Arabic between Formalization and Computation

Haytham El-Sayed

***Abstract*—This paper provides an attempt to apply a formal method (Montague grammar) to Arabic, as a pre-step towards its computation. Since semantic representation has to be compositional on the level of semantic processing, formalization based on Montague grammar can be utilized as a helpful and practical technique for the semantic construction of Arabic in Arabic understanding systems. As efforts of Arabic syntactic-semantic formalization are very limited, I hope that this approach might be a further motivation to redirect research to modern formal interpretation techniques for developing an adequate model of semantic processing for Arabic.**

***Index Terms*—Arabic NLP, formalization of Arabic, Montague grammar, syntax-semantics connectivity.**

I. Introduction

Many logicians and philosophers of language such as Rudolf Carnap [1]-[3], David Lewis [4], Saul Kripke [5], Richard Montague [6]-[8], Donald Davidson [9], among others, strongly believed in the capacity of formal languages and the applicability of formal methods to formalize and then interpret natural languages. In light of the hypothesis that natural languages can be formalized (i.e., amenable to receive formal interpretation, using formal languages and tools of logic) they positively applied their formal methods to different natural language groups. Successful syntactic-semantic formalization of different languages, such as English, Japanese and Thai, has led to promising achievements in the field of natural language processing (NLP).

Regarding to Arabic, my research did not discover any relevant logical literature on the question of how Arabic can be formally treated. That is, for Arabic, it was noted that the relationship of logic to language was an area much discussed in the medieval Arabic Aristotelian tradition (see [10], [11]), so the question has a historical pedigree, albeit one lacking continuity to the present. Therefore, the current work can be regarded as rebuilding the relationship between logic and Arabic language.

Moreover, for the last four decades concentration on Arabic formalisms and computation has focused on Arabic from the morphological and syntactical point of view (e.g., [12]-[16]). In this field, some success has been achieved. However, the importance of semantic processing for achieving the understanding capability, there were little works reported on semantic representation and semantic analysis of Arabic ([17]-[20]). Therefore, I believe that there is an elemental need to make more effort to develop an adequate model for semantic processing for Arabic and even no existing formal theory is capable to provide a complete and consistent account of all phenomena involved in Arabic semantic processing.

This paper is an attempt to provide a syntactic-semantic formalization of a fragment of Modern Standard Arabic (MSA) as a pre-step towards its computation. I do so using Montague grammar (henceforth MG) as a formal method. MG is a well-known mathematically motivated logical method that treats natural languages semantic and of its relation with syntax. In general, the syntactic component of MG is formulated in terms of a categorial system. The semantic component of MG is formulated indirectly in terms of model-theoretic semantics via an intermediate language of intensional logic.

Based on the widespread applicability of MG to variety of natural languages such as: English, Thai, Japanese and Romanian (e.g., [21]-[24]), and its use as a real and efficient tool in natural language processing (e.g., [25]), one can claim that MG can be regarded as a one way out of two current challenges of Arabic, which are syntactic-semantic formalization and computation on logical or non-statistical level, at the same time.

The current paper is divided into two main sections. In section one, I offer a formalization of a small fragment of Arabic quantification by providing a Montagovian syntactic calculus and then formal interpretation of the fragment. The process of formalization is illustrated by using some random Arabic quantified expressions. In section two, I proceed to explain how Arabic can be amenable to logical computation via such formalization.

II. Formalization of a Fragment of Arabic

A. The Syntax of the Fragment

As many commentators (e.g, [26]-[29]) have noted, determining a system of syntactic derivation is the first step towards a Montagovian formalization of a fragment of a language. In what fellows, I adopt Montague's method where the mathematics of Arabic sentence structure is defined in terms of syntactic categories and syntactic calculus for a fragment of Arabic quantification. Syntactic categories are divided into two types: primitive and complex categories. Table I below identifies both types of syntactic categories of my fragment.

In what fellows, I shall provide a syntactic calculus for some Arabic quantified expressions. The syntactic calculus below is based on MG in two respects: (a) Montague’s syntactic rules and (b) Montague’s method of syntactic derivation (i.e., tree analysis). The process of syntactic derivation in both examples (1) and (2) is provided using:

Manuscript received September 5, 2014; revised January 15, 2015.
Haytham El-Sayed is with the Department of Philosophy, South Valley University, Egypt (e-mail: haytham_e_e@ yahoo.com).

guidance table, and tree analysis, as following (note: read the expression in us using right-to-left direction).

TABLE I: SYNTACTIC CATEGORIES

Category index	Index definition	Category name
t	Primitive	Sentence
e	Primitive	Entity expression; or proper name
T	t/IV	Term (noun phrase)
IV	t/T	Intransitive verb phrase
TV	IV/T	Transitive verb phrase
CN	t//e	Common noun phrase
DET	T/CN	Determiner
CAN	CN/CN	Adjective
IV/t	IV/t	Sentence-taking verb phrase
IV//IV	IV/IV	IV-taking verb phrase

TABLE II: GUIDANCE TABLE OF EX.1

Ex.(1)

Syntactic category	CN	DET	IV
Expression in use	*aṭ-ṭullāb-i*	*jamī'-u*	*ḥaḍar-a*
Morpheme-by-morpheme gloss	DEF-student.PL-GEN	all-NOM	come.PST-3MS
Full translation	All students came.		

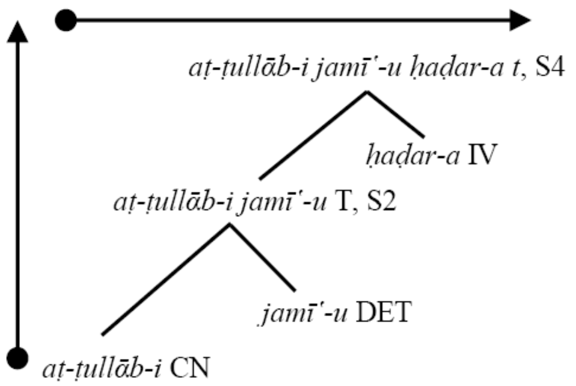

Fig. 1. Tree analysis of Ex.1.

Example (1) is a verbal sentence/jumla fiʿliyya, whose initial word is an intransitive verb (IV) [*ḥaḍar-a* /came]. To analyze the sentence, I start, using the bottom-to-up and left-to-right directions as illustrated by arrows, with categorizing [*aṭ-ṭullāb-i* /students] as CN. Next, using S2, CN [*aṭ-ṭullāb-i* /students] combines with DET [*jamī'-u* /all] to produce T [*aṭ-ṭullāb-i jamī'-u* /all students]. Then, using S4, T [*aṭ-ṭullāb-i jamī'-u* /all students] combines with IV [*ḥaḍar-a* /came] to produce t [*aṭ-ṭullāb-i jamī'-u ḥaḍar-a* /All students came] (shown in Table II, Fig. 1).

TABLE III: GUIDANCE TABLE OF EX.2

Ex.(2)

CN	DET	T	TV	T	IV//IV
rajul-i-n	*kull-u*	*hā*	*(an) yu-ḥibb-a*	*Sārah*	*ta-tamannā*
man-GEN-INDEF	every-NOM	3FS.OBJ	to 3m-love-S.SUBJ	Sarah-NOM	3FS-wish-IMP
Sarah wishes that every man would love her.					

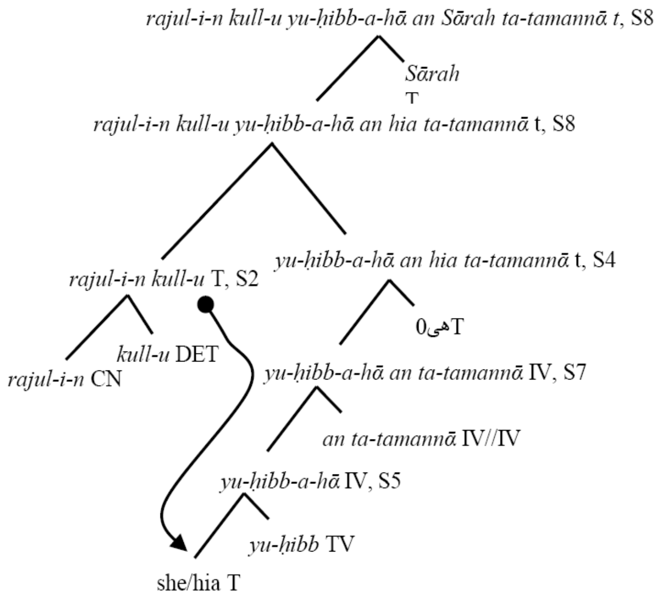

Fig. 2. Tree analysis of Ex.2.

Example (2) is slightly complicated since it is constructed of seven syntactic categories [T, TV, IV, IV//IV, CN, DET, t]. Therefore, in addition to the two known directions (i.e., bottom-up and left-to-right) a new direction will be involved in the process of syntactic derivation. The new direction is forward/backward, and will be illustrated throughout the process of syntactic derivation. To analyze Example (3) I start, as normal, with categorizing [*rajul-i-n* /man] as CN. Next, using S2 and forward direction, CN [*rajul-i-n* /man] combines with DET [*kull-u* /every] to produce T [*rajul-i-n kull-u* /every man]. Next, using S5 and backward direction illustrated by the arrow, T [*hia*/she] combines with TV [*yu-ḥibb* /love] to produce IV [*yu-ḥibb-a-hā* /love her]. Next, using S7 and forward direction, IV//IV [*an ta-tamannā* /wish to] combines with IV [*yu-ḥibb-a-hā* /love her] to produce IV [*yu-ḥibb-a-hā an ta-tamannā* /wish that love her]. Next, using S4, T [*hia*/she] combines with IV [*yu-ḥibb-a-hā an ta-tamannā* /wish to love her] to produce a sentence t [*yu-ḥibb-a-hā an hia ta-tamannā* /she wishes that love her]. Next, using S8, T [*rajul-i-n kull-u* /every man] combines with t [*yu-ḥibb-a-hā an hia ta-tamannā* /she wishes to love her] to produce t [*rajul-i-n kull-u yu-ḥibb-a-hā an hia ta-tamannā* /she wishes that every man loves her]. Then, using again S8, T [*Sārah* /Sarah] combines with t [*rajul-i-n kull-u yu-ḥibb-a-hā an hia ta-tamannā* /she wishes that every man loves her] to produce t [*rajul-i-n kull-u yu-ḥibb-a-hā an Sārah ta-tamannā* /Sarah wishes that every man loves her] (shown in Table III, Fig. 2).

B. Compositional Interpretation of the Fragment

Reference [29] shows, the most important features of MG are its use of model theoretic semantics which has commonly been used for the semantics of logical languages, and its adherence to the principle of compositionality—that is, the meaning of the whole is a function of the meanings of its parts and their mode of syntactic combination. I present below a compositional semantic interpretation of the fragment. Using Montague's rule-to-rule translation method, I shall show how the above examples of Arabic quantified expressions can be translated into the language of intensional logic. The process of translation is illustrated by Fig. 3 and Fig. 4 below.

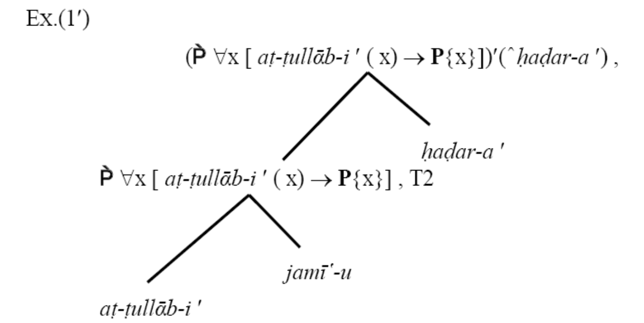

Fig. 3. Rule-to-rule translation of Ex.1.

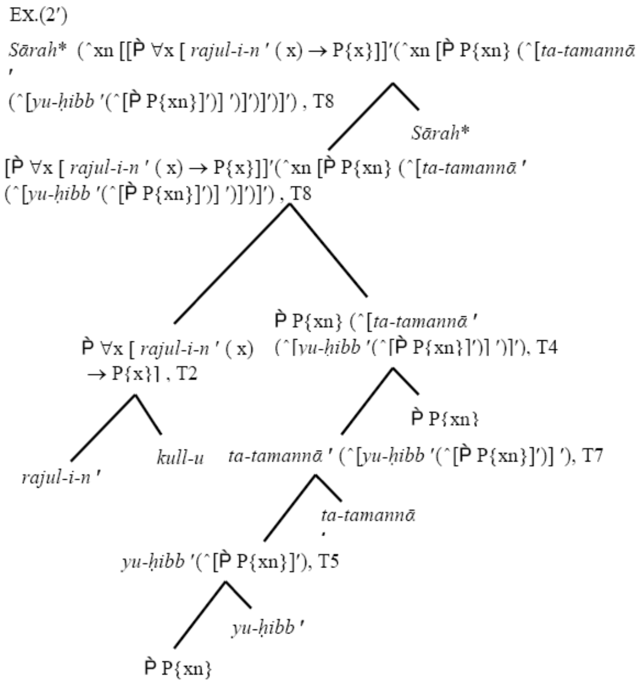

Fig. 4. rule-to-rule translation of Ex.2.

The most surprising outcome to emerge from this investigation is that MG is absolutely sufficient to provide a successful compositional interpretation to my fragment. That is, both procedures of syntactic derivation and semantic analysis of my fragment are in parallel with those used in MG. The results obtained from the preliminary syntactic and semantic analysis of my fragment have indicated that within the scope of my fragment no developments or extensions were required to the original theory.

The point to be made here is that, since the Montagovian framework is mainly based on analysis of English, which belongs to the Indo-European language family, success in applying the Montagovian framework to Arabic, which is Afro-Asiatic, was a ‘challenge’. As it was expected, in the light of the differences between Arabic and English, some adjustment needs to be made to Montague's approach -whether in details or more widely- to accommodate features of Arabic generally, and Arabic quantification specifically. Contrary to expectations, this study did not find a significant difficulty in producing a compositional interpretation to Arabic, using Montague original theory of grammar, without the need to additional factors. In addition, this study did not find significant differences, in apparatus or process of application of MG, between Arabic and cases of other language groups (i.e., [22]-[24] above). This increases the evidence for the hypothesis, according to Farghaly [30, p.43], that Arabic shares many properties with other natural languages, and interesting insights can be gained by analyzing the structure of Arabic by any of the prevailing linguistic theories or formalisms.

However, with a limited portion or fragment of Arabic quantified expressions, caution must be applied, as the findings might not be transferable to all linguistic phenomena of Arabic, or even to all Arabic quantified expressions. More specifically, MG is limited to treat certain quantified expressions in Arabic for two respects:

1) Montague’s syntactic and semantic rules (as presented in Montague’s PTQ) are designed in a way that does not fit well certain expressions of Arabic quantification, however they do fit well other such expressions. To be exact, in addition to the noun phrase construction, Arabic quantifiers are used in further different constructions such as: (a) determiner and demonstrative [DET+DEM], (b) determiner and term (i.e., proper noun) [DET+T], and (c) determiner and pronoun [DET+PRO].
2) Certain characteristics of Arabic can be regarded as a source of limitation since certain quantified expressions in Arabic are used in grammatical forms that are out of the range of MG. To be exact, based on ([31]-[33]), Arabic quantifiers are classified into three classes, according to morphological and syntactic aspects: (1) nominal quantifiers, (2) numerals and (3) phrasal quantifiers. Furthermore, the class of nominal quantifiers in itself is classified into three sub-classes, according to the range of quantity: (i) expressions of totality, (ii) expressions of majority and (iii) expressions of partiality. In each of these classes there are a number of quantifiers which can not be represented in terms of Montague's rules.

Both limitations require the creature of new and specific rules to treat certain constructions of Arabic, or proceed to post-Montague semantics.

III. Towards a Logical Computation of Arabic

In the field of applied logic, which concerns the uses of logical languages and methods in computer science, it has become increasingly difficult to ignore that logic does well enough in both programming and implementing several computational systems. That is, the second half of the 20th century witnessed an intensive interaction between logic and computer science, to the extent that computer scientists have utilized many logical languages and techniques in areas such as program verification, semantics of programming languages, automated theorem proving, and logic programming (see [34]-[37]). To be more exact, reference [38] shows, language (spoken and written) is central to all aspects of our communication. Therefore natural language processing systems, both current and future, are bound to play a crucial role in our communication with machines and even among ourselves [p. 1242]. Besides, there is no doubt that logic is considered an essential tool to construct these systems, and it is capable to do so because it has the formal notation, as reference [39] shows: "logic was developed as a formal notation to capture the essential properties of natural language and reasoning" [p. 515].

To illustrate, among various logical methods and tools, categorial grammar (CG) can be considered the main representative of logical methods that have widely used to accomplish computational systems of natural languages on the logical or non-statistical level. Historically, as reference [40] shows, CG come of a tradition of linguistic description rooted in philosophy of language, logic and algebra [p.1]. CG is not a single grammar but it is divided to many sorts such as: basic categorial grammar or "Lambek" categorial grammar, combinatory categorial grammar (CCG), unification categorial grammar (UCG) and MG as an extension to CG. It is no coincidence that the first, abortive flowering of categorial grammar for serious natural language description came with the first attempts at implementing natural language processing on a computer [Ibid]. That is to say, researchers in NLP have implemented systems and programs which based on CGs (eg., CATLOG system, MCGTOOLS system, The Natural Deduction CG Parser, ROSETTA machine translation).

Based on the current application, since Arabic shares other natural languages the possibility of formalization on the level of CGs, one can expect that the computational results that based on CGs of such languages could be extended to Arabic. That is, many computational systems and programs have been implemented in languages such as English, Japanese and Thai due to their amenability to CGs formalization. As illustrated above, Arabic is amenable to formalization, using one of CG extensions which is MG, and this increases the evidence for the hypothesis that logical computation of Arabic is considerably achievable.

IV. Conclusion

In this paper I attempted to present some results of my view of a formalization for a fragment of Arabic quantification. Based on the applicability of formal methods to Arabic, I believe that the progress in the field of NLP, that has been made in the last years, is also applicable to Arabic with some modifications. This formalization is based on MG, and it has been successfully used in several NLP systems to achieve deep syntax semantic Analysis. Unfortunately there are still little works reported from the Arabic computational linguistic community for semantic construction and formalization of Arabic. I hope that this approach might be a further motivation to redirect research to modern semantic construction technologies for developing an adequate model of semantic processing for Arabic.

Acknowledgment

This work has been based on my 2011 University of Leeds PhD thesis. I would like to thank my primary supervisor, Professor John Divers. Professor Divers has guided wisely and unfailingly every stage of my work. He has also given me so generously his expertise and support which has helped me overcome academic and non-academic difficulties. My greatest debt is to him. I would also like to thank my co-supervisor, Professor Janet Watson- University of Salford, for the invaluable conversations, pieces of advice and linguistic insights that have improved this work in countless ways.

References

[1] R. Carnap, *Introduction to Semantics*, London: Geoffrey Cumberlege, 1942.

[2] R. Carnap, *The Logical Foundation of Probability*, Chicago: University of Chicago Press, 1950.

[3] R. Carnap, *Meaning and Necessity: A Study in Semantics and Modal Logic*, 2nd ed., Chicago: University of Chicago Press, 1956.

[4] D. Lewis, “General semantics,” in *Montague Grammar*, B. Partee, Ed. London: Academic Press, 1976, pp. 1-50.

[5] S. Kripke, “Naming and necessity,” in *Semantics of Natural Language*, D. Davidson and G. F. Harman, Eds. Dordrecht: Reidel, 1972.

[6] R. Montague, “English as a formal language,” in *Formal Philosophy: Selected Papers of Richard Montague*, R. Thomason, Ed. London: Yale University Press, 1974, pp. 188-221.

[7] R. Montague, “Universal grammar,” in *Formal Philosophy: Selected Papers of Richard Montague*, R. Thomason, Ed. London: Yale University Press, 1974, pp. 222-246.

[8] R. Montague, “The proper treatment of quantification in ordinary English,” in *Formal Philosophy: Selected Papers of Richard Montague*, R. Thomason, Ed. London: Yale University Press, 1974, pp. 247-270.

[9] D. Davidson, *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, 1984.

[10] T. Street. (2013). Arabic and Islamic philosophy of language and logic. In *the Stanford Encyclopedia of Philosophy*. E. N. Zalta, Ed. [Online]. Available: http://plato.stanford.edu/archives/win2013/entries/arabic-islamic-language/

[11] O. Leaman, *An Introduction To Classical Islamic Philosophy*, Cambridge: Cambridge University Press, 2004.

[12] Y. Hlal, “Morphological analysis of Arabic speech,” presented at Conference on Computer Processing of the Arabic Language, Kuwait, 1985.

[13] A. F. Fehri, *Issues in the Structure of Arabic Clauses and Words*, Dordrecht & Boston: Kluwer Academic Publishers, 1993.

[14] E. Benmamoun, *the Feature Structure of Functional Categories: A Comparative Study of Arabic Dialects*, Oxford: Oxford University Press, 2000.

[15] K. Beesley, “Finite-state morphological analysis and generation of Arabic at Xerox research: Status and plans in 2001,” in *Proc. the ACL 39th Meeting, Workshop on Arabic Language Processing; Status and Prospects*, Toulouse, 2001, pp. 1–8.

[16] J. Aoun and Y. A. Li, *Essays on the Representational and Derivational Nature of Grammar: The Diversity of Wh-Constructions*, Boston Massachusetts: The MIT Press, 2003.

[17] B. AL-Johar and J. McGregor, “A logical meaning representation for Arabic (LMRA),” in *Proc. 15th National Computer Conference*, Riyadh, Saudi Arabia, 1997, pp. 31-40.

[18] B. Haddad and M. Yaseen, “Towards understanding Arabic: A logical approach for semantic representation,” presented at ACL/EACL01: Conference of the European Chapter, Workshop: Arabic Language Processing: Status and Prospects, 2001.

[19] H. El-Sayed, “Philosophical investigations into formalization and compositionality of language: A Montagovian analysis of Arabic quantification as a case study,” PhD Thesis, Dept. Philosophy, Univ. Leeds, October, 2011.

[20] H. El-Sayed, *Arabic as a Formal Language*, 2014.

[21] R. Montague, “The proper treatment of quantification in ordinary English,” in *Formal Philosophy: Selected Papers of Richard Montague*, R. Thomason, Ed. London: Yale University Press, 1974, pp. 247-270.

[22] A. Costachescu, “Three-place verbs in a MG for Romanian,” *Revue Roumaine De Linguistique*, vol. 31, no. 1, pp. 91-98, 1986.

[23] T. Sugimoto, “Transformational Montague grammatical studies of Japanese,” PhD thesis, Univ. Hawai'i, 1982.

[24] K. Godden, “Montague grammar and machine translation between Thai and English,” PhD thesis, Univ. Kansas, 1981.

[25] M. Rosetta, *Compositional Translation*, Dordrecht: Kluwer, 1994.

[26] A. Kao. (2009). Montague grammar. [Online]. Available: http://wwwpersonal.umich.edu/~akao/NLP_Paper.htm

[27] B. Partee and H. Hendriks, "Montague grammar," in *Handbook of Logic and Language*, J. Benthem and A. Meulen, Eds. Oxford: Elsevier, 1997, pp. 5-91.
[28] B. Partee, "Richard Montague (1930-1971)," in *Encyclopedia of Language and Linguistics*, K. Brown, Ed. 2nd ed. vol. 8, Oxford: Elsevier, 2006, pp. 255-57.
[29] T. Janssen. (2012). Montague Semantics. *The Stanford Encyclopedia of Philosophy*. [Online]. Available: http://plato.stanford.edu/archives/win2012/entries/montague-semantics/
[30] A. Farghaly, "The Arabic language, Arabic linguistics and Arabic computational linguistics," in *Arabic Computational Linguistics*, A. Farghaly, Ed. CSLI Publications, 2010, pp. 43-81.
[31] K. Ryding, *A Reference Grammar of Modern Standard Arabic*, Cambridge: Cambridge University Press, 2005.
[32] P. Hallman, "Quantifier," in *Encyclopedia of Arabic Language and Linguistics*, K. Versteegh, Ed. Vol. IV. BRILL, 2009, pp. 14-20.
[33] N. Wahba, *Encyclopedia of Arabic Grammar*, *Morphology*, *Rhetoric and Speech*, Cairo: Naguib Wahba, 2003.
[34] H. El-Sayed, "Predicate logic and its applications in artificial intelligence," M.A. Thesis, Dept. Philosophy, South Valley Univ., Qena, Egypt, April, 2004.
[35] H. El-Sayed, "Applied logic as an Arabian project," presented at 24th Annual Conference of Egyptian Philosophical Association: Contemporary Arabian Intellectual Projects, Cairo University, Egypt, December 7-9, 2013.
[36] E. Gradel and R. Kahle, *Computer Science Logic*, Berlin: Springer-Verlag, 2009.
[37] T. Icard and R. Muskens, *Interfaces: Explorations in Logic*, *Language and Computation*, Berlin: Springer-Verlag, 2010.
[38] A. Joshi, "Natural language processing," *Science*, vol. 253, no. 5025, pp. 1242-1249, 1991.
[39] J. Allen, *Natural Language Understanding*, Amsterdam: The Benjamin/Cummings Publishing Company, 1988.
[40] M. Wood, *Categorial Grammars*, London: Routledge, 1993.

Haytham El-Sayed was born in Al-Sharqiya, Egypt, on November 3, 1977. He has obtained a B.A. degree in philosophy from South Valley University, Egypt 1999, a M.A. degree in philosophy (symbolic logic) from South Valley University, Egypt in 2004, and a PhD degree in philosophy (symbolic logic and philosophy of language) from The University of Leeds, UK in 2011.

He is a postdoctoral researcher at SOAS University of London, UK, fully sponsored by the Egyptian Ministry of Higher Education (February-August 2015); he is also a 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); he was a PhD student at the University of Leeds, funded by the Egyptian Ministry of Higher Education (2007-2011); and he had been an assistant lecturer in the Department of Philosophy, Faculty of Arts, South Valley University, Qena, Egypt (2004-2007); he also was a 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 are *Basics of Symbolic Logic*, e-learning academic course, Cairo: National Centre of E-Learning, 2013; "Applied logic as an Arabian project," presented at 24th Annual Conference of Egyptian Philosophical Association: Contemporary Arabian Intellectual Projects, Cairo University, Egypt, December 7-9, 2013; "The philosophical and logical contributions to technology development: AI as a model," *Diogenes*, Cairo: Cairo University Press, 2014, issue 1, pp. 243-300; *Arabic As a Formal Language, Al-ʿarabya-tu lugha-tun ṣūri-ah*, 2014, [unpublished]. Dr. El-Sayed has been a member of the Egyptian Philosophical Association in Cairo, since 2013.