

Investing in artificial intelligence algorithms in processing
the Arabic language at the grammatical level

Imane HASSANI ¹⁽¹⁾

Language and communication lab. Ahmed Zabana University - Relizane,
(Algeria)

imane.hassani@univ-relizane.dz

brahimi BOUDAUD ⁽²⁾

Ahmed Zabana University - Relizane, (Algeria)

brahimiboud@gmail.com

Sending date :27/08/2024	Acceptance date :05/12/2024	date of publication :21/12/2024
------------------------------------	---------------------------------------	---

Abstract:

This study aims to employ the most important neural algorithms that contribute to processing the Arabic language automatically through the analysis and generation features, in order to shed light on their effectiveness at various linguistic levels, especially at the grammatical level, which must be simplified in the best neural smart ways, to facilitate neural smart ways, to facilitate its understanding and simplification for the computer.

Keywords: artificial intelligence, algorithms, analysis and generation, automatic processing, grammatical level

ملخص :

تهدف هذه الدراسة إلى توظيف أهم الخوارزميات العصبية التي تساهم في معالجة اللغة العربية بشكل آلي من خلال خاصيتي التحليل والتوليد، وذلك لتبسيط الضوء على فعاليتها في مختلف المستويات اللغوية، وخاصة على المستوى النحوي الذي يجب تبسيطه بأفضل الطرق الذكية العصبية، لتسهيل فهمه وتبسيطه للحاسوب.

الكلمات المفتاحية: الذكاء الاصطناعي، الخوارزميات، التحليل والتوليد، المعالجة الآلية، المستوى النحوي

¹Imane HASSANI

1. Introduction

Artificial intelligence technology first emerged in the 1950s, marking a turning point in human history, particularly in contemporary cognitive sciences. The goal was to simulate human intelligence by leveraging its systems and applications in all linguistic aspects, especially in oral and written language processing, using advanced techniques known as neural algorithms.

Thus, it is crucial to unravel the possibility of exploiting artificial intelligence algorithms in natural language processing applications automatically and to understand to what extent these algorithms in syntactic analysis are effective?

Therefore, researchers and specialists in computational linguistics have sought to use these computational neural techniques. To achieve this, we must innovate various smart applied methods to help us be more open to the information world.

Artificial Intelligence:

Artificial intelligence aspires to mimic human behavior by searching for techniques that grant artificial machines the capability to reach the level of human intelligence and understand its nature by creating computer applications capable of representing intelligent human behavior (Bonnet, 1993, p. 11).

That is, the ability of these machines to find solutions to certain problems based on the inferential data previously entered into the programs. It is a sophisticated term used to describe challenges that are hard to solve by computers. Therefore, scientists of human intelligence use multiple techniques to develop the level of computer intelligence, such as modeling and simulation to obtain programmable models of human intelligence (Talba & others, 1994, p. 29)

Artificial intelligence, then, is the science that deals with advanced systems that display properties similar to human behavior and includes both theoretical and practical aspects such as perception, natural language processing, reasoning, planning, learning, adaptation, logical inference, etc. It relies on a set of distinctive sciences such as computer science, engineering, mathematics, and linguistics to innovate and develop algorithms and create applications in machines to acquire intelligent behavior akin to human mental skills.

As a result of what has been mentioned, artificial intelligence aims to create a model that highlights a new trend in modern technological techniques. Scientists and researchers in this scientific field aspire to innovate smart applications and systems by: (Al-Omari, 2015, pp. 95-96)

1. Building a database that supports the ability to organize information and focus on empirical rules that are almost unavailable in sources.

2. Building a solid mechanism free from human emotions such as fear, disorder, and fatigue.

Foundations and Types of Artificial Intelligence:

Generally, the science of artificial intelligence relies on two main foundations (Mohammed, 2024):

A. Data Diagnosis: Diagnosing the problem into a form that the computer can understand and think about solving, meaning converting it into an understandable and analyzable form.

B. Intellectual Exploration: Examining the options available to the computer and evaluating them according to predefined criteria or those it inferred itself, then making the optimal decision for the solution.

Based on these two points, it is divided into four main categories: (Ali Hegazy)

1. Reactive Machines: These are interactive devices that operate based on their current experiences and do not store any part of the memory or past experiences.

2. Limited Memory: Certain parts of the memory are stored that help in interaction according to this memory, although they possess short-term memory.

3. Theory of Mind: This type is the most intelligent, interacting based on its understanding of people's thoughts and feelings and anticipating how they will treat it.

4. Self-awareness: This is an extension of the third type, Theory of Mind, and the most advanced, referring to the awareness shown in recognizing needs and internal states.

Artificial Intelligence Systems:

A. Expert Systems: These programs transfer human knowledge and expertise to the computer, enabling it to perform tasks that non-specialists in this field cannot execute. The idea is to feed the computer with as much knowledge as the specialist holds, then deal with this knowledge using search and inference tools to obtain results similar to those obtained by the human expert. (Al-Husseini, p. 3)

Thus, expert systems rely on a knowledge base containing field-related knowledge and facts. The specialized system focuses on understanding, analyzing, and processing problems rather than focusing on its computational efficiency.

B. Artificial Neural Networks: These computational applications are designed to mimic the way the human brain performs specific tasks. This is done through extensive parallel processing divided into simple processing units. These units are just computational elements called neurons or nodes with neural properties, storing practical information and experimental data by adjusting weights, making them available to the user. (Al-Hassanab, p. 3)

Thus, they act as a bridge to achieve communication between advanced applications in the field of artificial intelligence, thanks to their effective role in enhancing data processing and analysis efficiency and speed, becoming a prominent tool in both applied and theoretical fields.

C. Automated Natural Language Processing: Artificial intelligence scientists use the term natural language to refer to the languages used by humans, distinguishing them from unnatural languages like programming and computer languages that facilitate communication between humans and machines and among different machines.

Automated natural language processing involves a set of computational techniques aimed at analyzing and representing texts in natural languages at one or more levels of linguistic analysis to achieve a similarity that the machine can execute in a manner closely resembling natural language (Al-Dakhissi, 2020, p. 25). From a computational perspective, it relies on producing informational applications capable of processing natural languages automatically. (pierrete, 1998, p. 27)

Accordingly, computational linguistics is a branch of computer science and artificial intelligence concerned with developing algorithms that enable machines to interact with, understand, and generate human language.

Natural Language Understanding: This involves understanding and analyzing the syntactic and semantic aspects of spoken and written texts, making the machine capable of understanding the pattern of natural language and its inputs.

Natural Language Generation: In script programming, written or spoken texts are generated based on input data that mimics human behavior, meaning the machine interacts with language as an output of information. This system relies on the following tasks in natural language processing: (Al-Wazra, 2024)

1. Text Preparation: Importing relevant content from the database and organizing it.

2. Sentence Planning: Choosing appropriate words and forming meaningful phrases, considering punctuation and linking tools.

3. Text Realization: Preparing a composed text of multiple sentences while adhering to recognized linguistic rules and maintaining grammatical accuracy.

D. Intelligent Agent Systems: These are computer applications that achieve specific goals in a changing environment on behalf of other entities (whether computational or human) over a specified period without the need for direct and continuous supervision, aiming to convert goals into tasks. (krupansky, 2024)

It is certain that information derived from the surrounding environment is integrated, interpreted, and understood using data processing techniques, with plans and decisions made based on the algorithms used in executing these operations.

Therefore, capabilities are unified to provide effective and dynamic interaction with complex environments through harmonious and coherent voice recognition and understanding its natural language.

E. Fuzzy Logic Systems: This method relies on awareness and simulating the way humans perceive values using clear data.

When we talk about uncertain and not fully embodied knowledge in reality, we refer to a set of ideas and techniques used to express or prove it. (Yassin, 2017, p. 194)

Fuzzy logic techniques and systems are used with other integrated systems that operate with artificial intelligence techniques.

Algorithms:

Concept of Algorithm:

Algorithms are associated with the Arab mathematician Ja'far ibn Musa al-Khwarizmi, particularly in the field of intelligent electronic programming. They form a set of organized rules, starting from simple rules to complex ones, and then to even more complex rules (Ali, 1988). From this definition, it is clear that an algorithm is a set of specified rules and methods used to simplify mathematical symbols and transform them into linguistic rules that can be used and applied. Algorithms are based on two essential conditions:

1. Inputs: The values to be worked on should appear as inputs, which can be zero or more.

2. Outputs: The algorithm should produce expected actual results from its execution, with at least one value. (Obaidullah, 2024)

Algorithm Description:

There are several methods to explain algorithms or to solve and process a particular issue. These methods vary from individual to individual in terms of simplicity and precision. Some of these methods include (Al-Mousa, 2010, pp. 77-76) (Mariam, 2022, p. 25):

1. Writing algorithms in step-by-step format using a common language such as natural language, English, or other languages.
2. Writing algorithms using geometric bases, meaning representing them with agreed-upon diagrams, the most famous being flowcharts.
3. Writing algorithms using symbolic language represented in programming languages like C++ and English.

Grammar is a meeting point between linguistics and computer science. It seeks to solve the problem of linguistic engineering within the framework of knowledge, i.e., transferring mental knowledge to a coherent system compatible with the nature of artificial intelligence in processing natural languages. This is done in light of the newly created and available sources through multiple descriptive techniques (Al-Mousa, 2010, p. 71) (Abdul Rahman, 2009, p. 71). Therefore, grammar is one of the basics of linguistic engineering, playing a crucial role in any natural language processing operation, being the first step in the automated field.

Investing Algorithms for Processing Arabic Language at the Grammatical Level:

Automatic processing of the Arabic language aims to create an intelligent encyclopedia of linguistic studies that contributes to the application of scientific methodology and meets its automatic needs in all aspects, especially the grammatical aspect, to design a computer program that considers linguistic features. (Muneim, 2015, p. 181)

Stages of Automatic Processing of Syntactic Units:

This process involves several stages, starting from the original text and ending with a comprehensive syntactic and grammatical analysis of the sentence:

1. Text Segmentation: Dividing the text into small parts known as symbols or words.

2. Syntactic Analysis: Identifying grammatical elements of sentences such as root, pattern, suffixes, and prefixes.

3. Determining Grammatical Parts: Identifying the grammatical relationship between words in sentences.

4. Determining Structural Relationships: Such as the relationship between the verb and subject or the adjective and the noun.

5. Understanding Context: Understanding the overall context of sentences in terms of implicit and explicit meanings.

6. Determining Types of Syntactic Units in Sentences: At the grammatical level, it involves parsing and analyzing sentence structure, identifying grammatical functions of sentence components, and providing automated parsing.

It is controlled by a group of factors that include the following: (Al-Najm)

A - The linguistic segment covered by the automated system: It is based on the following:

- _ Identifying and classifying grammatical structures.
- _ Highlighting lexical vocabulary.
- _ Identifying grammatical forms.

Like: He and she who did well won the scholarship (Ali, 1988, p. 404).

Thus, the linguistic chip describes data and data for the computer to recognize and process.

Through these factors, it is determined:

- _ Syntactic analyzes of all the words that make up the sentence.
- _ The origins of words and their weights.
- _ Grammatical structures and prefixed and suffixed forms that give words their value.
- _ Highlighting the structural relationships according to the levels of grammatical rules that determine the genetic nature of words.

B_ Complexity system: It is based on two principles:

The first principle: The nature of the grammatical system used.

The second principle: The linguistic direction that the grammatical system focuses on, whether it is morphological grammatical or semantic grammatical. Based on this, the method of programming it is determined. (Ali, 1988, p. 404)

Through these factors, it is determined:

- _ Syntactic analyzes of all the words that make up the sentence.
- _ The origins of words and their weights.
- _ Grammatical structures and prefixed and suffixed forms that give words their value.
- _ Highlighting the structural relationships according to the levels of grammatical rules that determine the genetic nature of words.

Show quoted tax.

Algorithmic Grammar Analysis (Labbes, pp. 24-25):

The analysis is performed at two levels: (Al-Zakan, p. 33)

1. At the Single Word Level:

- ***Derivation or Generation:*** Transitioning from the root of the word to all words derived from it.

- **Analysis:** Transitioning from the word to its root, and determining its morphological, grammatical, and semantic status.

Moreover, the word determines the form of the grammatical processor, whether it is in the past, present, or imperative with its various forms: indicative, subjunctive, jussive, the emphasized, and the non-emphasized, the known and the unknown, the constructed, and the declined, the visible and the estimated. (Al-Zakan, p. 33)

The noun is the bare noun, but changes occur if more is derived from it. Additionally, there is the letter, which is the prefix and suffix attached to the verb and noun (Labbes, p. 19). Therefore, analysis at the word level depends on the property of the trilateral root, determining the pattern, and the nature of the word, whether it is a noun or a verb, highlighting its grammatical sign.

2. At the Sentence Level:

- Grammatical and Morphological Analysis of the Word: It relies on other linguistic parsers from morphology, phonology, and semantics through:

- Analyzing the Sentence Structure.
- Determining Structural Relationships.
- Parsing Words Within the Sentence.
- Determining Functions of Sentence Components. (Dawood, 2001, p. 275)

At the grammar level, the structure of sentences can be analyzed, parsed, and automatically generated. An example is the word: insertion.

- Analysis Elements: A + entered + the alif.
- Morphological Form: Af'al.
- Word Type: Source.
- Grammatical Case: Nomination.
- Word Root: D-Kh-L.
- Meaning: Highlighting a specific thing.

From this example, we conclude that the grammatical parser relies on:

- Determining Sentence Elements.

- Identifying Morphological Form, Source, Root, and Grammatical Case.
- Highlighting Its Position in the Sentence.

Thus, automated parsing occurs within the sentence as a whole.

Parsing Algorithms:

The focus is on names, whether inflected or built, and one form of verb, which is the present tense. The past tense and imperative are originally built, while nouns and present verbs share similar positions in nomination and accusative. Nouns uniquely have the genitive position, while present verbs are characterized by the jussive position. (Laqam & others, p. 3948)

Parsing Algorithms Include: (Ali, The Arabic Language and the Computer, 1988, pp. 404-405)

1. Database for Parsing:
 - A database for all names that express activity with the nominative present verb.
 - A database for all names that express objectivity with the accusative present verb.
 - A database for all names that express addition.
 - A special database for the jussive present verb with particles or nominal tools.
2. List of All Inflected and Built Words.
3. List of All Inflected Nouns and Verbs.
4. List of All Particles, Past Tense Verbs, Imperative Verbs, and Built Nouns That Require One Form in All Their Forms.
5. Morphological Description for Grammar Sections.
6. List for Organizing Final Vowel Marks of Words (Dual, Masculine Sound Plural, Five Nouns, Five Verbs, Jussive Present, Weak-Ending Verbs).
7. Automated Diacritical Marking of Linguistic Texts.
8. Spelling and Grammatical Checkers.

Each of these stages requires advanced techniques to deal with the complexities of the Arabic language. Executing these steps sequentially achieves

precise syntactic analysis, contributing to the understanding and analysis of texts according to the built-in data.

Grammar Generation Algorithms:

These are one of the main components for text generation. They are one of the applications whose importance grows with the expansion of artificial intelligence methods and knowledge organization rules (Al-Ansari, p. 271). The grammar generator is divided into two types:

1. Lexical Grammar Generation: It includes the lexical form of words and their various classifications (nouns, verbs, adjectives, etc.), creating linguistic rules to determine the syntactic structure of sentences.

2. Linguistic Grammar Generation: It aims to generate the grammatical structures of sentences as they exist in the linguistic text, determining how words interact with each other within a single sentence. This type is used in automatic translation and text generation.

To build an accurate grammatical generator, several aspects need consideration:

1. ***Classification of Words and Sentences.***
2. ***Understanding Contextual Usage.***
3. ***Providing Synonymy and Antonymy.***
4. ***Creating Linguistic Patterns and Models.***

1_ Displaying the text to be analyzed into linguistic rules and laws that were previously prepared and programmed in the automated device, then applying them to the text to be analysed.

2_ Storing a large amount of manually analyzed text in the computer, to serve as an example against which to measure it. The entered text is displayed on those models for comparison and the correct probabilities are deduced according to the levels of agreement.

After completing the previously mentioned stages, the results, words and sentences, are collected in tables and cells in the form of matrices, and a comprehensive structural model of all grammatical data is built.

Conclusion:

Ultimately, we arrive at a main idea: investing artificial intelligence algorithms in language levels, especially the grammatical level, has become necessary to keep pace with technological development and benefit from the productions of computational linguistics.

The research reached a set of results, which we summarize as follows:

_Artificial intelligence aspires to create automated models that closely resemble the human brain and mimic its behavior.

_ Automated processing of the Arabic language aims to create graphical bases and develop educational tools that contribute to simplifying linguistic rules, focusing on aspects of the language related to pronunciation, clarity, grammatical rules, and meanings.

_Syntactic analysis focuses on the linguistic parts within sentences and classifying them, with the aim of developing automated interfaces that interact naturally and smoothly with texts.

_ Grammatical processing depends on two principles: entering data (analysis) and outputting results (generation).

Grammatical analysis algorithms are based on linguistic databases that contribute to analyzing words and dividing them into basic units by understanding their linguistic structure. As for the generation process, it is concerned with producing linguistic knowledge rich in vocabulary, phrases, and texts.

Based on this, algorithms at the grammatical level seek to show the grammatical rules to achieve the artificial network in the Arabic language, by creating linguistic networks and converting linguistic units into mathematical symbols that the computer can understand.

What we see today in many automated programs is a lack of interest in the grammatical aspect, which is one of the basics of automated processing, and as a result, there is a lack of interest in communication between humans and machines.

Automated syntactic and grammatical processing of the Arabic language requires advanced tools and techniques to handle its complexities. These processes involve multiple stages, from simple text segmentation to comprehensive syntactic analysis, and require precise algorithms to achieve accurate and meaningful results. The development of these algorithms is crucial

for enhancing Arabic language processing in natural language understanding, machine translation, and other AI applications.

Automated syntactic and grammatical processing of the Arabic language requires advanced tools and techniques to handle its complexities. These processes involve multiple stages, from simple text segmentation to comprehensive syntactic analysis, and require precise algorithms to achieve accurate and meaningful results. The development of these algorithms is crucial for enhancing Arabic language processing in natural language understanding, machine translation, and other AI applications.

References:

Abdul Rahman, A. R. (2009). Introduction to Computational Linguistics0 1st edition. Saudi Arabia: Wajuh Publishing and Distribution.

Al-Ansari, A. M. (n.d.). The linguistic and grammatical lesson in light of the computer. D. T.

Al-Dakhissi, A. K. (2020). Computational Linguistics and the Automatic Processing of the Arabic Language. Irbid, Jordan: Modern Book World.

Al-Hassanab, Q. A.-H. (n.d.). in Artificial Neural Networks. Iraq: Imam Jaafar Al-Sadiq University.

Al-Husseini, Q. H. (n.d.). On Artificial Neural Networks.

Ali Hegazy, M. O. (n.d.). Artificial Intelligence.

Ali, N. (1988). The Arabic Language and the Computer. Kuwait: Publications of Arabization.

Ali, N. (1988). The Arabic Language and the Computer. Cairo: Osama Al-Khouli, Arabization for Publishing and Distribution.

Al-Mousa, A. A. (2010). Introduction to Computers and the Internet. King Faisal University, Al-Ahsa, 6th edition.

Al-Najm, S. (n.d.). Computational Linguistics.

Al-Omari, G. I. (2015). The Role of Information Technology and Knowledge Management in Building Organizational Memory, Arab Journal for Security Studies. Saudi Arabia: Naif Arab University for Security Sciences,, Vol. 27, Issue 52.

Al-Wazra, A. (2024, 05 24). Basics of Natural Language Processing. Kingdom of Saudi Arabia: TuwaiqWebnyar Academy. <http://tuwaiq.edu.sa>, www.youtube.com/@tuwaiqAcademy.

Al-Zakan, M. A. (n.d.). Linguistics and Arabic Language Programming in the Computer, Using the Arabic Language in Information Technology, ed. Mohamed Al-Hanash. International Journal of General Linguistics.

Bonnet, A. (1993). Artificial Intelligence – Its Reality and Future, trans. Ali Sabri Farghali. Kuwait: Publications of the World of Knowledge.

Dawood, M. M. (2001). Arabic and Modern Linguistics. Cairo: Gharib Publishing and Distribution House.

krupansky, J. (2024, 05 12). what is a soft ware Agent? <https://activity.com/agdef.htm>.

Labbes, J. T. (n.d.). Informatics and Language, Literature, and Civilization.

Laqam, A. A., & others. (n.d.). Computing the Arabic Language Between Reality and Aspirations – A Proposed Methodology for the Arabic Language Departments at Prince Sattam bin Abdulaziz University. Journal of Arabic Studies, Faculty of Dar Al Uloom, University of Minya.

Mariam, C. (2022). Lectures on the Subject of Computer Science. Algorithms.

Mohammed, K. (2024, 5 8). The Four Types of Artificial Intelligence. available on the website, accessed on, at 13:09.

Muneim, S. (2015). Computational Linguistics and Machine Translation – Some Theoretical and Practical Constants, 1st edition. Jordan: Modern Book World.

Obaidullah, A. F.-A. (2024, 06 16). Computer Algorithms. www_alaarabmag.com.

pierrette, B. (1998). traitement automatique des langues naturelles. Paris_Brucellosis: Aupelf-uref-Ed-Duculot.

Talba, F., & others. (1994). Computer and Artificial Intelligence . Modern Egyptian Office Press, Delta Computer Encyclopedia.

Yassin, S. G. (2017). Decision Support Systems. Amman, Jordan: Dar Al-Manhaj for Publishing and Distribution.