Teaching Basics Of Arabic Syntactic Analysis Using PALMYRA Tool

Ahmed abdelghany muhammed

Phonetics and linguistics depatment – faculty of Arts – Alexndria university

Senior lecturer in faculty of linguistic studies and human development – university malaysia kelantan

Abdelghany.ma@umk.edu.my

الإستشهاد المرجعى:

Ahmed Abdelghany Muhammed (2024) . Teaching Basics Of Arabic Syntactic Analysis Using PALMYRA Tool حولية كلية الآداب . جامعة بني سويف. مج 13.ج2.- ص ص 715- 740

Abstract

This research aims to teach the basics of Arabic syntactic analysis for native and non- native speakers of Arabic. In this research, we apply the structural linguistic approach through employing a tree viewer application called PALMYRA 2.4 (a platform-independent graphical tool for syntactic dependency annotation supporting languages that require complex morphological tokenization.) in teaching sentence analysis. PALMYRA is

configured to be compatible to any dependency parsing representation, and to enable annotating many different linguistic features. It uses an interface that relies on drag-and-drop utilities.

descriptors :

: syntactic analysis- Palmyra tool- Arabic language- POS tags- syntactic relations

Introduction

Arabic syntax is a challenging area in Arabic language sciences for native speaker learners, as well as non-native learners. This is because syntactically, Arabic language is a free word-order language, opposed to fixed-word order languages like English. At the morphology level, Arabic is morphologically rich and highly ambiguous, it has a complicated system of morphology characterized by affixational morphemes and complex morphological rules. Missing diacrtization in Arabic written materials makes the analysis process more challenging. One of the main reasons that help in syntactic analysis of Arabic sentences is understanding the meaning from the context and from the non-linguistic background, that is why ancestors argued that "Eʿrāb (syntactic analysis) is meaning", and that is why Automatic processing of Arabic is challenging.

Recently, learning Arabic has begun to gain more attention among learners across the Islamic world as well as the Western world. The Arabic language has shifted from the language of Islam only to being widely recognized as an international language. Thus, there has been an urgent need to innovate pedagogical approaches to teach Arabic grammar

What non-native speakers of the Arabic language need is to recognize the boundaries of different Arabic structural syntactic

components and their functions, so, the system of "Eʿrāb" with tree viewer is the most appropriate and effective way that serves their needs.

There are many grammatical representations that differ in terms of the linguistic theories on which they are based. One of these is the Universal Dependency representation (UD) (Nivre et al., 2016) that is applicable in a lot of languages. It is one of the most popular representations. The Penn Treebank representation (PTB) (Marcus et al., 1993) makes use of a constituency representation that is rich with morphological features. Many languages, including Arabic (Maamouri et al., 2003), English (Marcus et al., 1993), and Chinese (Xue et al., 2005) use this representation in syntactic analysis. PTB representation is very convenient to constituency parsing. The Prague Dependency Treebank (PDT) (Böhmová et al., 2003) is a representation used for annotating treebanks in Arabic (Hajic et al., 2004) and Czech (Böhmová et al., 2003).

Advantages of PALYMRA

- It runs on all hardware platforms or software architectures (platform-independent).
- It is easy to install.
- It can be used offline.
- It is lightweight.
- It is open-source.
- It is easy to use (requires minimal training).
- It lists all the trees sentences to be easily searched and displayed.
- It supports rich-morphology languages (like Arabic).
- It allows drag-and-drop in assigning POS tags and syntactic relations, and has the options of using both shortcut keys and buttons.

- It allows the user to make tokenization by splitting and merging nodes in the displayed trees.
- It introduces an adaptable configuration file that allows the user to configure the list of the syntactic relations and part of speech tags used in the representation.
- It can be extended to annotate other morphological features that go beyond POS tags and syntactic relations.

Related work

Many representation softwares have been developed in the context of syntactic analysis and morphological analysis. Most of these tools were created for natural language processing (NLP) purposes to improve automatic syntactic and morphological analysis. None of these tools were developed for language learning purposes. TrEd (Pajas and Tepánek, 2008), EasyTree (Little and Tratz, 2016), ConlluEditor, and the Quran corpus tree viewer tool are all examples of these representation softwares.

TrEd is a graph visualization software written in Perl. It has been used in many treebank projects including Prague Arabic Dependency Treebank (Hajic et al., 2004) and Columbia Arabic Treebank (CATiB) (Habash and Roth, 2009). The limitations of TrEd are that it is a standalone software and cannot be run through a web browser. In addition to it does not have a simple option for word tokenization. Also, it is difficult to install and learn.

EasyTree (Little and Tratz, 2016) is a light-weight software developed for annotating dependency trees in browsers. PALMYRA is developed on the base of it. One of the limitations of this tool is that it doesn’t support word tokenization. Also, it does not maintain sentence order of nodes.

ConlluEditor tool (Heinecke, 2019) is a fully graphical editor for manual annotation. It allows editing morphological features and syntactic relations. It has the functionality for merging and splitting words. Also, it displays multi-word tokens and empty nodes. One of its limitations is that it uses a client-server architecture.

Quran corpus tree viewer tool (Kais Dukes and Tim Buckwalter, 2010) is used in annotating the Quranic Arabic dependency treebank (QADT) and displaying Quran verses using dependency graphs. These graphs show how each word is related in the Quran verse and what syntactic role it plays in building up a complete syntactic tree. Quran corpus is an Arabic annotated linguistic resource presents the morphology and the syntactic role for each word in the Quran. It provides three levels of analysis: morphological annotation, a syntactic treebank and a semantic ontology. Each word of the Quran is tagged with its part-of-speech and several morphological features. The grammar framework adopted by the Quran Corpus is the traditional Arabic grammar of i'rab (إعراب).

The linguistic framework

PALMYRA takes a configuration file as an input. The configuration file specifies the various configuration and annotation options for the tool. The most important use of the config file is to specify the inventory of POS tags available for morphological tagging process, and the syntactic relations available for syntactic tagging process. One of the main advantages of PALMYRA tool is that it allows users to define their own morphological tag set and syntactic relations. In this way, user can propose different syntactic representations and can control the depth of the morphological and syntactic analysis.

Figure 1: POS Tags and syntactic relations

Part-Of-Speech (POS) tags

There are three main POS tags in the traditional Arabic grammar classification. These are noun, verb and particle. Each of these main POS tags can be divided into many subordinate classes according to morphological and semantic considerations.

Noun (اسم): a content word that has a meaning itself and not associated with time. Subordinate classes include:

Interrogative nouns (أسماء الاستفهام): like (كم - كيف - أين – متى – ماذا – ما – مَن – أيان - أنى)

Deverbal nouns (المصادر - اسم المفعول – اسم الفاعل): nouns that are derived from verbs or verb phrases, but that behave grammatically purely as nouns, not as verbs. However, in some cases it may be a head of subject or object.

Pronouns (الضمائر) include:

Implicit pronouns (الضمير المستتر) which, in turn, include:

Compulsory Implicit (واجب الاستتار): in cases of present verb referring to 1st person pronoun (I write "أكتب" – we write "نكتب"), present verb referring to 2nd person masculine singular pronoun (you [sg-mas] write "تكتب (أنتَ)"), and imperative referring to 2nd person masculine singular pronoun (write {you[sg - mas]} اكتب).

Optionally Implicit (جائز الاستتار): includes any other cases than compulsory implicit.

Prominent pronouns (الضمير الظاهر) which, in turn, include:

Attached pronouns (الضمير المتصل) which include:

Object pronouns (ضمائر المفعول) that include:

1st person (ـنا - ـي)

2nd person (ـكنّ - ـكم - ـكما - ـكِ - ـكَ)

3rd person (ـهنّ - ـهم - ـهما - ـها - ـه)

Subject pronouns (ضمائر الفاعل) which include 1st person pronouns (ـنا – ـتُ), 2nd person pronouns (ـي – ـين - ـنّ – ـتم – ـتما – ـتِ – ـتَ), and 3rd person pronouns (ـن - ـوا - ـون).

Detached pronouns (الضمير المنفصل) which include:

Accusative pronouns (ضمائر النصب) which, in turn, include:

1st person (إيانا – إياي),

2nd person (إياكنّ - إياكم – إياكما – إياكِ – إياكَ)

3rd person (إياه – إياها – إياهما - إياهم - إياهنّ)

Nominative pronouns (ضمائر الرفع) which, in turn, include:

1st person (أنا - نحن)

2nd person (أنت – أنتِ – أنتما - أنتم - أنتنّ)

3rd person (هو – هي – هما – هم – هنّ)

Proper nouns (اسم عَلَم): the names of a particular persons, countries, places, organizations, or things.

Verbal nouns (أسماء الأفعال) include:

Present verbal nouns (أوه – وي – بخٍ – أفٍّ)

Past verbal nouns (هيهات – شتان - سرعان)

Imperative verbal nouns (هيّا - مكانك – أمامك – – حَيَّ – آمين – صه – عليك – دونك – رويدك – هلمّ - هاؤم)

Relative pronouns (الذي – التي – اللذان -اللتان – الذين – اللاتي - ما)

Demonstrative pronouns (هذا – هذه – هذان – هاتان - هؤلاء - أولئك)

Conditional conjunctions (إذا – مَن – ما – أينما – متى – أين – أنّى - كيفما)

Interrogative pronouns: (مَن – ماذا – متى – أين - كيف)

Adverbs: (أمام – خلف – وراء – فوق – تحت – قبل – بعد – الآن – يسار - يمين)

Verb (فعل): a content word that has a meaning itself and is associated with time. Subordinate classes include:

Past verbs (الفعل الماضي)

Present verbs (الفعل المضارع)

Imperative verbs (فعل الأمر)

Passive voice verbs (صيغة البناء للمجهول في الأزمنة الثلاثة)

Verbal infinitive (المصدر المؤوّل): a combination of "أن + الفعل المضارع" or "ما + الفعل الماضي"

Incomplete verbs (الأفعال الناقصة) which include:

Kana and its sisters (كان وأخواتها): (كان – أصبح – أمسى – ظل - بات – مازال – ما برح ....)

Kada and sisters (كاد وأخواتها): (كاد – أوشك – كرب – عسى – طفق – بدأ – شرع .....)

Particle: a stop word that has no meaning in itself, but it gets its meaning from the surrounding context. Subordinate classes include:

Verb-like particles (الحروف المشبّهة بالفعل): (إنّ – أنّ - كأنّ – ليت – لعل).

"لا" as a verb-like particle (لا النافية للجنس)

Negation particles (لا - لن – لم - ليس)

Conditional particles (لو- لولا - لوما - إن)

Vocation particles (يا – أ – أيا – هيَا - أي)

Prepositions (عن – على – في – إلى – بـ - في)

Exception particle (إلا)

Interrogative particles (أ - هل)

Coordinating Conjunction (و – ثم – أو – فـ)

Infinitive particles (الحروف المصدرية): they precede verbs to combine an indirect Verbal infinitive structure (ما - أن).

Other particles (..... إمّا – أمّا – سـ – قد – سوف)

Syntactic relations

Performing syntactic analysis in the dependency framework involves two inter-related steps: attachment and labeling (Žabokrtský and Smrž, 2003; Habash and Rambow, 2004; Habash et al., 2007; Smrž

et al., 2008; Tounsi et al., 2009). The attachment indicates that there is a syntactic relationship between two words, “the head word” (or governing) and “the dependent word” (or governed). The labeling specifies the type of each attachment by assigning “labels” or “relations” for each attachment. For example, the “subject” relation may label the attachment of a dependent noun to a heading verb, where the noun is the subject of the verb. This section provides a review of the different syntactic relations used in learning basics of Arabic syntactic analysis textbook. These relations can be divided into five categories (Nominative relations - Accusative relations - Genitive relations – modifiers – subordinate clauses)

Nominative relations

Active subject (الفاعل)	It marks the explicit syntactic active subjects of verbs
Passive subject (نائب الفاعل)	It marks the explicit syntactic passive subjects of verbs
Subject-predicate (خبر المبتدأ)	It marks the predicate of the subject regardless it appears before or after the subject
Subject of incomplete verbs (اسم كان وأخواتها)	It marks the subject of incomplete verbs (كاد - كان وأخواتها - ما الحجازية - وأخواتها)
Predicate of verb-like particles (خبر إن وأخواتها)	It marks the predicate of verb-like particles (لا النافية - إن وأخواتها للجنس)

Table1: the nominative syntactic relations and their references

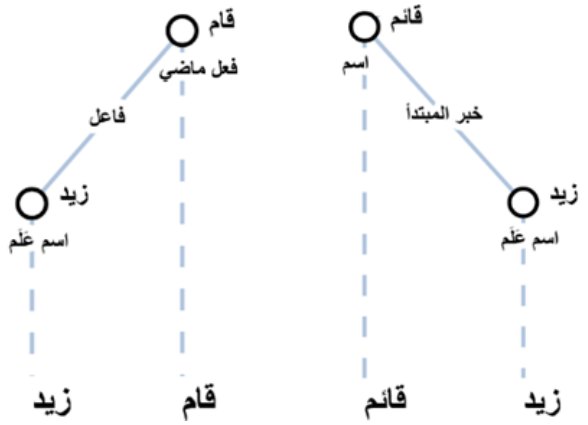


Figure2: Examples of nominative syntactic relations (Active subject "الفاعل" - Subject-predicate "خبر المبتدأ")

Accusative relations

Direct Object (المفعول به)	It marks the object of verbs and deverbal nouns
Absolute object (المفعول المطلق)	It marks the verbal noun (in the accusative form) that serves to emphasize the meaning of the main verb
Causative object (المفعول لأجله)	It marks the relation between the verb and the noun referring to the cause of doing this verb
Specification (التمييز)	It marks the specifier in the specification construction
Adverb (الحال)	It marks the noun that modifies a verb, and gives extra information about the manner of doing this verb

Subject of verb-like particles (اسم إن وأخواتها)	It marks the subject of verb-like particles (لا النافية للجنس- إن وأخواتها)
Predicate of incomplete verbs (خبر كان وأخواتها)	It marks the predicate of incomplete verbs (ما الحجازية - كاد وأخواتها- كان وأخواتها)

. Table2: the accusative syntactic relations and their references

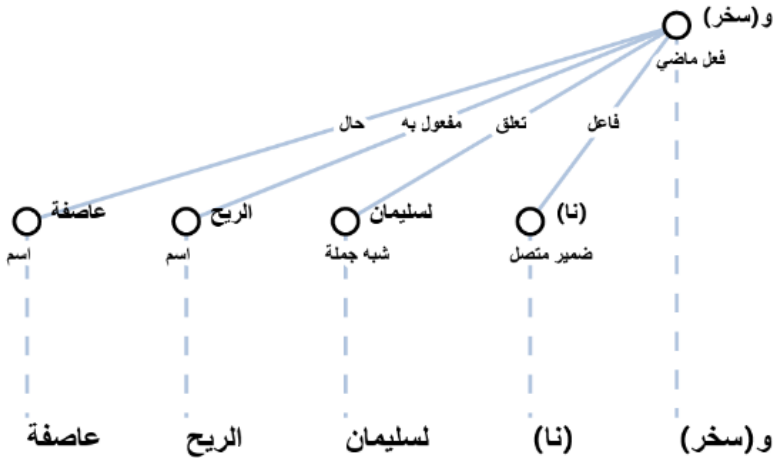


Figure3: Examples of accusative syntactic relations (Direct object "المفعول به" – Adverb "الحال")

Genitive relations

Possessive construction (الإضافة)	It marks the possessor in an "Edafa" construction.
Genitive noun (الجار والمجرور)	It marks the relation between the preposition and the following genitive noun (الاسم المجرور).

. Table3: the genitive syntactic relations and their references

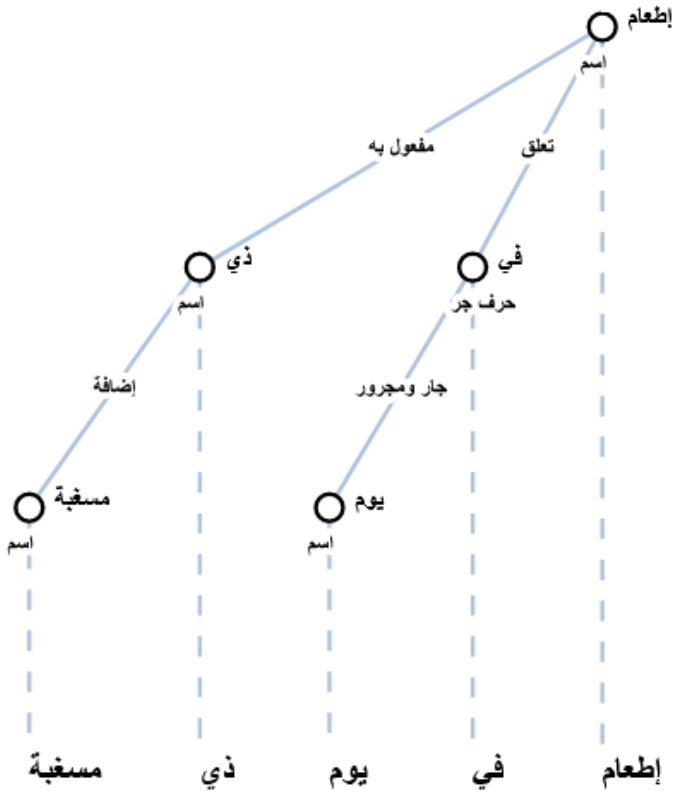


Figure4: Examples of genitive relations (Possessive construction "إضافة" – Genitive noun "جار ومجرور")

Modifiers

They refer to nouns that follow the preceding word in case ending and other morphological features. There are four modifiers in Arabic language, they are:

- Adjectival modification
- Apposition modification
- Emphasis modification
- Coordinating conjunction

Subordinate Clauses

Condition clause (جملة الشرط)	It marks the relation between the condition particle and the condition clause after it
Condition-consequence clause (جواب الشرط)	It marks the relation between the condition particle and the consequence clause.
Oath clause (جملة القسم)	It marks the relation between the oath particle and the oath clause after it
Oath-consequence clause (جواب القسم)	It marks the relation between the oath particle and the oath-consequence clause
Relative clause (جملة الصلة)	It marks the relation between the relative pronoun and the relative clause
Parenthetical clause (الجملة الاعتراضية)	It marks the relation between the head of the main sentence and the Parenthetical clause
Attachment clause (شبه الجملة المتعلق)	It marks the relation between the verb and the prepositional phrase attached to it

. Table4: subordinate clauses with their references

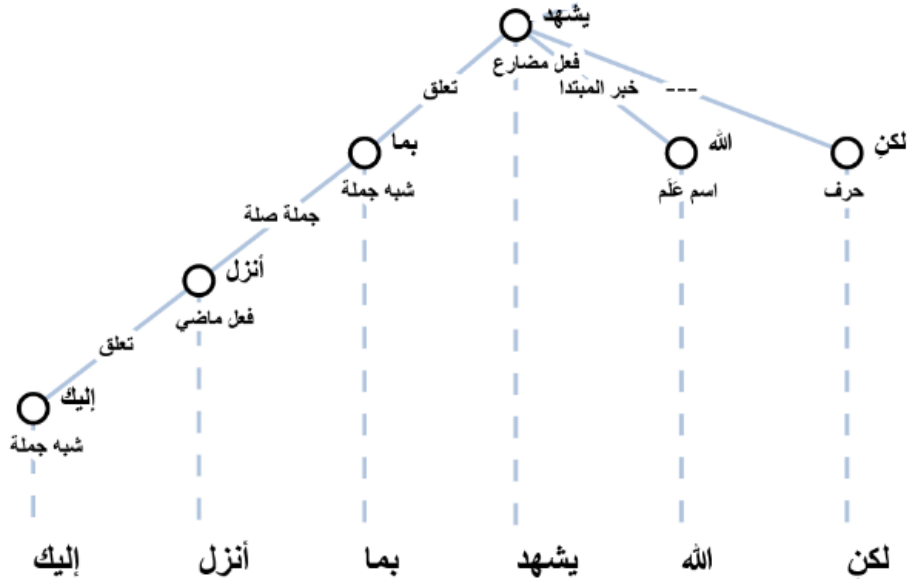
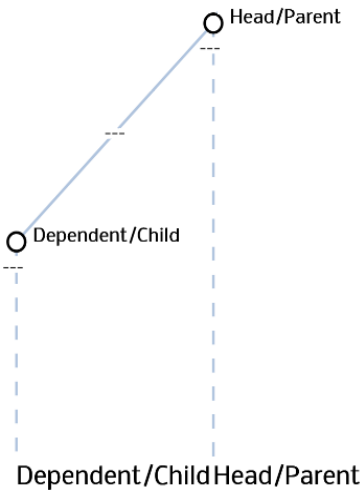


Figure5: Examples of subordinate clauses (attachment clause "شبه جملة متعلق" – relative clause "جملة صلة")

Head-dependent attachment

Each syntactic relation has a head/parent, and a dependent/child. In the tree, the circle parts are called nodes, and they represent a token in the sentence. Each node has the token name next to it, as well as the node's POS tag. The relation between the two nodes is on the line that connects them.



In Arabic, there are two types of sentences; verbal sentence and nominal sentence. Verbal sentence consists of a verb, usually followed by a subject, object, and other modifiers. In these sentences, the verb is the sentence head, and the subject, object, and other modifiers are dependents. The Basic Nominal Sentence consists of a subject and a predicate. In these sentences, the predicate is

the head of the subject, whether the sentence order is subject-predicate, or predicate-subject.

The Basic Nominal Sentence may be preceded by an incomplete verb (كان وأخواتها وكاد وأخواتها), a verb-like particle (إن وأخواتها), or functional negation particles like (ما الحجازية -لا النافية للجنس). In such cases, the preceding verb or particle heads both the subject and the predicate. Figure6: head-dependent attachment

Tokenization

Compared to English, Arabic has a rich and complex morphological system. Arabic base words inflect to eight features. Verbs inflect for aspect, mood, person and voice. Nouns and adjectives inflect for case, state, gender and number. Furthermore, inflected base words can attract various clitics. Clitical prefixes include determiners, particle proclitics, conjunctions and question particles in strict order. Clitical suffixes include pronominal modifiers. As a result of clitic attachment, morpho-syntactic interactions sometimes cause changes in spelling or pronunciations. Palmyra allows the user to fix the tokenization of the sentence, in terms of segmenting words correctly, normalizing them morpho-orthographically and correcting typographical errors

Segmentation

Affixes and clitics are marked with a “+” sign. The “+” signs are always attached to the affixes and clitics, and not the stem part of the word.

The following prefixes and proclitics are always segmented:

- Interrogative particle (أجئتم ← أ+ جئتم)
- Conjunctions (وقال ← و+ قال) - (فقال ← ف+ قال)

- Particles (كالحق ← ك+ الحق) - (بالله ← ب+ الله) - (لمدرسة ← ل+ مدرسة)
- Verbal modifier particles (ليكتب ← ل+ يكتب) - (سيكتب ← س+ يكتب)

The following suffixes and enclitics are always segmented:

- Pronominal enclitics - (رأيته ← رأيت +ه) - (بيوتكم ← بيوت +كم) - (يسمعني ← يسمع +ني)
- Relative and interrogative pronouns (ممَّا ← من +ما)

Normalization

Since decliticization can sometimes lead to malformed word forms, normalization is important for transforming the resulting tokens into their naturally uncliticized form. Words should be normalized to look the way it would if it didn't have any affixes/clitics attached to it. The following are some examples of this normalization:

● Alif-Lam (ال التعريف) : للكتاب (llktAb) 'for-the-book' → ل+ الكتاب (l+ AlktAb) (not ل+ لكتاب)

● Ta-Marbuta (التاء المربوطة) : مكتبتنا (mktbtnA) 'our library' → نا+ مكتبة (mktbħ +nA) (not مكتبت +نا)

● Alif-Maqsura (ألف التأنيث المقصورة) : مستشفاهم (mstšfAhm) 'their hospital' → مستشفى +هم (mstšfý +hm) (not مستشفا +هم)

● word-final Hamza (الهمزة المتطرفة) : بهائه / بهاءه / بهاؤه (bhAŵh / bhA'h / bhAŷh) 'his glory [nom./acc./gen.]' →ه+ بهاء (bhA' +h) (not بهاؤ ه+ / بهائ +ه)

● Waw of plurality (واو الجماعة): قالوها (qAlwhA) 'they-said-it' → قالوا +ها (qAlwA +hA) (not قالو ها)

Typographical errors

Some types of typos that you may find are:

- Punctuation written using numbers or letters:

"٠" instead of "." (full stop)

"ز" instead of "," (numeric commas)

- Punctuation that is not separated: "اليوم." ← "اليوم."
- Hamza variants (ا/أ/إ/آ) as in (الإسم ← الاسم) - (الى ← إلى)
- Ya/Alif Maqsura variants (ى - ي) as in ((مستشفي ← مستشفى) – (إلي ← إلى) – (فى ← في)
- Ta-Marbuta/ha' variants (ة/ه) as in (لة ← له) (سياره ← سيارة) -
- Missing spaces:

Negation particles, as in (مازلنا ← ما زلنا) - (لاسيما ← لا سيما) - (لابد ← لا بد)

Vocative particle, as in (يارجل ← يا رجل)

Between words, as in (رئيسالجمهورية ← رئيس الجمهورية)

- Extra spaces where they do not belong: رئ يس الجمهو رية ← (رئيس الجمهورية)

Conclusion and future work

The different approaches of teaching Arabic language unfortunately mainly focused on the accuracy of inflection, case ending, and the accuracy of the rules of Arabic syntax. Unfortunately, such tendency for teaching non-native speakers of Arabic is not practical enough or satisfactory without visualizing the syntactic relations between the head word and its dependents. Linking abstract rules with concrete head-modifier visualized relations makes the learning process

to be more practical, and the syntactic rules to be easy in understanding and learning. So, in this paper, PALMYRA is described as a graphical tool for visualizing and editing Arabic syntactic trees. The design of PALMYRA 2.0 enables it to be configured to work with all syntactic constructions. It has been specifically designed to support rich-morphology languages (like Arabic) and to visualize the complexities of syntactic constructions. It also allows for editing linguistic features through simple drop-down menus and text fields. Being built entirely using standard web technologies, PALMYRA runs on all major web browsers.

In the future, we plan to use PALMYRA as a part of an experimental study of comparing the performance of two groups of students; one group studies the Arabic grammar in the traditional approach, and the other studies in PALMYRA practical approach.

References

- Alena Böhmová, Jan Hajic, Eva Hajicová, and Barbora Hladká, 2003. The Prague Dependency Treebank, pages 103–127. Springer Netherlands, Dordrecht.
- Dima Taji and Nizar Habash. 2020. PALMYRA 2.0: A Configurable Multilingual Platform Independent Tool for Morphology and Syntax Annotation. In Proceedings of Universal Dependencies Workshop (UDW) 2020.
- Dukes, K., & Buckwalter, T. (2010, March). A dependency treebank of the Quran using traditional Arabic grammar. In *2010 the 7th International Conference on Informatics and Systems (INFOS)* (pp. 1-7). IEEE.
- Dukes, K., Atwell, E., & Sharaf, A. B. (2010, May). Syntactic annotation guidelines for the Quranic Arabic dependency

treebank. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.

- Heinecke, J. (2019, August). ConlluEditor: a fully graphical editor for Universal dependencies treebank files. In *Proceedings of the Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019)* (pp. 87-93).

- Javed, T., Habash, N., & Taji, D. (2018, May). Palmyra: A platform independent dependency annotation tool for morphologically rich languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

- Jan Hajic, Otakar Smrž, Petr Zemánek, Jan Šnaidauf, and Emanuel Beška. 2004. Prague Arabic Dependency ˇ Treebank: Development in Data and Tools. In Proceedings of the International Conference on Arabic Language Resources and Tools, pages 110–117, Cairo, Egypt. ELDA.

- Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal dependencies v1: A multilingual treebank collection. In Proceedings of the Language Resources and Evaluation Conference (LREC), Portorož, Slovenia.

- Lamia Tounsi, Mohammed Attia, and Josef van Genabith. 2009. Automatic treebank-based acquisition of Arabic LFG dependency structures. In Proceedings of the EACL 2009 Workshop on Computational Approaches to Semitic Languages, pages 45–52, Athens, Greece.

- Little, A. and Tratz, S. (2016). Easytree: A graphical tool for dependency tree annotation. In Proceedings of the Tenth

International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, 2016.

- Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn treebank. Computational Linguistics, 19(2):313–330.

- Mohamed Maamouri, Ann Bies, Hubert Jin, and Tim Buckwalter. 2003. Arabic treebank: Part 1 v 2.0. Linguistic Data Consortium (LDC2003T06).

- Najjar, M. (2020). Teaching Arabic Syntax for Non-Speakers: A Pragmatic Approach. *International Journal of Learning and Teaching*, *6*, 252-256.

- Naiwen Xue, Fei Xia, Fu-Dong Chiou, and Marta Palmer. 2005. The Penn Chinese Treebank: Phrase structure annotation of a large corpus. Natural Language Engineering, 11(02):207–238.

- Nizar Habash and Ryan Roth. 2009. CATiB: The Columbia Arabic Treebank. In Proceedings of the Joint Conference of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing (ACL-IJCNLP), pages 221–224, Suntec, Singapore.

- N. Habash, R. Faraj and R. Roth. 2009. Syntactic Annotation in the Columbia Arabic Treebank. In Conference on Arabic Language Resources and Tools, Cairo, Egypt.

- Nizar Habash, Ryan Gabbard, Owen Rambow, Seth Kulick, and Mitch Marcus. 2007. Determining case in Arabic: Learning complex linguistic behavior requires complex linguistic features. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), pages 1084–1092.

- Nizar Habash and Owen Rambow. 2004. Extracting a Tree Adjoining Grammar from the Penn Arabic Treebank. In Proceedings of Traitement Automatique du Langage Naturel (TALN-04), pages 277–284. Fez, Morocco.
- Otakar Smrž, Viktor Bielický, Iveta Kouˇrilová, Jakub Krácmar, Jan Hajic, and Petr Zemánek. 2008. Prague arabic dependency treebank: A word on the million words. In Proceedings of the Workshop on Arabic and Local Languages (LREC 2008), pages 16–23, Marrakech, Morocco.
- O. Smrž and J. Hajic. 2006. The Other Arabic Treebank: ˇ Prague Dependencies and Functions. In Ali Farghaly, editor, Arabic Computational Linguistics. CSLI Publications.
- Pajas, P. and Štepánek, J. (2008). Recent advances in a feature-rich framework for treebank annotation. In Proceedings of the 22Nd International Conference on Computational Linguistics - Volume 1, COLING '08, pages 673–680, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Salameh, M., Cherry, C., & Kondrak, G. (2013, June). Reversing morphological tokenization in English-to-Arabic SMT. In *Proceedings of the 2013 NAACL HLT Student Research Workshop* (pp. 47-53).
- Taji, D., Gizuli, J. E., & Habash, N. (2019). An Arabic dependency treebank in the travel domain. *arXiv preprint arXiv:1901.10188*.
- Zdenek Žabokrtský and Otakar Smrž. 2003. Arabic syntactic trees: from constituency to dependency. In Proceedings of the Eleventh Conference of the European Chapter of the Association for

Computational Linguistics (EACL'03) – Research Notes, Budapest, Hungar

المستخلص

يهدف هذا البحث إلى تعليم أساسيات التحليل النحوي العربي للناطقين بغير اللغة العربية بالأساس، وقد استخدم الباحث في هذا البحث التحليل النحوي التركيبي للجمل والمركبات العربية (Phrase Structure Grammar)، وذلك من خلال استخدام أحد تطبيقات التمثيل النحوي الشجري (PALMYRA 2.4)، وهو أداة تمثيل مستقلة لوسم وترميز أقسام الكلام والمركبات النحوية المختلفة، ومن مميزات هذه الأداة أنها تدعم اللغات التي تتطلب ترميزاً مورفولوجيا معقداً مثل اللغة العربية. وقد تم بناء هذا التطبيق بحيث يكون متوافقًا مع أي تمثيل تحليلي نحوي تبعي (dependency parsing representation)، ويستطيع هذا التطبيق أيضًا ترميز العديد من الخصائص اللغوية الأخرى (linguistic features) في النص، كما أنه يستخدم واجهة (interface) تعتمد على أدوات السحب والإفلات المساعدة.

الكلمات الدالة: التحليل النحوي– أداة PALMYRA– اللغة العربية– أقســـام الكلام–العلاقات النحوية.

تعليم أساسيات التحليل النحوي العربي باستخدام أداة التمثيل النحوي الشجري PALMYRA

إعداد

د. أحمد عبد الغني محمد عبد الغني

مدرس بقسم الصوتيات

كلية الآداب- جامعة الإسكندرية

أستاذ مساعد بكلية الدراسات اللغوية والتنمية البشرية-

جامعة ماليزيا كلنتان - ماليزيا