A Top-Down Chart Parser for Analyzing Arabic Sentences

Ahmad Al-Taani[1], Mohammed Msallam[2], and Sana Wedian[1]
[1]Department of Computer Science, Yarmouk University, Jordan
[2]Department of Computer Science, Al-Aqsa University, Palestine

Abstract: *Parsing of Arabic sentences is a necessary mechanism for many natural language processing applications such as machine translation, question answering, knowledge extraction and information retrieval. In this study, we present a top-down chart parser for parsing simple Arabic sentences, including nominal and verbal sentences within specific domain Arabic grammar. We used the Context Free Grammar (CFGs) to represent the Arabic grammar. We first developed the Arabic grammar rules that give precise description of grammatical sentences. Then, we implemented the parser that assigns grammatical structure to the input sentence. The parser is tested on sentences extracted from real documents. Experimental results showed the effectiveness of the proposed top-down chart parser for parsing modern standard Arabic sentences. From a practical perspective, the parser is able to satisfy syntactic constraints and reduce parsing ambiguity.*

Keywords: *NLP, Parser, chart parsing, top-down chart parser, context free grammar, syntactic structures.*

Received February 14, 2010; accepted August 10, 2010

1. Introduction

Arabic is the fourth most widely spoken language in the world. It belongs to the Semitic family of languages which differs from Indo-European languages in terms of its syntax, semantic and morphology. Although different spoken Arabic dialects exist throughout the Arab world, there is only one form of the written language found in printed works, and it is known as Standard Arabic [15].

Parsing is defined as the process of identifying the structure of a specific sentence according to a given grammar. The term *parser* is used in cases where the sentences are made up of information units of any kind.

Parsing Arabic sentences is a difficult task. The difficulty is due to the following reasons: first, the average length of an Arabic sentence is 20 to 30 words, and in some sentences, the number of words exceeds 100. Therefore, Arabic sentences, by nature, are long and complex. Second, the Arabic sentence is syntactically ambiguous and complex due to the frequent usage of grammatical relations, order of words and phrases, conjunctions, and other constructions such as diacritics (vowels), which is known in written Arabic as "altashkiil" [15]. Parsing Arabic sentences is a difficult task due to the following reasons [12]:

1. The length of the Arabic sentences and the complexity of the Arabic syntax.
2. The omission of diacritics (vowels) in written Arabic "altashkiil".
3. The free word order nature of Arabic sentences.
4. The presence of an elliptic personal pronoun "alDamiir Almustatir".

An important method for natural language parsing is called Chart Parsing. It consists of a tabular-based, top-down parsing algorithm [8]. The basic idea is to obtain a table that contains all substructures generated during the parsing process through different iterations until all possible structures for the sentence are obtained. The chart parser takes as an input a sentence which consists of a set of words and grammar rules that are referred to as production rules.

In recent years, research in the field of Natural Language Processing (NLP) systems has witnessed a significant improvement in an attempt to develop various approaches for parsing different languages. For Indo-European languages, specific parsing approaches had been explored in depth and had emphasized some trends such as lexical function grammars, deterministic parsing, and closer integration of syntax and semantics [10]. As a result, NLP systems for Indo-European systems had gained strength and power.

In this research, we present a top-down chart parser for parsing simple Arabic sentences, including nominal and verbal sentences within specific domain Arabic grammar. Context Free Grammar (CFG) is used to represent the Arabic grammar. CFGs Grammars are consisting entirely of rules with a single symbol on the left-hand side, called the mother. CFGs are a very important class of grammars for two reasons [11]:

- The formalism is powerful enough to describe most of the structure in natural languages.

- Yet it is restricted enough so that efficient parsers can be built to analyze sentences.

We first developed the Arabic grammar rules that give precise description of grammatical sentences. Then, we implemented the parser that assigns grammatical structure to the input sentence.

The rest of this paper is organized as follows. Section 2 reviews some important related work. Section 3 presents the proposed scheme. Section 4 discuses the experimental results of the system. Finally, we present the conclusions drawn from this study in section 5.

2. Related Work

For the last two decades concentration on the Arabic language processing has focused on morphological analysis. In this field, many working systems have been achieved [3, 4, 7, 10, 14] and many others.

In contrast, there were less works reported on syntactic analysis of Arabic. This is due to challenging features of the Arabic language such as high degree of ambiguity, complex Arabic syntax and absence of regular punctuation. Some progress has been made in recent years [2, 5, 9, 12, 13, 14, 15, 16], but there is still no general parser available for Arabic with sufficiently wide coverage. At present, no analyzer seems to be able to analyze ordinary real-world Arabic texts. Most systems simply select types of syntactic phenomena for treatment, with considerable lexical limitations. However, real world texts like article from newspaper, abstract from scientific journals or web pages usually contain all sorts of sentences which cause problems for parsers in assigning a suitable structure [13].

For the Arabic language, the development of Arabic parsing system focused mainly on the analysis of Arabic morphology [9]. Rafea *et al*. [14] analyzed and discussed the problem of implementing a morphological analyzer for inflected Arabic words. To implement their Arabic parser, they took the advantage of the already developed morphological analyzer by integrating it with the Arabic parser.

Al-Daoud *et al*. [2] proposed a framework to automate the parsing of Arabic sentences. The study focused on the simple verbal sentences. The proposed approach is divided into two phases; lexical analysis and syntax analysis. The proposed system assumes that the entered sentences are correct lexically and grammatically.

Attia [3, 4] investigates different methodologies to manage the problem of morphological and syntactic ambiguities in Arabic. He built an Arabic parser using Xerox linguistics environment which allows writing grammar rules and notations that follow the LFG formalisms. Attia tested his approach on short sentences randomly selected from a corpus of news articles; he claimed a performance of 92%.

Bataineh *et al*. [5] implemented a top-down parser with recursive transition network for the analyzing Arabic sentences. According to the authors, the system is tested on 77 sentences and gave a performance of 85.6%.

Daoud [7] presents a lossless compression algorithm based on the affix analysis that takes advantage of the statistical studies of the diacritical Arabic morphological features.

Othman *et al*. [12] used Unification Based Grammar (UBG) formalism to write the Arabic grammar rules using a bottom-up chart parser. The grammar consists of 170 rules. The rules are divided into 22 groups of rules each of which represents a grammatical category such as: object, subject, defined, conjunction form, substitution form etc., each grammar rule has the form: rule (LHS, RHS):- constrains. The grammar rules encode the syntactic and the semantic constrains that help in resolving the ambiguity of parsing Arabic sentences.

Shaalan *et al*. [15] developed an Arabic Parser for modern scientific text. The parser is written in Definite Clause Grammar (DCG) and is implemented to be part of a machine translation system. The development process of the authors' Arabic parser involved two steps. In the first step, all rules that make up the Arabic grammar and that give a precise account for a sentence to be grammatically correct are acquired. In the second step, the parser that assigns grammatical structure into input sentence was implemented.

Tounsi *et al*. [16, 17] presented a method for parsing Arabic sentences using Treebank-based parsers and automatic LFG f-structure annotation methodologies developed by Bikel's [6]. The modified approach learned ATB functional tags and merge phrasal categories with functional tags in the training data.

Most of the related work reported in this study concentrated on short sentences and used hand-crafted grammars, which are time-consuming to produce and difficult to scale to unrestricted data. Also, these approaches used traditional parsing techniques like top-down and bottom-up parsers demonstrated on simple verbal sentences or nominal sentences with short lengths.

3. The Proposed Approach

The main goal of the proposed parser is to provide a computerized system to parse Arabic sentences. In this section, we present the architecture of the system, and discuss the main parts that constitute the proposed top-down chart parser. The proposed top-down chart parsing scheme consists of three main steps: word classification, Arabic grammar identification and parsing. The architecture of the system is given in

Figure 1. In this figure, the arrows indicate the flow of information. Boxes are the modules of the system.

In word classification task, we use the three main categories that are used in Arabic language to distinguish between words. These categories are: Nouns, verbs and particles. In Arabic language, a noun is a word that describes a person, a thing, or an idea. Arabic verbs are similar to those in English. Although the tenses and aspects are different, the verb tenses can be classified as present, past and imperative. The particles are classified as prepositions, adverbs, conjunctions, interrogative particles, exceptions, and interjections.

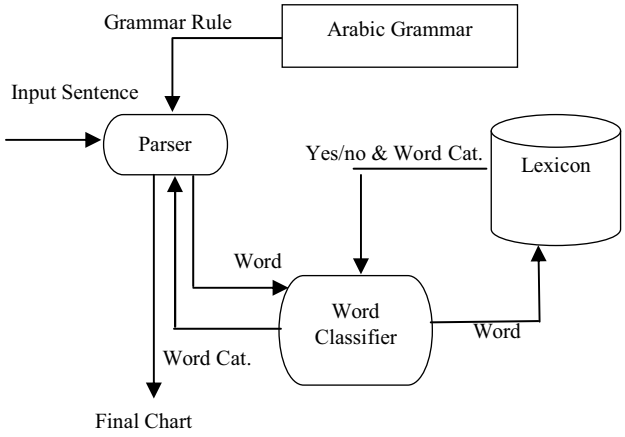

Figure 1. The architecture of the Arabic chart parser (adapted from [7]).

In Arabic language, words are usually formed as a sequence of (antefix, prefix, core, suffix, postfix). In the proposed chart parser, we classify words into nouns or verbs based on their affixes and some other rules. Table 1 shows some of the affixes that we have used in our work. Moreover, we have used more rules of patterns that clearly distinguish words from nouns [1]. Examples on these rules are mentioned in Table 2.

Table 1. The used prefixes and suffixes.

Class	Prefixes	Suffixes
Verb	سأ، ست، سي، فسأ، فست، فسي، ا، ن، ي، ت، س	ون، ي، ين، تما، تن، ني، وا، ان، ن
Noun	بال، فال، كال، لل، ال	اء ، تكن ، تك ، تها، تي، تهم، تكم، ته، تكما، ة، ات

Table 2. Pattern rules used to classify words.

Class	Pattern	Prefix	infix	Examples
Verb	إسْتَفْعَلَ	إسْت	-	استودع،استرحم، استنتج، استحسن
Noun	إسْتَفْعَالَ	إسْت	ا	استوداع، استرحام، استنتاج، استحسان

For the sake of Arabic grammar Identification, we used the CFGs to represent the structure of sentences in terms of what phrases are subparts of others. We classify the rules of grammar into two categorizes according to the type of Arabic sentence:

1. Nominal Sentences: The nominal sentence is defined as a sentence that begins with a noun. The parts of such types of sentences are the inchoative المبتدأ (al-mubtada') and the predicate الخبر (al-Khabar). Copulas or action-verbs may be used freely in this type of sentences. We have used the following rules to identify the nominal sentences:

   NS ➔ NP NP
   NS ➔ NP NS
   NS ➔ NP VP
   NS ➔ NP VS

2. Verbal Sentences: The verbal sentence is the sentence that begins with a verb [3]. It has the following structure: verb + subject + accusative object. We have used the following rules to identify verbal sentences:

   VS ➔ VP NP
   VS ➔ VP NS

3. Domain Words: In addition to the two main classes of sentences (nominal and verbal), the system takes into consideration the existence of particular domain words that uniquely identify the structure of the sentence. Domain words include verbs (الأفعـــال), pronouns (الضمائر) and particles.

There are two main types of verbs, the intransitive verb الفعل اللازم (Al fialu al-lāzim) and the transitive verb الفعل المتعدي. Intransitive verb takes only a subject الفاعِل, as in the sentence : جاء الوَلَد (the boy came). Transitive verb الفعل المتعدي (Al fialu al-mutaaddī) takes a subject الفاعِل and an accusative object, المَفْعول بهِ, as in the sentence: كتب التلميذ الدرس (the student wrote the lesson).

It is important to illustrate here that the system has the ability to identify the structure of a sentence if it has one verb or two non-consecutive verbs, like the sentence: جاء الذي كتب. In addition, the system can identify verbs of the present or past tenses.

The system can identify two main types of pronouns, the independent pronouns الضمائر المنفصلة (al-damā'ir Al-munfasila), that are written as a standalone words and cover only subject functions, and the linked pronouns الضمائر المتصلة (al-damā'ir al-muttasila) that are always written attached to the end of the word they refer to, and are used as personal pronouns subject, object and also have a possessive function. For example, كتابهم (their book) (plural masculine) = Kitabuhum, كتابكما (your book) (dual male or female) = Kitabukumaa. The Arabic independent and linked pronouns are illustrated in Tables 3 and 4, respectively.

The system can identify the structure of sentence that contains different types of particles, these include:

- Conjunctions between two pronouns or two nouns. For example: (and: wa و), (then: thoma ثم), (but: bl بل), (or: ao أو , am أم), (excluding: لا– ف – لكن – حتى).

- Exceptive particles, such as: But: (ada عدا, shoa سوى, ila إلا), (rather than: gher غير), (except: hasha حاشا), (khala خلا).
- Particles which introduce the present verb, such as: (not: Ln لن, Lm لم, Lma لما), (to: Ki كي), (shall "saofa" سوف, seen سين), (edn: إذن).
- Prepositions which introduce the noun, such as: (with: ma'a, مع), (on: ala على), (to: ila الى), (in: fi في), (about: an عن).

Table 3. Independent pronouns.

	Pronouns					
Nominative	He	huwa	هو	I	Ana	أنا
	She	heya	هي	We	nahnu	نحن
	They	humā	هما	You	anta	أنت
	They	Hum	هم	You	antumā	أنتما
	They	Hunna	هن	You (pl.)	antum	أنتم
				You	antuna	أنتن
Accusative	- إياها –إياه –إياكن –إياكم –إياكما –إياك –إيانا –إياي إياهن - إياهم –إياهما					

Table 4. Linked pronouns.

Pronouns	
The Imperfect verb with (you masculine dual, they masc. Dual, you masc. Plural, they masc. Plural, you feminine dual)	الأفعـال الخمسـة واو الجماعة- ألف الاثنين- نا - تاء الفاعل - ياء المخاطبة –
Possessive pronouns	هاء الغائب – كاف المخاطب –ياء المتكلم

3.1. System's Main Functions

3.1.1. Words Categorization

The system provides the user with the ability to perform both an automatic categorization and manual categorization. The selection of either method depends on the extent to which the automatic categorization succeeds. That is, the user might initially use the automatic categorization to break the sentence down into its main components. However, if the automatic categorization failed to find the correct components, then the user can perform the manual categorization, in which he/she will be to insert the component type (i.e., PRO, N, ART, etc.,) manually. To clarify these two methods, we have initially inserted the sentence: أنا و أنت متفقان في الرأي), as shown in Figure 2.

Figure 2. Parser interface.

From this figure, the user can either choose to automatically categorize the components of the inserted sentence, or to do this task manually. If the user chooses the automatic categorization, he must press "Categorization" button and the words categories will be displayed in the "Word Categorization" pane, as shown in Figure 3. If automatic categorization fails, however, the user can press "Manual Cat" button to manually insert the correct category as illustrated in Figure 4.

Figure 3. Word categorization.

Figure 4. Manual categorization.

3.1.2. The Parser

After that the user can use the parser where he must press the "Run Parser" button. Figure 5 shows each step for parsing the sentence to identify its structure.

Figure 5. Parsing sentence.

3.1.3. Final Chart

Finally, after the parsing process is completed, the user can view the final chart of the inserted sentence. Figure 5 shows the parsing process for the nominal sentence "أنا وأنت متفقان في الرأي". The final chart for this sentence

is shown in Figure 6. For purpose of explanation, the final chart for the given sentence is also given in Figure 7.

Figure 6. The chart for "أنا و أنت متفقان في الرأي".

<table>
<tr><td colspan="6">NS3 (rule 6 with NP2 & NS2)</td></tr>
<tr><td colspan="4"></td><td colspan="2">NS2 (rule 3 with NP3 & NP4)</td></tr>
<tr><td colspan="5"></td><td>NP4 (rule 14 with ARTN1 & N2)</td></tr>
<tr><td colspan="4">NS1 (rule 3 with NP2 & NP3)</td><td colspan="2"></td></tr>
<tr><td colspan="3"></td><td>NP3 (rule 12)</td><td colspan="2"></td></tr>
<tr><td colspan="3">NP2 (rule 17 with PRO1 & CONJ! & PRO2)</td><td colspan="3"></td></tr>
<tr><td>NP1 (rule 13)</td><td colspan="5"></td></tr>
<tr><td>PRO 1</td><td>CONJ 1</td><td>PRO 2</td><td>N 1</td><td>ARTN 1</td><td>N 2</td></tr>
<tr><td>أنا 1</td><td>و 2</td><td>أنت 3</td><td>متفقان 4</td><td>في 5</td><td>الرأي 6</td></tr>
</table>

Figure 7. The chart for "أنا و أنت متفقان في الرأي".

4. Experimental Results

The parser is tested on 70 sentences extracted from Arabic documents. Sentences have different sizes from 2 to 6 words. The performance of the system was very good in all experiment scenarios for the various sizes of sentences. Table 5 shows the performance of the parsing experiments while the results for word classification according to their affixes and other rules are presented in Table 6.

Table 5. The results of the parsing sentences.

Type	Total	Correct	%
Nominal Sentence	36	35	97.2 %
Verbal Sentence	34	31	91.2 %
Total	70	66	94.3 %

Table 6. The results of the word classification.

Type	Total	Correct	%
Noun	175	159	90.9%
Verb	42	39	92.9%
Pronoun	42	36	85.7%
Article (Noun)	45	42	93.3%
Article (Verb)	2	2	100%
Conjunction	7	7	100%
Total	313	285	91%

Comparison of related work was a difficult task in this study because most of the reported related work did not present any experimental results. Only three related work could be compared partially to our approach.

Bataineh and Bataineh [5] reported that 85.6% of 90 sentences were parsed successfully using their top-down parser, Tounisi *et al*. [16] reported about 77% parsing accuracy on parsing Arabic sentences using Treebank-based corpus. Ouersighni [13] has used 105 Arabic sentences to test the morphological analyser which gave an accuracy of about 89%, but he didn't report any results for the parser. In contrast, we reported an average accuracy of 94.3% using more efficient parsing method, top-down chart parser. The only previous approach similar to our approach is the work presented by Othman *et al*. [12] which used bottom-up chart parsing, but the authors didn't present any experimental results in their study; instead, they explained the bottom-up parser on a simple Arabic sentence of length 3 words.

5. Conclusions and Future Work

In this work, we have presented an efficient top-down chart parser for parsing simple Arabic sentences. We have used CFG to represent the Arabic grammar. We depended solely on Arabic grammar rules in parsing to determine the structure of sentence within specific domain of Arabic grammar. The grammar rules encode the syntactic and the semantic constrains that help in resolving the ambiguity of parsing Arabic sentences.

The proposed parsing technique will provide promising impact on many language applications such as question answering and machine translation, because the source sentences will be analyzed according to the grammar rules that represent their intended meaning. Thus, reduces syntactic and semantic ambiguity.

After the review of related work in this study, we have drawn the following conclusions:

1. Most of reported related work used traditional parsing techniques like top-down and bottom-up parsers with different methods for representing the grammar like recursive transition networks and GFS. These methods are not robust enough for natural languages like the Arabic language.
2. Most of these approaches used simple verbal sentences or nominal sentences with short lengths.

Despite all the above facts, our proposed top-down chart parser has the following advantages over existing approaches:

- It analyses both Arabic nominal and verbal sentences regardless the length of the sentence. The grammar and the Arabic sentences used in this study

are presented in Appendix A and Appendix B, respectively.

- It uses efficient parsing techniques, i.e., a top-down chart parser.
- Experimental results showed the effectiveness of the system for analysing both verbal and nominal sentences.

Future work will focus on the following related issues:

- Identifying passive or active voice verbs since in this research we did not take diacritics (vowels) into consideration.
- Identifying more than one linked pronoun in the word, as in the verb "قابلتـه" contains two linked pronouns, the first is "تاء الفاعل", and the second is "هاء الغائب".
- Identifying the existence of two consecutive prepositions as in the sentence: قل كل من عند الله.
- In addition, our system cannot identify particles that are used to introduce two present verbs in the sentence such as (من– ما – مهما – منى – أيان – أين – أيـنما – أني – حيثما - كيفما). This will be considered for future work.

References

[1] Abuleil S. and Evens M., "Discovering Lexical Information by Tagging Arabic Newspaper Text," *Journal of Computers and the Humanities*, vol. 36, no. 2, pp. 191-221, 1998.

[2] Al-Daoud E. and Basata A., "A Framework to Automate the Parsing of Arabic Language Sentences," *The International Arab Journal of Information Technology*, vol. 6, no. 2, pp. 191-195, 2009.

[3] Attia M., "An Ambiguity-Controlled Morphological Analyzer for Modern Standard Arabic Modeling Finite State Networks," *in Proceedings of The Challenge of Arabic for NLP/MT Conference*, London, pp. 155-159, 2006.

[4] Attia M., "Handling Arabic Morphological and Syntactic Ambiguity within the LFG Framework with a View to Machine Translation," *PhD Thesis*, University of Manchester, UK, 2008.

[5] Bataineh B. and Bataineh E., "An Efficient Recursive Transition Network Parser for Arabic Language," *in Proceedings of the World Congress on Engineering WCE*, UK, pp. 124-127, 2009.

[6] Bikel D., "Intricacies of Collins Parsing Model," *Journal of Computational Linguistics*, vol. 30, no. 4, pp. 253-259, 2004.

[7] Daoud A., "Morphological Analysis and Diacritical Arabic Text Compression," *the International Journal of ACM Jordan*, vol. 1, no. 1, pp. 41-47, 2010.

[8] Dick G. and Ceriel H., "Parsing Techniques, a Practical Guide," *Technical Report*, England, 1990.

[9] Ditters E., "A Formal Grammar for the Description of Sentence Structure in Modern Standard Arabic," *in Proceedings of Arabic NLP Workshop at ACL/EACL*, Ouersighni, pp. 44-46, 2001.

[10] Feddag A, "Arabic Morpho-Syntax and Semantic Parsing," *in Proceedings of the 3rd International Conference and Exhibition on Multi-lingual Computing*, UK, pp. 125-129, 1992.

[11] James A., *Natural Language Understanding*, Benjamin/Cummings, CA, 1995.

[12] Othman E., Shaalan K., and Rafea A., "A Chart Parser for Analyzing Modern Standard Arabic Sentence," *in Proceedings of the MT Summit IX Workshop on Machine Translation for Semitic Languages: Issues and Approaches*, USA, pp. 33-39, 2003.

[13] Ouersighni R., "Towards Developing A Robust Large-Scale Parser for Arabic Sentences," *in Proceedings of the International Arab Conference on Information Technology*, Tunisia, pp. 15-18, 2008.

[14] Rafea A. and Shaalan K., "Lexical Analysis of Inflected Arabic Words Using Exhaustive Search of an Augmented Transition Network," *Journal of Software Practice and Experience*, vol. 23, no. 6, pp. 567-588, 1993.

[15] Shaalan K., Farouk A., Rafea A., "Towards an Arabic Parser for Modern Scientific Text," *in Proceeding of the 2nd Conference on Language Engineering*, *Egyptian Society of Language Engineering*, Egypt, pp. 103-114, 1999.

[16] Tounsi L., Attia, M., and Genabith J., "Parsing Arabic Using Treebank-Based LFG Resources," *in Proceedings of the LFG09 Conference*, Cambridge, pp. 583-586, 2009.

[17] Tounsi L., Attia M., and Genabith J., "Automatic Treebank-Based Acquisition of Arabic LFG Dependency Structures," *in Proceedings European Chapter of the Association for Computational Linguistics*, *EACL Workshop Computational Approaches to Semitic Languages*, Athens, pp. 45-52, 2009.

Appendix A: Grammar Rules

1. VS --> VP NP
2. VS --> VP NS
3. NS --> NP NP
4. NS --> NP VP
5. NS --> NP VS
6. NS --> NP NS
7. VP --> ARTV V1
8. VP --> V
9. VP --> V1
10. VP --> V P
11. VP --> V1 P
12. NP --> N
13. NP --> PRO
14. NP --> ARTN N
15. NP --> ARTN PRO
16. NP --> N CONJ N
17. NP --> PRO CONJ PRO
18. NP --> N CONJ PRO
19. NP --> PRO CONJ N

Appendix B: Sentences Used For Testing

1. الجهل عاقبته وخيمة
2. فوق كل شجرة طائر
3. إن الصبر من خصال المؤمنين
4. أخوك معاقب من المعلم
5. الجنة تحت أقدام الأمهات
6. لولا العقاب لساء الأدب
7. عندي ثلاثون قلما
8. الكتاب رخيص لكنه مفيد
9. أولئك هم الفائزون
10. نحن العرب نرعى الذمم
11. لا فضيلة أنبل من المحبة
12. ما عاقل في القوم
13. رجل و امرأة فقيرة على الباب
14. كان سمير صديقنا
15. إن الخيانة عاقبتها العار
16. كتيب هذب أخلاقي
17. زيد غير قادم اليوم
18. علاء شاعر كاتب خطيب
19. سلام هي حتى مطلع الفجر
20. أنا و أنت متفقان في الرأي
21. على الآباء مسؤولية الأبناء
22. إلى المدرسة ذهب محمد و محمود
23. لكل عالم هفوة
24. الفتاة محبوبة تقديرا لسلوكها
25. هو الغفور الرحيم
26. حلية الأدب خير حلية
27. القاضيان يحكمان بالعدل
28. النجاح أساسه العمل
29. في بيتنا رجل
30. إن المصباح ضوؤه شديد
31. إن الله لقوي عزيز
32. كل عام و أنتم بخير
33. العدو هزم
34. المدرس أكثر من الطالب خبرة
35. بالقرية مدرسة واحدة
36. هيهات الأمل في النجاح
37. تقرر إعلان نتيجة الامتحان
38. سوف تعلن النتيجة بعد شهرين
39. احترم الناس يحترموك
40. تكلم الخطيب بشجاعة
41. أكل مؤيد التفاحة الحمراء
42. شرعت لإنقاذ الغريق من البحر
43. يعجبني في الحديقة أزهارها
44. كتب على الورقة
45. سمع طيورا تغرد
46. لم يكافأ إلا نحن
47. شرعت لإنقاذ الغريق الصغير
48. يئس القول شهادة الزور
49. قرأت أنا الدرس
50. ذهب زيد و يوسف إلى المدرسة
51. اشتر الثوب بدينار فنازلا
52. استيقظت القرية كلها جمعاء
53. جلست لابتغاء الراحة
54. ترد الغيوم شعاع الشمس
55. عاد الغائبون إلا دوابهم
56. زارني الذي أبوه خطيب
57. علمت أن الموت قريب
58. أهان أخي جاري
59. كسا أخوك الفقير
60. حضر المؤتمر أربعون مندوبا
61. إياك نعبد
62. مات الرشيد ثم المأمون
63. تعلم فايز في مصر
64. أنبأت محمدا الخبر مفصلا
65. مات خالد في العام الماضي
66. خلق الإنسان ناطقا
67. قابلت مدير الشركة
68. فاز فريقنا في المباراة
69. رأيت سيارة جارنا
70. عمل عمل متقن

Ahmad Al-Taani is an associate professor of artificial intelligence at Yarmouk University, Jordan. He received his PhD in computer vision from University of Dundee, UK in 1994. He was the dean of the Faculty of Science and Information Technology at Jadara University during his sabbatical leave in 2008. In 2009, he was appointed as the Vice Dean of the Faculty of Information Technology and Computer Science at Yarmouk University and in 2010, he was appointed as the Dean of the faculty. Dr. Al-Taani has several publications in refereed journals and conference proceedings. His research interest includes: artificial intelligence applications, pattern recognition, image processing, and Arabic language processing.

Mohammed Msallam received his BSc in educational computer from Al-Aqsa University, Gaza in 2002, and MSc in computer science from Yarmouk University, Jordan in 2009. Currently, he is a lecturer at the department of computer science at Al-Aqsa University. His research interest includes: evolutionary computation, swarm intelligence, and Arabic language processing.

Sana Wedian is a lecturer at the Department of Computer Science at Yarmouk University. She received her MSc in computer science in 2009 from Jordan University of Science and Technology. Her research interest includes: Arabic language processing, evolutionary computing, and web spam detection.