A Survey and Classification of Arabic Logic-Based Meaning Representations (ALMR)

Haytham El-Sayed

Abstract—This paper provides a comprehensive survey and classification scheme of attempts to provide logical treatment to Arabic semantics, thereby developing Arabic semantic processing, during the past thirty years. The efforts of Arabic logic-based meaning representation are very limited. However, Badr Al-Johar and Jim McGregor (1997), Bassam Haddad and Mustafa Yaseen (2001-2007), and Haytham El-Sayed (2011, 2015), can be considered good starting points. Shading the light on these works structure and limitations is needed for better improvements, and developing an adequate logic-based model, of semantic processing for Arabic. Establishing a research community that combines Arab logicians, linguists and computer scientists has become increasingly necessary for high level, and amount, of achievements in this prominent interdisciplinary research area.

Index Terms—Arabic LMR, syntax-semantics representation, Arabic semantic processing, Montague grammar.

I. INTRODUCTION

With the development of computational models of semantics, requirement of using logic in the semantic area of Natural Language Processing (NLP) is progressively prominent. In English, for instance, it has become increasingly difficult to ignore that logic (i.e., its tools and formal languages) does well enough in providing adequate meaning representations for natural language expressions. Thus, semantically logic-based computational systems have been established (e.g., [1]-[7]). Consequently and unsurprisingly, most logic-based meaning representations (LMRs) are based on English.

In contrast, the level and amount of achievements in Arabic semantically logic-based computational systems is regrettably different. That is, the “most effort has been extensively focused on morphological analysis, moderate on syntax, and relatively poor on semantics” [8]. However, what is called here Arabic Logic-Based Meaning Representation (henceforth ALMR) has significantly been given attention by Arab logicians and computer scientists alike. Especially during the last three decades, various ALMRs, towards Arabic semantics processing, have been provided. [9]-[16] can be regarded as the visible achievements so far.

Although a variety of ALMRs have been provided to a variety of linguistic items, virtually all of them seem to be relevant to the field of Arabic logic-based computational linguistics. Among other techniques, they involve syntactic analyses, Discourse Representation Theory (DRT), Montague Grammar (MG). Despite the apparent differences in the linguistic units treated and the logical tools or formal theories applied, it seems sensible to put all these representations under the same umbrella. To bring order to the variety of these representations, a general classification scheme is presented in this work. That is, a comprehensive survey of existing ALMRs is given, listing and describing attempts from 1997 until today.

Generally, this work has several, partly overlapping, goals, ranging from purely disciplinary (i.e., formal logic) to more interdisciplinary (i.e., logic and computer science) objectives, listed in this order:

- To give logicians, linguists and computer scientists a better understanding of the nature of current ALMRs.
- To provide a starting point for researchers, in these fields, interested in ALMR.
- To encourage Arabic NLP developers, especially in semantics area, establish inter-disciplinary research community.

The remainder of this paper is organized as follows: the next section presents some relevant background related to the research problem. In Sect.3, I present the survey that I have done and the results I have obtained. My proposed classification scheme is described in Sect.4. In the last section, I conclude my work and highlight some possible future work.

II. RELEVANT BACKGROUND

In general, in the fields of logic and linguistics, semantics is the study of the meaning of linguistic expressions. In logic, specifically in formal semantics, which is the “study of the meaning of natural language expressions suing tools and languages of symbolic logic” [17], natural languages have had well-established formal representation many years ago. As well as, in linguistics, specifically in computational linguistics, semantics is the level at which language makes contact with the real world. Thus, semantics is regarded as one of the linguistic levels that logic and computational linguistics can meet and interact.

Logicians and computational linguists share a common interest in, and challenge for, the logical treatment of semantics as a pre step towards its proper computation. However, whereas semantic processing is the most important part of natural language processing, it is the most difficult at the same time; as the natural language is not easily amenable to receive formal representation. In other words, most agree that, the construction and composition of
meaning representation formalisms for natural language plays a decisive role in the whole semantic processing progression. However, due to the higher complexity and subtlety of semantics, as [18] shows, “the amount of research in computational models of semantics is much smaller than other areas of NLP”.

If it is the case in English, the situation in Arabic is much disappointed. Namely, if Arabic is compared with other languages, it has received much less modern formalization interest. That is, during the past forty years, Arabic linguists and computer scientists merely focused on Arabic formalization and computation from the morphological and syntactical point of view (e.g., [19]-[23]). No doubt that the achievements in this level are really prominent. However, for the reason that semantic processing is highly important for achieving the understanding capability, much work need to be done on semantic representation and semantic analysis of Arabic.

To illustrate, justifications of the low level of research achievements in the area of Arabic computational models of semantics are expressed in the following quotation from [24]:

“No formal descriptions of Arabic syntax structure, neither those accounting for underlying relationships and dependencies, nor those enriched with a ‘semantic-feature’ level, go beyond the sentence as linguistic unit of description. Formal text grammar, coherent, consistent, and adequate description of semantic features, and outlines for the formal description of general language pragmatics and discourse particularities are almost non-existent” [p.458].

While waiting for the breakthrough, mention should be made of some early initiatives in the right directions. [9]-[16] are the visible attempts so far. Let us start to explore their overall structures and limitations in providing ALMRs, in a chronological order.

III. SURVEY ON ALMR

Over the past thirty years, several attempts to provide Arabic logic-based meaning representation have been done by computer scientists, in addition to one attempt has been done by an Arab logician, while the Arab linguists were absent from the scene.

In 1997, the computer scientist at the University of Sheffield Jim McGregor and his student Badr Al-Johar introduced, for the first time, a research concerned with developing an Arabic natural language interface to database systems, using the intermediate meaning representation approach. For that reason, they build ALMR notation as a representative for this approach for the Arabic language. As stated later in Al-Johar’s PhD thesis, “the syntactic treatments are based on Generalized Phrase Structure Grammar (GPSG), whereas the semantics are expressed in formal semantics theory” [25].

It worth noting that, several approaches had been proposed for semantics before McGregor and Al-Johar, and very few of them applied to Arabic. For instance, in the eighties, an Arabic Language Interpreter (ALI) built by Saad Mehdi. “He used the Semantic Marker and Selectional Restriction (SMSR) approach in the definitions of each sense of the word in the dictionary to represent the meaning of the sentence” [1, p.32].

Beside the technical limitations, all of the above systems have the following ‘reprsentational’ problems: (a) none of them can handle quantifiers, (b) none of them build a complete independent meaning representation for the whole query. More on this later.

At the beginning of the third millennium, there were another attempts done by computer scientists as well. From 2001-2007, Bassam Haddad and Mustafa Yaseen, from Petra University, have published the following research contributions1:


Based on investigation, one can claim that all of these works are the same, and can be considered as ‘one’ research contribution. There is no change in the content, and the only one difference between these publications is that they have been presented in different years, occasions and publication formats. All of these works address issues related to employing logic-based semantic composition as a meaning representation for Arabic within a unification-based syntax-semantics interface.

Four years later from Haddad’s latest publication, the Arab logicians entered the scene for the first time. In 2011, Haytham El-Sayed provides a Montagovian syntactic-semantic formalization of a limited portion of Arabic quantified expressions. By offering such formalization, El-Sayed claim that success in formalizing a fragment of Modern Standard Arabic could be a prominent pre-step towards its logic-based computation. Later, in [16], El-Sayed argue that, giving the extensive applicability of the Montagovian analysis to variety of Arabic quantified expressions, and its use as a real and efficient tool in natural language processing, one can claim that developing the Montagovian apparatus, or may using other categorial grammars, can be regarded as a one way out of two current challenges of Arabic, which are: (a) syntactic-semantic formalization, and (b) computation on logical or non-statistical level, at the same time [p.25].

IV. CLASSIFICATION SCHEME

Let us now turn to the nature of ALMRs. As it turns out, however, the survey provided in the previous section

1 For the five research contributions listed above, three of them had been offered by both Haddad and Yaseen, in a co-authorship format, whereas the rest had been solely offered by Haddad. However, one cannot see any considerable differences or advances in Haddad’s individual research to the extent that one cannot deal with the attempts as separate contributions.
chronologically lists the ALMR attempt over the past thirty years, and not so much about categorization of the attempts and their clear-cut properties. For that reason, a classification scheme is introduced in this section to categorize and describe the fundamental nature of each attempt.

The attempts mentioned above could be classified in three categories:
- syntactic meaning representation: [B. Al-Johar and J. McGregor];
- surface semantic representation: [B. Haddad and M. Yaseen];
- Analytical syntax-semantics representation [H. El-Sayad].
- In what follows, a description of each category is introduced.

A. Al-Johar and McGregor

In [9], Al-Johar and McGregor provide a syntactic meaning representation for a range of Arabic syntactic categories. Their representation of a linguistic expression is first processed syntactically through a parse tree, based on a set of syntax rules. As [9] shows, following analysing the syntactic component of the expression, the tree analysis is then transformed to the intermediate logical query by the semantic interpreter, thereby producing the database query [p.31].

Going into detail, for the purpose to develop an Arabic interface to database systems, Al-Johar and McGregor build LMRA notation for the sake of intermediate meaning representation. According to [9]:

“The mapping to database information specifies how logic predicates relate to database objects. In the case of an interface to a relational database, the simplest approach would be to link each logic predicate to an SQL statement” [p.36].

Based on Al-Johar and McGregor (1997), the syntactic categories in use are: proper noun, mammal common noun, non-mammal common noun, noun with adjective, intransitive and transitive verbs, and preposition. This means that the formalization accommodates a good range of Arabic linguistic expressions. To illustrate, sample of the final output of Al-Johar and McGregor’s LMRA is formulated as shown in the table below [quoted from 1, pp.37-38].

<table>
<thead>
<tr>
<th>Words and Phrases</th>
<th>LMRA logical Formulas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proper noun</td>
<td>logical constant</td>
</tr>
<tr>
<td>Ahmad</td>
<td>ahmad</td>
</tr>
<tr>
<td>Mammal common noun</td>
<td>one-place predicates</td>
</tr>
<tr>
<td>talib student</td>
<td>(λx)(talib(X) ∧ gender(X))</td>
</tr>
<tr>
<td>Non-mammal common noun</td>
<td>one-place predicate</td>
</tr>
<tr>
<td>madat</td>
<td>(λx) madat(X)</td>
</tr>
<tr>
<td>course</td>
<td></td>
</tr>
<tr>
<td>Noun with adjective</td>
<td>one-place predicates</td>
</tr>
<tr>
<td>talib mumtaz</td>
<td>joined by ‘and’</td>
</tr>
<tr>
<td>excellent student</td>
<td>(λx)(talib(X) ∧ mumtaz(X))</td>
</tr>
<tr>
<td>Intransitive verb</td>
<td>one-place predicate</td>
</tr>
<tr>
<td>tkharraja</td>
<td>(λx) tkharraja(x)</td>
</tr>
<tr>
<td>graduated</td>
<td></td>
</tr>
</tbody>
</table>

For better understanding of the applicability of the above mentioned formalization to Arabic expressions, consider the following example [quoted from 1, p.39].

<table>
<thead>
<tr>
<th>Arabic Sentence</th>
<th>English Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>hal altaleb ahmed darasa madat com301?</td>
<td>Is the student Ahmed studied course com301?</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>d.</td>
<td>In addition, [12] pointed out that, “despite the fact that standard predicate logic represents well-studied formal representation formalism, it does not provide any compositional facilities. Lambda calculus offers an important framework for achieving such a goal but merely for the meaning construction of Arabic sentences”.</td>
</tr>
<tr>
<td>B. Haddad and Yaseen</td>
<td></td>
</tr>
</tbody>
</table>

Based on investigation into their works [10]-[14], Haddad and Yaseen believe that “since semantic representation has to be compositional on the level of semantic processing, lambda calculus, based on DRT, can be utilized as a helpful and practical technique for the semantic construction of Arabic, in Arabic understanding systems” [12].

In addition, [12] pointed out that, “despite the fact that standard predicate logic represents well-studied formal representation formalism, it does not provide any compositional facilities. Lambda calculus offers an important framework for achieving such a goal but merely for the meaning construction of Arabic sentences”.

Haddad and Yaseen’s model for constructing a meaning representation of an Arabic sentence is mainly based on some compositional rules, such as the following quantification rule:

\[ \|\text{Quant}\| \Rightarrow \lambda R.\|S(\text{Quantifier}(R, S)) \]

[as R refers to ‘restriction’, and S refers to ‘scope’]

By using such a rule, the logic-based meaning of the following Arabic sentence, for instance, could be represented [quoted from 4].
Syntactic and semantic rules and derivations are achievable.

To Arabic, and a straightforward quantified expressions have indicated that MG is applicable semantics Montagovian analysis of a fragment of Arabic connectivity.

Application is exclusively based on the syntax-semantics the declarative sentences or expressions, and the core formulations (i.e., DRT and HDPSG), they never say whether there are good reasons for using them specifically.

C. H. El-Sayed

As indicated in [15]:

“The project of formalizing language (i.e., El-Sayed 2011) is directed at the cognitive core of language: syntax and semantics. Therefore, lots of aspects of natural languages and their uses are fairly beyond the scope of that kind of project” [p.10].

That is, the linguistic level approached in this treatment is the declarative sentences or expressions, and the core application is exclusively based on the syntax-semantics connectivity.

The findings achieved from the preliminary syntax-semantics Montagovian analysis of a fragment of Arabic quantified expressions have indicated that MG is applicable to Arabic, and a straightforward applications of Montague’s syntactic and semantic rules and derivations are achievable. To illustrate, consider the syntactic analysis of the following example:

<table>
<thead>
<tr>
<th>Category</th>
<th>CN</th>
<th>DET</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcription</td>
<td>al-musāfīr-In</td>
<td>ba’d-u</td>
<td>‘ād-a</td>
</tr>
<tr>
<td>Word-to-word translation</td>
<td>the passengers</td>
<td>Some</td>
<td>returned</td>
</tr>
<tr>
<td>Full translation</td>
<td>Some passengers returned.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the syntactic derivation, as well as the semantic representation, the directions of derivation go from down-to-up and right-to-left, using Montague’s syntactic rules.

\[
\begin{align*}
\text{al-musāfīr-In} & \quad \text{ba’d-u} & \quad \text{‘ād-a} & \quad S2 \\
\text{al-musāfīr-In} & \quad \text{ba’d-u} & \quad \text{‘ād-a IV} & \quad S4 \\
\text{al-musāfīr-In} & \quad \text{CN} & & \quad S2
\end{align*}
\]

To be more balanced, suffice it to say that the application of MG into Arabic was a primitive trial towards an Arabic logic-based meaning representation, which is oriented by syntax-semantics connectivity. With the reason that MG is limited to treat certain Arabic quantified expressions, we could not claim that all Arabic linguistic phenomena, or even all Arabic quantified expressions, can be accommodated within the Montagovian framework, thereby trying a more advanced formal framework (e.g., DRT) is highly recommended.

V. CONCLUSION

Research on semantic representation of Arabic language, and in particular on deep logic-based meaning analysis and formalization, considering Arabic semantics with its relation to syntax, has so far received very little research attention. In view of that, it is not a surprise to see that the achievements in the field of Arabic semantic processing are in the low level, comparing with other natural languages, especially English. In other words, in the absence of a better Arabic logic-based meaning representation, it is certainly difficult, if not outright impossible, to expect significant advances in Arabic semantics processing.

Most agree that Arabic semantic analysis has so far not been treated deeply enough, neither linguistically nor logically. In addition, most of the reported works here that treated the Arabic logic-based meaning analysis and formalization have been provided by computer scientists (e.g. McGregor; B. Al-Johar; Haddad and M. Yaseen) with limited experience of logic formalisms. This is obvious in their simple, straightforward, and misleading symbolism in some cases. One of the main factors for this deficiency and negligence might reside in the complexity of this field, and in the invisible collaboration between researchers working in the fields of artificial intelligence, Arabic language, logic and linguistics. In view of that, and giving the fact that ALMR is a prominent interdisciplinary research field, building an interdisciplinary-oriented research community for ALMR is the fundamental assurance of research high quality and productivity in this area of specialization.

It also has to be stressed on the finding, that in McGregor and Al-Johar’s and Haddad and Yaseen’s works neither have indication for future work in their contributions nor carried work in the same points in their publication lists. In contrast, following his PhD in 2011, El-Sayed established a long-term research project with the aim to provide a logical treatment of Arabic as a pre step towards its computation.
As [26] shows, this research project has been announced and approved as an “Arab project” in the 24th Annual Philosophical Symposium: Contemporary Arab Intellectual Projects, The Egyptian Philosophical Society (EPS), Cairo, Egypt, 7th December 2013. The current survey and classification can be considered as step forward of this project. In addition, one of the studies that El-Sayed has started in this whole project is his interdisciplinary postdoctoral research “Towards a Logical Grammatical Approach of Arabic”, which established at SOAS University of London since 2015. This research project brings together logic, philosophy of language and linguistics to explore the expressive power of logical languages and techniques to formally represent various levels of Arabic linguistic expressions. It is proposed to develop a logical grammatical approach that is adequate to Arabic discourse. Moving from sentences to discourse is a crucial step towards establishing a logical grammatical approach that adequate to treat various levels of Arabic linguistics.

Finally, suffice it to say that Arabic logic-based meaning representation is extremely important for achieving the high level of Arabic natural language processing. In view of that, much work need to be done on semantic representation and semantic analysis of Arabic, and this accordingly necessitate establishing a research community that combines Arab logicians, linguists and computer scientists, thereby more collaboration as well as high level and amount of achievements could be anticipated.

REFERENCES


Haytham El-Sayed was born in Al-Sharqiya, Egypt, on November 3rd 1977. He is an assistant professor of Philosophy at the Faculty of International Business and Humanities (FIBH), Egypt-Japan University of Science and Technology (E-JUST), Egypt. Dr. El-Sayed has obtained a B.A. degree in philosophy from South Valley University, Egypt 1999, M.A. degree in philosophy (symbolic logic) from South Valley University, Egypt 2004, and PhD in philosophy (symbolic logic and philosophy of language) from The University of Leeds, UK, 2011. He has the following research and university positions: Postdoctoral researcher at SOAS University of London, UK, fully sponsored by the Egyptian Ministry of Higher Education (February-August 2015) ; Lecturer of logic and philosophy of science and language at the Department of Philosophy, Faculty of Arts, South Valley University, Qena, Egypt (2011- till now); PhD student at the University of Leeds, funded by the Egyptian Ministry of Higher Education (2007-2011); Assistant Lecturer in the Department of Philosophy, Faculty of Arts, South Valley University, Qena, Egypt (2004-2007); Demonstrator in the Department of Philosophy, Faculty of Arts, South Valley University, Qena, Egypt (1999-2004). His current research interests are: pure and applied logic; formal semantics; Arabic computational linguistics; philosophy of language; philosophy of science. Dr. El-Sayed' major publications: Arabic as a Formal Language, Lambert Academic Publishing, Saarbrücken, Germany, 2016; “Quantum Theory between Philosophy and Science”, an Academic Debate with Mohamed Y. Motow, Vol. 10-12, April, Algeria, pp. 424-439, 2016; “Arabic between Formalization and Computation”, International Journal of
Languages, Literature and Linguistics (JLLL), Vol. 1, No. 1, March, pp. 25-29, 2015;