

A Nanosyntactic Analysis of Arabic Complementizers

Saleem Abdelhady

American University of the Middle East, Kuwait

Abstract—This study investigates the syntactic composition of complementizers in Arabic from a nanosyntactic perspective (Starke, 2009). The study unravels the dichotomy in the behavior of root complementizers in Arabic; it highlights how the selection of complementizers is impacted by presupposed information and the degree of certainty. In spoken varieties of Arabic, such as Jordanian Arabic and Lebanese Arabic, root complementizers are blocked from root positions, but such positions are accessible for Modern Standard Arabic. Refining Ross' (1970) hidden verbs theorem and Fassi Fehri's (2012) featural distinction, the study shows that if a complementizer is selected by different triggers of veridicality, such as *?anna* in MSA, *?inn* in Jordanian Arabic and *?anno* in Lebanese Arabic, the complementizer cannot appear without its trigger, whereas if a complementizer is selected by one trigger, such as *?inna* in MSA, the complementizer can be used in root positions without a trigger. Comparing the findings of the study with Baunz' (2018) universal hierarchy, we show that the hierarchy in its current status fails to account for Arabic data. The conclusion gives a stronger contribution for the semantic composition of complementizers.

Index Terms—nanosyntax, complementizers, Jordanian Arabic, The Minimalist Program, factivity

I. INTRODUCTION

This study investigates the syntactic composition of complementizers in Arabic. Comparing the complementizer systems of Modern Standard Arabic (MSA) with Jordanian Arabic (JA) and other varieties of Arabic shows that the systems are more complex than previously stated in the literature (Jarrah, 2019b; Ouhalla, 1994; Shlonsky, 1997; Soltan, 2006; Ross, 1970; Fassi Fehri, 2012) and creates a problem for the universal hierarchy of complementizers (Baunz, 2018) in its current status. The complexity can be realized clearly with the use of clause-initial Cs.

(1) *?inna* *al-banāt-i* *waṣal-na* (MSA)
 COMP DEF-girls-ACC arrived-3F.PL
 'Indeed, the girls arrived'.

(2)**?anna* *al-banāt-i* *waṣal-na* (MSA)
 COMP DEF-girls-ACC arrived-3F.PL
 'Indeed, the girls arrived'.

(3)**?inn-u* *?il-walad* *?iḍḡa* (JA)
 COMP-3M.SG DEF-boy.ACC arrived
 'Intended: the boy arrived'.

(4)**?anno* *?il-walad* *?iḍḡa* (LA)
 COMP DEF-boy.ACC arrived
 'Intended: the boy arrived'. (Jarrah, 2019b)

The examples above show that, in MSA, two forms of Cs are used in the system *?inna* and *?anna*. In JA, nevertheless, only one form is used *?inn*. A notable difference between the two systems is that, in MSA, it is acceptable to have the C in clause-initial position, yet the C must be *?inna*. The use of *?anna* leads to ungrammatical construction (2). However, the ungrammaticality of using *?anna* intersects with the ungrammaticality of using Cs in the same position in JA and LA; the use of the C *?inn* is unacceptable, which makes it behave like the C *?anna* in MSA. Fassi-Feheri (2012) claims that the distinction is based on embedding contexts. However, he overlooks the fact that both Cs can be embedded (5) in MSA but not in JA (6), whereas Ross (1970) attributes the behavior to the existence of a hidden performative verb of saying. Looking at data from JA questions the validity of his conclusion (6).

(5) (*?aqūlu*) *?inna* *al-banāt-i* *waṣal-na* (MSA)
 (say.I) COMP DEF-girls-ACC arrived-3F.PL
 '(I say) that the girls arrived'. (Ross, 1970)

(6)*baḡūl* *?innu* *?il-bānt* *?iḍḡ-ū* (JA)
 say.I COMP DEF-girls.ACC arrived-3F.PL

‘I say that the boy arrived’.

(Jarrah, 2019b)

The use of the C *ʔinn* along with an explicit verb of saying is acceptable in JA. However, assuming that the verb of saying is hidden does not lead to grammatical structure in this variety (compare (3) with (6)). Comparing the two systems highlights the following pressing research questions: (a) Why does the hidden performative verb hypothesis only apply to MSA but not JA Cs? And (b) How is the MSA C system different from that of Arabic varieties? If not, how do they fit the universal pattern? And how far can Baunz' (2018) universal hierarchy account for Arabic data?

The analysis is coached within the spirit of Nanosyntax (NS) (Starke, 2009), a new approach to grammar that postulates that not only sentences, clauses, phrases, and words are composed of binary structures but also morphemes may be decomposed into features; the way morphemes are composed is similar to the way sentences are formed. This study utilizes NS to explore the internal structure of declarative complementizers in Arabic (Baunz, 2018; Starke, 2009). The research aims to arrive at a proper fine-grained syntactic analysis of Arabic complementizers through exposing declarative complementizers to factual, semi-factual, desiderative verbs and testing constructions on subject and *wh*-extraction.

The study is structured as follows. Section 2 sets the scene. It highlights previous models on the selection of Cs in Arabic and other languages. Two models are presented, namely, logophoricity and factivity. Section 3 presents Nanosyntax as the theoretical framework of data analysis. Section 4 utilizes the framework toward the decomposition of Cs in MSA and its spoken varieties. Section 5 compares the findings of the study with other languages. The last section concludes the study and compares the findings with other languages.

II. SETTING THE SCENE

To pave the way for understanding the structure of complementizers in the Arabic language, we highlight previous models that accounted for the selection of complementizers in MSA and its spoken varieties. In addition, we present what impacts the selection of complementizers cross-linguistically.

A. The Selection of Arabic Complementizers

According to Fassi-Fehri (2012), among others, there are three complementizers in MSA: *ʔanna*, *ʔinna*, and *ʔan*¹. The complementizers are in complementary distribution; that is, if one complementizer is used in a specific context, the other complementizers are used in non-intersecting environments. The selection of Cs is not random. According to Fassi-Fehri (2012), the behavior of Cs is attributed to logophoricity, clause type (root or embedded), and selection of mood.

(7) *ʔata ʔakar-u* *ʔann-i* *qabal-tu-ka* *fi* *ʔas-suq-i*.
remember-1SG.SBJ that-me met-1SG.SBJ-you in the-market-GEN
‘I remember that I met you at the market’.

(MSA)

(8) *samuʕ-tu* *ʔanna* *ʔar-rijaal-a* *yaʔkul-u* *ʔas-samak-a*.
heard-1SG.SBJ COMP the-men-ACC eat-3PL.SBJ the-fish-ACC
‘I heard that the men are eating the fish’.

(MSA)

(9) *ʔaqul-u* *ʔinna* *ʔal-walad-a* *qad* *tarka* *ʔal-bajt-a*.
say-1SG.SBJ COMP the-boy-ACC PTCL leave.3SG.SBJ the-house-ACC
‘(I say) that the boy left the house’.

(MSA)
(Ross, 1970)

(10) *ʔāmsi* *tamnay-tu* *ʔān* *t-uḡādr-a* *ḡad-an*
yesterday hoped-I that 2-go-SUBJ tomorrow
Yesterday, I hoped that you would go tomorrow.

(Fassi-Fehri, 2012, p. 240)

The phenomenon of logophoricity refers to a “binding relation that may employ a morphologically different set of anaphoric forms, in the context where the referent is an entity whose speech, thoughts, or feelings are being reported” (Clements, 1975). Fassi-Fehri (2012) argues that C contains a logophoric feature, and that feature impacts the selection of Cs; Cs may vary because of the deictic valuation and interpretation of the logophoric feature. The feature shows a distinction between two representations of the pragmatic role of the speaker: a speaker of the matrix speech and a speaker of the embedded speech.

(11) *qaala-t l-ii l-fataat-u* *ʔinna-nii* *ʔuḥibb-u-ka*.
said-F to-me the-girl-NOM that-I I-like-IND-you
‘The girl said to me that she likes me’.

(12) *ʔaxbara-t-nii* *l-fataat-u* *ʔanna-haa* *t-uḥibb-u-nii*.
informed-F-me the-girl-NOM that-her F-like-IND-me

¹ According to Fassi Fehri (2012, p. 240) and Persson (2002), *ʔan* is in the C head not T. However, not all researchers agrees that *ʔan*. For Habib (2009), *ʔan* originates in T. We here adopt Fassi Fehri's (2012, p. 240) view.

‘The girl informed me that she likes me’.

- (13) **ʔaxbara-t-nii* *l-fataat-u* *ʔinna-haa* *t-uħibb-u-nii*.
 informed-F-me the-girl-NOM that-her F-like-IND-me
 ‘The girl informed me that she likes me’.

(Fassi-Fehri, 2012, p. 237)

In (11), the C *ʔinna* appears because the agent of the matrix speech is distinct from the agent of the action of the embedded speech. In (12), the matrix speech and the embedded speech have the same agent; hence, (13) is ungrammatical. In our terminology, there is a distinction between the speaker and the cognizant, and that distinction impacts the selection of Cs. Fassi-Fehri (2012) shows that logophoricity by itself is not enough for the selection of complementizers. The distinction between the three forms is realized based on their logophoricity, position in a clause, and selection of mood. Therefore, he argues that root and embedded clauses are not analogous to a matrix and embedded speech; that is, root Cs do not necessarily correspond to the matrix speech utterer and the like. Based on that conclusion, Fassi-Fehri (2012) argues that *ʔinna* can be a root C (14), and it can be embedded under *ʔaqulu* (11). In addition, *ʔinna* matches the sayer – matrix speech. However, *ʔanna* and *ʔan* cannot be used in root clauses (15); they must be embedded (16); *ʔanna* matches only the agent of the embedded speech, i.e., it is not logophoric. Like *ʔanna*, *ʔan* must be embedded but similar to *ʔinna*, it is a logophoric C (cf., (16) and (17)). Unlike *ʔanna*, *ʔan* can select a subjunctive verb, i.e., it can select mood (17).

- (14) *ʔinna-n-ii* *ʔ-uħibb-u-ka*.
 that-n-I I-like-IND-you
 ‘Indeed, I like you’.
- (15) **ʔanna-ha* *t-uħibb-u-n-ii*.
 That-her 3-like-IND-me
 ‘that she like me’.
- (16) *naada-t* *ʔan* *(u)-dxul*.
 call-F that come.in
 ‘She called: “come in”’.
- (17) *naada-t* *ʔan* *y-adxul-a*.
 call-F that 3-come.in-SUBJ
 ‘She called him to come in’.
- (18) **ʔan* *(u)-dxul*.
 that come.in
 ‘That “come in”’.

(Fassi-Fehri, 2012, p. 242)

The result of these distinctions is that the Arabic Cs can be categorized as follows:

- a. [+ Logophoric, +Root, + Case] = *ʔinna*
- b. [+ Logophoric, - Root, + Mood] = *ʔan*
- c. [- Logophoric, -Root, + Case] = *ʔanna*

(Fassi-Fehri, 2012, p. 243)

The problem of Fassi-Fehri’s (2012) presentation of the selection of Arabic complementizers is that it is not clear why *ʔinna* is the only C that can be used in root clauses, while other Cs are blocked from occupying this position. The mysterious behavior of Cs remains despite Fassi-Fehri’s (2012) thorough presentation. What we aim to figure out is why specific Cs may occupy clause-initial positions while other Cs are blocked from these positions. Notice that Fassi-Fehri’s (2012) analysis considers clause types (root or embedded) as a feature to set Cs apart, yet his analysis leaves the question open. We consider other factors that may impact the selection of Cs.

B. The Selection of Complementizers Cross-Linguistically

Hooper and Thompson (1973) set the first criterion that impacts the selection of Cs in English. They show that declarative Cs are restricted to specific verbs; that is, in English, the C *that* is used only with factive and non-factive verbs. Factive predicates are considered “as one of the canonical classes of presupposition triggers” (Beaver & Geurts, 2014); they affect the truth-value of propositions in their complements. Using a factive verb (19) establishes a presupposition to be true. A nonfactive verb (20), nevertheless, does not commit a speaker to the truth value of a proposition in the embedded clause. The following examples are illustrative.

- (19) I *know* that the boy is sick.
 Presupposition: The boy is sick.
- (20) I *claim* that the boy is sick.
 Presupposition: The boy might be sick or not.

Hooper and Thompson (1973) subdivide factive, and nonfactive verbs that take *that* complements into the five groups.

TABLE 1
FACTIVE AND NONFACTIVE VERBS

Nonfactive			Factive		
A	B	C	D	E	
say, report, exclaim, assert, claim, vow, be true, be certain, be sure, be obvious	suppose, think believe, expect, guess, imagine, it seems, it happens, it appears	be (un)likely, (im)possible, (im)probable, deny	resent, regret, be sorry, surprised, bother, odd, strange, interesting	realize, learn, find out, discover, know, see, recognize	

Nonfactive are included in groups A, B, and C. Group A incorporates verbs that introduce reported complements; such complements are asserted but not presupposed. Group B is similar to Group A, yet, it differs in that its complements can be subject to tag questions. Group C contains complements that are neither asserted nor presupposed. Group D and E include factive verbs. With factive verbs, “presupposition and assertion are usually assumed to be mutually exclusive” (Hooper & Thompson, 1973, p. 16). Group D contains emotive factives, verbs that express some emotive attitude toward the complement, such as *regret*. Group E includes epistemic factives, verbs, such as *know*, that relate knowledge or the degree of validation.

In addition, Baunz (2018) shows that the selection of Cs in French (FR), Serbo-Croatian (SC), and Modern Greek (MG) can be impacted by factivity. Unlike English, factive and non-factive verbs may select different forms of Cs. The examples (21-23) illustrate Cs selected by factive verbs, whereas the examples (24-26) exemplify Cs selected by non-factive verbs.

(21) *Je me rappelle que je t'ai rencontré au marché.* (FR)
I PRES remember that I PAST meet at market
'I remember that I met you at the market.'

(22) *Sjećam se da sam te upoznao na tržnici.* (SC)
remember 1.SG that AUX.PAST.1.SG you met.past.participle on the-market
'I remember that I met you at the market.'

(23) *Thimam-e pu se sinandisa stin aghora.* (MG)
remember-1SG that you-met-1SG at the market
'I remember that I met you at the market.'

(Roussou, 1992; cited in Baunz, 2018)

(24) *Paul a dit qu'il a vu Mary.* (FR)
Paul PAST say that PAST see Mary
'Paul said that he saw Mary.'

(25) *Pavao je rekao da je video Mariju.* (SC)
Paul AUX.PAST.3.SG said.PP that AUX.PAST.3.SG see.PP Mary
'Paul said that he saw Mary.'

(26) *O Pavlos ipe oti i Roxani efije.* (MG)
the Paul said-3SG that the Roxanne left-3SG
'Paul said that Roxanne left.'

(Giannakidou, 2009; cited in Baunz, 2018)

In FR, the C *que* is used with both factive and non-factive verbs ((21) and (24)). Similarly, in SC, the C *da* is selected regardless of factivity (22) and (25)). However, among the three languages, MG shows a different pattern; factive verbs select *pu*(23), whereas non-factive verbs select *oti* (26). In addition, MG utilizes a third form of C with desiderative non-factives, verbs that express a desire to do the act denoted by the speaker. While in FR utilizes the same C *que* and SC selects the C *da* with verbs such as *want*. In MG, *na* is used.

(27) *Je veux que Jean parte.* (FR)
I want that John leave.SUBJ
'I want John to leave.'

(28) *Želim da Ivan ode.* (SC)
want.1SG SUBJ John leave.3SG
'I want John to leave.'

(29) *Thel-o na fij-i o Kostas.* (MG)
want-1SG SUBJ leave-3SG the Kostas

‘I want John to leave’.

The implication of these patterns is that in FR and SC, one C is selected in three contexts, leading to a syncretic pattern. The syncretic pattern points toward a complex structure of Cs cross-linguistically; the complex structure which appears as one form is used for three functions in some languages, whereas in other languages, distinct forms are utilized in each context.

C. Extraction Across Factive/Non-Factive Predicates in Arabic

Jarrah (2019, p. 106) argues that factivity may impact embedded constructions in Arabic. Factive verbs block subject extraction (30), whereas non-factive verbs allow subjects to be extracted from their clausal complements (33)².

- (30) *ʔabuu-j nasa/ʕirif ʔmn-ha*
 father-my forgot/knew COMP-3F.SG
ʔil-marah ʔaxa ʔ-at ʔis-saaʕah.
 DEF-woman took-3F.SG.SBJ DEF-watch
 ‘My father forgot/knew that the woman took the watch’.

(Jarrah, 2019)

- (31) **mn ʔabuu-j nasa/ʕirif*
 who father-my forgot/knew
ʔmn-ha ʔaxa ʔ-at ʔis-saaʕah.
 COMP-3F.SG took-3F.SG.SBJ DEF-watch
 Intended: ‘Who did my father forget/know took the watch?’

(Jarrah, 2019)

- (32) *ʔabuu-j fakk/ faʕar/ɖann ʔmn-uh*
 father-my doubted/felt/thought COMP-3M.SG
ʔil-walad ʔaxa ʔ DEF-watch
 DEF-boy took.3M.SG.SBJ DEF-watch
 ‘My father thought/doubt/feel/that the boy took the watch’.

- (33) *min ʔabuu-j fakk/ faʕar/ɖann*
 who father-my doubted/felt/thought
ʔmn-uh ʔaxa ʔ ʔis-saaʕah.
 COMP-3M.SG took.3M.SG.SBJ DEF-watch
 ‘Who did my father believe/doubt/feel/guess took the watch?’

(Jarrah, 2019)

While the subdivision of non-factives might be a decisive factor, the type of non-factive verb does not impact subject extraction; that is, the use of the non-factive verb of saying such as *bagul* ‘say’ does not block subject extraction (34).

- (34) *miin ʔabuu-j bi-gul ʔmn-ha ʔaxa ʔ-at ʔis-saaʕah.*
 who father-my PART-said that-3F.SG took-3F.SG.SBJ DEF-watch
 ‘Who did my father say that took the watch?’

Jarrah (2019) concludes that in Arabic, factive verbs create strong syntactic islands, while nonfactive verbs do not. While Jarrah's (2019) observation may be applicable to JA, it cannot be generalized to capture data from MSA; the case of extraction of Cs in MSA is left unexplored in Jarrah's (2019a) overall conclusion. Consider the following examples.

- (35) *ʔal-wald-u ta ɖakara ʔanna ʔar-radʒul-a*
 DEF-boy-NOM remembered.3M.SG.SBJ that DEF-man-ACC
qad ʔaxaa ʔas-saʕat-a.
 PTCL took.3M.SG.SBJ DEF-watch-ACC
 ‘The boy remembered that the man took the watch’.

- (36) **man al-wald-u ta ɖakara ʔanna*
 who DEF-boy-NOM remembered.3SG.SBJ that
qad ʔaxa ɖa ʔas-saʕat-a.
 PTCL took.3M.SG.SBJ DEF-watch-ACC
 ‘Who did the boy remember that the man took the watch’.

Albeit a factive verb, *ta ɖakara* ‘remember’ blocks extraction in MSA - contra Jarrah (2019). That indicates that the verb is not directly responsible for the extraction phenomenon and its associated intricacies, that is to say, extraction itself depends on the C that is selected by specific types of verbs (i.e., factive and nonfactive). In fact, the verbs themselves are not the immediate c-commanding heads for the extracted subject; that is, C separates the verb from the subject, and as a functional head, the syntactic properties of C may affect the possibility of extraction. A piece of second

² The transcription is modified to pattern along with the examples presented throughout the study.

counterevidence appears in the use of nonfactive verbs of saying such as *qala* ‘say’; albeit a nonfactive verb, extraction is not possible in MSA.

(37) *ʔal-wald-u qala ʔinna al-radʒul-a qad ʔaxa ɗa ʔas-saʕat-a.*
 DEF-boy-NOM say.3M.SG.SBJ that DEF-man-ACC PTCL took.3M.SG.SBJ DEF-watch-ACC
 ‘The boy said that the man took the watch’.

(38)**man ʔal-wald-u qala ʔinna qad ʔaxa ɗa ʔas-saʕat-a.*
 who DEF-boy-NOM say.3M.SG.SBJ that PTCL took.3M.SG.SBJ DEF-watch-ACC
 ‘Who did the boy say that took the watch?’

(39)!*man ʔal-wald-u qala ʔinna-hu qad ʔaxa ɗa ʔas-saʕat-a.*
 who DEF-boy-NOM say.3SG that-he PAR took.3M.SG DEF-watch-ACC
 ‘Who did the boy say that took the watch?’

A third counterevidence appears with how extraction operates across clause-initial Cs; despite lacking a factive and a non-factive verb, extraction is not permitted with clause-initial Cs (41). The observation questions Ross’ (1970) conclusion related to the existence of a hidden verb of saying in such constructions (40).

(40) *ʔinna al-radʒul-a qad ʔaxa ɗa ʔas-saʕat-a.*
 that DEF-man-ACC PTCL took.3M.SG.SBJ DEF-watch-ACC
 ‘Who did (I say) that took the watch?’

(41)**man ʔinna qad ʔaxa ɗa ʔas-saʕat-a.*
 who that PTCL took.3M.SG.SBJ DEF-watch-ACC
 ‘Who did (I say) that took the watch?’

Moreover, even though it is tempting to conclude that Arabic is the only language that hides the performative verb, yet the use of the verb of saying can be relatively used with any complementizer to make it appear like this. However, since in English the use of *that* in clause-initial position is not licensed (42). The prediction cannot pour out from a universal description.

(42) I say that the boy left the house.

(43)*That the boy left the house.

The hidden verb hypothesis does not solve the problem. The use of the verb *ʔaqul* in MSA does not automatically trigger that use of the complementizer. Consider the following instances.

(44) *ʔaqulu ʔal-walad-u qad tarka ʔal-bajt-a.*
 say.1SG.SBJ DEF-boy-NOM PTCL left.3M.SG.SBJ DEF-house-ACC
 ‘I say (that) the boy left the house’.

MSA behaves differently from JA. The complementizers can stand in clause-initial positions and behave differently with non-factive verbs of saying. The observation is that the structure of MSA complementizers makes some complementizers able to stand with or without the presence of hidden verbs, which leaves the following question open: why do certain Cs occupy the root position, while such position is not valid for embedded Cs? We propose that as morphemes, Cs have internal syntactic-semantic features that impact the use of Cs in clause-initial positions.

III. THEORETICAL FRAMEWORK

This part is devoted at presenting a general overview of the theoretical framework that will be adopted for the current study, namely NS (Starke, 2009). NS views grammar as a composite of atomic elements; in a syntactic structure, terminal nodes are not words or morphemes but semantic features. These features are universal in some respect, and language diversity can be explained by understanding how these features are ordered. This view of grammar is developed from the Minimalist Program (Chomsky, 1995), Cartography (Cinque & Rizzi, 2010; Rizzi, 1997), and Distributed Morphology (Marantz, 1997).

In the Minimalist Program, Chomsky (1995) argues that, in the syntax proper, Merge operates recursively to build larger constructions by combining lexical and functional categories. In NS, the same machinery applies. However, Merge operates on smaller units, i.e., submorphemes; a composite of syntactic-semantic features is ordered hierarchically in a binary fashion for spelling out larger constructions. One of the essential elements in nanosyntactic analysis is defining the set of these features that will map onto the derivations of trees. Nanosyntacticians develop three ways to obtain syntactico-semantic features: non-accidental syncretism, containment, and semantic decomposition. For relevance, we present only the first two tests.

Non-accidental syncretism refers to having a morphological form that fulfills more than one grammatical function in a paradigm. Syncretism is a good indicator that one of the morphological forms must be derived from other forms. The right order of morphemes depends on cross-linguistic comparisons of data (Caha, 2009b, p. 99). For example, by examining case syncretism patterns found in Serbian, Slovene, Czech, Ukrainian, Arabic, Latin, and other languages,

Caha (2009b, p. 99) notices that syncretic patterns follow specific hierarchical order: Nominative > Accusative > Genitive > Dative > Instrumental > Comitative. He refers to this order as Universal (Case) Contiguity.

Morphological containment is developed under Distributed Morphology (Marantz, 1997). It is noticed that lexical items can be contained in one another. Thus, in English, for example, we cannot have a suppletive form in the superlative without having a suppletive form in the comparative as well. The adjective *good* becomes *better*, and only then it is realized as *best*. But the case cannot be **good > gooder > best* or **good > better > goodest*. This kind of morphological process is fruitful for nanosyntacticians as it provides them with indicators to which elements are bigger than others.

The core difference between NS and other approaches (cf., Cartography, Cinque & Rizzi, 2010, Rizzi, 1997; and Distributed Morphology, Marantz, 1997) is that “there cannot be a lexicon before syntax and hence syntax does not “project from the lexicon”, syntax rather creates lexical items” (Nanosyntax—What is it?).

Utilizing NS as a framework of analysis, Baunz (2018) proposes that Cs have a functional sequence of semantic atoms. The functional sequence of atoms can be predicted through looking at how factivity affects the selection of Cs in SC, MG, FR. Given the syncretic patterns of Cs across different classes of factivity (emotive factives (F1), semi-factives (F2), nonfactives (F3) and desideratives (F4)), Baunz (2018, p. 154) proposes that Cs have the following hierarchical order:

- (45)
- | | |
|---------------------------------------|------------|
| a. [F1P F1] | MG => /na/ |
| b. [F2P F2 [F1P F1]] | => /oti/ |
| c. [F3P F3 [F2P F2 [F1P F1]]] | => /pu/ |
| d. [F4P F4 [F3P F3 [F2P F2 [F1P F1]]] | => /pu/ |

Based on her view, verbs of factivity can hold three semantic meanings and leads to three types of presuppositions:

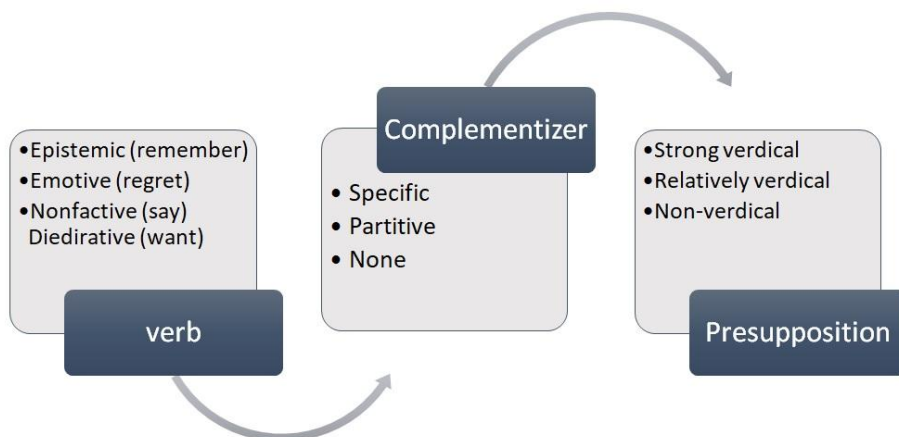


Figure 1. The Effect of Verbs on the Composition of Complementizers

The given diagram shows that there are four verbs distinguished according to their level of factivity. The first two verbs ‘remember’ and ‘regret’ are factive verbs; that is, they show that the speaker is sure to some extent about the content embedded in the utterance. The speaker using the verb ‘remember’ indicates that he/she knows specific information about the embedded utterance (a person cannot remember a situation which has not taken place in the actual world depending on his/her experience). The speaker utilizing the verb ‘regret’ does not provide the same high level of specificity denoted by the verb ‘remember’ simply because the verb ‘regret’ does not indicate the speaker is sure about the action since he/she may build his/her experience on information not necessarily correct, as this information depends on the experience of other agents involved in the actual activity the speaker feels sorry about. The last two types of verbs, viz., ‘say,’ and ‘want’ do not involve any level of specificity. In other words, the speaker using either verb does not suggest that the given information is specific to any extent, he/she assumes a situation to have taken place in the past (i.e., ‘say’) or desire it to take place in the future (i.e., ‘want’). These differences between verbs depending on the denoted specific information are co-related with the different presuppositions according to veridicality (i.e., the assertion of the truth of the utterance). While the first verb ‘remember’ shows strong assertion concerning the truth of the utterance, the verb ‘regret’ does not have the same strength of assertion implied by ‘remember’; hence, it is relatively veridical. The last two verbs, viz., ‘say’ and ‘want’, do not have any assertion, as their semantics implicates; therefore, they are non-veridical.

NS has not been widely utilized in many studies dedicated specifically to Arabic, as far as the authors know, there are only two studies (Abdel-Razaq, 2015; Saeed, 2014), and only the latter uses NS as the primary approach for data analysis. Saeed (2014) utilizes NS toward the decomposition of prepositions. Her study shows prepositions in Arabic are distinguished in two ways at the sub-morphemic level: (i) true prepositions always lexicalize place domain and (ii) semi-prepositions are distributed among path domain (for more details see Saeed, 2014). We deploy the non-accidental

syncretism test to arrive at the functional sequence of Cs in Arabic and compare it with that proposed in the literature (cf., Baunz, 2018). This approach would enable us to arrive at the structure of Cs in Arabic, and it will enable explaining the odd behavior of Cs by matching their internal structure with their external behavior.

IV. ANALYSIS

According to Baunz (2018), the differences between factive verbs are co-related with the strength of presuppositions (i.e., veridicality). While the first verb ‘remember’ shows strong assertion concerning the truth of the utterance, the verb ‘regret’ does not have the same strength of assertion implied by ‘remember’; hence, it is relatively veridical. The last two verbs, viz., ‘say’ and ‘want’ do not have any assertion, as their semantics implicates; therefore, they are non-veridical. In this section, we look at how different degrees of assertion are implicated through the selection of Cs in Arabic.

In MSA, factive verbs (e.g., *ʔata ʔakar* ‘remember’ (46)) and emotive factive verbs (e.g., *nadm* ‘regret’(47)) select *ʔanna*. Non-factive verbs select two Cs; all non-factive verbs (e.g., *ʔɖun* ‘think’(48)) select *ʔanna*, but the non-factive verb of saying (e.g., *ʔaqul* ‘say’(49)) selects *ʔmna*. Desiderative verbs (e.g., *ʔurid* ‘want’(50)) select *ʔan*.

(46) *ta ʔakar-tu ʔanna ʔal-radʒul-a qad ʔaxaa ʔas-saʕat-a.*
remembered-3M.SG that DEF-man-ACC PTCL took.3SG.SBJ DEF-watch-ACC
‘The boy remembered that the man took the watch’. (MSA)

(47) *nadm-tu ʔanna ʔal-walad-a qad tarka ʔal-bajt-a.*
regretted-I that the-boy-ACC PTCL left.3M.SG.SBJ the-house-ACC
‘I regretted that she came to our house’. (MSA)

(48) *ʔɖunnu ʔanna ʔal-walad-a qad tarka ʔal-bajt-a.*
think. 1SG.SBJ that the-boy-ACC PTCL left.3M.SG.SBJ the-house-ACC
‘I think that the boy left the house’. (MSA)

(49) *ʔaqulu ʔmna ʔal-walad-a qad tarka ʔal-bajt-a.*
say.1SG.SBJ that the-boy-ACC PTCL left.3M.SG.SBJ the-house-ACC
‘I say that the boy left the house’. (MSA)

(50) *ʔuridu ʔan j-uyadira ʔal-manzil-a.*
want.1SG.SBJ SUBJ PRES-leave.3M.SG.SBJ the-house-ACC
‘I want him to leave the house’. (MSA)

In JA and LA, only one C is used across all types of verbs. In JA, *ʔmn* is used with factives (51), emotive factives (52), non-factives, including verbs of non-factive verbs of saying (54), and desiderative verbs (55), while in LA *ʔanno* is selected.

(51) *ta ʔakar-t ʔmn-ha ʔil-bmt ʔiɖa-t maʕ-na.*
remember-I COMP-3F.SG the-girl came-3F.SG with-us
‘I remembered that the girl came with us’. (JA)

(52) *nadm-t ʔmn-ha ʔil-bmt ʔiɖa-t maʕ-na.*
regretted-I COMP-3F.SG the-girl came-3F.SG with-us
‘I regretted that the girl came with us’. (JA)

(53) *baɖun ʔmn-u ʔil-walad ʔiɖa.*
think.I COMP-3M.SG the-boy arrived.3M.SG.SBJ
‘I think that the boy left the house’. (JA)

(54) *bagul ʔmn-u ʔil-walad ʔiɖa.*
say.1SG COMP-3M.SG the-boy arrived.3M.SG.SBJ
‘I say that the boy arrived’. (JA)

(55) *bidi ʔmn-u ʔil-walad j-idzi hassa.*
want.1SG COMP-3M.SG the-boy PRES-come.3M.SG.SBJ now
‘I want the boy to come now’. (JA)

Across Arabic varieties, the complementizer systems show almost a similar pattern. In declarative structures, two complementizers *le* and *ta* are used in Sason Arabic (SnA) (Akku, 2018)³:

(56) *a-ref le Kemal ja.*
1SG-know COMP Kemal came.3M.SG
'I know that Kemal came'. (SnA)

(57) *ali iddz i-si le a-habb-u.*
Ali claim 3M-do COMP 1SG-love-him
'Ali claims that I love him'. (SnA)

(58) *ma-sima-tu le go zya.*
NEG-heard-1.SG COMP came.3.PL children
'I did not hear that the children came'. (SnA)

(59) *irə-nni leyla ta tə-či.*
want-1SG Leyla SUBJ 3F-come
'I want Leyla to come'. (SnA)

The following table summarizes the findings and orders Cs occupying the same contexts in different varieties of Arabic.

TABLE 2
SYNCRETIC USE OF COMPLEMENTIZERS CROSS-LINGUISTICALLY

	Emotive Factive	Semi-factive/nonfactive	Nonfactive 'say'	Desiderative
MSA	ʔanna	ʔanna	ʔinna	ʔan
SnA	le	le	le	ta
JA	ʔin	ʔin	ʔin	ʔin
LA	ʔanno	ʔanno	ʔanno	ʔanno

Table 2 shows that in MSA *ʔanna* is syncretic with emotive and semi factive verbs, in SnA *le* appears in three cells, viz., emotive, semi- and non-factives. In JA and LA, the Cs *ʔin* and *ʔanno* are syncretic across all types of verbs. In Arabic, the syncretic patterns can be represented following the same mechanism as in (60).

(60)

	MSA	JA	LA	SnA
a. [F1P F1]	=> /ʔan/	/ʔin/	/ʔanno/	/ta/
b. [F2P F2 [F1P F1]]	=> /ʔinna/	/ʔin/	/ʔanno/	/le/
c. [F3P F3 [F2P F2 [F1P F1]]]	=> /ʔanna/	/ʔin/	/ʔanno/	/le/
d. [F4P F4 [F3P F3 [F2P F2 [F1P F1]]]]	=> /ʔanna/	/ʔin/	/ʔanno/	/le/

The Arabic data supports the ordering of F1>F2>F3>F4. This pattern has two implications. First, it avoids *ABA patterns of syncretism. Second, it shows morphological containment; Cs can be realized as morphological patterns that contain each other.

Relative to Baunz' (2018) description of veridicality, in MSA, since *ʔanna* is used with factive and non-factive verbs, it is selected by strongly veridical (46), relatively veridical (47) and non-veridical verbs (48). *ʔinna* is used only with the non-factive verb of saying; hence, *ʔinna* is selected by a non-veridical verb. In addition, *ʔan* is selected by a non-veridical verb of wanting. The data suggests that the C system in MSA departs slightly from Baunz (2018). While the system matches veridical verbs, the system shows three divisions in non-veridical ones. Accordingly, in MSA, non-veridical verbs of saying differ from other non-veridical verbs. The former is non-veridical relative to the cognizant, the sayer, while the later is non-veridical relative to the speaker. The two types do exist along with non-veridical verbs of wanting noticed by Baunz (2018). The system of MSA can be represented as thus.

(61) Veridical NV speaker NV cognizant NV subjunctive
ʔanna ʔanna ʔinna ʔan

Ross's (1970) pioneering work has been questioned on the validity of the assumption that points toward embedding clauses with speech act projections. While many disagree with Ross' idea of embedding specific predicates (Speas & Tenny, 2003, p. 323), studies confirm the idea of embedding of speech act roles (Haegeman, 2014, Wiltschko & Heim, 2016) and support its existence through the different types of Arabic Cs (Speas & Tenny, 2003, p. 323-324) that can undergo embedding. We argue that embedding speaker roles overlooks a central component, the cognizant, sayer. According to Vandelanotte (2004), the sayer differs from the speaker and can contribute to the grammaticality of structures. His view can be straightforwardly illustrated in cases of reported speech. See the example in (62) for the temporal deixis shift.

(62) {Speaker {he is late.}} >
{Sayer {He was late.}} >

³ The data is insufficient to provide final conclusions. However, it enables drawing upon predictions to be verified.

{Sayer {he was late {Speaker {he is late}}}}

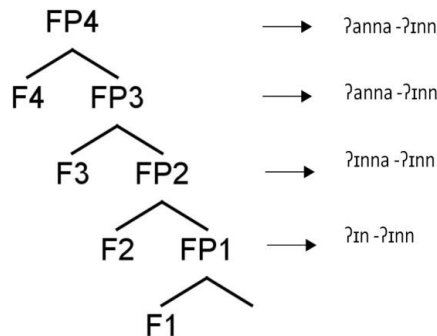
The view that a cognizant can be part of syntactic structures can enhance the Speech Act Layer proposed by Haegeman (2014) and Wiltschko and Heim (2016). The impact of factivity and its correlated interpretation across veridicality shows that Cs in Arabic may trigger different strengths of presuppositions. See the table below.

TABLE 3
THE IMPACT OF VERIDICALITY ON ARABIC COMPLEMENTIZERS

	Veridical	Nonveridical speaker	Nonveridical cognizant	nonveridical subj
SA	ʔanna	ʔanna	ʔinna	ʔan
JA	ʔinn	ʔinn	ʔinn	ʔinn
LA	ʔanno	ʔanno	ʔanno	ʔanno

Contiguous complementizers cannot be used in a clause-initial position as their function can overlap with other complementizers. How veridicality is transmitted affects the choice of the complementizer. That is, the same complementizers have different morpho-phonological realizations in the given Arabic varieties. While in MSA different complementizers are used for diverse functions, hence, they occupy different positions in the tree, and these positions are restricted in the sense that the same complementizer cannot be used in more than its designated functional head (with the exception of *ʔanna* which has two functions). The same situation does not exist in JA and LA since there is one and only one lexical form used to cover all the semantic functions related to the different types of complementizers in MSA. The same form cannot be used to cover all the functions pertained to the diverse complementizers to avoid ambiguity and vague representation of the presupposed information. The result we may reach is that the more functions the lexical form is supposed to serve the more restricted environments it can occur in since that requires more cognitive processing for the same lexical expression to be integrated into the structure from the side of the speaker, and more cognitive decoding it demands from the listener. There are four semantic levels pertaining to the veridical functions of the complementizers associated with particular verbs, as explained in (60) above. These different levels can be represented as in (63).

(63)



To make this representation more concrete and related to the presupposition argued above, we reflect how this mapping encodes specificity and the orientation. FP1 complementizers and FP2 complementizers can be used in MSA without a verb since they are not contiguous. That explains why *ʔinna* can be used in MSA in an initial position; however, in JA and LA, it is banned. In MSA, we argue that it is spelled out as FP2 while in JA, it can be spelled-out as FP1, FP2, FP3, and FP4 due to its syncretic form. Moreover, the internal structure of *ʔinna* enables it to be used without a verb of saying being present as presupposed information can be recovered from the semantic composition of the complementizer itself. *ʔanna*, on the other hand, cannot be used in the initial position for the same reason like JA *ʔinn*.

Since *ʔanna* contains two semantic blocks of veridicality (FP3 and FP4), recovering a presupposition would not be possible because the veridicality could target either FP4 or FP3. In this sense, it becomes evident that Arabic complementizers are not motivated solely by the use of a verb that is only found in Arabic, but it is more related to their use as triggers of presuppositions.

The advantage of this prediction over Ross’s (1970) conclusion is that even though he rightfully predicts the use of *ʔinna* after the cognizant verb of saying, he does not resolve the problem of why this is not applicable to other verbs.

- (64) [ʔurid ‘want’ {hidden} [[ʔan/*ʔanna/*ʔinna
 [ʔataʕʕzb ‘wonder’ {hidden} [ʔanna/*ʔinna/*ʔan
 [ʔaʕta ʔr ‘remember’ {hidden} [ʔanna/*ʔinna/*ʔan
 [ʔaql ‘say’ {hidden} [*ʔanna/ʔinna/*ʔan

We argue that the reason is attributed to the same factor that blocks and allows extraction in complex constructions, viz., the complex morpho-semantic composition of complementizers establishes them as strong morphemes that can be used in clause-initial positions, positions that enable the recoverability of presupposed information.

V. IMPLICATIONS

Comparing the C system in Arabic with that proposed by Baunz (2018), we show that the fundamental insight of is fruitful for describing Arabic Cs; nevertheless, the proposed universal hierarchy does not fit neatly. In MSA, *?anna* appears with factives and non-factives; however, it cannot be used with all non-factives; that is, instead of using *?anna* with the non-factive verb of saying, *?inna* is used. This can be illustrated in the table below

TABLE 4
CROSS LINGUISTIC COMPARISON OF THE SELECTION OF CS

	Emotive Factive	Semi-factive	Nonfactive	Desiderative
MG	pu	pu/oti	oti	na
SA	?anna	?anna	?anna/?inna	?an
SnA	le	le	le	ta
JA	?in	?in	?in	?in
LA	?anno	?anno	?anno	?anno
FR	que	que	que	que
SC	da	da	da	da

While the Cs in MSA depart from the distinction above; they meet with Cs cross-linguistically, if non-factives are broken into two types: speaker-oriented non-factives and cognizant-oriented non-factives. In that sense, *?inna* is distinguished from *?anna*. The distinction is supported by looking at how veridicality interacts with factivity as we have illustrated in Table 4 above. We, therefore, predict that it is more appropriated to follow the following pattern: veridical > nonveridical SPEAKER > nonveridical COGNIZANT > nonveridical SUBJ to end up with an inclusive hierarchy. The change to that universal picture is supported by the fact that nonveridically should not be limited to the speaker, but it could be viewed from the point of view of the cognizant.

VI. CONCLUSIONS

The study concludes with three observations. First, the choice of Arabic complementizers is not random and is affected by a presupposition that can be triggered on the side of the speaker, hearer and cognizant. Thus, Arabic complementizers are strongly affected by veridicality. Second, the study shows that MSA has more articulation of complementizers than JA and LA. Therefore, it permits the use of *?inna* in initial positions. The research also attributes the ungrammaticality of using *?anna* in that position to its size. *?anna* is argued to be bigger and shares contiguous semantic blocks, viz., it can be used with veridical and nonveridical SPEAKER. Thus, it can trigger different presuppositions, unlike *?inna*. The same argument is extended to JA. *?inn* cannot be used because it is syncretic across four semantic domains: veridical > nonveridical speaker > nonveridical cognizant > nonveridical subj. Using it, therefore, without a verb that delimits the different kinds of presuppositions can lead to an uncontrolled interpretation of presuppositions. Third, the study shows that Arabic complementizers follow the universal pattern proposed by Baunz (2018) and refines the theorem of hidden verbs that have been proposed by Ross (1970).

Compared with the findings reached in this paper concerning the distinctions above, Fassi-Fehri (2012) provides a relatively supportive distinction between Cs based on the suitability of the complementizer to exist in a specific position within the clause and its syntactic properties (i.e., Mood, Logophoricity and Case). We find that the given distinctions above are informative and supportive of the distinctions proposed in the paper. The study shows that Baunz's (2018) distinction of non-factive verbs should be extended since the C system of MSA shows that there are two types of nonveridically that is associated with non-factive verbs in Arabic, one related to the speaker while the other is related to the cognizant. The conclusion shows that Fassi-Fehri's (2012) logophoricity feature impacts the decomposition of Cs and provides a clear path for understanding when a C can obtain the +/- Root distinction. More data from other Arabic varieties, e.g., Moroccan Arabic, Algerian Arabic, Sudanese Arabic, Gulf Arabic, can strengthen these conclusions, yet we leave for further studies.

REFERENCES

- [1] Abdel-Razaq, I. (2015). *Who is What and What is Who: The Morphosyntax of Arabic WH*. Cambridge Scholars Publishing.
- [2] Akku, F. (2018). *The left Periphery of Sason Arabic* (unpublished). University of Pennsylvania.
- [3] Baunz, L. (2018). Decomposing Complementizers: The functional sequence of French, Modern Greek, Serbo-Croatian, and Bulgarian Complementizers. In L. Baunz, K. De Clercq, L. Haegman, & E. Lander (Eds.), *Exploring Nanosyntax*. Oxford University Press.
- [4] Beaver, D. I., & Geurts, B. (2014). Presupposition. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2014). Retrieved June 2018 from <https://plato.stanford.edu/archives/win2014/entries/presupposition/>
- [5] Chomsky, N. (1995). *The Minimalist Program*. MIT Press.
- [6] Cinque, G., & Rizzi, L. (2010). *Mapping Spatial PPs: The Cartography of Syntactic Structures*. Oxford University Press.
- [7] Clements, G. (1975). The logophoric pronoun in Ewe: Its role in discourse. *Journal of West African Languages*, (1)1, 141–177.
- [8] Fassi-Fehri, A. F. (2012). *Key Features and Parameters in Arabic Grammar*. John Benjamins Publishing.
- [9] Habib, R. (2009). The Syntax of the Standard Arabic Particles ?an and ?anna. *Selected Papers from the 2006 Cyprus Syntaxfest*. Cambridge Scholars Publishing. Newcastle upon Tyne.

- [10] Haegeman, L. (2014). West Flemish verb-based discourse markers and the articulation of the speech act layer. *Studia Linguistica*, 68(1), 116–139.
- [11] Hooper, J. B., & Thompson, S. A. (1973). On the Applicability of Root Transformations. *Linguistic Inquiry*, 4(4), 465–497. Retrieved from JSTOR.
- [12] Jarrah, M. (2019). Factivity and subject extraction in Jordanian Arabic. *Lingua*, 219, 106–126. <https://doi.org/10.1016/j.lingua.2018.12.002>
- [13] Marantz, A. (1997). No Escape from Syntax: Don't Try Morphological Analysis in the Privacy of Your Own Lexicon. *University of Pennsylvania Working Papers in Linguistics*, 27.
- [14] Nanosyntax—*What is it*. (n.d.). Retrieved October 13, 2019, from <http://nanosyntax.auf.net/whatis.html>
- [15] Persson M., (2002). *Sentential Object Complements in Modern Standard Arabic*. Almqvist & Wiksell International. Sweden.
- [16] Rizzi, L. (1997). The Fine Structure of the Left Periphery. In *Kluwer International Handbooks of Linguistics. Elements of Grammar* (pp. 281–337). https://doi.org/10.1007/978-94-011-5420-8_7
- [17] Ross, J. (1970). On Declarative Sentences. In R. Jacobs & P. Rosenbaum, *Readings in English Transformational Grammar* (pp. 22–227). Ginn and Company.
- [18] Saeed, S. (2014). *The Syntax and Semantics of Arabic Spatial Prepositions*. *Nordlyd*, (1)23, 1-20.
- [19] Starke, M. (2009). Nanosyntax A short primer to a new approach to language. *Nordlyd*, (1)6, 1-10. .
- [20] Vandelanotte, L. (2004). Deixis and grounding in speech and thought representation. *Journal of Pragmatics*, 36(3), 489–520. <https://doi.org/10.1016/j.pragma.2003.10.003>
- [21] Wiltschko, M., & Heim, J. (2016). The syntax of confirmationals: A neo-performative analysis. In G. Kaltenböck, E. Keizer, & A. Lohmann (Eds.), *Studies in Language Companion Series* (Vol. 178, pp. 305–340). <https://doi.org/10.1075/slcs.178.11wil>

Saleem Abdelhady is an assistant professor of English Language and Communication at the American University of the Middle East, Kuwait. Dr. Abdelhady obtained his Ph.D. from Memorial University of Newfoundland, St. John's Canada. He has published several articles in sociolinguistics, syntax, and pragmatics.