Phasehood of Wh-Questions in Modern Standard Arabic

By

Haroon Nasser Alsager

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved March 2017 by the
Graduate Supervisory Committee:

Elly van Gelderen, Chair
Karen Adams
Mark A. James

ARIZONA STATE UNIVERSITY

May 2017

ABSTRACT

Wh-questions have been widely discussed in different languages such as English, Mandarin Chinese, Italian, and Russian, but little attention has been paid to the structure of wh-questions in Modern Standard Arabic (MSA). Thus, this dissertation attempts to analyze the structure of wh-questions using the current frameworks: *Minimalism* and *Cartography*.

In the late 1990s, Chomsky established the Minimalist Program which aims to describe the clause structure in as simple and economic mechanism as possible, and he advanced his famous research program to include phase theory, which aims to restrict the syntactic operations. On the other side, Rizzi (1997, 2001) proposed the Cartographic approach. In this approach, Rizzi attempted to analyze the left periphery domain in detail, and suggested the split CP hypothesis. Following those two approaches, Ginsburg (2009) and Totsuka (2015) unified them into one approach and suggested that ForceP, TopicP, and IntP are phasal domain while FocusP, FinP, and WhP are not. An overview of the Chomskyan model and Rizzi's approach has been provided in Chapter 2. Also, this dissertation discussed the unified approach by Ginsburg (2009) and Totsuka (2015).

In addition to the overview of the general frameworks, this dissertation discussed the clause structure such as the word order, left periphery domain (i.e., CP), and resumption in MSA. Furthermore, Chapter 2 presented the earlier studies on the wh-questions in MSA and highlighted the major gap which this dissertation attempts to fill. In these studies the structure of wh-questions in MSA were mis-analyzed because the

surface structure of the nine wh-questions might look the same, but, in fact, they are not. Therefore, this dissertation attempts to (re)study the structure of wh-questions with taking into consideration the resumption and [definiteness].

In Chapter 3, the methodology and corpus analysis, which is used in collecting the wh-questions in MSA, are discussed. Finally, Chapter 4 analyzed the corpus findings based on the unified approach by Ginsburg (2009) and Totsuka (2015) and showed some evidence that *man* 'who' and *ayy* 'which' questions in MSA are in phasal phrase (i.e., IntP) while the rest of wh-questions are not.

TABLE OF CONTENTS

iv

LIST OF TABLES

ABBREVIATIONS

Abbreviations Used in Glosses

2	second person
3	third person
ACC	accusative
COMP	complementizer
F	feminine
FUT	future
M	masculine
NOM	nominative
Pl	plural
S	singular
CA	Classical Arabic
CP	complementizer phrase
D	determiner
ABC	Aoun, Benmamoun, & Choueiri
IP	inflectional phrase
TP	tense phrase
MP	Minimalist Program
MSA	Modern Standard Arabic
NP	noun phrase
PP	prepositional phrase

SVO	subject verb object
T	tense
VP	verb phrase
VSO	verb subject object
PIC	Phase Impenetrability Condition
RP	Resumptive pronoun
FL	Faculty of Language
C-I	Conceptual-Intentional system
LF	Logic Form
SM	Sensorimotor system
PF	Phonetic Form
LexicalA	Lexical Array
N	Numeration
uF	uninterpretable feature
iF	interpretable feature
Q	Question feature (interragtive feature)
Wh	wh- feature
BYU	Brigham Young University
N/A	not avalible

CHAPTER 1: INTRODUCTION

In this chapter, the purpose of the inquiry is introduced. Also presented is the scope of the research, considered a framework, and the language investigated. Furthermore, the methodology and the corpus analysis used in the dissertation for the data collection are explained. Finally, the chapter ends with a roadmap for the remainder of the dissertation, and a summary of the main concerns and results is provided near the end of this chapter.

Purpose of the Study

During the past two decades, numerous research papers discussing issues in Modern Standard Arabic or its many dialects have been written, e.g. on word order, agreement, negation, and case theory, yet little attention has been paid to wh- questions. In this paper, the structure of wh- questions in Modern Standard Arabic is discussed. Also analyzed are wh- questions based on the Minimalist Program (MP) by Chomsky (2000, 2001, 2008, 2013, 2015) and the Cartography approach by Rizzi (1997, 2001). Advancing the Minimalist program, Chomsky suggested phase theory, which restricts the syntactic operations. In contrast, Rizzi (1997, 2001) proposed the split C(omplementizer) P(hrase), the aim of which was to describe in detail the discourse domain (CP). Thus, in this dissertation, it was attempted to combine the two approaches (i.e., Chomsky's approach and Rizzi's approach) and to explain why we have flexibility in some of the wh-questions (i.e., maan 'who', and ayyu 'which') and not in other wh-questions.

Scope of the Research

One of the crucial questions that has received considerable attention in the literature of syntax during the last four decades is the interrogative structure (i.e., yes/no and wh-questions) within the syntactic theoretical frameworks. Chomsky (1995) talked about the [Q]uestion feature, which is a strong feature and located in the head of the CP, and it is responsible for triggering the yes/no or the wh-words to the CP head. Furthermore, Rizzi (1997, 2001) stated that three layers in his system for the interrogative items, (a) InterrogativeP, (b) FocusP, and (c) WhP. In the literature of Arabic syntax, most of the works dealt with the wh-questions as if they had same structure and/or the same syntactic behavior. Thus, in this dissertation the aim was to answer the following questions:

1. What is the structure of the nominal wh-questions in simple and complex clauses?
2. What is the structure of the adverbial wh-questions in simple and complex clauses?
3. What is the relation between gap and resumptive strategy, and wh-words?
4. What is the role of the resumptive pronoun in phase theory?

Language Investigated

Modern Standard Arabic (MSA) is the language which is investigated in this dissertation. MSA is one of the Semitic languages which descended from the Afro-Asiatic family of languages. It is the native language of more than 300 million people living in Middle East and north Africa (Gordon, 2005).

Many researchers claimed that MSA is derived from Classical Arabic (CA). Both MSA and CA share morphological and syntactic structures; however, each exhibits some

differences from the other in respect to vocabularies and stylistic features. Versteegh (1984) claimed that CA was the standard version of Arabic during the Islamic era. He mentioned that CA has remained unchanged between the 7th and 20th centuries due to the dominating belief (i.e., Islamic era).

In 1973, El-Said Badawi was the first researcher who investigated Arabic and its dialect, using his knowledge of the traditional Arabic grammar combined with the modern linguistic theories. He worked on Egyptian Arabic, and he established a new term in the literature of Arabic syntax called *fusha Al-asr* (Modern Standard Arabic).

Generally speaking, MSA is the formal language that is used in many Arab countries. It is the only variety of Arabic that is used in school. It is the language for the media and communication. Also, it is the language used in almost all printed documents, such as newspapers, magazines, books, and official government letters.

Methodology and Corpus Analysis

Wh-questions in the literature of syntax have been characterized as nominal and adverbial wh-questions. In the literature of Arabic syntax, the nominal wh-questions are *man* 'who', *ayyu* 'which', *ma* and *matha* 'what', and *kam* 'how many/much' while the adverbials are *mataa* 'when', *ayn* 'where', *kayf* 'how', and *limatha* 'why'. In order to investigate the syntactic differences between those two groups, the structure of nominal and adverbial wh-questions were analyzed, taking into consideration wh-movement, resumptive pronouns and gaps, and the other functional projections in the left periphery domain.

3

To analyze the structure of wh-questions in MSA, an Arabic corpus was utilized, that focused on the formal texts. The Brigham Young University (BYU) Arabic Corpus, current the largest Arabic corpus, was used. It is a free web-search engine that has 173,600,000 tokens. The BYU Arabic corpus is an untagged and unparsed corpus, which means that it has only raw texts. The corpus is divided into six main genres. In this dissertation, only three of those genres that fulfill the definition of MSA were included. I used 100 examples from each genre and for each of the wh-questions.

Most of the data appear in the simple or relative clauses (which is fine; however). To strengthen the analysis, I chose to have another source of data. Examples were used for the wh-questions in different complex clauses that the BYU Arabic corpus did not provide as wh-words in *anna*-clauses. Most of the examples provided are grammatical, based (a) on my judgment as a native speaker of Arabic from Saudi Arabia, (b) on other native speakers from Saudi Arabia, Jordan, Palestine, Egypt, and Iraq; or (c) because they can be found in traditional and/or modern grammar books.

The concordance and structure frequency of wh-questions have been studied. The findings of corpus analysis show an interesting and clear division between *man* and *ayy-a/u* question words and the rest of the wh-questions.

Organization

This dissertation is organized as follows: Chapter 2 is focused on the theoretical framework. First, an overview of the first approach used in this dissertation, the Minimalist Program is presented. Following that, an advanced version of MP, the Phase Theory, is discussed. Next is an introduction to the second approach used, which is the

split CP hypothesis. After introducing the two approaches, it was time to unify them in one approach or framework, which is the phasehood in the CP domain. This is followed by a discussion of the syntactic issues that play a major role in the structure of Wh-questions in MSA, such as clause structure in MSA, resumptive pronouns (RPs; i.e., base-generation approach vs. movement approach), functional projections in the CP domain, and wh-island. Chapter 2 concludes with the earlier studies on Wh-questions in MSA.

In Chapter 3, the methodology and corpus analysis, which is used in collecting the Wh-questions in MSA, are discussed. In this chapter, a brief overview of corpus analysis as one of the linguistics methodologies is introduced. Following that, there is a discussion of the corpus that is used in this study. In this dissertation, the corpus-assisted approach was used, which means that the corpus is one of the data collection sources. As mentioned in this dissertation, the BYU Arabic Corpus was used as one of data sources, as were grammatical examples based either on my and other native speakers' judgment or that can be found in traditional and/or modern grammar books. Finally, concordance and structure frequency were the corpus tools that were used to answer this dissertation's questions.

In Chapter 4, Chapters 2 and 3 were combined by analyzing the corpus findings based on the theoretical frameworks that were established this dissertation. Also included in Chapter 4, in order to strengthen my analysis, are the non-corpus findings. Chapter 4 is divided into two main sections. In the first section, the wh-phasal phrase,

such as the *man* 'who' and *ayy-a/u* 'which' are analyzed. In the second section, the non-wh-phasal phrase is studied.

Finally, in Chapter 5, the dissertation is concluded with a summary and some suggestions for further studies.

CHAPTER 2: WH-QUESTIONS: THEORETICAL FRAMEWORKS

In this chapter, an overview of the general frameworks which were adopted for this dissertation, such as the MP, the phase theory by Chomsky (1995, recent work), and the split CP are discussed.   The two approaches were started in the mid-1990s. The phase theory had been developed in Chomsky's works (2000, 2001, 2008, 2013, 2015), while the split CP hypothesis was advocated in Rizzi (1997, 2001). The phase theory (or the MP) is focused on the elementary syntactic operations ,while Rizzi's system is focused on drawing as detailed as possible maps of the syntactic relations. Section (2.4) discusses The phasehood in CP is discussed in a later section. Following that is a discussion about the clause structure in Modern Standard Arabic (MSA), the earlier works on the wh-question and left dislocated structures in Modern Standard Arabic. Finally, as part of the wh-structure, the major developments in the analysis of resumption will be discussed.

Minimalist Program

The syntactic theories have gone through several changes during the last five decades. The Minimalist Program is one of the syntactic frameworks that were developed by Chomsky (1995, 2001, 2004, 2008, 2013, 2015). Chomsky started the MP by defining the faculty of language (FL). He stated that the Faculty of Language has two interfaces: (a) the Conceptual-Intentional system (C-I or LF) and (b) the Sensorimotor system (SM or PF), with the syntax governing the relation between the two interfaces.

The MP is based on the bottom-up derivational structure. According to Chomsky (1995, p. 225), the derivational structure starts by *Select* the lexical items (syntactic

7

objects) from the *Lexical Array* (LexicalA) (or *Numeration* (N)). The syntactic objects

are combined by the syntactic operation *Merge*. After that, they are mapped to the C-I

and SM interfaces. The mapping operation of the syntactic objects to C-I and SM is

called *Transfer* or *Spell-Out*. Example (1) shows the syntactic operation *Merge* in MP:

(1) Merge (a, b) => {a, b}.

a   b
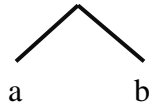
Merge (y, (a,b)) => {y{a, b}}

y

a   b
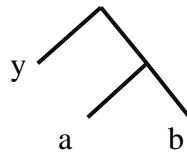
According to MP, *Merge* has two types: (a) External merge (EM), and (b) Internal

merge (IM). The external merge is responsible for combining two lexical items from the

LexicalA in a set, and this set could be also combined with another lexical item from the

LexicalA by the EM operation. On the other hand, internal merge (IM) is known as

movement, which could be defined as moving/internal merging of lexical item from an

established set to another/higher position to satisfy *Agree*/*Valuation* operation. Examples

(2) a. and (2) b. show the External Merge and Example (2) c. presents the Internal Merge

in Chomsky's system.

(2) a. Select X and Y from N.

Merge X and Y.

```
        X
       / \
      X   Y
```
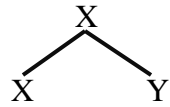
b. Select Z from N

Merge Z with X

```
       / \
      Z   X
         / \
        X   Y
```
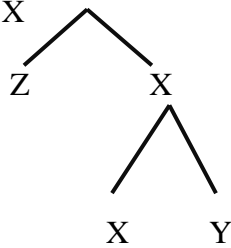
c. Select Y from X

Merge X and Y

Move/IM of Y to check/value the unvalued feature [uF]

```
        XP
       /  \
      Y    X
          / \
         X   Y
```
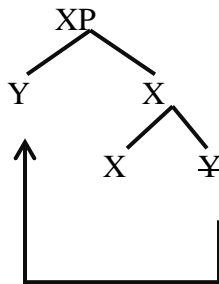
An important operation in the Minimalist Program is the *Agree/feature checking* operation. There are two types of features: (a) the interpretable features and (b) the uninterpretable features. The interpretable features [iF] have a semantic content while the uninterpretable features [uF] do not have a semantic content. The uninterpretable features need to search/Probe for their interpretable features/Goal in order to value/check

9

their [uF] before they are deleted. This operation (i.e., Probe-Goal configuration) is

called *Agree/feature checking* in MP (Chomsky, 1995).

To sum up, the phrase structure initiates the derivation by the *Select* operation,

which picks the lexical items from the *Lexical Array/ Numeration* to build the clausal

structure. Following the *Select*, *Merge* as a minimalist operation, combines two lexical

items from the *Lexical Array* using the EM as in Example (3). After that, the

uninterpretable features must be valued/checked through the minimalist operation called

Agree. The valuation could be processed through the IM, after which the [uF] could be

deleted after the valuation as in Example (4). Finally, the clause could now be mapped to

the C-I and SM interfaces through the minimalist operation called *Transfer/Spell-Out* as

in Example (5).

(3) Lexical Array = {book, bought, Ali, the}

(4)

TP
Ali   T'
[EPP]
[iPhi]
[uCase]
T   vP
[NOM]
[iT]
[uPhi]
Ali   v'
v   VP
V   DP
bought   the book
[ACC]   [iPhi]
[uPhi]   [uCase]

(5)

CP
C'   => Spell-Out/Transfer
C   TP

(Citko, 2014)
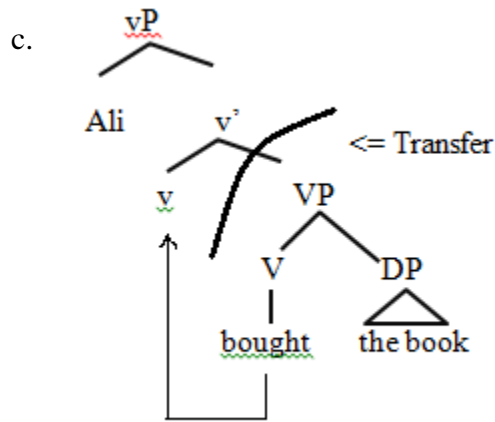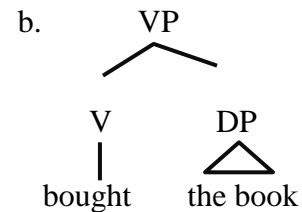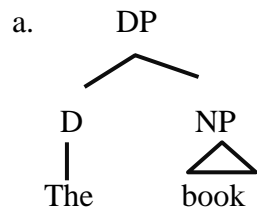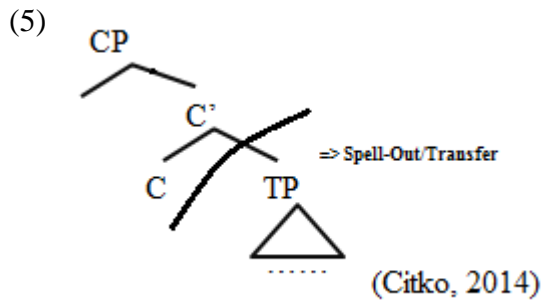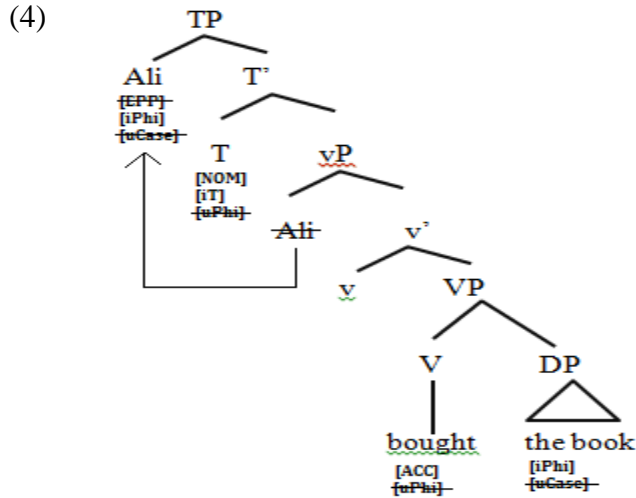
Cartographic Approach

In the early studies of syntax, the clause structure consisted of three main layers
(CP, TP, and VP). These layers were described as follows in Example (6):

(6) a. VerbP is the lexical layer, which is headed by the verb, and it is the layer

where the theta roles take place.

b. TenseP is the inflectional layer, which is headed by an abstract T, and it is

responsible for assigning features such as case and phi-features.

c. ComplementizerP is the pragmatic layer, which is headed by C, and it is

responsible for hosting different types of complementizers such, as

question markers, relative pronouns, focalized and/or topicalized elements,

and so forth.

This assumption of having only three layers turned out to be an oversimplification of the clause structure. As a result, each of these layers went through different stages of classification. Kayne (1994) and Larson (1988) suggested that the VP layer should split to more than one projection. It could consist of vP, VP, and RootP. Pollock (1989) followed the notion of split layers and suggested the IP should also split into AgrP, TP, AspP, and MoodP. Finally, Rizzi (1997, 2001) and others claimed that the CP must follow the same path as in the VP and IP. He claimed that the CP consisted of the following layers as in Example (7):

(7)　　　　　　　ForceP (TopicP) FocusP (TopicP) FinP IP VP.

(Rizzi, 1997, p. 297)

12
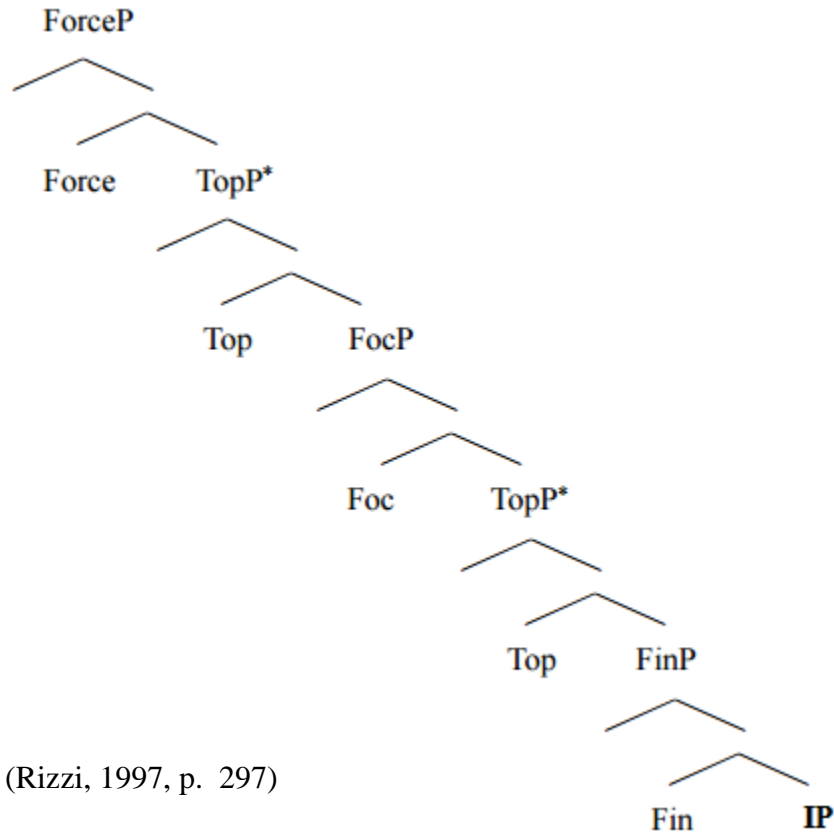
Rizzi (1997, 2001) described the ForceP and FinP as the selectional layer, which is responsible for communicating between the higher and lower structural layers. These two layers are responsible both for showing the properties of the verbal layer in the FinP and the mood of the clause in ForceP (i.e., indicative, interrogative, imperative, etc.). For example, if the complementizer *that*/*che* is used in a clause in English/Italian, its embedded clause must be a finite clause; and if the complementizer *for*/*di* is used in a clause in English/Italian, the embedded clause must be a non-finite clause. Example (8) shows that there is a relation/connection between the Fin/Force head and the T head by choosing its finiteness feature [+/-Fin].

(8) a. Credo   che loro apprezzerebbero molto il tuo libro

      'I believe that they would appreciate your book very much'

   b. Credo   di apprezzare   molto il tuo libro

      'I believe   of to appreciate your book very much'

(Rizzi, 1997, p. 288)

Rizzi (1997) discussed the position of the two complementizers, *che* and *di*, in Italian. He found out that the complementizer *che* must precede a TopicP as in Examples (9) a and b, while the complementizer *di* must be preceded by a TopicP as in Examples (9) c and d).

(9) a. Credo   **che** il tuo libro, loro lo   apprezzerebbero molto

      'I believe that your book, they would appreciate   it a lot'

   b. *Credo,   il tuo libro, **che** loro lo   apprezzerebbero molto

      'I believe, your book, that they would appreciate   it a lot'

13

c. *Credo **di** il tuo libro, apprezzarlo molto

'I believe 'of' your book to appreciate it a lot'

d. Credo, il tuo libro, **di** apprezzarlo molto

'I believe, your book, 'of' to appreciate it a lot'

This suggests that the comp *che* is sitting in the matrix head of the CP (i.e., in the Force

head), while the comp *di* is located in the lowest head of the CP (i.e., in the Fin head).

On the other hand, Rizzi (1997) mentioned that as we have restricted and fixed

layers, such as the ForceP and FinP, we also have free and independent layers in the CP

systems, such as TopicP and FocusP. He pointed out that we can have more than one

topicalized element in the same clause, while there must be one focalized element in a

clause. Furthermore, the topicalized element and the focalized element could occur

together in the same clause. Example (10) shows the hierarchical structure of TopicP and

FocusP in Rizzi's system.

(10) a. Il libro, a Gianni, domani, gliero daròsenz'altro

"The book, to John, tomorrow, I'll give it to him for sure"

b. *A GIANNI IL LIBRO darò (non a Piero, l'articolo)

"TO JOHN THE BOOK I'll give, not to Piero ,the article"

c. A Gianni, QUESTO, domani, gli dovrete dire

"To Gianni, THIS, tomorrow, you should tell him"

An important component of Rizzi's (1997) approach is the position/behavior of

the Wh-word in the split CP hypothesis. He pointed out that the wh-word could be

preceded by a topicalized element, but not followed by a TopicP. Furthermore, the wh-

word must not appear in a clause with a focalized element. Example (11) presents the

position of the wh-word che cosa "what" in the split CP.

(11) a. A Gianni, che cosa gli hai detto?

      To Gianni, what   did you tell him?

   b. *Che cosa, a Gianni, gli hai detto?

      What,   to Gianni, did you tell him?

   c. *A GIANNI che cosa hai detto (, non a Piero)?

      TO GIANNI what did you tell (, non to Piero)?

   d. *Che cosa A GIANNI hai detto (,non a Piero)?

      What   TO GIANNI did you tell (, not to Piero)?

                                    (Rizzi, 1997, p. 291)

To summarize the previous part, Rizzi (1997) provided the syntactic hierarchy of

the split CP hypothesis as in Example (12).

(12) a. Credo che a Gianni, QUESTO, domani, gli dovremmo dire

         C   Top   Foc   Top   IP

    "I believe that to Gianni, THIS, tomorrow we should say"

   b. Credo che, domani, QUESTO, a Gianni, gli dovremmo dire

         C   Top   Foc   Top   IP

c. Credo che domani, a Gianni, QUESTO, gli dovremmo dire

      C   Top   TopFoc   IP

d. Credo che a Gianni, domani, QUESTO, gli dovremmo dire

      C   Top   TopFoc   IP

e. Credo che QUESTO, a Gianni, domani, gli dovremmo dire

      C   Foc   Top   Top   IP

f. Credo che QUESTO, domani, a Gianni, gli dovremmo dire

      C   Foc   Top   Top   IP   (Rizzi, 1997, p. 295-6)

In 2001, Rizzi updated and expanded the CP layer to include two more layers to the 1997 layers. He added the InterrogativeP and the WhP. So, Example 7 could be revised to the following, as shown in Example (13).

(13)   ForceP (TopicP) IntP (TopicP) FocusP (TopicP) WhP (TopicP) FinP IP

      VP.

Rizzi found out that the yes/no particle *se*, in an embedded clause in Italian, could be followed by a focus element and preceded by a topic, while the force element, such as *che* in Italian, could be followed by a focus and not preceded by anything. Thus, this case suggested having a new layer between the ForceP and FocusP that would be labelled as IntP. Example (14) shows the structure hierarchy of the embedded yes/no particle *se* in Italian.

(14) a. Mi domando **se** QUESTO gli volessero dire (non qualcos' altro).

      'I   wonder if THIS   they wanted to say to him, not something else.'

b. *Mi domando QUESTO **se** gli volessero dire (non qualcos' altro).

    'I   wonder   THIS   if they wanted to say to him, not something else.'

c. Non so **se**, a Gianni, avrebbero potuto dirgli la verit`a.

  'I don't know if to Gianni, they could have said the truth.'

d.   Non so, a Gianni, **se** avrebbero potuto dirgli la verit`a.

  'I don't know, to Gianni, if they could have said the truth.'

<div align="right">(Rizzi 2001:289)</div>

The same thing happened with WhP. There are some wh-questions in embedded clauses that could be preceded by a focus element and must not be followed by a focus element. This case suggested that there should also be a layer between the FocusP and FinP labelled as WhP, which is responsible to host the wh-operators in the embedded wh-questions. Example (15) presents the position of the embedded wh-questions in the split CP hypothesis

(15)   a. *? Mi domando **a chi** QUESTO abbiano detto (non qualcos'altro)

        'I wonder **to whom** THIS they have said (not somethin else)'

    b. *? Mi domando QUESTO **a chi** abbiano detto (non qualcos'altro)

        'I wonder THIS **to whom** they have said (not something else)

    c. Mi domando A GIANNI **che cosa** abbiano detto (non a Piero)

        'I wonder TO GIANNI **what** they have said (not to Piero)

    d. *? Mi domando **checosa** A GIANNI abbiano detto (non a Piero)

        'I wonder **what** TO GIANNI they have said (not to Piero)

<div align="right">(Rizzi, 2001, p.290)</div>

To sum up, in 2001 Rizzi revised his analysis of the left periphery structure/split CP, and he concluded that the CP layer should be divided into sub-layers. There should be a Force layer which is responsible to select the mood of the clause. The ForceP is followed by an IntP, which is loci of the yes/no particles. After the yes/no layer, there should be the FocusP, which is the home for the wh-questions in the main clause and the focalized elements. In contrast, the wh-questions in an embedded clause have their own layer, which is called WhP, and it could be preceded by a FocusP. The last layer in the split CP is the FinP, which is responsible for selecting the finiteness of its clause [+/-Fin]. Finally, between all of the previous layers we could have TopicP. From the new version of Rizzi's system, it could be noted that there are three layers for the [+Q/Wh], which are IntP, FocusP, and the embedded WhP.

At the end of this section, the Cartographic approach was developed about the same time as the minimalist approach by Chomsky (1995, 2001, 2008, 2013, 2015). Cartography could be defined as the approach which assigns each functional category to a particular position in the hierarchal structure. As stated earlier, Kayne (1984), Larson (1988), Pollock (1989), and Rizzi (1997, 2001) are the first authors who began the cartographic approach by splitting the traditional view of clausal structure (i.e., VP, TP/IP, and CP) into mini layers such as vP, VP and RootP for the lexical layer (VP), TP, AspP, and MoodP for the inflectional layer (IP/TP), and ForceP, FocusP, TopicP, and FinitenessP for the pragmatic/discourse layer (CP). The focus in this dissertation is on the pragmatic/discourse layer (CP), which is responsible for assigning or hosting the Q(uestion) feature and Wh- feature. Chomsky (1995) and Rizzi (1997, 2001) mentioned

that CP layer is the host projection of the interrogative particles and wh-question. Thus, in this dissertation, the phase theory by Chomsky and the split CP hypothesis by Rizzi are applied on the Wh-questions in Modern Standard Arabic.

Phase Theory

In the Minimalist Program, Chomsky (1995, 2000, 2001) introduced four syntactic operations that control the derivation as follows: (a) Lexical Array/Numeration, (b) Select, (c) Merge (IM and EM), and (d) Transfer/Spell-out. From these operations, phase theory was introduced as a solution to a theoretical issue that appeared from the External Merge over Internal Merge principle (i.e., from the Merge over Move (MOM) principle). The main idea behind the MOM principle is, when a derivation faces a decision to choose between the two syntactic operations Merge and Move, Merge always wins in this equation. For example, if we want to analyze the following sentence in Example (16) using the 1995 version of MP, we will face a problem with violating the Merge over Move principle

(16)   There are likely to be many parrots at the clay lick right now.

   a. N = {There, are, likely, to, be, many, parrots, at, the, clay, lick, right, now}

   b. Merge [V be] with the [DP many parrots …]

   c. Merge [T to] with the VP [be ….]

   d. Move [DP many parrots] to Spec of TP to check EPP

   e. Merge [Adj likely] to [TP many parrots to be …]

   f. Merge [V are] to [AdjP likely many parrots to be …]

g. Merge [T] to [VP are likely many parrots …] and move *are* to T

h. Merge the expletive *there* to the Spec of the matrix TP to check the EPP.

<div align="right">(Citko, 2014, pp. 24-27).</div>

Up to now, everything seems accurate; however, (16d) violates the Merge over Move principle because Chomsky (2000)said that, in the choice of selecting between the syntactic operation Merge and Move, it is preferred to choose Merge over Move. Thus, we could reanalyze Example (16) by considering the MOM principle, as shown in Example (17).

(17)   There are likely to be many parrots at the clay lick right now.

a. N = {There, are, likely, to, be, many, parrots, at, the, clay, lick, right, now}

b. Merge [V be] with the [DP many parrots …]

c. Merge [T to] with the VP [be ….]

d. Merge the expletive *there* to Spec of TP to check EPP

e. Merge [Adj likely] to [TP there to be many parrots…]

f. Merge [V are] to [AdjP likely there to be …]

g. Merge [T] to [VP are likely there to be …] and move *are* to T

h. Move the expletive *there* to the Spec of the matrix TP to check the EPP

<div align="right">(Citko, 2014, pp. 24-27)</div>

In Chomsky's 2001 work, he gave examples of the MOM dilemma when the derivation needs to choose between the Merge and Move, and he concluded that we need to redefine the Lexical Array/Numeration as follows:

Suppose we select LA as before … Suppose further that at each stage of the derivation of a subset LAi is extracted, placed in a[n] active memory (the 'work space'), and submitted to the procedure L. When LAi is exhausted, the computation may proceed if possible. Or it may return to LA and extract LAj, proceeding as before. (Chomsky, 2001, p. 106)

In other words, instead of initiating the whole sentence in one Lexical Array, we could divide the sentence in sub-arrays. Furthermore, this could solve the issue of preference of Merge over Move because the two arguments are in two different sub-arrays unless the two arguments are located in the same subarray; therefore, we need to select Merge over Move. These sub-arrays are referred to as phases. Therefore, the general concept of MP remains unchanged (i.e., the four syntactic operations which control the derivation) except the LA has been redefined to sub-arrays instead of one big LA (i.e., (a) Subarrays/Phases, (b) Select, (c) Merge, and (d) Spell-Out/Transfer).

In Chomsky (2001), the interfaces (i.e., the Conceptual-Intentional system (C-I or LF) and the Sensorimotor system (SM or PF)) have an impact in defining the phases. The following quotation explains the role of C-I and SM in the phase theory:

Ideally, phases should have a natural characterization in terms of IC: they should be semantically and phonologically coherent and independent. At SEM, vP and CP (but not TP) are propositional constructions: vP has full argument structure, and CP is the minimal construction that includes Tense and event structure and (at the matrix, at least) force. At PHON, these categories are relatively isolable (in clefts, VPmovement, etc.). These properties do not, however, yield exactly the right distinctions: vP with v nontransitive is relatively isolated and is a domain for QR, though these cannot be phases for Spell-Out. Call these weak phases. Then the strong phases are those that have anan EPP-position as an escape hatch for movement and are, therefore, the smallest constructions that qualify for Spell-Out. (Chomsky, 2004,p. 124)

From the previous definition of phases, Chomsky suggested that vP and CP are the phases (i.e., the transitive and unergative vPs), while the TP and VP are not phases.

21

He also categorized the phase head as the loci of the uninterpretable features. Thus, if we want to revise the four syntactic operations that Chomsky talked about, they should be: (a) Sub-arrays (phases), (b) Select, (c) Merge (i.e., IM and EM), and, finally, (d) Transfer/ Spell-Out. After defining the phases, Chomsky assumed that there should be a condition which governs the relations between the Transfer/Spell-Out and the phases (i.e., there should be a condition that governs the relation between the first with the final operation in the computation process). He assumed the Phase Impenetrability Condition (PIC), and there are two versions of PIC (i.e., strong PIC and weak PIC) as in Example (18):

> (18)   a.   In phase α with head H, the domain of H is not accessible to
>
>   operations outside α; only H and its edge are accessible to such
>
>   operations.   (Chomsky, 2000, p. 108).
>
>   b.   The domain of H is not accessible to operations at ZP; only H and its
>
>   edge are accessible to such operations.   (Chomsky, 2001, p. 14)

According to Example (18a), the phase head became closed immediately after the merge of the next head. For example, the complement of vP became inaccessible after the merge of TP head. However, in (18b) the complement of the phase head is still active and accessible until the merge of the next phase head. This means that the complement of vP is still accessible to TP until the merge of the CP head (i.e., C).

Up until now, this section has discussed the phasehood of the functional heads below the CP. An important question arises about the phasehood of the functional heads in the left periphery domain (i.e., CP domain). Chomsky (2008, p. 143) mentioned that C

is a phase head and "C is shorthand for the region that Rizzi (1997) calls the 'left periphery,' possibly involving feature spread from fewer functional heads (maybe only one)." This means that C could spread or donate its phasehood to all or only one of its daughters (i.e., Force, Topic, Focus, and Fin). In the following section, the phasehood in the Rizzi's system (i.e., split CP) is discussed.

On the Phasehood in CP

As noted earlier, Chomsky (2000, 2001) talked about the phase theory and suggested that C and v are the phase heads, while T and V are nonphase heads. Moreover, Chomsky (2008, p. 143) mentioned that C as the phase head is the shorthand for Rizzi's system. Following that, Totsuka (2015) claimed that, in the split CP, the Force and Topic heads are phase heads, while Focus and Fin are nonphase heads. Furthermore, I argue that only the highest Topic is a phase head, while other Topic heads in the same clause are nonphase heads. Example (19) combines the phase theory with Rizzi's system.

(19)

(Totsuka, 2015, p. 16)

23
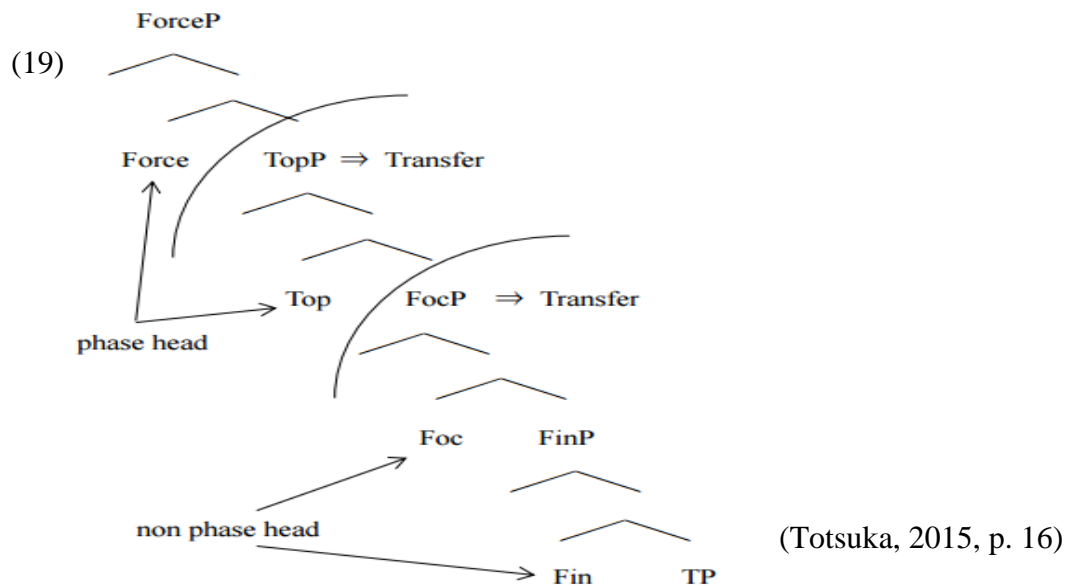
Furthermore, Ginsburg (2009) worked on the interrogative features cross linguistically, and he claimed that IntPs in Rizzi's system are phase heads and he called the IntP as TypeP. Thus, from Ginsburg's (2009) and Totsuka's (2015) analysis, it can be concluded that, out of the three interrogative layers in Rizzi's system (1997, 2001), that is, IntP, FocusP, and WhP, only the IntP are the phasal phrase, while the other two phrases are not phasal domains.

Following Totsuka's and Ginsburg's analysis, let us revisit Rizzi's system and study why there are grammatical clauses while others are ungrammatical. First I will look at the Topic-Focus asymmetry. I will replicate Example (10) in Example (20) to show how phases work in the split CP hypothesis.

(20) a. Il libro, a Gianni, domani, gli erodaròsenz'altro

  "The book, to John, tomorrow, I'll give it to him for sure"

   Topic Topic Topic TP …

 b. *A GIANNI IL LIBRO darò (non a Piero, l'articolo)

  "TO JOHN THE BOOK I'll give, not to Piero ,the article"

  Focus   Topic …

 c. A Gianni, QUESTO, domani, gli dovrete dire

  "To Gianni, THIS, tomorrow, you should tell him"

   Topic Focus Topic TP …

              (Rizzi, 1997, pp. 290-1).

Examples (20a and c), are grammatical because in (20a) the highest Topic is the phase head, while the middle and lowest Topics are not, and they could move from the

24

embedded clause to the final destination to check their feature before the phase head

merges (i.e., highest Topic head) and they would not violate PIC. However, in (20c) the

same scenario happens that the highest Topic head is the phase head, and the focalized

and topicalized element could move to their node and check their feature before the phase

head has merged, and they would not also violate PIC. However, in (20b) the sentence is

ungrammatical because there is only one Topic head, which is a phase head, and it is

located below the FocusP. The topicalized element in (20b) moves to the phase head

Topic and the whole sentence will be inaccessible and will Spell-Out before the focalized

element moves to its head, and this would crash the C-I system and violate the PIC.

Second, Rizzi (1997) mentioned that the complementizer *che* is in the Force head

while the *di* is in the Fin head. In Example (21), Example (9) is replicated to show and

study the phasehood of the full set of Rizzi's system (i.e. ForceP, TopicP, FocusP, and

FinP).

(21) a. Credo   **che** il tuo libro, loro lo apprezzerebbero molto

        "I believe that your book, they   would appreciate it a lot"

    b. *Credo, il tuo libro,   **che** loro lo apprezzerebbero molto

        "I believe, your book, that they would appreciate it a lot

    c. *Credo   **di** il tuo libro,   apprezzarlo molto

        "I believe 'of' your book to appreciate it a lot"

    d. Credo,   il tuo libro, **di**   apprezzarlo molto

        "I believe, your book, 'of' to appreciate it a lot"

                            (Rizzi, 1997, p. 288).

Examples (21a and d), are grammatical because in (21a) the topicalized element would move to its projection to check its feature; then, it would Spell-Out its complement. The complementizer *che* will merge to its projection; after that, it will Transfer/Spell-Out the whole sentence and this scenario would not violate the PIC. In (21d), the complementizer *di* merges in the Fin head, and it is a non-phase head which means that the topicalized element could raise/move to the spec of Topic to check its feature; then it will, as a phase projection, Spell-Out the whole embedded clause. On the other hand, Examples (21d and c) are ungrammatical because, in both cases, the phasal domain follows the nonphase domain, which means that the whole sentence Spell-Out/Transfer to C-I and PF interferes before the topicalized element as in Example (21c), or the complementizer *di* as in Example (21d) could move to its layer, and these cases would violate the PIC.

Finally, Rizzi (2001) revisited his analysis and he claimed that the left periphery cross linguistically should have this hierarchy as in Example (13). I represent it here in Example (22).

(22)   ForceP (TopicP) IntP (TopicP) FocusP (TopicP) WhP (TopicP) FinP IP

VP.

I studied the phasehood of IntP as in Ginsburg's (2009) analysis, and left other functional projections for further research. Ginsburg (2009) claimed that IntP is a phasal domain and that can be seen, as in the following Example (23).

(23) a. Mi domando **se** QUESTO glivolessero dire (non qualcos' altro).

'I wonder   if   THIS they wanted to say to him, not something else.'

b. *Mi domando QUESTO **se** glivolessero dire (non qualcos' altro).

   'I wonder   THIS   if they wanted to say to him, not something else.'

c. Non so **se**, a Gianni, avrebbero potuto dirgli la verit`a.

   'I don't know if to Gianni, they could have said the truth.'

d. Non so, a Gianni, **se** avrebbero potuto dirgli la verit`a.

   'I don't know, to Gianni, if they could have said the truth'

   (Rizzi, 2001, p. 289)

Examples (23a, c and d) are grammatical sentences because, in all of these examples, the nonphase phrase does not precede any of these phasal domains as IntP*se* in (23a and c), and TopicP in (23d), while (23b) is ungrammatical because it violates the PIC. In (23b) the focalized element *Questo* precedes the IntP*se*, which means that, during the derivation, *se* merges in its phase head to check the [+Q] feature, and, as a phase head it Spell-Out its complement at the moment of valuing its uninterpretable feature(s); then, its complement would not be accessible to any other higher functional projection, so the focalized element could not move out of IntP domain and the [Focus] feature will be unvalued, and this would crash the computational process.

Before moving to Arabic, I would like to analyze the left periphery structure in English based on the unified approach (i.e., the phase theory in the split CP hypothesis). van Gelderen (2004) mentioned that topicalized and focalized elements could occur in the same clause; however, there is a restriction in their word order. She said that the wh-word could appear after the topicalized element, but not before it, as in Example (24).

27

(24)   a. *What to Leonard should we say on his birthday?

b. To Leonard what should we say on his birthday?

The restriction in the clause hierarchy could be explained by the phase theory. Therefore,

Example (24a) is ungrammatical because the WhP is followed by a TopicP *to Leonard*,

which is a phase head, and this phase head would Spell-Out the whole clause before the

WhP could move to its head to check its feature; thus, PIC would be violated and this

would lead to a crash in the C-I system. Examples (24a and b) show the Spell-Out

operation for both (24a and b).

(25)   a.

b.

To sum up, Chomsky (2000, 2001) mentioned that C and v are phase heads while T and V are not. Chomsky (2008) also mentioned that C is the short hand of Rizzi's system. Following this claim, Ginsburg (2009) and Totsuka (2015) assumed that Force, Topic, and Int are phase heads, while Focus and Fin are nonphase heads. Some examples have been provided to present the phasehood in the split CP. The nonphasal domain must not precede a phasal-domain because it violates the PIC, as in Chomsky (2000); however, the phase head could precede a nonphase and phase head.

In this dissertation, Totsuka's (2015) and Ginsburg's (2009) assumption on the phasehood in the left periphery domain have been adopted. I claim that wh-questions are divided into two projections: (a) nominal wh-questions are on the Int projection (i.e., phase head), and (b) adverbial wh-questions are on the focus projection (i.e., nonphase

head). The main concertation in this dissertation is the two functional projections,

FocusP (nonphaseP) and IntP (phaseP), while I left the ForceP and FinP because they are

out of the scope of this study. In the following section, earlier studies and works on the

wh-questions are discussed.

Clause Structure in MSA

In this section, the ways in which verb movement, EPP, and phi-features impact

feature-checking mechanisms and derivations in MSA are discussed. Furthermore, it is

argued that the deep structure word order of interrogative clauses in Modern Standard

Arabic is subject-verb-object (SVO), while the surface structure has a verb-subject-object

(VSO) word order that is a result of strong features from which the VSO word order in

the surface structure is derived.

Scholars such as Doner (2013), Aoun, Benmamoun, and Choueiri (2010),

Rouveret (2010), Al-Horais (2009), Soltan (2007), McCloskey (2001), Carnie, Harley,

and Pyatt (2000), Ouhalla (1994, 1996), AL-Shorafat (1998), Aoun, Benmamoun, and

Sportiche (1994), and Mohammed (1989) have investigated verb movement in a variety

of languages having VSO word order. Their studies examined feature-checking systems,

such as EPP, phi-features, T, and case, in order to determine the surface word order for

each of the languages considered. Mohammed (1989), Ouhalla (1994), Benmamoun, and

Sportiche (1994), Soltan (2007), and Aoun, Benmamoun, and Choueiri (2010) claimed

that MSA has both a VSO word order and an SVO word order. This indicates some

differences in the feature-checking system or the presence of a strong/weak feature which

is responsible for the two word orders in MSA. Additionally, Alsaeedi (2015) asserted

that MSA exhibits both VSO and SVO word orders, with the basic word order being

SVO, while the VSO word order is the grammaticalized form.

Four different linguistic approaches have been used to analyze the structure of

declarative clauses in MSA and its dialects. The first approach relates to verb movement.

Benmamoun and Sportiche (1994), and Carnie et al. (2000) agree that there is a strong

feature higher than the tense phrase (TP) that requires the verb to move from the head of

the verb phrase (VP) to the head of the TP to get the tense feature (iT). In their analyses,

the verb then needs to move to a higher head to check this strong feature (i.e., V-to-T-to-

X/F/C). Also in their analyses, there is an EPP feature which requires an argument to

appear in the spec of the TP. In Benmamoun (1994), the higher head is identified as a

head of the focus phrase (FocusP), which is located between the complementizer phrase

(CP) and the TP and aims to fulfill the semantic differences between VSO and SVO word

order. Carnie et al. (2000) examin Irish and conclude that, in order to derive the VSO

word order in Irish, there must be a strong feature (+Finiteness) which is located in the

head of the CP. This causes the verb to move from V-to-T-to-C to check the finiteness

feature in the head of the CP. Consequently, the finiteness feature in the head of the CP

is reason for VSO word order in Irish.

The second approach used in analyzing the declarative clause structure argues for

the presence of a weak EPP feature. This approach is supported by Mohammed (1989)

and FassiFehri (1989) for MSA and McCloskey (2001) for Irish. Their research indicated

that neither Arabic nor Irish requires the subject to move from the spec of the VP to the

spec of the TP because the EPP is a weak feature. The weak EPP does not motivate the

subject to move; thus, the weakness of the EPP causes the VSO word order while the subject in the SVO word order in a topic, and the attached suffix on the verb is a resumptive pronoun, not an agreement. Thus, the VSO word order is a result of weak EPP in MSA, and the SVO is a result of the [+topic] feature which triggers the subject to move higher than the verb and left a resumptive pronoun in its foot.

The third approach used in analyzing the declarative clause structure argues for the split of the TP in MSA. Ouhalla (1994) introduced the idea that the TP splits into Agreement Phrases (i.e., AgrPs and AgrPo). He claimed that, in VSO word order, the AgrPs is lower than the TP and derives the VSO word order. However, in SVO languages such as English, the AgrPs is higher than the TP with SVO word order as the result.

The fourth approach used in analyzing the declarative clause structure is the assertion that there are two syntactic structures in MSA. Soltan (2007) discussed the existence of two different syntactic structures with two different feature-checking systems. According to Soltan (2007), since the SVO word order is a subject-complement structure, then the subject is base-generated in the spec of the TP and satisfies the EPP feature while the verb moves from VP-to-TP in order to check the tense feature [iT] and to be in spec-head agreement with the subject in order to get the full set of phi-features (i.e., full agreement between the subject and the verb in person, number, and gender). Furthermore, Soltan (2007) stated that VSO word order has a different feature-checking system than the SVO structure. He claimed that in the VSO structure there is no EPP, so there is no requirement for the subject to move to the spec of the TP. Soltan (2007)

designed his analysis based on Mohammed's (1989) analysis, and Ouhalla's (1994)

analysis.

Examples (26), (27), and (28) illustrate the different syntactic structures as

presented by Soltan (2007) to show the two feature-checking systems.

(26) a. alnisa   akl-na   altufah   b. akl-t   alnisa   altufah

    the-women ate-3PlF the-apples   ate-defaultF the-women the-apple
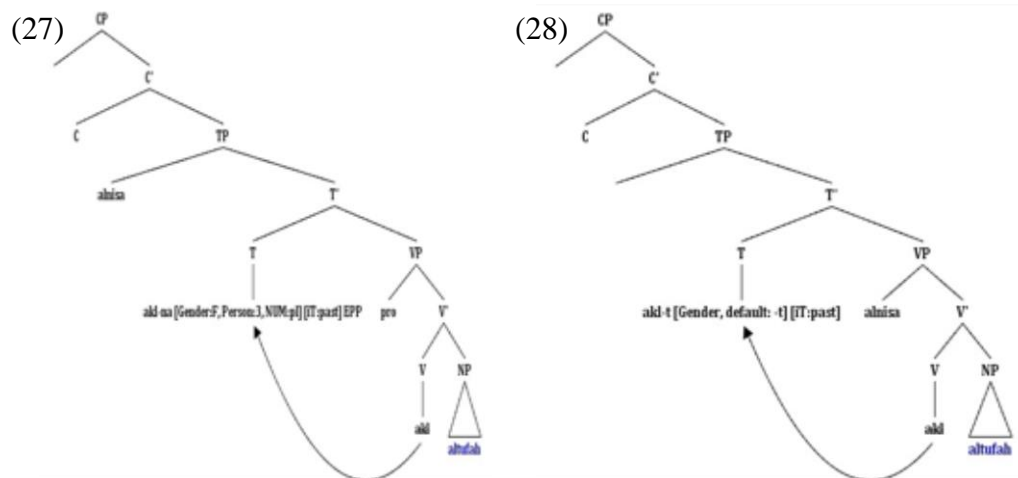
The feature-checking systems in the two syntactic structures are different from

one another, which results in MSA having both an SVO and a VSO word order. In SVO

word order, there is a strong EPP which requires an argument to move to the spec of the

TP to fulfill this feature. Also, the phi-features (i.e., gender, number, and person) are

satisfied through the spec-head relationship. However, in VSO word order, in Soltan's

(2007) analysis, there is no EPP, so there is no need for an argument to show in the spec

of the TP, and the phi-features are modified to have only gender and a default feature which he defined as third person singular.

From the approaches presented in this section, I chose the second approach to apply to my analysis of the syntax of yes/no questions in MSA: (a) Mohammed's (1989) analysis of feature checking in MSA and (b) McCloskey's (2001) analysis of Irish showing that the EPP feature is weak in VSO word order with no motivation for the subject to move from the spec of the VP to the spec of the TP. Examples (29a) and (29b) illustrate verb movement in VSO and SVO as presented in Mohammed's and McCloskey's analyses:

(29) a. In VSO word order, the V moves from the VP to the TP, and there
is a weak EPP.

b. In SVO word order, the V moves from the VP to the TP, and there
is a further movement for the S to move to the spec of the TP
to check the strong EPP.

The feature-checking system in Mohammed's (1989) analysis has interpretable tense (iT), which requires the verb element to move from the head of the VP to the head of the TP to give value to the uninterpretable tense (uT) feature on the verb element. The EPP feature is weak in VSO word order; thus, the subject remains in the spec of the VP. However, in SVO word order, the EPP is strong; therefore, the subject must move to the spec of the TP to check the EPP feature.

Additionally, Mohammed (1989) discussed phi-features and the agreement system in MSA. He asserted that in SVO word order in MSA, the subject and the verb

34

are in the spec-head relation; therefore, there is full agreement on the verb which means

that verbs in SVO word order show the full set of phi-features–person, gender, and

number. However, in MSA, the VSO word order has partial agreement on the verb

element because the verb is in the head of the TP while the subject is in a lower position

(i.e., the head of the VP), and they are not in spec-head relation. In this case, the default

phi-features–third person, singular, and gender–appear on the verb. Examples (30)

through (33) show the feature-checking system as outlined by Mohammed (1989) in VSO

and SVO word order in MSA.

(30)akl-Ø   altulab   altufah   {VSO}

      ate-3SM   the-students   the-apples

(31)
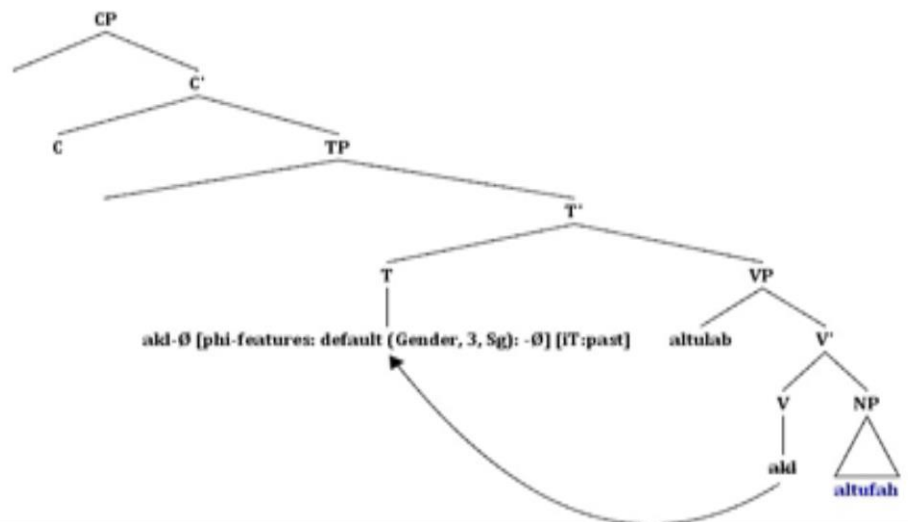
(32) altulab   akl-u   altufah   {SVO}

      the-students   ate-3PlM   the-apples

35

(33)
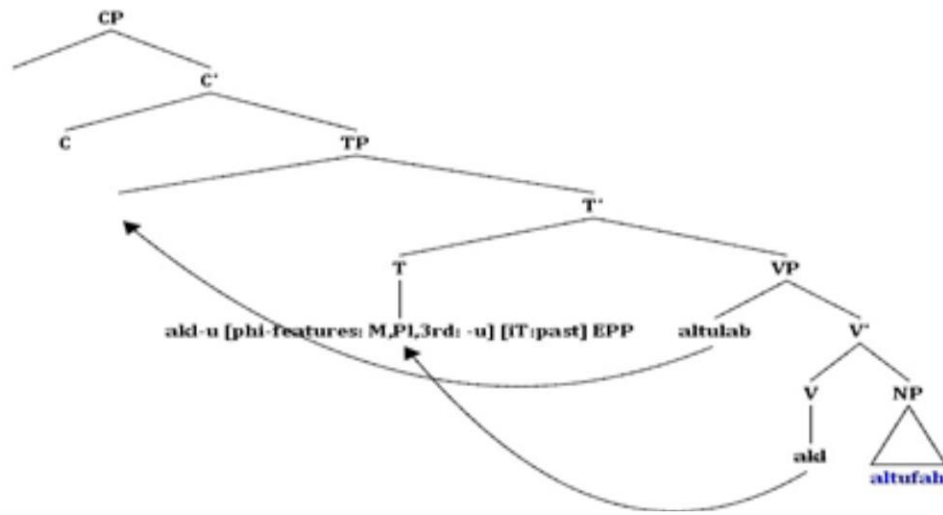
I chose the weak feature approaches by Mohammed (1989) and McCloskey (2001) because these analyses are clear and do not require many layers or movements to satisfy the feature-checking system and derive the surface word order. In contrast, Benmamoun's (1994) and Ouhalla's (1994) analyses of the feature-checking system in MSA have unnecessary movements and too many layers. Furthermore, the analyses by Benmamoun (1994) and Ouhalla (1994) present a challenge in explaining the partial agreement and the VP-to-TP-to-Foc/Agr movement. Soltan (2007) presented a more promising analysis in that it has the advantage of distinguishing the meaning of the two word orders by the two syntactic structures; however, explaining the dual agreement is problematic. Also, Soltan (2007) introduced two mechanisms for the phi-features which is unnecessary because he divides the phi-features into two sets, one containing person and number, and the other containing gender.

Thus, I chose Mohammed (1989), FassiFehri (1989) and McCloskey (2001) who claimed that the EPP feature in VSO languages is weak and the SVO word order is a

result of the [+topic] feature. Accordingly, there is no need for the argument to show in the spec of the TP and the raising of the verb element to the head of the TP to check the phi-features, and tense is what derives the surface VSO word order. Furthermore, Alsaeedi (2015), asserted that the basic word order in MSA is VSO due to the weakness of the EPP feature, as well as that the EPP feature has been grammaticalized to a strong feature, which derives the SVO word order in late MSA or visa versa. Consequently, according to these authors, the existence of both VSO and SVO word orders in MSA is the result of the grammaticalization of the EPP feature. In the following section, the clausal hierarchy with a focus on the CP layer, which is the main issue of this dissertation, is discussed.

Earlier Studies on wh-questions in MSA.

Historically, there were two main Arabic grammar schools that analyzed the Arabic clause structure and they are: (a) Basra school, which is represented by Sibawayh in his grammatical work *alkitab* "the book," and (b) Kufa school, which is represented by Kisaie in his Arabic grammar book *muktasar fi alnahu* "Introduction of Syntax." Furthermore, there were other traditional Arab grammarians who did not belong to these schools but had an impact on the development of the Arabic syntax, such as ibn Hishamin in his book *Qataralnada*. These Arabic scholars and others talked about the question of the structure in MSA, and they divided the information structure into two types: (a) yes/no questions such as *hal* and *ʔa-* (they called them *hurufalistifham* "question particles/letters/pronouns"), which are out of the scope of this study, and (b) wh-questions such as man "who", matta "when", kam "how," and so forth (they called them

a *asmaalistifham* " question nouns"). They all agreed that yes/no and wh-questions must be followed by a verb except the interrogative particle *ʔa-*. It could be followed by a verb or a noun. The following examples in (34) provided the structure of questions in MSA based on the traditional schools.

(34)   a. hal ja'-a   Ali-un.

           Q   came-3SM Ali-NOM

          'Did Ali come?'

    b. *ʔa*-ja'-a   Ali-un.

        Q-came-3SM   Ali-NOM

        'Did Ali come?'

    c. *hal   Ali-un   ja'-a.

           Q   Ali-NOM came-3SM

    d. ayna   thahab-a   Ali-un.

        Where went-3SM   Ali-NOM

         'Where did Ali go?'

    e. *ayna   Ali-un   thahab-a.

         Where Ali-NOM went-3SM

    f. man qabal-a   Ali-un.

        Who met-3SM Ali-NOM

         'Who did Ali meet? '

g. *man Ali-an qabal-a

Who Ali-ACC said-3SM   (Fargal, 1986, pp. 27-28)

They analyzed the previous examples and concluded that the ill-formedness of
(34c, e, and g) is because of having a noun phrase following the interrogative
word/particle. They did not try to correct them or to present the grammatical version of
these examples. Furthermore, the Basra school analyzed the question word as in (34f) as
a topic based-generated in the matrix projection while the rest of the structure is the
comment of the question word. In contrast, the Kufa's school analyzed the wh-question
based on its deep structure, so in (34f) they would analyzed "Ali" as the topic and "qabal-
a who" as the comment of the topic. Thus, Kufa's syntacticians would be the first
scholars (or maybe one of the first scholars) who captured the concept of the wh-
movement, while the Basra scholars supported the non-movement approach. To sum up,
both schools were descriptive. They did not look deeply into the structure of the
interrogative in MSA.

Following the traditional Arabic grammarians, the modern linguists of Arabic
such as FassiFehri (1982), Wahba (1984), Demirdache (1991), Aoun and Benmamoun
(1998), Shlonsky (2000, 2002), Aoun and Choueiri (1996, 1999, 2000), Aoun and Li
(2003), Sells (1984), and Aoun, Choueiri, and Hornstein (2001), Aoun, Benmamoun, and
Choueiri [henceforth ABC] (2010), among others, studied the wh-questions as part of the
CP domain (i.e., they have not carefully looked at or analyzed the deep structure of the
wh-question in MSA), and they concluded that the wh-questions in MSA could be
divided into two types: (a) nominal wh-questions such as *man* "who," *ʔayy-u/a* "which,"

39

ma "what," and *matha*"what," and (b) adverbial wh-questions such as *ʔayn*"where,"

mataa "when," *kam*" how many/much," *kifa* "how," and *limatha* "why." Furthermore, as

for those scholars, both types should move from their lower position in the tree to their

final destination (i.e., to FocusP). Example (35) shows the nominal and adverbial wh-

structure in MSA.

(35)   a. man/ʔayya   mariiDin zaarat   Naadia __?

            who/which patient   visited.3FS Nadia

            'Who/which patient did Nadia visit?'

        b. man/ʔayya   mariiDin zaarat-**hu**   Naadia?

            Who/which patient   visited.3FS-3MS Nadia

             'Who/which patient did Nadia visit?'

        c. maadaa/ʔayya kitaabin ʔistarat   Laila fi-l-maktabati?

            what/which   book   bought.3FS Laila in-the-bookstore

            'What/which book did Laila buy at the bookstore?'

        d. ʔayna dahabtum   baʕda   l-ʁadaaʔi?

             where went.2Pl   after   the-lunch

             'Where did you go after lunch?'

        e. mataa haDartum l-masrahiyyata?

            when   saw.2Pl   the-play

            'When did you see the play?'

        f. kayfa/limaadaa mazzaqta   l-kitaaba?

            how/why   tore.2MS   the-book

'How/why did you tear the book?'

g. kayfa   naSilu   ?ilaa l-mat¯hafi?

   how   arrive.1Pl to   the-museum

   'How do we get to the museum?'

h. kam   kitaabin qara?a   t-talaamiidu?

   how many book   read.3MS the-students

   'How many books did the students read?'

i. kam   dafaʕtum li-taSliihi s-sayyaarati

   how much   paid.2Pl   to-fixing   the-car

   How much did you pay to fix the car?'   (ABC, 2010, p. 132)

From Example (35), we can see that the nominal wh-words could be related to a gap, as in (35a), or to a resumptive pronoun, as in (35b), while the adverbial wh-words must be related to only a gap. Thus, the main difference between these two types, as the syntacticians of Arabic mentioned, is that the adverbial wh-questions are related to gap strategy and they are derived by movement approach. On the other hand, the nominal wh-questions have the option to relate to a gap or resumptive pronoun in some cases, but in other cases, such as wh- with relative operator or wh- followed by a focalized element nominal wh-question, must be related to the resumptive strategy. In the case of resumptive pronouns, the nominal wh-questions are derived by the based-generated approach.

Furthermore, those scholars are following Chomsky (1995, p. 291) and Radford (1997, p. 108) who claimed that the interrogative structure has the [+Q/wh] feature,

which triggers the wh-words to move or base-generated in the CP head. Also, they are

adopting Rizzi's system (i.e., split CP), and they claimed the [+Q/wh] feature in MSA is

in the FocusP. Thus, scholars of Arabic syntax claimed that the wh-questions must either

move or be base-generated in FocusP in order to check the [+Q/wh]. In the following

section, the topicalization and focalization in MSA will be discussed because they are one

of the major roles in determining the accurate projection for the wh-questions in MSA.

Left Periphery Structure in MSA

As part of studying the wh-question in MSA,provided in this dissertation was an

overview of the structure of TopicP and FocusP because they interact with wh-structure,

and they would strengthen my claim that nominal wh-questions are in a phasal phrase

(i.e., IntP), while adverbial wh-questions are in a non-phasal phrase (i.e., FocusP).

Bakir (1980), Ayoub (1981), Demirdache (1988), FassiFehri (1993), Khalaili

(1994), Ouhalla (1994b), Shonlsky (1994a and b, 1996, 1997, 2000), and Aoun,

Benmamoun, and Choureiri (2010), among others, studied the structure of the left

periphery in MSA and they concluded that the topicalized elements must be related to

resumptive strategy, be [+definite], are insensitive to islands (i.e., they are following non-

movement approach), and they show the [+nominative] case marker except in *anna*

clauses. In *anna* clauses, the topicalized element, which follows the *anna*

complementizer must show the [+accusative] because of the [+finiteness] feature on the

anna. On the other hand, most of these researchers mentioned that focalized elements in

MSA are related to gap strategy, could be indefinite or definite, are sensitive to islands

(i.e., they are following movement approach), and they retain their original case marker.

42

Ayoub (1981) mentioned that the definite focalized elements may be related to resumptive strategy. The following examples in (36) present the topicalization in Modern Standard Arabic.

(36)   a. at-tilmiidat-u   raʔaa-ha   Saami l-baariha

the-student.FS-Nom saw.3MS-3SF Sami the-yesterday

'The student, Sami saw her yesterday.'

b. naadia,mataa   raʔaa-ha   Saami?

Nadia when   saw.3MS-3FS Sami

Nadia, when did Sami see her?'

c. *mataa Naadia   raʔaa-ha   Saami?

when Nadia   saw.3MS-3FS Sami

'Nadia, when did Sami see her?'

d. zaʕamtu   ʔanna   r-risaalat-a   al-walad-u   kataba-ha

claimed.1S that   the-letter-Acc the-boy-Nom   wrote.3MS-it

I claimed that the letter, the boy wrote it.'

e. zaʕamtu   ʔanna   al-walad-a   r-risaalat-u   kataba-ha

claimed.1S that   the-boy-Acc the-letter-Nom wrote.3MS-it

I claimed that the boy, the letter, he wrote it.'

f.   al-qaSiidat-u   ʔallafa-ha   ʕomar

the-poem-Nom   wrote.3MS-it Omar

The poem, Omar wrote it.'

g. *al-qaSiidat-u   ʔallafa-Ø   ʕomar

43

the-poem-Nom wrote.3MS   Omar

The poem, Omar wrote it.'

h. *qaSiidat-un   ʔallafa-ha   ʕomar

poem-NOM   wrote.3MS-it   Omar

A poem, Omar wrote it.'

i. haada l-kitaab-u   qaraʔtu-hu   mundu   muddat-in

this   the-book-NOM   read.1S-it   from   while-Gen

This book, I read it a while ago.'

j. huwa   raʔaytu-hu   fi-l-madiinat-i

He,   saw.1S-3MS   in-the-city-Gen

Him, I saw him in the city.'   (ABC, 2010, p. 191-195)

k. Zayd-un   raʔy-ta   al-rajula alathi   darab-hu

Zayd-NOM you-saw the-man   who   hit-3MS

"Zayd, you saw the man who hit him"   (Ayoub, 1981)

From the previous examples, it can be seen that the topicalized elements must be [+definite] (i.e., names, strong pronouns, or determiner/demonstrative phrase), and related to resumptive pronouns. Thus, Examples (36a, d, e, and f) are grammatical because the topicalized element is [+definite] and there is a resumptive; however, (36g and h) are ungrammatical because there is no resumptive pronoun in the first and [-definite] on the second. Additionally, (36b and c) provides evidence for Totsuka's analysis that the Topic is a phase head and Focus is a nonphase head. In (36b), the TopicP is sitting at the matrix projection and the FocusP is following it; therefore, there is

no PIC violation, while in (36c) the vice versa violates the PIC (note: the wh-word in these examples is an adverbial wh-question). In most of the TopicP, the topicalized element shows the [+nominative] case marker; however, in *anna* clauses the topicalized element, which is following the complementizer *anna* must show the [+accusative] case marker because the complementizer is a case assigner and it has the [+finiteness] feature. Finally, (36k) shows that the topicaltization structure in MSA is insensitive to islands.

After studying the topicalization structure in MSA, I would like to look at the focalization in MSA, which provides evidence on the position of the wh-question as well as the topicalization does. As stated earlier, the focus structure in MSA saves its original case marker, is sensitive to islands, and it could be a [+/-definite]. Most of the scholars, such as Bakir (1980), Moutaouakil (1989), Ouhalla (1994b), and Shlonsky (2000), who worked on the left periphery structure mentioned that focalization in MSA is related to a gap strategy; however, Ayoub (1981) mentioned in his dissertation that the [+definite] focalized element has the optionality to relate to a gap or to a resumptive pronoun. The following Example (37) would clear the picture on the focalization structure in MSA.

(37)   a. shay-an   shariba   Zayd-un

             tea-ACC.indef   drank.3ms Zayd-NOM

           'It was tea that Zayd drank.'(ABC, 2010, p. 202)

        b. kitaab-an   wajad-a   Mohammad-un

           book-ACC.indef   found-3MS   Mohammad-NOM

           "A book Mohammad found."

45

c. al-kitaab-a   wajad-a   Mohammad-un

  the-book-ACC found-3MS   Mohammad-NOM

  "The book Mohammad found."   (Bakir, 1980)

d. al-kitaab-a   wajad-a-hu   Mohammad-un

  the-book-ACC found-3MS-3S Mohammad-NOM

  "The book Mohammad found it."   (Ayoub, 1981)

e. *Zayd-an   raʔy-ta al-rajula   alathi darab-(hu)

  Zayd-ACC you-saw the-man   who   hit-(3MS)

  "Zayd you saw the man who hit (him)"   (Ayoub, 1981)

f. *ʔayna   Saalim-an   qaabala   Khalid-un

    Where   Salim-ACC met.3ms   Khalid-Nom

g. *Saalim-an   ʔayna   qaabala Khalid-un

    Salim-ACC   where   met.3ms Khalid-Nom

h. FaaTimat-u   l-wardat-a   ʔaʕTaa-ha   Saalim-un

  Fatima-Nom   the-rose-ACC gave.3ms-her Salim-Nom

  'Fatima, the rose Salim gave her.'

i. *l-wardat-a   FaaTimat-u   ʔaʕTaa-ha   Saalim-un

    the-rose-ACC Fatima-Nom   gave.3ms-her Salim-Nom

    'Fatima, the rose Salim gave her.'   (ABC, 2010, p. 204)

j. zaʕamtu ʔanna r-risaalat-a   kataba-**ha**   l-walad-u

  claimed.1s that   the-letter-ACC wrote.3ms-it the-boy-Nom

  'I claimed that the letter, the boy wrote it.'

k. *zaʕamtu ʔanna r-risaalat-a kataba l-walad-u

claimed.1s that the-letter-ACC wrote.3ms the-boy-Nom

'I claimed that the letter, the boy wrote.'

(Shonlsky, 2000, p. 336)

In Examples (37a-c), the focalized elements, as scholars of Arabic syntax stated, could be [-definite] as in (29a and b), or [+definite] as in (37c). Ayoub (1981) was the first researcher who noticed the relation between the [+definite] focalized element and the resumptive pronoun, as shown in (37d). Example (37e) shows that FocusP is sensitive to islands, and that makes it is ungrammatical. Examples (37f and g) present that wh-questions cannot be preceded by or follow a FocusP (note: these examples only use the adverbial wh-question). Examples (37h and i) provide evidence for Totsuka's (2015) anaylsis that FocusP, as a nonphasal phrase, could not precede a TopicP as a phasal phrase or it violates PIC. Finally, the focalized element in the *anna* clause, as in Examples (37j and k) must be related to the resumptive strategy because, as Shonlsky (2000) assumed, there should be a [+definite] feature that has to be checked by the resumptive pronoun.

To conclude this section, researchers have distinguished between the focalization and topicalization by a set of characteristics as follow:

1. A Focalized element has the option to relate to a resumptive or to a gap based on [+/-definite], while topicalization must always be related to a resumptive pronoun.

47

2. FocusP is sensitive to islands, so focus DPs can not cross/move over CP; however, TopicP does not show any sensitivity to islands.

3. There is no limit on the occurrence of the TopicP, while there must be only one FocusP in a sentence.

4. A focalized element retains its original case marker, while the topicalized lexical must show the [+nominative] case marker.

5. Finally, FocusP cannot precede a TopicP, otherwise it violates PIC. On the other hand, TopicP can precede or follow a FocusP.

An important component of wh-structure in MSA is resumptive pronouns. In the following section, an analysis of resumptive pronouns will be presented.

Resumptive Pronouns

Resumption is a saving device of an extraction process in which an extracted element does not allow a gap in its foot, but a resumptive pronoun. Resumptive pronouns can be found in variable positions, as in focalization, topicalization, relative, and wh-constructions. For example, English uses resumptive strategy as a saving device to rescue the derivation from violating general constraints, as in the following Example (38).

(38) a. I just saw **a girl** who Long John's claim that **she** was a Venusian made all the headlines.

b. The only kind of **car** which I can never seem to get **its** carburetor adjusted right is them Stanley Steamers.

(Ross, 1967, p. 432-433)

On the other hand, English does not allow a resumptive pronoun to appear in the original position of the extracted element, but a gap when there is no need for it and there is no constraining violation, as shown in Example (39).

(39)   a. *I saw the boy that Mary loves **him.**

        b. I saw the boy that Mary loves ___.

<div align="right">(Rouveret, 2011, p. 2)</div>

Ross (1967) and Sells (1984) discussed resumption and found out that resumptive pronouns appear in a antecedent's trace in order to avoid a locality constraining violation, or Empty Category Principle (ECP) violation, or in a language which does not allow preposition stranding or extraction out of possessive construction. Thus, there are several factors which could make resumptive pronouns obligatory, optional, or not allowed. Furthermore, there are semantic factors which play an important role in the appearance of resumptive pronouns as the [+definiteness] and specificity effect/reading/ interpretation.

In this dissertation, the theory of resumption as not the main focus; however, resumptive construction has some interaction with wh-trace, and, to be more specific, there is a relation between wh-trace (i.e., gap or resumptive pronoun) in MSA on the one hand, and, on the other hand, avoiding a locality constraining violation and showing [+definiteness].

Sharvit (1999) defined resumptive pronouns as follows:

> Resumptive pronouns have a dual nature. In some ways they are like traces, in others like 'regular' pronouns. Like wh-traces, they need to be bound from an A'-position. Unlike traces, they are not subject to Bounding constraints. In addition, like traces (and A-bound pronouns), resumptive pronouns are interpreted as bound variables (bound by the wh-phrase in wh-questions and by

the relative operator in relative clauses). But in some sense they have less freedom of interpretation, compared to traces (p. 591).

This dual stats ouf resmptives urged researchers such as Ross (1967), Tallerman (1983), Borer (1984), Sells (1984), McCloskey (1990), Demirdache (1991), Shonlsy (1992), Aoun, Choueiri, and Hornstein (2001), Aoun and li (2003), and Boeckx (2003) to assume three approaches in analyzing resumptive construction.

The first approach is the nonmovement approach. Borer (1984), Sells (1984), and McClosky (1990) worked on the resumption in relative, left dislocation, focalization, and wh-constructions, and claimed that resumptive pronouns are like regular pronouns. Their claim is based on three categories of resumptive pronouns, as they do not show any sensitivity to island, they do not give rise to Weak Crossover effects, and they are governed by Highest Subject Restriction. In turn, Example (40) illustrates these three factors on Irish.

(40)   a. na danta   sin   nach bhfuil fhios   againn cen ait ar cumadh **iad**

       the poems these C-neg is   knowledge at-us   what place C were-

       composed **them**.

     "Those poems that we do not know where they are composed"

     b. an fear so   ar mharibh a bhean fein   **e**.

      the man this C killed   his-own-wife him.

      "this man that his own wife killed"

     c. * an fear   a raibh **se** broeite

       The man C was   he ill

      "The man that was ill"   (McClosky, 1990, p.76-77)

Thus, those researchers suggested that the resumptives merge the derivation with verb projection, while the dislocated phrase, relative marker, or the wh-phrase are directly merged to the left periphery position (i.e., nonmovement approach/base-generated approach). Furthermore, the relation between the resumptive pronoun and its antecedent is established via binding relation at the S-structure or LF. On the other hand, they claimed that, in gap cases, wh-phrases, dislocated phrases, and relative markers are following the movement approach and the gap would be the representation of antecedent-trace construction. For Borer, Sells, and McClosky, gaps are the true trace, which means that there is a movement, while resumptives are regular pronouns, which means that they are following the base-generated approach.

The second approach in the analysis of resumption is the last resort approach. From the theory of resumption, we can see that gap strategy is used when there is no island constraining violation; however, on the other side of the coin, resumptive strategy is used when there is an island/locality constraining violation in order to save the derivation. Thus, Kroch (1982), Shlonsky (1992), and Aoun (2000) proposed that "resumptive pronoun[s] are never freely generated" which means that resumptive pronouns appear in a position when a gap cannot, so they see resumptive as last resort and following the movement approach. More recently, van Urk (2016) adopted the last resort strategy in the framework of phase theory in Dinka. He discovered that left dislocated phrases use resumptive strategy if they are separated from their trace by PP, DP, or CP, but not by vP (i.e., if the antecedent is separated from its trace by phasal phrase (but not vP), the resumptive pronoun must appear to save the derivation). The

main issue with the last resort approach remains that it cannot predict or expect to have

optionality/freedom between resumptive pronouns and gap as in Hebrew and Irish, but it

is not the case in MSA because resumptives appear in MSA when there is an island

violation or it needs to check the [+definiteness] feature/interpretation.

The third approach is the special kind of movement approach. The nonmovement

approach by McClosky (1990; i.e., the first approach) was the dominant approach during

the 1980s and the 1990s. However, in McClosky's (1990) paper, he discussed the

relation between Strong Crossover effects and resumptive structure, which gives a sign

that resumption could follow the movement approach. Furthermore, Tallerman (1983) in

Welsh and Tuller (1986) in Hausa discovered that resumptive pronouns show island

sensitivity, which means that they are also adopting the movement approach. Thus,

Demirdache (1991), Torrego (1995), Rouveret (1994), (Uriagereka, 1995a, 2005),

Sportiche (1996, 1998), Aoun and Benmamoun (1998), Belletti (1999, 2005), Aounand

Choueiri (2000), Cecchetto (2000), Aoun, Choueiri, and Hornstein (2001), Kayne (2002),

Aoun and Li (2003), and Boeckx (2003) proposed the big DP structure that first merge in

the derivation with a resumptive pronoun sitting in the head of DP and the dislocated

phrase, relative marker, and wh-phrase in the complement/spec of the big DP head as in

Examples (41a and b).

(41)   a.

b.

52
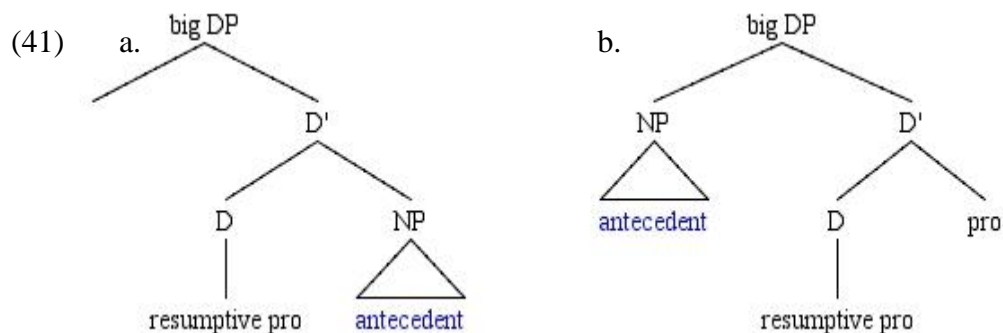
To sum up, the third approach, as the previous approaches, could not explain the alternation/optionality between the gap strategy and resumptive strategy. Rouveret (2011) summarized this continuous issue as follows:

> If movement is involved in the derivation of resumptive structure, the problem raised by the divide between gap and resumption is displaced. The choice is not between moving wh-expression or an operator-like element from the A'-bound position and merging a resumptive pronoun into the same position, but between moving in the overt syntax and moving at LF, as Demirdache (1991) proposes, or between moving a whole constituent to the periphery and extracting a subpart of it, as Rouveret (1994) suggests, or between spelling out and not spelling out the trace of movement"

To summarize this chapter, an overview of the phase theory by Chomsky (2000, 2001, 2008) and the split CP hypothesis by Rizzi (1997, 2001) have been presented. Following these two approaches, I discuss Ginsburg (2009) and Totsuka (2015) who combined these two frameworks and suggested that ForceP, TopicP, and IntP are phasal phrases while FocusP, FinP, and the embedded WhP are nonphasal phrases. Fassi Fehri (1982), Demirdache (1991), Shlonsky (2000, 2002), Aoun, Benmamoun, and Choueiri (2010), and others analyzed the structure of wh-questions in MSA and discovered that there are two types of wh-questions. I am suggesting that nominal wh-questions are phasal phrases while adverbial wh-question are non-phasal phrases. Finally, I have looked at the resumptive and gap structure and pointed out that resumption is one of the continuous problems in the literature; however, in this dissertation, the last resort approach was adopted. In the following chapter, the methodology which has been used will be explained.

CHAPTER 3: CORPUS ANALYSIS OF WH-QUESTIONs IN MSA

Introduction

Chomsky (1965) claimed that studying only the observed use of language cannot establish the linguistic field. In fact, he saw linguistics as two main divisions, which are linguistic competence (in which he was interested), and linguistic performance. Chomsky (1965, p. 4) defined linguistic competence as "the speaker-hearer knowledge of his/her language." while the linguistic performance is "the actual use of language in concrete situations. (p. 18)" Thus, performance cannot directly reflect the linguistic competence unless in an ideal situation. In a later work (Chomsky, 1986) revisited the definition and the distinction between competence and performance and renamed them as Internal-Language (I-Language) for the former and External-Language (E-Language) for the latter. Although performance data would be the available tool to test and evaluate the linguistic competence, so in the real world syntacticians have long mainly focused their analyses and data collection on the intuitions of native speakers. Currently, however, some research papers, theses, and dissertations have been based on corpus/corpora in order to have a real picture of linguistic competence (Butler, 2004). Conrad (2010) discussed the corpora and corpus analysis in the syntax (grammar) discipline and concluded that there are some grammatical/syntactic features that could not be tested or studied by using solely a native speaker who has only to say "grammatical/ungrammatical" and/or "acceptable/unacceptable." Thus, there should be an accurate tool that accurately represents the examined language. In the next section. I

will briefly discuss corpus analysis. That is followed by a description of the corpus used in this dissertation. Finally presented are the corpus findings from the data collection.

Corpus Analysis

The origin of the word *corpus* can be traced back to Latin, and the plural of the word corpus, in Latin, is corpora (and it is sometimes referred to as corpuses). which means *body*. Thus, corpus in the linguistic field would mean the *body* of language (Baker, 2010). Corpora and corpus analysis became popular during the evolution of personal computers in the 1990s. The Expert Advisory Group on Language Engineering Standard (EAGLES; 1996) defines corpus as follows:

> A corpus is a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language.

Thus, a corpus is seen as authentic materials (texts, videos, pictures, and/or audio records) that are stored in different registers (Biber, Conrad, andReppen, 1998). Leech (1992), McEnery and Wilson (1996), Tognini-Bonelli (2001), Hoey (2005), and Teubett (2005) discussed the corpus analysis and agreed that corpora are a new approach/method(ology) in studying and investigating linguistic phenomena. Furthermore, Tognini-Bonelli (2001) noted that there two types of the corpus analysis, which are the corpus-based approach and the corpus-driven approach. The *corpus-based approach* uses corpora/corpus as a source of examples to check the frequency of a linguistic feature in a small set of data while the *corpus-driven approach* uses all the data in the selected corpus, and, from that, the research would build and show the linguistic pattern the researcher found in this corpus. Following Tognini-Bonelli, Partington (2006) added one more type in the corpus analysis, which is the *corpus-assisted approach*. This

approach uses corpora/corpus as one of data sources in investigating linguistic phenomena. In this dissertation, as stated in Chapter 1, a combination of two types of data sets were used. They are: (a) corpus data, and (b) data from Arabic speakers based on their intuition. Thus, in this dissertation the corpus-assisted approach is used. In addition to the corpus analysis approaches, Paul Baker (2010) stated that, within the corpus analysis, there are sub-methods or sub-approaches that would help researchers to address and investigate their research questions. These sub-methods are (a) frequency, (b) collocation, (c) concordance, (d) keywords, and (e) dispersion. Thus, researchers need to choose one or more of these sub-methods in order to achieve/solve their concerns. For the purpose of this dissertation, concordance and frequency were chosen to present the structure of Wh-questions in MSA. In the following section, the Brigham Young University Arabic Corpus is presented with general information about it.

BYU Arabic Corpus

The Brigham Young University (BYU) Arabic Corpus is the largest accessible and free Arabic corpus and is based on Al-Sulaiti's homepage.   The BYU Arabic Corpus is an untagged and unparsed corpus. It has 173,600,000 tokens, and is divided into six main registers as follows: (a) newspapers, (b) modern literature, (c) nonfiction, (d) Egyptian Colloquial, (e) premodern, and (f) Arabic learner. Texts in this corpus were collected from almost all the Arab countries, such as Egypt, Saudi Arabia, Kuwait, Qatar, Sudan, Algeria, Syria, Iraq, Lebanon, and Palestine, and from different time periods (i.e., newspapers: from 1995-2001, modern literature from 2012, nonfiction from 1990-2010, Egyptian Colloquial from 2005 and 2014, premodern from the Islamic era, and Arabic

learner (not mentioned). It is a web-based search tool, so every time researchers must enter a string of Roman letters and/or characters and choose one or more/all register(s) in which they want to search. Table 1 shows the findings of the word *matha* (what) in the BYU Arabic Corpus. After entering the string (e.g., *matha* [word]), the result would be as shown in Table 1.

Table 1

Findings of the Word matha (what) in the BYU Arabic Corpus

As in Table 1, the corpus shows three of Baker's submethods, which are frequency (as in the summary and subsections), concordance (as in the citation), and collocation (as in the word before/after and citation). The advantage of using the BYU Arabic corpus is its number of tokens and accessibility. The disadvantage of this corpus

is that it shows raw texts (i.e., untagged and unparsed texts) so, for example, in Table 1 *matha* (what) has 2,484 occurrences in the modern literature register, which means that the researcher has to check one by one to see if they are a real wh-question or not. *Matha* is one of the easiest examples, but what about *man* (who). In Arabic, *man* could be the preposition (from), or the wh-question (who), so in this case the results have to be checked one by one to make sure the right examples are included. For the purpose of this dissertation, I only chose three registers: modern literature, nonfiction, and newspapers because they are the only registers that fulfill the definition of MSA. Because the concern in this dissertation the structure of wh-questions in MSA, I focused on the concordance while I showed the structure frequency of each of the wh-questions. Finally, due to the disadvantages and limitations of the BYU Arabic corpus, I decided to use the first 100 examples for each of the wh-questions in MSA. The following section presents the findings of the corpus analysis.

Corpus Findings

The purpose in this section is to present the findings of the corpus analysis, the aim of which aim is to fulfill the two sub-aims stated in Chapter 1 (sub-aim1: to compare between the gap and resumptive strategy by counting the structure frequency; sub-aim2: to determine the final destination of wh-questions by studying the concordance of the left periphery domain). Unfortunately, the corpus did not provide any example of extracted wh-words out of complicated clauses (islands) such as *anna-clauses*, complex NP clauses, or wh-clauses. Thus, the corpus only shows wh-questions either in small/simple

clauses or in relative clauses. However, the data show systematic patterns which are discussed in the following paragraphs.

The BYU Arabic corpus analysis provides evidence for the claim in this dissertation that wh-questions in MSA should be divided into phasal phrase and non-phasal phrase. The results could be divided into two major groups: (a) wh-questions that have the optionality to choose between gap and resumptive pronoun strategy in one case, or are forced to choose the resumptive pronoun strategy only in other cases, such as in *man* (who) and *ayy-a/u* (which) situations; and (b) wh-questions which are related only to gap strategy. The results are compiled in tables which are organized based on the frequency of the resumptive pronoun and the gap in the whole data of each wh-question. In addition to the structure frequency, the results also show the position of each wh-question in the clausal hierarchy by providing the relationship between wh-questions and the other functional projections in the left periphery.

Table 2 presents the findings of *man* (who) in the corpus analysis. It shows a comparison between the occurrence of gap and resumptive pronouns in the who-questions. Also, the findings show the position of the who-question in the left periphery domain.

Table 2

Structure of man-questions (who-questions)

man 'who'	Nonfiction	Modern Literature	Newspapers
Occurrence	100	100	100
Preceded by TopicP			4
Followed by FocusP	12	5	6
Resumptive in RC	54	50	64
Resumptive in non-RC	15	9	26
Gap	31	41	0

The findings here are straightforward. They show the relationship between the *man*-clauses with gap and resumptive pronouns, and with other split CP functional projections in three different genres/registers. In the nonfiction genre, there are 100 occurrences of *man*–questions. Thirty-one of these used the gap strategy, while 15 used the resumptive pronoun strategy. In both of those cases the wh-questions were in non-relative clauses (i.e., in simple/small clauses). In the relative clauses, the resumptive pronoun strategy was the only option used in the *man*-question structure. Fifty-four out of the 100 occurrences were who-questions in their final position while there is a resumptive pronoun sitting in its thematic position. Furthermore, there are 12 focalized elements following the *man*-question which provide evidence that *man* is not sitting in the FocusP because, theoretically, there cannot be two focalized elements in the same clause.

In the modern literature genre, almost the same scenario happened. In nonrelative clauses, the two strategies were used. Forty-one of the 50-small clauses were the gap in the thematic position of the wh-question, while 9 out of the total instances of the nonrelative clauses were resumptive pronouns in the *man*-questions. In the relative clauses, the only strategy used was resumptive pronouns. As in the non-fiction register, there were five focalized elements which were following the *man*-question.

The last register included in this analysis was the newspaper. There were 100 occurrences of *man*-question. Sixty-four of the total instances were *man*–structure in relative clauses. In that context, as in the other register, the resumptive pronoun was the only option for the wh-question to make the structure well-formed. Otherwise, the structure would ungrammatical. In the nonrelative context, the frequency of resumptive pronouns was higher than the gap, which is the contradictory picture of the results in the other genres. There were 26 occurrences of resumptive pronouns in the simple structure, while there were 10 occurrences of the gap in the rest of data. As in the other two genres, focalization appeared in the *man*-structure in the newspaper context. There were 10 instances of focalized elements following the *man*-question. Example (1) shows the *man*-resumptive structure while Example (2) presents the *man*-gap structure. Example (3) shows the *who*-word in the relative context.

(1) man mina   tadunu-**hu**   musib

     who   from-us   think-2SM-3SM   right

    "From us who do you think have the right?"

(2) man qatal__   alrajulkhafir Bunduq

who killed.3SM the-man.NOM chief Bunduq

"Hey chief Bunduq, who killed the man?"

(3) ya khafir Bunduq man allathi aʕtiyt-**hu**   hathih albarqiyah

hey chief Bunduq who that   gave-2SM-3SM this   the-letter

"Hey chief Bunduq, who did you give this letter?"

(BYU Arabic Corpus)

To sum up, in the non-relative clauses the wh-question in MSA has the option to choose between the gap or resumptive pronoun. On the other hand, in relative clauses the wh-question has only one option to select, which is the resumptive pronoun, in order to have a well-formed structure. Also, focalized elements could appear in the wh-structure in MSA, which gave us a sign that *man* (who) cannot be in the FocusP. Table 3 discusses the findings of the which-structure in MSA.

Table 3

Structure of ayy-a/u-questions (which questions)

ayy-a/u 'which'	Nonfiction	Modern Literature	Newspapers
Occurrence	100	100	100
Preceded by TopicP		6	
Followed by FocusP	11	14	11
Resumptive in RC	55	64	66
Resumptive in non-RC	17	6	19
Gap	45	36	34

The findings about the which-question in the corpus analysis showed some similarity between the *ayy-a/u* questions (which) and *man* questions (who). In all of the genres, *ayya/u*-questions have the optionality of choosing between gap and resumptive strategies in the nonrelative clauses (i.e., small/simple clauses). There were 17 occurrences of resumptive pronouns in the nonfiction context, 6 resumptive pronouns in the modern literature, and 19 resumptive structures in the newspaper genre. On the other side, there were 45 occurrences of the gap structure in the nonfiction genre, 36 instances of gap strategy in the modern literature context, and 34 gaps in the newspaper discourse. As in the who-question, there were also focalized elements following the *ayy-a/u* word, which suggests that the ayy-a*/u* question should not be in the FocusP. Examples (4), (5), and (6) show the structure of *which*.

(4) ayya   siyarh turid   _____

    which-ACC car   want-2SM

    "which car do you want?"

(5) ayyu   ʔitijah   tuʔidin-**h**

    which-NOM   way   support-2SF-3S(it)

     "which way do you support"

(6) ayyu   brnamj allathi taʕtabr-h   min aʕmalikum

    which-NOM   show   that   condsider-2SM-3S(it) from works-your

    alnajiha

     successful?   (BYU Arabic Corpus)

    "which show do you consider it as one of your successful shows?"

63

Interestingly, the findings of the corpus analysis showed that there were seven topicalized elements in the modern literature discourse which precede the *which*-word. This gives us a hint that *who* and *which* in MSA must sit in a functional project higher than FocusP and lower/in TopicP. Tables 4 and 5 present the data for the *ma*-question(what) and the *kam*-question.

Table 4

Structure of ma-questions (what-questions)

ma 'what'	Nonfiction	Modern Literature	Newspapers
Occurrence	100	100	100
Preceded by TopicP	6	11	2
Followed by FocusP			
Resumptive in RC	48	76	90
Gap	52	24	10

Table 5

Structure of kam-questions (how many/much-questions)

kam 'how many/much'	Nonfiction	Modern Literature	Newspapers
Occurrence	100	100	100
Preceded by TopicP	5	11	
Followed by FocusP/TopicP			
Resumptive in RC	36	27	52
Gap	64	73	48

The question words *ma* and *kam* share the same syntactic behavior in that they

select gap in the nonrelative clauses and resumptive strategy in the relative clauses. As in

the previous wh-questions, there were 100 *ma*-structures which were examined. The

findings of the *ma*-questions show a clear distinction between *ma* in a relative context

and in a nonrelative context. *Ma*-questions in relative clauses selected only the

resumptive strategy. Thus, in relative clauses, there were 48 instances of *ma*-resumptive

structure in the nonfiction genre, 76 occurrences of *ma*-resumptive in the modern

literature genre, and 90 occurrences of *ma*-resumptive structure in newpapers discourse.

On the other side, the *ma*-question picked the gap strategy for the non-relative clauses.

Thus, you cannot find a resumptive pronoun in a non-relative clause in a *ma*-question.

Finally, the data from Table 4 shows that topicalization appeared in the *ma*-question

while there is no instance of focalization. Examples (7) and (8) show the structure of the

ma-question in non-relative and relative clauses while Examples (9) and (10) present the *kam*-structure as in the corpus .

(7) ma   allathi qala-**h**   alraʔis   fi almuqabalh

what that   said-3SM-3S(it) the-president in the-meeting

"what did the president say in the meeting?"

(8) ma   raʔiyuk fi madumun alqisah

what your-opinion in context   the-story

"what is your opinion about the context of the story?"

(9) kam marah daxalt   alsijin

how times   entered-2SM the-prison

"how many times did you go to prison?"

(10) kam ʕadad alʕamalit allati tujurun-ha fi alyaum

how   many surgeries that did.3PlM-3S(it) in the-day

"how many surgeries did you make per day?"   (BYU Arabic Corpus)

The data of the corpus analysis would suggest that *ma* (what) should sit in the FocusP, and it can be preceded, but not followed, by a TopicP. Table 6 presents the corpus findings of the *matha*-questions in MSA.

Table 6

Structure of matha-questions (what-questions)

matha 'what'	Non-fictions	Modern Literature	Newspapers
Occurrence	100	100	100
Preceded by TopicP		14	6
Followed by FocusP/TopicP			
Resumptive in RC	N/A	N/A	N/A
Gap	100	100	100

The *matha*-word is the grammaticalized word of *ma*-word. *Matha* could be divided into two parts: (a) *ma,* which is the question word of *wha;* and (b) *tha* which is the demonstrative pronoun of *this*. This grammaticalized version lost some of its features. Thus, I assumed that *ma* has [+definite/neutral] in its first stage, and became [-definite] in the grammaitcalized stage (i.e., in *matha*). The [-definiteness] means relative clauses/operator cannot appear in the *matha*-structure because it is ungrammatical in MSA to have an indefinite noun phrase with relative clauses. Back to corpus analysis of *matha*, the findings show that *matha* only uses the gap strategy in all of the genres. Example (11) presents the *matha*-structure.

    (11) matha kunt   tafʕal__   fi alharah

      what   was.2SM doing.2SM in the-neighborhood

     "what were you doing in the neighborhood?"

The findings also show that topicalized elements could precede the *matha*-word, as in the modern literature register there were 14 occurrences of TopicP, and as in the newspapers register there were 6 occurrence of topicalized elements. The findings did not show any instances of focalization, which suggests that the *matha*-word must sit in the FocusP. Tables (7-10) provide the exact interpretation of their results.

Table 7

Structure of mataa-questions (when-questions

mataa 'when'	Nonfiction	Modern Literature	Newspapers
Occurrence	100	100	100
Preceded by TopicP	17	32	6
Followed by FocusP/TopicP			
Resumptive in RC	N/A	N/A	N/A
Gap	100	100	100

Table 8

Structure of ayn-questions (where-questions)

ayn 'where'	Non-fictions	Modern Literature	Newspapers
Occurrence	100	100	100
Preceded by TopicP	28	19	8
Followed by FocusP/TopicP			
Resumptive in RC	N/A	N/A	N/A
Gap	100	100	100

Table 9

Structure of kayf-questions (how-questions)

kayf 'how'	Nonfiction	Modern Literature	Newspapers
Occurrence	100	100	100
Preceded by TopicP	5	17	3
Followed by FocusP/TopicP			
Resumptive in RC	N/A	N/A	N/A
Gap	100	100	100

Table 10

Structure of limatha-questions (why-questions)

limatha 'why'	Nonfiction	Modern Literature	Newspapers
Occurrence	100	100	100
Preceded by TopicP	4	13	1
Followed by FocusP/TopicP			
Resumptive in RC	N/A	N/A	N/A
Gap	100	100	100

Villiers and Poeper (1995) discussed the relation between the wh-movement and relative clauses. They mentioned that extraction out of relative clauses or linking wh-words to the verb inside the relative clauses is impossible and makes the clause ungrammatical. Thus, the data in Tables 7-10 did not show any instances of adverbial wh-questions extracted out of the relative clauses, and/or linked to the embedded verb (i.e., the verb inside the relative clauses). Furthermore, in non-relative clauses the adverbial wh-questions always select the gap strategy; otherwise it would be ungrammatical. Examples (12-15) present the structure of the adverbial wh-questions.

(12) ayn   jamʕt-uk__   allati taʕalmt   fi-ha

where   university-your   that   learned.2SM   in-it

"Where is your university which you learned in it?"

(13) mataa tanqul   nashat-uk   ila masir _____

when move.2SM.Fut   work-your to   Egypt

"When will you move your work to Egypt?"

(14) kayf   harb   Mamdouh kharj masir

how escaped.3SM   Mamdouh out   Egypt

"how did Mamdouh escape Egypt?"

(15) limatha lam yahsil   hatha alfilm ʕala aljazih

why   not earn.3SM   this   movie on   the prize

"why did not this movie earn the prize?"

The data showed that topicalized elements could precede the adverbial wh-words, and there is no case of focalized elements in the adverbial wh-structure. This suggests that adverbial wh-words are located in the FocusP.

To sum up, the findings of the corpus analysis showed that there is some distinction between the *man* 'who'and *ayy-a/u* 'which' as one group and the rest of the wh-words (i.e., adverbial wh-words, *ma*'what,' and *matha* 'what') as the other group. The data showed that the *who/which*-question has some regulations in the relative or nonrelative clauses. In relative clauses they have to appear in wh-resumptive structure, while in nonrelative structure they have the freedom to choose between gap and resumptive pronoun. As for the other group, because the adverbial wh-words cannot link to the verb of the relative clauses, and the indefiniteness (i.e., [-definite]) of *matha* 'what,' those wh-words have to pick the gap strategy. For the *ma*-question'what' and *kam*-question 'how many/much,' they are in the grey area, but they are still close to the

71

second group. The *Ma*-and-*kam*-question in relative clauses has to appear in the *ma/kam*-resumptive structure, while in the non-relative structure they select the gap strategy.

In Chapter 4, some explanation of the corpus analysis findings based on the theoretical frameworks established in Chapter 2 is provided. Also included are some examples from sources other than the corpus in order to make the picture much clearer.

CHAPTER 4: ANALYSIS OF WH-QUESTION IN MSA

Introduction

After having looked at the data from the BYU Arabic corpus, the final task was to provide some explanation of the structure of wh-questions. As this dissertation stated earlier, there are two sources of data collection: (a) BYU Arabic corpus, and (b) traditional/modern grammar books. The aim in this chapter is to evaluate the structure of wh-questions from these two sources based on the framework and theories established in Chapter 2.

This chapter starts with recapping certain aspects of Chapter 2, some preliminary discussion of the structure of wh-questions, and a description of the main findings and results from the corpus analysis in Chapter 3. Following that is an analysis of the wh-questions in MSA. The analysis starts by applying the theoretical framework on the nominal wh-questions *man* and *ayy,* then on the other nominal wh-questions *ma* and *kam,* and, finally, on the adverbial wh-questions.

In Chapter 2, a brief summary of the modern Chomskyan theory and the split CP hypothesis by Rizzi (1997, 2001) have been presented. Chomsky (2000, 2001, 2004, and 2008) advanced the Minimalist Program. He said that deriving a sentence, from a lexical array to interfaces, is a set of chunks. Thus, syntax sends those chunks to the C-I system and the SM system in a series of cycles, but not all at once. Those chunks are called phases. Chomsky suggested that CP and vP are phasal phrases, while TP and VP are not. Chomsky attempts, in the MP and phase theory, to minimize and simplify the computational effort in the derivation of sentences. On the other hand, Rizzi (1997,

2001) had a deep and elaborate analysis of the clause structure. Rizzi thought that CP

projection is not enough to describe the discourse and pragmatic domain. Thus, he

proposed the split CP hypothesis. He suggested the CP should be split into four

functional projections (i.e., ForceP, TopicP, FocusP, and FinP). In 2001, Rizzi revised

his analysis and suggested two more layers (i.e., IntP and WhP) and an unlimited number

of TopicPs in the split CP hypothesis. Following those two approaches, Ginsburg (2009)

and Totsuka (2015) proposed a unified approach which reintroduced the split CP based

on the phase theory. The first sign of the unified approach was Chomsky's (2008) when

he mentioned that CP as a phasal phrase is the shorthand of Rizzi's split CP hypothesis.

Ginsburg (2009) worked on the interrogative feature and structure cross-linguistically,

and he claimed that IntP, FocusP, and WhP, which are the final destination of the

interrogative lexical in Rizzi's system, are not the same in phase theory. He said that IntP

is the phasal phrase while FocusP and WhP are not. Totsuka (2015) worked on Rizzi's

system based on the phase theory. He proposed that, based on Chomsky's definition and

diagnostic of the phase head, ForceP and TopicP are phasal domains while FocusP and

FinP are not. Taking those two claims into consideration, (1) represents split CP based

on the phase theory.

(1)

(Totsuka, 2015, p. 16)
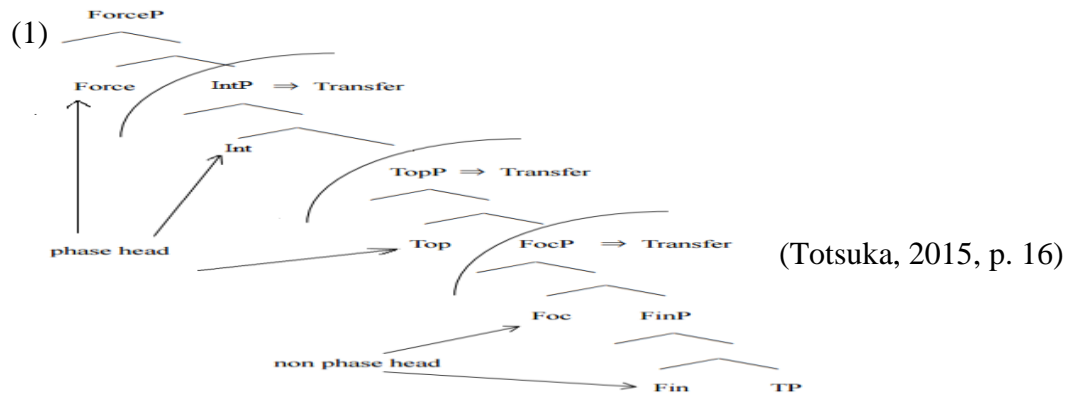
In addition, Chapter 2 gave a brief overview of the clause structure in MSA with a careful look at the structure of wh-questions in MSA. FassiFehri (1982), Wahba (1984), Demirdache (1991), Shlonsky (2000, 2002), and ABC (2010) studied the wh-structure and concluded that wh-questions are divided into two types: (i) nominal wh-questions, and (ii) adverbial wh-questions. The former could be linked with gap or resumptive strategy, while the latter has only the gap strategy option. Furthermore, in Chapter 2 the resumptive and gap strategies were looked at because they play a major role in the structure of wh-questions in MSA. In non-relative clauses, topicalized elements and a phasal phrase, which is separated from its original foot by another phasal phrase, must choose only the resumptive strategy. Furthermore, definite focalized elements in non-relative clauses have the option to choose between the gap and resumptive strategy. In relative clauses, the relativized NP could also have the option to choose between gap and resumptive strategy except the adjunct relativized NP could only pick the gap strategy.

In Chapter 3, the methodology which was used in this dissertation was discussed. A corpus-assisted approach has been adopted. For this dissertation, the BYU Arabic corpus was chosen to fulfill the goals of this study. It is the largest and free Arabic corpus on MSA. The wh-questions in the BYU Arabic corpus were studied, and the corpus showed that the data could be divided into wh-questions moved out of relative clauses and non-relative clauses. The corpus analysis showed that adverbial wh-questions must be linked to the gap strategy and could be preceded by a TopicP, but not followed by a FocusP or TopicP. On the other hand, the nominal wh-questions showed

that in relative context they all are linked to resumptive strategy, while in non-relative

context *man* 'who' and *ayy*'which' are linked to a resumptive pronoun while the *ma*

'what' and *kam* 'how many/much' pick the gap. The following sections analyze the wh-

questions in MSA, starting with the *man* and *ayya* structure, then the *kam*-questions and

ma-questions including the grammaticalized version of *what* in MSA (i.e., *matha*-

questions), and finally with the adverbial wh-questions.

man & ayy questions in MSA

Having established the framework and theories in Chapter 2, and Chapter 3

showed the patterns in the structure of wh-questions; it is now time to analyze the wh-

structures. In this section, the nominal wh-questions *man* and *ayy* in MSA are discussed.

The following examples (2-21) are from the BYU Arabic corpus. They only showed the

structure of those two wh-questions in relative and non-relative clauses. Examples (2-5)

present the *man*-questions in simple structures.

(2) man ʔaʕta-kum hatha alħaq?

who gave-2MP this   right

'who gave you this right?'

(3) man qatal   alrajul   khafirBunduq?

who killed.3SM the-man.NOM chief Bunduq

'hey chief Bunduq, who killed the man?'

(4) man ʔathina   lahum bi-aldukhoul fi manzili?

who allowed.3SM   them   to-enter   in my-house

'who allowed them to enter my house?'

(5) man yunqith   alshaʕb   alʕiraqi?

who rescue.3SM the-people the-Iraqi

'who rescues/helps the Iraqi people?'

It might be noticed that, in Examples (2-5), *man* 'who' is extracted from the subject position. In those examples, the rich agreement of the verbs could be resumptive and analyzed as a Big DP, or just as any agreement, and it would be a result of spec-head configuration. Choosing between resumptive or agreement in these examples would not make any change in the analysis, and this issue is a result of the continuing debate about the word order in MSA. Personally, this researcher prefers the resumptive option because it would go smoothly with Mohammed's (1989), FassiFehri's (1989), and McClosky's (2001) analysis for the word order in VSO and SVO languages. Thus, when *man* and *ayy* questions do not have a phasal phrase blocking their phasal projection from its original foot, they have the option to choose between the resumptive and gap. If there is a phasal phrase blocking *man* and *ayy* from their original foot, then they would go with the resumptive option to save the derivation from crash. The proposed analysis for Example (5) is shown in (5.a) for the spec-head configuration and in (5.b) for the resumptive strategy.

(5.a) CP/IntP

```
(5.a)  CP/IntP
           /\
          /  \
       C'/Int'        <= Transfer
        /\
       /  \
   C/Int    TP
       \    /\
    man'who'   T'
              /\
             /  \
            T    vP
            |    /\
        yunqith  man'who'    v'
                            /\
                           /  \
                          v    VP     <= Transfer
                          |    /\
                     yunqith  V    DP
                              |    /\
                            nqth  alshaʕb alʕiraqi
```
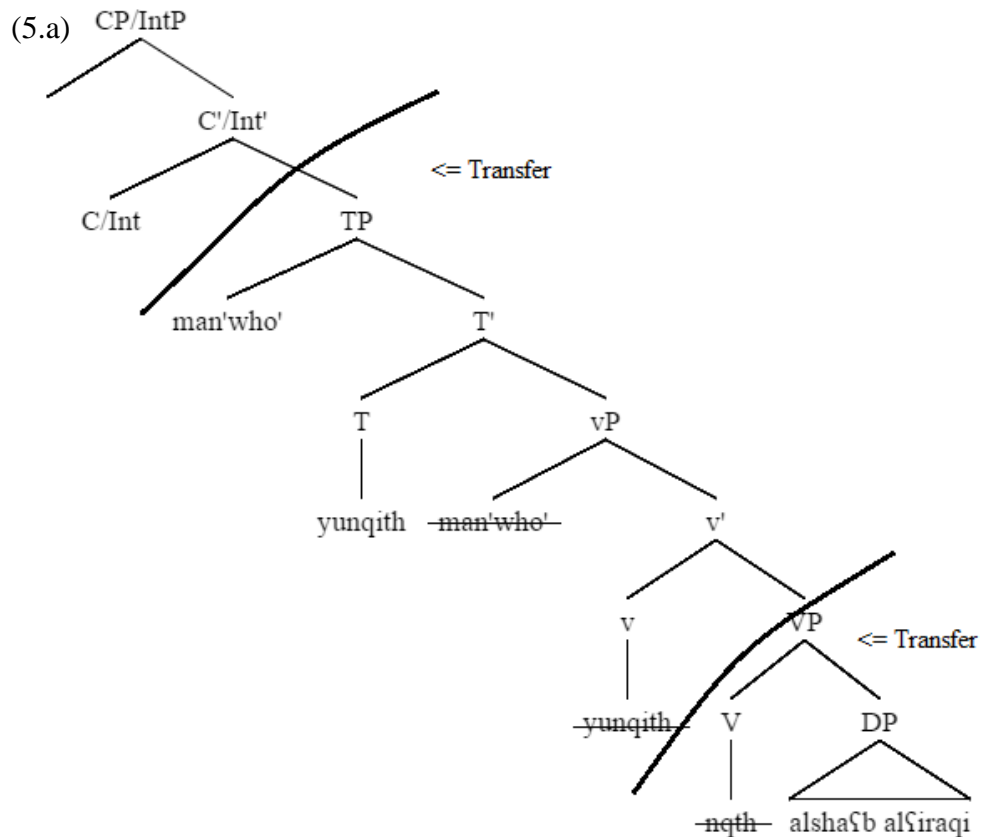
In Example (5.a), the derivation starts with selecting the determiner phrase

alshaʕb from the lexical array and merging it with the adjective phrase *alʕiraqi*, and

labels their projection as a DP. Then, the DP is merged with the verb *nqth* and projecting

the VP. The big VP merges with the small v, which is a phase head. The verb *nqth*

moves to the small v and, at the same moment, the phase head v sends the VP to the

interfaces. The word *man* is merging the [spec, vP] and labeling the whole projection as

a vP. Now it is time for the T to merge, which has the [tense] feature. In Arabic, the

verbs move to the T to check their [T] feature, and in (5.a) the verb *nqth* moves to the T

to check the [present] feature. The wh-word *man* moves to the [spec, TP]. The wh-word

checks the EPP feature on the T, and because the wh-word and the verb are in the spec-

head relation, the verb shows the full set of the phi-features (i.e., person, gender, and number) on the verb, and it became *yunqith*. [3rd,S,M]. Because the EPP would not allow the subject (i.e., wh-word) to move higher, the C head or Int head donate their feature (i.e., the [Q/wh] feature) to the T head, transfer the TP to the interfaces, and deletes itself.

(5.b)

ForceP

Force'

Force   IntP   <= Transfer

man   Int'

Int   TP   <= Transfer

~~man~~ RP.3SM   T'

leaving resumptive pro

T   vP

yunqith ~~man'who'~~   v'

v   VP   <= Transfer

yunqith   V   DP

nqth   alshaʕb alʕiraqi
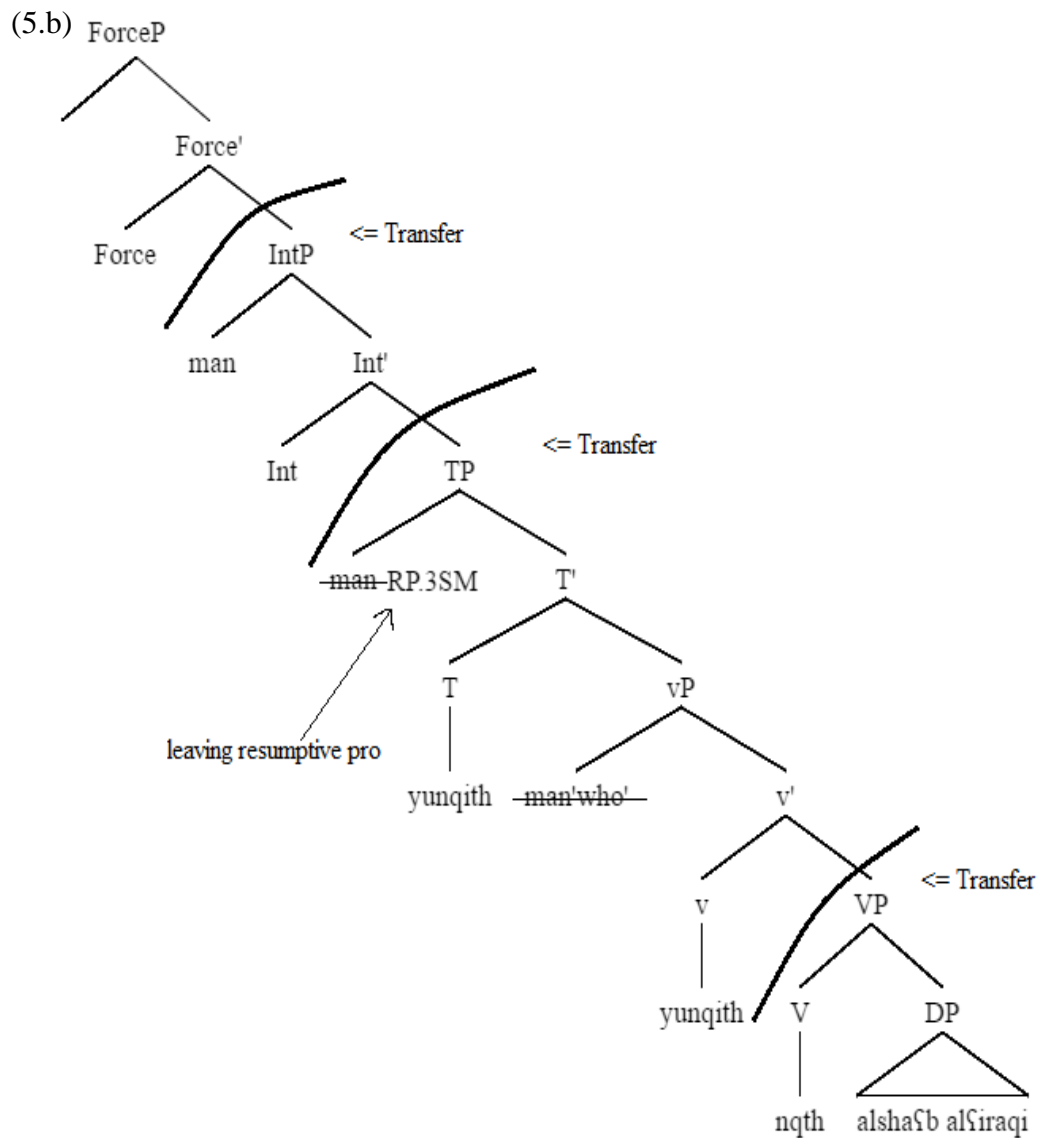
In (5.b), the same scenario would happen except that the wh-word (i.e. *man* 'who') and the resumptive pronoun, which has the [3rd, S, M] features, merge to the

derivation as a Big DP (i.e., the resumptive pro is sitting in the head of the Big DP, and the wh-phrase is sitting in the complement of the resumptive pro). The wh-phrase moves all the way to the spec of IntP to check the [Q, Wh] features while the RP stays in the Big DP to check the EPP feature. Int head is a phase head and transfers the TP, and as in Ginsburg's (2009) and Totsuka's (2015) proposals that ForceP is a phasal phrase, its main role is to show if the sentence is an embedded clause or not, and to transfer the whole clause to the interfaces.

For the *ayy*-questions 'which', the following Examples (6-9) present the structure of which-questions in simple structures. The interesting thing about *ayy* is that it shows the case marker on it. If *ayy* chooses to keep its original case marker (i.e., either ACC/NOM), it is then picking the gap strategy. However, if it chooses/has to leave a resumptive pronoun in its original foot, the nominative case marker will appear on the *ayy*. Examples (6 and 7) show that *ayy* picked the gap strategy and kept its original case marker on its way to IntP to check the [Q/Wh] features (i.e., leaving the gap in its original foot when it moves to higher projection for feature checking). The syntactic tree for the gap strategy derivation is provided in Example (7.a) for (7). Examples (8and 9), on the other hand, show that *ayy* picked the resumptive strategy and has a nominative marker as its case marker. Also, a structure tree is shown in Example (9.a) for example (9).

(6) ayy-a   Saudiyah nurid?

    which-ACC   Saudi   want.1Pl

    'which Saudi Arabia do we want?'

(7) Saeed ayy-a   saiyrat-in   yurid?

80

Saeed, which-ACC car-GEN   want.3SM

'Saeed, which car do you want?'

(7.a) ForceP

ayy-a saiyrat-in
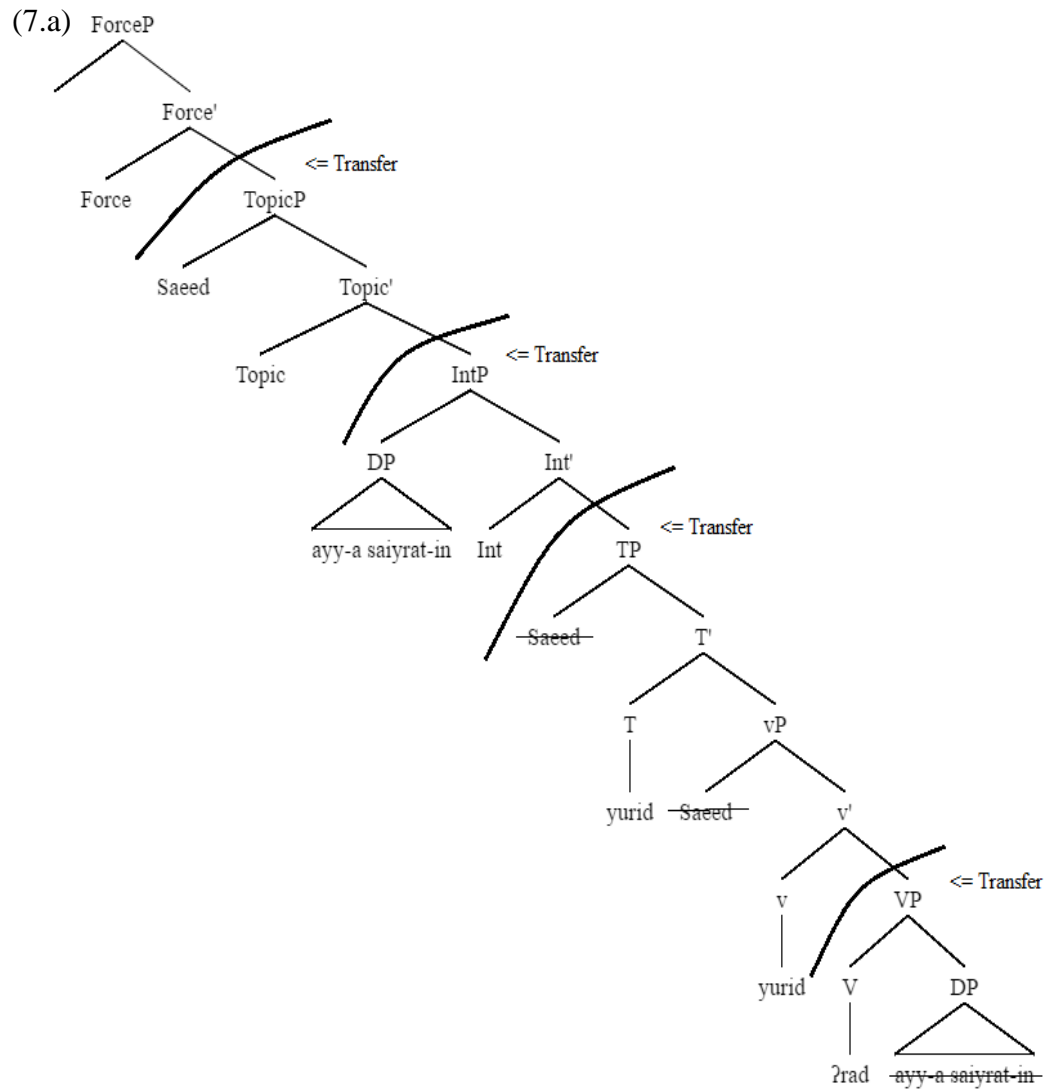
In Example (7.a), the derivation started with merging the verb *ʔrad* 'want' with

the wh-phrase (i.e., the DP). After that, the phase head v merges with VP and, at the

same moment, the verb moves from the V to the v, the wh-phrase moves to the specifier

position of the phase head v, and the phase head v spells-out/transfers the VP to the

interfaces. Then, the subject merges to the spec of phase head v. Now, it is time for the

T head to merge, and the verb moves to it to check the [tense] feature, the subject moves

to spec of the T head to check the EPP feature, and the phase head v now could label its

projection as vP. The phase head Int merges to the TP, the wh-phrase moves to spec of

Int head to check the [Q/Wh] features, and the DP (i.e., Saeed phrase) moves to spec of

the Int head to escape from spelling-out (i.e., not violating the PIC) and leaving the

resumptive pronoun, which values the phi-features on the verb in the T head, and the

phase head Int transfers/spells-out the TP. The Topic head merges to the IntP, the subject

moves to the spec of the Topic head to check the [Topic] feature, and the phase head

Topic spells-out the IntP to the interfaces. Finally, the Force head merges and

transfers/spells-out the whole clause to the interfaces.

(8) ayy-u   albaramj   taʕtabir-un-ah   min aʕmalikum   alnajiha?

     which-NOM the shows consider-2PlM-3PlM from your-shows successful

     'which shows do you consider them as your successful shows?'

(9) ayy-u   almanatiq   wajadtu-ha   mutalawithah?

     which-NOM   the-areas   found.2PlM-3PlM   polluted

     'which areas did you find them polluted? '

82

(9.a)

ForceP
├─ Force'
│   ├─ Force
│   └─ IntP   <= Transfer
│   ├─ DP
│   │   └─ ayy-u almanatiq1
│   └─ Int'
│   ├─ Int
│   └─ TP   <= Transfer
│   ├─ RP.2PlM
│   └─ T'
│   ├─ T
│   │   └─ wajadtu
│   └─ vP
│   ├─ RP.2PlM
│   └─ v'   <= Transfer
│   ├─ v
│   │   └─ wajadtu
│   └─ VP
│   ├─ V
│   │   └─ wajd
│   └─ DP
│   └─ -ha1 mutalawithah
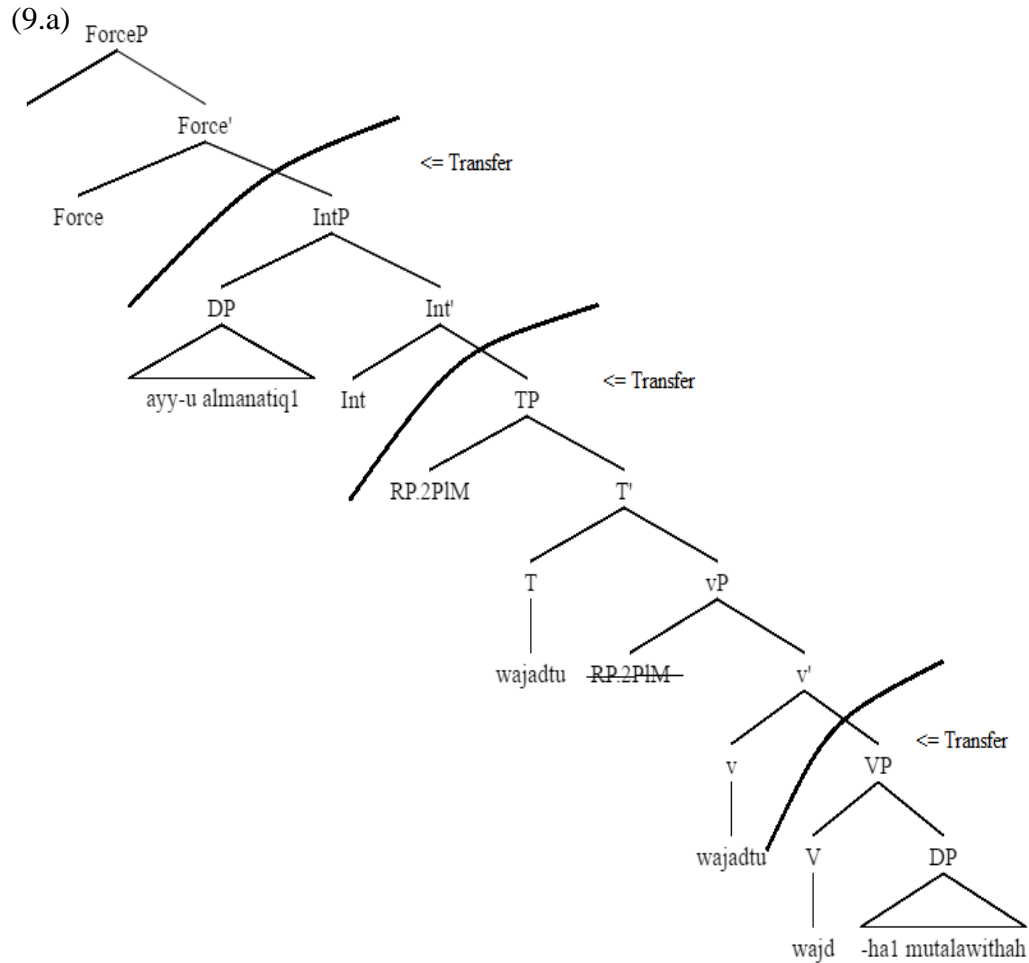
In Example (9.a), the same scenario happens. However, the wh-phrase and the resumptive pronoun enter the derivation as one unit (i.e., as a Big DP) and the wh-phrase raises to a higher functional projection while the RP sits in the Big DP and gets the original case marker while the wh-word *ayy* gets the nominative case marker.

Before moving to more complicated structures, Fargal (1989), and FassiFehri (1993) discussed topicalization and focalization in the two word orders SVO and VSO in MSA. They brought interesting examples about the wh-movement regarding these two structures, and found that it is ungrammatical to have a TopicP following the wh-

questions. For example, (10.a) shows the sentence in a non-topic context, (10.b&c) show the topic phrase in different positions, and, finally, (10.d) presents the grammatical version of (10.c).

(10) a. man   akala   Al-asad-u?

who   eat.3SM the-lion-NOM

'who did the lion eat?'

b. al-asad-u   man   akala(gap/-h)?

the-lion.NOM   who   eat.3SM

'The lion, who did it/he eat?'

c. *man   al-asad-u   akala

who   the-lion-NOM eat.3SM

d. man al-asad-u   akala-h

who the-lion-NOM   eat.3SM-3SM

'who did the lion eat him?'   (Fargal, 1989, p.62)

From previous examples, in this chapter it is proposed that *man* and *ayy* are sitting in a phase domain because its phase head requires having a clear way between the wh-word and its trace. Otherwise, the wh-word needs to have a resumptive, or (10.c) would be the result.

The following Examples (11&12) are the most important data from the BYU Arabic corpus because they provide evidence that *ayy* and *man* are not sitting in the FocusP as the other wh-words do in MSA, but in a projection higher than the FocusP. This projection must be either a ForceP, TopicP, or IntP. All of those projections are

84

phasal domain, and it should be the IntP because it is the home of [Q/Wh] features. As

mentioned in Chapter 2, *man* and *ayy* are phasal phrases, and they have the flexibility to

choose between gap and resumptive strategy in a non-blocking situation; however, in

case of a phase blocking between *man* or *ayy* and its original foot, the wh-words pick the

resumptive strategy. Also, it was noticed that, in case of definiteness, *man* and *ayy*

choose the resumptive option. It might be that definiteness is part of the narrow syntax in

MSA. The definiteness will not be discussed in this dissertation because it is beyond the

scope of this study. In other words, in a non-blocking environment, the wh-questions

alternate between the gap structure and the Big DP structure, while in

blocking/definiteness cases, the wh-questions have only the last resort approach, which

means that the wh-questions use RP to save the derivation from crash. The derivation of

Example (11) based on the phasehood in the split CP is shown in Example (11.a).

(11) man alrisalt   faqat   ʔaʕtat-ha   iyah   khafirBunduq?

who the-letter only   gave.2SM-3SF 3SM(him)   chief Bunduq

who did you give only the letter to, chief Bunduq?

(11.a) ForceP

```
ForceP
  Force'
    Force
    IntP          <= Transfer
      DP
        man2
      Int'
        Int
        FocusP    <= Transfer
          DP
            alrisalt faqat1
          Focus'
            Focus
            TP
              RP.2SM
              T'
                T
                  ʔaʕtat  RP.2SM
                vP
                  v'
                    v
                      ʔaʕta
                    VP            <= Transfer
                      V
                      DP
                        -ha1 iyah2 khafir Bunduq
```
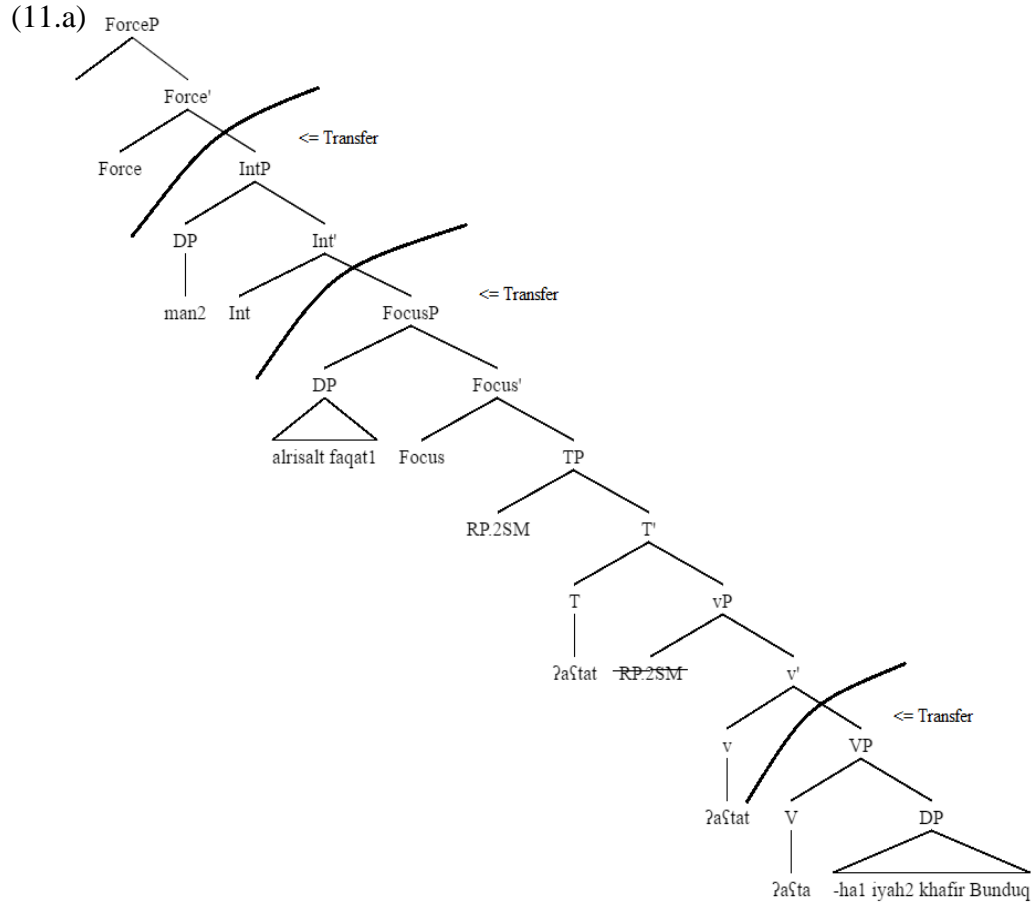
(12) ayy-u dawr sadiqk faqat kana yuʔad-ih amam AGT?

  which-NOM role your friend only was.3SM perform-3SM in-front-of AGT

  'which character did only your friend make/perform in front of the Arab Got

  Talent?'

Example (12) would have the same derivation as in Example (11.a), but with the

wh-word *ayy*. In Examples (11&12), both of them showed that the nominal wh-questions

man and *ayy* are located in the IntP, which is a phasal phrase to check the [Q/Wh]

features. Furthermore, in non-blocking situations the wh-questions could choose between

the gap and resumptive in their original foot. Definiteness might play a role in the

narrow syntax in MSA, which acts as a blocker as phases do, or it might reduce the *ayy* and *man*'s options to only the resumption.

It is now time to move to *man-* and *ayy*-questions in relative clauses. Examples (13-21) show the extraction of wh-questions out of relative clauses. In relative clauses, it is predictable for the wh-questions in those examples to pick the resumption option because there is a relative operator sitting in a phase domain and blocking the wh-word from its extracted position. The syntactic derivation of a wh-extraction out of a relative clause, as in Example (22), appears in tree form in (22.a).

(13) ya khafirBunduq man allathi aʕtiyt-hu   hathih albarqiya?

hey chief Bunduq who that   gave-2SM-3SM this   the-letter

'hey chief Bunduq, who did you give this letter?'

(14) man allathi yurisil   hathih albarqiya?

who that   send.3SM this   the-letter

'who sends this letter?'

(15) man allathi akbara-ka   bi-alkitab?

who that   told.3SM-3SM about-the-book

'who told you about the book?'

(16) man   allathi yuhasibu-hum ʕala khataya-hum?

who that   punish-3PlM   on sins-their

'who punishes them on their sins/mistakes'

(17) man allathi yurid   ʔan yukhbir   bi-hathih almaʕlumat Ali?

who that   want.3SM   to   tell.3SM   with-this information Ali

87

'who wants to tell Ali this information?'

(18) ayy-u   alfariqain   allathin   la   yuʔmnun   bi-libralih?

which-NOM the-groups(dual) that.3PlM not believe.3PlM in-liberal

'which groups do not believe in Mohammed?'

(19) ayy-u   alhalat   allatit   asdadʕi   alʕilag   fi almashifa?

which-NOM the-cases that.3SF   need.3PlF   the-medicine in the-hospital

'which cases need to be treated in the hospital?'

(20) ayy-u   dimuqratih allati tatahdath   ʕan-ha?

which-NOM   democracy   that talk.prog.2SM about-3SF(it)

'which democracy are you talking about?'

(21) ayy-u   ʔanwaʕ alsiyahah allati yuridu-ha   alyamen?

which-NOM types   tourism   that   want.3SM-3SF(it) the-Yemen

(Lit Translation: 'which tourism does Yemen want?'/ 'what kind of tourism

does Yemen want?')

(22) ayy-u   alakhani   allati tasmaʕ-ha   katheran

which-NOM   the-songs that   listen.2SM-3SM(it) frequently

'which song do you listen to frequently?'

(22.a) ForceP

Force'

<= Transfer

Force   IntP

Int'

<= Transfer

Int   ForceP

Force'

<= Transfer

Force   TP

allati   RP [2,S,M]   T'

T   vP

tasmaʕ   RP [2,S,M]   v'

<= Transfer

v   AspP

tasmaʕ   -ha   Asp'

Asp   VP

katheran   V   Big DP

smaʕ   D'

D   ayy alakhani

-ha [3,S,M]
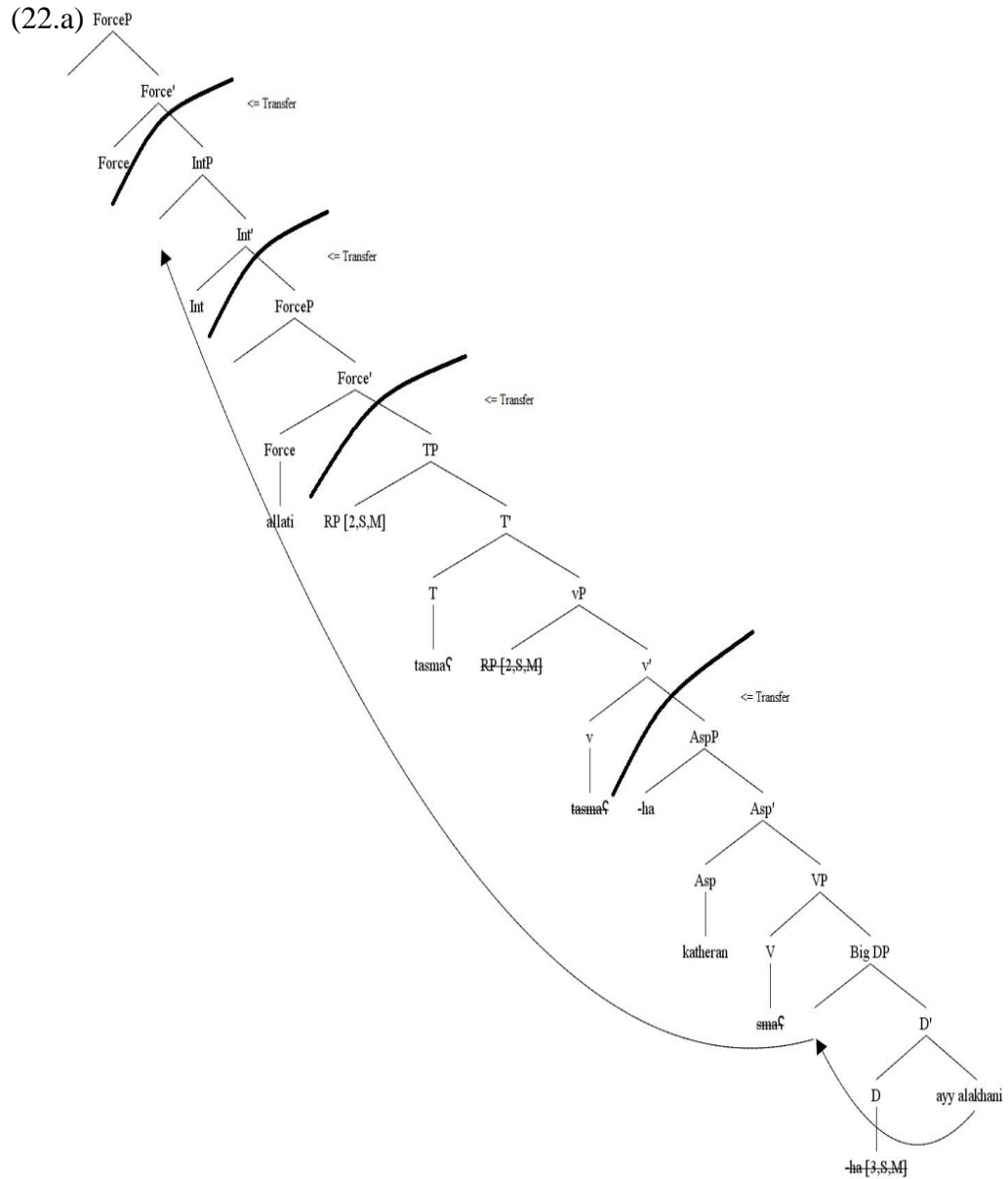
The final analysis in the *man* and *ayy* section discusses the wh-extraction out of *anna*-clauses. The following Examples (23-27) are from the literature and native Arabic speakers. The wh-extraction out of *anna*-clauses supports the researcher's assumption that blocking between the *man* and *ayy* in IntP as a phasal phrase and its original foot

89

causes the derivation to pick the resumptive strategy only as a last resort approach to save the derivation from crash. The syntactic tree for Example (23) is provided in (23.a).

(23) ʔayy-u   l-ʔawlaad-I   ħasiba   Aħmad-u   ʔanna-hum dahabuu?

which-NOM the-boys-GEN thought.3SM Ahmad-NOM that-3PlM left.3PlM

'Which boys did Ahmad think had left?'

(23.a)

here the derivation picks the last resort option (i.e. RP) becasue there is a phasal phrase blocking between the wh- and its trace.

(24) man taʕtaqidu   ʔanna l-ʔawlaad-a   qaabaluu-h?

who think.2SM that   the-boys-ACC met.3PM-3SM

'Who do you think that the boys have met?'   (Alotaibi&Borsley, 2013)

90

(25) man akkada   Salim-un   ʔanna-hu raʔa   dabʕ-an   fi alkhabah

who confirmed.3SM Salim-NOM that-3SM saw.3SM hyena-ACC in the-

forest

(Lit translation:'who did Salim confirm that _(he) saw a hyena in the

forest?')

Up to now, this section has analyzed the *man-* and *ayy*-questions in MSA. The

data from the BYU Arabic corpus have shown that those two wh-questions are sitting in a

functional projection higher than the FocusP and lower than ForceP and TopicP, which

means that they are located in the phasal phrase IntP. Furthermore, in a non-blocking

situation, the data have shown that *man* and *ayy* could choose between the gap and

resumption option. In the gap strategy, *man* and *ayy* show an accusative case marker and

leave nothing in its trace, while in the resumption they display the nominative case

marker and leave the resumptive pronoun, which gets the accusative case marker. On the

other hand, the data in blocking cases have shown that *man* and *ayy* have only one option,

which is the resumptive strategy, to pick. Otherwise, the computational process could not

run and we will get a [syntax error message]. The following section analyzes the *ma*,

kam, and *matha* questions. The structure of those wh-questions looks like the structure of

man and *ayy*, but they are not. In the following section, the reasons that do not make *ma*,

kam, and *matha* locate in a phasal phrase while they have the gap and resumption option

are discussed

91

ma, kam, & matha Questions in MSA

Having analyzed the first group of the nominal wh-questions, it is now the time to analyze the rest of the nominal wh-words (i.e. *ma*, *kam*, and *matha*). *kam* and *ma* nominal wh-questions do agree with *man* and *ayy* in their surface structure (i.e., they have a resumptive pronoun in relative clauses and/or in blocking context); however, they, in fact, differ in their syntactic (deep) structure.

The corpus analysis in Chapter 3 showed that the *ma*-questions and *kam*-questions cannot be followed/preceded by a FocusP, but can only be preceded by a TopicP, which is a phasal domain. This suggests that *ma*, *kam*, and *matha* are not sitting in the same projection as *man* and *ayy*. In fact, they are located in the FocusP as in the English wh-questions, and the reason for that is the [+focus] feature in these wh-words is very strong, while *man* and *ayy* are a lack of [focus] feature. Furthermore, focalization in MSA could have resumptives. Ayoub (1981) and Shonlsky (2000) mentioned that [+definite] focalized elements could alternate between the gap and resumptive strategy while [-definite] focalized elements have only the gap option. Thus, we can see in the *kam* and *ma* structures that they could choose between the resumptive and gap strategy while the *matha* has only one option, which is the gap strategy, because it has a [-definite] feature. So, the appearance of a resumptive pronoun in those wh-questions is not the result of a phase head because they are not sitting in a phasal phrase. It is part of the focalization structure in MSA. The following Examples (26-41) are from the BYU Arabic Corpus.

Examples (26-33) show the structure of the *ma* and *kam* in non-relative clauses, while Examples (35-42) present the structure of *kam* and *ma* in relative context.

(26) ma   raʔiyuk   fi madumun alqisah?

what your-opinion in context   the-story

'what is your opinion about the context of the story?'

(27) ma   hiya qisat alkitab   alaswad   ʔithan?

what   be   story the-book the-black so

'So, what is the story of the black book?'

(28) ma   huwwa mustaqabal alraismaliyah fi asia?

what   be   future   capitalism   in Asia

'what is the future of Capitalism in Asia?'

(29) ma   alsabab   wara   hathih almasaiyb?

what   the-reason behind these   problems

'what is the reason behind these problems?'

(30) kam marah   daxalt   alsijin?

how times   entered-2SM the-prison

'how many times did you go to prison?'

(31) kam diflan siunjiban?

how child future.make.3Pl

'how many kids they will have?'

(32) kam anfaqat   hathih aljamʕiyat   ʕala alʔiʕlanat?

how spend.3SF   these   organizations on   the-advertisement

'how much do these organizations spend on the advertisement?'

93

(33) kam shakhsin mat   fi afrikiya bisabab   fiqdan almeiyah   alnathifah

how person   died in Africa because of lack   the-waters the-clean

'how many person are dying in Africa because of lacking of clean water?'

The previous examples showed that the *ma* and *kam* are picking the gap strategy in the non-relative clauses, which is fine in the focalization structure in MSA. The following Examples (34a-b) also show the *ma* and *kam* in non-relative context, but they are picking the resumptive options using the Big DP approach. These examples are acceptable based on the judgment of Arabic native speakers. The syntactic tree for Example (34.d) is shown in (34.e).

(34) a. ?ma   ?ishtri-h   min alsouq?

what   bought.2SM-3SM   from the-mall

'what did you buy from the mall?'

b. ?ma   akalt-h   albariha?

what   ate.2SM-3SM   yesterday

'what did you eat yesterday?'

c. ?kam tufaha akalt-ha   albariha?

how   apple ate.2SM-3SM   yesterday

'how many apples did you eat yesterday?'

d. ?kam siyaratun ?ishtri-ha   Asama min almaʕard

how   car   bought.3SF-3SF   Asama from auto-show

'how many cars did Asama buy from the auto show?'

(34.e) ForceP

Force'

<= Transfer

Force   FocusP

DP   Focus'

kam siyaratun   Focus   TP

T'

T   vP

ishtri-ha   Asma   v'

v   VP   <= Transfer

ishtri-ha   V'

V   PP

shira   Big DP   min almaʕard

DP   D'

kam siyaratun   D

-ha
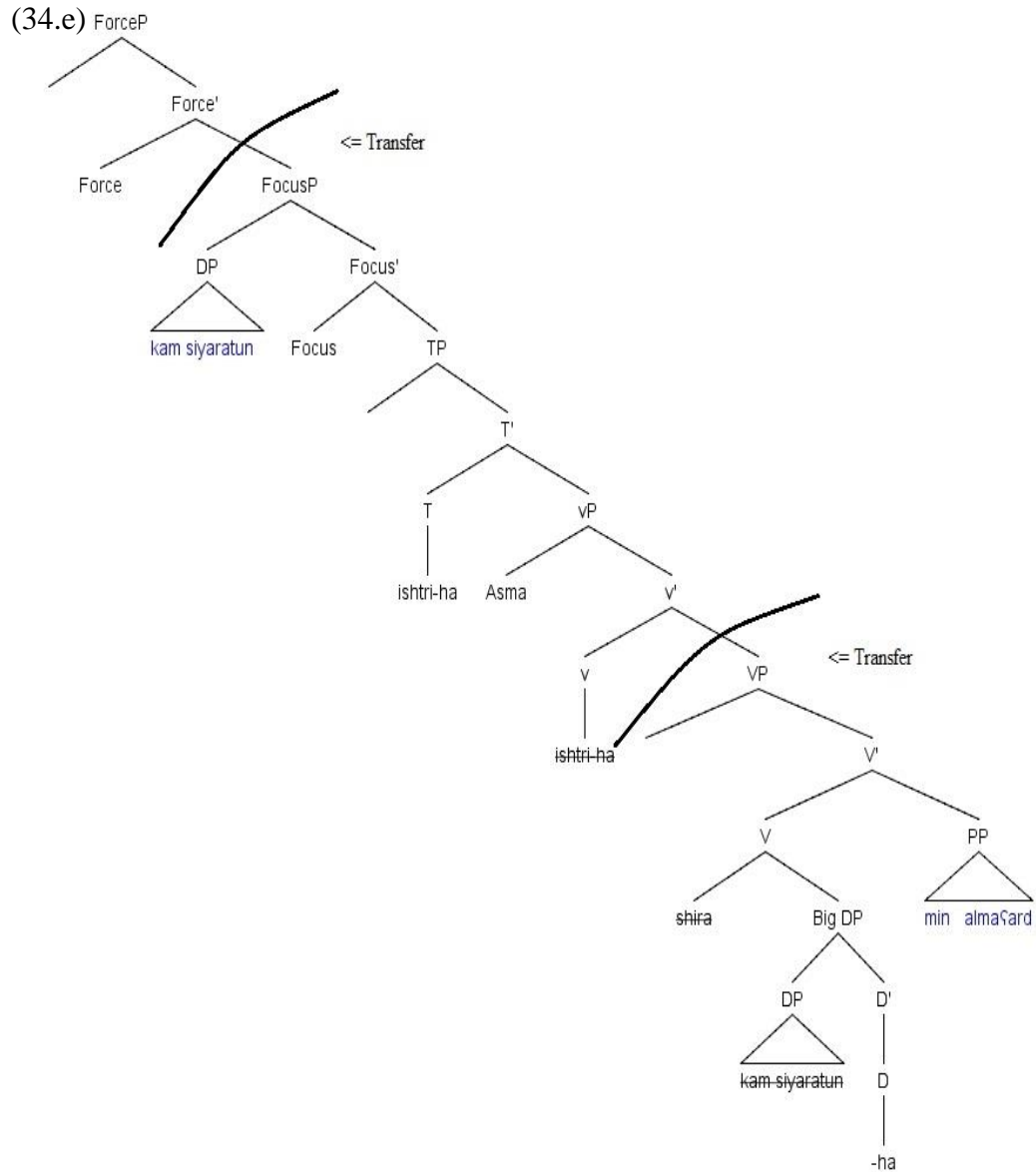
For the *ma* and *kam* questions in relative context, the data from the corpus in Examples (35-42) showed that those wh-questions are choosing the resumptive strategy. On the other hand, Examples in (43) present those wh-questions in relative clauses with choosing the gap strategy in relative clauses.

(35) ma   allathi qala-h   alraʔis   fi almuqabalh?

95

what   that   said-3SM-3S(it) the-president in the-meeting

'what did the president say in the meeting?'

(36) ma   alshiyʕ   aljadid allati taqul-h   almasrhiyah?

what   the-thing the-new that   saying.3SF-3SF the-play

'what is the new thing that the play is talking about?'

(37) ma   allathi   tukhifi-h?

what   that   hide.2SM-3SM

'what do you hide?'

(38) ma   allathi   yantathir-uk   hunak?

what   that   waiting.3SM-2SM   there

'what is waiting for you there?'

(39) kam   ʕadad alʕamalit allati tujurun-ha   fi alyaum?

how   many   surgeries that did.3PlM-3S(it) in the-day

'how many surgeries did you do per day?'

(40) kam ʕadad almuwathaf allathin   siyahtafith-un   bi-ʕamali-hum?

how many the-workers that.3PlM   future.keep-3PlM with-work-their

'how many workers will keep their work?'

(41) kam ʕadad   alkhuyul allati ladik   alʕaan?

how many the-horses that   have.2SM   now

'how many horses do you have now?'

(42) kam ʕadad almaqalat   allati katabt-ha   hata   alʕaan?

how many the-articles   that   wrote.2SM-3SM   until now

96

'how many articles have you written until now?'

(43) a. ?ma allathi ishtra   min alsouq?

   what that   bought.3SM   from the-mall

   'what did he buy from the mall?'

   b. ?ma   allathi akhatht   min   Asama?

   what   that   took.2SM   from   Asama

   'what did you take from Asama?'

   c. kam ʕadad almaqalat allati katabt   hata   alʕaan?

   how many the-articles that   wrote.2SM   until now

   'how many articles have you written until now?'

   d. kam   ʕadad alʕamalit allati tujurun   fi alyaum?

   how   many   surgeries that   did.3PlM   in the-day

   'how many surgeries did you do per day?'

To summarize this part, the nominal wh-questions are divided into two groups: (i) the first group is the phasal wh-questions (i.e. *man* and *ayy*), and (ii) the second group is the non-phasal wh-questions (i.e. *ma* and *kam*). Both of these groups could be linked to a gap or a resumptive pronoun based on the data from the BYU Arabic Corpus or based on the judgment of the Arabic native speakers. Thus, the main difference between these two groups is that the non-phasal wh-questions could alternate between the resumptive and the gap strategy in simple and relative context, while the phasal wh-questions (i.e. *man* and *kam*) could alternate between the RP and the gap option in a non-blocking context,

but in a blocking context, as in relative or *anna*-clauses, they have to choose only the resumptive strategy for the phase head's requirement.

The last nominal wh-question that must be discussed is the *matha*-question. It is the grammaticalized version of the *ma*-question. It could be divided into *ma,* which is the wh-word, and *tha,* which is the demonstrative 'this' in MSA. The interesting thing about this wh-question is that it lost its [definiteness] features; thus, it is ungrammatical to have *matha* in a relative clause because of [-definiteness] on *matha*. Also, because of the [-definiteness] on *matha*, *matha* has only one option, which is the gap strategy, no matter if *matha* is extracted out of a simple or complex clause such as *anna*-clauses. Table (6) in Chapter 3 showed that the occurrence of *matha,* which is linked to RP, is zero. The following Examples (44-45) present the *matha*-questions, from the BYU Arabic Corpus, while Example (46) is grammatical based on the Arabic native speakers.

(44) matha kunta   tafʕal   dikhil mabna   alwisarah?

      what   be.2SM   doing.2SM   inside building   the-ministry

      'what were you doing inside the ministry's building?'

(45) matha yajri   fi hatha albalad?

      what   happening   in this   the-town/country

      'what is happening in this town/country?'

(46) matha tadunu   ʔanna Ali-an   akal

      what   think.2SM   that   Ali-ACC   ate.3SM

      (Lit: 'what do you think that Ali ate?')

The following section is the second type of wh-questions which are the adverbials.

Adverbial wh-questions in MSA

Up to this point, we have seen in the nominal wh-questions that they have some similarity in their surface structure, but in their deep structure they do not. In fact, there are the phasehood's requirments versus the [focus] feature. As this dissertation stated previously, all the adverbial wh-questions are linked to the gap strategy, and they might be preceded by a TopicP, but not followed by a TopicP, which in the framework of this dissertation (i.e. phasehood in the split CP) means that the adverbials are sitting in a non-phasal projection and it should be the FocusP, like the English wh-questions, because they must not be followed by a phasal phrase such as the TopicP or the structure would violate the PIC. The following Examples (47-54) are also from the BYU Arabic Corpus.

(47) ayn   jamʕt-uk   allati taʕalmt   fi-ha?

where   university-your that learned.2SM in-it

'Where is your university which you learned in it?'

(48) ayn   kunti   tajulisin?

where   be.2SF   siting.2SF

'where were you sitting?'

(49) mataa tanqul   nashat-uki   lamasir?

when move.2SM.Fut work-your to Egypt

'When will you move your work to Egypt?'

(50) mataa badaʔ   ihtimam-uk   bi alʕamal   alsiyasi?

99

when started.3SM   concern/interest-your with the-field the-politics

'when have you started your interest in politics?'

(51) kayf harb   Mamdouh kharj masir?

how escaped.3SM   Mamdouh out   Egypt

'how did Mamdouh escape Egypt?'

(52) kayf nahmi   abnaʔ-na   min asdiqa ʔalsouʔʔ

how save.1Pl children-our   from friends   the-bad

'how can we protect our kids from bad friends?'

(53) limatha lam yahsil   hatha alfilm ʕala aljazih?

why   not earn.3SM   this   movie on   the prize

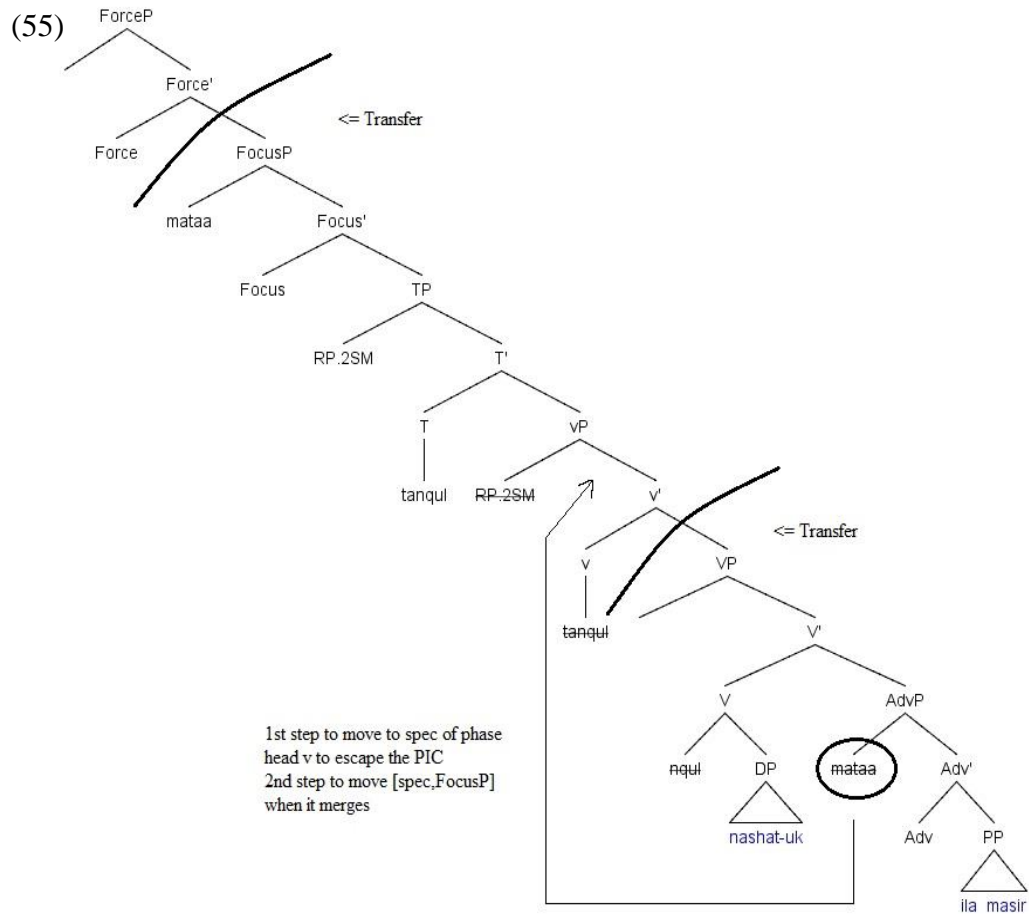'why did not this movie earn the prize?'

(54) limatha fakart   bi   alʕamal   alsahafi?

why   think.2SM   with the-filed   the-media

'why have you decided to work in the media?'

In Example (55), the structure tree is provided for the adverbial wh-questions in

Example (49)

(55)

ForceP
  Force'
    <= Transfer
    Force   FocusP
              mataa   Focus'
                      Focus   TP
                              RP.2SM   T'
                                        T   vP
                                        |
                                     tanqul   RP.2SM   v'
                                                      v   VP
                                                      |
                                                   tanqul   <= Transfer
                                                                          V'
                                                               V   AdvP
                                                          nqul   DP   mataa   Adv'
                                                                  nashat-uk   Adv   PP
                                                                                    ila masir

1st step to move to spec of phase
head v to escape the PIC
2nd step to move [spec,FocusP]
when it merges

To summarize what has been discussed in this chapter, analyses of the nominal

and adverbial wh-questions have been provided. We have seen that nominal wh-

questions in MSA share the wh-resumptive and wh-gap structure. However, they do not

share the same reasons behind the appearance of a resumptive pronoun and/or gap in the

wh-structure. The *man* and *ayy* questions in the data from the corpus showed that they

could be followed by a FocusP and TopicP and preceded by a TopicP. Thus, based on

Rizzi's system and the phasehood in the split CP by Ginsburg (2009) and Totsuka (2015),

those wh-questions must be located in the IntP, which is a phasal phrase. Furthermore,

this phasal projection requires a clear way from its original foot to its final position (i.e.,

101

no phase domain blocking between the wh- and its traces) in order to choose between the gap and resumptive strategy. Otherwise, it has to pick only the resumptive option. On the other hand, *ma* and *kam* are nominal wh-questions, but they are not phasal wh-phrases. They have a strong [focus] feature which requires them to raise to the FocusP while *man* and *ayy* lack a [focus] feature. For this, the *man* and *ayy* questions move all the way to IntP to check [Q/Wh]. The appearance of a resumptive pronoun in the *ma* and *kam* structure is due to the [definintess] feature. In MSA, the [definite] focalization structure could be linked to the gap and to the resumptive. On the other hand, *matha* cannot be linked to a resumptive pronoun because it has [-definiteness]. Finally, this chapter has provided an analysis for the adverbial wh-questions. There is not something interesting about them. They are just the like the English wh-questions. They raise from their original projection to the FocusP to check the [focus] feature while the [Q/Wh] are checked in prob-goal mechanism (i.e. Agree operation).

CHAPTER 5: CONCLUSION

In the first section of this concluding chapter, summaries are offered of previous chapters and the main goal of this dissertation is highlighted. In the second section, some ideas that have some relation with (/involved in) wh-questions are discussed, but they are out of the scope of this dissertation, and could be good topics for future research or dissertations. Finally, this chapter ends with a discussion of some limitations.

Chapter Summaries & Contributions

In this dissertation, the syntactic structure of wh-questions in Modern Standard Arabic was examined. Wh-questions have been studied in some research papers, but as part of the discourse/pragmatic (i.e., CP) domain's studies. However, there are a few research studies focused only on wh-questions in MSA (e.g., ABC, 2010; Aoun & Benmamoun, 1998; Aoun, Choureiri & Hornstein, 2001; Demirdache, 1991; Fargal, 1989). In these studies the structure of wh-questions in MSA were mis-analyzed because the surface structure of the nine wh-questions might look the same, but, in fact, they are not. For example, the hierarchical structure of the CP domain is not the same in those wh-questions, and the reason(s) behind picking the gap and resumptive strategy is (are) also not the same. Thus, the aim of this dissertation was to fill this gap by (re)studying and (re)analyzing the structure of wh-questions using the current frameworks (i.e., the MP and Cartographic approach).

In Chapter 2, a brief overview of the Chomskyan model (MP), and Rizzi's system was provided. In 1995, Chomsky proposed his famous framework, the Minimalist Program. He stated that the Faculty of Language could be divided into Logic Form and

Phonetic Form, while the syntax maintains the relation between those two parts.

Chomsky mentioned that, in syntax, there are four operations that run the computational

process, and they are *Select*, *Merge*, *Agree*, and, finally, *Transfer/Spell-Out*. In 2001,

Chomsky advanced the Minimalist Program to include phase theory, which aims to

restrict the syntactic operations. At the same period, Rizzi (1997, 2001) and Cinque

(1999) worked on the functional projections and concluded with the Cartographic

approach, the aim of which was to analyze clauses in detail. Chomsky, on the other hand,

preferred to describe clause structure in as simple and economic mechanism as possible.

Following those two approaches, Ginsburg (2009) and Totsuka (2015) unified them into

one approach. Ginsburg (2009) worked on the interrogative structure. Ginsburg

proposed that, using split CP hypothesis by Rizzi (1997, 2001) and the phase theory by

Chomsky (2001, 2004, 2008, 2013, 2015), the IntP should be TypeP and has all the

clausal mood while ForceP has only one role to show, which is if the clause is embedded

or not, and *Transfer* the whole clause to the interfaces. Totsuka (2015) basically

reintroduced the split CP hypothesis based on the phase theory. Totsuka proposed that

ForceP and TopicP are phasal phrases while FocusP and FinP are not.

In addition to the general overview of the current framework, in this dissertation

the clause structure in MSA was discussed, and it presented the main and famous four

approaches in analyzing the clause structure in MSA and/or its dialects. In this

dissertation, Mohammed's (1989), FassiFehri's (1989), and McClosky's (2001) analysis

about the clause structure in VSO languages was adopted. In Chapter 2, before talking

about wh-questions, the left periphery domain in MSA was discussed and it reintroduced

104

topicalization and focalization based on the unified approach by Ginsburg (2009) and Totsuka (2015), and it was discovered that phase theory (phase heads) play(s) a major role in deciding the hierarchical structure of clauses cross-linguistically and in MSA specifically. A focalized element preceding a TopicP cannot be found because TopicP is a phasal phrase and would *Spell*-Out/*Transfer* the clause before the focalized element escapes the phasal domain (i.e., we can not have FocusP preceding TopicP, otherwise FocusP would violate the PIC). Furthermore, in Chapter 2 the history of wh-questions' studies in MSA, are presented, and, finally, the relation between resumption and wh-structure was discussed. In this chapter, it was shown that resumptive pronouns could be due to the [definiteness] feature on the focalized element, or it could be a phase head's requirement, otherwise the clause would crash.

In Chapter 3, the methodology which was used in this dissertation was presented. The corpus-assisted approach was used in order to fulfill the purpose of this study. Tognini-Bonelli (2006) defined the corpus-assisted approach as a method which uses corpus (corpora) as one of the data sources. The Brigham Young University Arabic Corpus (BYU Arabic Corpus) was used in this study. It is the largest and free written Arabic corpus. It covers different periods of times and genres. It is focused only on Modern Standard Arabic (note: recently, some spoken and written files on the Egyptian Arabic and L2 of Arabic were added). For this dissertation, I chose to have 100 wh-structures for each of the wh-questions in MSA. They were picked randomly. In fact, I selected the first 100 wh-questions for each one and analyzed them. The data from the corpus analysis showed interesting results about the structure of wh-questions. The

results could be divided into wh-questions which are sitting in phasal phrase, as *ayy* and *man*, and wh-questions which are not, as in the rest of the wh-questions.

In Chapter 4, the attempt was made to link between what has been said in the literature (i.e., the unified approach by Ginsburg (2009) and Totsuka (2015)) and the corpus analysis in Chapter 2. In this chapter it was discovered that [focus], [Q/wh], [definiteness] features, and phase head's requirement plays a major role in the wh-derivation. It has been noted that extraction of the nominal wh-questions *kam*, *ma*, and *matha* questions out of a complex clause would be possible with or without the resumptive pronoun. On the other hand, extraction of *man* and *ayy* questions out of complex clauses, such as *anna*-clauses, *allathi*-clauses (relative clauses), would force these two wh-questions to pick the resumptive pronoun. If this were not done, it would crash because, in cases where there is a phasal phrase blocking between *ayy* and *man* and its original foot, those wh-questions have to save their derivation by selecting the resumption. Finally, Chapter 4 closes with an analysis of the adverbial wh-questions in MSA. It has been seen that adverbial wh-questions are like any English wh-questions. They are linked to gap in simple and complex contexts. They are sitting in FocusP because a focalized element and adverbial wh-question cannot be in the same clause, and they cannot be in phasal domain because they cannot appear before a TopicP, otherwise they would violate PIC.

Limitations and Direction for Future Research

Although in this dissertation I studied and analyzed an important structure in the left periphery domain, there are still remaining topics and issues that need to be studied,

106

such as the resumption in Modern Standard Arabic and/or its dialect. Also, research studies should be conducted on the historical changes in the wh-questions. Are they phrases or heads? What are the features they have/lost? An important topic that needs to be studied is the [definiteness] in Arabic. And, in relation with that, there should be a research study done on the relation between resumption and [definiteness], and phasal domain.

For the limitations, it was mentioned in Chapter 3 that the BYU Arabic Corpus is the largest, free, and accessible Arabic corpus currently available. However, it faces some technical and programming issues, such as it has only raw texts. They are not parsed or tagged, which makes the data collection not a happy and fun game. Also, the BYU Arabic Corpus does not cover Arabic in all of its periods, but only in certain time spots. Moreover, it does not cover all registers. Despite the limitations of the BYU Arabic Corpus, I am personally thankful to Dr. Dilworth B. Parkinson, its creator, for making Arabic accessible and available for online users.

REFERENCES

AL-Horais, N. (2009). A minimalist program to agreement in Arabic. *Newcastle Working Papers in Linguistics*, 15, 1-22.

Al-Shorafat, M. (1998) The minimalist program and the structure of Arabic clauses in an AGR-based model. *Papers and Studies in Contrastive Linguistics*, 34, 123-139.

Al-Saeedi, M. (2015). *The Rise of New Copulas in Arabic.* (Master Thesis). Arizona State University, Tempe.

Aoun, J., Benmamoun, E., Choueiri, L. (2010). *The syntax of Arabic*. New York: Cambridge University Press.

Aoun, J, and Benmamoun E. (1998). Minimality, Reconstruction, and PF Movement. *Linguistic Inquiry*, 29: 569–592.

Aoun, J., Benmamoun, E., & Sportiche, D. (1994). Agreement and conjunction in some varieties of Arabic. *Linguistic Inquiry*, 2, 195-220.

Aoun, J., and Choueiri, L. (1996). Resumption and Last Resort. Unpublished manuscript, USC, Los Angeles.

Aoun, J., Choueiri, L. and Hornstein, N. (2001). Resumption, movement, and derivational economy. *Linguistic Inquiry*, 32: 371–403.

Aoun, J. and Li, Y. A. (2003). *Essays on the Representational and Derivational Nature of Grammar: The Diversity of Wh-Constructions*. Cambridge, Mass.: MIT Press.

Ayoub, G.(1981). *Structure de la Phrase en Arabe Standard*. (Doctoral dissertation) Universit´e de Paris VII, France.

Bahloul, M.(1996). Extending the NegP Hypothesis: Evidence from Standard Arabic. In M. Eid, (ed.), *Perspectives on Arabic Linguistics: Papers from the annual symposium on Arabic Linguistics*, 8, 31–47. Amsterdam: John Benjamins.

Baker, P. (2010) *Sociolinguistics and Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Bakir, M. (1980). *Aspects of Clause Structure in Arabic*. (Doctoral dissertation). Indiana University, Bloomington.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics. Investigating language structure and use*. Cambridge: Cambridge University Press.

Boeckx, C. (2003). *Islands and Constrains. Resumption as Stranding*. Amsterdam: John Benjamins.

Butler, C. (2004) Corpus studies and functional linguistic theories. In M. Bednarek, M. Hilpert, and J. L. Mackenzie (Eds.) *Functions of Language*. Amsterdam: John Benjamins.

Carnie A., Harley, H., and Pyatt, E. (2000). VSO order as raising out of IP? some evidence from Old Irish. In A. Carnie and E. Guilfoyle (Eds.) *The Syntax of Verb Initial Languages* (39-59). New York: Oxford University Press.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. Greenwood Publishing Group.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge: MA: MIT Press.

Chomsky, N (2000). Minimalist inquiries: the framework. *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*. In Roger Martin, David Michaels and Juan Uriagereka (Eds). 89-155, MIT Press, Cambridge, MA.

Chomsky, N. (2001). Derivation by phase. *Ken Hale: ALife inlanguage*. In Michael Kenstowicz (Ed). 1-50, MIT Press, CambridgeMA.

Chomsky, N. (2004). Beyond explanatory adequacy. *Structure and Beyond:The Cartography of Syntactic Structures*. In AdrianaBelletti (Ed), 104–131, Oxford University Press, Oxford.References

Chomsky, N. (2007). Approaching UG from below. Interfaces+Recursion= Language? *Chomsky's Minimalism and the View from Syntax-Semantics*. In Uli Sauerland and Hans-MartinGärtner (Eds). 1-29, Mouton de Gruyter, Berlin.

Chomsky, N. (2008) On phases. *Foundational Issues in LinguisticTheory: Essays in Honor of Jean-Roger Vergnaud*. In RobertFreidin, Carlos P. Otero, and Maria Luisa Zubizarreta (Eds). 133-166,MIT Press, Cambridge, MA.

Chomsky, N. (2013). Problems of projections. *Lingua*, 130, 33-49

Chomsky, N. (2015). Problems of projection: Extensions. In E. Di Domenico, C.

Hamann, and S. Matteini (Eds.), *Structures, strategies and beyond: Studies in honour of Adriana Belletti* (pp. 1-16). Linguistik Aktuell/Linguistics Today, 223. Amsterdam, The Netherlands: John Benjamins. DOI: 10.1075/la.223.01cho

Citko, B. (2014). *Phase Theory: Introduction*. Cambridge: Cambridge University Press.

Conrad, S. (2010). What can a corpus tell us about grammar? In M. McCarthy & A. O'Keeffe (Eds.), *Routledge Handbook of Corpus Linguistics*. Milton Park, UK: Routledge.

Demirdache, H., (1991). *Resumptive Chains in Restrictive Relatives, Appositives, and Dislocation Structures*. (Doctoral dissertation) MIT, Cambridge, Mass.

Doner, J. (2013). A typology of EPP-checking mechanisms. *McGill Working Papers in Linguistics*, 23 (1), 24-37.

EAGLES, (1996). http://www.ilc.cnr.it/EAGLES/corpintr/node13.html

Eid, M. (1989). Verbless sentences in Arabic and Hebrew. In C. Bernard and M. Eid (Eds.) *Perspectives on Arabic Linguistics: Papers from the Annual Symposium on Arabic Linguistics*, 3, 32-62.

Fargal, M. A. (1986). *The Syntax of Wh-Questions and Related Matters in Arabic.* (Doctoral Dissertation). Indiana University

Fassi Fehri, A (1982). *Linguistique Arabe: Forme et Interpr´etation*. Rabat: Publications de la Facult´e des Lettres et Sciences Humaines.

Fassi Fehri, A. (1989). Generalized IP structure, case, and VS order. *MIT Working Papers in Linguistics*, 10, 75-111.

Ginsburg, J. R. (2009). *Interrogative Features*. (Doctoral Dissertation). University of Arizona, Tuscon.

Gordon, R. G., Jr. (ed.) (2005). *Ethnologue: Languages of the World, 15th Edition*.Dallas, Tex.: SIL International. Online version: www.ethnologue.com/

Hary, B. (1996). The Importance of the Language Continuum in Arabic Multiglossia. In A. Elgibali (ed.), *Understanding Arabic: Essays in Contemporary Arabic Linguistics in Honor of El-Said Badawi*. Cairo: The American University in Cairo Press. (69-90).

Hoey, M. (2005). *Lexical Priming: A New theory of Words and Language*. London/New York: Routledge

Kayne, R.(1994) *The Antisymmetry of Syntax*, MIT Press, Cambridge, MA.

Larson, R. (1985) Bare-NP adverb. *Linguistic Inquiry*14, 595.

Leech, G. (1992), Corpora and theories of linguistic performance. In J. Svartvik (ed.), *Directions in corpus linguistics: proceedings of Nobel symposium*, 82, Berlin and New York, Mouton de Gruyter, 125-148.

McCloskey, James (2001). The distribution of subject properties in Irish. In W. D. Davies and S. Dubinsky (Eds.), *Objects and Other Subjects: Grammatical Functions, Functional Categories, and Configurationality* (157-192). Dordrecht, Netherlands: Kluwer Academic Publishers.

McEnery and Wilson (1996). *Corpus Linguistics*. Edinburgh University Press

Mohammed (1989) The sentence structure of Arabic. Doctoral dissertation, USC, Los Angeles, CA.

Moutaouakil, A. (1989). *Pragmatic Functions in a Functional Grammar of Arabic*. Dordrecht: Foris.

Ouhalla, Jamal (1994). Verb movement and word order in Arabic. In D. Lightfoot , & N. Hornstein (Eds.), *Verb movement* (41-72). New York: Cambridge University Press.

Ouhalla, Jamal (1996). Remarks on the binding properties of Wh-pronouns. *Linguistic Inquiry*, 27(4), 676-707.

Parkinson, D. http://arabicorpus.byu.edu/index.php

Partington, A. (2006). *The Linguistics of Laughter: A Corpus-Assisted Study of Laughter Talk*. London and New York: Routledge.

Pollock, Jean-Yves, 1989. Verb movement, UG and the structure of IP. *Linguistic Inquiry*, 20: 365–424.

Radford, A. (1997). *Syntax: a minimalist introduction*. Cambridge: Cambridge University Press.

Rizzi, L. (1990). Speculation on verb second. In J. Mascaró and M. Nespor (Eds.), *Grammar in Progress: GLOW Essays for Henk van Riemsdijk* (375-386). Doreecht: Foris.

Rizzi, L. (1997). The Fine Structure of the Left Periphery. In Liliane Haegeman (ed.), *Elements of Grammar* (281-337). Dordrecht: Kluwer.

Rizzi, L. (2001). On the position `Int(errogative)' in the Left Periphery of the Clause. In Cinque & Salvi (Eds.), *Current Studies in Italian Syntax*. 287-296. Amsterdam: Elsevier.

Ross, J. (1967). *Constraints on Variables in Syntax*. (Doctoral dissertation). MIT, Cambridge, Mass.

Rouveret, A. (1994). *Syntaxe du gallois*. Paris: CNRS

Rouveret, A. (2010). On verb-subject languages. *Lingua*, 120, 232-263.

Rouveret, A. (2011) *Resumptive Pronouns at the Interfaces*. Amsterdam: Benjamins.

Ryding, K. (2005). A reference grammar of Modern Standard Arabic. New York: Cambridge University Press.

Sells, P, (1984). *Syntax and Semantics of Resumptive Pronouns.* (Doctoral dissertation). University of Massachussets, Amherst.

Sharvit, Y. (1999). Resumptive pronouns in relative clauses. *Natural Language and Linguistics Theory*, 17, 587-618.

Shlonsky, U (2000). Remarks on the complementizer layer of Standard Arabic. In J. Lecarme, J. Lowenstamm, and U. Shlonsky (Eds.), *Current Issues in Linguistic Theory* (325-345). Amsterdam: John Benjamin Publishing Company.

Soltan, Usama. (2007). On Formal Feature Licensing in Minimalism: Aspects of Standard Arabic Morphosyntax. Ph.D dissertation, University of Maryland, College Park.

Soltan, U. (2011). On issues of Arabic syntax: An essay in syntactic argumentation. *Brill's Annual of Afroasiatic Languages and Linguistics*, 3, 236-280.

Tallerman, M. (1983), Islands constrains in Welsh. *York Papers in Linguistics*, 10, 197-204

Teubert, W. (2005). My version of corpus linguistics. *International Journal of Corpus Linguistics*, 10 (1), 1–13..

Tognini-Bonelli, E. (2001). Corpus Linguistics at Work. Amsterdam/Philadelphia: John Benjamins.

Totsuka, M (2015). *On Phasehood of Functional Caterogries in the Left Periphery*. (Doctoral Dissertation). Tohoku University, Tohoku.

van Gelderen, E. (2004). *Grammaticalization as Economy*, John Benjamins, Amsterdam.

Versteegh, K. (1984). *Pidginization and Creolisation . The Case of Arabic.*Amsterdam-. Philadelphia, John Benjamins Publishing.

Villiers, J., & Poeper, T. (1995). Relative clauses are barriers to wh-movement for young children. *J. Child Lang.*, 22, 389-404.

Wahba, W. (1984). *Wh-constructions in Egyptian Arabic*. (Doctoral dissertation). University of Illinois, Urbana-Champaign