "المـــبتدأ و الخـــبر": A subject refers what or whom the sentence is about. The predicate is the part of the sentence that tells something about the subject.

Sentences are always started with subject and not Verb; this subject either can be proper-noun, or pronoun. The next rule recognizes subject directly and clearly:

- R12 (s) →NP {PN + SN | PSN | ADJ} + Complement
- R13 (s) →NP {PRN + SN | PSN | ADJ} + Complement
- R14 (s) → NP {PRN + SN | PSN + PR} + Complement

There are four types for Predicate.

Single word:

R15 (s) →NP {PN + SN | ADJ}

Nominal phrase:

R16 (s)→ NP {PN} + NP + Complement

Verbal phrase:

R17 (s)→ NP {PN} + VP + Complement

Preposition phrase:

R18 (s) →NP {PN} + PP + Complement

Algorithm: The technique has two groups of rules to process grammatical relation. The first group is chunking, this stage do recognize the sentence to three phrases, Noun Phrase (NP), Preposition Phrase (PP) and Verb Phrase (VP). Second group is grammatical relation extraction this step do recognize the sentence after first stage to many functions of Arabic language, focus this system on three main functions in Arabic are Subject, object and predicate.

First phase (Chunking)
Begin
1. Read text
2. Tokenization
3. Take word to store in lexicon
4. Search for the word in the lexicon
5. If found then
6. Return the corresponding tag (Marching with the Rule of first group)

Then
Go to second phase
7.   Else
Back to step 2
End
Second phrase (GRs)
Begin
1.   List sequence of tags corresponds to each phrase (First phase)
2.   Ignore the tag of ambiguity word
3.   Compare a sequence of tags with the Rule of step 1
4.   When one grammar rule matched
5.   Get to functions Arabic grammar
End

In this phase, the proposed system will recognize grammatical relations. The same theme of rules that has been discussed in chapter 3 is implemented to recognize the three main grammatical functions (subject, object and predicate) with others functions.
Implementation of algorithm:

Step 1: In put sentence (3.1 …. 3.7) Referred to in chapter 3.
Step 2: Split sentences to word by word.
Step 3: Store the words from output (Step 2) in lexicon
Step 4: Given tag to the words in (Step 3) Part Of Speech (POS).
Step 5: Split to Phrase, in this (Step) Arab-GR technique made up of clause to input sentence. Phrase is one word or more than words made up of gathers.
Step 6: Substantiality step to recognize grammatical functions. Dependence on (STEP 5) that Arab-GR technique make search in phrases get away of lexicon vocabulary (POS), if find any word tagged then stop.
Step 7: Repeat (Step 1) to (Step 6) on the next sentence.

RESULTS

In order to evaluate the accuracy of the proposed Arab-GR system, the results of system had been compared with human judges. In Arabic there are no standard methods for automatic measurement of grammars. Hence, it is compensated by a manual evaluation for checking the grammar accuracy.

The following steps describe the evaluation methodology:

- Run the system on the input sentence test
- Obtain the output system and compare it with the human results

- Classify errors that appeared from both results
- Assign a suitable F-score for both situations errors and correctness grammar functions (Subject, Predicate and Object). A rang of F-score between 0-100%
- Compute the percentage of the total F-score for both above situations.

Experiment: The purpose of this experiment is to investigate whether the Arab-GR system is sufficient enough for extracting of grammatical relation in Arabic sentences. As discussed in chapter IV the rules based method was employed for extracting the grammatical function in Arabic sentences. The accuracy test was conducted using data set consisting of 80 sentences. It has been randomly selected (in house data set).

The best evaluation method has been implemented for each output list of the classifications of grammatical relations (Subject, Object and Predicate). Ultimately, the 80 real sentences of different lengths were successfully parsed. This phase allowed the syntax to significantly mature, as it is exposed it to the sense of real-life data and deal with high levels of complexity and variations. However, this strategy is limited only to Subject, Object and predicate.

As explained with graph below, the Arab-GR has achieved 83.60% of accuracy.

DISCUSSION

The main objective of this research is to investigate the extraction of the grammatical relations from Arabic sentences, to achieve the objective, of designing the Arab-GR system.

The study had recognized subject, object and predicate to benefit the natural language fields such as: Information Retrieval (IR), Question Answer Applications (QA), Named Entity Recognition (NER), Speech Synthesis and Recognition (SSR), Machine Translation (MT), Index Term Generation (ITG),

Rule-Based approaches are witnessing a renewed interest in NLP applications in an attempt to solve common problems. It deals with problems faced by Arab-GR systems such as different sentences with the same meaning, The Latent Personal Pronoun and The Connected Personal Pronoun. It comprises some of the earliest perspectives for solving the shallow parser difficulty of recognizing chunking as a tagging task. The hand-written and easily comprehended Rule-Based approaches are used in extraction of grammatical relation. But a heavy reliance of the quality and size of training corpora is needed or the machine learning techniques. When the training set and the testing data

are associated with the same domain, they generally offer better outcome.

CONCLUSION

Depending on Rule-Based approach, this research was divided into two phases. First was shallow parsing. This stage was attempted to enhance the shallow parser (Chunking). While the second phase extracted grammatical relation.

The chunking segmented sentence to phrases. In this Noun Phrase (NP), Verb Phrase (VP) and Preposition Phrase (PP) were identified.

F-scores against the length of sentence, the graph clearly presents and compares a grammatical functions performance analysis against

The Fig. 2 shows the result for three main functions in comparison with the human expert, we got 86% as subject, 84% as object and 80% as predicate.

These phrases were used in the process of recognizing the grammatical elements in second phase of this research.

In the Second phase the GRs were developed in two steps.

Initially, the rules that comprise syntax for Arabic that provides an accurate syntactic relation of a sentence that had been obtained. The syntax had been established particularly for the reason of comprehending scientific Arabic text. However this has the benefit that the syntax can be tailored to the particular necessities of the scientific field. Alternatively, we had attempted to espouse broad resolutions as much as possible, as these augment the possibilities that the syntax can be used in other fields as well. Accordingly, in devising the syntax we sought stability among short-term and long-term goals.
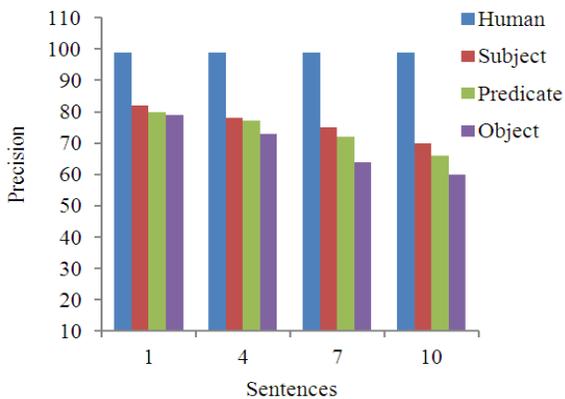

Fig. 2: Evaluation graph of compare human with Arab-GR system

The syntax fall into either one of the two categories: simple sentence or compound sentence. The former is not linked to any another sentence, instead it might entrench another sentence. The later is more than a simple sentence linked with a juxtaposition of article (أداةالعطف). There are three classes of simple sentences: nominal, verbal and special sentences. The special sentences are either the unique verbs (Kana and his sisters كانوأخواتها.), or unique particles ('Inna and his sisters أن وأخواتها.).

The second step implements the parser that allocates grammatical structure on input sentence. As the system has been built as a complete module it can be adapted towards any other related systems.

REFERENCES

Abney, S., 1996. Partial Parsing Via Finite-State Cascades. J. Natural Language Eng., 2: 337-344. DOI: 10.1017/S1351324997001599

Ahmed, A.F.., 1999. Developing an Arabic parser in a multilingual machine translation system. M.Sc. Thesis, CAIRO University.

Albared, M., N. Omar and M. Ab Aziz, 2009. International review on computers and software. Arabic Part Speech Disambiguation, 4: 517-532.

Albared, M., N. Omar and M.J.A. Aziz, 2011. Developing a competitive HMM Arabic POS tagger using small training corpora. Intell. Inform. Database Syst. Lecture Notes Comput. Sci.., 6591: 288-296. DOI: 10.1007/978-3-642-20039-7_29

Albared, M., N. Omar, M.J.A. Aziz and M.Z.A. Nazri, 2010. Automatic part of speech tagging for Arabic: An experiment using bigram hidden Markov model. Rough Set Knowl. Technol. Lecture Notes Comput. Sci., 640: 361-370. DOI: 10.1007/978-3-642-16248-0_52

Attia, M.A., 2008. Handling Arabic Morphological and Syntactic Ambiguity within the LFG Framework with a View to Machine Translation. PhD Dissertation, University of Manchester, UK.

Aziz, M.J.A., F. Ahmad, A.A.A. Ghani and R. Mahmod, 2006. Pola grammar technique for grammatical relation extraction in Malay language. Malaysian J. Comput. Sci., 19: 59-72.

Diab, M., K. Hacioglu and D. Jurafsky, 2004. Automatic tagging of Arabic text: From raw text to base phrase chunks. Proceedings of the 5th Meeting of the North American Chapter of the Association for Computational Linguistics/Human Language Technologies Conference (HLT-NAACL' 04), PA, USA., pp: 149-152.

Diab, M., K. Hacioglu and D. Jurafsky, 2007. Automatic processing of modern standard Arabic Text. Arabic Computat. Morphol., 38: 159-179. DOI: 10.1007/978-1-4020-6046-5_9

Diab, M.T., 2009. Second Generation AMIRA Tools for Arabic Processing: Fast and Robust Tokenization, POS Tagging and base phrase chunking. Columbia University, New York.

Grover, C. and R. Tobin, 2006. Rule-based chunking and reusability. University of Edinburgh.

Hinrichs, E.W. and J.S. Trushkina, 2002. Forging agreement: Morphological disambiguation of noun phrases. Res. Language Comput., 2: 621-648. DOI: 10.1007/s11168-004-7432-2

Kholy, A.E. and N. Habash, 2010. Orthographic and morphological processing for English-Arabic statistical machine translation. Columbia University.

Kinyon, A., 2001. A language--independent shallow--parser compiler. Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, (ACL '01), ACM, Stroudsburg, PA, USA., pp: 330-337. DOI: 10.3115/1073012.1073055

Loftsson, H., 2007. Tagging and parsing Icelandic text. University of Sheffield, UK.

Neumann, G., C. Braun and J. Piskorski, 2000. A divide-and-conquer strategy for shallow parsing of German free texts. The Pennsylvania State University.

Oliva, K. and V. Petkevifc, 2002. Morphological and syntactic tagging of Slavonic languages. A Lecture at Empirical Linguistics and NLP, Sozopol, Bulgaria.

Piasecki, M., 2006. Hand-Written and automatically extracted rules for polish tagger. Proceedings of the 9th International Conference on Text, Speech and Dialogue, (TSD' 06), ACM, Springer-Verlag Berlin, Heidelberg, pp: 205-212. DOI: 10.1007/11846406_26

Shaalan, K., 2010. Rule-based approach in Arabic natural language processing. Int. J. Inform. Commun. Technol., 3: 11-19.