Verbal Complementation in Egyptian Colloquial Arabic
An LFG account

Shaimaa ElSadek

A thesis submitted for the degree of Doctor of Philosophy

Department of Language and Linguistics

University of Essex

December, 2016

Abstract

This study provides description and analysis of some verbal complementation patterns in Egyptian Colloquial Arabic (ECA), namely the Auxiliary *kān*, Causative *ḫalla*, phasal verbs and modals. Each verb is represented by a set of sentences extracted from a 5 million word corpus of ECA online texts that was built for the purpose of the current study using the Sketch Engine tool. These verbal complements are described and analysed within the principles of LFG syntactic theory, and represented in a grammar fragment implemented using the XLE tool.

The analysis shows that both tense and aspect can be expressed verbs in ECA, where in simple tense forms the verb carries tense only, while in compound tense, the main predicate marks tense and occupies I while the following lexical verb marks grammatical aspect and occupies V. The bi- prefix marks present tense on verbs in I and imperfect aspect on verbs in V, as well as a HAB/PROG feature. The bare Imperfective verb form is treated as a non-finite verb in ECA, where it can not occupy I and is marked by **VFORM=BARE**. All of the verbal constructions analysed are bi-clausal structures, however, they show differences regarding the kind of control relation. Functional control was attested in constructions where the main predicate is the auxiliary *kān*, the causative verb *ḫalla*, phasal verbs, as well as non-inflecting modals. Anaphoric control was attested only with inflecting modals, with the modal *yiʔdar* 'able' showing a case of obligatory anaphoric control.

This is, to my knowledge, the first study which attempts to develop a grammar for ECA using the XLE platform. It provides an insight over the issues correlated with developing this grammar, which could be a step towards including ECA into the ParGram project in order to develop broad coverage grammars for a bigger number of languages.

Acknowledgements

This PhD thesis is the result of an exciting journey full of highs and lows, through which I have learned a lot and got to know some really beautiful people, and this is a humble expression of gratitude to all of them. On top of the list is my supervisor, Prof.Louisa Sadler, who has helped and guided me in every step of the way, encouraged me to push myself further than I thought I can go, and supported me in many ways. She is a real role model to me, and I am proud to be her student. I am also lucky to be in close contact with Dr. Doug Arnold, who has helped me solve many technical issues with my XLE, as well as overcome a lot of the PhD lows with his words of wisdom, and I have to quote here "It only has to be too good to fail". He is definitely my PhD Guru. I am also grateful to my examiners, Prof. Miriam Butt and Prof. Robert Borsley for their insights and valuable comments, which helped me improve my work and reassured me that my work was really too good to fail. I will do my best to keep up the good work.

I would also like to thank the Ministry of Higher Education in Egypt and the Egyptian Cultural Bureau in London for providing me with the scholarship that faciliated my stay and study in the UK. I loved being a student at the University of Essex and I am thankful for the staff members of the department of Language and Linguistics for their constant help and support. The PhD journey would not be complete without the company, and I was lucky to get to know great people as fellow students at the department with whom I shared lots of laughs and worries as well as seminars and research groups. On top of this list is Paloma Carretero Garcia, my friend and companion with whom I started and ended this journey, who gave me the emotional support when I needed it most, and with whom I shared lots of laughs and crazy adventures as well as achievements and sometimes tears, as well as a marvelous Japanese experience. She was my family in Essex.

Last but by no means least, my deepest gratitude goes to the three pillars of my life. My mom, who encouraged me every step of the way, my cousin Sarah, who believed in me and helped me go through a lot. And Mohamed, my life partner, husband, and best friend, who supported me in every possible way, believed in me, endured my moods and provided me with whatever I needed to finish this work, his presence made this journey much more exciting, and I dedicate this work to him.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

There are three main varieties of Arabic available in any Arabic speaking environment, which are Colloquial Arabic, Modern Standard Arabic and Classical Arabic (Beshr, 1998). Classical Arabic is the language of heritage, and Modern Standard Arabic (MSA) is the unified Arabic variety used in formal discourse across the Arab world, while colloquial Arabic is the spoken variety which native speakers of Arabic learn as their mother tongue (ElSadek, 2010). Despite the vast amount of literature on Arabic language and description of its linguistic structures, most of this literature is devoted to Modern Standard Arabic, while studies of the colloquials mostly fall within the domain of dialectology (ElSadek, 2010). This is partly motivated by the negative attitude towards the colloquials which are sometimes referd to as a ""common, vulgar language" *'ammiyya* as against the "correct, classical language" *fusha*" (Versteegh, 2004, p.1741). Some scholars also claim that most of the Arabic dialects have not developed a lot far from the classical variety, which obscures some of the salient properties of these dialects, especially in the field of syntax (Versteegh, 2004). All these factors led to the scarcity of linguistic descriptions of the colloquials, especially in terms of analyzing its syntactic structures.

This is the gap which the current study aims to remedy a part of, where the main focus of this study is on the description and analysis of some of the Egyptian Colloquial Arabic (ECA)

verbal complementation patterns, namely the Auxiliary *kān*, Causative *ḫalla*, phasal verbs and modals. Data on which the description is based represents ECA as it is currently used through collecting a corpus of written ECA texts. Each of the verbal structures is then analysed through the Lexical Functional Grammar (LFG) syntactic theory, and then the analysis is implemented through the XLE program. This method of description and analysis of ECA verbal structures is the first of its kind as far as I know, where the LFG treatments of Arabic structures in the literature focus on MSA as well as some structures in different dialects of Arabic, while there were no similar works describing these verbal structures in ECA within the domain of LFG.

This study also serves to tackle some interesting questions about structures in ECA, such as whether the imperfective verb form represents a non-finite form in ECA despite the fact that it inflects for person, gender and number as other finite verb forms[1]. This is also related to the question of the function of the bi-prefix which denotes progressive and habitual aspects on verbs in ECA, while it is absent from MSA and most of the other dialects of Arabic. Another question is related to miscreant agreement that is attested with some verbs such as the auxiliary *kān* and phasal verbs as well as modals, where some of these verbs can optionally agree with the subject, rendering grammatical sentences whether it agrees or not. And last but not least, this study discusses the role of corpus-based studies in providing authentic data as the basis for analysis, especially for a diverse and widely used dialect such as ECA that undergoes continuous changes in its structures that need to be properly described and analysed, as well as the big percentage of code-mixing between the colloquial and MSA in its written form, which needs checking by a native speaker that the structures used are in fact colloquial[2].

This introductory chapter provides background information for the rest of the thesis, starting by section 1.1 which defines the notion of verbal predicates and their complementation patterns, which is the main focus of the current study. While section 1.2 describes the linguistic situation

---

[1]This is discussed in more details in section 1.4.
[2]Discussed in more details in section 1.3.

in Egypt and defines ECA. Section 1.3 provides information about the corpus built to serve as the data source for the current study, as well as the issue of finding dialect corpora for Arabic. Section 1.4 discusses the issue of the non-finite verb form in ECA, and whether this is represented by the imperfective verb form or not[3]. Section 1.5 discusses the notions of anaphoric and functional control and the basis for defining each within the principles of LFG syntactic theory. These priniciples will be the basis for the analysis of verbal predicates throughout the current study. Finally, section 1.6 provides an outline for the rest of the thesis.

1.1 Verbal complementation

Tallerman (2011, p.39) states that: "The major function of verbs is to express what is known as 'predication'. A PREDICATE expresses an 'event' in the sentence, which may be quite literally an event (such as collapse or explode) but also includes actions, processes, situations, states and so on." Within the theory of LFG, verbal predicates are assumed to provide the essential backbone of a clause, since they are the elements that require presence of the arguments necessary to make the clause grammatical (Butt et al., 1999). This property of verbs is manifested through the lexical entry which specifies the subcategorisation frame of the verb and the clause structure of its complements, which is central to any account of grammar of a langauage, as stated in Butt et al. (1999, p.43) "the determination of a verbs subcategorization frame and the writing of verbal entries constitutes a central part of any grammar development effort."

This illustrates the importance of studying verbal complementation and finding common patterns by which a verb selects its complements, which is the motivation behind the current work. The choice of verbs to investigate was mainly driven by verbs in Egyptian Colloquial Arabic (ECA) which lack adequate grammatical study, and which are known to have specific complementation patterns in other languages. These are:

---

[3]Refered to in the rest of the study as the bare Imperfective verb form.

- The auxilliary *kān* which expresses compound tense in ECA, as in:

  (1) ʕali kān   bi-yzāker   fi el-bēt
      Ali   be.PV.3SGM   BI-study.IPFV.3SGM   in the-house
      Ali was studying at home

- The causative verb *ḫalla* which forms the analytic/syntactic causative structure in ECA, as in:

  (2) Mona ḫallet   ʕali yizāker   fi el-bēt
      Mona make.PV.3SGF   Ali   study.IPFV.3SGM   in the-house
      Mona made Ali study at home

- Phasal verbs in ECA, as in:

  (3) ʕali badaʔ   yizāker   fi el-bēt
      Ali   start.PV.3SGM   study.IPFV.3SGM   in the-house
      Ali started studying at home

- Modals in ECA, as in:

  (4) ʕali lāzim yizāker   fi el-bēt
      Ali   must   study.IPFV.3SGM   in the-house
      Ali must study at home

Complementation is defined by Noonan (2007, p.52) as "the syntactic situation that arises when a notional sentence or predication is an argument of a predicate." Different verbs across languages vary regarding both the number and kinds of complement types they can have, which can differ in terms of the morphological form of the predicate as well as the syntactic relation between the predicate and its arguments represented in the complement construction (Noonan, 2007).

In his comparative study of complementation patterns in French, English and Arabic, Abdelmoumene (1988, p.16) states that "complements are distinguished by: major sentence types such as statements, questions, commands, etc., and by other structures such as finite and non-finite

forms; subjectless and subjectful clauses; tense/aspect variations and different occurrences of complementisers." These criteria are investigated for each of the verbs in the current study, with special attention to the notion of non-finite forms in ECA and which forms can be said to be non-finite, and the possible markings of morphosyntactic tense/ grammatical aspect on both the main predicate and its complement. While the question of subject vs subjectless clauses was not tackled in the current study due to the property of Arabic language as being a kind of pro-drop or null-subject language, having rich verbal morphology where the subject can be indicated as an inflection on the verb. The verbal paradigms available make the subject pronoun redundant, and therefore application of this criteria is tricky (Abdelmoumene, 1988).

1.2 Egyptian Colloquial Arabic

Egyptian Colloquial Arabic is also known as Cairo Arabic or Egyptian Arabic. Although it is originally the dialect of Cairo, the capital of Egypt, ECA is adopted by speakers in numerous urban areas all over Egypt as the main linguistic variety used in everyday life. It is also used extensively in the Egyptian media and in some written contexts such as novels, folk poetry and memoirs (Woidich, 2006). The variety of ECA investigated in the current study corresponds to what Badawi (2012) describes as *ʕāmiyyat al-motanawirīn* translated here as 'Colloquial of the enlightened'. This dialect was also refered to in Azer (1980)'s study by the letters ECA which stand for Egyptian Colloquial Arabic, however it was described as spoken Cairene Arabic. This is defined as the variety that is refered to in scholarly and scientific articles about Egyptian Colloquial Arabic, it is used by a wide variety of social classes and educational levels as the day to day conversational vehicle, and it is the mother tongue of Egyptian children among the educated classes (Badawi, 2012). I chose this language variety as it is my native dialect, being born and raised in Cairo, it is also the variety mostly used in colloquial written text from which the data is extracted.

Among the linguistic features that distinguish this dialect is its morphological and syntactic structures that do not follow the rules of Modern Standard Arabic, such as the prevalence of SVO structures more than VSO, the free position of wh words in questions, as well as adding the *bi-* prefix to imperfective verb forms to indicate habit, and *ha* or *ħa*[4] to indicate future (Badawi, 2012). The use of the *bi-* prefix which marks the indicative in spoken varieties of Arabic often represents "one of the as yet unanswered questions of Neo-Arabic structure. It is an aspect of a general problem of describing exclusively spoken languages, i.e., all Neo-Arabic varieties with the exception of Maltese" (Maas, 2009, p.129). The function of this prefix will be tackled in more details with respect to tense and aspectual marking in ECA in section 2.3.1.

Despite the fact that ECA is not usually used in formal contexts in Egypt such as education and newspapers, it has a wealth of resources in the spoken medium, in the form of TV programs, movies, theatre plays and songs. It has also recently been used in the written form in limited media such as online magazines, blogs, forums, social media as well as a few books, however most of these resources include lots of code mixing between ECA and MSA. The prevalence of this code mixing is clear in Tawakol (2008)'s study of the usage of the bi-prefix, which marks habitual or progressive aspect in ECA, on verbs of MSA (refered to in her study as *fusha* or literary Arabic) which are grouped under the term *'B-Semi-Fusha-verbs (BSFV)'* and are used in semi-formal contexts such as television programs. This is the reason the current study is based on a corpus collected from specific websites such as blog posts and online magazines that are written mostly in the colloquial, and wherever there exists data from other varieties of Arabic, such as the case with some tweets which are in different varieties of Arabic, I made sure as a native speaker to filter any sentences that are not in ECA so as to make sure the data is as representative of ECA as possible. Section 1.3 below discusses the process of data collection and building the corpus in more details.

---

[4]Both of *ha* and *ħa* are free variants of the future tense prefix, for example both *haktib* and *ħaktib* mean "I willl write".

1.3   The corpus

Corpus linguistics can be defined as "dealing with some set of machine-readable texts which is deemed an appropriate basis on which to study a specific set of research questions. The set of texts or *corpus* dealt with is usually of a size which defies analysis by hand and eye alone within any reasonable timeframe. It is the large scale of the data used that explains the use of machine-readable text" (McEnery and Hardie, 2012, p.1). This property of enabling the collection and searching rapidly through a big set of data is the main advantage in using a corpus to answer a specific research question. And this is exactly why this study is based on corpus data, so as to ensure that the data on which the analysis is based represents the actual use of the verbal predicates investigated in the dialect, as accurately as possible. It also added a wealth of different structures that raised some interesting questions about the data, which was better than dealing only on my knowledge as a native speaker of the dialect. A note to make here is that the current study is a corpus-based study and not a corpus-driven study, where the corpus data is used merely to explore and describe how the verbs investigated are used in the dialect, and therefore building the corpus was only a method serving this aim rather than building the corpus as an aim on its own.

"Arabic can still be considered a relatively resource poor language when compared to other languages such as English, and having access to freely available corpora will definitely improve the current Arabic NLP technologies" (Zaghouani, 2014, p.26). This issue is clearly shown in the study of dialects of Arabic, where the only freely available corpora for dialectal Arabic are the 'Tunisian Dialect Corpus' and the 'Arabic Multi Dialect text Corpora' that includes 2 million words collected from webpages representing Gulf, Levantive, North Africa and Egyptian Arabic (Zaghouani, 2014). The difficulty in finding a corpus representative of ECA as the source for my data led to the need to build one. Using the world wide web as the source of data is one of the data collection methods for building corpora known as *'Web as Corpus'*, it "takes as its starting point a massive collection of data that is ever-growing, and uses it for the study of

language" (McEnery and Hardie, 2012, p.7). The fact that the amount of data available to investigate on the internet is massive led to the need for finding a tool that can help collect the needed structures, save and make it easier to search through it all. And the tool used for that was Sketch Engine, which is described in section 1.3.1.

1.3.1   Sketch Engine

Sketch Engine[5] is an online corpus software interface that offers corpora in various languages, as well as tools for concordancing and creating corpora. It is used for Lexicography by various well-known dictionary publishers such as Oxford University Press and Cambridge University Press (Arts et al., 2014). It was also used by Arts et al. (2014) in creating a webcrawled corpus of 5.8-billion Modern Standard Arabic words called *'arTenTen'*. Sketch engine provides a tool called *'WebBootCaT'* which can create a corpus from relevant texts on the internet. The user can specify either some input words that define the topic to be searched for (also called *'seed words'*), or a list of URLs from which data is downloaded. The downloaded data is then "deduplicated, cleaned, spam text or non-text is eliminated to obtain high quality text material"[6]. Sketch engine then tokenises and tags the data and the corpus is then compiled and saved permanently on the user's profile.

I have used the *'WebBootCaT'* tool in Sketch Engine to build a corpus of ECA data that represents the current use of the dialect. The seed words used for the web search were the main verbal predicates investigated in the current study, where each verb was represented in its different forms including different inflections and in some cases different possible spellings. For example the auxiliary *kān* 'be' had the seed words *kān* 'he was', *kānit* 'she was', *konna* 'we were', *ħankūn* 'we will be', *kont* 'I was' and so on. Figure 1.1 shows an example screen shot of the WebBootCat

---

[5]https://www.sketchengine.co.uk

[6]https://www.sketchengine.co.uk/user-guide/user-manual/corpora/create-a-corpus-from-the-web/- Accessed on 4/6/2016.

and choosing the seed words for the auxiliary *kān*.

Figure 1.1: WebBootCaT Seed Words

1.3.2 Data collection

The resultant corpus size is around 5 million tokens, but the actual number of ECA words is less than this number, as it was manually filtered for code switching with MSA structures, as well as English words and numbers which were all removed. The corpus included written text from twitter as well as blogs which were validated by the researcher as instances of the dialect and were entered into the sktech engine tool. Each of the verbs was then searched for through the corpus in all its inflected forms for different person, gender and number inflections as well as temporal markings, this was intended to get the most possible number of outcome sentences. Sketch Engine provided the concordance for each sentence, which was useful in checking the context in which the sentence occured and making sure it is colloquial. All the sentences were then saved and the structures in which the verbal predicate occured were analyzed. Verbs were

searched for through the corpus using Arabic orthography as this is how the online texts are written, however I have then represented those examples in IPA symbols for the purposes of the current study. Figure 1.2 shows a screen shot for the search for the phasal verb *ʔaʕad* 'go on' and some example sentences.

Figure 1.2: Sample Sentences

Depending on the world wide web as the source of data has its problems, where the web is a mixed medium for both edited and non-edited text, and therefore some of its contents may contain errors of different sorts (McEnery and Hardie, 2012). Most of the errors found in the corpus were related to orthography issues, where sometimes the prefix *bi-* was written separated from the verb by a space (although the letter *b* in Arabic is always connected to other letters of the same word). This was remedied easily by searching for the seed words in this written form (with the *b* unattached) so as to get the biggest possible number of sentences. Other problems that I have encountered while collecting the data are caused by a word having different spellings, or cases of homography where the same written word form has two different meanings. Examples are:

- Lack of diacritisation led to the search for the phasal verb *fidil* 'remain.PV.3SGM' to get also results for the noun *fadl* 'favour'. This is shown in figure 1.3, where although having the same word form in orthography, only the first sentence represents the phasal verb, while the rest of the sentences represent the noun 'favour'.

Figure 1.3: Issues in Orthography

- Different spellings are sometimes available for some forms, such as the modal forms *ʕāwiz* and *ʕāyiz*, where both are free variations for the modal that means 'want'.

- The future *ħa-* prefix is also sometimes written as *ha-*, therefore any verb in the future tense had two seed words, for example *ħakūn* and *hakūn* both mean 'I will be'. Examples for these seed words are shown in figure 1.1.

These cases, in addition to instances of other dialects of Arabic occuring in the data were manually filtered based on my knowledge as a native speaker of ECA, where non-relevant results were removed from the data. Another choice made was to include only sentences where the verbal predicate (auxiliary, causative, phasal verb or modal) was followed by another verb, refered to throughout the study as the lexical verb to differentiate it from the main verbal predicate. This choice is motivated by the interest in investigating the question of the non-finite verb in ECA, where in most cases these verbal predicates were followed by verbs only in the bare imperfective form while other verbal predicates allowed both the bare form and other finite verb forms, examples are 5 where the phasal verb *baʔa* 'become' is followed by a bi-prefixed verb, while in example 6 the modal *yinfaʕ* 'can' allows only bare imperfective verb form:

(5) baʔa   bi-yemʃi   zay el-batrīq
    become.PV.3SGM BI-walk.IPFV.3SGM like the-penguin

Lit: He became (such that) he walks like a penguin

He walks like a penguin

(6) w   dah yinfaʕ   yitʕimil   fi zarf ḫamas sinīn
    and this can.IPFV.3SGM do.PASS.IPFV.3SGM in time five   years

And this can be done in five years time

It is also the case that the auxiliary *kān* expresses compound tense only when it is followed by a lexical verb, and since the main objective of describing this auxiliary is to account for tense and aspect in ECA, only sentences having the auxiliary followed by lexical verb were included. As for phasal verbs, there were many cases where the same verb form was used as a phasal verb in some cases and as a main (lexical) verb in others, such as the case with the phasal verb *ʔaʕad* where the phasal verb means 'go on', while the lexical use means 'sit'. However the phasal verb occurs preceding a bare form verb (as in 8 below) while the lexical meaning occurs preceding an NP or PP (as in 7 below).

(7) ʔaʕadt   ʕala el-kanaba el-gedīda
    sit.PV.1SG on   the-sofa   the-new

I sat on the new sofa

(8) ʔaʕadt   ʔazākir   fi el-beit   lḥad nos el-lēl
    go.on.PV.1SG study.IPFV.1SG in the-house until half the-night

I kept studying in the house until midnight

More examples for phasal verbs used in their lexical meaning are shown in section 4.2.3. Finally, investigating the case where two verbs are following each other and both share one or more arguments gives interesting insights on the nature of the control relation between the two verbs, which is important for analyzing these verbs in the current study. The issue of the bare imperfective verb form in ECA as a non-finite verb form, and the literature review in this respect is covered in the next section, 1.4.

1.4   Non-finite form

Verbs in ECA can occur in a number of forms expressing various markings for morhposyntactic tense as well as grammatical aspect[7], these are:

- Perfective, as in:

  (9) ʔaḥmad katab   el-gawāb
      Ahmed write.PV.3SGM the-letter
      Ahmed wrote the letter

- bi-prefixed Imperfective, as in:

  (10) ʔaḥmad bi-yekteb   el-gawāb
       Ahmed BI-write.IPFV.3SGM the-letter
       Ahmed is writing the letter

- ḥa-prefixed Imperfective, as in:

  (11) ʔaḥmad ḥa-yekteb   el-gawāb
       Ahmed FUT-write.IPFV.3SGM the-letter
       Ahmed will write the letter

- Bare Imperfective, as in:

  (12) ʔaḥmad ʕāwez   yekteb   el-gawāb
       Ahmed want.PTCP.SGM write.IPFV.3SGM the-letter
       Ahmed wants to write the letter

Note that the verbal complement in example 12 occured in the bare Imperfective form in an embedded clause after the modal ʕāwez 'want'. The embedded lexical verb occured in the bare Imperfective form in a lot of instances of the data, and in some cases it was the only form possible for the verbal complement of certain predicates. This raises the question of whether this bare Imperfective constitutes the non-finite form of verbs in ECA, which is a problematic

---

[7]These forms and their morphosyntactic tense markings are discussed in further details in section 2.3.

issue here as most grammars of ECA refer to this form simply as the 'bare form' which is defined as an "imperfect verb occuring with no aspect prefix" (Abdel-Massih et al., 1979, p.275). Description of this form is mainly concerned with the fact that it indicates subjunctive meaning[8] (Abdel-Massih et al., 1979), or simply stating it as the 'imperfective' (Hanna, 1967), while the notion of finiteness is often avoided.

In her study of Finiteness in Jordanian Arabic, Al-Aqarbeh (2011, p.233) states that "the review of the literature on finiteness in Arabic indicates that research in this area is limited. Among the few studies that exist, there is no consensus on the properties of finiteness in Arabic." The consensus among previous studies is that morphologically all verb forms inflect for agreement and tense/aspect and therefore are all finite, while syntactically only verbs that raise to TP are finite while others are non-finite (Al-Aqarbeh, 2011). And since Al-Aqarbeh (2011)'s study assumed that Jordanian Arabic verbs do not encode tense distinctions but only aspect, therefore the study concluded that the notion of finiteness does not apply to Jordanian Arabic. However, the current study assumes that verbs inflect for both tense and aspect, and therefore needs to investigate what a non-finite verb form is in ECA.

Finite verbs are also defined as having the properties of: "(a) the ability to license a referential subject (usually reduced to the ability to license Nominative), and (b) the ability to have a full or absolute Tense interpretation" (Fehri, 2012, p.243). With respect to Arabic, Fehri (2012) assumes that all verb forms are morphologically finite as they inflect for features such as tense/mood and subject, while there are some other semantic and syntactic criteria that can be used to identify whether the clause is finite. Therefore he concludes that "There are no real infinitives, and no participles, but only finite forms which are syntactically ambiguous: they are used in contexts where non-finite forms would occur in Indo-European, in addition to finite contexts." (Fehri, 2012, p.244)

---

[8]The notion 'subjunctive' is not considered here in more semantic details, but it is rather used to show that these forms are not used as the main verbs in indicative sentences.

Sells (2007) differentiates between four different uses for the term finite, that of a value feature of the verb form, another is for the grammatical property of the clause, and the pattern for behaviour of the verb carrying this property such as negation and agreement, and another for the use of the clause such as assertion or subjunctive. This is useful here as it shows that the morphological form is not the only criteria for defining a form as 'finite'. These views imply a distinction between form and function, which seems to be the most appropriate way to deal with the issue of finiteness in Arabic, where although the form of the verb agrees with the subject as finite verbs do, it does not function as a finite verb.

This is the position adopted by Hallman (2015) where he presents evidence that "supports the view of Arabic verb morphology in which the imperfective form marks neither an aspectual nor a temporal category. Rather, it is the default form of the verb that occurs when the verb is not marked perfective, and therefore corresponds both morphologically and semantically to the English non-finite verb form" (Hallman, 2015, p.129). This is based on evidence such as that in some modal contexts, the imperfective verb carries only the lexical meaning of the verb, while it does not denote any temporal or aspectual information related to the imperfective morphology. This leads to the conclusion that Arabic is a language where agreement is possible on non-finite verbs, such as the case with Portuguese, also refered to as an 'inflected infinitive' language (Hallman, 2015).

In Jelinek (1981)'s account of Egyptian Arabic, non-finite embedded clauses are described as ones that employ a verb that is inflected in the subjunctive IMPF (imperfective) which lacks any bi- or Ha- prefix, which corresponds to the bare imperfective accounted for in the current study. She states that: "There are two non-finite verb paradigms in Egyptian Arabic: the imperative and the subjunctive (traditionally called the 'Imperfect without Prefix')" (Jelinek, 1981, p.13).

As is apparent from the discussion of finiteness above, the notion needs a more thorough investigation that might be beyond the focus of the current study. And therefore the notion of "bare imperfective" will be maintained here and the details of the notion of finiteness will be avoided for the time being. However, this form is considered in the current study to be non-finite as it lacks the morphological inflection for tense and aspect as opposed to other finite verb forms, and the clause headed by this verb form also lacks the grammatical tense feature and is used in subjunctive uses as mentioned in the previous studies on Egyptian Arabic. This assumption also follows Jelinek (1981) describing the bare imperfective verb form in ECA as non-finite, as well as Hallman (2015) assuming that the imperfective in Arabic is non-finite.

1.5   Defining Control relation within LFG

The definition of a control relation in LFG thoery is dependent upon the specification of the structure as either mono or bi-clausal. A mono-clausal structure is one in which the grammatical functional structure is that of a single predicate (a nuclear PRED) with only one subject (Butt, 1995). Serial verbs are an example of mono-clausal structures, where both verbs in the serial verb construction share the same tense, aspect, mood and polarity (Aikhenvald and Dixon, 2006). Another example of mono-clausal construction is complex predicates, where both verbs behave as a single predicate in terms of agreement and anaphora, and they are both controlled with the same subject, such as the case with the permissive in Urdu (Butt, 1995). It is important to note here that these are properties of the grammatical function of the structure rather than the phrase structure, where in LFG, "it is possible to show that complex predicates must be simple with respect to grammatical functions (relations), but may be either simple or complex with regard to c-(onstituency) structure (phrase structure)" (Butt, 1995, p.3). The term 'bi-clausal' on the other hand describes the case where two verbs occur in a sequence but where each of them represents a unique predicate, with its own properties in terms of agreement,

negation and temporal marking, and where each could have its own subject. It is also possible for both verbs to share the same subject, which is described as a 'control relation'.

Control implies the presence of several properties such as: clause boundary, double assignment of grammatical relation between the controller in matrix clause and the controllee (seen as a missing/implicit argument) in the subordinate clause (Kroeger, 2004). It is also correlated with the presence of a non-finite verb with no overt SUBJ in the subordinate clause (Falk, 2001), however, this characterstic is debatable, which is further discussed in section 1.5.2.

Control can be Lexically or Syntactically determined, where lexically determined control occurs with verbs which require specification of an XCOMP in its lexical entry, and they are called 'Control predicates'. The kind of control relation is represented in the lexical entry through a 'Control Equation' (Kroeger, 2004). This is the kind of control which is tackled in the current study, where each verb will be investigated as to the kinds of complements it takes and how this can be described in terms of the control relation. Characteristics of each kind of control relations is represented in section1.5.1 to maintain a clear definition of each, as this will form the basis for investigation of the kind of control relation for each verb throughout the current study. While section 1.5.2 is concerned with discussion of finite control, which is important for the description of control relations in some of the control verbs investigated in the current study.

Another important feature of defining the syntactic structure of the analysed verbs is deciding whether the main verb occurs in *I* or *V*, which will be crucial for representing the constituent structure of the sentence. Within LFG theory, *I* is a functional category which is usually occupied by a finite verbal element that functions as the (categorical) head of the sentence (*IP*). In English for example, *I* is "the category of temporal/aspectual finite auxiliary and modal verbs" (Bresnan, 2001, p.99). *IP* is the projection of *I* which often corresponds to a sentence in many languages, where in English for example, "the tensed auxiliary verb appears in I, and the rest of the verb complex appears inside the VP" (Dalrymple, 2001, p.61).

Each of the verbs analysed in the current study shall be compared against these definitions of control relations and functional categories, where these tests are to be applied to the structures in which the verb occurs with the attempt to answer the following questions:

- Is the verb a control predicate?

- Is it Anaphoric or Functional control?

- How would the f-structure of the verb look like?

- What is the control equation of that verb's lexical entry (if found)?

1.5.1   Kinds of Control relations within LFG

In her account of the principles of LFG theory, Bresnan (1982) defines the control relation as "a relation of referential dependence between an un-expressed subject (the controlled element) and an expressed or unexpressed constituent (the controller); the referential properties of the controlled element, including possibly the property of having no reference at all, are determined by those of the controller" (Bresnan, 1982, p.317). One of the main factors in differentiating between functional and anaphoric control is the relation between the controller and the controlled element (controllee), where functional control implies the complete identity between the f-structures of the controller and the controllee, while anaphoric control implies the mere identity of reference between them (Bresnan, 1982).

Other properties in which functional control can be distinguished from anaphoric control are related to the controllee, where in functional control the controllee must always have a specific grammatical function in the subordinate clause in a specific language (SUBJ in English) and it can not be overt, where no independent subject argument is allowed in the subordinate clause

(Kroeger, 2004). In anaphoric control, however, the controllee can be any direct argument other than the subject and it can be expressed overtly as an independent noun or pronoun in the subordinate clause (Kroeger, 2004). In terms of their complements, a functional control predicate takes an XCOMP as its argument while an anaphoric control predicate takes a COMP as its argument (Falk, 2001). Examples for functional control predicates in English are the verbs 'try'[9] and 'presuade', while examples for anaphoric control predicates are the verbs 'agree' and 'want'.

Another pair of terms that define different kinds of control relations are Equi and Raising. Different theories of transformational grammar have different analysis for raising, while in LFG theory raising is categorised as a case of functional control (Falk, 2001). Raising verbs have a number of characteristics such as that they do not select their subjects, and that idiomatic expressions retain their meanings under raising, as well as the fact that "when the embedded verb of a raising predicate is passivised with its arguments switched, such that the embedded object becomes the matrix subject, the meaning of the passive sentence is similar to that of the active sentence" (Ademola-Adeoye, 2011, p.35). Raising predicates are also chracterised by their ability to take dummy subjects or expletives (such as it, there), where there is no sematic role assigned to the subject of a raising predicate (also called a non-thematic subject) (Kroeger, 2004). Raising can be either to subject or to object, and the complement of the raising predicate is an open function (XCOMP) (Falk, 2001). Examples for raising predicates in English are the verbs 'seem' and 'begin' (Kroeger, 2004).

Equi constructions, on the other hand, are "control constructions in which the controller bears a thematic relation to its verb" (Falk, 2001, p.118). Unlike the case with raising, the subject of Equi constructions can not be an expletive or an idiomatic expression, where the subject is assigned a semantic role and therefore it can undergo some selectional restrictions (Kroeger, 2004). Despite categorising Equi in his textbook for LFG as a case of anaphoric control, Falk

---

[9]According to Falk (2001) , while Dalrymple (2001) lists 'try' as an example for anaphoric control, as shown in the discussion of Equi below.

(2001) states that Equi can represent either anaphoric control or functional control. A Functional control of the type Equi construction takes a thematic subject as well as an XCOMP complement, examples in English are the verbs 'plan' and 'agree' (Kroeger, 2004). In her account of the principles of LFG theory, Dalrymple (2001) presents evidence that the English verb 'try' represents a case of obligatory anaphoric control, she also argues that "examination of a larger range of languages shows that more variation is found: some languages have two types of equi verbs, some specifying anaphoric control and some specifying functional control" (Dalrymple, 2001, p.328).

1.5.2   Finite Control

Control is often correlated with the concept of finiteness in the literature of sytnactic studies, where "the basic view is maintained that obligatory control (OC)- necessary coreference between a matrix argument and an embedded null subject- is characteristic of nonfinite (or infinitival) clauses, that is, clauses uninflected for tense and agreement" (Landau, 2004, p.811). However, some of the data investigated in the current study seems to contradict with this assumption, where for example the phasal verbs *rigiʕ* 'return', *baʔa* 'become', *fidil* 'remain' and *ʔaʕad* 'go.on' (discussed in section 4.3.1) can take both finite and non-finite complements, and a control relation is attested between the phasal verb and its verbal complement. These phasal verbs might then be candidates for the phenomenon of 'finite control'.

Finite control can be defined as the case of obligatory coreference between an overt matrix argument and an embedded null subject of a finite clause complement. This has been attested in several languages such as Korean, Greek, Spanish, Finnish, Persian and Balkan languages (Lee, 2009). In her cross-linguistic study of finite control, Ademola-Adeoye (2011, p.40) refers to finite subject-to-subject raising as 'Hyperraising' and defines it as: "raising of a DP from the embedded subject position of a finite complement to the matrix subject position". She presents data from different languages which exhibit raising out of finite complements, stating that in

some languages hyperraising is correlated with subjunctive complements (such as the case in Greek and Persian), where subjunctive complements are assumed to be 'defective' in some sense, which makes them comparable to non-finite raising contructions in a language such as English (Ademola-Adeoye, 2011). While some other languages exhibit hyperraising out of finite indicative complements (such as the case in Turkish, Finnish, Bantu and Moroccan Arabic), examples from Morrocan Arabic for the raising predicate *ttshab-li* 'seemed to me' are cited as follows:

(13) a. ttshab-li [beli zat mmi]
seemed.3SGM-to-1SG COMP came.3SG.FEM mother.1SG

It seeemed to me that my mother came [(Ademola-Adeoye, 2011, p.78)[10]]

b. Ttshab-et-li mmi [beli zat t]
seemed.3SGF-to-1SG mother COMP came.3SG.FEM

It seeemed to me that my mother came [(Ademola-Adeoye, 2011, p.78)]

Where in both cases *ttshab-li* takes a finite complement introduced by the complementiser *beli*, and there is a subject coreference between the matrix clause and the embedded finite clause. The position of *mmi* 'mother' preceding the complementiser indicates a clear case of hyperraising (Ademola-Adeoye, 2011).

Landau (2004) accounts for finite control in Hebrew which can have control into finite as well as non-finite complements. He goes on to discuss some characteristics of finite control in Hebrew, such as taking complements in the future tense only, as well as a null subject that is obligatorily controlled by a matrix argument. The matrix verb defines the kind of control relation in Hebrew, where for example commisive and directive verbs allow finite control, while other kinds of verbs such as factives do not allow finite control (Landau, 2004). Finite control is investigated against data for some of the verbal predicates in the current study which seem to exhibit raising into finite as well as non-finite complements.

---

[10]The Moroccan Arabic examples are cited from Ura, H (1994). Varieties of raising and their implications in the theory of case and agreement. In MIT Occasional Papers in Linguistics (Vol.7). Massachusetts: MIT.

1.6   Outline

This thesis is devoted to the detailed description and analysis of four different predicates inves-
tigating the complementation patterns of each. Chapter 2 presents a detailed account for tense
and aspect in ECA, and the function of the auxiliary *kān* in expressing compound tense forms.
It attempts to answer questions such as whether any form of *kān* can be combined with any
form of the following lexical verb or are there any restrictions on their co-occurrence? and what
are the possible tense and aspectual features each of these components can represent? and how
to represent these combinations in the LFG framework in order to reach an analysis for the
auxiliary *kān* in compound tense forms?

Chapter 3 investigates the different forms of complements of the causative verb *ḫalla* in ECA
with the aim of providing an analysis for this structure within the LFG syntactic theory, where
the debate arises as to whether the causative verb and the lexical verb form one complex pred-
icate or two predicates, a question which is not yet tackled in the literature for ECA. In order
to reach this aim, a thorough investigation of the behaviour of this structure is done, answering
questions such as what kinds of verbal complements does this causative verb take? what are
the possible tense and aspect combinations of these complements? how do these constructions
behave under negation? and the behaviour of the arguments of both verbs and whether they
share the same subject or object, and what does this say about the control relation between the
causative and the following lexical verb.

Chapter 4 investigates the syntactic behaviour of phasal verbs in ECA, finding out what are the
possible complements these verbs can take and how to represent these structures syntactically
within the principles of LFG. The list of verbs include: *badaʔ* 'begin', *baʔa* 'become', *fiḍil* 're-
main', *liḥiʔ* 'catch', *ʔarrab* 'be.near', *battal* 'stop', *rigiʕ* 'return' and *ʔaʕad* 'go.on'. These verbs
are classified into three sets according to the different patterns of behaviours of each set of verbs,
the complements they take and syntactic behaviour of each. With respect to the syntactic anal-

ysis, the main question to be answered was the status of phasal verbs as complex predicates, or whether each of the phasal verb and the following lexical verb form a bi-clausal structure, and the kind of control relation between the two verbs within the principles of LFG syntactic theory.

Chapter 5 is concerned with describing the syntactic behaviour of modals in ECA, where they can be verbs, participles or nouns, with both inflecting and non-inflecting forms. The behaviour of modals has not been the focus of previous work on the syntax of ECA (with the notable exception of Azer (1980)). The modals investigated are: *yiʔdar* 'able', *yinfaʕ* 'can', *ʕāwiz* 'want', *lāzim* 'must', *darūri* 'should', *ʔel-mafrūd* 'the-ought-to', *nāwi* 'intend', *mosammem* 'insist', *nefsi* 'wish', *momken* 'possible', as well as *yimken* 'may'. These modals are categorised into three groups according to the structures in which they occur, which are: non-inflecting modals, inflecting modals, as well as yimken which is the only modal followed by finite form verbs. The chapter shows how members of each of these categories behave in terms of the types of complements they take as well as their ability to have an independent subject. These differences in patterns of behaviours lead to different analysis of each category, which is illustrated using LFG tools and represented in the lexical entry and f-structure of each.

Chapter 6 describes the XLE as a tool developed for testing LFG grammars of different languages, and how this is applied to the grammar developed for these verbal structures in ECA. It describes every section of the grammar file, and the list of grammatical sentences that can be analysed using this grammar and the ungrammatical sentences that are not allowed. It also mentions the issues in implementation and how I managed to overcome some of them while some others present a challenge to the grammar. Finally, Chapter 7 provides a conclusion and discussion of some of the main questions that were posed in the introduction and how the thesis managed to answer them, as well as some questions that are still open for future research.

Chapter 2

Aux *kān* - Tense and Aspect in E.C.A

Tense and aspect are considered by Tallerman (2011, p.41) to be "the most common morphosyntactic categories associated with verbs". But despite the great amount of studies investigating how such temporal relations such as tense and aspect are grammaticalized in language, many issues in that respect remain unresolved. One of which is the way in which morphosyntactic tenses are expressed in different languages, therefore grammatical theories accounting for tense must fulfill the tasks of both assigning verb forms to the appropriate morphosyntactic tenses, as well as explaining how expressions can be used to convey information about event time in languages without tense. This study aims to define the concepts of tense and aspect, and to investigate how they are represented in Egyptian Colloquial Arabic (ECA), as well as providing a preliminary sketch for the grammatical expression of temporal and aspectual notions in this dialect.

The current chapter is divided as follows: Section 2.1 defines the notion of tense including specification of different tenses as well as the Reichenbach's system which is adopted in the current study, followed by section 2.2 which defines the notion of grammatical aspect necessary to define tense and aspect as they are expressed on verbs in ECA. Section 2.3 describes the main verb forms in ECA regarding the marking of grammatical tense and aspect on each, as well as

description of two peculiar properties which are the prefix *bi* and the auxiliary *kān*. Section 2.4 provides a literature review of the main accounts of tense and aspect, while section 2.4.1 discusses some issues that are correlated with the study of tense and aspect in ECA, which the current study attempts to describe and (possibly) resolve. Section 2.5 presents the data on which the current study is based, divided into simple and compound tenses with description of each set of sentences and the verb forms and temporal interpretation of each. Section 2.6 presents a discussion of the data including a list of all the possible and impossible combinations between the auxiliary and the verb, as well as comparison between findings of the current study and conclusions drawn in other studies on tense and aspect in ECA. Section 2.7 provides the analysis for the structures described, including my account of tense and aspect in ECA and the function of the *bi* prefix and auxiliary *kān*, as well as the lexical entries, f-structures as well as c-structures of the verbal combinations. Finally, section 2.8 summarises the main points tackled in the chapter and provides the conclusions.

2.1   Tense

Tense is defined by Comrie (1985, p.9) as the "grammaticalized expression of location in time". Grammaticalization (grammatical representation) of tense varies from one language to another in the ways used to convey these temporal contrasts. Van Lambalgen and Hamm (2005) state that there are two properties that can define a language as having grammaticalized tense, the first is that it distinguishes between past and non-past time reference, and the second is to have an obligatory expression of tense that is morphologically bound on verbs. There are a few languages lacking those properties, which are said to be non-tensed, and the temporal location is lexically instead of grammatically expressed (Van Lambalgen and Hamm, 2005). Burmese is an example of such a language which lacks the grammatical category of tense, but rather has some particles signalling features of polarity, mood, and aspect (Comrie, 2009).

Despite these differences among languages, the verb is the most widely used tool to express

tense contrasts, as stated by Binnick (1991, p.3) "To this day the verb is thought of as a 'time-word'..It is that part of speech which is concerned with distinctions of time, that is, with tense." Therefore, the study of grammaticalized tense is mainly concerned with the ways tense is marked on verbs, and the temporal reference of these markings. This is achieved through the distinction between the notion 'time reference' and the morpho-syntactic notion 'tense', which is necessary to show the difference between meanings and functions of the various categories used to express tense (Eisele, 1990b). An example is the present perfect in English as in "John has left London", although this refers to an action in the past, it is often used to show some kind of correlation to the present, moreover, it is not possible to say "John has left London yesterday" (Klein, 1992)[1]. This raises the issue of distinguishing between different tenses and specifying the interpretations of each, which is the focus of the following section.

2.1.1 Different tenses

The world is experienced as having three times, which are: the present, that is experienced now, the past which has gone by, and the future which is to happen next, and tenses are therefore the signs used to express these times (Binnick, 1991). Languages generally use special marking for each of these tenses, as the case where speakers of English can distinguish between three times corresponding to three tenses, which are past, present and future. This is stated by Binnick (1991, p.4) as: "The three times are thought of as segments of an indefinitely long line passing through the point of the present."

The previously mentioned example of speakers of English, although it seems a simple generalization, can be used to show the problems often associated with the study of tense systems in languages. This is clear through the fact that in terms of morphological structure, English verbs can be inflected only for present and past tenses, while the future tense lacks any kind

---

[1]This issue is discussed in more details in section 2.2 where the issue of the 'perfect' aspect is tackled.

of verbal morphology equivalent to those of present and past tenses, and hence, English verbs from this perspective can be assumed to have only two tenses (Tallerman, 2011).

Another issue is the definition of meanings of the forms used to express tense, where the same verb form can be used to denote events at various points in time. This issue has also been addressed as the difference between the term 'tense', referring to morphological marking on language structures (mostly verbs), and the term 'time reference' which refers to the role played by each of these forms in defining the location in time of the action, event or state described by the verb (Brustad, 2000).

This study will not be concerned with the first issue, that of defining the number of tenses. This is because despite the fact that grammarians assumed different number of tenses in languages, from dozens to only a couple, the three-tense model seems to be the most satisfying and widely used across studies of tense (Binnick, 1991). The second issue will be the main focus of this study, where the following sections will be devoted to describing the morphosyntatic tenses that are expressed by verbs of Egyptian Arabic, starting by definition of the Reichenbach system which is used in the current study to distinguish between different tenses.

2.1.2   Reichenbach's system

Reichenbach's system for the definition of tenses is used in various frameworks as the basis for the linguistic analysis of tense and time reference (Eisele, 1990b). The most important property in this system is that it is highly referential, based on a theory of points in time rather than relying on point of view which is mostly subjective and often problematic in cases where different tenses might be used to express different viewpoints for the same event. As stated by Binnick (1991, p.111) ".Reichenbach defines tenses as relations holding between S ('now', the time of speech act), E (the time of the event or state of affairs), and a third, more abstract time R (the reference point, the temporal 'point of view')."

Therefore, this system defines tenses in terms of E,R, S and the relations that hold between them such as precedes, follows and coincides with. An example of the possible combinations of the simple tense is "if E precedes R it necessarily precedes S, and a past tense results; similarly, if E follows R, it follows S, and the result is the future. If all three coincide, the present results" (Binnick, 1991, p.112). These notions can be further used in establishing the distinction between simple (absolute) and compound (relative-absolute) and complex (relative) tenses which vary in the way E,R and S are expressed and related to each other (AlKhawalda, 1997). Relative tenses are defined through the relation between R and S, while in simple tenses R coincides with S (Binnick, 1991). Both simple and compound tenses will be tackled in the coming sections, where simple tense is marked on the lexical verb itself while compound tense is formed through combining the auxiliary with the lexical verb to assign these points in time, which is further elaborated in the following sections.

It is often proposed to define the temporal relation in terms of a pair of orderings, one representing the relation between E and R and another representing the relation between R and S rather than a strict order of the three parameters with each other (Binnick, 1991). For example, defining future perfect as the case where R follows S (to represent future) and E precedes R (representing perfect), without specifying the relation between E and S. Defining temporal relations as a pair of relations has the advantage of separating between tense and aspect. This is clear through the following observations (Binnick, 1991):

- Tense is concerned with the relation between R and S. Where R precedes S in past tense, R and S coincide in the present tense, and R follows S in the future tense.

- Aspect is concerned with the relation between R and E. Where E precedes R in anterior tenses, E and R coincide in simple tenses, and E follows R in all posterior tenses[2] (Binnick,

---

[2]According to (Binnick, 1991), Reichenbach uses the term "anterior" to refer to cases where E precedes S (corresponding in traditional notions to prefectivity) while "posterior" refers to cases where E follows S.

1991).

Despite the fact that this system provides a useful tool for description of possible temporal relations in natural languages, it does not in itself give a full account of the tense/aspect system in a certain language. This issue is the main concern of the current study, where such tools will be used in an attempt to reach a comprehensive account of the tense and aspect system in ECA, and the ways tense and aspect are represented on both the auxiliary and lexical verbs as well as the meanings associated with each use. An important note to make here is that the term 'aspect' can be used to refer to either lexical or grammatical aspect, while its use in the current study is restricted to grammatical aspect only, where the semantic aspectual distinctions are expressed using grammatical means (Comrie, 1976). This is the focus of the coming section (2.2) which defines grammatical aspect and provides some examples from the literature for the different aspectual distinctions.

2.2   Aspect

Comrie (1976, p.3) defines aspect as "different ways of viewing the internal temporal constituency of a situation." As Binnick (1991) describes it, the term 'aspect' is etymologically related to the root 'spect-' which corresponds to the terms 'view, see'. This is related to the fact that different aspects correspond to different views about the same situation. Example: 'I visited Aunt Martha and saw your picture' is in perfective aspect, viewing the situation as a completed whole, while 'I was visiting Aunt Martha and saw your picture' is in imperfective, viewing the situation as an ongoing action (Binnick, 1991). The distinction between perfective and imperfective aspects is also captured in Comrie (1976, p.16) as: "perfectivity indicates the view of the situation as a single whole, without distinction of the various separate phases that make up that situation; while the imperfective pays essential attention to the internal structure of the situation."

Another basis for differentiation between these two aspects is the use, where there is a set of uses for each across various languages as defined by Binnick (1991, p.156): "The imperfect(ive) has continual, habitual, and generic uses in many languages, while the perfect(ive) has punctual, iterative, and resultative uses." These relatively agreed-upon uses will be the basis of the treatment of grammatical aspect in the current study. An additional way of distinction between these aspects is finding equivalent terms referring to each, where the perfective aspect can also be referred to as non-progressive, punctuative, dynamic, completive, transitory; which differs from the imperfective referred to as progressive, habitual, static and permanent (Binnick, 1991). This study is not concerned with comparisons between these methods of defining aspects, but will rather refer to any of these methods that prove useful in describing how aspects are marked on E.C.A verbs and how those markings are used to refer to different meanings.

Another more controversial term related to aspect is the 'perfect', which is described in English as a "complex morpho-syntactic construction made of an auxiliary ("have", "be"), followed by a past participle, as in "Jamie has eaten all the chocolate biscuits"(Ritz, 2012, p.882). It can be confused with the term 'perfective', however in the current study 'perfective' refers to the completive aspect, in contrast with the imperfective aspect, while 'perfect' will be used only to denote the interpretation of some forms in the examples such as those corresponding to the English 'past perfect'. Defining the perfect whether as a tense or aspect, and identifying its meanings across languages has been the subject of debate among scholars, due to the various range of meanings and uses associated with these forms, especially the English present perfect, which is more constrained than the past and future perfect forms (Ritz, 2012). An example of the English 'present perfect puzzle' is provided by Klein (1992) where the sentence "John has left his wife" clearly refers to an event in the past, however it seems to be somehow relevant to the present, moreover, it is impossible to add past time adverbials to the sentence as in "*John has left his wife last year". This leads to Klein (2009)'s view of the perfect as a separate dimen-

sion of temporal expression which is distinguished from the usual tense and aspectual markings (past, present, future, prefective and imperfective), stating that "the perfect forms constitute a tense system on their own, based on auxiliary marking" Klein (2009, p.54). The main features associated with the 'perfect' are its relevance to the current situation which can be described as "past in the scope of the present" (Ritz, 2012, p.888) as well as its incompatibility with past time adverbials unless they extend to the moment of speech (Ritz, 2012).

The discussion of the puzzle of the present perfect in English raises the question of whether ECA has some kind of morphosyntactic form dedicated for the 'perfect'? Trying to answer this question is quite tricky, as the present perfect generally refers to a different range of situations that can be denoted using a wide variety of verb forms in ECA, in addition to the fact that some of the meanings do not hold for ECA sentences. Following is a brief illustration of the different uses of the 'perfect' as mentioned by (Comrie, 1976) and its possible equivalents in ECA[3]:

- The perfect of result, where the event has relevance to utterance time 'now' (John has arrived). This in ECA can be expressed using the past tense verb form which will be ambiguous between past tense and present perfect in English:

  (1) jon   wasal
       John arrive.PV.3SGM
       John arrived/John has arrived

- Experiental perfect, where the event is less relevant to utterance time (Mary has run). This in ECA will also be represented by the past tense marking, as in:

  (2) mari giret
       Mary run.PV.3SGM
       Mary ran/Mary has run

---

[3]These sentences are merely translations of the examples in Comrie (1976).

- Perfect of persistent situation, where the occurrence of the event is rather implied, and in some languages as German the present tense is used to convey this meaning (I've been waiting for hours). This in ECA will be expressed using a participle and an adverb to show the interval through which the situation holds:

  (3) ʔana mistanni   men sāʕa
      I   wait.PTCP.SGM from hour
      I've been waiting for an hour

- Perfect of recent past, where current relevance is related to recency (I have seen some old classmates recently). This will also be expressed using the participle as well as an adverb to show the recent past (hot topic as described in (Ritz, 2012)):

  (4) ʔana lessa me2ābel   soħāb-i
      I   just meet.PTCP.SGM friends-my
      I've just met my friends

Examples 1 and 2 express the present perfect interpretation by the verb in the perfective form, while examples 3 and 4 express the present perfect using participles. Although these examples are in no way exhaustive of the present perfect in ECA, they show that there does not seem to be a specific form to express the perfect in ECA. The variability of the forms used to express the prefect in ECA, and the fact that they do not express exactly the same distinctions as those in the English examples, as well as the lack of use of the auxiliary *kān* in all of these forms makes it irrelevant to the scope of the current study, and therefore the only aspects that shall be investigated here are the perfective and imperfective.

In order to avoid confusion, a distinction is made between the morpho-syntactic form and the interpretation in discussing these aspects, where the terms "Imperfective" and "Perfective" (with capital initial letters) are used throughout to refer to the morphological form[4], while the terms "imperfective" and "perfective" aspects (no capitals) are used to refer to the interpretation of

---

[4]Which might be used to denote either a tense or aspect.

the form. It is important here to clarify that these terms are used to refer only to morphological forms and how they are used to express graammatical aspect. While Arabic does not have an imperfective/perfective distinction that plays out syntactically and semantically in terms of what is refered to as "viewpoint aspect", where some languages offer a choice between perfective, imperfective and neutral aspect, while other languages have no overt viewpoint morphemes and therefore do not offer such a choice of viewpoint (Smith, 1991).

Having been through definitions of tense and aspect and some of the issues in the study of each, section 2.3 below presents the main morpho-syntactic forms in which tense and aspect can be expressed on verbs in ECA, along with a description of two grammatical features used in ECA in relation to tense and aspect that are relevant to the current study, these are the *bi-* prefix and the auxiliary *kān*.

2.3   Morphosyntactic tense on ECA verbs

There are various ways in which a verb can be inflected for tense/aspect in ECA. Table (2.1) below aims to show these different forms as they are refered to in the current study, along with an example of each and its meaning. Note that the meaning here is given informally though translational equivalents, with no characterisation of any tense or aspectual features, as this will be further tackled at the end of the current chapter.

Form	Example	Translation
Perfective	katab	he wrote
bi-prefixed	bi-yiktib	he writes/ is writing
ħa-prefixed	ħa-yiktib	he will write
Imperfective (Bare)	yiktib	write

Table 2.1: Verb forms in ECA

Note that the bi-prefixed form carries both meanings, the habitual and the progressive. This is

a property of this prefix in ECA that will be further investigated in section 2.3.1 below. Another note is that the Imperfective is also called Bare, this is due to it lacking the bi-prefix. This notion is defined in (Abdel-Massih et al., 1979) as: "A "bare form" is an imperfect verb occuring with no aspect prefix..The bare form usually indicates subjunctive meaning and results from deleting a redundant aspect prefix"(Abdel-Massih et al., 1979, 275). The term subjunctive is not further tackled in the current study, however it is used to show that the bare form occurs in contexts other than the indicative which makes it eligible to be the non-finite verb form in ECA. The uses of this form and its status as a non-finite verb form is discussed in more details in section 1.4. Following is a description of the bi-prefix in ECA.

2.3.1   bi-prefix

In a comprehensive study of Egyptian Arabic, Abdel-Massih et al. (1979) defines the bi-prefix as indicating viewpoint aspect on Imperfective verb forms, with the exact indication differing according to the verb. According to Abdel-Massih et al. (1979), the prefix bi- indicates habitual occurrence when added to stative verbs or verbs of change of location, while it indicates either continuity[5] or repetition on all other verbs. This difference is illustrated using the examples below cited in Abdel-Massih et al. (1979, p.268):

(5) bi-yifham   b-suʕūba
     BI-comprehend.IPFV.3SGM with-difficulty
     He comprehends with difficulty

Where the bi prefix on the stative verb (comprehend) gives a habitual interpretation, as opposed to:

(6) ʔuxt-i   b-tidris   dilwaʔti
     sister-my BI-study.IPFV.3SGM now
     My sister is studying now

---

[5]The term 'continuity' is used here to refer to the same concept as 'progressive', where it is defined in Abdel-Massih et al. (1979, p.268) as " 'continuity' being viewed as the feature of an action or an event in progress".

Where the verb here indicates "either continuity or repetitiness" Abdel-Massih et al. (1979, p.268). These examples show that the bi-prefix gives habitual interpretation on stative verbs and progressive interpretation on action verbs. This assumption is further verified in the current study through the analysis of the auxiliary *kān* and its combination with various verb forms.

In his study concerning the marking of aspect on verbs in ECA, Eisele (1990a) differentiates between the use of the bi-prefix on stative and non-stative verbs, where he indicates that "Stative verbs are characterised as having only a habitual reading in the bi-imperfect form, while other types of verbs may have both a habitual as well as a real present time reading" (Eisele, 1990a, p.196), this is shown in example 7 below quoted from (Eisele, 1990a) where the stative auxiliary verb *kān* in the bi-imperfective refers to a non-specific time reference which corresponds to the habitual reading:

(7) lamma ʔāgi,   bi-ykūn   bi-yiktib
     when   come.IPFV.1SGM BI-be.IPFV.3SGM BI-write.IPFV.3SGM
     When(ever) I come, he is (always,usually) writing

Defining the meaning associated with the present tense forms as either habitual or progressive is dependent on the interaction between three factors, which are: formal (grammatical) aspect, lexical aspect, as well as the pragmatics of present-time reference (Eisele, 1990b). Where he states that in ECA, "it is the interaction of the formal aspect of the present-tense form, the bi-imperfect, with the lexical aspect of the verb which determines whether the bi-imperfect may be used only with non-specific time reference (i.e., with only a habitual reading), or with both specific and non-specific time-reference (i.e., with both a habitual and a non-habitual reading)"(Eisele, 1990b, p.198). In other words, the fact that bi-prefixed stative verbs can only have the habitual reading is assumed to be the result of the contradiction between aspectual natures of a stative verb and a processive form (represented by bi-prefix), where a process entails a change of state, which is absent in the case of a stative verb, and therefore the bi-prefix can only mark the habitual reading on stative verbs(Eisele, 1990a). While in relation to its interpre-

tation with all other verbs, the 'bi-imperfect' form can have both a present continuous reading as in the English progressive, as well as a habitual reading as the English simple present (Eisele, 1990a).

In a dedicated study for the role of *bi-* prefix in Egyptian Colloquial Arabic, Tawakol (2008) defines the b- prefix[6] as an aspectual marker which is used as a prefix for imperfective verb forms in Egyptian Colloquial Arabic to refer to either habitual actions or an action which is currently being undertaken. The study was concerned only with verbs called BSFV (B-Semi-Fusha Verbs) which occur in code mixing between Standard and Egyptian Colloquial Arabic, occurring in what is referred to as 'ammiyat Al-Muthaqqafin'[7]. And therefore the forms of the verbs used is different from those used in the current study, including verbs such as *bi-nastakmel* 'we continue' for which the equivalent in ECA would be *bi-nkammel*. Despite this difference in verbs, the study reached a number of conclusions about the use of the bi-prefix that are considered relevant to the current analysis. One of which is that the *bi-* prefix is used as a tense marker on a variety of BSFV verbs in code-switching environments, but this is constrained by a set of rules that govern its appearance, and prohibits its use in certain cases. For example, the *bi-* prefix was not used when the verb was preceded by some Fusha structures such as a Fusha relative pronoun, interrogative or adverb. Conditions prohibiting the use of the prefix includes also cases where the verb is preceded by Fusha time adverbials, modals, verbs of beginning, particles contradicting habitual action, as well as where the verb carries a conditional or subjunctive mood. These findings are useful for the current analysis as they show the dominance of usage of *bi-* prefix on verb forms in ECA, that it even extends to some BSFV and is considered one of the main characteristics of this language variety (Semi-fusha) as mentioned in (Badawi, 2012). The restrictions on usage of this prefix are also in accordance with results of the current analysis, where absence of the bi- prefix is often correlated with a non-indicative use while its

---

[6]The *bi-* prefix mentioned in the current study is the same as the b- prefix in Tawakol (2008).

[7]This term is used by(Badawi, 2012) to refer to the variety of Arabic which is intermediate between Fusha (Modern Standard Arabic) and colloquial, which is used in some formal contexts by educated people. This variety includes a mixture of features from both standard and colloquial Arabic on various levels.

presence can mark habitual/progressive readings.

In her study of tense, mood and aspect in both M.S.A and E.C.A, Moshref (2012) indicates that the *bi-* prefix marks ECA imperfectives covering habitual, iterative or progressive aspects. The *bi-* prefix is assumed to be used in the indicative mood, while it is absent in nonfactual statements and jussive imperative moods. As to its interpretation, the *bi-* prefix is defined in Moshref (2012) as an aspectual marker which is attached to the imperfective verb form, indicated as PROG or HAB. On the other hand, she shows that "bi- prefix is dropped in several contexts, such as cases where the time reference is unspecific(..), or when its presence is redundant because it comes in co-ordination with a preceding *b-* or *ħa-*" prefixed verb, as in her example: (Moshref, 2012, p.95).

(8) bi-yesħa   el-sāʕa   sabʕa w   yuxrug   tamanya
    BI-wake.up.IPFV.3SGM the-hour seven and go.out.IPFV.3SGM eight

    He wakes up at seven o'clock and goes out at eight

The bare Imperfective verb[8] here does not denote any temporal information other than that of the bi-prefixed verb, where the whole sentence denotes a present habitual meaning, which raises a question about the wide scope of the bi-prefixed verb as the main predicate in the sentence and shows that the bare imperfective verb lacks any specific temporal interpretation. The bare imperfective is also mentioned in Moshref's study as expressing deontic modality, as in the example below (Moshref, 2012, p.126):

(9) lāzim yurūħ
    must go.IPFV.3SGM

    He must go

Therefore, the assumptions agreed upon among all of the previously mentioned studies regarding bi-prefix in E.C.A are that it is added to the Imperfective verb form to indicate either

---

[8]This term will be used throughout the study to refer to the Imperfective form lacking the bi-prefix, discussed in more details in section 1.4.

habitual or progressive readings, except stative verbs where only the habitual reading is possible. And its absence results in what is referred to in this study as the 'bare Imperfective' form, which is often correlated with a non-factual or modal reading, and occurs in contexts other than the indicative mood, such as subjunctive or jussive imperative. Studies concerned with code-switching between ECA and MSA shed light on how pervasive the use of this prefix is, extending in certain contexts to verb forms which are not part of the ECA inventory.

Following is a description of another important marker of morphosyntactic tense/ grammatical aspect, that is the auxiliary *kān* which is used to form compound tense in ECA. Section 2.3.2 below discusses the auxiliary and provides some example sentences for its use, as well as the literature review of *kān* in ECA.

2.3.2 Auxiliary *kān*

kān is defined as "carrier of a tense, an aspect, or a mood distinction which would otherwise have no carrier" (Abdel-Massih et al., 1979, p.277). This is illustrated through its effect of acting as the verbal form that adds tense and aspect to equational sentences that otherwise lack any verb, as in:

(10) ʕali kān   taʕbān
     Ali be.PV.3SGM tired.PTCP.SGM
     Ali was tired

(11) ʕali kān   f-el-bēt
     Ali be.PV.3SGM in-the-house
     Ali was in the house

It also alters the temporal and aspectual properties of a verbal sentence by acting as the carrier of a new temporal feature (Abdel-Massih et al., 1979, 278), as in:

(12) ʕali ħa-ykūn   bi-yedris   lamma newsal
     Ali FUT-be.IPFV.3SGM BI-study.IPFV.3SGM when arrive.IPFV.1PL
     Ali will be studying when we arrive

The auxiliary here is marked with the prefix *ħa* which marks future tense and puts the whole event in the future progressive, lack of the auxiliary will make it present progressive, as in Abdel-Massih et al. (1979)'s example:

(13) ʕali bi-yedris
      Ali BI-study.IPFV.3SGM
      Ali is studying

Another slightly different function of *kān* is that of adding a subjunctive meaning to a sentence when it occurs in the imperfective form *ykūn*, which is referred to as the 'bare form' of the auxiliary, as in Abdel-Massih et al. (1979, p.278)'s example:

(14) yikūn   bi-yħebbaha?
      be.IPFV.3SGM BI-love.IPFV.3SGM
      Could it be that he loves her?

"Auxiliaries cross-linguistically are a closed class of verbal elements. They generally have developed from main verbs such as be, stay, have or go" (Butt et al., 1999, p.60). This definition can be further augmented by Jelinek (1981)'s account of auxiliaries as constituents that contain a specified set of elements marking tense and/or modality[9]. Jelinek argues that there is a constituent of Egyptian Arabic sentences that meets the criteria set by Steele for defining AUX category, in that it is the only constituent that can mark tense in the sentence. Members of this category include: "the finite inflections of the auxiliary verb KWN, the copula; the particles marking sentential negation; and certain pronouns marking person subject" (Jelinek, 1981, p.9).

According to Jelinek (1981)'s account of the category AUX in Egyptian Arabic, there is only one auxiliary verb in ECA, referred to in her work as KWN[10]. It is an auxiliary verb that may correspond to the English verbs "be" or "have" according to aspect and voice of the sentence

---

[9]This account of auxiliaries across languages is adopted by Jelinek from: Steele et al. (1981) An Encyclopedia of AUX: A study in cross-linguistic equivalence. Cambridge,MA. MIT Press.

[10]This is the root of the auxiliary verb *kān* in Arabic, and is used in Jelinek's thesis to refer to the same auxiliary mentioned in the current study as *kān*.

(Jelinek, 1981). This auxiliary is a regular verb in that it inflects for person, number and gender as other members of the class 'verb', while it differs from other verbs in that it is the only member of this class that can occur with various predicates in the language to mark tense contrasts, which makes the auxiliary KWN a verb which has distinct semantic and syntactic properties that are peculiar to it (Jelinek, 1981).

Eisele (1992) provided an account of auxiliaries in Egyptian Arabic which raises questions about Jelinek's arguments and criteria for having a distinct syntactic category AUX in Egyptian Arabic. He observes that evidence of distinct syntactic behaviour that is characterstic of the English AUX for example, such as lack of do-support, tag question, negative contraction, auxiliary reduction or V' deletion, is absent in the case of the ECA auxiliary KWN (Eisele, 1992). This is clear through the fact that KWN is not the only verb that can be negated in ECA (any verb can), and its inversion with subject does not form a yes-no question, which is different from English AUX (Eisele, 1992). But despite this, Eisele's account supports the strongest of Jelinek's claims of AUX category for KWN in that it designates tense in compound tenses, which although is true, but it is not specific to *kān*, where many other verbs as *fedel* 'remain' and *baʔa* 'become' can fulfill this feature (Eisele, 1992)[11]. This is clear in his statement that "in many compound-verb phrases the highest verb often carries the deictic time reference, whether or not the verb is *kān* 'to be' " (Eisele, 1992, p.154).

The main claim in Eisele's work is that AUX does not form a distinct syntactic category as it behaves like other verbs in the main dynamic syntactic properties such as negation and movement, as shown in the examples below where both the auxiliary and the lexical verb can be negated using the same negation circumfix in ECA:

(15) Mona ma-ʔakalet-ʃ
      Mona NEG-eat.PV.3SGF-NEG

---

[11]In Eisele (1992)'s account, *fedel* and *baʔa* are included within the general term 'auxiliaries', while the current study provides a distinct description for them as phasal verbs in Chapter 4.

Mona didn't eat

(16) Mona ma-kanet-ʃ   bi-tākol
      Mona NEG-be.PV.3SGF-NEG BI-eat.IPFV.3SGF

      Mona wasn't eating

However, Eisele (1992) states that AUX does form a distinct lexical category; where verbs belonging to this subclass of verbs (auxiliaries) share specific features that are not common among other verbs. These features are represented as follows, following Eisele (1992), with special emphasis on how *kān* behaves with regard to each:

- The degree to which the auxiliary forms a syntactic unit with the following verb. This is clear in whether the auxiliary subcategorises for a complementiser or not (+/-COMP). In this respect, *kān* does not subcategorise for a complementiser but is rather followed directly by the verb with no intervening clauses, as shown in:

  (17) ʕali kān   bi-yelʕab
        Ali be.PV.3SGM BI-play.IPFV.3SGM
        Ali was playing

  (18) *ʕali kān   ʔin   bi-yelʕab
        Ali   be.PV.3SGM that BI-play.IPFV.3SGM

- Whether there is subject corefrentiality between the auxiliary and the following verb, showing the degree to which both verbs form a semantic unit (+/- COREF). Regarding this feature, *kān* obligatorily co-refers to the same subject as the following verb.

  (19) ʔana smeʕt-ak   b-tetkallim
        I   hear.PV.1SG-you BI-talk.IPFV.3SGM
        I heard you talking

  (20) *ʔana kont   b-tetkallim
        I   be.PV.1SG BI-talk.IPFV.3SGM

- The type of time reference the embedded verb can have, which shows the degree to which the verb depends on the auxiliary for its deictic time reference, also referred to as 'temporal discreteness' (+/-DEICTIC). In this regard, the verb following *kān* is described by Eisele (1992, p.161) as "may not have an independent deictic time reference". Detailed discussion of this feature is presented in section 2.7 where the temporal/aspectual markings on each of the auxiliary and the lexical verb are analysed.

The category AUX is described by Eisele (1992) as a fuzzy category having some core members and other peripheral ones according to the number of features characteristic of these verbs. "Thus, 'auxiliary-hood' reflects a continuum of features, with some verbs being more auxiliary-like, and others being less auxiliary-like" (Eisele, 1992, 160). However, the feature of having a specific subcategorisation frame for the verb is decsribed by Eisele (1992) as the most important property defining a verb as being a member of the category AUX, and therefore *kān* is considered to be the most prototypical auxiliary of these verbs (Eisele, 1992). Despite the fact that the current study is not particularly concerned with specification of the category AUX, the subcategorisation frame of each of the verbs discussed in the current study (*kān*, the causative, phasal verbs and modals) is a very important feature in the verb's description and analysis. However the term auxiliary is used in the current study to refer to *kān* only, as the most prototypical auxiliary.

In her reference grammar of Modern Standard Arabic, Ryding (2005) accounts for *kaan-a*[12] as an auxiliary that is combined with another verb to form a compound tense form. It combines with different verb forms to convey different interpretations, where it is used in conjunction with present tense main verbs to convey habitual action in the past, or with verbs expressing states of mind such as knowing, feeling, understanding or liking to express states that extend over a period of time in the past (Ryding, 2005). These are often expressed in the past continuous tense in Arabic rather than simple past as in English, forming the past progressive (Ryding,

---

[12]This is the form of the verb in Modern Standard Arabic, which corresponds to *kān* in E.C.A, losing the final vowel.

2005). *kān* can also be used with the past tense main verb to express the past perfect form, or the future tense of the main verb to describe an unreal or contrary-to-fact condition, where an action could have taken place but it didn't (Ryding, 2005). Present or future tense form of *kān* can also be combined with a past tense main verb to give the future perfect form (Ryding, 2005). All these interpretations for the possible combinations between *kān* and the lexical verb are accounted for in section 2.5.2, where all possible forms of the auxiliary *kān* are combined with all possible forms of the lexical verb to produce compound tense forms, and the interpretation along with examples of each is provided.

In their account of negation in Modern Standard Arabic, Sharif and Sadler (2009) provided an analysis through the LFG framework for negation in simple and compound tenses in Modern Standard Arabic. This discussion is useful for the current study as it mentions the auxiliary *kān* as an auxiliary verb which combines in its finite form with various indicative verb forms to form compound tenses in which the auxiliary signals the feature TENSE and the verb signals one of the aspectual features PRF (completed), PROG (continuative) or PROSP (prospective) (Sharif and Sadler, 2009). These distinctions between aspectual features as well as the LFG analysis for these forms will be adopted and further discussed through analysis of data in the current study.

To sum up the discussion so far, auxiliaries in ECA form a subclass of verbs, which although behaves in many ways as other verbs in this dialect, but they have their distinct semantic and syntactic properties, the most important of which is signalling tense contrasts in compound tense forms. Regarding features of auxiliary-hood, *kān* is the most typical auxiliary, which can be combined with various forms of verbs to form compound tense structures that can designate different moods such as indicative or subjunctive according to the morphological structure of the auxiliary and the verb combined with it.

Having discussed two distinctive properties of ECA morphosyntactic tense forms, which are the

bi-prefix and the auxiliary *kān*, section 2.4 below provides an overview of the previous studies of tense and aspect in Arabic in general and ECA in particular, as well as a discussion of some of the main issues in the study of tense and aspect in ECA.

2.4   Accounts of tense and aspect in ECA

The question of how tense and aspect are represented in Arabic language has been the focus of many studies in that area. The most prominent finding relating all these studies is that both tense and aspect are represented in some way in Arabic structures. One of those is Comrie (1976)'s account which states that: "In written Arabic, there are two sets of forms, traditionally referred to variously as aspects, tenses, or states, and distinguished either as Perfect and Imperfect, or as Perfective and Imperfective"(Comrie, 1976, p.78). Another brief account is found in Binnick (1991)'s investigation of literature on temporal reference in Semitic languages which are assumed to be tenseless, where this assumption is seriously questioned. Binnick (1991) claims that Semitic languages have two finite tense forms, one unprefixed (*faʕal* '(he) did') and the other is prefixed (*ya-fʕal* '(he) is doing/does)[13]. This claim is supported by data from a survey investigating the temporal meanings of Arabic verb forms which concludes that 98.7% of the *faʕal* forms refer to the past and only 1.93% referring to future, and 72.34% of *yafʕal* forms refer to present and 14.89% to future and 12.77% refer to past tense.

The question of the interpretation of the morphosyntactic tense forms in Arabic as denoting tense or grammatical aspect is one of the main controversial issues in the study of tense and aspect in Arabic. Some accounts such as Fehri (2012)'s study of the correlation between aspect and time reference in different languages, with special emphasis on Arabic language, providing examples from M.S.A to support the description. According to Fehri's classification of languages which lack either tense or aspect marking where the distinction between these markings

---

[13]These forms are inflections of the MSA verb 'to do', with *faʕala* as the Perfective and *yafʕalu* as the Imperfective, which are commonly used in Arabic grammars to refer to the general template of the verb irrespective of a specific root.

is not clear, Arabic belongs to a class referred to as A-type in which the aspectual value is associated with a specified time reference marked on the verb (Fehri, 2012). And therefore, according to this, the conclusion is drawn that Arabic is considered a 'tense' language rather than an 'aspectual' one. The problem still persists, according to Fehri, in the "ambiguous use of the same finite verbal form for Past or Perfect (or non-past, Imperfect) to express Anteriority (or non-Anteriority) of Reference Time(=RT) with respect to either Utterance Time (UT) or Event Time(=ET). Perfectivity (or imperfectivity), on the other hand, is not morphologically expressed, being either correlated with past or perfect T (or non-past, imperfect)" (Fehri, 2012, p.4).

Another different view is adopted by studies such as that of Al-Aqarbeh (2011) differentiates between syntactic and semantic tense and aspect, where syntactic tense/aspect refers to the morphological grammaticalisation on the verb, while semantic tense/aspect refers to how these forms are interpreted which is also affected by the context and pragmatic information. The assumption adopted in Al-Aqarbeh (2011)'s study is that if the same form can be used in different temporal contexts (with past, present and future adverbials for example) then it does not encode tense, but only aspect. The conclusion drawn is that verb forms in Jordanian Arabic encode only aspect, and not tense, and absolute tense is established in the context as a pragmatic implicature. The reason the debate is ongoing between the two opposite views on tense and aspect in Arabic is attributed to the fact that most of the scholars were concerned mainly by the form of the verb rather than by analyzing how it is used and interpreted in language structures, and therefore missing important distinctions such as that between "the simple tense (where the verb is alone in the sentence), the compound tense (where the verb is accompanied by the auxiliary verb k-w-n) and the complex tense (which is usually expressed by dependent clauses)" (AlKhawalda, 1997, p.32). Examples of cases in MSA where the Imperfective verb form *yafʕalu* can be used to denote different tenses such as past, present and future are (AlKhawalda, 1997, p.32):

(21) lam   yalʕabu   al-waladu
    NEG.PAST play.IPFV.3SGM the-boy
    The boy didn't play

(22) yalʕabu   al-waladu
    play.IPFV.3SGM the-boy
    The boy plays/ is playing

(23) lan   yalʕaba   al-waladu
    NEG.FUT play.IPFV.3SGM the-boy
    The boy will not play

It is important to note here that the problems in specifying tense and aspect might be different between MSA and ECA. This is shown in the ECA sentences correspondent to examples (21), (22) and (23) mentioned above in (AlKhawalda, 1997), where the different tenses are not denoted by the same form, shown as follows:

(24) el-walad ma-lʕeb-ʃ
    the-boy NEG-play.PV.3SGM-NEG
    The boy didn't play

This example corresponds to the MSA example (21), but here the verb has the Perfective verb form which denotes past tense, combined with the circumfix (ma - ʃ) denoting negation in ECA.

(25) el-walad bi-yelʕab
    the-boy BI-play.IPFV.3SGM
    The boy plays/is playing

This corresponds to example (22) above, but here the verb has the bi prefix which is commonly used in ECA to denote present tense, as well as habitual or progressive aspect.[14]

(26) el-walad meʃ ħa-yelʕab
    the-boy not FUT-play.IPFV.3SGM
    The boy will not play

---

[14]This prefix is investigated in more details in section 2.3.1.

This corresponds to example (23) above, while here the verb is in the *ħa* prefixed verb form which is used to denote future tense, and it is preceded by the ECA negation marker *meʃ*. These examples show that the problem mentioned in (AlKhawalda, 1997) of using the same Imperfective verb form to denote different tenses does not exist in ECA due to the presence of distinct verb forms for each temporal interpretation. This will prove useful in the coming sections where the temporal denotation of verb forms in ECA will be specified in more details.

Comparison between the two opposite views on temporal marking in Arabic (tense only versus aspect only) needs a more thorough investigation of the syntactic and semantic properties of tense and aspect in Arabic, while the focus in the current chapter is merely on the ways in which morphosyntactic tense and grammatical aspect are expressed on the verb forms in ECA, giving special attention to the auxiliary *kān* and the ways in which it can be used to form compound tense forms. In addition to the fact that there are a number of differences between the expression of morphosyntactic tense and grammatical aspect on verb forms in MSA and ECA, as shown with the MSA examples (21, 22 and 23) and their counterparts in ECA.

Other accounts were devoted to the investigation of tense and aspect in ECA as a distinct language variety, where data is provided and analyzed in an attempt to reach a coherent picture about marking of grammatical tense and aspect on ECA structures. One of the works in this respect is Eisele (1990b)'s investigation of tense and aspect in ECA through which he provides and analyzes data in an attempt to formalize a comprehensive model for how tense and aspect are represented in this language variety, providing explanations for the temporal and aspectual characteristics of ECA. The main assumption is that ECA verbs are marked for both time reference and grammatical aspect, with means for referring to the three tenses (past, present and future) as well as habitual and progressive aspects. This is shown through the distinction made between deictic and non-deictic time references in compound tense forms, where deictic time reference can be marked on time adverbials as well as the first/uppermost verb of the

verb phrase, while non-deictic time reference is also marked on time adverbials as well as the lower member of the verb phrase. This is shown in example (27) below, where the auxiliary *kān* represents deictic time reference denoting past tense here, while the following verb represents non-deictic time reference, in this case progressive[15] (Eisele, 1990b).

(27) kān   bi-yiktib
      be.PV.3SGM BI-write.IPFV.3SGM
      (He) was writing

Another useful distinction is that made between complex/compound and non-complex tense[16], where in complex tenses (formed by combination of an auxiliary and a verb), deictic time reference will be marked on the auxiliary, and information about non-deictic time reference will be marked on the following lexical verb (as shown in example 27 above). While non-complex (simple) tenses mark both deictic and non-deictic time references on the same form, such as example (28) below where the verb is in the *bi-* prefixed verb form denoting both present and progressive.

(28) bi-yiktib
      BI-write.IPFV.3SGM
      (He) is writing

Following is section 2.4.1 which includes a brief discussion of some of the main issues facing the study of tense and aspect in ECA. Tackling these issues through description and analysis of the data is among the aims of the current study.

2.4.1   Issues in studying tense and aspect in Egyptian Colloquial Arabic

The question of how to interpret the morphosyntactic temporal marking on Arabic structures as tense, or aspect, or both, has been a controversial issue. Where traditional linguists view Arabic

---

[15]Despite the fact that the terms "deictic" and "non-deictic" time reference are not adopted in the current study, their meanings still hold. Where the auxiliary and verb are treated as having different kinds of markings, where the auxiliary marks tense while the verb marks grammatical aspect.

[16]Note that both terms "compound" and "complex" tenses are used interchangeably in Eisele's study.

as not having tense, but rather grammatical aspect expressed as either perfect or impefect. This view influenced the study of morphosyntactic tense marking in dialects too, including ECA (Eisele, 1990b). This controversy may be the result of the many problems which are often associated with attempts to assign tense and aspect to Arabic in general, and particularly ECA verb forms, such as those stated by Eisele (1990b) where:

- Embedding of almost all verb forms with the temporal auxiliary *kān* to give different tenses such as:

  (29) kān   katab
       be.PV.3SGM write.PV.3SGM
       (He) has written

  (30) kān   bi-yekteb
       be.PV.3SGM BI-write.IPFV.3SGM
       (He) was writing

  (31) kān   ħa-yekteb
       be.PV.3SGM FUT-write.IPFV.3SGM
       (He) was going to write

- The Perfective form can sometimes be used to give a present perfect tense, as in:

  (32) ḫallast   el-ʃoɣl   delwaʔti
       finish.PV.1SG the-work now
       I've finished the work now

- The *bi* prefixed Imperfective form is used to refer to habitual and progressive aspects, as in:

  (33) ʔaħmad bi-yzāker   delwaʔti
       Ahmed   BI-study.IPFV.3SGM now
       Ahmed is studying now

  (34) ʔaħmad bi-zāker   kol   yōm
       Ahmed   BI-study.IPFV.3SGM every day
       Ahmed studies everyday

- The AP (active participle) form can be used to indicate different tense interpretations, as in:

(35) ʔana fāhim
     I   understand.PTCP.SGM
     I understand

Example 35 expresses an aspect similar to progressive, denoting an ongoing, present state of affairs. This is different from the interpretation in:

(36) ʔana ʔāri   el-kitāb el-sana ʔilli   fātet
     I   read.PTCP.SGM the-book the-year which pass.PV.3SGF
     I read the book last year

(37) ʔana msāfer   faransa bokra
     I   travel.PTCP.SGM France tomorrow
     I'm travelling to France tomorrow

These problems led to the conclusion that most of the descriptive studies "have been deficient so far in two areas: in the definitions of tense, time reference and aspect, and in the full examination of the data concerning tense, time reference and aspect in CA" (Eisele, 1990b, p.176). This is where the current study aims to fill the gap, through providing a description and analysis for the expression of morphosyntactic tense and grammatical aspect in ECA. The next questions to ask is whether any form of *kān* can be combined with any form of the verb or are there any restrictions on their co-occurrence? and what are the possible tense and aspectual features each of these components can represent? and eventually how to represent these combinations in the LFG framework? These are the questions which the current chapter attempts to answer, starting by section 2.5 which is devoted to description of the data on which this study is based, showing the various patterns in which morphosyntactic tense and grammatical aspect can be marked on verbs in ECA, including a separate account for simple tense forms as well as the ways in which the auxiliary *kān* can be combined with the lexical verb to form compound tense forms.

2.5   Data

Despite the differences in definitions and interpretations of each of the notions 'tense' and 'aspect', the study of both seem to be greatly interrelated, since both are often marked on the same structures. Therefore the current study will account for each whenever it is marked on the verb, while preserving the differences between features of morphosyntactic tense and those of grammatical aspect. Tenses will be distinguished into either simple, which corresponds to cases where only the lexical verb marking is used to indicate tense (discussed in section 2.5.1 below), or compound, where the verb is combined with the auxiliary *kān* to derive the required tense and aspectual interpretations (described in section 2.5.2).

This study is only concerned with structures and functions of inflected verb forms, where combination of the verb with time adverbials or the use of participles will not be addressed here. In the current study, the term 'tense' will be used to refer to the morpho-syntactic property rather than the direct temporal reference of the form discussed. This will allow for distinguishing between the morpho-syntactic marking of the verb form and its semantic interpretation, where although both might be illustrated, the main focus of the current study is on the morpho-syntactic marking of the verb form. Morphosyntactic tense marking of verbs in the example sentences is represented in each example in boldface, such as **simple past**, while the complete set of possible tense/aspect markings on all simple and compound constructions and the morphosyntactic term for each is presented later in table 2.4, after the detailed description of the data. Glossing of each of the sentences includes mentioning the equivalent tense form for each verb in English, to avoid using the aspectual terms 'perfective' and 'imperfective' in referring to tense, so it is not confused with the analysis given. Along the same line, the terms **Imperfective** and **Perfective** with initial upper case letters are used to refer to the morphological form of the verb, as in Perfective *katab* 'wrote' which can refer here to a past tense verb, while the grammatical aspect is refered to by the terms **imperfective** and **perfective** in lower case letters, as in perfective *katab* refering to a completed event as used in *kān katab* '(He) has writ-

ten'.

The data is composed of sentences extracted from the corpus of ECA online texts that was built for the current study[17], where I have searched for the verb in its different tense/aspect morphosyntactic forms and different person, number and gender inflections so as to include different varieties of the inflected verb form. Although I am a native speaker of Egyptian Colloquial Arabic, some of the structures mentioned in the data were problematic as to reaching a decision on whether these structures are actually used in the dialect or not. For these structures, I have checked the corpus specifically for such an example sentence, where if the structure occured even for once, then its included, while non-occurence at all throughout the corpus in addition to its questioning by a native speaker was considered enough reason to rule it out. This was an important step in order to avoid forming rules that might not be really implemented in the dialect, and therefore forming what can possibly be ill-formed structures. Each sentence in the data below includes the example in transcription, glossing, translation into English and some additional notes on the interpretation(s) of the sentence as well as the corresponding equivalent in Reichenbach's terms, namely R, S and E and the relation between them, as described previously in section 2.1.2.

2.5.1   Simple tense forms

Arabic verbs can be marked for simple tense in a variety of different forms, these are shown along with the interpretations in section 2.3. Simple tense generally expresses the relation between event time (E) and speech time (S) (AlKhawalda, 1997). Lexical verbs are distinguished into static vs. dynamic, where states are defined by Kearns (2011, p.157) as being "internally uniform, one moment being much like another, and have no forward movement or natural beginning or end." This distinction can affect the interpretation of the tense marking on the verb, therefore

---

[17]Details about the corpus in section 1.3.

each morphosyntactic tense form will be illustrated for both a dynamic and a static verb. Sentences illustrating simple tense forms are shown as follows:

- The **simple past** in ECA is encoded by the Perfective verb form, such as:

(38) katab   kalām saʕīd   ʕa-l-fēsbuk
     write.PV.3SGM talk   happy on-the-facebook
     (He) wrote happy thoughts on facebook

The same form applies to stative verbs, as in:

(39) fehemt   el-sirr
     understand.PV.1SG the-secret
     I got the secret!

In Reichenbach's terms, both R and E coincide and precede S.

- The **simple present** in ECA is encoded by the *bi-* prefixed Imperfective form, as in:

(40) fawwāz bi-yelʕab   maʕ ʔawlād-u
     Fawwaz BI-play.IPFV.3SGM with kids-his
     Fawwaz plays/is playing with his kids

This form can refer to the event in two different ways, either as an ongoing process (also referred to as 'progressive' PROG), or it can be used in a habitual sense (HAB), where he always plays with his kids. The distinction is usually made clear through the context in which the sentence occurs. Although R,E and S all coincide, the habitual aspect refers to various points where the event recurs, while progressive aspect refers to a certain point through the event where it is in progress. Therefore this form could mark either **habitual present** or **progressive present** shown respectively as follows:

(41) fawwāz bi-yelʕab   maʕ ʔawlād-u kol   yōm
     Fawwaz BI-play.IPFV.3SGM with kids-his   every day
     Fawwaz plays with his kids everyday

(42) ?ana ʃāyif   fawwāz bi-yelʕab   maʕ ?awlād-u delwa?ti
     I   see.PTCP.SGM Fawwaz BI-play.IPFV.3SGM with kids-his   now
     I can see Fawwaz playing with his kids now

The same structure when used with stative verbs can give only the habitual reading as in

the **habitual present** form, such as:

(43) ħasan   bi-yħeb   el-ʃukulāta
     Hassan BI-love.IPFV.3SGM the-chocolate
     Hassan loves chocolate

The same form can also be used to indicate general statements or facts, as in:

(44) ?el-?amar bi-ynawwar   el-saħara
     the-moon BI-light.IPFV.3SGM the-desert
     The moon lightens the desert

The bare Imperfective form, on the other hand, can only be used in examples such as:

(45) fawwaz bi-yħib   yi-lʕab   maʕ ?awlād-u
     Fawaz BI-love.IPFV.3SGM play.IPFV.3SGM with kids-his
     Fawaz loves to play with his kids

The stative verb *bi-ħib* 'loves' is bi-prefixed, while the action verb 'playing' is in the bare

form, but only the *bi-* prefixed verb marks the time of the action as present, while the

bare imperfective denotes the action being described.

- **Future tense** is marked in ECA by the *ħa-* prefixed form. In Reichenbach's terms for

  future tense, S precedes both R and E. The examples are as follows:

(46) ħa-netkallem   ketīr
     FUT-talk.IPFV.1PL lots
     (We) will talk a lot

This sentence refers to a situation that is due to take place in the future, where S precedes

E and R. The prefixed form can also occur with stative verbs, such as:

(47) ʔana meʃ ħa-ʔbal   el-ʔihāna
     I   NEG FUT-accept.IPFV.1SG the-insult
     I will not accept the insult

To sum up the current section, simple tense is marked on the main verb of the sentence, which in this case denotes tense as well as some aspectual features in the case of *bi-* prefixed verbs. Simple tense verb forms in ECA along with the morphosyntactic tense marking of each are shown in table 2.2:

Form	Morphosyntactic Tense Marking
Perfective	past
bi-prefixed	present progressive/habitual
ħa-prefixed	future

Table 2.2: Simple tense

Note that the bare imperfective form is not included as a simple tense form, this is because it does not occur as a main verb, but rather as an embedded verb. This is illustrated through the following examples:

(48) ʔel-nās   ʔilli ʕayza   teʕmel   mozahrāt
     the-people who want.PTCP.SGF do.IPFV.3SGF demonstrations
     The people who want to make demonstrations

(49) badaʔ   yiktib   ʔimēl tawīl
     start.PV.3SGM write.IPFV.3SGM email long
     He started to write a long email

Further discussion of the status of this bare form as a non-finite verb is discussed in more details in section 1.4.

2.5.2   Compound tense forms

The data represented in this section includes sentences where the auxiliary *kān* is used in combination with the lexical verb. This combination results in a compound tense form, where the

auxiliary denotes tense (specifying relation between R and S), while the lexical verb denotes aspect (specifying relation between E and R). This is the position adopted by many studies concerned with tense and aspect in ECA, such as Jelinek (1981) who claims that in sentences where both tense and aspect are marked in a sentence in Egyptian Arabic, AUX marks tense while the PREDICATE marks aspect.

The data represents combinations between various forms of the auxiliary *kān* and the lexical verb including different types of verbs (stative vs. dynamic), changes in the marking of grammatical aspect on lexical verb, and different morphosyntactic tense markings on the auxiliary. Since the main focus in this part of the study is about the auxiliary and how different forms can give different tense markings, occurences of each of the combinations was indexed according to the form of the auxiliary. The most problematic issue was the present tense marking, where both the bare Imperfective and bi-prefixed forms are collectively refered to in previous studies as 'Imperfect' which might imply that they have the same interpretation, while each of these forms was investigated separately in the current study and I have specified the meanings associated with each and their different uses in the data.

The following sections represent all the possible structures of compound tense forms in ECA, showing the ways in which morphosyntactic tense and grammatical aspect are marked on the auxiliary and main verb respectively, as well as an example of this structure from the ECA corpus and its interpretation. The terms 'Perfective', 'Imperfective' and 'Prospective' are used here to refer to the form of the lexical verb. Section 2.7.2 will then be concerned with defining how these notions correspond to specific tense and aspectual markings, while the current section is concerned mainly with description of the data, and therefore these terms are used only to refer to the form of the verb, while its interpretation is shown in relation to each example. Division of the sections is based on the form of the auxiliary in the compound tense construction, either *kān* 'was', *bikūn* 'being' or *ħaykūn* 'will be'.

2.5.2.1   *kān*

This form of the auxiliary is also known as the Perfective form, note here that 'Perfective' refers to the form only and not any aspectual marking, as this auxiliary marks past tense in compound tense forms. The range of verb forms with which it can be combined are represented as follows, divided according to the form of the lexical verb into 'Perfective', 'bi-prefixed', 'bare' and 'Prospective' with stative and action verbs represented for each[18].

- Perfective verb form:

    (50) kān   ʕamal   ʔabl-aha   film
         be.PV.3SGM do.PV.3SGM before-it.3SGF movie
         He had done a movie before it (year 2000)

    This sentence denotes the time at which an event occured (making the movie) before the time refered to (the year 2000). And therefore the interpretation here resembles that of the English **past perfect**, where E precedes R, and both precede S.

- Perfective form-Stative verb:

    (51) law kān   ħabb-eni   kān   ʔal-l-i
         if   be.PV.3SGM love.PV.3SGM-me be.PV.3SGM tell.PV.3SGM-to-me
         If he had loved me he would've told me

    The example here is also **past perfect**, where the loving event expressed by past perfect precedes the telling event, and both precede speech time.

- *bi-* prefixed Imperfective verb form:

    (52) ʔaħmad ʔismaʕīl kān   bi-yḫattat   l-hogūm   ʕa-l-ħudūd
         Ahmed Ismail   be.PV.3SGM BI-plan.IPFV.3SGM for-attack on-the-borders
         el-masreyya
         the-Egyptian

---

[18]Representing both stative and action verbs is specifically useful in the case of bi-prefixed verbs, as it determines the progressive or habitual aspect associated with the verb form

Ahmed Ismail was planning for an attack on the Egyptian borders

This sentence is in the **past progressive**, where R refers to a certain point at which the event E is in progress. The same form can be used to give a habitual reading, such as in:

(53) ʔel-barnāmeg kān   bi-yitzāʕ   ʕa-l-MBC   fi nafs waʔt
     the-program   be.PV.3SGM BI-show.IPFV.PASS.3SGM on-the-MBC at same time
     el-mosalsal
     the-series
     The program was shown on MBC (a TV channel) at the same time as the series

Here the same construction marks **past habitual**, where the program aired at the same time each day. Here, R refers to various points at which E recurs. This is a property of *bi-* prefixed verbs of giving either a habitual or progressive reading, which can only be clarified through context. Note also that this combination (*kān* + *bi-* prefixed verb) had the highest frequency of occurence among all the other verb forms. The exact frequency of each verb form is beyond the focus of the current study, and therefore will not be calculated. However, it helps demonstrate the variability in use between the different forms, which might explain why some forms are less acceptable than others.

- *bi-* prefixed Imperfective form- Stative verb:

  (54) howwa kān   bi-yħebbe-ni   bas ʔana kont   b-aħeb
       he   be.PV.3SGM BI-love.IPFV.3SGM-me but I   be.PV.1SG BI-love.IPFV.1SG
       ħad   tāni
       someone else
       He used to love me but I used to love someone else

The verbal combination here is in the **past habitual** where R and E coincide, and that is due to it being a stative verb which can not be seen as a process but rather can only have the habitual interpretation denoting a characteristic (Eisele, 1990b). An important note to make here is that the interpretations provided here refer only to the form of the morphosyntactic tense/ viewpoint aspect, which is presented in the English translation,

while it does not necessarily correspond to the tense and aspectual markers on each verb, as this will be dealt with in more details in the analysis. For example **past habitual** does not necessarily mean that the aspect feature on the verb will be marked as habitual, but it might occur as a binary feature to show the intended meaning. These decisions will be taken considering the overall structure and possible markings on different verbs in the data.

- Bare imperfective verb form:

(55) kont   ʔaʕod   fi-l-matʕam   m-asmaʕ-ʃ   sōt
     be.PV.1SG sit.IPFV.1SG in-the-restaurant NEG-listen.IPFV.1SG-NEG sound
     I used to sit in the restaurant and not hear any sound

The interpretation of this sentence is somehow problematic, where it refers to the state of the person as 'sitting', which seems to correspond to the general interpretation of the imperfective aspect which can be subdivided into other categories such as habitual or continuous according to the forms expressed in a language (Comrie, 1976), therefore this construction is marked here as **past imperfective**. The habitual or progressive meaning can only be inferred here through the context in which the sentence occurs, as in a narrative. On the other hand, the progressive or habitual viewpoint aspect can be expressed on the sentence if the bi-prefixed verb was used, as in the following constructed[19] example:

(56) kont   b-aʕod   fi-l-matʕam   kol yōm
     be.PV.1SG BI-sit.IPFV.1SG in-the-restaurant every day
     I used to sit in the restaurant every day   [constructed]

This bare Imperfective form was questioned by the researcher and occured very few times in the corpus, especially in comparison to the bi-prefixed verb. This low frequency of occurence is an indication about the restrictions on its use, in comparison to the bi-prefixed

---

[19]A constructed example is one that was not extracted from the corpus, but which I (as a native speaker) added to demonstrate a certain point. The constructed example is marked by [constructed].

form for example which occured frequently in the corpus and is used more freely. As a
native speaker I do not use this bare Imperfective form as the main verb (as in sentence
55), but I would percieve it as an 'odd' form of the language, where I can understand
what it means and might hear it in some idiolects or in old movies, but would not use it
to express an event in the indicative. Discussion of this bare form occurs in more details
in relation to non-finiteness in section 1.4.

- Bare imperfective form- Stative verb:

  The combination between the auxiliary and the bare imperfective form of stative verbs
  was found in the corpus only in one occurence, which is the following question:

  (57) kont   teʕraf-o   maʕrefa   ʃaẖseyya?
        be.PV.2SGM know.IPFV.2SGM-him knowledge personal?
        Have you known him personally?

  This example represents the **past imperfective** just like the case in example 55, as here
  the correlation with a stative verb seems to focus on the meaning of the verb as a state
  or characteristic, where the question asks whether the listener used to know him in the
  past. An interesting point to make is that the use of the bare imperfective here is not as
  questionable as that in 55, where the use of the bi- prefixed verb is ungrammatical in this
  construction, as in:

  (58) *kont   b-teʕraf-o   maʕrefa   ʃaẖseyya?
        be.PV.2SGM BI-know.IPFV.2SGM-him knowledge personal?
        [constructed]

  This shows that the use of the bare Imperfective is indeed restricted to certain contexts,
  such as questions and non-factual situations, while the bi-prefixed verb is used more in
  indicative uses. The combination between *kān* and stative bare Imperfective verbs also
  had low frequency of occurence in the corpus, which maybe the result of the low frequency

of occurence of stative verbs in compound tenses generally in addition to the rare usage of the bare Imperfective form in compound tense. Note also that both stative and non-stative verbs have the same interpretation in terms of morphosyntactic tense and grammatical aspect, which shows that this bare form is different from the bi-prefixed form, where bi-prefixed stative and non-stative verbs differ in their aspectual interpretations between habitual and/or progressive aspects.

- Prospective (*ħa-* prefixed) verb form:

This term is used to refer to the verb in the form of *ħa-* prefixed Imperfective, which is used to mark future tense in simple tense verbs as shown previously in section 2.5.1. The term 'Prospective' is used instead of 'future tense' in order to unify the notions describing form of the verb, and it is adopted from Jelinek (1981) who calls this form the HA-imperfect marking prospective aspect. Note however, that this form is glossed in the examples as 'FUT' only because this is the standard notion used in the Leipzig glossing rules which are adopted throughout the current study. In Reichenbach's terms, the prospective form expresses a case where R precedes E (R<E), as shown in:

(59) konna   ħa-nmūt
     be.PV.1PL FUT-die.IPFV.1PL
     We were going to die

Here the sentence refers to something that was going to happen in the past but it did not, it is also refered to in AlKhawalda (1997) as 'future in the past' as it includes only R preceding S. The terminology adoped here to refer to this structure is **past prospective** following Eisele (1990b) as it shows both the tense (past) as well as grammatical aspect (prospective).

- Prospective (*ħa-* prefixed) form- Stative verb:

(60) ʔana kont   ħa-ʕraf   mnēn?
     I   be.PV.1SG FUT-know.IPFV.1SG from.where?
     How would I've ever known?

This sentence although it takes the form of a question, but it is still an instance of the **past prospective** where R<S and R<E. This was the only example found in the corpus for a stative verb in this form.

2.5.2.2 *bi-kūn*

This form can also be refered to as the *bi-* prefixed Imperfective, which is again a term refering merely to the form of the verb. However, this form is unique in that it does not seem to mark tense as straightfowardly as other forms of the auxiliary do, where it is expected to mark present tense due to its form[20], but in actual use it refers to a situation that cannot be defined at a certain point of time. This is also mentioned in Eisele (1990b) as a property of *bikūn* as giving "a non-specific time reference, and thus is understood with a habitual reading"(Eisele, 1990b, p.198). This interpretation is assumed by the current study to be the reason for the special marking of *bikūn*, which is further shown in section 2.7.3.

This form was the most restricted in use among forms of the auxiliary, and it was questioned by the researcher and investigated through the corpus to verify its use and interpretation. This might be a property of the present tense marking on the auxiliary and the limitation of forms of the verb with which it combines, this seems to apply even in MSA as cited in AlKhawalda (1997, p.272) that "The present form of the auxiliary verb *yakuunu*, except in certain contexts such as the usage of modal verbs, is invisible in the present perfect, the present continuous, and in copular sentences." The only cases where the auxiliary *bi-kūn* was used in compound tense construction are presented as follows, divided according to the form of the following lexical verb into:

- Perfective verb form:

  (61) leh   el-rāgel   ma-yetgawwez-ʃ   ʔarmalt-u? lʔan-u bi-ykūn
       why the-man NEG-marry.IPFV.3SGM-NEG widow-his? as-he   BI-be.IPFV.3SGM

---

[20]With analogy to the cases of bi-prefixed Imperfective lexical verbs in I marked as present, such as example (40) above.

māt
die.PV.3SGM
Why wouldn't a man marry his widow? Because he would then be dead

This verbal construction marks **present perfect**, however in interpretation, the sentence refers to an imaginary situation, where the question asks about a generic property of humans. The reason for this might be the complexity of marking on *bikūn* as shown earlier as well as the peculiar nature of present perfect in ECA as stated by Eisele (1990b) that it is used in a different way from its English equivalent, and can be combined with a wide variety of time adverbials that might affect its interpretation. When the verb is inchoative (as the case here), the perfect verb form has stronger tendency to be interpreted as resultative, denoting entering into a state (Eisele, 1990b). This seems to be the case with present perfect in English too, where it is described by Kearns (2011) to be semantically very complex as its time reference can span both past and present resulting in an unidentified event time.

- *bi-* prefixed Imperfective verb form:

(62) mīn ʔaktar lāʕib bi-txāf   menn-o   lamma bi-ykūn
      who more   player BI-fear.IPFV.2SGM from-him when   BI-be.IPFV.3SGM
      bi-yilʕab   ʔoddām farīʔ-ak?
      BI-play.IPFV.3SGM infront   team-your?
      Who is the player from whom you're most afraid when he's playing against your

      team?

This question is asking about the state of the player, it may also be considered a generic question that does not include reference to any specific point in time, but is rather questioning the state whenever it occurs. Here the morphosyntactic tense/aspect is hard to pinpoint, as per the marking on the forms it refers to present habitual or present progressive according to context, while it differs from other bi-prefixed lexical verbs in the simple tense form, as in:

(63) 2a7mad bi-yilʕab   kōra
     Ahmed BI-play.IPFV.3SGM ball
     Ahmed plays/is playing football

The difference lies in the fact that example 62 refers to a general case which does not relate to a certain point in time, but rather can be used to refer to various points in time where the event can take place. This might be due to the special status of *bikūn* which gives an interpretation of general state. And so for now it will be refered to as **present imperfective**.

- Prospective (*ħa-* prefixed) verb form:

(64) ħatta lamma masr   bi-tkūn   ħa-telʕab   ma-bahtamm-eʃ
     even when Egypt BI-be.IPFV.3SGF FUT-play.IPFV.3SGF NEG-care.1SG-NEG
     Even when Egypt is about to play, I don't care

This sentence also refers to the general case where an event is about to happen, without specifying the exact time for the situation. It can be refered to here as **prospective present** following Eisele (1990b), where R precedes E (R<E) in the aspectual marking.

In all these sentences, the auxiliary *bi-kūn* assigns a deictic time reference in a way that is different from the case of the past tense auxiliary *kān*, but rather gives the sentences a somewhat 'characteristic' interpretation as refered to in Eisele (1990b), where it is marking an undelimited or unspecific time reference as opposed to a non-habitual reading where the event has a delimited time reference. The sentences here address the state in case a potential event happens, without stating a specific point in time where it occurs. Note also the correlation between the auxiliary and *lamma*, which is a particle referring to possible points in time where the event may occur, similar to the English 'whenever'. A further note to make here is that all these examples tend to carry a certain mood that can be called for now 'conditional', but it is a result of the restricted sentence structures in which this form occured in the corpus that is causing this modal effect, as it is missing from other possible examples such as:

(65) kol   yōm fi nafs el-maʕād b-akūn   b-atfarrag   3ala el-telifizyōn
     every day in same the-time BI-be.IPFV.1SG BI-watch.IPFV.1SG on   the-television

     Every day at the same time (I am being at the state of) watching TV   [constructed]

(66) kol   yōm fi nafs el-maʕād b-akūn   ʔāзid   fi el-bēt
     every day in same the-time BI-be.IPFV.1SG sit.PTCP.SGM in the-house

     Every day at the same time (I am being at the state of) sitting at home   [constructed]

Where each of these sentences refer to a situation that happens at different points in time.
Note that in this section of compound tense structures with *bi-kūn* no distinction was made
between stative and non-stative lexical verbs, this is due to the low frequency of occurence of
this auxiliary and therefore all the sentences in which it occured in the corpus were included as
examples here.

2.5.2.3   Future tense *ħa-ykūn*:

This form is also known as the *ħa-* prefixed Imperfective form, and it denotes that S precedes R
(S<R) which describes future tense, while the relation between E and R is denoted by the form
of the following verb which marks grammatical aspect. The range of verb forms with which
ħa-ykūn can be combined are represented as follows:

- Perfective verb form:

  (67) waʔt-aha ħa-nkūn   darab-na   ʔamrīka w-el-ʔittiħād   el-ʔoroppi
       time-her FUT-be.IPFV.1PL hit.PV.1PL America and-the-union the-European
       By then, we will have hit (strongly affected) America and the European Union

  This sentence refers to a situation that may happen in the future but referring to the
  event as completed, which can be described as **future perfect**. Therefore, the auxiliary
  marks future tense while the verb marks the grammatical aspect perfect.

- Perfective form- Stative verb:

(68) ħa-neʔdar   nastafīd   men baʔd   kuwayyes leʔen-ena
     FUT-able.IPFV.1PL benefit.IPFV.1PL from each.other well   because-us
     ħā-nkūn   fehemna   baʕd   ẖalās
     FUT-be.IPFV.1PL understand.PV.1PL each.other finally
     We will then be able to benefit well from each other as we will have already understood

     each other

This sentence refers to a situation that will occur in the future based on the fact that an-
other event had already finished, where the state represented by the verb (understanding)
will be established. Therefore this is another example for **future perfect**.

- bi-prefixed Imperfective verb form:

(69) ħa-nkūn   ennaharda b-nelʕab   fi maʕhad gūta   el-sāʕa sabʕa
     FUT-be.IPFV.1PL today   BI-play.IPFV.1PL in institute Goethe the-hour seven
     We will be playing (performing) today at Goethe Institute at 7 O'clock

This sentence refers to the future point in time at which the situation will be taking place
as an ongoing process, corresponding to the **future progressive**. Therefore here the
auxiliary marks future tense (S<R) while the verb marks progressive aspect (E=R). This
verb can also mark habituality in a different context (although no example was cited for
it through the corpus), where it would refer to the future time where this action would be
a habit expressed by a different time adverbial or when bi- is prefixed to a stative verb,
as in the constructed example:

(70) el-sana el-gayya   7a-tkūn   bi-tefham   ʔalmāni
     the-year the-coming FUT-be.IPFV.2SGM BI-understand.IPFV.2SGM German
     kowayyes
     good
     Next year you will understand German well   [constructed]

- bi-prefixed form Stative Imperfective verb:

(71) mesh maʕʕūl ʔin   ʔeħna ħa-nkūn   bi-nefham   ʔaktar men
     NEG   possible that we   FUT-be.IPFV.1PL BI-understand.IPFV.1PL more   from
     ʔidāret   el-ʔahli
     administration the-Ahli
     It's not possible that we will know better than the Ahli (football team) administration

This structure marks **future habitual** which refers to multiple points in time at which the event recurs as a characteristic. However, this sentence refers to a counter-factual situation (irrealis), where although the auxiliary has the future tense marking form but it does not refer to a situation that is going to actually happen in the future, but rather to a state that is not considered possible. While the verb marks that state itself as a characteristic, which is the interpretation of the bi-prefixed stative verb.

To sum up the current section, compound tense in ECA is marked through a combination between the auxiliary and the lexical verb, the auxiliary marks tense while the lexical verb marks grammatical aspect. The set of possible combinations between the auxiliary and the lexical verb in their different forms, along with morphosyntactic tense marking of each is ilustrated in table 2.3.

Auxiliary Form	Verb Form	Morphosyntactic Tense Marking
Perfective	Perfective	past perfect
Perfective	bi-prefixed	past progressive/habitual
Perfective	bi-prefixed (stative)	past habitual
Perfective	bare Imperfective	past imperfective
Perfective	ħa-prefixed	past prospective
bi-prefixed	Perfective	present perfect
bi-prefixed	bi-prefixed	present imperfective
bi-prefixed	ħa-prefixed	prospective present
ħa-prefixed	Perfective	future perfect
ħa-prefixed	bi-prefixed	future progressive/habitual
ħa-prefixed	bi-prefixed (stative)	future habitual

Table 2.3: Compound tense

Section 2.6 below presents a discussion of the simple and compound tense data, including list-

ing all the morphosyntactic tense forms for the verb in simple tense, the possible combinations between different forms of the auxiliary and the lexical verb, as well as the impossible combinations that did not occur in the corpus. It also includes comparison between the findings in the current study and those of other studies about tense and aspect in ECA, namely Jelinek (1981)'s and Eisele (1990b)'s accounts of ECA.

2.6   Discussing the data

The data in section 2.5 represented all the possible morphosyntactic tense constructions in ECA, including simple tense which is marked on the main verb, as well as compound tense formed through the combination between the auxiliary and the lexical verb. Table 2.4 below demonstrates all of the possible verb forms that mark tense and aspect in ECA as was illustrated in the data, including both simple and compound tense forms:

Auxiliary Form	Verb Form	Morphosyntactic Tense Marking
	Perfective	simple past
	bi-prefixed	present habitual|present progressive
	bi-prefixed (stative)	present habitual
	ħa-prefixed	future
Perfective	Perfective	past perfect
Perfective	bi-prefixed	past habitual| past progressive
Perfective	bi-prefixed (stative)	past habitual
Perfective	bare Imperfective	past imperfective
Perfective	ħa-prefixed	past prospective
bi-prefixed	Perfective	present perfect
bi-prefixed	bi-prefixed	present imperfective
bi-prefixed	ħa-prefixed	prospective present
ħa-prefixed	Perfective	future perfect
ħa-prefixed	bi-prefixed	future habitual | future progressive
ħa-prefixed	bi-prefixed (stative)	future habitual

Table 2.4: Possible forms

On the other hand, other tentatively possible combinations are missing from the data as they

were questioned regarding their usage in the dialect. These are the combinations which were searched for in the corpus, but did not get any results, including the following combinations:

- bi-prefixed auxiliary with bare Imperfective lexical verb, which would tentatively be:

(72) *bi-ykūn   yilʕab
    BI-be.IPFV.3SGM play.IPFV.3SGM

- *ħa-* prefixed auxiliary with *ħa*-prefixed lexical verb, this combination did not occur at all throughout the searches. Despite the fact that there is no explanation for its lack of usage, as there are instances of succession of two perfective forms for example in the past perfect, so there is no reason to stop the succession of two *ħa*-prefixed imperfective forms. It is also easy to deduct the resultant meaning and usage in the suitable context, this construction would be like:

(73) ħa-ykūn   ħa-yilʕab
    FUT-be.IPFV.3SGM FUT-play.IPFV.3SGM
    (He) will be going to (will) play   [constructed]

- Any cases of bare Imperfective form of the auxiliary, as these occur only in dependent clauses and therefore the expression of tense and aspect on them is defined by the clause which they are part of, and so they were excluded from the data. An example is the case where the bare imperfective form of the auxiliary *yikūn* occurs as the complement of modals, as in:

(74) ʔaħmad momken ykūn   taʕbān
    Ahmed   might   be.IPFV.3SGM tired.PTCP.SGM
    Ahmed might be tired

- Cases where the sentence has an extra modal interpretation, where it refers to an imaginary situation and therefore can not refer to a certain point in terms of tense/aspectual marking, such as:

(75) kont   ʔaʔros   nafsi   ʔashal?
     be.PV.1SG pinch.IPFV.1SG myself easier?
     Would it've been easier if I just pinched myself?

This sentence comes in the form of a question, where the speaker is wondering what would have happened in an imaginary situation, which adds a modal interpretation to it. Due to the complexity of adding the modal function, where the form *yikūn* is not adding any temporal information to the phrase, such sentences were excluded from the analysis.

Another important note to consider about the data is that only a subset of the possible combinations were actually adressed in previous studies of ECA. Where for example Jelinek (1981) and Eisele (1990b) mentioned only a subset of the forms tackled in the current account. A comparison between these studies and the current account is provided in table 2.5 below.

Auxiliary Form	Verb Form	Current Account	Jelinek (1981)	Eisele (1990b)
Perfective	Perfective	✓	✓	✓
Perfective	bi-prefixed	✓	✓	-
Perfective	bare Imperfective	✓	-	-
Perfective	ħa-prefixed	✓	✓	✓
bi-prefixed	Perfective	✓	-	✓
bi-prefixed	bi-prefixed	✓	-	-
bi-prefixed	bare Imperfective	-	-	-
bi-prefixed	ħa-prefixed	✓	-	?
ħa-prefixed	Perfective	✓	✓	✓
ħa-prefixed	bi-prefixed	✓	✓	-
ħa-prefixed	bare Imperfective	-	-	-
ħa-prefixed	ħa-prefixed	-	*	?

Table 2.5: Comparison between accounts

In table (2.5), the symbol (✓) refers to forms that occurred, (-) refers to forms that did not occur, (?) refers to forms mentioned as questionable, while (*) refers to forms mentioned as ungrammatical. These differences between studies indicate that there is no full account of the whole range of possible combinations of tense and aspectual marking on verbs in ECA, and

while different studies might agree on some common structures as being used, they still differ
as to the possibility of many others, and this is where the account presented in the current
study proves useful. Analysis of these structures is another area in which previous studies differ
greatly, and this will be the focus of section 2.7 below.

2.7   Analysis

This section presents analysis of the findings of the current chapter based on the data discussed
previously and relating it to the literature of tense and aspect in ECA. It starts by setting the
base for marking of the auxiliary and lexical verb, followed by section 2.7.2 in which the tense
and aspect notions are re-defined with respect to data of ECA. Then the *bi*-prefix and the status
of the auxiliary *kān* are re-visited in order to reach some conclusions on their role in defining
morphosyntactic tense and grammatical aspect in ECA. Finally, an LFG analysis of the verbal
constructions is presented in the form of the lexical entries, f-structures as well as c-structures
of some example sentences.

2.7.1   Marking of Auxiliary and verb

"In any verb group, formal tense is marked only once, and always appears on the first verb in
the sequence"(Kearns, 2011, p.177). Despite the fact that this statement is mentioned in Kearns
(2011)'s account of the English verb group, it applies to the current analysis of ECA where the
main verb occupies *I* while its verbal complement occupies *V* [21]. Where in simple verb forms,
the verb carries the tense feature, occupying I. While in compound tense forms, the temporal
marking of the auxiliary functions as the tense marker of the whole verbal sequence, while
the following lexical verb preserves marking of grammatical aspect, in which case the auxiliary
occupies I while the lexical verb occupies V. This follows from Eisele (1990b)'s account of

---

[21]This assumption also extends to apply to modals in the current study, which despite the fact that they do
not carry tense, they serve as the main predicates and carry the modal property of the sentence and therefore
occupy *I*. This is further described in section 5.4.4.

compound tenses as: "the embedding of simple verb forms under *kān* is not indicative of their not being 'tenses', but rather of a syntactic fact: that CA (and Arabic in general) allows more than one finite verb per clause. This goes for compound tenses as well as for 'raising' verbs. In such cases, the higher or embedding verb serves to indicate the deictic time reference of the sentence" (Eisele, 1990b, p.187). This assumption makes it possible to account for the compound tense verbal combinations where the auxiliary and verb differ in tense and aspectual markings but can still give the intended interpretation, such as the case of future perfect where the auxiliary has the *ħa-* prefixed Imperfective form while the verb is in the Perfective from, resulting in **future perfect**. This is the position adopted in Sharif and Sadler (2009) stating that: "Notice that compound verbs may involve the combination of Perfective form and Imperfective form verbs. No feature clash results because the Perfective/Imperfective distinction is one of morphological form rather than f-structure feature content: as we have seen, a Perfective form verb conveys distinctions of tense when it occurs initial in the verbal sequence, and conveys distinctions of aspect when it is non-initial" (Sharif and Sadler, 2009, p.20).

2.7.2 Defining the notions for ECA

This is the step where the notions introduced earlier regarding tense and aspect markings in ECA are revised against the data to see if they are accounted for accurately, this is a preliminary step that will form the base for the analysis of these structures and then their implementation within XLE, presented in details in chapter 6. Following the treatment of tense and aspect in Butt et al. (1999), each of the dimensions of tense and aspect are treated separately so as to express the overt distinctions expressed in the language with no attempt to provide a semantic description, where such semantic information can be represented in a separate level distinct from f-structure. A set of notions denoting the marking of morphosyntactic tense and grammatical aspect were used for that puropse, and shown as follows:

2.7.2.1   Tense distinctions

The feature TENSE encodes the distinctions past (PAST), present (PRES) and future (FUT), following the analysis in Butt et al. (1999). Eisele (1990b) refers to this feature as deictic temporal reference whose interpretation depends on the context of utterance including speech time, and it can be expressed through verbal tense (grammatical means), time adverbials (non-grammatical means) as well as pragmatic or contextual devices. It is important to note here that this study is concerned only with grammatical means to express these distinctions, represented through morphological marking on the main verb in simple tense and on the auxiliary in compound tense.

Although most of the previous studies on tense and aspect in Arabic generally ignore these tense distinctions, and especially the future tense, the current study maintains these distinctions as they are all morphologically represented on the verb (for example with *ħa-* prefixed on verbs unlike the use of 'will' in English). In Reichenbach's terms, tense is concerned with the relation between R and S differentiating between tenses in the following way (Binnick, 1991):

- PAST denotes the case where R preecdes S (R<S), and is associated with the Perfective form.

- PRESENT denotes the case where R coincides with S (R=S), and is associated with the *bi-* prefixed Imperfective form.

- FUTURE denotes the case where R follows S (S<R), and is associated with the *ħa-* prefixed Imperfective form.

A note to make here is that these forms are standardised for any verb that occurs in I and carries tense, therefore the feature TENSE is associated with any verb in I, and it can have one of three values: PAST, PRESENT, or FUTURE. These tense distinctions on the verb in I will be shown later in the lexical entries of the verbs, and will be implemented in the XLE

grammar in chapter 6. The only exception is the *bi*-prefixed Imperfective auxiliary *bikūn* which gives present tense with an added generic meaning, that refers to the event not taking place at speech time, but rather at different points in time. Grammatical aspect is the focus of the following section.

2.7.2.2   Aspect distinctions

In Reichenbach's terms, aspect is concerned with the relation between R and E. Where E precedes R in anterior tenses, E and R coincide in simple tenses, and E follows R in all posterior tenses (Binnick, 1991). Eisele refers to this feature as 'non-deictic' temporal reference, which is a "verbal non-deictic time reference, functions as an embedded time reference" (Eisele, 1990b, p.182).

The feature ASPECT is marked on the lexical verb in V, while it is also expressed on bi-prefixed Imperfective verbs which always have the habitual and/or progressive interpretations, whether in I or V. Expression of morphosyntactic tense and grammatical aspect on bi-prefixed verbs will be further tackled in more details in section 2.7.3. Note that the lexical verb in V can occur as a complement to one of many verbs in I, including the auxiliary *kān* (described in the current chapter), the causative, phasal verbs and modals, which are covered in subsequent chapters of this study. Lexical entries for all of the verbs in I and V are represented in the XLE grammar implementation in chapter 6.

The feature ASPECT can carry one of the markings perfect (PERF)[22], continuative (PROG) or prospective (PROSP), following the analysis in (Sharif and Sadler, 2009). This distinction is adopted in the current analysis, with the addition of the habitual aspect (HAB), which although it is often represented by the same form as the progressive, the context where the verb occurs makes the distinction clear, as well as the fact that stative verbs can only be marked for the habitual aspect. The aspectual feature imperfect (IMPF) was also used in the current analysis

---

[22]Note that the term 'perfective' written in lower case, indicates the grammatical aspect itself rather than merely the form as in 'Perfective'.

as it is the standard term used in most of the previous studies. However in previous studies it is used as a cover term for progressive, habitual and prospective aspects, while this study differentiates between these different aspects as shown below:

- PERF refers to perfective aspect where E precedes R, this is also refered to as anterior aspect in Binnick (1991), and is associated with the Perfective form.

- The bare Imperfective form carries the aspect IMPF where E corresponds to R. This refers to the use of this form as the lexical verb to denote a state or characteristic that holds at the tense specified by the auxiliary. Although the bare form is assumed to be non-finite in the current study, defining this IMPF feature in compound tense forms is important in order to identify the imperfective aspect associated with the use of this verb in:

(76) kont   teʕraf-o   maʕrefa   ʃaḥseyya?
     be.PV.2SGM know.IPFV.2SGM-him knowledge personal?
     Have you known him personally?

Where if the bare verb did not carry any aspectual feature, then the morphosyntactic tense marking of example 76 would be 'past', which yields inaccurate results, so it is important to differentiate the 'past imperfective' in example 76 from the simple past in:

(77) ʕereft-o   ʔemta?
     know.PV.2SGM-him when?
     When did you know him?

Therefore this seems to be a case of syncretism between the non-finite bare Imperfective form[23] and the bare Imperfective form used in compound tense to denote imperfective aspect. An overview of the bare Imperfective form verb and the contexts in which it occured throughout the current study is presented in section 7.2.3.

---

[23]Discussed in details in section 1.4 and occuring as a verbal complement to the causative verb, phasal verbs and some modals throughout this study.

- The *bi*-prefixed Imperfective verb form can be used to mark either habitual (HAB) or progressive (PROG) aspects, and the context where it occurs serves to specify the intended aspect. In both aspects, R coincides with E, but the difference lies in defining the point/s at which the reference time is situated. Where in both cases E is seen a continuum of points in time, and R in HAB aspect refers to various points at which this event is happening, while R in PROG aspect refers to a certain point at which the event is taking place. *bi*-Imperfective stative verbs form a special case in that they can only mark habitual aspect as mentioned in a number of studies of tense and aspect in ECA, such as Eisele (1990b) and Abdel-Massih et al. (1979).

- PROSP refers to prospective aspect where R precedes E (R<E). This forms motivation for including it as a different aspect from the imperfect where R and E are simultaneous to each other and the event is seen as a continuum of points in time to which R refers (Eisele, 1990b). PROSP aspect is marked on *ḥa*-prefixed Imperfective forms.

2.7.3 The *bi*-prefix revisited

The *bi*-prefix has a crucial role in determining the temporal and aspectual features on the verbs. Where its absence makes the verb unable to carry tense and therefore it can be assumed to be a non-finite form of the verb, refered to here as the bare Imperfective form[24]. It also adds progressive and/or habitual aspects to the verb whether in I or V. These roles can be summarised as follows:

- The *bi*-prefixed auxiliary *bi-kūn* denotes present tense which is correlated with a generic or extended meaning, refering to a situation that can happen at various points in time rather than only at speech time. This might correspond to the perfect of 'persistent situation' which refers to "a state holding throughout an interval" (Ritz, 2012, p.883). This is the case mentioned in Eisele (1990b) as associated with a non-specific time reference that

---

[24]Detailed discussion of the bare Imperfective verb as a non-finite form in ECA is presented in section 1.4.

is used to mark a habitual reading or denote a characteristic. This is also manifested in Jelinek (1981)'s account of AUX as a syntactic category whose member can not be inflected for present tense, and that the imperfective of KWN[25] is a subjunctive form that does not appear in AUX (Jelinek, 1981). Therefore, *bikūn* does not mark tense in the same way as *kān* and *ħay-kūn*, where it adds a habitual reading that fails to place the event at a specific point in time, which is often correlated with a subjunctive meaning. The modal interpretation correlated with this structure and refered to as subjunctive (Abdel-Massih et al., 1979) might be a function of the semantic interpretation of the form, but this notational issue would not be adressed more here.

- *bi*-prefix attaches to the main verb (in I) in simple tense forms to denote present tense as well as adding aspectual features of progressive and/or habitual. If the verb is a stative verb then only the habitual interpretation is possible, while for non-stative verbs, both habitual and progressive interpretations are possible. These multiple interpretations lead to the *bi*-prefixed Imperfective form being ambiguous, which is an issue tackled in more details later in chapter 6 which includes XLE implementation of this form.

- *bi*-prefix attaches to the lexical verb following the auxiliary (in V) to add the habitual and/or progressive features, where a stative verb denotes only habitual aspect while a non-stative verb denotes both habitual and progressive aspects.

The *bi*-prefix therefore is very important in allowing the verb to carry tense, where its absence yields a bare Imperfective verb form which fails to have any tense marking, and this might be the reason none of the sentences in the data had this bare Imperfective form as the main verb for a non-dependent clause, where it occured only in dependent clauses or as a verbal complement. The *bi*-prefix is also unique in its ability to mark both tense and aspect on main verbs in simple tense forms. This might also explain why the *bi*-prefixed auxiliary has this peculiar interpretation of a generic state or characteristic, where it might be attributed to the

---

[25]KWN in Jelinek (1981) corresponds to the auxiliary *kān* in the current study.

habitual aspect added to the auxiliary (as it is a stative verb) although auxiliaries are supposed to carry tense only. Section 2.7.4 below discusses characteristics of the auxiliary *kān* focusing on its syntactic analysis in terms of the LFG syntactic theory.

2.7.4 Status of the Auxiliary

Auxiliaries are a class of verbs that are considered controversial in many languages as to whether they should be treated differently from 'main verbs' and how this analysis should look like. In LFG theory, there are generally two possiblities for analyzing an auxiliary, for one of them the auxiliary is a distinct predicate which subcategorizes for a SUBJ and XCOMP and so it is treated as a raising verb, while in the other analyses it is considered just an element specifying a feature such as tense or aspect, and therefore it does not behave as a predicate and would not have a subcategorization frame (Butt et al., 1999). Adoption of the second approach means analyzing auxiliaries as having a flat f-structure that combines tense and aspect information across the verbs to form the complex tense, as well as hierarchical c-structure that can account for relevant syntactic properties (such as VP deletion or fronting) (Butt et al., 1999). In cases where there are restrictions on the forms in which consecutive auxiliaries and verbs can occur, the flat f-structure fails to state these restrictions as the auxiliary does not have unique subcategorization properties of its own, but is rather analyzed as part of the f-structure of the main verb (Butt et al., 1999). And this is the problem with analyzing *kān* in its different forms as a feature, as there are some restrictions on the combinations between different forms of the auxiliary and the following verb as in the lack of the combination between two *ħa-* prefixed imperfectives as shown in the analysis section above.

Falk (2008) uses LFG notation to analyze auxiliaries in English either through the 'aux-predicate' analysis where the auxiliary is the head subcategorising for the following verb phrase as its complement, or the 'aux-feature' analysis where the lexical verb is the head and the auxiliary is just a marker for the tense feature. Different auxiliaries are analysed differently according to

their syntactic behaviour, where 'will' and 'progressive be' are analyzed as aux-predicates while 'do', 'have' and other modals such as 'would' are analyzed as aux-features according to Falk (2008). A thorough investigation of the criteria on which these conclusions are based was followed in the current analysis, and the assumption is that the auxiliary *kān* shall be treated as an aux-predicate. Some of the criteria on which this assumption is based are as follows:

- Distinct modifiers. Falk (2008) assumes that in a biclausal structure, each of the auxiliary and verb can have distinct modifiers. This applies for *kān* in examples such as[26]:

(78) bokra,   el-rāgel   ħa-ykūn   gah   embāreħ
      tomorrow, the-man FUT-be.IPFV.3SGM come.PV.3SGM yesterday
      Tomorrow, the man will have already came yesterday   [constructed]

The interpretation of this sentence assumes that tomorrow, the man will be in a state that he had already came the day before. The point here is to show that the past tense verb can be modified with a compatible adverb as it occupies a different clause from the auxiliary which can have its own modifier.

- Different interpretations. One of the criteria on which 'do' is analyzed as a feature is that it seems to have the same interpretation as the synthetic form of the perfective verb (where 'took'='did take' in terms of tense interpretations). But this is not the case for the auxiliary *kān*, where compound tense forms give different interpretations from their simple counterparts, even in the case of succession of the perfective auxiliary marking past tense and the bare imperfective which has no tense marking. The difference is shown in the examples below:

(79) ʔana ʔilli ʔaʕadt   hena el-ʔawwel
      I   who   sit.PV.1SG here the-first
      I'm the one who sat here first!   [constructed]

---

[26]This example is constructed based on what is presented in Falk (2008) as evidence for progressive BE in English having an aux-predicate analysis.

(80) nefs-i   kont   ʔaʕod   henāk
      wish-my be.PV.1SG sit.IPFV.1SG there
      I wish I had sit there   [constructed]

- The auxiliary *kān* is predicative in the sense that it contrasts with other verbs such as the phasal verbs 'begin' and 'remain' in specifying the relation between the SUBJ and the verb. This goes in accordance with the analysis in Falk (2008) of the progressive BE in English where it is assumed to be equivalent to the main verb BE in expressing the existence of state where the verb occupies a V-to-I position. The difference is shown in the following examples:

(81) Mona badaʔit   tezāker   kowayyes
      Mona begin.PV.3SGF study.IPFV.3SGF well
      Mona started to study well   [constructed]

(82) Mona kānit   bi-tzāker   kowayyes
      Mona be.PV.3SGF BI-study.IPFV.3SGF well
      Mona was studying/used to study well   [constructed]

Note also that the phasal verb 'begin' requires the following lexical verb to be in the bare Imperfective form, while the auxiliary *kānit* is followed by a *bi*-prefixed verb. More discussion on the phasal verbs and their sub-categorisation patterns is the focus of chapter 4.

- Negation. The assumption is that if the auxiliary and the verb are different predicates then each could be negated separately giving different interpretations. This is a rather tricky test, as by applying the test to examples 83 and 84 below, both sentences seem to negate the same situation, however the difference is that the second one can give the additional interpretation of the person being in 'the state of' not eating meat, as a property or characteristic of that person such as being a vegetarian.

(83) ʔana ʕāref   ʔinn-ak ma-kont-eʃ   bi-tākol   laħma
      I   know.PTCP.SGM that-you NEG-be.PV.2SGM-NEG BI-eat.IPFV.2SGM meat
      I know that you weren't eating meat

(84) ʔana ʕāref   ʔinn-ak   kont   ma-b-tākol-ʃ   laħma
     I   know.PTCP.SGM that-you be.PV.2SGM NEG-BI-eat.IPFV.2SGM-NEG meat
     I know that you were not eating meat

Adopting the aux-predicate analysis for the auxiliary *kān* as the analysis of progressive BE in

Falk (2008) implies that XCOMP is the only thematic argument for the auxiliary, while SUBJ

is a non-thematic argument. Another point to note here is that further evidence need to be

established as to whether *bikūn* should be treated differently from *kān* and *ħa-ykūn* due to the

peculiar way in which it defines tense for the verbal sequence, as well as the fact that it is more

restricted in the forms with which it occurs.

In terms of the syntactic annotation, tense is always expressed on verbs in I, whether it is the

auxiliary or the lexical verb, while the verb in V expresses aspect only. The only special cases

are:

- *bi*-prefixed Imperfective verb forms, in which the grammatical features habitual and/or

  progressive can be marked on the verb in I in addition to present tense. This is accounted

  for in the lexical entries of the verbs using the features PROG and HAB. Using these

  binary features as an addition to the aspect value IMPF (in V) aims to resolve the

  ambiguity arising from having two different values (PROG/HAB) for the same feature

  of grammatical aspect. For example the bi-prefixed verb *bi-yiktib* 'writes/writing' would

  have the following features of grammatical aspect included in its lexical entry:

  ASPECT = IMPF

  HAB = +

  PROG = +

  Otherwise, if HAB and PROG are given as values for the feature of grammatical aspect,

  then the same verb *bi-yiktib* 'writes/writing' would have both values for the same feature,

as in: ASPECT=HAB/PROG which leads to ambiguity.

- Restrictions on occurence of the *bi*-prefixed *bikūn* which is used only in certain contexts refering to events that have no specific event time, as in example 62 which is repeated below:

(85) mīn ʔaktar lāʕib   bi-txāf   menn-o   lamma bi-ykūn
     who more   player BI-fear.IPFV.2SGM from-him when   BI-be.IPFV.3SGM
     bi-yilʕab   ʔoddām farīʔ-ak?
     BI-play.IPFV.3SGM infront   team-your?
     Who is the player from whom you're most afraid when he's playing against your

     team?

This is related to the peculiar interpretation of the compound tense forms having *bikūn* as a state or characteristic rather than merely denoting present tense. This is accounted for as an added habitual feature that is absent from all other forms of the auxiliary, where other forms of the auxiliary do not carry any markings for aspect. In this respect, *bikūn* can be assumed to not behave as an auxiliary, which is clear in its inability to substitute other forms of the auxiliary as shown in the examples below:

(86) ʔaḥmad kān   bi-yelʕab   gitār f-el-madrasa
     Ahmed be.PV.3SGM BI-play.IPFV.3SGM guitar at-the-school
     Ahmed used to play guitar at school   [constructed]

(87) ʔaḥmad ḥa-ykūn   bi-yelʕab   gitār bokra   f-el-nādi
     Ahmed FUT-be.IPFV.3SGM BI-play.IPFV.3SGM guitar tomorrow at-the-club
     Ahmed will be playing guitar tomorrow at the club   [constructed]

(88) *ʔaḥmad bi-ykūn   bi-yelʕab   gitār
     Ahmed   BI-be.IPFV.3SGM BI-play.IPFV.3SGM guitar

Therefore *bikūn* is treated as a verb form that occurs in I, which carries both present tense as well as habitual aspect, due to prefixation of bi- to *kān* which is a stative verb. The lexical entry for *bikūn* would specify a value for TENSE=PRESENT as well as adding a

feature HAB=+[27]. This might be the reason for the peculiar interpretation of *bikūn* as well as the restrictions on its use.

- Restriction on occurence of the bare Imperfective verb form as main verb, where it occured only in dependent clauses throughout the data on which the study is based. This is accounted for by having an entry for the bare verb form only as V, which does not carry tense but rather only specifies the feature (↑VFORM)=BARE in the lexical entry of the bare imperfective verb. This conforms with the analysis adopted for complementation patterns of other verbs in the current study that can only take verbal complements in the bare Imperfective verb form. This is represented in the XLE grammar implementation in chapter 6 where some verbs (such as the phasal *badaʔ* 'begin') can only be followed by a bare Imperfective verb form, and therefore specification of this feature BARE on the verb form is neccessary.

Having established the basis on which the analysis is based and the findings and discussion of the data, section 2.7.5 below presents the lexical entries for all of the verbs discussed in the current chapter, including different forms of the auxiliary and the lexical verb. These lexical entries will later on be implemented in the XLE grammar fragment of ECA, represented in chapter 6.

2.7.5   Lexical Entries

This section represents the lexical entry for each of the verbs and auxiliary forms analyzed earlier, focusing on the marking of tense and aspect on each and ignoring information such as person and number inflections, for ease of presentation, while all of these information are presented in the XLE grammar fragment, accounted for in chapter 6.

- Auxiliary

---

[27]The lexical entry is presented in section 2.7.5.

kān: I

(↑PRED)= 'be (↑ SUBJ)<(↑ XCOMP)>'

(↑ SUBJ) = (↑ XCOMP SUBJ)

(↑TENSE) = PAST

bi-kūn: I

(↑PRED) = 'be (↑ SUBJ)<(↑ XCOMP)>'

(↑ SUBJ) = (↑ XCOMP SUBJ)

(↑TENSE) = PRESENT

(↑HAB)= +

ħay-kūn: I

(↑PRED)= 'be (↑ SUBJ)<(↑ XCOMP)>'

(↑ SUBJ) = (↑ XCOMP SUBJ)

(↑TENSE) = FUTURE

- Perfective non-stative verb

  katab: I

  (↑PRED)= 'write <(↑ SUBJ)(↑ OBJ)>'

  (↑TENSE) = PAST

  katab: V

  (↑PRED)= 'write <(↑ SUBJ)(↑ OBJ)>'

  (↑ASPECT) = PERF

- Perfective stative verb

fehem: I

(↑PRED)= 'understand <(↑ SUBJ)>(↑ OBJ)'

(↑TENSE) = PAST

fehem: V

(↑PRED)= 'understand <(↑ SUBJ)>(↑ OBJ)'

(↑ASPECT) = PERF

- *bi*-prefixed non-stative verb

  bi-yiktib: I

  (↑PRED)= 'write <(↑ SUBJ)(↑ OBJ)>'

  (↑TENSE) = PRESENT

  (↑HAB) = +

  (↑PROG) = +

  bi-yiktib: V

  (↑PRED)= 'write <(↑ SUBJ)(↑ OBJ)>'

  (↑ASPECT) = IMPF

  (↑HAB) = +

  (↑PROG) = +

- *bi*-prefixed stative verb

  bi-ḥib: I

  (↑PRED)= 'love <(↑ SUBJ)(↑ OBJ)>'

  (↑TENSE) = PRESENT

  (↑HAB) = +

bi-ħib: V

(↑PRED)= 'love <(↑ SUBJ)(↑ OBJ)>'

(↑ASPECT) = IMPF

(↑HAB) = +

- Bare Imperfective verb

ʔaʕod: V

(↑PRED)= 'sit <(↑ SUBJ)>'

(↑ASPECT) = IMPF

(↑VFORM) = BARE

This specification on the VFORM of the bare imperfective verb is important in cases where the verb in I can only have complement verbs in the bare imperfective form, such as the case with phasal verbs and modals investigated in the later chapter of the current study.

- *ħa*-prefixed non-stative verb

ħa-yiktib: I

(↑PRED)= 'write <(↑ SUBJ)(↑ OBJ)>'

(↑TENSE) = FUTURE

ħa-yiktib: V

(↑PRED)= 'write <(↑ SUBJ)(↑ OBJ)>'

(↑ASPECT) = PROSP

- ħ*a*-prefixed stative verb

  ħa-ʔbal: I

  (↑PRED)= 'accept <(↑ SUBJ)>(↑ OBJ)'

  (↑TENSE) = FUTURE

  ħa-ʔbal: V

  (↑PRED)= 'accept <(↑ SUBJ)>(↑ OBJ)'

  (↑ASPECT) = PROSP

These lexical entries are implemented in the XLE grammar fragment in chapter 6 which includes the whole lexicon that is capable of producing the list of grammatical sentences that constitute the data on which this study is based. An important note to make here is that the rules for assigning tense and aspect where the verb in I carries tense while that in V carries aspect, will also apply to the other verbs which are tackled in the current study, namely the causative, phasal verbs and modals. The following section includes the f-structure of some example sentences representing tense/aspect in ECA.

Another alternative to the analysis provided here is including the morphological component, where the lexical entry would be in the form of the underlying lexeme to which morphemes can be added to create the lexical item. So for example the prefix ħ*a* will give the information **TENSE=FUT** or **ASPECT=PROSP**, and the lexical entry of the verb would be the lexeme **YEKTEB** to which the prefix ħ*a* can be added to create the verb **ħayekteb**. Although this might offer a fuller analysis in terms of the morphological components of the verbs, it also is faced with the complexity of the root and pattern of Arabic morphology, where the root of the word carries its main lexical information. For example the root **KTB** represents the lexeme **WRITE**, which can be either realized in the perfective form **KATAB** or the (bare) imperfective form **YEKTEB** to which affixes can be added. Since the current analysis is mainly

concerned with the expression of grammatical tense and aspect on the verbs concerned and how its syntactic structure can be represented in a grammar of ECA, such a morphologically detailed analysis was not used here.

2.7.6   F-structure

The f-structures below represent different ways in which tense and aspect are expressed on the verbs as well as different possible combinations between the auxiliary and lexical verb. However, these are only a sample of the sentences showing different simple and compound tense structures, while the whole list of f-structures are adopted in the XLE implementation presented in chapter 6, where the grammar fragment is capable of producing the full list of sentences. For each of the sample sentences below, the example is glossed followed by the f-structure, starting with present progressive/habitual, past habitual and finally past prospective.

(89) ʔel-walad bi-yekteb   el-gawāb
      the-boy   BI-write.IPFV.3SGM the-letter

      The boy writes/is writing the letter

$$
\begin{bmatrix}
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \textit{`boy'} \\ \text{NUM} & \textit{SG} \\ \text{GEND} & \textit{MASC} \end{bmatrix} \\
\text{PRED} & \textit{`write <SUBJ, OBJ>'} \\
\text{TENSE} & \textit{PRESENT} \\
\text{HAB} & + \\
\text{PROG} & + \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \textit{`letter'} \end{bmatrix}
\end{bmatrix}
$$

(90) howwa kān   bi-yħebb-ni
      he   be.PV.3SGM   BI-love.IPFV.3SGM-me

      He used to love me

$$
\begin{bmatrix}
\text{SUBJ} & _i & \begin{bmatrix} \text{PRED } PRO \\ \text{NUM } SG \\ \text{GEND } MASC \end{bmatrix} \\
\text{PRED} & \text{`be SUBJ <XCOMP>'} \\
\text{TENSE} & PAST \\
\text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{`love <SUBJ, OBJ>'} \\ \text{ASPECT} & IMPF \\ \text{HAB} & + \\ \text{SUBJ} & _i \\ \text{OBJ} & \begin{bmatrix} \text{PRED } PRO \end{bmatrix} \end{bmatrix}
\end{bmatrix}
$$

(91) konna   ħa-nmūt
     be.PV.1PL FUT-die.IPFV.1PL

     We were going to die

$$
\begin{bmatrix}
\text{PRED} & \text{`be SUBJ <XCOMP>'} \\
\text{SUBJ} & _i & \begin{bmatrix} \text{PRED } PRO \\ \text{NUM } PL \end{bmatrix} \\
\text{TENSE} & PAST \\
\text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{`die <SUBJ>'} \\ \text{ASPECT} & PROSP \\ \text{SUBJ} & _i \end{bmatrix}
\end{bmatrix}
$$

2.7.7   C-structure

The typical c-structure for sentences provided in the data is as follows:

(92) ʔel-walad katab   el-gawāb
     the-boy   write.PV.3SGM the-letter

     The boy wrote the letter

```
                              IP
                    ┌──────────────┴──────────────┐
                   NP                             I'
            (↑SUBJ) = ↓                        ↑ = ↓
                    │                  ┌──────────┴──────────┐
                    │                  I                    VP
               el-walad             ↑ = ↓                 ↑ = ↓
                                      │                     │
                                      │                    NP
                                    katab             (↑OBJ) = ↓
                                                           │
                                                       el-gawaab
```
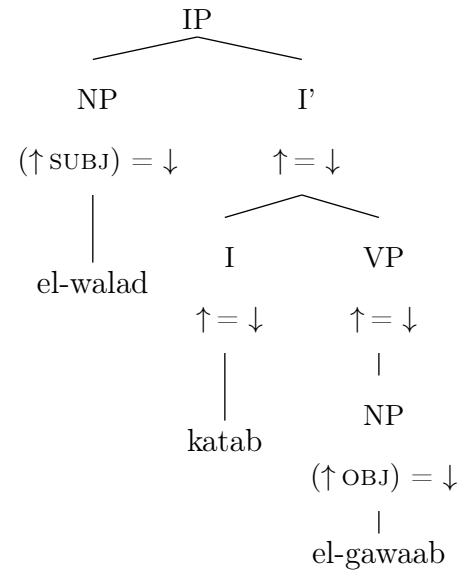
(93) Ɂaħmad kān   bi-yḵattat   el-hugūm
      Ahmed   be.PV.3SGM BI-plan.IPFV.3SGM the-attack

      Ahmed was planning the attack

```
                              IP
                    ┌──────────────┴──────────────┐
                   NP                             I'
            (↑SUBJ) = ↓                        ↑ = ↓
                    │                  ┌──────────┴──────────┐
                    │                  I                    VP
                 Ɂaħmad             ↑ = ↓              (↑XCOMP) = ↓
                                      │              ┌────────┴────────┐
                                      │              V                NP
                                    kān            ↑ = ↓          (↑OBJ) = ↓
                                                     │                │
                                                     │                │
                                                 bi-ḵattat        el-hugūm
```
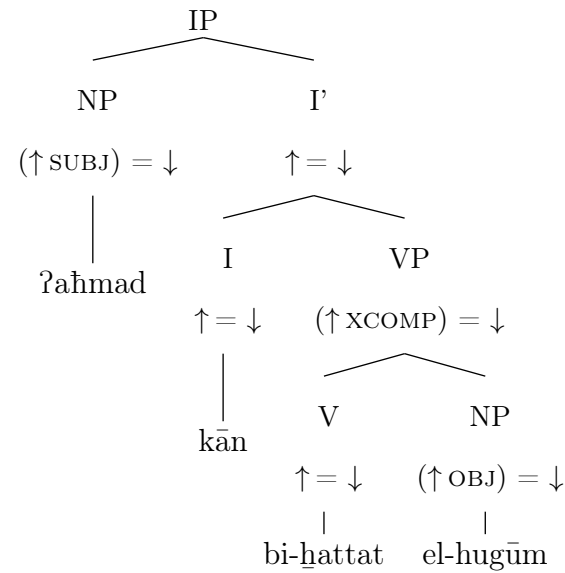
(94) ħa-nkūn   ennaharda bi-nelʕab   fi maʕhad gūta
      FUT-be.IPFV.1PL today   BI-play.IPFV.1PL in institute Goethe

      Today we will be playing in Goethe Institute

IP
|
I'
↑ = ↓

I   ADVP   VP
↑ = ↓   (↑ ADJ) = ↓   (↑ XCOMP) = ↓

                                    V   PP
ħankūn   ennharda
                              ↑ = ↓   (↑ ADJ) = ↓

                                              P   NP
                              b-nelʕab
                                          ↑ = ↓   (↑ OBJ) = ↓
                                            | |
                                            fi   maʕhad-gūta
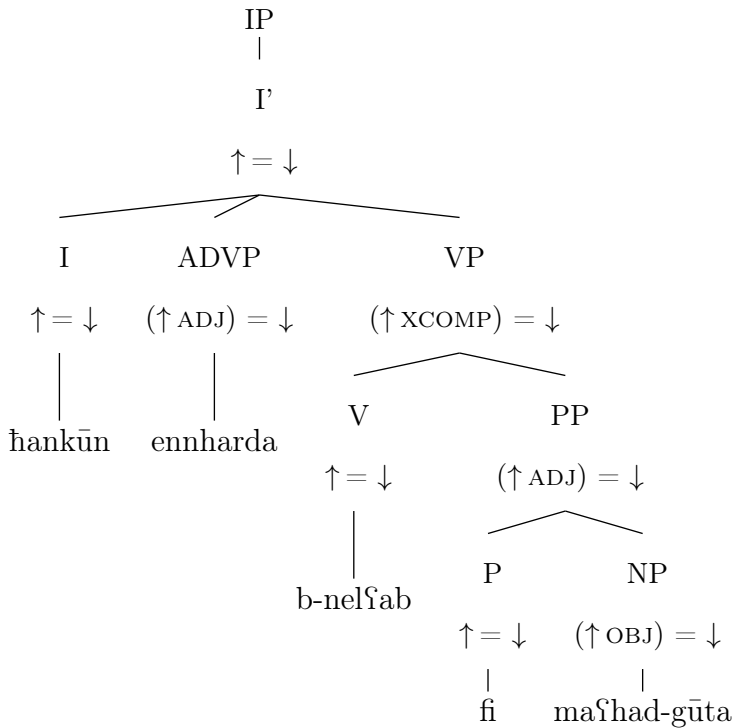
2.8   Conclusions

One of the main questions raised in this chapter was concerned with what are the possible forms that can be used to express tense and aspect in ECA, and specifically in compound tense where different forms of the auxiliary are combined with different forms of the lexical verb, so what are the possible combos? and what are the ungrammatical ones? and how to account for both? The current chapter attempted to answer these questions through presenting the corpus based data and analysing each of the sentences in order to describe how morphosyntactic tense and grammatical aspect are expressed on each verb form. This section therefore aims to present the conclusions about how tense and aspect are expressed on verbs in ECA.

In simple tense forms, the main verb occupies I and carries tense only, with the addition of the features HAB and/or PROG on bi-prefixed Imperfective verbs. While in compound tense, the lexical verb carries grammatical aspect only. Expression of morphosyntactic tense and grammatical aspect on the lexical verb in its different forms in I and V are shown in table 2.6 below:

Verb Form	I	V
Perfective	past	perfect
bare Imperfective	-	imperfective
bi-prefixed	present HAB+ PROG +	imperfective HAB+ PROG+
bi-prefixed (stative)	present HAB+	imperfective HAB+
ħa-prefixed	future	prospective

Table 2.6: Summary-Verb

Specifying that either morphosyntactic tense OR grammatical aspect are expressed on the verb, but not both, has the advantage that it avoids the confusion resulting from having both tense and aspect expressed on the same form as well as showing the difference between simple and compound tense, as in:

(95) 2aħmad ʔakal
     Ahmed   eat.PV.3SGM

     Ahmed ate   [constructed]

Which marks simple past, specifying that an event happening at a time prior to speech time, where E<S. This is different from the following:

(96) 2aħmad kān   ʔakal   ʔabl   el-ħafla
     Ahmed   be.PV.3SGM eat.PV.3SGM before the-party
     Ahmed has eaten before the party   [constructed]

Which marks past perfect, where R<E<S. Assuming that the simple verb carries both tense and aspect would yield the same result (past perfect) in both cases, which is wrong. Note also that in table 2.6 the bare Imperfective verb form has an empty slot in I, this is due to the fact that there were no examples in the data on which the study is based where the bare Imperfective verb occured as the main verb, where its occurence is correlated with modal or non-indicative uses, and therefore often assumed to be a non-finite form of the verb. More disussion of the bare Imperfective verb as a non-finite form in ECA in section 1.4.

On the other hand, *kān* is the auxiliary which marks tense in compound tense forms in ECA, it occupies I while the lexical verb occupies V. Table 2.7 below shows the different forms in which the auxiliary *kān* expresses morphosyntactic tense in compound tense forms:

Aux Form	I
Perfective	past
bare Imperfective	-
bi-prefixed	present HAB+
ħa-prefixed	future

Table 2.7: Summary-Aux

Note that the auxiliary in I does not occur in the bare Imperfective verb form, as this form lacks marking of morphosyntactic tense, and therefore contradicts with the function of the auxiliary as tense carrier in compound tense forms. Combining the different forms and morphosyntactic marking of the auxiliary in I (listed in table 2.7) with the different forms of the lexical verb in V (listed in table 2.6) results in producing 11 possible compound tense structures (listed previously in table 2.4). Similarly, ungrammatical structures would be impossible to produce due to the lack of required morphosyntactic tense/ aspect marking on the auxiliary/verb. For example the bare Imperfective form auxiliary can not combine with any lexical verb to form compound tense structures, because the auxiliary does not carry any tense marking in the bare Imperfective form. This is accounted for in more details in chapter 6 where the grammar fragment is capable of producing all and only the possible compound tense combinations. On the other hand, some combinations had no examples in the corpus data although they seem to be grammatical, such as the future prospective form:

(97) #ħa-ykūn   ħa-yeʕmel
      FUT-be.IPFV.3SGM FUT-do.IPFV.3SGM

      will be going to do   [constructed]

Although this combination did not occur in the corpus data, it did not seem ungrammatical to me (as a native speaker). Therefore, there seems to be some other factor defining which

forms of the auxiliary and verb can be combined together rather than just the form of the verb marking or its tense/aspectual interpretation, which might be an interesting point to tackle in future research.

The different forms of compound tense in ECA has not been covered fully in any of the previous studies, such as the case in Jelinek (1981) and Eisele (1990b) who agree on the occurence of only a subset of these forms, shown in table 2.5. And this is where the current account proves useful as to showing the whole range of possible combinations as well as examples and interpretation of each. Regarding forms of the auxiliary, *yikūn* is the auxiliary mentioned in Abdel-Massih et al. (1979) to add a subjunctive meaning, *bikūn* can also be correlated with such a modal interpretation in the examples analyzed here. It is not yet clear whether this meaning is related to the overall structure of the sentence (where they often appear in conditional sentences) or if it is an interpretation inherent in these verb forms carrying a modal marking. This a point that needs further investigation in future studies.

A final point to make here is concerned with the *bi-* prefix[28], which is a rather important feature in specifying tense and aspect in ECA as it fulfills the following functions:

- Marks present tense on main verbs in I, plus a HAB feature on stative and both HAB and PROG on non-stative verbs.

- Marks present tense on auxiliary plus a HAB feature, which adds a meaning of an event that happens at different points in time to the compound tense structure (as something always happening).

- Marks imperfective aspect on verbs in V, plus a HAB feature on stative verbs and both HAB and PROG on non-stative verbs.

- Bare Imperfective verb forms which lack the *bi-* prefix can not occur in I, as the tense

---

[28]Discussed in details in sections 2.3.1 and 2.7.3.

feature will be missing. Therefore bare forms occur only in dependent clauses and are marked by (↑VFORM)=BARE as a standard in anlaysis of all verbs in the current study.

- Absence of the *bi-* prefix on verbs in V marks imperfective aspect, with no HAB or PROG features. Which gives only the meaning of the state or characteristic denoted by the verb. This feature might explain why the use of this form seems to be more restricted than the *bi*-prefixed form in compound tense, where it only occurs following perfective forms of the auxiliary. It is also judged to be the marked option to use, where the unmarked form is the *bi*-prefixed form.

(98) kont   b-anām   badri
     be.PV.1SG BI-sleep.IPFV.1SG early
     I used to sleep early

This clearly describes a habit, which can be more clear through adding the adverbial 'everyday', as in:

(99) kont   b-anām   badri kol   yōm
     be.PV.1SG BI-sleep.IPFV.1SG early every day
     I used to sleep early every day

While the bare Imperfective verb is questionable in this context:

(100) ??kont   ʔnām   badri
       be.PV.1SG sleep.IPFV.1SG early
       I (was) sleep early

Insead, it might be used in a modal context, refering to an event that might or might not happen, as in:

(101) kont   ʔnām   badri yaʕni?!!
       be.PV.1SG sleep.IPFV.1SG early
       Should I've slept early?!!

Chapter 3

Causative *ḫalla*

Talmy (2000, p.481) defines a basic causative situation as one that "consists of three main components: a simple event (that is, one that would otherwise be considered autonomous), something that immediately causes the event, and the causal relation between the two." Therefore, the core arguments in a causative construction include a causer, which is the agent of the predicate of cause , and a causee, which is the agent of the caused event (Payne, 1997, p.176). This causer can refer to "a person (e.g. 'Mary made me laugh') or an abstract thing (e.g. 'The heat of the sun made me feel dizzy') or 'John's lecture made me feel sleepy') or an event, coded through a complement clause (e.g. 'Walking all day made me tired')(Dixon, 2000, p.32). However, Talmy (2000) argues that in such cases where the event is caused by a 'force' such as 'the wind' in 'The wind broke the window', it is better to account for the situation as a case of 'event causation' rather than agent or instrument. Accordingly, the underlying structure of the causative situation would consist of a 'causing event' and a 'caused event'.

In spite of the importance of these notions in defining the causative situation, they are mainly concerned with the semantic concept underlying causation, and therefore will not be further tackled here. However, the current chapter is concerned with the ways in which these causative situations are expressed using the syntactic means in ECA. Starting by section 3.1 below which

97

discusses the types of causative constructions across languages, followed by section 3.2 which focuses on *ḥalla* as the verb used in ECA for expression of analytic causatives. Section 3.3 presents a number of studies in the literature of causatives within the domain of syntax, whereas section 3.4 presents a description of the corpus data for the causative verb, on which the study is based. Section 3.5 goes on to discuss in details the different forms of the causative verb and its behaviour based on the corpus data discussed in terms of the different verb forms, the arguments it can have, its negation patterns as well as investigation of whether the causative construction forms a mono or bi-clausal structure. Analysis of the causative construction within the syntactic theory of LFG is presented in section 3.6 along with the lexical entry of the verb as well as the f-structure and c-structure of the sentences analysed. And finally section 3.7 includes the conclusions reached about the causative construction in ECA.

3.1   Types of causative constructions

Causative constructions can be divided into three types: lexical, morphological/synthetic and periphrastic/analytic causatives (Payne, 1997; Hoffman, 1991). Lexical causatives are verbs which include the cause predicate in their lexical meaning, such as the English verbs 'feed' and 'kill' (Payne, 1997). While morphological causatives are formed through a change in the form of the verb, such as the case in Standard Arabic where the causative can be formed through ablaut (changing the stem vowel of the verb to /a/) or gemination of the middle consonant of the root, as in *ḥariba* 'be destroyed' for which the causative is either *ḥaraba* 'destroy' or *ḥarraba* 'destroy' (Hallman, 2006). Analytic causatives, on the other hand, are expressed through a separate causative verb whose complements refer to the caused event, such as the English verbs 'make' and 'cause' (Payne, 1997).

A further distinction is made between causative constructions formed of two verbs, that by (Dixon, 2000) where they can either be formed of two verbs behaving as a single predicate

(refered to also as an analytic causative), or a periphrastic causative which is formed of two verbs in two different clauses, where the causative verb occupies the main clause and the lexical verb is in its complement clause [1]. However, the term 'analytic causative' is used more widely in the literature where it refers generally to a causative construction formed of two verbs irrespective of the number of predicates they form. An example is Comrie (1989)'s definition of the analytic causative as: "The prototypical case of the analytic causative is where there are separate predicates expressing the notion of causation and the predicate of the effect" (Comrie, 1989, p.167). Therefore the term 'analytic causative' will be used to refer to the causative construction formed of _ḫalla_ followed by a lexical verb in the current study, as opposed to 'synthetic (morphological) causative'. The status of the _ḫalla_ analytic construction as either mono or bi-clausal will be further investigated in section 3.6.1, where the results of the data description will be applied in order to reach the best analysis for the causative construction in question.

Morphological/synthetic causatives are formed in ECA by a productive rule of gemination of the second consonant of the root of the verb to form second binyan (also called form II) verbs which represent the causative of their corresponding first binyan (form I) verbs (Watson, 2002). This is the same strategy used in Modern Standard Arabic for morphological causatives in addition to Ablaut, while the latter is not used in ECA. Examples for synthetic causatives in ECA are shown below, where example (1) shows the synthetic second binyan causative, while example (2) shows its corresponding first binyan verb:

(1) ʔana saḥḥēt   el-walad
    I   cause.wake.up.PV.1SG the-boy

    I woke the boy up

(2) ʔel-walad siḥi
    the-boy   wake.up.PV.3SGM

    The boy woke up

---

[1]Criteria on which this distinction is based will be tackled in more details in section 3.3.

A similar meaning to that of (1) can be expressed by an analytic form of the causative construction (also called 'syntactic form') using causative *ḥalla*, as in:

(3) ana ḥa-ḫalli   el-walad yisḥa
    I   make.FUT.1SG the-boy wake.up.IPFV.3SGM

    I will wake the boy up

Despite the fact that both the morphological and the analytic causative constructions seem synonymous, this study is concerned only with analytic causatives, focusing on how the causative verb *ḫalla* is used to form causative constructions in ECA. Section 3.2 below is concerned with description of the ECA verb *ḫalla*.

3.2   *ḫalla*

This chapter is concerned with the complementation patterns of the causative verb *ḫalla*, therefore only the analytic form where causative *ḫalla* is used will be considered. This use of *ḫalla* corresponds to the causative English verb 'make' in its syntactic use, where it marks the causative meaning in structures such as 'I made him cry'. Dixon (2000) defines a prototypical causative construction as one that involves "a morphological process, or a verb which only has an abstract, causative meaning (or a lexical pair whose members are in causative relation). In English, *make* only has causative meaning while *order* also refers to an act of speaking. In view of this, *Mary made John go* is treated as a causative construction, but *Mary ordered John to go* is not." (Dixon, 2000, p.32). In English, 'make' is the more prototypical causative verb than 'cause', which implies indirect causation and is less common than 'make' (Dixon, 2000). The causative use of verbs like the English 'make' is a common way of forming causative constructions in different languages of the world (Moreno, 1993).

An important note regarding *ḫalla* is that it can be used to express both permissive and causative meanings, as in:

(4) keda   enti ħa-t̲ẖalli   el-rāgel yidrab-ek
    this.way you make/let.FUT.2SGF the-man hit.IPFV.3SGM-you

    This way you'll make the man hit you! OR This way you'll let the man hit you!

Both meanings are possible with the same structure under different contexts, but the only difference in this respect is that only the permissive meaning arises if we substitute ẖalla with sāb, as in:

(5) keda   enti ħa-tsībi   el-rāgel yidrab-ek
    this-way you leave.FUT.2SGF the-man hit.IPFV.3SGM-you

    This way you'll let the man hit you!

This phenomena is not unique to ECA, as it is attested in many languages where the same periphrastic causative construction can carry either causative or permission granting meanings (Shibatani, 1973). The permissive use can also be correlated with a hortative mood[2] (denoting a wish) and especially when it occurs in the second person, as in:

(6) ẖalli   ʔel-ʔayyām temʃi   tabīʕi
    let.IMP.2SG the-days   walk.IPFV.3SGF normal

    Let life pass normally

The verb behaviour in terms of tense and aspect and negation patterns is the same for both causative and permissive meanings, and therefore no distinction will be made in this work for those meanings. However, the focus will be on describing causative ẖalla only.

3.3   Literature review

Causativisation is often studied as a valency changing derivation, which prototypically applies to intransitive verbs (and transitives as well) to derive transitives, introducing a new causer

---

[2]The term 'hortative mood' is used here only descriptively to account for the meaning of the sentence and showing this unique use, while I am not attmepting to refer to any semantic or syntactic properties relating to this term in ECA.

argument (Dixon and Aikhenvald, 2000). And although analytic constructions are attested in a number of languages, their frequency of occurence cross-linguistically is relatively low (Comrie, 1989). These might be the reasons why morphological causatives have much more studies than analytic causatives, as is stated by Stefanowitsch (2001): "Analytic causatives have not received much attention in the literature, compared with the vast literature on morphological and lexical causatives".

This applies to the study of analytic causatives in Arabic, where the studies in the literature are mainly concerned with morphological causatives in Arabic. For example, in Benmamoun (1991)'s study of Causatives in Morroccan Arabic the focus is only on morphologically-derived causatives that are formed through attaching a causative affix to the verb in the form of gemination of the second consonant of the root (as the case in ECA, shown in example 1). Another study is Fehri (2012)'s extensive study on Arabic grammar in which only morphological causatives are mentioned, where causativisation is a result of root-based and morpheme-based reduplications. On the other hand, none of the studies on the grammar of ECA mentioned the analytic causative construction, hence comes the importance of the current study. Due to the lack of studies on causative *ḥalla* in Arabic, the rest of the literature review is devoted to studies concerned with analytic causatives in different languages, and how they are treated in terms of syntactic structure.

In his study of analytic causatives within a construction grammar approach, Stefanowitsch (2001) describes the structure of analytic causatives as consisting of a finite matrix verb expressing the causation event, and taking the arguments subject (representing the causer), object (representing the causee) as well as an embedded non-finite verb phrase. Properties of analytic causatives include their ability to occur with any kind of verb or construction, and that the subject "typically encodes the agent of an event prior to the actual causing event on the causation chain" (Stefanowitsch, 2001, p.36). This difference can also be explained within the

'Iconicity Principle' defined by Escamilla (2012) as "a correlation between the degree of formal compactness of the linguistic material encoding the causative macroevent and the perceived directness of the relationship between causing event and caused event: i.e., shorter forms, on the whole, were posited to encode more direct causation than longer forms, as in the classic English examples *I killed him.* [direct causation] vs. *I caused him to die.* [less direct causation]." (Escamilla, 2012, p.5). This difference affects the choice of the causative used in Hupa (analyzed in Escamilla (2012)'s work) where the syntactic causative construction is used in cases where the causee is preceived as having some degree of control over the caused event (Escamilla, 2012). However, these differences require a detailed semantic comparison between synthetic and analytic causatives which is beyond the focus of the current study, where only the syntactic structure of analytic causatives in ECA will be described. It is important here only as a motivation for not using any of the literature on morphological causatives in Arabic, as it differs in many aspects of its structure from the analytic causatives on which this study is based.

Another more relevant distinction is that between analytic and periphrastic constructions, which is defined by Dixon (2000, p.34) in the following criteria:

- Causatives formed of two verbs in one predicate. Where the causative and the lexical verb are in the same clause, and they share the same features such as polarity as well as tense and aspect. This could either be a serial verb construction as in Tariana language where both verbs share the same subject, or cases where the two verbs form one predicate as in the French causative verb 'faire', where the lexical verb is always in an infinitive form and the causee can not intervene bewteen the two verbs but instead appears in oblique function.

- Periphrastic causatives. Where each of the verbs occupy different clauses, with the causative in the main clause and the lexical verb in the subordinate/complement clause. The causee of a periphrastic causative can be marked for its function either in the subordinate clause only (as in Macushi where it carries ergative case in the subordinate clause),

or in the main clause only (as in English where it is the object of the causative verb and therefore marked accusative) or in both clauses (as in Canela-Kraho where it is an object of the causative verb and subject of the lexical verb).

Evidence for the bi-clausal properties of periphrastic constructions were cited by Shibatani (1973) in his attempt to differentiate between the underlying structure of lexical and periphrastic causatives in Korean (but he also mentions that these differences are attested in a number of different languages, including English). Such evidence includes cases where adverbial modification is ambiguous between the two events represented by the two clauses, which is absent in cases of lexical causative constructions in Korean. The examples cited in (Shibatani, 1973, p.287) are translated into their equivalents in ECA below, where the difference still holds:

(7) el-ʔom   ḥallet   el-walad yākul   el-sāʕa   ʕaʃara
    the-mother make.PV.3SGF the-boy eat.IPFV.3SGM the-hour ten
    The mother made the child eat at 10 o'clock

There is ambiguity here as to which event took place at 10 o'clock, either the event of the mother making the child eat (the causing event), or that she told him to eat at any point while the event of him eating took place at 10 o'clock (the caused event). This ambiguity is not present in the use of morphological causative in:

(8) el-ʔom   ʔakkelit   el-walad el-sāʕa   ʕaʃara
    the-mother feed.PV.3SGF the-boy the-hour ten
    The mother fed the child at 10 o'clock

Where here there is only one event that took place, that of feeding the child. This is taken by Shibatani (1973) as evidence for "an embedding analysis of periphrastic causatives, for the scope of adverbial modification can be accounted for more straightforwadly by positing two underlying sentences, one embedded into another" (Shibatani, 1973)[3].This is also taken here as evidence that different types of causatives have different syntactic structures, where the literature section

---

[3]This will be discussed again with regards to the *ḥalla* construction in section 3.6.1 to prove that the structure is indeed bi-clausal.

here is only concerned with analytic or periphrastic constructions.

In terms of arguments of a causative construction, Alsina (1992) considers the causative in some languages to be a "three-place predicate in which the causer (or agent) acts on an individual, the patient, in bringing about an event, of which this individual is itself an argument" (Alsina, 1992, p.521). The patient of the causer might map to different grammatical functions in the embedded predicate according to its meaning, where it can function as the logical subject if it represents the participant controlling the caused event, while it functions as the logical object (patient or theme) if it represents a participant affected by the caused event (Alsina, 1992). This mapping reflects on the question of whether the causee is an argument of the cause predicate, which is investigated regarding _ḥalla_ in section 3.5.2.1. The semantic arguments are then mapped onto their corresponding participants using syntactic functions within the lexical mapping theory, however, argument structure mapping is not within the focus of the current study.

One of the analytic causative constructions that has extensive studies throughout the literature is the French causative predicate 'faire', which is described in Yates (2002, p.393) as: "In simple terms, the causative predicate 'faire' (to do, to make) takes an infinitive verb as complement, and apparently shares its arguments with the embedded predicate's arguments. If the embedded verb is intransitive, its subject will be expressed as the causative's direct object. If the infinitive verb is transitive, its subject will be expressed either as an indirect a-object or as a par-phrase (the typical agent complement phrase in French). These are the possible functions that can be occupied by what has been called the 'causee' role. All other arguments keep their grammatical function but seem to be complements to the causative verb." The predominate analysis in LFG for the French causative is that it is mono-clausal, where the two verbs ('faire' and the embedded verb) behave as one predicate sharing the same arguments (Lovestrand, 2009). Contrary to this view, Yates (2002) argues that this structure is in fact bi-clausal based on facts such as case assignment of the causee and clitic placement on 'faire' and the lexical verb in which the

two verbs behave as belonging to different clauses. This proposal is further investigated in

Lovestrand (2009)'s work in which tests for subjecthood are applied to the causee under the

assumption that bi-clausal sturctures involve control, where the causee functions as the OBJ

of the matrix clause as well as the SUBJ of the complement clause, and therefore proving that

the causee in the French causative construction indeed has some properties of the grammatical

subject provided evidence for the structure being bi-clausal (Lovestrand, 2009). Such tests

for subjecthood of the causee as well as evidence for the bi-clausal structure of the causative

construction will be used in the analysis section with respect to *ḫalla* causative constructions in

ECA. Section 3.4 presents the corpus-based data for causative *ḫalla* and its usage in sentences

of ECA, on which the analysis will be based.

3.4   Data

The following are a list of sentences which were extracted from the ECA corpus that I have

compiled using online texts in ECA from resources such as blogs, magazines and twitter[4]. I

have conducted a search for *ḫalla* through the corpus in all its inflections regarding person,

number, gender and tense and aspectual markings as well as cases where it was combined with

the negation affix. This serves to get as much data as possible for this verb in all its contexts,

so as to have a clear idea about the different structures in which it can be used. The search

results included only the cases where *ḫalla* was followed by a lexical verb, excluding cases where

it functions as the main verb with nominal or adjectival complements.

Following are some examples for the sentences that constitute the set of data for *ḫalla*, starting

by sentences 9 to 13 where the causer is an inanimate entity:

(9) el-eḥterāf   ʔātel l-el-mawheba, bi-yḫallī-k   teġanni
     the-professionalism killer for-the-talent, BI-make.IPFV.3SGM-you sing.IPFV.2SGM
     ʕaʃān   tedfaʕ   el-ʔigār
     because pay.IPFV.2SGM the-rent

---

[4]Section 1.3 includes more details about the corpus and its compilation.

Professionalism kills talent, it makes you sing only to pay the rent

(10) el-mobarrer el-waħīd ʔilli   yẖallī-k   teʃtem   bani-ʔādam
     the-reason   the-only which make.IPFV.3SGM-you curse.IPFV.2SGM son-Adam
     ʔinn-o   ykūn   damm-o   tʔīl
     that-him be.IPFV.3SGM blood-his heavy
     The only reason that makes you curse a person is that he's a bad person

(11) ʔana magnūn l-el-ħad   ʔilli   yẖallī-ni   ʔalbes   w
     I   crazy   to-the-extent which make.IPFV.3SGM-me dress.IPFV.1SG and
     ʔanzel   ʔarūħ   el-ʃoġl maʕ ʔin-ni ʃoġl-i   baẖallas-o
     descend.IPFV.1SG go.IPFV.1SG the-work with that-me work-my finish.BI.IPFV.1SG-it
     men ʕala el-ʔinternet
     from on   the-internet
     I'm crazy to the extent that makes me get dressed and go to work although I finish my

     work on the internet

Note that example 11 has a sequence of three verbs in the complement of *ẖalla*. They take the
form of coordination between the first one *ʔalbes* 'dress' and the following sequence of two verbs
ʔanzel ʔarūħ 'descend go'. This whole structure is often used as a collocation which roughly
means simply 'to go'.

(12) ʔilli   ẖallā-ni   ʔāgi   ʔin-ni ʕreft   ʔinn-ak taʕbān
     which make.PV.3SGM-me come.IPFV.1SG that-I know.PV.1SG that-you tired
     What made me come is that I knew that you're tired

(13) feʕlan   ħa-tkūn   el-ħāga el-waħīda ʔilli   ħa-tẖallī-ni
     Actually FUT-be.IPFV.3SGF the-thing the-only   which FUT-make.IPFV.3SGF-me
     ʔastaħmel   fekret el-dirāsa men ʔawwil   w   gdīd
     tolerate.IPFV.1SG idea   the-study from beginning and new
     Actually this will be the only thing that will make me tolerate the idea of studying all

     over again

Note that in all of the previous examples, the causer is an inanimate entity such as profession-
alism, extent of madness, mokery, a relative clause representing the causing event, as well as an
unspecified entity in 13.

However, this is not always the case, where the causer can of course refer to a human as in the
following examples:

(14) howwa ʔilli ḫallā-ni   ʔatʔakkid   ʔin meʃ kol zobbāt el-ʃorta   weḥʃīn
     he   who make.PV.3SGM-me be.sure.IPFV.1SG that not all officers the-police bad
     He's the one who made me sure that not all policemen are bad

Note also that in all of the previous examples the causee is expressed as a pronoun attached to
causative *ḫalla*, however this is not always the case, as shown in examples 15 to 18 below where
the causee is expressed as a free NP:

(15) taryaʔa w hzār   ḫallo   el-walad ynām   nōm   el-wāḥed ʔilli
     Irony   and mocking make.PV.3PL the-boy sleep.IPFV.3SGM sleepiness the-one that
     meʃ lāʔi   ḥad   yesmaʕ-o
     not find.PTCP.SGM someone listen.IPFV.3SGM-him
     The irony and mocking made the boy sleep as a person who can't find anyone to listen to
     him (feeling lonely and sad)

(16) ʃukran ʔinn-ak   ḫallēt   Sundus torbot   bēn   ʕelaʔet-ha
     thanks that-you make.PV.2SGM Sundus relate.IPFV.3SGF between relationship-her
     b-iyya   w   fekret el-ḥob
     with-me and idea   the-love
     Thanks for making Sundus relate between her relationship with me and the concept of
     love

(17) ʔēh el-sabab   ʔilli   ḫalla   gōz-ek   yebʕed   ʕann-ik
     what the-reason which make.PV.3SGM husband-your get.away.IPFV.3SGM from-you
     What is the reason that made your husband leave you?

(18) keda   enti ḥa-tḫalli   el-rāgel yidrab-ek
     this.way you FUT-make.IPFV.2SGF the-man hit.IPFV.3SGM-you
     This way you'll make the man beat you!

In all of these cases, the NP expressing the causee can be expressed either by an NP or an
attached pronoun with no change in meaning, except for the vagueness of the person to whom
the pronoun refers, as is the case with all pronouns. This is shown by:

(19) keda   enti ḥa-ṯẖalli-h   yidrab-ek
     this-way you FUT-make.IPFV.2SGF-him hit.IPFV.3SGM-you
     This way you'll make him beat you!

ẖalla occured in a number of different temporal and aspectual markings in the previous exam-
ples[5], it can also occur in the imperative form, as in:

(20) ẖalli   sāmi yibʕed   ʕan el-siyāsa
     make.IMP.2SG Sami get.away.IPFV.3SGM from the-politics
     Make Sami stay away from politics

(21) ẖallī-ha   tḥess   ʔinn-ak feʕlan bi-tetʔallem
     make.IMP.2SG-her feel.IPFV.3SGF that-you really BI-in.pain.IPFV.2SGM
     Make her feel that you're in real pain

While the lexical verb occured in the bare imperfective form[6] in all of the sentences, except
in only one of the corpus examples where the lexical verb occured in the perfective form, the
sentence is:

(22) ʔuslūb-u   maʕā-na ẖallā-ni   baʔēt   ʕāwiz
     method-his with-us   make.PV.3SGM-me become.PV.1SG want.PTCP.SGM
     ʔadrab-u
     hit.IPFV.1SGM-him
     His manner of behaviour with us made me want to hit him

Note that the verbal complement of the causative is a phasal verb baʔēt 'become.PV.1SG' fol-
lowed by the modal ʕāwiz 'want.PTCP.SGM', however they are used here merely to denote the
state of the speaker at the time of speech, while the caused event is expressed by the lexical
verb ʔadrab-u 'hit.IPFV.1SGM-him'. Discussion of these verb forms is tackled in more details in
section 3.5.4.

---

[5]Discussed further in section 3.5.4.
[6]The bare imperfective form is considered to be the non-finite verb form in ECA, this is tackled in more
details in section 1.4.

Although I (as a native speaker) think the structure in 23 is un-natural, in fact three sentences in the corpus included the complementiser *ʔin* heading the complement clause of *ḫalla*, the sentence is[7]:

(23) ʔintāg   el-betrōl fi el-ʕirāʔ howwa ʔilli   ḫalla   ʔinn-aha
      production the-petrol in the-Iraq it   which make.PV.3SGM that-it
      tebʔa   qowwa ʔiqtisādeyya mohemma
      become.IPFV.3SGF force economic   important
      The oil production in Iraq is what made it become an important economic force

However, this structure seems to bear the meaning of causing a state rather than acting on a causee. This is shown by example 24 below which was a result of my search on Twitter for this specific structure *ḫalla ʔinn* 'made that':

(24) mīn ḫalla   ʔin el-nās   men beʕīd ʔaħla   qaʕda ʕamma,
      who make.PV.3SGM that the-people from far   more.beautiful rule general,
      bel-ʕaks   fi nās   ʔamma tʔarrab-l-aha   tlaʔī-ha
      with-contrary in people when   be.near.IPFV.2SGM-to-her find.IPFV.2SGM-her
      ʔagmal   w ʔagmal
      more.beautiful and more.beautiful
      Who made it a fact that people are more beautiful from far away a general rule, on the
      contrary, there are people who get more beautiful the more you get near them

This sentence has no causee at all, but it rather represents causation of a state or idea and wondering about who made it into a fact. It is hard here to account for the meaning of the sentence accurately as it does not indicate a specific event, instead it refers to a general concept (that people are more beuatiful when you don't know them that well) which the speaker is questioning. Another sentence is:

(25) ya   gidʕān mark ḫalla   ʔin ma-ħadd-eʃ   yeʔfel   el-fēsbuk
      hey people Mark make.PV.3SGM that NEG-someone-NEG close.IPFV.3SGM the-facebook
      Oh people, Mark made it that no one closes facebook

---

[7]This sentence is a simplified version of the context in which the structure occured, where in the corpus it is a long paragraph with no breaks, therefore the sections preceding the causative structure were simplified without affecting the meaning or syntactic structure of the casuative construction.

The same kind of reference to a generic state is used in example 25, where the speaker is indicating that Mark Zuckerbeg (the founder of facebook) is adding some new features and techniques that are causing people to not close facebook (as it is so tempting..). Therefore the caused event is actually a state in which different people are entered rather than having a specific causee. This seems to be correlated with the use of the complementiser, which will be further tackled in more details in section 3.5.5.

It is also possible to have an adjunct intervening between the causative verb and its complement phrase, as in the corpus examples:

(26) mālik ʔilli ḫallā-ni   bas ʔatkallim   ʕan   el-mawdūʕ badri
     Malik who make.PV.3SGM-me just talk.IPFV.1SG about the-subject early
     Malik made me only talk about it earlier

(27) ḫallā-ni   men ġēr   ma yoʔsod   ʔaħtarem   nafsi
     make.PV.3SGM-me from other that intend.IPFV.3SGM respect.IPFV.1SG myself
     He made me -without intending- have more respect for myself

Note that here the phrase 'without intending' is represented in the translation as an interjection, while it occured in the written corpus without any commas or punctuation marks to show such an interjection. However it is a common structure in ECA and it is commonly used as an interjection.

The following section, 3.5, is concerned with discussion of the data as well as investigating the behaviour of causative _ḫalla_, focusing on aspects such as negation patterns of the verb, its arguments, as well as having adverbs or complementisers in the causative structure.

3.5   Behaviour of the verb

This section includes investigation of the behaviour of the causative construction, in terms of negation, arguments of _ḫalla_, the possiblity of having other constituents inserted between _ḫalla_

and its verbal complement, the verb forms in which both *halla* and the lexical verb can occur, as well as the possiblity of having a complementiser in the causative structure. These criteria are important to understand the relation between *halla* and its verbal complement, in terms of the structure being mono or bi-clausal as well as defining the type of control relation between the two verbs. This will be the basis for reaching the analysis that best decribes the behviour of the causative construction.

Examples used in this section are mostly not taken from the corpus, due to the fact that not all of these structures were attested in the corpus. Therefore the sentences used in this section are either modifications of the corpus sentences, by adding negation or inserting an adverb for example, or sentences that I have created (as a native speaker of ECA) to investigate this structure. The latter sentences will be marked as [Constructed] to show the difference.

3.5.1   Negation

Negation of *halla* negates the causative meaning, as in:

(28) ma-tḫall-īʃ   sāmi yibʕed   ʕan el-siyāsa
      NEG-make.IMP.2SG-NEG Sami get.away.IPFV.3SGM from the-politics
      Don't make Sami stay away from politics = Don't be the reason he gets away from politics

While negation of the lexical verb gives only an interpretation of causation of a negative action "make X not-do", as in:

(29) ḫalli   sāmi ma-yibʕed-ʃ   ʕan el-siyāsa
      make.IMP.2SG Sami NEG-get.away.IPFV.3SGM-NEG from the-politics
      Make Sami not stay away from politics = Keep Sami involved in politics

This difference in meaning between the two sentences is subtle, however the difference between the two cases of negation is actually a difference between the event negated, where in 28 the

negation is for the causative event meaning "don't make X do", while in 29 the caused event is negated meaning "make X not-do". This difference might give evidence for each of the verbs forming a distinct predicate from the other, where negation on one does not scope over the other.

3.5.2   Arguments

This section aims to find out what are the arguments of *ḥalla*, specifically investigating whether the causee functions as its object, and whether the causer is the thematic subject of *ḥalla*. Section 3.5.2.1 investigates the first of these issues.

3.5.2.1   Is the causee an OBJ argument of *ḥalla*?

In all of the previous examples, the causee occurs in a position in which it can be considered object of *ḥalla* as well as subject of the following lexical verb. It might be represented as a separate noun or as a pronoun attached to *ḥalla* in OBJ position, with no difference in meaning[8]. The function of the causee as subject of the lexical verb seems straighforward, while the question remains as to whether it also represents the syntactic OBJ of the causative verb.

Answering this question needs inspection of the syntactic behaviour of objects and the way they are defined with repsect to the verb. This issue was tackled in Joshi (1993)'s thesis in an attempt to find criteria for defining the grammatical object in Marathi, where the grammatical object is defined as the argument having one or more of the proto-patient properties, most of these properties are adopted from the Dowty (1991) model that includes properties such as undergoing change of state/location, being an incremental theme and being affected causally by another participant (Joshi, 1993, p.80).

---

[8]Note that if the causee is realised as a pronoun, it is required to be attached to *ḥalla*, so it can not occur as a free pronoun in OBJ position.

Camilleri et al. (2014) also used some tests to prove the status of the object for some ditransitive verbs, these tests include affixation of the pronominal object to the verb, as well as promotion of the object to subject position in passive constructions. These criteria were applied to the causee as a possible object of causative *ḫalla* represented in the previous examples to show whether it really occupies the grammatical function of OBJ for the causative verb or not. Application of each of these criteria is represented as follows:

- Having proto-patient properties: Since all of the structures in question here involve the causative *ḫalla*, therefore the causee can be said to be affected causally by the event of causation in all of the examples, and therefore exhibits a proto-patient property.

- Incorporation of pronominal objects to the verb: This is clear in cases of negation where causative *ḫalla* is negated using the circumfix *(ma- -ʃ)*, where the suffixal negative marker is attached after the pronominal object, such as:

(30) ma-tḫalli-hū-ʃ   yidrab-ek!
      NEG-make.IMP.2SGF-him-NEG   hit.IPFV.3SGM-you
      Don't make him hit you!

Here the pronoun referring to the causee (him) is attached to the causative verb and precedes the negative suffix, which is the position for the object of *ḫalla*. This might form evidence that this argument is in fact a grammatical object for the causative verb.

- Passivisation: This test is based on the assumption that "a primary object is expected to be able to promote to subject under passivisation" (Camilleri et al., 2014, p.34). Applying this test to the examples illustrated here would not give the expected result of the causee functioning as the object. This is due to the fact that the causative verb *ḫalla* cannot be passivised in the first place, where passives are formed in ECA by adding the prefix *ʔit-* to the verb, and applying this process to the causative would give *ʔit-ḫalla* which does not

mean 'be-made' but rather means 'abandon', which is a different verb in the active voice. Therefore this test is unapplicable to the case in question here.

To sum up, results of the objecthood tests applied so far show that the causee functions as object of the causative verb, where the causee functions both as object of the causative and subject of the following lexical verb.

3.5.2.2   Is the SUBJ of *ḫalla* a thematic argument?

The SUBJ of *ḫalla* can be either a human (denoting volition) such as the case in examples 18 and 20, or an inanimate/abstract entity as in examples 10 and 13. This raises a question as to whether the SUBJ of the causative is a thematic argument in the lack of an animate volitional agent for the verb.

Such a question was raised by Joshi (1993) in defining the grammatical subject in Marathi, and the conclusion was drawn that "if a verb has a volitional argument, that argument is its grammatical subject. If the verb has no volitional argument, then any argument of the verb that possesses a P- Patient property can be its grammatical subject." (Joshi, 1993, p.92). A similar approach to defining the grammatical functions according to their thematic roles is the properties set by Dowty (1991) for defining the proto-agent, such as the volitional involvement in the event, sentience as well as causing a change of state in another participant.

Application of these criteria to the current examples of causative constructions shows that whether or not the agent is represented as an animate volitional causer, it functions as the grammatical subject. This can be shown in examples in which the inanimate causers behave similarly to the animate volitional causers, where it can be substitued by a noun representing the subject, such as the following example which is parallel to example 11:

(31) momken gunūn-i   yḫallī-ni   ʔalbes   w   anzel
      possible madness-my make.IPFV.3SGM-me dress.IPFV.1SG and descend.IPFV.1SG

ʔarūḥ   el-ʃoġl   maʕ ʔin-ni   ʃoġl-i   ba-ḫallas-o   men ʕala
go.IPFV.1SG the-work with that-me work-my finish.BI.IPFV.1SG-it from on
el-ʔinternet
the-internet

My madness can make me get dressed and go to work although I finish my work on the

internet

Therefore this suggests that the causer actually fulfills the position of the grammatical and

thematic subject of the cause predicate, whether it represents an animate volitional causer or

not. This resembles the case in English where inanimate causers can be subjects, unlike in

Marathi where they are expressed as obliques or adverbial adjuncts (Joshi, 1993). The same

criteria can also be applied to the role of the causee as the grammatical subject of the lexical

verb.

3.5.3 Insertions

This section is concerned with the possiblity of having arguments or adjuncts inserted between

ḫalla and its verbal complement, which could provide evidence that each of the verbs constitutes

a separate predicate. This is illustrated by cases in which an adverbial element occurs between

the causative verb and the lexical verb, as in the corpus example 27 repeated below:

(32) ḫallā-ni   men ġēr   ma yoʔsod   ʔaḥtarem   nafsi
     make.PV.3SGM-me from other that intend.IPFV.3SGM respect.IPFV.1SG myself
     He made me -without intending- have more respect for myself

Here the phrase *men ġēr ma yoʔsod* 'without intending' represents a manner adverbial phrase.

Other adverbs can also be added to a causative structure, as in:

(33) ʔilli   ḫallā-ni   ʔimbāreḥ ʔāgi   ʔin-ni ʕreft   ʔinn-ak
     which make.PV.3SGM-me yesterday come.IPFV.1SG that-I know.PV.1SG that-you
     taʕbān
     tired.PTCP.SGM
     What made me yesterday come is that I knew that you're tired

Where the adverb *ʔimbāreħ* 'yesterday' occurs between the causative verb and the lexical verb. It is also possible to substitute the pronoun for causee with an NP, as in:

(34) ʔilli ḫallā   Sawsan ʔimbāreħ tīgi   ʔin-aha ʕerfet
     which make.PV.3SGM Sawsan yesterday come.IPFV.3SGF that-she know.PV.3SGF
     ʔinn-ak taʕbān
     that-you tired.PTCP.SGM
     What made Sawsan yesterday come is that she knew that you're tired

Adverbs can also occur in different positions, whether before *ḫalla* or after the lexical verb, where in each case it modifies a different event:

(35) ʔilli   ʔimbāreħ ḫallā-ni   ʔāgi   ʔin-ni ʕreft   ʔinn-ak
     which yesterday make.PV.3SGM-me come.IPFV.1SG that-I know.PV.1SG that-you
     taʕbān
     tired.PTCP.SGM
     Yesterday, what made me come is that I knew that you're tired

Here the adverb refers to the time of the causation event, which happened yesterday, while there is no indication of the time of the caused event, where the speaker might have actually come today. While the case is different with example 36 below where the adverb specifies the time of the caused event, meaning that the speaker's coming event occured yesterday:

(36) ʔilli   ḫallā-ni   ʔāgi   ʔimbāreħ ʔin-ni ʕreft   ʔinn-ak
     which make.PV.3SGM-me come.IPFV.1SG yesterday that-I know.PV.1SG that-you
     taʕbān
     tired.PTCP.SGM
     What made me come yesterday is that I knew that you're tired

3.5.4 Verb forms

Causative *ḫalla* can occur in all tenses, where it occured as *bi*-prefixed imperfective (denoting present habitual/progressive) in example 9, as bare imperfective (representing the infinite verb form) in 10 and 11, as *ħa*-prefixed (denoting future tense) in 13 and 18, and as perfective (denoting past tense) in all of the other examples. It also occured in the imperative verb form

in examples 20 and 21. A note to make here is that these forms are interchangeable, where
the choice of the temporal marking depends mainly on the meaning intended while there is no
restriction on the verb form of *ḫalla* in terms of the expression of tense and aspect. This applies
to all of the forms of *ḫalla* except the bare imperfective, which occurs in specific clause types as
it represents a non-fnite verb form[9].

On the other hand, the lexical verb occured in the bare imperfective form in all of the examples,
except for example 27 repeated here, where it occured in the perfective verb form.

(37) ʔuslūb-u   maʕ-āna ḫallā-ni   baʔēt   ʕāwiz
      method-his with-us   make.PV.3SGM-me become.PV.1SG want.PTCP.SGM
      ʔadrab-u
      hit.IPFV.1SGM-him
      His manner of behaviour with us made me want to hit him

In this example the temporal interpretation of the whole structure is past perfective, where
the causative verb marks past tense while the following verb denotes perfective aspect[10]. The
lexical verb can also occur in the *bi*-prefixed form, such as example 15 repreated below with a
different temporal marking on the lexical verb:

(38) taryaʔa w hzār   ḫallo   el-walad bi-ynām   nōm   el-wāħed
      irony   and mocking make.PV.3PL the-boy   BI-sleep.IPFV.3SGM sleepiness the-one
      ʔilli meʃ lāʔi   ħad   yesmaʕ-o
      that NEG find.PTCP.SGM someone listen.IPFV.3SGM-him
      The irony and mocking made the boy sleep as a person who can't find anyone to listen to
      him (feeling lonely and sad)

Despite the fact that the free translation here is the same as that of 15, the glossing is different
as the lexical verb is bi-prefixed, this is reflected in a slight difference in meaning that I could
not express accurately in the free translation. The *bi*-prefixed verb adds a habitual aspect to

---

[9]The bare form and its function as a non-finite verb in ECA is tackled in more details in section 1.4.
[10]This is based on the analysis of tense and aspect in compound tense forms presented in chapter 2.

the structure, where the whole structure is in the past habitual denoting that the past causative event made the child get into the habit of sleeping in this manner, while the structure in 15 lacks this habitual aspect and merely denotes causing the state.

This difference is clear in example 39 below where the adverb *lḥad delwaʔti* 'uptil now' which is used to refer to a state that extends till the present is added to the *bi*-prefixed lexical verb structure, and the habitual meaning still holds, where the sentence refers to a habit of sleeping that is continuing till now:

(39) taryaʔa w   hzār   ḫallo   el-walad bi-ynām   lḥad delwaʔti
      irony   and mocking make.PV.3PL the-boy BI-sleep.IPFV.3SGM uptil now
      nōm   el-wāḥed ʔilli meʃ lāʔi   ḥad   yesmaʕ-o
      sleepiness the-one   that NEG find.PTCP.SGM someone listen.IPFV.3SGM-him
      The irony and mocking made the boy sleep uptil now as a person who can't find anyone
      to listen to him (feeling lonely and sad)

These examples show a significant difference in the expression of tense/aspect between *ḫalla* and the lexical verb, and this might form evidence for the structure being bi-clausal, where each verb has its own temporal marking, with an indication that *ḫalla* marks tense while the following lexical verb marks aspect.

The only cases in which both *ḫalla* and the lexical verb occured in the bare imperfective form were in examples 10 and 11 where both are in embedded clauses following *ʔilli* 'which/who'. This does not give the sentence the present tense marking, but rather adds a modal interpretation to the whole structure, which is clear in the possiblity of adding the modals *lāzim* 'must' or *momken* 'possible' preceding the causative construction in the same sentence, as in:

(40) ʔel-mobarrer el-waḥīd ʔilli   momken yḫallī-k   teʃtem   bani
      the-reason   the-only which possible make.IPFV.3SGM-you curse.IPFV.2SGM son
      ʔādam ʔinn-o   ykūn   damm-o tʔīl
      Adam that-him be.IPFV.3SGM blood-his heavy
      The only reason that might make you curse a person is that he is a bad person

While this modal interpretation is not possible if _halla_ is in a tensed form, such as the _bi_-prefixed in:

(41) *ʔel-mobarrer el-waħīd ʔilli   momken bi-yhallī-k   teʃtem   bani
      the-reason   the-only which possible BI-make.IPFV.3SGM-you curse.IPFV.2SGM son
      ʔādam ʔinn-o   ykūn   damm-o tʔīl
      Adam that-him be.IPFV.3SGM blood-his heavy

However, this is a result of the fact that verbal complements of the modal _momken_ 'possible' can only be in the bare imperfective form[11].

Another point to note here is that when the auxilliary _kān_ precedes _halla_, the only form that seems possible for the lexical verb is the bare imperfective, while _halla_ can occur in different tenses. The resultant temporal interpretation is the result of the combination between the form of _kān_ and _halla_, as the lexical verb is in the bare imperfective form anyway, so it does not give any specific temporal indication. This is shown by the following example:

(42) feʕlan   di el-ħāga el-waħīda ʔilli   kān-et   ħa-thallī-ni
      actually this the-thing the-only   which be.PV.3SGF FUT-make.IPFV.3SGF-me
      ʔastaħmel   fekret el-dirāsa men ʔawwil   w   gdid
      tolerate.IPFV.1SG idea   the-study from beginning and new
      Actually this is the only thing that would've made me tolerate the idea of studying all
      over again

This structure refers to a past prospective event, where something was going to happen but it did not. This is the result of the combination between the perfective form of the auxilliary denoting past tense and the _ħa_-prefixed form of _halla_ denoting prospective aspect, which gives the past prospective compound tense interpretation.

3.5.5   Complementiser _ʔin_

This section is concerned with cases where the complement clause of the causative is headed by the complementiser _ʔin_ 'that', these are the corpus examples 23, 24 and 25 repeated below:

---

[11]This is shown in more details in the investigation of the modal _momken_ in section 5.3.3.3.

(43) ʔintāg   el-betrōl   fi el-ʔirāʔ howwa ʔilli   ẖalla   ʔinn-aha
     production the-petrol in the-Iraq it   which make.PV.3SGM that-her
     tebʔa   qowwa ʔiqtisādeyya mohemma
     become.IPFV.3SGF force   economic   important
     The oil production in Iraq is what made it become an important economic force

(44) mīn ẖalla   ʔin el-nās   men beʕīd ʔaħla   qaʕda ʕamma,
     who make.PV.3SGM that the-people from far   more.beautiful rule   general,
     bel-ʕaks   fi nās   ʔamma tʔarrab-l-aha   tlaʔī-ha
     with-contrary in people when   be.near.IPFV.2SGM-to-her find.IPFV.2SGM-her
     ʔagmal   w   ʔagmal
     more.beautiful and more.beautiful
     Who made it a fact that people are more beautiful from far away a general rule, on the

     contrary, there're people who get more beautiful the more you get near them

(45) ya   gidʕān mark ẖalla   ʔin ma-ħadd-eʃ   yeʔfel   el-fēsbuk
     hey people Mark make.PV.3SGM that NEG-someone-NEG close.IPFV.3SGM the-facebook
     Oh people, Mark made it that no one closes facebook

The use of the complementiser in the complement clause of ẖalla is percieved as un-natural or
weird by the native speaker, where the normal structure would drop the complementiser while
preserving the same meaning, as in:

(46) ʔintāg   el-betrōl   fi el-ʔirāʔ howwa ʔilli   ẖallā-ha   tebʔa
     production the-petrol in the-Iraq it   which make.PV.3SGM-her become.IPFV.3SGF
     qowwa ʔiqtisādeyya mohemma
     force   economic   important
     The oil production in Iraq is what made it become an important economic force

The meaning in these examples indicates causation of a state, where it gives the meaning of caus-
ing a proposition to be true rather than affecting the causee to perform an action, which is the
interpretation of the sentence without the complementiser. This difference somewhat resembles
the distinction between direct and indirect causation, where direct causation is one in which "a
causer directly causes a caused event. Direct causation is contrasted with indirect causation, in
which a causer brings about an event through some mediating factors." (Matsumoto, 1992, 194).

This is further shown by the fact that adding the complementiser to any of the other corpus examples results in the sentence being unacceptable. Note also that in cases where the complementiser *ʔin* is used, the incorporation of the OBJ within the causative verb does not seem to be possible anymore. This is clear in the ungrammaticality of the example:

(47) *ʔintāg   el-betrōl   fi el-ʔirāʔ howwa ʔilli   ma-ẖalla-hā-ʃ   ʔinn-aha
     production the-petrol in the-Iraq it   which NEG-make.PV.3SGM-her-NEG that-it
     tebʔa   qowwa ʔiqtisādeyya mohemma
     become.IPFV.3SGF force   economic   important

While the counterpart where the complementiser is absent yields a grammatical sentence.

(48) ʔintāg   el-betrōl   fi el-ʔirāʔ howwa ʔilli   ma-ẖallā-hā-ʃ
     production the-petrol in the-Iraq it   which NEG-make.PV.3SGM-her-NEG
     tebʔa   qowwa ʔiqtisādeyya mohemma
     become.IPFV.3SGF force   economic   important
     The oil production in Iraq is what didn't make it become an important economic force

Therefore the causee in sentences taking a complement phrase headed by *ʔin* is not an argument of the causative, but rather functions solely as the subject of the lexical verb. This eliminates the control relation between the object of the causative and subject of the lexical verb which is attested in the corpus examples lacking the complementiser, and therefore sentences having *ʔin* shall have a different analysis than the rest of the corpus examples. These issues will be tackled in more details in section 3.6 where the analysis will be provided, as well as 3.7 where conclusions will be drawn.

The following section, 3.6, aims at using the results of the investigation of the data sentences in order to reach the analysis that can best account for the analytic causative structure in ECA.

3.6   Analysis

This section is devoted to the syntactic analysis of the causative structure represented in the previous examples. The analysis is based on the behaviour patterns investigated in the data,

which helps reach the correct account of the syntactic structure of the causative construction within the principles of the LFG theory. Each of the following sections is devoted to answering a question about the strucuture, where 3.6.1 investigates whether the structure is mono or bi-clausal, while section 3.6.2 describes the kind of control relation between *ḫalla* and the lexical verb clauses. This analysis is eventually represented in terms of the verbs' lexical entry in section 3.6.3, f-structure in section 3.6.4 and c-structure in section 3.6.5.

3.6.1 Is the structure mono or bi-clausal?

The current section investigates whether each of the causative *ḫalla* and the following lexical verb shall be treated as separate predicates with separate subcategorisation frames and forming different clauses, or as a case of two verbs forming one predicate. Answering this question is based on the investigation of a number of aspects of behaviour of the causative construction with regards to similar tests found in the literature, such as Falk (2008)'s analysis of English auxiliaries either as 'aux-predicate' or 'aux-feature'.

The syntactic behaviour of the causative and the lexical verb suggests that each of these verbs forms a separate clause, criteria on which this assumption is based include:

- Distinct modifiers: Where the assumption is that if each of the verbs could have distinct modifiers, then this shows that each has a separate clause. This is shown in section 3.5.3 where different adverbs are inserted between *ḫalla* and the lexical verb, and showing different interpretations according to the clause modified. This is exemplified by example 49 below where the causing event is modified by the time adverbial 'yesterday' while the caused event is modified by 'quickly'.

(49) Ɂilli   ḫallā-ni   Ɂimbāreḥ Ɂāgi   bsorʕa Ɂin-ni ʕreft
     which make.PV.3SGM-me yesterday come.IPFV.1SG quickly that-I know.PV.1SG
     Ɂinn-ak   taʕbān
     that-you tired.PTCP.SGM

What made me yesterday come quickly is that I knew that you're tired

- Different interpretations: Where each of the verbs clearly has a distinct interpretation, with _ḫalla_ denoting the causation while the following lexical verb denotes the caused event. Another feature in which they differ is in the marking of morphosyntactic tense/ grammatical aspect, where the causative verb marking tense while the lexical verb marking grammatical aspect[12].

- Negation: Where the assumption is that if each of the verbs could be negated separately giving different interpretations, then each forms a separate clause. This is clear in section 3.5.1 which shows that negation of _ḫalla_ negates the causative meaning, while negation of the lexical verb gives the meaning of causation of a negative action. It is also possible to have negation on both _ḫalla_ and the following lexical verb, as in:

(50) ma-tẖallī-ʃ   el-walad ma-yetkallem-ʃ   tāni ʕaʃān
      NEG-make.IMP.2SG-NEG the-boy NEG-talk.IPFV.3SGM-NEG again because
      bi-yḫāf   menn-ak
      BI-afraid.IPFV.3SGM from-you
      Don't make the boy stop talking because he's afraid of you   [Constructed]

Therefore, the criteria of having a clause boundary between the causative and the lexical verb is met, which makes the construction a periphrastic construction based on the distinction criteria posed by Dixon (2000) as mentioned in more details in section 3.3 earlier. Section 3.6.2 below aims to answer the second question about the analysis of the causative construction, which is about the kind of control relation between the causative verb and the lexical verb.

3.6.2   Kind of control relation

In LFG terms, "Control refers to a relation of referential dependence between an un-expressed subject (the controlled element) and an expressed or unexpressed constituent (the controller); the referential properties of the controlled element, including possibly the property of having no

---

[12]Based on the analysis of compound tense forms discussed in more details in section 2.7.

reference at all, are determined by those of the controller" (Bresnan, 1982, p.317). Control can be divided into two kinds according to the relation between the controller and the controllee, where complete identity between f-structures of the controller and the controllee represents a case of functional control, while in cases of anaphoric control, the controller and the controlled element merely share the same referential properties (Bresnan, 1982).

Observations about the behaviour of the causative contruction suggest that there is indeed a control relation attested between the OBJ of *ḫalla* which represents the controller, and the SUBJ of the following lexical verb (representing the controlled SUBJ). The question remains as to the nature of this control relation, whether anaphoric or functional control, which will be the focus of the current section. Based on the criteria used to differentiate between anaphoric and functional control in the LFG literature such as works by (Kroeger, 2004; Falk, 2001; Dalrymple, 2001), the causative structure represents functional control as it satisfies the following:

- The controllee can not be overt, where it is ungrammatical to have an independent SUBJ pronoun in the subordinate clause. This is shown by the ungrammaticality of:

  (51) *ʃukran ʔinn-ak ḫallet-ha   **Sundus** torbot   bēn
       thanks that-you make.PV.2SGM-her Sundus relate.IPFV.3SGF between
       ʕelaʔet-ha   b-iyya   w fekret el-ḥob
       relationship-her with-me and idea   the-love

  Where the sentence is ungrammatical whether the pronoun refers to Sondos or not, which rules out anaphoric control.

- The controllee must be SUBJ of the lexical verb, as is shown in the ungrammaticality of:

  (52) *keda   enti ḥa-tḫalli-h   ʔadrab-u
       this.way you FUT-make.IPFV.2SGF-him hit.IPFV.1SG-him

  Where here the pronoun which co-refers to the causee (the man) functions also as the object of the lexical verb.

- The controller must be argument of the matrix clause. This is shown by the function of the causee as object of the causative verb as illustrated in section 3.5.2.1.

- The ability to have an expletive subject for the lexical verb, as in examples 53 and 54 below which were a result of a specific search for the structure *halla fīh* on twitter:

  (53) Ɂana ħāsis   Ɂin rabbena ḫalla   fī-h sēf   w ʃeta ʕaʃān
       I   feel.PTCP.SGM that God   make.PV.3SGM in-it summer and winter because
       yisbit   l-el-nās   Ɂin ma-fī-ʃ   ħāga bteʕgeb-hom
       prove.PV.3SGM to-the-people that NEG-in-NEG thing admire.BI.IPFV.3SGF-them
       I feel that God made (there is) summer and winter to prove to people that they could

       never be satisfied

  (54) ma-tʃatt-eʃ   nafsak b-el-tafkīr   fi Ɂay marħala ħēr   Ɂilli
       NEG-distract.IPFV.2SGM-NEG yourself by-the-thinking in any stage   other which
       Ɂenta fī-ha, bas ḫalli   fī-h ḫutūt ʕamma ma-tensa-hā-ʃ
       you in-it, but make.IMP.2SGM in-it lines general NEG-forget.IPFV.2SGM-it-NEG
       Don't distract yourself by thinking about any stage other than the one you're in, but

       keep some broad lines that you shall not forget

- The idiomatic meaning can be preserved in the controlled clause, as in:

  (55) babā-ha howwa illi   ḫalla   el-Ɂatr yfot-ha
       dad-her he   who make.PV.3SGM the-train miss.IPFV.3SGM-her
       Her father is the one who made her a spinster (Lit: He made the train miss her)

       [constructed]

Therefore, application of these tests prove that the structures where *halla* occurs involve functional control of the type raising, where the causee is both subject of the lexical verb and non-thematic object of the causative verb, therefore representing a case of subject to object raising.

However, in cases where the complement clause is headed by the complementiser *Ɂin* as shown in section 3.5.5, there is no such control relation at all, as the causee does not fill any grammatical

function to the causative verb, and therefore represents a simple case of causative verb having a complement clause (COMP). Hence this case would have a separate lexical entry for the causative verb, as shown in section 3.6.3 below were lexical entries for *ḥalla* are presented.

3.6.3   Lexical Entries

This section represents the lexical entry of the causative contructions based on results of the analysis. An example for the lexical entry for causative *ḥalla* is presented as follows:

ḥalla: I

(↑PRED)= 'make <(↑ SUBJ), (↑ XCOMP)> (↑ OBJ)

(↑ OBJ) = (↑ XCOMP SUBJ)

(↑ TENSE)= PAST

This represents the causative verb in cases where no complementiser is used. The OBJ of the causative is the same as the SUBJ of the lexical verb. The lexical verb functions as the main predicate of the XCOMP and it adds the gramamtical aspect. For most of the corpus examples (all except for example 27), the embedded lexical verb is in the bare imperfective verb form, however, it is possible to have other verb forms following the causative verb, such as the case in 27 as well as the bi-prefixed verb in example 38. Therefore there is no need to add any constraints on the verb form of the lexical verb, unlike the case with other phasal verbs and modals investigated in subsequent chapters. The lexical entry for the lexical verb would be:

ʔatkallem: V

(↑PRED)= 'talk <(↑ SUBJ)>'

(↑VFORM) = BARE

baʔēt: V

(↑PRED)= 'become <(↑ SUBJ)>'

(↑ASPECT) = PERF

(↑HAB) = +

Note that the verb *baʔēt* is the same as the phasal verb mentioned in section 4.3.1 where it is the main predicate in the sentence and occupies I in the c-structure, while here it occured in example 27 as a lexical verb and therefore it has a separate lexical entry where it occupies V and carries the grammatical aspect 'perfect'.

A different lexical entry is needed for the causative verb *ḫalla* in sentences where it can be followed by the complementiser *ʔin*, represented as follows:

ḫalla: I

(↑PRED)= 'make <(↑ SUBJ)(↑ COMP)>'

(↑ TENSE)= PAST

Here the causee is not a grammatical object of the causative verb, forming a two-place predicate, where the causer initiates a state represented by the sentence introduced by the complementiser *ʔin*.

These lexical entries are implemented in the XLE grammar fragment and presented in more details in chapter 6 as well as the full list of lexical entries for the grammar fragment in appendix A. While section 3.6.4 below presents the f-structures of the causative construction.

3.6.4   F-structures

This section presents the f-structure for some of the examples for the causative structure, while the f-structures of the full list of sentences in the data are implemented in the XLE grammar

fragment and presented in chapter 6. For each of the sample sentences below, the example is glossed followed by the f-structure.

(56) keda   enti ħa-tḫalli   el-rāgel   yidrab-ek
     this-way you FUT-make.IPFV.2SGF the-man hit.IPFV.3SGM-you

     This way you'll make the man beat you!

$$
\begin{bmatrix}
\text{PRED} & \text{'make} <\textit{SUBJ, XCOMP}>\text{' OBJ} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED 'you'} \end{bmatrix} \\
\text{OBJ}[1] & \begin{bmatrix} \text{PRED 'man'} \end{bmatrix} \\
\text{TENSE} & \textit{FUTURE} \\
\text{ADJ} & \begin{bmatrix} \text{PRED 'this.way'} \end{bmatrix} \\
\text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{'hit} <\textit{SUBJ, OBJ}>\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED } [1] \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED } PRO \end{bmatrix} \\ \text{VFORM} & \textit{BARE} \end{bmatrix}
\end{bmatrix}
$$

(57) ḫalla   ʔinn-aha tebʔa   qowwa
     make.PV.3SGM that-it   become.IPFV.3SGF force

     Made it become a force

$$
\begin{bmatrix}
\text{PRED} & \text{'make} <\textit{SUBJ, COMP}>\text{'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED 'PRO'} \end{bmatrix} \\
\text{TENSE} & \textit{PAST} \\
\text{COMP} & \begin{bmatrix} \text{PRED} & \text{'become} <\textit{SUBJ, OBJ}>\text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED 'PRO'} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED } FORCE \end{bmatrix} \\ \text{VFORM} & \textit{BARE} \end{bmatrix}
\end{bmatrix}
$$

3.6.5 C-structures

This section presents the c-structure for a sample sentence from the data, which is implemented into the XLE grammar fragment in chapter 6. Each example is glossed followed by the c-structure.

(58) ḥalla-ni   ʔaħterem   nafs-i
      make.PV.3SGM-me respect.IPFV.1SG self-my

He made me respect myself

```
              IP
               |
              I'
            ↑ = ↓
          ╱        ╲
        I             VP
      ↑ = ↓      (↑ XCOMP) = ↓
        |            ╱      ╲
     ḥallāni       V          NP
              ↑ = ↓      (↑ OBJ) = ↓
                |             |
             2a7terem       nafsi
```
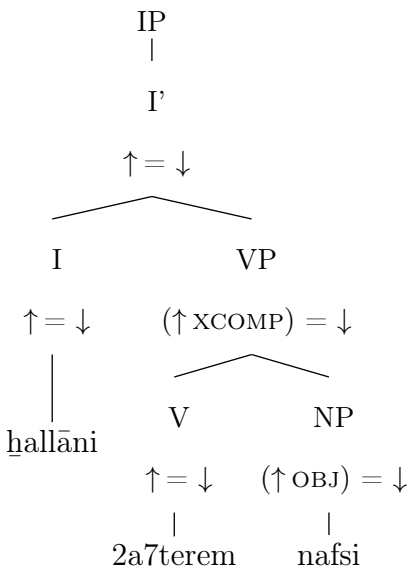
(59) mark ḥalla   ʔin   ma-7add-eʃ   yeʔfel   el-fēsbuk
      Mark make.PV.3SGM that NEG-someone-NEG close.IPFV.3SGM the-facebook

Mark made it that no one closes facebook

```
                        IP
              ┌─────────┴─────────┐
             NP                    I'
         (↑SUBJ) = ↓            ↑ = ↓
             │              ┌─────┴─────┐
            Mark            I            CP
                          ↑ = ↓     (↑COMP) = ↓
                            │             │
                          halla          C'
                                        ↑ = ↓
                                   ┌──────┴──────┐
                                   C              IP
                                 ↑ = ↓          ↑ = ↓
                                   │        ┌─────┴─────┐
                                  ʔin      NP            VP
                                      (↑SUBJ) = ↓      ↑ = ↓
                                           │       ┌─────┴─────┐
                                       maħaddesh   V            NP
                                                 ↑ = ↓      (↑OBJ) = ↓
                                                   │             │
                                                 yeʔfel      elfacebook
```
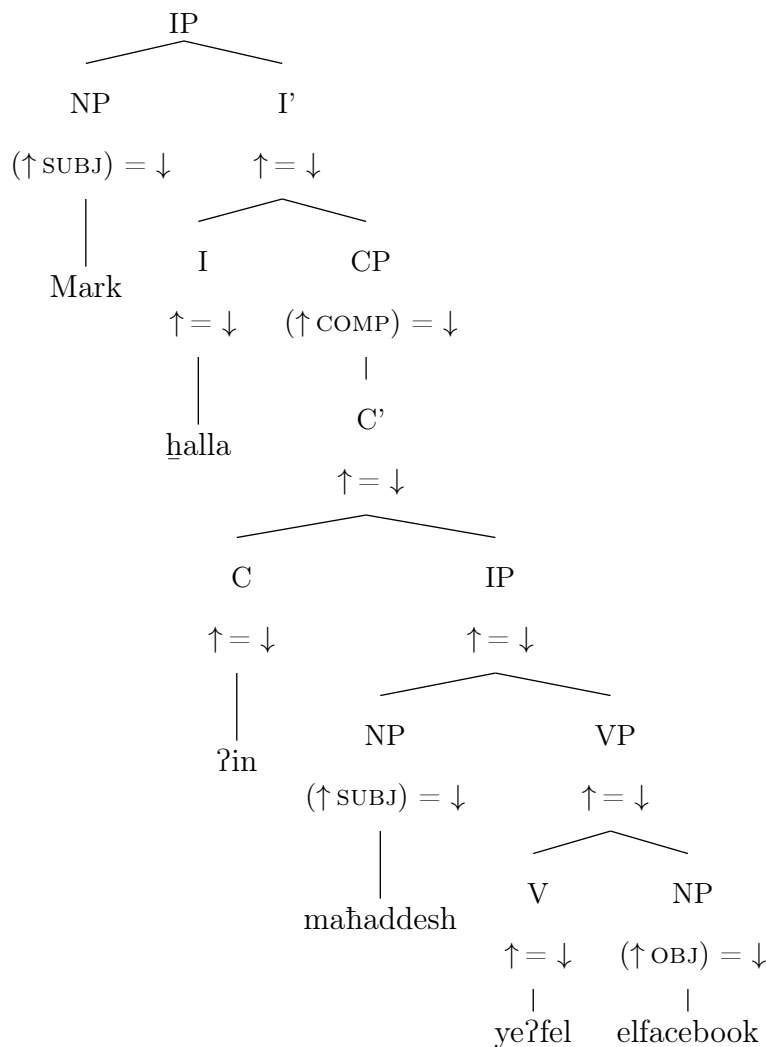
3.7   Conclusions

The data analysed in the current study for structures headed by causative *ḥalla* show that the structure is indeed periphrastic, which is a bi-clausal structure where the causative verb heads the matrix clause and the lexical verb occupies its complement clause (Dixon, 2000). This structure is common among causatives in different languages, such as the French 'faire' (Yates, 2002), korean periphrastic causatives (Shibatani, 1973) as well as Japanese morphological causatives which are bi-clausal structures where the complement of the base verb is an XCOMP (Matsumoto, 1998). In LFG terms, the relation between the two clauses represents functional control of the type raising.

A further distinction is made between two types of bi-clausal causatives, such as the case in Japanese where explicit permissive causatives are equi predicates that subcategorize for a causer, a causee and an event argument. Implicit causatives, on the other hand, are raising predicates that subcategorize for a cause and event arguments (Matsumoto, 1992). This distinction is somewhat reflected in ECA through the difference between the few cases where *halla* takes a complement clause that is headed by the complemetiser *ʔin*, and other sentences in which *halla* acts on the causee and where the presence of *ʔin* leads to unacceptability of the strucuture. However, the analysis in ECA shows that sentences having the complementiser *ʔin* show no control relation between the matrix and subordinate clauses, while sentences lacking *ʔin* are cases of functional control of the type raising. This is due to properties such as its ability to have an expletive subject.

The distinction between the two causative structures might also correspond to the distinction between direct and indirect causation, which is characterised by the presence of an explicit or implicit intermediate agent, and the difference in time interval between the causing event and the caused event, as well as associating a patientive causee with direct causation and an agentive causee with indirect causation (Shibatani, 2002). Accordingly, causative contructions having the complementiser *ʔin* represent indirect causation where the causee functions only as an agent in the subordinate clause, while other constructions lacking *ʔin* represent direct causation, where the causee functions as an object of the causative verb, as well as subject of the lexical verb.

Chapter 4

Phasal verbs

This chapter is concerned with the description and analysis of a group of verbs that are refered to in the literature using the terms "aspectualizers", "aspectual verbs" , "phasal verbs" or "begin class verbs". For the purposes of this study we can take these terms to be interchangeable, however closer inspection of defintions of these terms shows that each focuses on a specific aspect in defining this class of verbs. In using 'aspectual verbs' and 'begin-class verbs' the focus is on the semantic properties of these verbs and their aspectual denotation. While the terms 'aspectualizers' and 'phasal verbs' are often used to refer to this class of verbs in a broader sense, listing them and showing properties of their complements, and therefore both are often used interchangeably. However, the term 'phasal verbs' is adopted in the current study which is mainly concerned with listing some verbs of this class in ECA and describing their syntactic properties.

The chapter starts by section 4.1 which provides an overview on the literature of phasal verbs describing their main properties and how they are treated in terms of the syntactic analysis. Section 4.2 presents the corpus-based data for phasal verbs and describing the criteria on which the verbs are chosen as well as the list of ECA phasal verbs included in the study. Section 4.3 presents the sentences divided into three sets according to the type of complements of the

pahsal verbs, and showing the differents patterns of behaviour of the verb in terms of negation, arguments as well as the forms each verb can take. This leads to the syntactic analysis presented in section 4.4 where the control relation between the phasal verb and the lexical verb will be defined, as well as providing the lexical entries, f-structures and c-structures of the sentences. Section 4.5 includes the conclusions.

4.1 Literature review

This section presents the literature review about phasal verbs, starting by 4.1.1 where definitions of phasal verbs in a number of studies is presented. Followed by section 4.1.2 which accounts for different properties of phasal verbs as mentioned in the literature. And finally section 4.1.3 mentions different accounts for complements of phasal verbs. These sections are important as they form the base for the description of phasal verbs in the current study, which investigates aspects such as the kinds of complements phasal verbs can have and the syntactic account for phasal verbs as control verbs. Comparison between the findings of the current study and the literature on phasal verbs is presented in section 4.5.

4.1.1 Definitions

Noonan (2007) defines phasal verbs as follows: "Phasal predicates refer to the phase of an act or state: its inception, continuation, or termination, and are represented in English by forms such as *begin, start, continue, keep on, finish, stop*, and *cease*. In this category, we should also place *repeat* and *resume*, predicates with an iterative sense "(Noonan, 2007, p.139). This account was mainly concerned with the complementation patterns of different classes of verbs, phasal verbs being one of them, where members of this class can have different complement types even if they share the same meaning, and where tense and aspect contrasts of complements of phasal predicates can lead to contrasts in meaning (Noonan, 2007).

The term 'aspectualizers' is used to refer to "verbs that refer either to the beginning (e.g., begin, start), continuation (e.g., continue, keep) or final part of a situation (e.g., end, finish)" (Nagy, 2009, p.229). While Newmeyer (1975) referred to the same class of verbs using the terms 'aspectual verbs' and 'begin-class verbs' and described the aspectual properties of their complements. Aspectual verbs are defined as "lexical items whose semantic role is to function as one-place predicates of arguments which contain entire propositions"(Newmeyer, 1975, p.8). According to this definition, these verbs must have a proposition as their only argument, and the event described in the complement is these verbs only predicate and logical subject, therefore a verb such as 'happen' is considered aspectual in English (Newmeyer, 1975).

Begin-class verbs, on the other hand, are defined as "verbs of initiation, duration, and cessation. Semantically, they assert occurrence or non-occurrence of their associated propositions with respect to one or more points in time. For example, *begin, start* and *commence* assert occurrence after an implied non-occurrence. *Stop, finish* and *end* assert non-occurrence after a presupposed occurrence. *Keep* and *continue* assert occurrence after the temporal reference point, where occurrence before that point is normally presupposed. *Resume* asserts occurrence after presupposed non-occurrence, which was itself preceded by a presupposed occurrence" (Newmeyer, 1975, p.25).

As these definitions suggest, there are various ways to account for this class of verbs according to the viewpoint of the definition, but the verbs included in this class are largely the same, refered to in this study as 'phasal verbs'. Section 4.1.2 below is concerned with specifying some of the properties of phasal verbs as mentioned in the literature.

4.1.2   Properties of phasal verbs

This section is concerned with the different ways in which phasal verbs are treated in the literature, where they can be accounted for as auxiliaries, complex predicates or raising verbs.

4.1.2.1   Phasal verbs as auxiliaries

In his account of auxiliary-hood in Egyptian Colloquial Arabic (ECA), Eisele (1992) treats phasal verbs as members of the lexical category of auxiliaries. This is based on the notion that they form compound-verb phrases where they represent the highest verb which carries the deictic time reference, and in that they resemble *kān*, which is an auxiliary in ECA. This is specifically true of the phasal verb *yibʔa* 'to become', for which Eisele (1992) shows examples of embedding a perfective verb, a bare imperfective, active participle and a predicate nominal, this shows that it can indicate a variety of temporal contrasts through embedding most types of predicates (Eisele, 1992). Therefore, based on Eisele (1992)'s account, phasal verbs represent the head of a complex verb construction in which they carry the decitic time reference. The category of auxiliary-hood is also described as a fuzzy category with a continuum of features, where some members are prototypical auxiliaries, such as *kān*, while other members are more peripheral, such as modals and phasal verbs (Eisele, 1992).

The verbs investigated in Eisele (1992)'s study as phasal verbs (refered to as aspectualizers) were: *fidil* 'remain' , *ʔaʕad* 'continue', *baʔa* 'become', *ʕād* 'go back'[1], *ʔidir* 'can', *ʔarrab* 'be nearly', *ʕirif* 'know how', *liħiʔ* 'catch', *ʔibtada* 'begin'[2], *ħāwil* 'try', *battal* 'stop'. The verb *baʔa* 'become' is also included but it is considered by Eisele (1992) to be a real auxiliary, just as *kān* 'be'.

However, some of these verbs were not used in the current study, as they do not really represent phasal verbs either in their meanings or in syntactic behaviour, such as *ħāwil* 'try' and *ʕirif*

---

[1]This verb is not used in ECA as a phasal verb as far as I can judge, it is used in MSA. Therefore it was not included in the current study.

[2]This phasal verb is investigated in the current study as *badaʔ*, where both are different templates for the same verb which means 'begin'.

'know'.

There are a number of features that Eisele (1992) uses to describe the behaviour of phasal verbs as belonging to the category of auxiliaries, these are listed below in order of importance for distinguishing verbs as belonging to the lexical class of auxiliary verbs:

- They do not subcategorize for a complementiser, and this is listed as the most important feature of an auxiliary verb. However, *badaʔ* 'begin', *ħāwil* 'try' and *battal* 'stop' are listed as having this feature optional, where they might have a complementiser following them (Eisele, 1992).

- Subject coreferentiality between the phasal verb and the following lexical verb is obligatory. This feature is shared among all of the phasal verbs, and it is mentioned as the second most important feature of being an auxiliary according to Eisele (1992).

- The lexical verb following the auxiliary may not have an independent temporal reference (deictic time reference in Eisele (1992)'s terms). All of the phasal verbs mentioned in his study share this property of lacking a discrete deictic time reference for its complements.

This description shows that although phasal verbs share some behaviours with the auxiliary *kān*, they do not form a coherent set as they show different behaviours as is shown above. These behaviours will be further investigated in the data section of this chapter, to show if they still hold for the data set used in the current study, and this will then be used to investigate how these verbs should be analyzed syntactically.

This is close to the description of these verbs in Michaelis (1998) as 'semi-auxiliaries', where it is stated that: "In English, the inceptive aspect is expressed by means of the 'semi-auxiliaries' *begin* and *start*. These verbs qualify as semi-auxiliaries (SAs) because they partake of semantic properties of true auxiliaries while also exhibiting certain behavioral properties of main verb. SAs resemble main verbs in the following respects: they do not undergo subject-auxiliary inversion, and they do not provide a locus for negative attachment; instead, SAs require do-support

in each of these cases. SAs share with auxiliaries the following semantic property: they are subject-raising predicates. That is, the particular SA does not impose any thematic role specification upon the subject argument, this specification being supplied by the downstairs verb" (Michaelis, 1998, p.85). This study, however, is based on examples from English, and therefore the criteria might not apply to the current set of phasal verbs.

4.1.2.2 Phasal verbs as complex predicates

According to (Eisele, 1992), aspectualizers form compound-verb phrases where they represent the highest verb which carries the deictic time reference, and in that they resemble *kān*. This is true for verbs such as *yibʔa* where: "*yibʔa* (imperfect of *baʔa* 'be, become') is used similarly to *kān*, in that it can embed most types of predicates to indicate temporal contrasts" (Eisele, 1992, p.154). Here it is clear that *baʔa* is treated by Eisele (1992) as another form of *kān* which is a true auxiliary, rather than as a phasal verb. *baʔa* is investigated in more details in section 4.3.1.

Maas (2009) has a similar account, where these verbs are treated as parts of a complex predicate. They are called 'coverbs' and are treated similarly to *kān* which also functions as a temporal marker for the whole construction.

4.1.2.3 Phasal verbs as raising verbs

In Michaelis (1998)'s account of these verbs as semi-auxiliaries (SAs), the semantic property that these verbs were assumed to share with auxiliaries is that "they are subject-raising predicates. That is, the particular SA does not impose any thematic role specification upon the subject argument, this specification being supplied by the downstairs verb" (Michaelis, 1998, p.85).

This view is also adopted in Noonan (2007) which gives the English example "(i) *R*am began to

eat meat, where Ram is clearly the subject of began. It is generally assumed that, in cases like (i), the subject of the complement has been raised to be matrix subject [see Newmeyer (1975), Langacker (1995) for discussion of this issue]" (Noonan, 2007, p.141).

Investigation of phasal verbs as raising predicates is tackled in more details in section 4.4.2 with respect to the data of the current study.

4.1.3 Complements of phasal verbs

In his account of complementation patterns of different verb classes, Noonan (2007) states that different phasal verbs might have different kinds of complements depending on the aspect to which the verb is inherently associated, such as "inception with inceptive (inchoative) aspect, continuation with progressive (durative) aspect, and termination with perfective (completive) aspect" (Noonan, 2007, p.140). These variations are also true of phasal predicates sharing the same meaning, where they might have different complement types, and tense and aspectual contrasts of complements of phasal predicates can lead to contrasts in meaning (Noonan, 2007). This is an important aspect of the description of phasal verbs, and it forms the basis for the categorisation of phasal verbs in the current study, shown in more details in section 4.3.

Eisele (1992) also pointed out that phasal verbs might have some restrictions on the kinds of complements they may have, where the following verb might not have an independent deictic time reference or they may be specified for their ability to be followed by a complementiser. For example the phasal verbs *fidil* 'remain' and *ʔaʕad* 'continue', *baʔa* 'become', *ʕād* 'go back', *ʔidir* 'can', *ʔarrab* 'be nearly', *ʕirif* 'know how', *liħiʔ* 'catch' do not take a complementiser, they have subject coreference with their complements and the embedded verb is not temporally discrete. While the verbs *ibtada* 'begin', *ħāwil* 'try' and *battal* 'stop' share the same properties with the first group, except that they may or may not take a complementiser. Comparison between the findings of Eisele (1992) and the current study is presented in section 4.5.

The following section, 4.2, includes the full list of the phasal verbs investigated in the current study, as well as criteria for choosing the verbs and extraction of the data sentences from the corpus.

4.2   Data

4.2.1   Choosing the verbs

The first stage of choosing the verbs was investigating the list of phasal verbs described in the previously mentioned studies which have counterparts in ECA, based on my knowledge as a native speaker of ECA. Since Eisele (1992)'s account was to my knowledge the most comprehensive account of phasal verbs in ECA, it formed the basis for choosing the list of phasal verbs to be investigated in the current study. However, some of the verbs mentioned in Eisele (1992)'s study were excluded as they do not represent phasal verbs, these are *ʔidir* 'can' and *ʕirif* 'know', where they do not denote a phase of an event or state. The verb *ʕād* 'go back' was also excluded as it is not used as a phasal verb in ECA, though it is a phasal verb in MSA.

I have also added the verb *kammil* 'continue' to the list of phasal verbs, based on its meaning which is listed as a phasal verb in other languages. However by searching for this verb through the corpus I found it had no verbal complements at all, were it was followed by nouns in all of the corpus sentences, representing the lexical use of the verb[3], and therefore it was excluded from the list of the verbs investigated in the study. The next section presents the full list of phasal verbs in ECA which are described in the current study.

---

[3]More on the lexical uses for phasal verbs is mentioned in section 4.2.3.

4.2.2 Verbs covered

The phasal verbs under investigation in the current study are: *badaʔ* 'begin', *baʔa* 'become', *fidil* 'remain', *liħiʔ* 'catch', *ʔarrab* 'be.near', *battal* 'stop', *rigiʕ* 'return', *ʔaʕad* 'go.on'.

The next step was to conduct a search for each of these phasal verbs through the ECA corpus that I have built for the purposes of this study[4]. I have searched for each of the phasal verbs in its different inflections in terms of number, gender, person, different markings of tense/aspect as well as different possible spellings. The sentences were then checked for the kind of complements each of the phasal verbs has, whether it is followed by a verb (finite or non-finite), participle, noun, prepositional phrase or a complementiser. This led to excluding the sentences in which the verb is used in its lexical sense rather than as a phasal verb. The is described in more details in the coming section, 4.2.3.

4.2.3 Phasal vs. Lexical verb

For some of the phasal verbs there were two meanings, one for the phasal verb and another for a lexical verb, but both meanings are written similarly and that is why they both occured in the search results. This is the case with the verb *ʔaʕad*, where the phasal verb means 'go-on', while the lexical use means 'sit'. In such cases, only the sentences representing the phasal verb were included in the data, and all other uses were excluded. There are cases also where the phasal verb can have the same meaning both as a lexical and a phasal verb, however the structure of the sentence would make it clear if the verb is lexical or phasal. For example in sentence 1 the coordination between *ʔirgaʕ* and *ʔeħki* seems to be between two lexical verbs:

(1) ʔirgaʕ   w ʔeħki   le-l-ʔiḫwa   btūʕ   lagnet   el-ʔamn
   return.IMP.2SGM and tell.IMP.2SGM to-the-brothers belonging committee the-security

   Go back and tell the security committe people

---

[4]Details about the corpus are in section 1.3.

Example 1 indicates two successive actions, one of returning somehwere and the other of telling the security people, which is different from the phasal verb in:

(2) ʔirgaʕ   ʔiħki   le-l-ʔiḫwa   btūʕ   lagnet   el-ʔamn
    return.IMP.2SGM tell.IMP.2SGM to-the-brothers belonging committee the-security
    Go back to telling the security committe people

Which means going back to telling the security people. Sentences such as 1 were therefore excluded from the data. I have also excluded other instances where any of the verbs was used in the lexical meaning, having nouns, adjectives or prepositional phrases as its complements. Such as:

(3) rigiʕ   el-beit
    return.PV.3SGM the-house
    He returned to the house

(4) el-doktōr rigiʕ   min   ʔagazt-u
    the-doctor return.PV.3SGM from vacation-his
    The doctor returned from his vacation

(5) yaʕni   el-leʕb baʔa   ʕala kbīr
    meaning the-play become.PV.3SGM on   big
    So it's now a big plan 'LIT: And now the playing became big'

baʔa 'become' was the only verb that occured having adjectives following it, as in:

(6) fi el-ħaml   b-abʔa   ʕasabeyya   w   ħassāsa   ʔawi
    in the-pregnancy BI-become.IPFV.1SG nervous.1SGF and sensitive.1SGF very
    During pregnancy I become very nervous and sensitive

baʔa can also be used in ECA as a discourse marker, as in:

(7) ʃufu   baʔa ħa-tgīb   fulūs ʔad   ʔēh
    see.IMP.2PL then FUT-get.IPFV.3SGF money amount what
    See then how much money she'll get

Where it does not denote a phasal verb, but rather an emphasis on the need to do something, where it can be dropped with no significant change in meaning, as shown below:

(8) ʃufu   ħa-tgīb   fulūs ʔad   ʔēh
     see.IMP.2PL FUT-get.IPFV.3SGF money amount what
     See how much money she'll get

Although the majority of the sentences in which *baʔa* occured in the corpus represented its use as a discourse marker, all such sentences were deleted from the data and I was careful to include only the sentences which represent its use as a phasal verb. The phasal verb is distinguished by its ability to inflect for the shared subject as the following lexical verb, while the discourse marker does not inflect. Example 27 in the data represents the phasal use of the verb.

ḫallas 'finish' was deleted from the list of phasal verbs although it was folllowed by verbs, but in all of these sentences it occupied a different clause than the following verb, so it did not form a phasal verb complementation, but rather a sequence of two lexical verbs, as in:

(9) w   baʕd-ma ḫallast   roħt   el-beit
     and after   finish.PV.1SG go.PV.1SG the-home
     After I finished I went home

Where *ḫallas* 'finish' here is part of an adjunct clause followed by the lexical verb, and the whole sentence represents two different successive actions, which can be easily separated by a pause or comma between 'finishing' and 'going'. This shows that this is a lexical use of the verb rather than a phasal verb. All of these sentences were excluded from the data as they represent instances of lexical verbs, not the phasal verb.

The list of sentences included in the data all represent phasal verbs as they are used in ECA, and where the phasal verb takes a verbal complement. Section 4.3 to follow presents examples for these sentences, grouped according to the behaviour of phasal verbs in terms of the kind of complements they take.

4.3   Behaviour of Phasal verbs

In this section, phasal verbs are divided into sets according to the behaviour of each, and the main criteria in distinguishing between the verbs and dividing them into different sets will be the complements each can take. Where the first group in section 4.3.1 are phasal verbs that can take complements of finite verbs, non-finite verbs and participles. While the second group is presented in section 4.3.2 for phasal verbs that take verbal complements of non-finite and imperative form verbs. While section 4.3.3 presents the phasal verb *battal* 'stop' which can only take verbal complements of non-finite verbs. Further tests are applied to each set to show its different patterns of behaviour, including negation, insertions and defining the arguments. Understanding these patterns help reach the suitable analysis for the phasal verbs.

4.3.1   *rigiʕ* 'return', *baʔa* 'become', *fidil* 'remain' and *ʔaʕad* 'go.on'

These are the only phasal verbs taking finite verbs, non-finite verbs and participles as verbal complements. Examples 10 and 11 show finite verb complements, examples 13 to 15 show non-finite verb complements, while examples 16 and 18 show participles as complements of the phasal verbs.

(10) baʔa   bi-yemʃi   zay el-batrīq
      become.PV.3SGM walk.BI-IPFV.3SGM like the-penguin

      Lit: He became (such that) he walks like a penguin

      He walks like a penguin

This implies the subject getting into a habit of walking like a penguin, which he did not have before.

(11) rigiʕna   ʔithaneʔna   men yum-ēn
      return.PV.1PL fight.PV.1PL from days-two

      And we went back to fight two days ago

This sentence indicates that we used to fight before and then stopped, but two days ago, we had a fight and that is how we know that we went back to the habit of fighting. It is also possible to have the same sentence with the lexical verb in the bare imperfective form:

(12) rigiʕna   nitḫāneʔ   men yum-ēn
     return.PV.1PL fight.IPFV.1PL from days-two
     And we went back to fighting two days ago

The only difference between the two sentences is that example 11 implies that an actual fight took place 2 days ago, while example 12 does not carry such an implication.

Phasal verbs can also take lexical verbs in the bare imperfective form[5] as their complements, as in:

(13) w   rigiʕ   y-salli   fi el-fasl
     and return.PV.3SGM pray.IPFV.3SGM in the-class
     And he went back to praying in class

(14) ʔaʕadna   nitmaʃʃa   ʃiwayya ʔana w   ʕabd-elħamīd
     go.on.PV.1PL walk.IPFV.1PL some   me   and Abdel-Hamid
     We kept on walking me and Abdel-Hamid

(15) w   fidil   yigri   wara el-nās
     and remain.PV.3SGM run.IPFV.3SGM behind the-people
     And he kept on running after the people

This set of phasal verbs can also be followed by participles, as in:

(16) maħaddeʃ baʔa   fāhim   bi-titkallef   kām
     nobody   become.PV.3SGM understand.PTCP.SGM BI-cost.IPFV.3SGF how.much
     Nobody any longer understands how much it costs!

This sentence indicates that there was a time that people knew how much things costs, but something happened that they do not know this information anymore. This change of state is indicated by the phasal verb *baʔa*, which is shown by the contrast with:

---

[5]Representing the non-finite verb form in ECA.

(17) maħaddeʃ fāhim   bi-titkallef   kām
      nobody   understand.PTCP.SGM BI-cost.IPFV.3SGF how.much
      Nobody understands how much it costs!

Which states the fact that nobody knows how much the cost is. Another example for participles
as complements is:

(18) w   fidilna   ʔaʕdīn   l-ħad   be-l-lēl
      and remain.PV.1PL sit.PTCP.PL to-extent at-the-night
      And we remained sitting till the night

Although participles are treated in Aboul-Fetouh (1969) as nouns, they have some distinct be-
haviours that make them resemble non-finite form verbs, such as having a relationship between
the verb stem and the form of active/passive participle, they are inflected for gender and num-
ber, and the fact that active participles can occur in verbal slots with time reference such as
following the auxiliary (Aboul-Fetouh, 1969).

This view is further supported by Gary and Gamal-Eldin (1982) in which participles are listed
as the non-finite forms of verbs, where both active and passive participles inflect for num-
ber and gender. Participles are defined in (Gary and Gamal-Eldin, 1982) as the non-finite
forms, and they are said to be voice marked in the stem, while they are not clearly marked for
tense/aspect, as noted in: "Non-finite forms do not overtly signal tense, though being marked
for imperfect/perfect aspects, they are time-coded. Thus the passive participle is compatible
with past time reference only, and the active participle with past and non-past" (Gary and
Gamal-Eldin, 1982, p.101). Therefore, sentences where phasal verbs are followed by participles
are included in the current study.

Following are the tests used to investigate the behaviour of this group of phasal verbs, including
negation, arguments, insertions and verb forms. Unlike the corpus examples mentioned above,
most of the sentences presented in the tests sections are corpus data that I have modified a bit,

as in adding negation to a verb to show the behaviour of the verbs under negation. While in case where the whole sentence is constructed it is marked by [constructed].

4.3.1.1 Negation

Negation of the phasal verb gives a different meaning than negation of the following lexical verb, as in:

(19) ʔirgaʕ   ʔiʃtaɣal   tāni
    return.IMP.2SGM work.IMP.2SGM again
    Go back to working again

(20) ʔirgaʕ   ma-tiʃtaɣal-ʃ   tāni
    return.IMP.2SGM NEG-work.IPFV.2SGM-NEG again
    Go back to not working again

(21) ma-tirgaʕ-ʃ   tiʃtaɣal   tāni
    NEG-return.IPFV.2SGM-NEG work.IPFV.2SGM again
    Do not go back to working again

Both the phasal verb and the lexical verb can be negated in the same sentence[6] as shown in:

(22) ma-tirgaʕ-ʃ   ma-tiʃtaɣal-ʃ   tāni
    NEG-return.IPFV.2SGM-NEG NEG-work.IPFV.2SGM-NEG again
    Do not go back to not working again (which means something like 'work!)

This might form evidence that each verb forms a separate clause, where negation on one does not scope over the other, but each can be negated separately giving a different meaning. Note also that negation can occur on any verb form, not only the imperative, such as the following sentence where the phasal verb is in the perfective form:

(23) rigiʕ   ma-yiʃtaɣal-ʃ   tāni
    return.PV.3SGM NEG-work.IPFV.3SGM-NEG again
    He went back to not working again

---

[6]Note that when the imperative verb is negated, the stem has the same form as the bare imperfective and that is why it is glossed as IMPFV in the examples. While the negated verb as a whole represents an imperative form, which is shown by the free translation.

It is important to note here that in a sequence of imperative verbs, negating one of them is possible only when the lexical verb is in the non-finite form, as in the ungrammaticality of:

(24) *ʔirgaʕ   ma-ʔiʃtaɣal-ʃ   tāni
     return.IMP.2SGM NEG-work.IMP.2SGM-NEG again

This might be due to restrictions on the occurence of the negation circumfix on the imperative verb forms, as is shown in example 20 above where the negation is on the bare (non-finite) form. This goes in accordance with Benmamoun (2000)'s account of negation of imperatives in Arabic, stating that the verb form in the positive imperative is different from the negative imperative, which is rerpesented by the bare imperfective in Egyptian Arabic. Another difference also is that the positive imperative lacks the person agreement prefix, while it is present on the negative imperative. And therefore the positive imperative can be said to use a different verbal paradigm than the negative imperative (Benmamoun, 2000).

4.3.1.2   Arguments

All of these phasal verbs share the same subject with the following lexical verb in all of the examples, as in:

(25) w   baʔa   bi-yetfeteħ   fi ʔai waʔt
     and become.PV.3SGM BI-open.IPFV.3SGM in any time
     And it now opens anytime

Where the subject is the same for both verbs, the phasal verb and the following lexical verb. All of the examples for the phasal verb *baʔa* occured in the 3SGM form as their subjects were 3SGM, however the phasal verb occurs in any form that agrees with the subject, as in the 3SGF form in:

(26) baʔit   b-temʃi   zay el-batrīq
     become.PV.3SGF BI-walk.IPFV.3SGF like the-penguin
     She (became) walking like a penguin

The phasal verb obligatorily agrees with the subject in person, number and gender, as shown by the ungrammaticality of:

(27) *baʔa   b-temʃi   zay el-batrīq
     become.PV.3SGM BI-walk.IPFV.3SGF like the-penguin

4.3.1.3   Insertions

It is possible for an adverb to occur between the phasal verb and the lexical verb, as in:

(28) ħafdal   tūl ʕomri ʔākul   kintāki yaʕni?
     remain.FUT.1SG all age   eat.IPFV.1SG Kentucky as-if?

     Will I stay all my life eating Kentucky(KFC)?

(29) baʔēt   dayman b-atʔaḫḫar
     become.PV.1SG always   BI-be.late.IPFV.1SG

     I became always (being) late

It is also possible to have some adverbs that place one of the verbs in a different temporal frame from the other, as in:

(30) w rigiʕ   imbāriħ y-tmarran   kol yōm
     and return.PV.3SGM yesterday train.IPFV.3SGM every day

     And he went back yesterday to training every day

This means that the returning to the habit of training happened yesterday, but the training itself is happening every day from now on. This might show further evidence of the structure being bi-clausal, where each verb has a different temporal indication.

The adverb can also occur between the phasal verb and the following participle, as in:

(31) fawwāz baʔa   dayman lābis   el-galabeyya
     Fawwaz become.PV.3SGM always   wear.PTCP.SGM the-galabeyya

     Fawwaz (became) always wearing the galabeyya (gown)

4.3.1.4   Verb forms

Each of the phasal verbs in this group can occur with lexical verb complements having one of different forms of tense/aspectual marking, however there are some restrictions on the form of the lexical verb according to the marking of the phasal verb, the possible combinations are listed as follows:

Phasal verb	Verbal complement
Perfective	Perfective/ bi-prefixed/Bare/Participle
bi-prefixed	bi-prefixed/Bare/Participle
ħa-prefixed	bi-prefixed/Bare/Participle
Imperative	Imperative

Table 4.1: Verb forms for *rigiʕ*, *baʔa*, *fidil* and *ʔaʕad*

Table 4.1 shows that all forms of the phasal verbs can be followed by a bi-prefixed or a bare Imperfective form lexical verb, except for restrictions on the imperative which is discussed in examples 35 and 36. The reason why phasal verbs in this group can have bi-prefixed lexical verbs as their complements[7] might be the fact that they are associated with a progressive aspect (continuation), where the bi-prefixed verb form is also used to denote progressive or habitual aspect in ECA. This might also indicate that the marking on the lexical verb denotes aspect rather than tense.

Each of the combinations in table 4.1 gives a different interpretation, corresponding to those in compound tense structures discussed in section 2.6, where the phasal verb marks tense while the lexical verb marks aspect[8]. For example the combination between Perfective phasal verb and a Perfective lexical verb gives an interpretation of past perfect, presented in example 11, repeated below:

---

[7]Unlike the phasal verbs in sections 4.3.2 and 4.3.3 which can only be followed by a bare Imperfective lexical verb form.

[8]This is accounted for in the analysis section 4.4 where the phasal verb occupies I marking tense while the lexical verb occupies V and marks grammatical aspect.

(32) rigiʕna   ʔithaneʔna   men yum-ēn
     return.PV.1PL fight.PV.1PL from days-two
     We went back to fight two days ago

The interpretation here is past perfect, where past tense is marked on *rigiʕna* 'we went back' and the lexical verb *ʔithaneʔna* refers to the completed event of fighting. While the interpretation is different if the lexical verb is marked as bi-prefixed, as in example 33:

(33) rigiʕna   b-nethāneʔ   kol   yōm
     return.PV.1PL BI-fight.IPFV.1PL every day
     We went back to fighting every day

Here the sentence refers to a past habitual event, where the phasal verb refers to a past tense event while the lexical verb adds the habitual aspect, hence the addition of the adverb *kol yōm* every day. A third possible combination is:

(34) ħa-nirgaʕ   nethāneʔ   tāni
     FUT-return.IPFV.1PL fight.IPFV.1PL again
     We will get back to fighting!

Example 34 marks future tense on the phasal verb while the lexical verb is in the bare Imperfective form.

The bare imperfective (non-finite) form of the lexical verb can occur following any form of the phasal verb, except for the imperative. The imperative form of the phasal verb can only take imperative lexical verbs as their complements, therefore it is ungrammatical to have an imperative phasal verb followed by bare imperfective (except in negation). Example 35 shows the combination between an imperative phasal verb and an imperative lexical verb, while example 36 shows the combination between an imperative phasal verb and a lexical verb in the bare Imperfective fom, which yeilds an ungrammatical sentence.

(35) ʔirgaʕ   ʔiħki   le-l-ʔiħwa   btūʕ   lagnet   el-ʔamn
     return.IMP.2SGM tell.IMP.2SGM to-the-brothers belonging committee the-security
     Go back to telling the security committe people

(36) *ʔirgaʕ   tiħki   le-l-ʔiḫwa   btūʕ   lagnet   el-ʔamn
      return.IMP.2SGM tell.IPFV.2SGM to-the-brothers belonging committee the-security

In his study of the morphology of ECA, Aboul-Fetouh (1969) claims that the imperative is
not a tense or aspect marking, but rather forms a separate paradigm of verb forms having a
prefix *ʔi* attached to the stem of the bare imperfective, and a suffix showing gender and number
agreement. This restriction on the form of the lexical verb following the imperative phasal verb
seem to be a case of obligatory agreement in form between the two imperative verbs, as in cases
of pseudocoordination and re-analysis in Lodrup (2014), which is attested with some motion
and aspectual verbs and where imperative agreement is said to be the most common form of
verbal agreement.

Section 4.3.2 below includes description of the data for the second group of phasal verbs. The
order of presentation is the same as this group, where example sentences are shown for each
phasal verb, followed by tests for negation, arguments, insertions and investigating the different
possible forms of the phasal verbs and their verbal complements.

4.3.2   *badaʔ* 'begin', *liħiʔ* 'catch' and *ʔarrab* 'be.near'

This group of phasal verbs can only have verbal complements of non-finite verbs, as well as
sequences of imperatives. Examples 37 to 39 show the phasal verb followed by a lexical verb
in the bare Imperfective form as its complement, while example 40 shows an imperative phasal
verb followed by an imperative lexical verb as its complement.

(37) ʔalb saʕd badaʔ   yitħarrak
      heart Saad start.PV.3SGM move.IPFV.3SGM
      Saad's heart started to move (beat)

(38) ʔana ʔarrabt   ʔamūt   min el-ḫōf
      I   be.near.PV.1SG die.IPFV.1SG from the-fear
      I'm almost scared to death!

(39) lamma ʔalħaʔ   ʔarrawwaħ   ʕaʃān   ʕand-ena ħafla ʔennaharda
     when   catch.IPFV.1SG go.home.IPFV.1SG because at-us   party today

     I need to go home soon as we have a party tonight

While an imperative phasal verb is followed by an imperative lexical verb, as in:

(40) ʔilħaʔu   ʔeʃteru   w ʔeʕmelu   ʃēr
     catch.IMP.2PL buy.IMP.2PL and do.IMP.2PL share

     Go buy now (before the offer ends) and share (on twitter)

4.3.2.1   Negation

Negation of the phasal verb gives a different meaning from negation of the lexical verb, as in:

(41) ʔaħmad badaʔ   yismaʕ   mazzīka ɣarbi
     Ahmed start.PV.3SGM listen.IPFV.3SGM music   western

     Ahmed started listening to Western music

(42) ʔaħmad ma-badaʔ-ʃ   yismaʕ   mazzīka ɣarbi   lessa
     Ahmed NEG-start.PV.3SGM-NEG listen.IPFV.3SGM music   western yet

     Ahmed haven't started listening to Western music yet

(43) ʔaħmad badaʔ   ma-yismaʕ-ʃ   ɣēr   mazzīka ɣarbi
     Ahmed start.PV.3SGM NEG-listen.IPFV.3SGM-NEG except music   western

     Ahmed started not listening to any music except Western music

4.3.2.2   Arguments

In all of the sentences, the phasal verb and the following lexical verb share the same subject, as in:

(44) ʔana ʔarrabt   ʔamūt   min el-ḫōf
     I   be.near.PV.1SG die.IPFV.1SG from the-fear

     I'm almost scared to death!

However objects of the lexical verb seem to be arguments of the lexical verb only, and not the phasal verb:

(45) bada?   yiktib   ?imēl tawīl
     begin.PV.3SGM write.IPFV.3SGM email long

     He started to write a long email

(46) el-?imēl   bada?   yitkitib
     the-email begin.PV.3SGM write.PASS.IPFV.3SGM

     The email started to be written

Here only the lexical verb is passivised, while the phasal verb is in the active form.

4.3.2.3   Insertions

Adverbial material may separate the phasal verb from the verbal complement, this applies to all phasal verbs in this group, as shown by the following examples.

(47) bada?   waħda waħda yiktiʃif   tarī?-u fi el-ħayāh
     begin.PV.3SGM one   one   discover.IPFV.3SGM way-his in the-life

     He started to discover his way in life a bit by bit

Here the phasal verb is followed by two identical successive adverbs *waħda waħda* showing the manner. They are used this way in ECA, where it is not possible to use only one of them to give the same meaning, instead it will just mean the number one.

It is also possible to have an adverb denoting a different temporal marking for each of the verbs, as in:

(48) bada?t   embāriħ ?azākir   l-ħad   el-sāʕa sabʕa be-llēl
     begin.PV.1SG yesterday study.IPFV.1SG to-extent the-hour seven at-night

     I started yesterday to study till 7 at night

This means that yesterday I started being in the habit of studying till 7 at night, which might be extending till now (the habit of studying till 7 at night, not the beginning). This again shows that each verb can give a separate marking for morphoyntactic tense/grammatical aspect.

4.3.2.4   Verb Forms

As with the previous group of verbs, the phasal verb can occur in any possible marking for morphosyntactic tense/grammatical aspect, while the following lexical verb occurs only in the bare Imperfective form for this set of verbs. The only exception is the imperative phasal verb, where if the phasal verb is in the imperative form, the following lexical verb should also occur in the imperative. The possible forms of verbs are illustrated as follows:

Phasal verb	Verbal complement
Perfective	Bare
bi-prefixed	Bare
ħa-prefixed	Bare
Imperative	Imperative

Table 4.2: Verb forms for *badaʔ*, *liħiʔ* and *ʔarrab*

These phasal verbs all have the inceptive aspect inherent in their meanings, which might be the reason they only take bare Imperfective form lexical verbs as their complements, as the latter does not mark a grammatical aspect that may contradict with that of inception, such as the perfective or progressive. However, validation of this assumption needs a deeper semantic analysis which is beyond the focus of the current study.

The following section, 4.3.3, presents description of the data for the phasal verb *battal*, which differs in the kinds of verbal complements it takes from the other phasal verbs, and this is why it is included in a separate section. The order of presentation is the same as the previous groups of phasal verbs, where it starts by some example sentences from the corpus data, followed by tests for investigating the behaviour of this phasal verb in terms of negation, insertions, its arguments and the possible forms of the phasal verb and its verbal complements.

4.3.3 *battal* 'stop'

This phasal verb can only take non-finite form verbs as its verbal complements, while no participles or sequence of imperatives are allowed as its complements. Example 49 shows *battal* with a bare Imperfective verb complement.

(49) el-ʃorta   battalit   tiʕazzib   fi el-gamāʕāt el-ʔislameyya
     the-police stop.PV.3SGF torture.IPFV.3SGF in the-groups the-Islamic
     The police stopped torturing the Islamic groups

Unlike the rest of the phasal verbs included in the current study, the imperative form of *battal* takes a bare Imperfective lexical verb as its complement, as shown in:

(50) battal   tifakkar   kitīr ʔizzāy momken tebʔa   ʕazīm
     stop.IMP.2SGM think.IPFV.2SGM lots how   possible become.IPFV.2SGM great
     Stop thinking a lot about how to be great!

While it is ungrammatical for the imperative form of *battal* to take a verbal complement in the imperative form, as shown by the ungrammaticality of:

(51) *battal   fakkar   kitīr ʔizzāy momken tebʔa   ʕazīm
     stop.IMP.2SGM think.IMP.2SGM lots how   might   become.IPFV.2SGM great

4.3.3.1   Negation

As with all of the previous phasal verbs, negation of *battal* gives a different meaning from negation of the following lexical verb, as in:

(52) mobail ħasan   battal   yiʕmil   sōt
     mobile Hassan stop.PV.3SGM do.IPFV.3SGM sound
     Hassan's mobile stopped making a sound

(53) mobail ħasan   ma-battal-ʃ   yiʕmil   sōt
     mobile Hassan NEG-stop.PV.3SGM-NEG do.IPFV.3SGM sound
     Hassan's mobile didn't stop making a sound

(54) mobail ħasan battal   ma-yiʕmil-ʃ   sōt
     mobile Hassan stop.PV.3SGM NEG-do.IPFV.3SGM-NEG sound

     Hassan's mobile stopped not-making a sound (being silent)

It is also possible to have negation on both *battal* and its verbal complement, as in:

(55) mobail ħasan ma-battal-ʃ   ma-yiʕmil-ʃ   sōt
     mobile Hassan NEG-stop.PV.3SGM-NEG NEG-do.IPFV.3SGM-NEG sound

     Hassan's mobile didn't stop not-making a sound (being silent)

4.3.3.2 Arguments

battal shares the same subject with its verbal complement in all of the sentences, as shown in example 56 where the speaker is the subject of both verbs:

(56) battalt   ʔalʕab   baskit
     stop.PV.1SG play.IPFV.1SG basketball

     I stopped playing basketball

While arguments of the lexical verb are not arguments of *battal*, this is shown by the inability of the object of the lexical verb to occur as a pronoun on the phasal verb, while it can be incorporated with the lexical verb only, as in:

(57) battalt   ʔalʕab-u
     stop.PV.1SG play.IPFV.1SG-it

     I stopped playing it

(58) *battalt-u   ʔalʕab-u
     stop.PV.1SG-it play.IPFV.1SG-it

4.3.3.3 Insertions

None of the sentences analysed for *battal* had any adverbs separating it from its lexical verb complement, however it is possible as shown by example 59 where I have added an adverb to the corpus sentence:

(59) mobail ħasan battal   fagʔa   yiʕmil   sōt
mobile Hassan stop.PV.3SGM suddenly do.IPFV.3SGM sound

Hassan's mobile has suddenly stopped making a sound

This adverb can occur in any of multiple positions across the sentence with no change in meaning, as shown in:

(60) mobail ħasan battal   yiʕmil   sōt   fagʔa
mobile Hassan stop.PV.3SGM make.IPFV.3SGM sound suddenly

Hassan's mobile has suddenly stopped making a sound

4.3.3.4 Verb forms

As mentioned at the beginning of section 4.3.3, *battal* can only have non-finite verbal complements, even when *battal* is in the imperative form. These are shown in table 4.3:

Phasal verb	Verbal complement
Perfective	Bare
bi-prefixed	Bare
ħa-prefixed	Bare
Imperative	Bare

Table 4.3: Verb forms for *battal*

In terms of the lexical aspect of the phasal verb, *battal* is the only phasal verb having the completive aspect inherent in it. However, I am not sure this has any relation to the fact that *battal* can only have non-finite complements. Explaining this property of *battal* might be interesting for future work on phasal verbs.

Investigation of the phasal verbs included a detailed account of the ways in which each behaves in terms of negation, their ability to have adverbial elements preceding its verbal complements, the way they share arguments with the lexical verb, and the forms of the phasal verbs and their verbal complements. As a result of this account, phasal verbs are grouped into three sets according to the forms of the verbal complements each can take: either non-finite verb forms

only, non-finite forms as well as imperative verbs, or finite, non-finite, as well as imperative verbs. All of the phasal verbs can be negated separately from their complements, can have adverbs separating them from their verbal complements, share the same subject with their lexical verb complements. These findings can help reach the suitable syntactic analysis for phasal verbs in ECA, which is the focus of section 4.4 to follow.

4.4 Analysis

This section is devoted to reaching the suitable analysis for phasal verbs in ECA based on the patterns of behaviour of the data discussed in section 4.3. Questions about the behaviour of the phasal verbs showed no particular difference between the three sets, in terms of negation, insertions and arguments. The main difference was in the range of possible verb forms for verbal complements of each set, which seems to be related to the lexical aspect inherent in each of the sets, where the division into inception, continuation and termination seem to be consistent with the possible verb forms for complements of each set.

The analysis is divided into subsections according to the questions asked about the analysis of phasal verbs. Section 4.4.1 aims to answer the question of whether the phasal verb and the following lexical verb form a mono-clausal or bi-clausal structure, section 4.4.2 investigates the kind of control relation between the two verbs, while section 4.4.3 investigates how information about morphosyntactic tense and grammatical aspect are specified on the phasal verb construction. The analysis is then presented in terms of the lexical entries of the verbs in section 4.4.4, the f-structures in section 4.4.5, as well as the c-structures in section 4.4.6.

The proposed analysis is then implemented using XLE in chapter 6, so as to devise a set of rules that are capable of producing all and only grammatical sentences, as well as showing the correct analysis for each. This grammar is capable of producing the correct f-structures and

c-structures for the analyzed sentences, as well as devise a way to account for the restrictions on the verb forms of the verbal complements and their combinations with different phasal verbs.

4.4.1   Mono or Bi-clausal?

This section aims to investigate whether the structure formed of the phasal verbs followed by verbal complements represent a mono-clausal or a bi-clausal structure. I will investigate the properties of two types of mono-clausal structures, which are serial verbs and complex predicates, in an attempt to either prove that phasal verbs belong to one of those types or eliminate the possiblity of a mono-clausal structure for phasal verb constructions in ECA. I start with the investigation of properties of serial verbs, where Aikhenvald and Dixon (2006) accounts for some of the properties of serial verbs, that include sharing the same tense, aspect, mood and polarity on both verbs of the serial verb construction (SVC), where some languages might mark tense, aspect or mood similarly on both verbs, or on only one of the verbs in SVC. Regarding negation, "There can only be one negator per SVC. It can either have the whole construction as its scope (...) or part of the construction" (Aikhenvald and Dixon, 2006, p.8).

Investigation of the behaviour of phasal verbs shows that they do not form a serial verb construction, this is based on evidence such as the fact that for each of these verbs, negation of the phasal verb gives a different meaning from negation of the lexical verb, where either one of the verbs or both can be negated giving different meanings. In addition to the possibility of having distinct modifiers for each of the phasal verb and its verbal complement, including different temporal/aspectual markings. Another piece of evidence also is the fact that the arguments of the lexical verb are not arguments of the phasal verb, where they can not be passivised or incorporated with the phasal verbs as pronouns, as shown in:

(61) badaʔt   abʕat-l-aha   gawāb kol   yūm
      start.PV.1SG send.IPFV.1SG-to-her letter   every day

      I started sending her a letter everyday   [constructed]

(62) *bada?t-l-ha   ab\Sat   gawāb kol   yūm
      start.PV.1SG-to-her send.IPFV.1SG letter   every day

[constructed]

Another kind of mono-clausal construction is complex predicates, which are described in Butt (1995) as having a complex argument structure (two or more semantic heads contribute arguments) while the grammatical functional structure is that of a simple predicate, where there is only a single predicate (a nuclear PRED) and a single subject. An example of a complex predicate is the permissive in Urdu, which is a mono-clausal structure where both verbs behave as a single predicate in terms of agreement and anaphora, and they are both controlled with the same subject (Butt, 1995).

This resembles the behaviour of the phasal verbs under discussion here, where the phasal verb shares the subject with the following lexical verb, and so they both behave as a single clause with respect to anaphora (as they share the same subject). However they are not considered complex predicates based on evidence from behaviours such as having separate arguments (other than the subject) and different negation patterns and temporal markings and the different modifiers each can have. Investigation of properties of serial verbs and complex predicates leads to elimination of the possiblity of a mono-clausal structure for phasal verbs, and therefore suggests that the structure is indeed bi-clausal while the subject is shared between the phasal verb and the lexical verb, which suggests a control relation betwen the two verbs, investigated in section 4.4.2 to follow.

4.4.2   Kind of control relation

In LFG terms, control implies the presence of several properties such as: clause boundary between a matrix clause and a subordinate clause, as well as an NP that is an argument of both clauses, representing the controller in the matrix clause and the controllee (seen as a miss-

ing/implicit argument) in the subordinate clause (Kroeger, 2004). It is also correlated with the presence of a non-finite verb with no overt SUBJ in the subordinate clause (Falk, 2001)[9].

Since all of these are properties of the phasal verbs constructions which are investigated in the current chapter, the question to ask here is what is the kind of control relation between the two verbs, whether it is anaphoric control or functional control. Properties of each in (Kroeger (2004); Falk (2001); Dalrymple (2001)) were tested against the behaviour of phasal verbs under investigation. Evidence seem to show that there is a functional control relation of the type raising, based on criteria such as:

- The controller is the subject in the matrix clause, which corresponds to the controllee as the subject of the subordinate clause. And the controllee can not be given an overt expression.

- An idiom part can be the SUBJ without losing the idiomatic meaning, as in:

(63) ʔīd-i   rigʕet   takol-ni   kol   ma   ʔaʃūf   sigāra
      hand-my return.PV.3SGF eat.IPFV.3SGF-me every when see.IPFV.1SG cigarette
      I went back to having an urge to smoke (LIT: My hand started to eat me every time
      I see a cigarette)

The structure of the complement clause is compatible with an XCOMP analysis, as no complementiser is possible in the complement clause of any of the phasal verbs. This is shown by the ungrammaticality of:

(64) *baʔit   ʔin-aha b-temʃi   zay el-batrīq
      become.PV.3SGF that-her BI-walk.IPFV.3SGF like the-penguin

(65) *battalt   ʔin-i ʔalʕab   baskit
      stop.PV.1SG that-I play.IPFV.1SG basketball

---

[9]However, control can also occur into finite clauses, as discussed in section 1.5.2

Another question to ask is whether the subject is a thematic argument of the phasal verb or not. Dowty (1991) discusses some properties of the proto-agent that can be a thematic argument of a verb, including the volitional involvement in the event, sentience, causing a change of state in another participant, and the ability to exist independently of the event (Dowty, 1991).

Application of these properties to the subject of phasal verbs seem to show that the subject is not a thematic argument of the phasal verb. This is shown by examples such as:

(66) battal-it   timattar?
     stop.PV.3SGF rain.IPFV.3SGF?
     Did it stop raining?   [constructed]

In this sentence the subject is translated as 'it', which refers to whatever is raining, this can be substituted by a noun refering to the world/weather, as in:

(67) el-gaw   baʔa   saʔʕa ʔawi
     the-weather become.PV.3SGM cold   very
     It became very cold   [constructed]

(68) el-donia   regʕet   timattar   tāni
     the-world return.PV.3SGF rain.IPFV.3SGF again
     It went back to raining   [constructed]

In all these sentences, the noun refers to some state of the weather as raining or being cold, which involves no volitional agent. Therefore, the subject of the phasal verb is assumed to be non-thematic.

To sum up the discussion on control so far, the fact that the embedded subject can not be expressed rules out optional anaphoric control, while the idiom test suggests functional control with a non-thematic subject, hence raising. This analysis is represented in the lexical entry for the phasal verb, in section 4.4.4, the f-structure presented in section 4.4.5, as well as the c-structure in section 4.4.6. However, these are only examples for the representations, while

the whole list of lexical entries, f-structures and c-structures are included in the XLE grammar
fragment described in chapter 6. The following section 4.4.3 describes how information about
tense and grammatical aspect are specified on the phasal verb construction.

4.4.3   On tense and aspect

Marking of morphosyntactic tense and grammatical aspect on the phasal verb construction
follows the same concept adopted for compound tense in chapter 2, where the phasal verb
denotes tense while the lexical verb denotes grammatical aspect. This is shown in the contrast
between the examples 69 and 70 shown below:

(69) rigiʕna   ʔitḫaneʔna   men yum-ēn
     return.PV.1PL fight.PV.1PL from days-two
     And we went back to fight two days ago

This sentence indicates that we used to fight before and then stopped, but two days ago, we
had a fight and that is how we know that we went back to the habit of fighting, which gives the
sentence a punctual and perfective meaning. This is different if the lexical verb occurs in the
bare imperfective, as in:

(70) rigiʕna   nitḫāniʔ
     return.PV.1PL fight.IPFV.1PL
     And we went back to fighting

Where this sentence lacks the perfective denotation, but just states that we went back to fight-
ing without giving any indication on whether a real fight occured or not. This is crucial to the
assumption held in the current work that the embedded verb (also refered to here as lexical
verb) denotes aspect, which is perfective in 69 due to the lexical verb having the perfective form.
Note also that this meaning is inherent irrespective of the presence of the temporal adverb *men
yum-ēn* 'two days ago', which further shows that aspect is a property of the lexical verb.

The same assumption of specification of tense on the phasal verb while the lexical verb adds gramamtical aspect also applies to cases where the lexical verb adds a habitual aspect to the sentence, as in:

(71) baʔēna   dayman bi-nitʔaḫḫar
      become.PV.1PL always   BI-late.IPFV.1PL

      We (became that) we are always late   [constructed]

Where here the phasal verb puts the whole event of becoming in the past tense while the lexical verb adds the habitual aspect as it is in the form of bi-prefixed verb. This shows in the f-structure of the construction as adding the feature HAB to the XCOMP, shown in section 4.4.5.

4.4.4 Lexical Entry

Results of the previous tests of behaviour show that the ECA phasal verbs investigated in the current study are raising verbs, and the differences between the complements each set can have shows no evidence to treat them differently in terms of the analysis. Phasal verbs occupy I and carry morphosyntactic tense marking, this applies to all of the phasal verbs investigated in this chapter, while they differ in some features of their lexical entry according to the properties associated with each set, this is exemplified with each lexical entry.

The first lexical entry is for a phasal verb from the first group discussed in section 4.3.1, namely the verbs: *rigiʕ* 'return', *baʔa* 'become', *fidil* 'remain' and *ʔaʕad* 'go.on'. The lexical entry specifies the PRED value, the control relation, the morphosyntactic tense marking of the verb, as well as the restriction on imperative phasal verbs to take only imperative verbal complements:

fidil: I

(↑PRED)= 'remain< (↑ XCOMP)> (↑ SUBJ) '

(↑ SUBJ) = (↑ XCOMP SUBJ)

(↑ TENSE)= PAST

ʔirgaʕ: I

(↑PRED)= 'return< (↑ XCOMP)> (↑ SUBJ) '

(↑ SUBJ) = (↑ XCOMP SUBJ)

(↑ VFORM)= IMP

(↑ XCOMP VFORM)=c IMP

While a member of the second group of phasal verbs discussed in section 4.3.2, namely the verbs *badaʔ* 'begin', *liħiʔ* 'catch' and *ʔarrab* 'be.near' needs additional specification of the form of the verbal complement as non-finite verb only, except for imperatives which are followed by imperatives only:

badaʔ: I

(↑PRED)= 'begin< (↑ XCOMP)> (↑ SUBJ) '

(↑ SUBJ) = (↑ XCOMP SUBJ)

(↑ TENSE)= PAST

(↑ XCOMP VFORM)=c BARE

ʔilħaʔ: I

(↑PRED)= 'catch< (↑ XCOMP)> (↑ SUBJ) '

(↑ SUBJ) = (↑ XCOMP SUBJ)

(↑ VFORM)= IMP

(↑ XCOMP VFORM)=c IMP

battal 'stop', discussed in section 4.3.3, is the only phasal verb with the specification to take only bare Imperfective form verbs in its complement, in all cases including the imperative:

battalit: I

(↑PRED)= 'stop<(↑ XCOMP)> (↑ SUBJ) '

(↑ SUBJ) = (↑ XCOMP SUBJ)

(↑ TENSE)= PAST

(↑ XCOMP VFORM)=c BARE

battalu: I

(↑PRED)= 'stop<(↑ XCOMP)> (↑ SUBJ) '

(↑ SUBJ) = (↑ XCOMP SUBJ)

(↑ VFORM)= IMP

(↑ XCOMP VFORM)=c BARE

On the other hand, lexical verbs occuring as complements to the phasal verbs occur in V and carry grammatical aspect if they are finite, else they are specified as having bare Imperfective form:

yigri: V

(↑PRED)= 'run<(↑ SUBJ)>'

(↑VFORM) = BARE

2ithane2na: V

(↑PRED)= 'fight<(↑ SUBJ)>'

(↑ASPECT) = PERF

4.4.5   F-structure

This section includes examples for f-structures of the data sentences for phasal verbs, where each sentences is glossed followed by the f-structure. The first f-structure represents phasal verbs of the first group represented in section 4.3.1, these phasal verbs can take verbal complements of either finite or non-finite form verbs.

(72) rigiʕna   ʔitḫaneʔna
     return.PV.1PL fight.PV.1PL

     And we went back to fight

$$
\begin{bmatrix}
\text{PRED} & \text{`return <XCOMP> SUBJ'} \\
\text{SUBJ [1]} & \begin{bmatrix} \text{PRED `PRO'} \end{bmatrix} \\
\text{TENSE} & PAST \\
\text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{`fight <SUBJ>'} \\ \text{SUBJ [1]} \\ \text{ASPECT} & PERF \end{bmatrix}
\end{bmatrix}
$$

(73) baʔēna   dayman bi-nitʔaḫḫar
     become.PV.1PL always   BI-late.IPFV.1PL

     We (became that) we are always late   [constructed]

$$
\begin{bmatrix}
\text{PRED} & \text{`become <XCOMP> SUBJ'} \\
\text{SUBJ [1]} & \begin{bmatrix} \text{PRED `PRO'} \end{bmatrix} \\
\text{TENSE} & PAST \\
\text{XCOMP} & \begin{bmatrix} \text{PRED} & \text{`be.late <SUBJ>'} \\ \text{SUBJ [1]} \\ \text{ASPECT} & IMPF \\ \text{HAB} & + \end{bmatrix} \\
\text{ADJUNCT} & \begin{bmatrix} \text{PRED `always'} \end{bmatrix}
\end{bmatrix}
$$

While phasal verbs of the second group, represented in section 4.3.2, can take verbal complements in the bare Imperfective form only, and their f-structure is shown below.

(74) badaʔ   yiktib   ʔimēl
     begin.PV.3SGM write.IPFV.3SGM email

     He started to write an email

$$
\begin{bmatrix}
\text{PRED} & \text{`start <XCOMP> SUBJ'} \\
\text{SUBJ [1]} & \begin{bmatrix} \text{PRED } \textit{`PRO'} \end{bmatrix} \\
\text{TENSE} & \textit{PAST} \\
\text{XCOMP} & \begin{bmatrix}
\text{PRED} & \text{`write <SUBJ, OBJ>'} \\
\text{SUBJ [1]} & \\
\text{OBJ} & \begin{bmatrix} \text{PRED } \textit{`email'} \end{bmatrix} \\
\text{VFORM} & \textit{BARE}
\end{bmatrix}
\end{bmatrix}
$$

4.4.6   C-structure

Despite the fact that the phasal verbs are divided into three groups based on the kinds of verbal complements each may have, they share the same c-structure representation, where the phasal verb occupies I and the verbal complement occupies V as part of the XCOMP. This is shown by the c-structure below:

(75) battalt   ʔalʕab   baskit
     stop.PV.1SG play.IPFV.1SG basketball

     I stopped playing basketball

```
                    IP
                    |
                    I'
                    |
                 ↑ = ↓
               /        \
          I                VP
       ↑ = ↓           (↑XCOMP) = ↓
          |              /        \
        battalt       V            NP
                   ↑ = ↓        (↑OBJ) = ↓
                      |             |
                   ʔalʕab        baskit
```
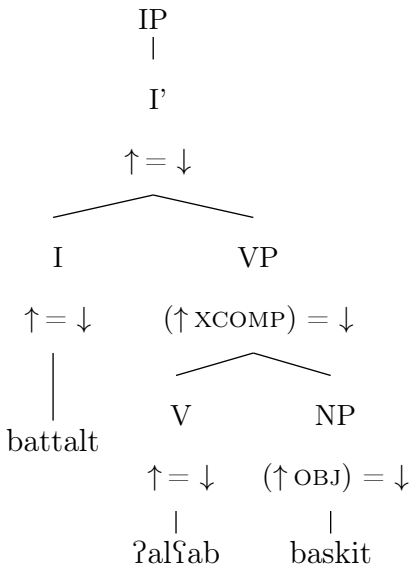
Having presented the data and their different patterns of behaviour as well as the analysis for the phasal verbs throughout sections of this chapter, section 4.5 below concludes the main findings of the chapter, including comparison between the findings of the current study and previous accounts of phasal verbs in the literature, as well as the main features of the syntactic analysis for these structures.

4.5   Conclusions

The current chapter is concerned mainly with the investigation of the behaviour of phasal verbs in ECA, answering questions such as whether they can be divided into groups according to the forms of their verbal complements, and whether these groups would have different syntactic representations in terms of their lexical entries, f-structures and c-strucutures. Results show that phasal verbs do indeed show three different patterns of behaviour in terms of their verbal complements, where one group of phasal verbs take either finite or non-finite form verbal complements, while the second group can have non-finite verbal complements for all verb forms except the imperative phasal verb which takes imperative verbal complements, while the phasal verb *battal* can have only non-finite verbal complements for all forms of phasal verbs, including the imperative.

There are a number of properties of phasal verbs investigated in the current study that agree with Eisele (1992)'s account for phasal verbs in ECA, such as:

- They do not subcategorize for a complementiser, and this applies to all phasal verbs investigated in the current study (unlike in Eisele (1992) where there were exceptions).

- Obligatory subject coreferentiality between the phasal verb and the lexical verb, which is shown in the control relation between the main clause and the subordinate clause.

- The phasal verb specifies tense (deictic time reference) while the lexical verb can only specify grammatical aspect (non-deictic time reference). This is shown by the specfication of tense and aspect in section 4.4.3.

- The selectional restriction mentioned in Eisele (1992) for *fidil* 'remain' as allowing active participles as its complement holds for all of the other phasal verbs in the first group of phasal verbs in the current study in section 4.3.1, including the verbs: *rigiʕ* 'return', *baʔa* 'become', *fidil* 'remain' and *ʔaʕad* 'go.on', while it does not hold for the other two sets of phasal verbs.

- However the features specifying phasal verbs as auxilliaries were not investigated here, as it seemed irrelevant to the purposes of the current study.

With respect to the syntactic analysis, the status of phasal verbs as complex predicates was investigated within the principles of LFG, and evidence showed that the structure of the phasal verb and the lexical verb is indeed bi-clausal, which contradicts with their status as complex predicates. This seems to contradict with some of the previous studies mentioned here which described phasal verbs as complex predicates (Eisele (1992); Maas (2009)), although these studies did not provide much syntactic evidence as to why the phasal verbs are considered complex predicates, so it seemed to be a rather general term based on the fact that these verbs co-occur

as 'coverbs' as described by Maas (2009), and they have only one deictic time reference speci-fied by the phasal verb as described by Eisele (1992). Therefore, the analysis of phasal verbs as raising verbs holds for the verbs analyzed in the current study, which agrees with the accounts for phasal verbs in (Michaelis (1998); Noonan (2007)).

Chapter 5

Modals

"Modality is one of the 'golden oldies' among the basic notions in the semantic analysis of language. But, in spite of this, it also remains one of the most problematic and controversial notions: there is no consensus on how to define and characterise it, let alone on how to apply definitions in the empirical analysis of data." (Nuyts, 2005, p.5). Crosslinguistically, modals pose a challenge to empirical studies due to the fact that they often have a defective paradigm (Bhatt et al., 2011). It is also the case in some languages that modals can be expressed by non-verbal elements, such as the case with modals in Egyptian Colloquial Arabic where they can be verbs, nouns or participles, with both inflecting and non-inflecting forms. However, this is not an exception, as stated by (Hacquard, 2011): "Modality is expressed by many categories of lexical items: adverbs like *maybe*, nouns like *possibility*, adjectives like *possible*, or auxiliary verbs like *must, may, should* or *have to*."

This chapter aims to investigate the forms that can be used in Egyptian Colloquial Arabic to express modality, as well as describing the structures in which those modals are used. I start with a brief description of the notion of modality in section 5.1, then a discussion of the literature of modals in ECA in section 5.2. The list of modals in ECA as well as description of the corpus-based data is presented in section 5.3, where I have categorised the modals based

on the kind of complements they may have, as well as whether they inflect to agree with the subject or not. Section 5.4 includes answering a number of questions about the structures in which modals are used in ECA, with the aim of reaching the suitable analysis for these modals, which is then presented in terms of lexical entries, f-structures and c-structures. Finally, section 5.5 concludes the main findings of this chapter.

5.1   Defining modals

A good starting point in order to define modality is to account for the set of semantic categories that express modality, which include: dynamic, deontic and epistemic modality (Nuyts, 2005). Following is a brief account of each of these categories as described in Nuyts (2005):

- Dynamic modality (also called 'inherent modality' or 'facultative modality'): which expresses ability, where "the subject is able to perform the action expressed by the main verb in the clause" (Nuyts, 2005, 7). Examples in English are the modal verbs 'can' and 'able to'. ECA dynamic modals are: *yiʔdar* 'able' and *yinfaʕ* 'can'.

- Deontic modality: "may be defined as an indication of the degree of moral desirability of the state of affairs expressed in the utterance, typically but not necessarily on behalf of the speaker" (Nuyts, 2005, p.9). This includes a gradual scale between notions including permission, desirability, acceptability, obligation, unacceptability. Examples in English include the modal verbs: 'must', 'may', 'insist'. It is debatable whether notions such as volition and intention can also be included under deontic modality, these include the English auxiliary-like 'want' and the auxiliary 'will'. ECA deontic modals are: *ʕāwiz* 'want', *lāzim* 'must', *darūri* 'should', *el-mafrūd* 'the-ought-to', *nāwi* 'intend', *mosammem* 'insist' and *nefsi* 'wish'.

- Epistemic modality: is defined as expressing "the degree of probability of the state of affairs" (Nuyts, 2005, p.10), an English example is 'maybe'. ECA epistemic modals are: *momken* 'possible' and *yimken* 'may'.

These definitions are used only as a heuristic to delimit the range of my investigation and define the set of modals that can be included in the current study, where the modals under investigation are categorised in section 5.3 on syntactic and morphosyntactic grounds rather than on their semantic properties. Next, section 5.2 accounts for the literature review on modals in ECA.

5.2   Literature review

In his extensive study of modal verbs in Egyptian Arabic, Azer (1980) used sentences elicited from informants as well as texts and plays written in Egyptian Arabic to investigate the semantic and syntactic properties of modals. The set of modals analyzed include: *lāzim* 'must', *labud* 'have to', *darūri* 'should', *el-mafrūd* 'ought to', *momken* and *yisaħ* 'may' , *muħtamal* and *gāyiz* and *yimkin* 'can', *ʕāwiz* 'want', *yiʔdar* 'can', *yinwi* 'intend', *yuʔʕud* 'remain', *yifdal* 'keep.on', *yiħāwil* 'try', *yinsa* 'forget', *yiftikir* 'remember', *yirūħ* 'go', *yiħib* 'love'. Although he does mention some of the phonological and syntactic properties of modals in his study, the focus of Azer (1980)'s study is mainly on the semantics and pragmatics of these structures. According to Azer (1980), modals are divided into two main categories: Non-inflective modals and inflective modals, and each of these categories is divided into sub-categories according to the type of modality each member expresses (obligation, possibility, etc.). The data in his study was based on elicited sentences by informants as well as texts and plays written in the dialect, however, the dialect variety represented is different from that analyzed in my work, as Azer (1980)'s sentences represent an old variety of colloquial Arabic, where many of these structures are no longer used and did not figure in the corpus search I have conducted as part of my current study. An example is *muħtamal* 'maybe' which is an MSA word that is not used as a modal in ECA nowadays, as well as *darūri* 'should' which is not used productively in ECA as shown in more details in section 5.3.1. Azer (1980) also included some verbs that do not seem to express modality, such as *yiħib* 'love' which is a stative verb and *yifdal* 'keep.on' which is a phasal verb that is investigated as part of the current study in section 4.3.1. These factors led to the many

differences between my current work and Azer (1980)'s study, though it formed a good starting point for investigating modals in the dialect.

Despite the fact that modality can be expressed through verbal forms, these verbal expressions seem to form a class of modal verbs that can be distinguished from other verbs in a language in terms of their morphosyntactic properties which are lacking from other verbal classes. This is the position adopted by Zayed (1983) in his study of modals in Literary Arabic[1], where he assumes that modals form a subclass of verbs where they share some properties with main verbs, such as having an imperfective form, negation using *lā* as well as interrogation with wh-words. Examples for the modals investigated by (Zayed, 1983) are: *yagib* 'must/have to', *yanbaghi* 'have to', *yumkin* 'possible' and *yastatīʕ* 'can/be able to', where they occur in the Imperfective verb form[2], they can be negated using *lā* as in *lā yagiba* 'mustn't', and they can be interrogated using the yes-no question marker *hal*, as in *hal yagiba?* 'Is it necessary?' (Zayed, 1983).

On the other hand, these modals differ from main verbs in other properties such as having no perfect form or imperative form, lacking person inflection, the imperfect form being referentially neutral where it only occurs in one form, for example *yagib* 'must' which does not agree with the subject as in *ʔana yagib* 'I must', *hiyya yagib* 'she must' (Zayed, 1983). Some of these properties can be applied to modals in ECA while others can not be applied such as forming question using *hal* as this question marker is not used in ECA.

Al-Sabbagh et al. (2014) was the only study which investigated modality in modern Egyptian Arabic, however it is a corpus-driven study where the main focus is on the process of building an annotated corpus for epistemic modality called '3arif', formed of tweets in Egyptian Arabic and Modern Standard Arabic (Al-Sabbagh et al., 2014). However this corpus is discourse based, where the aim of building it was "to automatically identify the key players of Twitter's political discourse in countries of political unrest such as Egypt" (Al-Sabbagh et al., 2014, 1521). Despite

---

[1]The term 'Literary Arabic' is used to refer to the written form of Standard Arabic.
[2]Referred to in the current study as the bare Imperfective form.

the fact that this corpus was not available for me to use, it shows the lack of resources that represent the modern use of ECA, and specifically modality, which shows the importance of having corpus-based studies such as my current work.

Having defined the notion of modality and presented a number of different studies in the literature for modals in Arabic as well as ECA, the next section 5.3 includes the full list of ECA modals investigated in the current study and the sentences in which they occured in the corpus as well as description of the behaviour of each set of modals.

5.3   Data

The data on which the discussion in this chapter is based is comprised of sentences in which modals were used in ECA texts online. Each of the sentences is the result of a web based search for modals through the corpus that I have constructed for the purposes of the current study, details about building of the corpus and the methodology of data collection is presented in section 1.3. The corpus-based search included each modal in its different contexts such as negation and different tenses. For example the inflecting modal *yiʔdar* 'able' was searched for in forms like *ma-yeʔdar-ʃ* 'he's not able to', *teʔdar* 'she can' and *ħaneʔdar* 'we will be able to'. Detailed description of the process of searching through the corpus and examples for seed words used is included in section 1.3.2. Choosing the list of ECA modals to be investigated in the current study is based mainly on the modals mentioned in Azer (1980)'s work after filtering out the ones that are not used in the dialect any more, based on my judgement as a native speaker.

Due to the tendency of the modals included in the data to have slight changes in meaning between different languages, the translation of the sentences into English was sometimes tricky. Therefore in such cases, I have tried to give an idiomatic free translation which captures the nuances of meaning and usage as closely as possible to the English free translation, where the

meaning given in the glossing is the closest to the knowledge of the native speaker, while the translation might differ according to its use in the sentence. This is shown by example 1 where the modal *lāzim* is glossed as 'must' while in the sentence it is translated as 'should':

(1) kont   lāzim ʔaʕmil   kors   mokaθθaf maʕ mostafa
     be.PV.1SG must   do.IPFV.1SG course intensive with Mostafa

     I should've made an intensive course with Mostafa   [constructed]

Following is a representation of the modals in ECA, divided according to their structures into subsets including inflecting modals and non-inflecting modals. For each modal, the set of sentences which occured in the corpus including this modal followed by a lexical verb are presented, along with sentences showing its behaviour under negation, when it is combined with past tense auxiliary, whether it can have a noun/adverb inserted between the modal and the following lexical verb, whether it can be followed by a complementiser and its ability to have arguments different from those of the following lexical verb. Answering these questions aims to help reach the best analysis for the structures in which these modals occur, whether they are mono or bi-clausal, as well as the kind of raising/control relation between the modal and the following verb. Following is a detailed account of each of the modals investigated and examples for sentences in which they occur in the corpus.

5.3.1   *darūri* 'necessary'

The modal *darūri* 'necessary' is a highly marginal form in contemporary ECA on which this study is based, it is used more frequently in MSA while in ECA it is not used as productively as the other modals. This is based on the fact that it occured in only two sentences in the corpus, which is an indication on its restricted use. *darūri* has the morphosyntactic properties of a noun meaning 'necessary', where it can occur in the plural form *darūriyyāt* 'neccessities' (Wehr, 1976), as shown in the MSA example:

(2) ʔal-māʔ   w-al-hawāʔ men darūriyyāt al-ḥayāh
     the-water and-the-air from neccesseties the-life

     Water and air are among the life neccessities   [constructed]

Modal *darūri* is an invariant form that does not inflect in any way, and does not occur in the

plural form in ECA in its modal use, moreover, it does not seem to even have a subject unlike

the case with other non-inflecting modals, this is shown by the difference between examples (3)

and (4) below:

(3) ʔaḥmad lāzim yīgi
     Ahmed must come.IPFV.3SGM

     Ahmed must come   [constructed]

(4) #ʔaḥmad darūri   yīgi
     Ahmed   necessary come.IPFV.3SGM

                                                                      [constructed]

These are sentences that I have constructed to illustrate my point, they show that *darūri* is

not used in the same context as *lāzim* which is a non-inflecting modal that can take a noun

in subject position, however adding a noun in subject position for *darūri* yields sentence (4)

unacceptable[3]. The only two sentences in the corpus found for *darūri* 'necessary' were:

(5) lāzim el-motadayyen yizlim,   darūri   yaʕni?
     must the-religious   oppress.IPFV.3SGM, necessary as.in

     Is it a must that the religious person would oppress people, is it necessary?!

(6) ʔai ḥad ʕand-u maʕlūma   yʔol-l-i   darūri
     any one at-him information tell.IPFV.3SGM-to-me necessary

     Any one who has information should let me know

In both sentences, *darūri* does not take a verbal complement, but is rather used to emphasize the

meaning of obligation rather than introduce it. It expresses some kind of speech act resembling

the function of the imperative verb but it is not the main predicate of the sentence, where it

can be dropped with no change in meaning, as in:

---

[3]Note however that example (4) would be totally acceptable according to Azer (1980)'s study which represents
ECA in the time of the study which was published in the 1980s. This contrast in acceptability is very tricky to
pinpoint as it shows differences among the same dialect spoken at different times in history.

(7) ʔai ħad ʕand-u maʕlūma   yʔol-l-i
    any one at-him information tell.IPFV.3SGM-to-me
    Any one who has information should let me know

This tends to be the common use for *darūri* in contemporary spoken Egyptian Arabic. Due to this restriction on its use, it was not included as a modal in the current study.

5.3.2 *yimken* 'may'

This modal has the invariant form of a bare Imperfective verb with the default non-referential 3SGM agreement, and its meaning indicates possibility. It represents a separate category in the current study due to the fact that it is the only modal that can be followed by both finite and non-finite verbal complements[4], as shown in the following corpus sentences:

(8) yimken   yilāʔi   ʔilli yiʃrab-u
    may.IPFV.3SGM find.IPFV.3SGM that drink.IPFV.3SGM-it
    Maybe he'll find something to drink

(9) yimken   bi-yħāsib   el-nās   men gēb-u
    may.IPFV.3SGM BI-pay.IPFV.3SGM the-people from pocket-his
    Maybe he's paying the people from his own money

(10) yimken   ʔabil-hum   w   ana ma-ʕraf-ʃ
    may.IPFV.3SGM meet.PV.3SGM-them and I   NEG-know.IPFV.1SG-NEG
    Maybe he met them but I don't know

yimken occurs in the invariant 3SGM form and does not inflect with the subject, as shown in example 11 below where the noun preceding *yimken* in the subject position is 3SGF, and the verb *tebʔa* 'become' agrees with it, while *yimken* is in the 3SGM form:

(11) heyya yimken   tebʔa   ʃatra f-el-ʕulūm   bas gahla
    she   may.IPFV.3SGM become.IPFV.3SGF clever in-the-sciences but ignorant
    b-ħoʔoʔ-ha
    with-rights-her
    She might be clever at the sciences, but she's ignorant about her rights

---

[4]The bare imperfective verb form is considered to be the non-finite form in ECA, this is discussed in more details in section 1.4.

When *yimken* is followed by a finite verb form as in example 10 above, it seems to resemble the behaviour of a modal adverbial such as 'possibly' in English. However, the position where *yimken* can occur in the sentence is not as free as would be expected from such an adverbial, as is shown in:

(12) *ʔabil-hum   yimken w   ana ma-ʕraf-ʃ
     meet.PV.3SGM-them may   and I   NEG-know.IPFV.1SG-NEG

This shows that *yimken* is not an adverb but is rather used as a modal expression. It can also take the bare imperfective form of the auxiliary *ykūn* as its verbal complement, as in:

(13) yimken   ykūn   misāfir
     may.IPFV.3SGM be.IPFV.3SGM travel.PTCP.SGM
     Maybe he's travelling

This usage of the auxiliary *ykūn* in the bare imperfective form to denote a state is very common following modals as will be shown in the examples to follow, however it is not peculiar to modals only, where it can also occur following other verbs such as the stative verb *yiʕraf* 'know' in:

(14) ʔel-dakatra bi-yeʕrafu   ykūnu   hadyīn waʔt el-ʔazma
     the-doctors BI-know.IPFV.3PL be.IPFV.3PL calm   time the-crisis
     Doctors know how to be calm during a crisis

These examples show that *yimken* is a modal expression that can have verbal complements in both finite and non-finite forms. Following is a description of the different patterns of behaviour of this modal, in terms of the possiblity of having adverbial expressions between the modal and its verbal complement in section 5.3.2.1.1, its negation patterns in section 5.3.2.1.2, how it forms past tense in section 5.3.2.1.3, its arguments in section 5.3.2.1.4 and finally section 5.3.2.1.5 provides a summary for the patterns of behaviour of *yimken* investigated in the current study.

5.3.2.1 Behaviour of *yimken*

5.3.2.1.1 Insertions The subject can occur between the modal *yimken* and its verbal complement, as in:

(15) yimken   bāba ʔabil-hum   w   ana ma-ʕraf-ʃ
      may.IPFV.3SGM dad   meet.PV.3SGM-them and I   NEG-know.IPFV.1SG-NEG
      Maybe dad met them but I don't know

Adverbs can also occur between the modal and the lexical verb, as in:

(16) yimken   ʔaxīran ʔabil-hum   w   ana ma-ʕraf-ʃ
      may.IPFV.3SGM finally   meet.PV.3SGM-them and I   NEG-know.IPFV.1SG-NEG
      Maybe finally he met them but I don't know

Note that the position of the adverb is free in the sentence, where it can occur in different positions without changing the meaning of the sentence, while the occurence of the modal is more restricted. This shows that *yimken* is not classified as an adverb. This property is tackled in more details in section 5.4 to prove the status of modals as heads of the phrases where they occur.

5.3.2.1.2   Negation   Negation seems to occur only on the lexical verb following *yimken*, as in:

(17) yimken   ma-ykūn-ʃ   misāfir
      may.IPFV.3SGM NEG-be.IPFV.3SGM-NEG travel.PTCP.SGM
      Maybe he's not travelling

Negation on *yimken* gives a different form which resembles the MSA modal *yumken* 'can' mentioned in Zayed (1983)'s study of modals in Literary Arabic, it is not used in ECA as a modal verb but it occurs only in this negated form, negated using *la* 'not', as in[5]:

(18) la   yumken   ʔansa   mawqif-hom   fi yanāyir
      NEG may.MSA.IPFV.3SGM forget.IPFV.1SG attitude-their in January
      I can't forget their attitude in January

(19) la   yumken   yiʔdaru   yisrifu   ʕala el-daʕāya
      NEG may.MSA.IPFV.3SGM can.IPFV.3PL afford.IPFV.3PL on   the-advertisement
      They can't afford the advertisement

---

[5]*yumkin* is glossed as 'may.MSA' to show that this is the form used in Modern Standard Arabic and differentiate it from the form in which it is used as a modal in ECA, which is *yimken*.

This form *la yumken* is used in ECA only in this structure, negated and meaning something is impossible. *la* is used as a negative particle for imperfective verb forms in MSA where it immediately precedes the verb (Sharif and Sadler, 2009). It is also mentioned in Zayed (1983)'s study as a common characteristic between modals and other classes of verbs in Literary Arabic, where they all can be negated using negation using *lā*, as in the example *lā yagiba* 'mustn't'. However, this negation pattern is different from negation on verbs in ECA, which is expressed using the prefix *meʃ* or the circumfix *ma-ʃ* according to the form of the verb (Gary and Gamal-Eldin, 1982). Example 49 repeated here shows negation on the modal *yinfaʕ* 'can' using the circumfix *ma-ʃ*:

(20) ʔakīd el-wā7id ma-yenfaʕ-ʃ yʕīʃ lwaħdo
sure the-one NEG-can.IPFV.3SGM-NEG live.IPFV.3SGM own
For sure one can't live on his own= One has to live with people

On the other hand, there is only one case where the negative prefix *meʃ* can combine with *yimken*, which is when it is used in questions, as in example 21 below:

(21) meʃ yimken ykūn da howwa el-maʃrūʕ ʔilli b-neħlam
NEG may.IPFV.3SGM be.IPFV.3SGM this he the-project that BI-dream.IPFV.1PL
bi-h?
with-it?
Couldn't this be the project of our dreams?

Example 21 does not negate the posibility (as is expected negating *yimken*), but rather denotes a question where the speaker wonders if some event might take place, so it can also mean "This could be the project of our dreams". While to negate the possiblity I would use the modal *momken* [6] in its negated form instead, as in:

(22) meʃ momken ykūn da howwa el-maʃrūʕ ʔilli b-neħlam bi-h
NEG possible be.IPFV.3SGM this he the-project that BI-dream.IPFV.1PL with-it
This couldn't be the project of our dreams!

These examples show that the negative forms for *yimken* seem to be defective, where the form *la yumken* is borrowed from MSA, while the form *meʃ yimken* is used only in specific contexts.

---

[6]This modal is represented in details in section 5.3.3.3.

5.3.2.1.3   Past   The past tense auxiliary expressing tense can occur either preceding or fol-
lowing *yimken* and it refers to a situation that did not take place. Although the corpus did not
have any examples I find the structure possible as in the examples below:

(23) kānu   yimken   yisamḥū-k   law kont   ʔeʕtazart
      be.PV.3PL may.IPFV.3SGM forgive.IPFV.3PL-you if   be.PV.2SGM apologise.PV.2SGM
      Maybe they would've forgiven you if you've apologised   [constructed]

Note that *yimken* occurs in the default 3SGM form although the auxiliary agrees with the 3PL
subject. It is also possible for the auxiliary to occur in the default 3SGM form preceding yimken:

(24) kān   yimken   yisamḥū-k   law kont   ʔeʕtazart
      be.PV.3SGM may.IPFV.3SGM forgive.IPFV.3PL-you if   be.PV.2SGM apologise.PV.2SGM
      Maybe they would've forgiven you if you've apologised= It's possible that.. [constructed]

While when the auxiliary occurs following *yimken*, agreement is obligatory:

(25) yimken   kānu   yisamḥū-k   law kont   ʔeʕtazart
      may.IPFV.3SGM be.PV.3PL forgive.IPFV.3PL-you if   be.PV.2SGM apologise.PV.2SGM
      Maybe they would've forgiven you if you've apologised   [constructed]

(26) *yimken   kān   yisamḥ-ūk   law kont   ʔeʕtazart
      may.IPFV.3SGM be.PV.3SGM forgive.IPFV.3PL-you if   be.PV.2SGM apologise.PV.2SGM
                                                                             [constructed]

Note that example (24) can also be translated as 'it's possible', which might be evidence that
the modal forms a raising structure. The analysis of *yimken* as a raising predicate is discussed
in more details in section 5.4.3.1.

5.3.2.1.4   Arguments   This section investigates whether the modal *yimken* can have argu-
ments different from those of its verbal complement. *yimken* is a non-inflecting modal that
shows different patterns of behaviour from those of lexical verbs, where it can not be passivised
and does not occur in the imperative. However, examples show that the subject of *yimken*
should always be the same as that of the following lexical verb:

(27) ʔaħmad yimken   yīgi   ʔennaharda
     Ahmed may.IPFV.3SGM come.IPFV.3SGM today

     Ahmed might come today   [constructed]

Here *Ahmad* is the subject shared between the modal and the following lexical verb, while it is ungrammatical to have different subjects, as in:

(28) *ʔaħmad yimken   Mona tīgi   ʔennaharda
     Ahmed   may.IPFV.3SGM Mona come.IPFV.3SGF today

                                                              [constructed]

Despite the fact that the subject is shared between *yimken* and its vebral complement, it seems to lack any thematic function with respect to *yimken*, which is shown by the fact that it does not agree with the subject in:

(29) yimken   Mona tīgi   ennharda
     may.IPFV.3SGM Mona come.IPFV.3SGF today

     Mona might come today= It's probable that Mona comes today   [constructed]

This suggests that the subject is a non-thematic argument of *yimken*. Further discussion of the status of *yimken* as a raising predicate is tackled in more details in section 5.4.3.1. On the other hand, *yimken* does not share arguments of its verbal complement other than the subject, as shown in the ungrammaticality of:

(30) *yimken-l-aha   baʕat   gawāb
     may.IPFV.3SGM-to-her send.PV.3SGM letter

                                                              [constructed]

As opposed to the case where the recipient pronoun is attached to the lexical verb, yielding a grammatical structure, as in:

(31) yimken   baʕat-l-aha   gawāb
     may.IPFV.3SGM send.PV.3SGM-to-her letter

     Maybe he sent a letter to her   [constructed]

Therefore the data shows that although *yimken* obligatorily shares the same subject with its verbal complement, other arguments of the lexical verb are not arguments of *yimken*. This will be useful in the discussion of the structure as bi-clausal, presented in section 5.4.1. A summary of the behaviour of *yimken* as investigated by the data so far is presented in section 5.3.2.1.5 below.

5.3.2.1.5   Summary   Table 5.1 summarises the behaviour of *yimken* as described in the preceding sections. The column 'inflects' refer to its ability to inflect, 'followed by finite' refers to its ability to be followed by finite form lexical verbs, 'Insertions' refer to its ability to have nouns/adverbs inserted between the modal and the lexical verb, 'Negation' refers to whether the position of negation (on the modal or on the lexical verb) gives different meanings or not, and 'Different SUBJ' refer to whether the modal can have a SUBJ argument different from that of the following lexical verb or not.

Inflects	Followed by finite	Insertions	Negation	Different SUBJ
No	Yes	Yes	Yes[7]	No

Table 5.1: Behaviour of *yimken*

5.3.3   Non-inflecting modals

This section includes modals which do not show agreement with the subject in either gender, number, or person, but instead have an invariant form. They also share the property of having verbal complements only in the bare Imperfective form. The modals are the verb *yinfaʕ* 'can', participle *lāzim* 'must', the definite noun *el-mafrūd* 'the-ought-to' and the noun *momken* 'possible'. Example sentences are represented for each modal separately in the following subsections, while the behaviour of the whole set of non-inflecting modals in terms of negation, arguments, insertions and past tense is represented in section 5.3.3.5.

---

[7]Note the change in form of *yimken* under negation, in section 5.3.2.1.2.

5.3.3.1 *lāzim* 'must'

lāzim is a modal that is used to imply an external obligation stating how things should be done. Morphologically, it has the form of the participle from the root *l-z-m*, however it is not used productively as a participle in ECA. Example sentences for the modal *lāzim* in the corpus are:

(32) w   lāzim ykūn   ḥad   ʔiqāma   b-niyyet   el-hegra
      and must be.IPFV.3SGM take.PV.3SGM residency with-intention the-immigration
      And he must've had the residency with the intention of immigration

(33) w   lāzim tibʔa   fārid   dahr-ak   w   ʔinta ʔāʕid
      and must become.IPFV.2SGM straight.PTCP.SGM back-your and you   sit.PTCP.SGM
      And you must keep your back straight while sitting

The meaning of *lāzim* in these examples can be translated as 'should' with no major difference in meaning as far as this study is concerned, therefore both 'must' and 'should' might be used interchangeably in the following sentences.

A slightly stronger degree of obligation might be refered to in:

(34) da   kalb marīd w   lāzim yitḥebes   fi ʔafas
      this dog sick   and must lock.PASS.IPFV.3SGM in cage
      This is a sick dog that must be locked up in a cage

This sentence can be used to give an order, where the speaker is able to lock the dog up, and in this case it represents an authoritative use.

5.3.3.2 *el-mafrūd* 'the-ought.to'

The modal *el-mafrūd* is morphologically a passive participle prefixed with the definite article *el-*, the passive participle is derived from the root *f-r-d* denoting obligation, and hence it is glossed as 'the-ought-to'. It has an invariant form that does not show agreement with the subject. *el-mafrūd* is often used in a higher register of the dialect with lots of code-switching with MSA, however the context of the sentences is colloquial, as shown by the example:

(35) meʃ el-mafrūd   ʔattisil   elʔān b-zumalāʔ-i
     NEG the-ought.to call.IPFV.1SG now   with-colleagues-my

     I'm not supposed to call my colleagues now

Note here the use of the words *elʔān* 'now' and *zumalāʔ-i* 'my colleagues' which are MSA words
that have other equivalents in the dialect, however the sentence occured within an ECA context
as shown by the use of the negation particle *meʃ*.

el-mafrūd can also be used to give the meaning 'supposedly', in which case it is followed by the
finite form of the lexical verb:

(36) homma dōl   el-nās   ʔilli ʔinti el-mafrūd   b-toʕʕodi   maʕā-hom
     they   these the-people that you   the-ought.to BI-sit.IPFV.2SGF with-them

     These are the people that you (supposedly) sit with

Note that in example 36 the modal has a different meaning in the translation than that glossed
'ought to', where here it means a hypothetical situation rather than an obligation. This meaning
is also mentioned in Azer (1980) as the common use for *el-mafrūd*: " This non-inflective modal
is never used in the logical inference sense. It is very much like 'ought to' in English. Apart
from expressing 'duty' as its basic meaning, this modal can also express 'predictivity'. To many
speakers of this dialect (myself included) this predictivity sense amounts to nothing more than
an expression of likelihood" (Azer, 1980, p.111).

Therefore this represents a different meaning for the modal which is marked by the form of
the verbal complement of the modal, where if the lexical verb occurs in a finite form then the
modal would express likelihood, while if the lexical verb occurs in the bare Imperfective verb
form then the modal *el-mafrūd* would express obligation, as in:

(37) homma dōl   el-nās   ʔilli ʔinti el-mafrud   toʕʕodi   maʕāhom
     they   these the-people that you   the-ought.to sit.IPFV.2SGF with-them

     These are the people that you're supposed to sit with

The structure of the modal *el-mafrūd* investigated in this section is the one denoting obligation, where it behaves similiarly to the other non-inflecting modals which are followed by non-finite verbal complements. The other use meaning 'supposedly' shall be represented as a separate form having a different lexical entry.

5.3.3.3   *momken* 'possible'

The modal *momken* expresses possibility, and it was the most frequent modal as it retrieved the biggest number of results for modals in the corpus. It does not agree with the subject and is always followed by a bare Imperfective verb.

(38) momken yinsā-ni   baʕd fatra
      possible forget.IPFV.3SGM-me after while

      He might forget me after a while

(39) ʔiħna momken nittifiʔ   fi noʔta wa7da
      we   possible agree.IPFV.1PL in point one

      We may agree in one point

Note that *momken* is glossed here as 'possible', as it is morphologically a noun, unlike *yemken* which is derived from the same root *m-k-n* and also indicates possiblity, however *yemken* is morphologically a verb in the bare Imperfective form, and hence it is glossed as 'may.IPFV.3SGM'.

5.3.3.4   *yinfaʕ* 'can'

The meaning of the modal *yinfaʕ* resembles that of *yiʔdar*[8] to a great extent, where both express ability. However *yiʔdar* is glossed as 'able' as it is used more to refer to human ability to do something, while *yinfaʕ* is glossed as 'can' as it is used often to refer to inanimate capability, or something as being generally possible. *yinfaʕ* is morphologically a verb in the bare imperfective form, however it has the invariant 3SGM form as shown in:

---

[8]*yiʔdar* is an inflecting modal discussed in section 5.3.4.1.

(40) yinfaʕ   ʔagadded-u   bas meʃ delwaʔti
      can.IPFV.3SGM renew.IPFV.1SG-it but NEG now

      I can renew it, but not now

(41) w   dah yinfaʕ   yitʕimil   fi zarf ḥamas sinīn
      and this can.IPFV.3SGM do.PASS.IPFV.3SGM in time five   years

      And this can be done in five years time

The verbal complement of *yinfaʕ* must always occur in the bare Imperfective form, as is shown

in the ungrammaticality of:

(42) *w   dah yinfaʕ   ʔitʕamal   fi zarf ḥamas sinīn
      and this can.IPFV.3SGM do.PASS.PV.3SGM in time five   years

Example 43 shows the case where *yinfaʕ* is followed by a finite form lexical verb, however each

of the verbs occupy a different clause:

(43) law yinfaʕ   ʔolī-l-i
      if   can.IPFV.3SGM tell.IMP.2SGF-to-me

      If it can be done, tell me

Therefore such sentences are not considered as falling within the domain of the current study

as the verb following the modal is not its verbal complement, but rather occupies a different

clause. Having listed the non-inflecting modals along with examples for each, the following

section, 5.3.3.5, aims to investigate the different patterns of behaviour of non-inflecting modals,

in terms of negation, insertions, arguments, past tense formation, as well as their ability to take

a complement clause headed by a complementiser.

5.3.3.5   Behaviour of non-inflecting modals

This section aims to show the common behaviours among non-inflecting modals as well as the

differences between them regarding a number of criteria such as: negation, combination with

past tense auxiliaries as well as the possibility for an adverbial expression to separate the modal

from its verbal complement. Each of these criteria is discussed below.

5.3.3.5.1 Insertions It is possible for the subject to occur between the modal and the lexical verb, this is true for all of the non-inflecting modals:

(44) kān   el-mafrūd   morsi yedẖol
     be.PV.3SGM the.ought.to morsi enter.IPFV.3SGM
     Morsi should've entered

The subject can also occur as an NP formed of two nouns, representing a construct state, as in example 45 below where the subject is *Ɂasħāb el-bēt* 'owners of the house':

(45) w   lāzim Ɂasħāb el-bēt   yirāʕu   Ɂinn-ohom yiftaħu   mawadīʕ
     and must owners the-house consider.IPFV.3PL that-they   open.IPFV.3PL topics
     mohemma
     important
     And the hosts should consider opening interesting topics

It is also possible for an adverb to occur between the modal and the lexical verb, as in:

(46) kān   momken giddan ma-ykammel-ʃ
     be.PV.3SGM possible very   NEG-continue.IPFV.3SGM-NEG
     It was (very) possible that he doesn't continue

5.3.3.5.2 Negation can occur either on the modal or the main verb, with each giving a different meaning.

(47) lāzim ma-ykon-ʃ   fi-h darar
     must NEG-be.IPFV.3SGM-NEG in-it harm
     There must be no harm in it

Here the sentence expresses an obligation to not have any harm in it, while sentence 48 denies this kind of obligation:

(48) meʃ lāzim ykūn   fi-h darar
     NEG must be.IPFV.3SGM in-it harm
     It doesn't have to be harmful

Negation occurs as a circumfix on *yinfaʕ*:

(49) ʔakīd el-wā7id ma-yenfaʕ-ʃ   yʕīʃ   lwaħdo
      sure   the-one   NEG-can.IPFV.3SGM-NEG live.IPFV.3SGM own
      For sure one can't live on his own= One has to live with people

It can also occur on the following lexical verb, giving a different meaning, as in:

(50) ʔakīd el-wā7id yenfaʕ   ma-yʕiʃ-ʃ   lwaħdo
      sure   the-one   can.IPFV.3SGM NEG-live.IPFV.3SGM-NEG own
      For sure one can live not on his own= One can live with people

The same contrast is found with negation on sentences having *momken*, as in the contrast shown

below:

(51) ʔinta meʃ momken tekūn   teʕraf-u
      You   NEG possible be.PV.2SGM know.IPFV.2SGM-him
      It's not possible that you know him

(52) ʔinta momken ma-tkun-ʃ   teʕraf-u
      You   possible NEG-be.PV.2SGM-NEG know.IPFV.2SGM-him
      It's possible that you don't know him

However, negation on *elmafrūd* causes a less subtle change in meaning than that on *lāzim* and

momken, as shown in the difference between examples 53 and 54:

(53) meʃ el-mafrūd   neʕmel   zayy-u
      NEG the-ought.to do.IPFV.1PL   like-his
      We're not supposed to do what he does/We don't have to do what he does

Sentence 53 is ambiguous, this is caused by negation of the modal *el-mafrūd* which could either

mean obligation (in which case its negation would lead to prohibition) and it can also mean that

something is supposed to happen as in example 36, in which case its negation would mean that

something is not supposed to happen. While sentence 54 below can only have one meaning,

which is prohibiting the action from taking place. This resembles the meaning associated with

negation of *lāzim*, however *el-mafrūd* carries a lesser degree of obligation, and therefore it is

translated as 'supposed to'.

(54) el-mafrūd   ma-neʕmel-ʃ   zayy-u
     the-ought.to NEG-do.IPFV.1PL-NEG like-his
     We're supposed to not do what he does

This does not seem to contradict with what Azer (1980) mentions about negation of *el-mafrūd*,

stating that "these two negative alternatives do not result in drastic meaning changes" (Azer,

1980, p.113). Where he assumes the meaning is not different while giving different translations

for each of the sentences, as shown in the examples cited from Azer (1980)'s study:

(55) meʃ ʔel-mafrūd tesūʔ   men-gheir roxsa
     NEG the-ought   drive.IPFV.2SGM without   license
     You're not supposed to drive without a license   (Azer, 1980)

(56) ʔel-mafrūd ma-tsuʔ-ʃ   men-gheir roxsa
     the-ought   NEG-drive.IPFV.2SGM-NEG without   license
     You're supposed to not drive without a license   (Azer, 1980)

It is also possible to have negation on both the modal and the following lexical verb, as in the

corpus example:

(57) ma-yenfaʕ-ʃ   ma-ʔol-l-ū-ʃ   ʕala maʃaʕr-i
     NEG-can.IPFV.3SGM-NEG NEG-tell.IPFV.1SG-to-him-NEG on   feelings-my
     I can't not tell him about my feelings= I have to tell him about my feelings

Although the difference in meaning between sentences where negation occurs on the modal as

opposed to the lexical verb is not always straightforward, however these examples show that each

of the modal and the lexical verb behave independently regarding negation, this is particularly

clear in example 57 where both verbs are negated.

5.3.3.5.3 Past tense   The past tense auxiliary can be combined with the non-inflecting

modals to form past tense.

(58) kānu   momken yiẖallasu   el-matʃ   men badri
     be.PV.3PL possible finish.IPFV.3PL the-match from early
     They could've finished the match a while ago

(59) fi el-nezām   el-ʔadīm kān   yenfaʕ   tḥawwel   koleyyet-ak
     in the-system the-old   be.PV.3SGM can.IPFV.3SGM change.IPFV.2SGM college-your
     baʕd sana
     after year
     In the old system it was possible to change your college after one year

(60) kān   lāzim tewāgih   ʔaḫū-ha
     be.PV.3SGM must confront.IPFV.3SGF brother-her
     She should've confonted her brother

Note here that the auxiliary occured in the default 3SGM form, while in the same context in
the corpus, sentence 61 occured where the auxiliary agrees with the subject, with no difference
in meaning, as in:

(61) kānit   lāzim tewagh-uh   w   tʕarraf-u   ʔinn-aha
     be.PV.3SGF must confront.IPFV.3SGF-him and make.know.IPFV.3SGF-him that-she
     ʕerifit
     know.PV.3SGF
     She should've confonted him and let him know she (already) knew

This shows that the agreement on the auxiliary in that case is in free variation, where it can
either agree with the subject or occur in the default form with no difference in meaning. However
it would be ungrmmatical in any of these sentences to have the default form of the auxiliary
when it follows the modal, as shown in the ungrammaticality of:

(62) *lāzim kān   tewāgih   ʔaḫū-ha
     must   be.PV.3SGM confront.IPFV.3SGF brother-her

This contrast shows that the auxiliary obligatorily inflects for the subject of the clause when it
occurs after the modal. While on the contrary, when the auxiliary precedes the modal it can
occur either inflected or in the default form. However occurence of the finite form of the auxiliary
following these modals might be more restricted than that preceeding it. This assumption follows
from the fact that there were no corpus examples with the past tense auxiliary following the
modal, where the only cases in the corpus had the auxiliary in the bare Imperfective form and
it expresses a state, as in:

(63) ʔinta momken tekūn   teʕraf-u
     You   possible be.IPFV.2SGM know.IPFV.2SGM-him
     You might know him

It is also possible to have the auxiliary *kān* both preceding and following the modal, where the auxiliary preceding the modal is in the perfective form and expresses past tense, while the one following the modal is in the bare imperfective form and expresses the state, as in:

(64) ʔabʔa   b-akdeb   ʕalē-ki law ʔolt-el-ek   ʔin-ni mafīʃ   fatra
     become.IPFV.1SG BI-lie.IPFV.1SG on-you if   tell.PV.1SG-to-you that-I there's.no time
     men ħayāt-i kont   momken ʔakūn   maʃdūd   l-bent moʕayyana
     from life-my be.PV.1SG possible be.IPFV.1SG attracted to-girl specific
     I'd be lying if I told you that there isn't a time in my life in which I might've been attracted
     to a specific girl

This agrees with my judgement as a native speaker that the default position for the auxiliary would be preceding the modal, although it is still grammatical for it to be postponed till after the modal. It also agrees with Azer (1980)'s account of modals, stating that: "When the auxiliary 'kaan' is employed as a tense carrier it occurs in pre- or post-modal positions, though in my idiolect the pre-modal position is always preferred" (Azer, 1980, p.17)

5.3.3.5.4 Complementiser   : Unlike other modals in this group, *el-mafrūd* can be followed by a complementiser, as in:

(65) el-bēt   kān   malyān nās   kitīr el-mafrūd   ʔin homma ʔasħāb el-bent
     the-house be.PV.3SGM full   people many the-ought.to that they   friends the-girl
     di
     this
     The house was full of people who (supposedly) that they are this girl's friends

Here *el-mafrūd* does not denote obligation, but instead denotes a similar meaning to 36 above, refering to a hypothetical situation, or an idea that is believed to hold, while in reality it might not be true. Another corpus example where the modal *el-mafrūd* takes a phrase headed by a complementiser as its complement is:

(66) meʃ el-mafrūd   ʔin ʔintimāʔ-i   el-dīni   yifriʔ   maʕ-āk
     NEG the-ought.to that pertinence-my the-religious differ.IPFV.3SGM with-you
     My religious beliefs shouldn't make any difference to you

Here the modal *el-mafrūd* is followed by the complementiser *ʔin* heading a clause formed of an
NP followed by a VP. This structure is frequently used with *el-mafrūd* in ECA and it refers to
obligation, while occurence of the complementiser after any of the other modals in this group is
either ungrammatical or highly questioned, and it did not produce any examples in my corpus
search for any of the other modals in this group. This structure, where the modal takes a
complementiser phrase as its complement, is represented in the lexical entries in section 5.4.4.

5.3.3.5.5 Arguments Examples show that the subject of the non-inflecting modals *ʔel-*
mafrūd, *lāzim*, *momken* and *yinfaʕ* should always be the same as that of the following lexical
verb. This is shown using the constructed example:

(67) ʔaħmad lāzim yīgi   ennaharda
     Ahmed may come.IPFV.3SGM today
     Ahmed must come today   [constructed]

Here Ahmad is the subject shared between the modal and its verbal complement, while it is
ungrammatical to have different subjects, as in:

(68) *ʔaħmad lāzim Mona tīgi   ennaharda
     Ahmed   must Mona come.IPFV.3SGF today

It is also possible to have any of the non-inflecting modals in this group with an expletive in
subject position, as in:

(69) lāzim ma-ykon-ʃ   fi-h darar
     must NEG-be.IPFV.3SGM-NEG in-it harm
     There must be no harm in it

In terms of sharing the arguments between the modal and its verbal complement, each of the
non-inflecting modals in this group only shares the subject with its verbal complement, while
arguments of the lexical verb are not arguments of the modal, as shown by the example:

(70) *lāzim-l-aha ʔabʕat   gawāb
     must-to-her send.IPFV.1SG letter

[constructed]

As opposed to the case where the recipient pronoun is attached to the lexical verb, yielding a grammatical structure, as in:

(71) lāzim ʔabʕat-l-aha   gawāb
     must send.IPFV.1SG-to-her letter
     I must send a letter to her   [constructed]

These examples show that non-inflecting modals share only the subject with their verbal complements, this issue is tackled in more details regarding the status of the structure as bi-clausal in section 5.4.1.

5.3.3.5.6   Summary   Table 5.2 summarises the behaviour of non-inflecting modals as described in the preceding sections.

Modal	Inflects	Followed by finite	Insertions	Negation	Different SUBJ
lāzim	No	No	Yes	Yes	No
2el-mafrūd	No	No	Yes	Yes(subtle)	No
momken	No	No	Yes	Yes	No
yinfaʕ	No	No	Yes	Yes	No

Table 5.2: Behaviour of Non-inflecting modals

Section 5.3.4 below investigates the final set of modals in the current study, which are modals that show agreement with their subjects, and hence they are refered to as 'inflecting modals'.

5.3.4   Inflecting modals

This set includes modals which inflect to agree with the subject in either gender, number and person if they are verbs, or only in gender and number if they are participles. They also share the property of being followed only by non-finite form verbs in all cases, where they can have the

perfective auxiliary only preceding them to express past tense. A property of the verbal modals
that differ from other classes of verbs is that they can not occur in the imperative form. The
inflecting modals are: *yiʔdar* 'able' which is a verb, *ʕāwiz* 'want', *mosammem* 'insist' and *nāwi*
'intend' which are participles, as well as the noun *nefs* 'wish'. Data for each of these modals is
represented separately in sections 5.3.4.1 to 5.3.4.5, while behaviour of the whole set of inflecting
modals in terms of negation, insertions and arguments is presented in section 5.3.4.6.

5.3.4.1   *yiʔdar* 'able'

yiʔdar is morphologically a verb that agrees with the subject in person, number and gender,
therefore it is glossed as a verb in the example sentences.

(72) ʔahl-ik   yiʔdaru   yisrifu   ʕalē-ki walla meʃ ħa-yeʔdaru?
      family-your able.IPFV.3PL pay.IPFV.3PL on-you or   NEG FUT-can.IPFV.3PL
      Can your family pay your expenses or they won't be able to?

(73) tiʔdar   tibʕat   rasāyil   maggāneyya men el-mawqiʕ
      able.IPFV.2SGM send.IPFV.2SGM messages free   from the-site
      You can send free messages (SMS) from the website

In examples 72 and 73, the modal *yiʔdar* is the main verb of the sentence and it is in the bare
imperfective form, which might raise a question about the status of the bare imperfective verb
form as a non-finite verb. However, example 72 is a question while example 73 denotes the
modal use, and these are some of the non-indicative contexts in which the bare imperfective
verb form can occur as the main verb in the sentence. More details on specifying the bare form
as a non-finite verb form is presented in section 1.4.

The modal *yiʔdar* can also occur in finite forms, such as:

(74) b-tiʔdar   titħammil   el-masʔūleyya
      BI-able.IPFV.3SGF endure.IPFV.3SGF the-responsibility
      She's responsible

(75) w   lʔenn-aha   b-titekel   ʕala rabena b-tiʔdar   teʕaddi
     and because-she BI-count.IPFV.3SGF   on   God   BI-able.IPFV.3SGF cross.IPFV.3SGF
     w   tkammil
     and continue.IPFV.3SGF

     And since she counts on God, she 's able to go on and continue her path

Although it did not occur in the corpus, it is also possible for *yiʔdar* to occur in the future tense, as in the constructed example:

(76) meʃ ʕārif   ħ-aʔdar   ʔarūħ   el-gamʕa   bokra   walla
     NEG know.PTCP.SGM   FUT-able.IPFV.1SG   go.IPFV.1SG   the-university tomorrow   or
     laʔ
     not

     I'm not sure if I'll be able to go to university tomorrow or not   [constructed]

However, in all of these cases, the modal shall always be followed by a non-finite lexical verb, having the bare imperfective verb form, as in the ungrammaticality of:

(77) *meʃ ʕārif   ħ-aʔdar   ħ-arūħ   el-gamʕa   bokra
      NEG know.PTCP.SGM   FUT-able.IPFV.1SG   FUT-go.IPFV.1SG   the-university tomorrow
     walla laʔ
     or   not

                                                                             [constructed]

5.3.4.2   *ʕāwiz* **'want'**

The modal *ʕāwiz* is morphologically an active participle, and therefore it agrees with the subject in number and gender.

(78) ʕāwiz   ʔaʔoll-ak   ʕala ħāga
     want.PTCP.SGM   tell.IPFV.1SGM-you   on   thing

     I want to tell you something

It is also possible to have another pronunciation for the same modal, pronounced as *ʕāyiz*, replacing only the semi vowel. Both variants have the same meaning and use[9]:

---

[9]I have used both variants to search for the modal within the corpus, as mentioned in section 1.3.2 among the orthographic issues I have faced while collecting the data.

(79) ʔana ʕāyiz   ʔasʔal-ak   feʕlan ʕan-ha
     I   want.PTCP.SGM ask.IPFV.1SG-you really about-her

     I really want to ask you about her

(80) ʕayza   ʔaʃūf   bāba
     want.PTCP.SGF see.IPFV.1SG dad

     I want to see dad

5.3.4.3   *nāwi* 'intend'

The modal *nāwi* is also an active participle, and therefore it inflects for number and gender.

(81) nāwi   yiʕmel   ʔeh   baʕd.ma yetḥarrag?
     intend.PTCP.SGM do.IPFV.3SGM what after   graduate.IPFV.3SGM

     What does he intend to do after he graduates?

(82) ʔana nawya   ʔaḵtefi   kām   yōm
     I   intend.PTCP.SGF disappear.IPFV.1SG number day

     I intend to disappear for a few days

The lexical verb following *nāwi* can only occur in the bare imperfective form, as is shown by the ungrammaticality of:

(83) *ʔana nawya   ḥ-aḵtefi   kām   yōm
     I   intend.PTCP.SGF FUT-disappear.IPFV.1SG number day

There is another verbal form for this modal, which has the same meaning and is derived from the same root, but is morphologically a verb, which is *yinwi*. The following examples show this modal verb in the perfective form *nawēt*:

(84) ʔana ḵalās nawēt   ʔatkallem
     I   finally intend.PV.1SG talk.IPFV.1SG

     I finally intended to speak

(85) ḵalās ʔana nawēt   ʔaʕīʃ   fi el-mādi
     finally I   intend.PV.1SG live.IPFV.1SG in the-past

     I finally intended to live in the past

This verbal form occured only twice in the corpus, which is less frequent than the participle form *nāwi*. It can occur in the finite form and therefore can function as the main verb in indicative sentences, but it can only be followed by bare imperfective form lexical verbs, as in the ungrammaticality of:

(86) *ʔana ḫalāṣ nawēt ḥ-atkallem
     I   finally intend.PV.1SG FUT-talk.IPFV.1SG

It behaves similarly to the participle form *nāwi* and occurs in the same position, and therefore it does not have a separate entry in the current study, but will rather be considered a variant form of the same modal *nāwi*.

5.3.4.4 *mosammem* 'insist'

The modal *mosammem* is morphologically an active participle which agrees with the subject in gender and number, and it can also be followed by verbal complements in the bare imperfective form only.

(87) ʔinta mosammem teʕmi ʕēn-ak w ʕaʔl-ak
     you.SGM insist.PTCP.SGM blind.IPFV.2SGM eye-your and mind-your
     You insist on blinding your eyes and mind= You insist on ignoring the facts

(88) ʔinti mosammema teftaḥi mawdūʕ meʃ ḥa-tnāʔiʃ fī-h
     you.SGF insist.PTCP.SGF open.IPFV.2SGF topic NEG discuss.FUT.1SG in-it
     You insist on opening a topic that I won't discuss

5.3.4.5 *nefs-i* 'wish'

The modal *nefs-i* is treated in Azer (1980) as an inflecting modal particle, which is derived from the noun *nafs* 'self'. The justification for treating this modal as a particle is not relevant, as it is based on its distribution properties and its ability to have discontinuous negation, which is shared with other inflecting modals (Azer, 1980). On the other hand, *nefs-i* resembles the structure of a construct state which expresses possession in ECA, where an NP is formed of a possessed noun (the noun *nafs*) followed by a possessor (the attached pronoun *-i* 'my'), where

it literally means 'wish-owner' (in the case of *nefs-i* it means 'my wish'). Therefore it is glossed in the current study as a noun to which a pronoun is suffixed, as there was not enough reason for treating it as a particle.

(89) nefs-i   ʔaʃūf   fi masr kōra zay ʔoroppa
      wish-my see.IPFV.1SG in Egypt ball like Europe
      I wish I'd see in Egypt football as that in Europe

Although all the corpus examples for this modal had the first person suffix, it can inflect in different ways according to person, number and gender of the possessor, as shown in the constructed example:

(90) mona nefs-aha tsāfer   barra
      Mona wish-her travel.IPFV.3SGF outside
      Mona wishes to travel abroad   [constructed]

Having presented the list of inflecting modals included in the current study and some example sentences for each, I now turn to investigating the behaviour of this set of modals, in section 5.3.4.6.

5.3.4.6   Behaviour of inflecting modals

Inflecting modals share common aspects of their behaviour in terms of negation, sharing their arguments with the verbal complements, having adverbs preceding their verbal complements, whether or not they can take complement clauses headed by the complementiser *ʔin*, as well as combining with the auxiliary to form past tense. Each of these topics is discussed in the following subsections 5.3.4.6.1 to 5.3.4.6.5. Investigating these behaviours help reach a better understanding of inflecting modals in ECA, and hence develop the suitable analysis for these structures.

5.3.4.6.1   Insertions   It is possible to have adverbials between the inflecting modal and the following lexical verb:

(91) ya   mohannad sāʕid   māma ʕaʃān tiʔdar   baʕd keda
(Calling.particle) Mohannad help.IMP.2SGM mom   because able.IPFV.2SGM after this
tsāʕid   merāt-ak
help.IPFV.2SGM wife-your

Mohannad, help mom so you're able to help your wife afterwards

It is also possible for the subject to occur between the modal and the lexical verb:

(92) yiʔdaru   ʔahl-ik   yisrifu   ʕalee-ki walla meʃ ħa-yeʔdaru?
able.IPFV.3PL family-your spend.IPFV.3PL on-you   or   NEG FUT-able.IPFV.3PL

Can your family pay your expenses or they won't be able to?

5.3.4.6.2   Negation   Negation of the modal ʕāwiz can occur either on the modal or the following lexical verb with no change in meaning:

(93) meʃ ʕāwiz   ʔamūt fi farʃet-i
NEG want.PTCP.SGM die.1SG in bed-my

I don't want to die in my bed

(94) ʕāwiz   ma-mut-ʃ   fi farʃet-i
want.PTCP.SGM NEG-die.1SG-NEG in bed-my

I don't want to die in my bed

While with all of the other inflecting modals, negating the modal gives a different meaning from negating the lexical verb:

(95) ʔinta ma-tiʔdar-ʃ   tigbir   ħad   yiʕmil   ħāga
You   NEG-able.IPFV.2SGM-NEG force.IPFV.2SGM someone do.IPFV.3SGM something
howwa meʃ ʕayez-ha
he   NEG want.PTCP.SGM-it

You can't force someone to do something he doesn't want to do   [constructed]

(96) ʔinta tiʔdar   ma-tʔol-ʃ   l-ħad   ʔasrār-ak   law meʃ
You   able.IPFV.2SGM NEG-tell.IPFV.2SGM-NEG to-someone secrets-your if   NEG
ʕāyiz
want.PTCP.SGM

You can avoid telling anyone your secrets if you don't want to   [constructed]

However the difference in meaning can be subtle, as in:

(97) howwa meʃ nāwi   yoxtob-ha
     he   NEG intend.PTCP.SGM get.engaged.IPFV.3SGM-her
     He doesn't intend to get engaged to her (but he may)

Sentence 97 negates the intention to get engaged, however it leaves room for the action to still

take place. Unlike sentence 98 below that expresses the intention to not-do something, which

indicates a will for not-getting engaged.

(98) howwa nāwi   ma-yoxtob-hā-ʃ
     he   intend.PTCP.SGM NEG-get.engaged.IPFV.3SGM-her-NEG
     He intends to not get engaged to her

Although the difference in meaning here is subtle, it is still evident in the suitable contexts.

This difference in meaning is also evident with negation of the modals *mosammem* and *nefsi*.

(99) ʔaħmad mosammem   ma-yroħ-ʃ   el-gamʕa
     Ahmed insist.PTCP.SGM NEG-go.IPFV.3SGM-NEG the-university
     Ahmed insists on not going to university   [constructed]

(100) ʔaħmad meʃ mosammem   yroħ   el-gamʕa
      Ahmed NEG insist.PTCP.SGM go.IPFV.3SGM the-university
      Ahmed doesn't insist on going to university (but he may)   [constructed]

5.3.4.6.3   Past tense   The Perfective form auxiliary can occur preceding the inflecting modals

to express past tense, as in:

(101) howwa kān   ʕāyiz   yingaħ   w   feʕlan nagaħ
      he   be.PV.3SGM want.PTCP.SGM succeed.IPFV.3SGM and indeed suceed.PV.3SGM
      He wanted to succeed and indeed he did

(102) kān   mosammem   ʔageb-ha
      be.PV.3SGM insist.PTCP.SGM bring.IPFV.1SG-it
      He insisted that I should bring it

(103) homma kānu   yiʔdaru   yeʕzilū-h   zay mħammad nagīb
      they be.PV.3PL able.IPFV.3PL isolate.IPFV.3PL-him like Mohammed Naguib
      They could've isolated him as they did to Mohammed Naguib (a former Egyptian presi-
      dent)

The agreement on the auxiliary is obligatory even though it precedes the modal, this is shown

by the contrast between sentences 104 and 105:

(104) ʔana kont   nāwi   ʔabīʕ   l-uh   laptop
      I   be.PV.1SG intend.PTCP.SGM sell.IPFV.1SG to-him laptop
      I intended to sell him a laptop

(105) *ʔana kān   nāwi   ʔabīʕ   l-uh   laptop
      I   be.PV.3SGM intend.PTCP.SGM sell.IPFV.1SG to-him laptop

However, it is not possible to have the past tense auxiliary following any of the inflecting modals,

as in the ungrammaticality of:

(106) *homma yiʔdaru   kānu   yeʕzilū-h   zay mħammad nagīb
      they able.IPFV.3PL be.PV.3PL isolate.IPFV.3PL-him like Mohammed Naguib

5.3.4.6.4   Complementiser   The only inflecting modal that had corpus sentences where it

was followed by a complementiser is the modal *mosammem*:

(107) kān   mosammem   ʔin-ena nedfaʕ   ʕeʃrīn bas
      be.PV.3SGM insist.PTCP.SGM that-we pay.IPFV.1PL twenty only
      He insisted that we pay only twenty

It is also possible to have an NP following the complementiser:

(108) ʔinti ʔilli mosammema   ʔin kelmet rogūʕ   morsi fi el-roʔya
      you   who insist.PTCP.SGF that word returning Morsi in the-vision
      You're the one who insists that the word 'Morsi's return' is in the dream

Despite the fact that *mosammem* is the only case of an inflecting modal followed by a comple-

mentiser in the corpus, it is also possible for the other inflecting modals to have a complement

clause headed by a complementiser, except for ʕāwiz, where it is less acceptable[10]. Constructed examples for acceptable cases of inflecting modals followed by complementiser are shown below:

(109) kān   nāwi   ʔin-ena nedfaʕ   ʕeʃrīn bas
       be.PV.3SGM intend.PTCP.SGM that-we pay.IPFV.1PL twenty only

       He intended that we pay only twenty   [constructed]

(110) kān   nefs-u   ʔin-ena nedfaʕ   ʕeʃrīn bas
       be.PV.3SGM wish-his that-we pay.IPFV.1PL twenty only

       He wished that we pay only twenty   [constructed]

(111) kān   yiʔdar   ʔinn-u yedfaʕ   ʕeʃrīn bas
       be.PV.3SGM able.IPFV.3SGM that-he pay.IPFV.3SGM twenty only

       He was able to pay only twenty   [constructed]

(112) #kān   ʕāwiz   ʔin-ena nedfaʕ   ʕeʃrīn bas
       be.PV.3SGM want.PTCP.SGM that-we pay.IPFV.1PL twenty only

       He wanted that we pay only twenty   [constructed]

5.3.4.6.5   Arguments   Examples show that each of the inflecting modals can share the same subject with their verbal complements, as in the constructed example:

(113) ʔaħmad ʕāwiz   yīgi   ennaharda
       Ahmed   want.PTCP.SGM come.IPFV.3SGM today

       Ahmed wants to come today   [constructed]

Where Ahmed is the subject shared between the modal and its verbal complement. However, the inflecting modals ʕāwiz, nāwi, mosammem and nefsi can also have a different subject from that of the following lexical verb.

(114) ʔaħmad ʕāwiz   Mona tīgi   ennharda
       Ahmed   want.PTCP.SGM Mona come.IPFV.3SGF today

       Ahmed wants Mona to come today   [constructed]

---

[10]Based on my judegement as a native speaker.

(115) Ɂaħmad nefs-u   Mona tīgi   ennharda
      Ahmed   wish-his Mona come.IPFV.3SGF today
      Ahmed wishes that Mona would come today   [constructed]

It is also possible to have an independent subject for the lexical verb after the modals *mosammem*

and *nāwi*, but it is much more acceptable if the embedded clause is headed by a complementiser,

as in:

(116) Ɂaħmad mosammem   #(Ɂin) Mona tīgi   ennharda
      Ahmed   insist.PTCP.SGM (that) Mona come.IPFV.3SGF today
      Ahmed insists that Mona shall come today   [constructed]

(117) Ɂaħmad nāwi   #(Ɂin) Mona tīgi   ennharda
      Ahmed   intend.PTCP.SGM (that) Mona come.IPFV.3SGF today
      Ahmed intends that Mona shall come today   [constructed]

While it is ungrammatical for the modal *yiɁdar* to have an independent subject for the lexical

verb either with or without a complementiser:

(118) *Ɂaħmad yiɁdar   Mona tīgi   ennharda
      Ahmed   able.IPFV.3SGM Mona come.IPFV.3SGF today

(119) *Ɂaħmad yiɁdar   Ɂin   Mona tīgi   ennharda
      Ahmed   able.IPFV.3SGM that Mona come.IPFV.3SGF today
This shows that the modal *yiɁdar* can have a different analysis from that of the other inflecting

modals, which is represented in section 5.4.3.3.

5.3.4.6.6   Summary   Table 5.3 summarises the behaviour of the inflecting modals as de-

scribed in the preceding sections.

Modal	Inflects	Followed by finite	Insertions	Negation	Different SUBJ
yiɁdar	Yes	No	Yes	Yes	No
ʕāwiz	Yes	No	Yes	No	Yes
nāwi	Yes	No	Yes	Yes(subtle)	Yes
mosammem	Yes	No	Yes	Yes	Yes
nefsi	Yes	No	Yes	Yes	Yes

Table 5.3: Behaviour of Inflecting modals

As shown in table 5.3, all of the inflecting modals can have independent subjects from the subject of their verbal complement, except for *yiʔdar*. This is tackeld in section 5.4.3.3 where *yiʔdar* is treated in a different way from the rest of inflecting modals in terms of its analysis. Another difference is that *ʕāwiz* is the only inflecting modal that does not exhibit a change in meaning depending on the position of negation (either on the modal or on the lexcial verb). The importance of this property and how it affects the analysis of the structure as mono or bi-clausal is presented in section 5.4.1. While the following section, 5.4, is concerned with answering questions about the analysis of structures for all of the sets of modals investigated in the current study, as well as representing this analysis in LFG terms.

5.4   Analysis

This section is devoted to reaching the suitable analysis for the modal constructions investigated in the current study based on the patterns of behaviour of the data. This is shown in table 5.4 which summarises the different patterns of behaviour common among each of the modal categories described in the preceding sections.

Modal	Followed by finite	Insertions	Negation	Different SUBJ
yimken	Yes	Yes	Yes	No
Non-inflecting	No	Yes	Yes	No
Inflecting	No	Yes	Yes (except ʕāwiz)	Yes (except yiʔdar)

Table 5.4: Behaviour summary

Table 5.4 illustrates that there was no difference among the sets regarding insertions and negation, where all the modals investigated can have adverbs or subjects occuring between the modal and the following lexical verb, and they all[11] undergo meaning changes depending on the position of the negation, whether on the modal or the following lexical verb. In fact negation is assumed to be a weak point in differentiating between modals, where Hacquard (2011) argues

---

[11]Except the inflecting modal *ʕāwiz*

that several factors interfere with the way negation is interpreted on modals, such that negation is not a strong enough basis to differentiate between types of modals. Therefore ʕāwiz is not excluded from the analysis based on the exception it shows regarding negation.

On the other hand, the modals investigated differ in terms of their ability to be followed by finite verbs as well as their ability to have different subjects for each of the modal and the lexical verb. *yimken* is the only modal that can be followed by finite verbs while other non-inflecting modals and inflecting modals can only have verbal complements in the bare imperfective form. On the other hand, inflecting modals are the only group of modals which can take a subject different from that of its verbal complement. These differences in behaviour will be used to answer questions about the structures in which each of these group of modals occur, with the aim of reaching the analysis for these structures.

The main questions to ask about the modals are whether the modal and the following lexical verb form a mono-clausal or bi-clausal structure? and what is the kind of control relation between them? The questions asked about the patterns of behaviour of the structures resembles those adopted for other verbal complementation patterns in the current study, such as phasal verbs in chapter 4. The analysis is divided into subsections according to the topic investigated, where the question of whether the modal structures are mono or bi-clausal is investigated in section 5.4.1, while section 5.4.2 investigates the status of the subject as argument of the modal and its verbal complement. The kind of control relation between the modal and its verbal complement is investigated in section 5.4.3, this control relation is then represented in terms of the lexical entries in section 5.4.4, the f-structures in section 5.4.5, as well as the c-structures in section 5.4.6. Section 5.5 includes the conclusions of the investigations of modals in ECA in the current study.

5.4.1 Mono or Bi-clausal?

A well-known mono-clausal structure found in some languages is serial verbs, which are charac-
terised by the two verbs sharing common properties such as mood, aspect, tense and negation,
where negation occurs on only one of the verbs in the serial verb construction (Aikhenvald and
Dixon, 2006). Mono-clausal constructions can also be complex predicates, which have the gram-
matical functional structure of a single predicate and a single subject (Butt, 1995). Therefore
establishing whether the structure is mono or bi-clausal relates to the features it shares with
serial verbs or complex predicates constructions.

Mono-clausal structures are characterised by the property of sharing the same tense, aspect,
mood and polarity on both verbs of the serial verb construction (SVC), where the serial verb
construction can only have one negation marker on either of its verbs (Aikhenvald and Dixon,
2006). Since most of the modals investigated in the current study are not verbs and therefore
carry no tense/aspectual marking, the characteristic of sharing the same tense and aspect be-
tween the modal and its verbal complement is not applicable in the current study. Regarding
negation, on the other hand, behaviour of the modals shows that the position of negation either
on the modal or the main verb affects the meaning, which shows that negation does not scope
over both the modal and the verb. The corpus example 57 illustrated again as 120 below also
shows that it is possible to have negation on both the modal and the following lexical verb,
which is a property that serial verbs do not have:

(120) ma-yenfaʕ-ʃ   ma-ʔol-l-ū-ʃ   ʕala maʃaʕr-i
       NEG-can.IPFV.3SGM-NEG   NEG-tell.IPFV.1SG-to-him-NEG   on   feelings-my

       I can't not tell him about my feelings= I have to tell him about my feelings

This is possible with any of the modals investigated, including the modal ʕāwiz[12] where negation
can occur both on the modal and its verbal complement, as in the constructed example:

---

[12]Which is listed in table 5.4 as an exception to other modals in terms of change of meaning depending on
the position of negation.

(121) ʔana meʃ ʕāwiz ma-kallem-ak-ʃ, ʔana mesh ʕārif
    I NEG want.PTCP.SGM NEG-talk.IPFV.1SG-you-NEG, I NEG know.PTCP.SGM
    ʔakallem-ak
    talk.IPFV.1SG-you

    It's not that I don't want to talk to you, it's that I can't talk to you   [constructed]

Properties of a complex predicate where both verbs behave as a single predicate in terms of agreement and anaphora can be misleading in the case of modals investigated in the current study, as the modals that obligatorily share the same subject with the lexical verb are non-inflecting, where they lack any form of agreement with the subject. And in fact often the subject does not seem to hold any thematic relation to the modal. Regarding arguments, non-inflecting modals do not share arguments with the lexical verb:

(122) yinfaʕ ʔabʕat-l-ohom gawāb?
    can.IPFV.3SGM send.IPFV.1SG-to-them letter?

    Can I send them a letter?   [constructed]

(123) *yinfaʕ-l-ohom ʔabʕat gawāb?
    can.IPFV.3SGM-to-them send.IPFV.1SG letter?

    [constructed]

These examples are evidence against modal sentences having a mono-clausal structure, and therefore they are assumed to have a bi-clausal structure where the subject may be shared between the modal and the following lexical verb. The next question being, what is the kind of relation between the two clauses? and what are the arguments of each? These questions will be tackled in the coming sections.

5.4.2 Arguments

A question is raised regarding the arguments of these modals, which is whether the subject of the lexical verb represents the object of the modal or not. Examples show that the subject of the lexical verb can not be incorporated as a pronoun attached to the modal, as in:

(124) *yimken-ha tīgi   Ɂennaharda
      may-her   come.IPFV.3SGF today

This is also ungrammatical for inflecting modals which can have an independent subject different from that of the lexical verb:

(125) *Ɂaḥmad mosammem-ha   tsāfir
      Ahmed   insist.PTCP.SGM-her travel.IPFV.3SGF

This is different from the examples of verbal complementation for the causative _ḫalla_ investigated in chapter 3, in which the SUBJ of the lexical verb is also an OBJ argument of the causative and it can be incorporated as a pronoun attached to it. An example of this causative structure is:

(126) Ɂaḥmad ḫallā-ha   tsāfir
      Ahmed make.PV.3SGM-her travel.IPFV.3SGF

      Ahmed made her travel   [constructed]

The only grammatical equivalent for this structure is with the modal _ʕāwiz_, where it is possible to incorporate the subject of the lexical verb with _ʕāwiz_, as in the constructed example:

(127) Ɂaḥmad ʕāwiz-ha   tsāfir
      Ahmed want.PTCP.SGM-her travel.IPFV.3SGF

      Ahmed wants her to travel   [constructed]

However, other tests proving the status of the downstairs SUBJ as OBJ of the modal can not be applied to _ʕāwiz_, such as passivisation (as _ʕāwiz_ is a participle which can not be passivised) and incorporation of this pronoun into the negation circumfix (as _ʕāwiz_ can only be negated using the prefix meʃ). Therefore the assumption is that the independent subject of the lexical verb is not an object of the modal.

Another important question is whether the SUBJ is a thematic argument of the modal or not. The NP representing the subject of modals usually occurs in the canonical position for SUBJ, as shown previously in the corpus example 39, repeated below:

(128) ʔeħna momken nittifiʔ fi noʔta wa7da
    we   possible agree.IPFV.1PL in point one
    We may agree in one point

The NP 'we' occurs in the canonical position for subjects, preceding the modal *momken*, and it does not seem fronted, unlike example 129 below where the SUBJ seems misplaced:

(129) #momken ʔiħna nittifiʔ fi noʔta wa7da
    possible   we agree.IPFV.1PL in point one
    We may agree in one point

It is also possible for the auxiliary *kān* to agree with the subjects when it precedes any of the modals, however the agreement is optional for non-inflecting modals and obligatory for inflecting modals. The corpus examples 60 and 61 which are repeated below illustrate that agreement between the auxiliary and the SUBJ is indeed optional:

(130) kān   lāzim tewāgih   ʔaḫū-ha
    be.PV.3SGM must confront.IPFV.3SGF brother-her
    She should've confonted her brother

(131) kānit   lāzim tewagh-uh   w tʕarraf-u   ʔinn-aha
    be.PV.3SGF must confront.IPFV.3SGF-him and make.know.IPFV.3SGF-him that-she
    ʕerifit
    know.PV.3SGF
    She should've confonted him and let him know she (already) knew

This leaves a question about whether non-inflecting modals are subjectless predicates, however it also leads to the conclusion that the subject of non-inflecting modals is non-thematic, especially while assuming they are raising predicates as discussed in section 5.4.3.1.

While inflecting modals show a different behaviour, where agreement is **obligatory** between the auxiliary and the SUBJ, this is shown by the corpus examples 104 and 105, repreated below:

(132) ʔana kont   nāwi   ʔabīʕ   l-uh   laptop
    I   be.PV.1SG intend.PTCP.SGM sell.IPFV.1SG to-him laptop
    I intended to sell him a laptop

(133) *ʔana kā̄n   nā̄wi   ʔabī̄ʕ   l-uh   laptop
       I   be.PV.3SGM intend.PTCP.SGM sell.IPFV.1SG to-him laptop

This shows that inflecting modals shall have a different analysis from that of non-inflecting modals, where the SUBJ of inflecting modals is thematic with the possiblity of having an anaphoric control relation, as discussed in section 5.4.3.2. The kind of control relation for each group of modals is investigated in details in the following section, 5.4.3.

5.4.3   Kind of control relation

This section aims to answer the question about the kind of control relation between the modal and its verbal complement, whether it is anaphoric or functional control. Properties of each in (Kroeger (2004); Falk (2001); Dalrymple (2001)) in addition to the analysis of modals in (Hacquard (2011); Bhatt et al. (2011)) were applied to each of the modals, and the results show that they do not all share the same analysis. Where non-inflecting modals are assumed to represent functional control of the type raising, while inflecting modals, on the other hand, represent anaphoric control. Having different analyses for different modals is not surprising, as this is the case with Urdu/Hindi modals where some are analyzed as raising while others are analyzed as functional control constructions (Bhatt et al., 2011).

The different analyses for each group of modals and the motivation for each is represented in the following sections. Where section 5.4.3.1 includes the raising analysis for non-inflecting modals, section 5.4.3.2 investigates how inflecting modals are analysed as anaphoric control, while section 5.4.3.3 investigates the motivation for analysing *yiʔdar* as an Equi verb, representing the only case of obligatory anaphoric control for the ECA modals investigated in the current study.

5.4.3.1   Functional Control: Raising

All of the non-inflecting modals (*lā̄zim*, *ʔel-mafrū̄d*, *momken*, *yinfaʕ* in addition to *yimken*) are analysed as cases of functional control of the type raising. This is based on the results of two

tests which are used by Hacquard (2011) to prove that modals are raising predicates, illustrated as follows:

- Their ability to have expletive/dummy subjects, as in the constructed examples[13]:

  (134) lāzim   ykūn   fi-h mayya hena
        must be.IPFV.3SGM in-it water   here
        There must be water here   [constructed]

  (135) yimken   ykūn   fi-h mayya hena
        may.IPFV.3SGM be.IPFV.3SGM in-it water   here
        Maybe there's water here   [constructed]

  (136) ʔel-mafrūd   ykūn   fi-h mayya hena
        the-ought.to be.IPFV.3SGM in-it water   here
        There should be water here   [constructed]

- Their ability to be integrated into idiom chunks without losing the idiomatic meaning, as in:

  (137) ʔel-fās   momken toʔaʕ   f-el-rās
        the-hoe possible fall.IPFV.3SGF in-the-head
        Things might go past the point of no return. LIT:The hoe might fall in the head

        [constructed]

On the other hand, Inflecting modals are not able to have expletive/dummy subjects, as shown by the ungrammaticality of:

(138) *yiʔdar   ykūn   fi-h mayya hena
      able.IPFV.3SGM be.IPFV.3SGM in-it water   here

However, a superficialy similar sentence, but grammatical sentence would be:

(139) nefs-u   ykūn   fi-h mayya hena
      wish-his be.IPFV.3SGM in-it water   here
      He wishes there's water here   [constructed]

---

[13]The same sentences are used throughout each of the tests while changing the modal only so as to eliminate any other contextual factors that might alter the meaning or affect the analysis.

Where although this sentence is grammatical with an inflecting modal, it actually has an implicit
subject refered to by the SGM possessive suffix on the modal 'nefs-u' (his), so it actually means
'he wishes that there is water here' which shows that the sentence still can not have an expletive
subject. Therefore, inflecting modals can not have dummy subjects, and will need a different
kind of analysis, represented in the following section, 5.4.3.2.

5.4.3.2   Anaphoric control

The inflecting modals *ʕāwiz*, *nāwi*, *mosammem* and *nefs-i* (all of the inflecting modals **except**
yiʔdar 'able') are analysed as cases of anaphoric control, this is based on factors such as:

- Presence of an independent subject in the subordinate clause, as in:

(140) ʔaħmad ʕāwiz   Mona tsāfir
       Ahmed   want.PTCP.SGM Mona travel.IPFV.3SGF
       Ahmed wants Mona to travel   [constructed]

Mona here is the subject of the lexical verb, while it is not an argument of the modal.

(141) nefs-i   trūħi   el-madrasa
       wish-my go.IPFV.2SGF the-school
       I wish you go to school   [constructed]

Where the subject of the modal (the speaker) is different from that of the verbal comple-
ment (you).

- The possibility of having the complementiser *ʔin* heading the complement clause:

(142) ʔaħmad mosammem   ʔin Mona tsāfir
       Ahmed   insist.PTCP.SGM that Mona travel.IPFV.3SGF
       Ahmed insists that Mona travels   [constructed]

It is also possible to have a pronoun attached to the complementiser representing the
SUBJ of the modal, as in:

(143) ʔaħmad nāwi   ʔinn-uh ysāfir
      Ahmed intend.PTCP.SGM that-he travel.IPFV.3SGM
      Ahmed intends to travel   [constructed]

The analysis of inflecting modals as anaphoric control predicates whether or not their complement clauses are headed by the complementiser *ʔin* follows Dalrymple (2001)'s analysis of the English verb 'try' as an anaphoric control predicate in sentences like 'David tried to leave', where the VP maps into a COMP (Dalrymple, 2001, p.328). A similar analysis for inflecting modals in the current study is represented as the c-structure of example 150, included in section 5.4.6.

Therefore, inflecting modals (except *yiʔdar*) are analysed as anaphoric control predicates, while the analysis of *yiʔdar* is represented separately in section 5.4.3.3.

5.4.3.3   Equi

A different behaviour is attested with the inflecting modal *yiʔdar* 'able', where it can not take a dummy subject and can not be used in idioms, therefore, excluding the raising analysis, as shown by the following examples:

(144) *yiʔdar   ykūn   fī-h mayya hena
     able.IPFV.3SGM be.IPFV.3SGM in-it water   here

                                               [constructed]

But yet *yiʔdar* can not have an independent subject in the subordinate clause as other inflecting modals (therefore excluding optional anaphoric control), this is shown by the ungrammaticality of:

(145) *ʔaħmad yiʔdar   Mona tsāfir
     Ahmed   able.IPFV.3SGM Mona travel.IPFV.3SGF

                                               [constructed]

It can only have a pronoun attached to the complementiser which **obligatorily** represents the subject of *yiʔdar*:

(146) ʔaħmad yiʔdar   ʔinn-uh ysāfir
        Ahmed able.IPFV.3SGM that-he travel.IPFV.3SGM

        Ahmed is able to travel   [constructed]

Therefore *yiʔdar* behaves as an Equi verb which follows Dalrymple (2001, p.327)'s statement

that: "the controller in an equi construction is semantically as well as syntactically selected by

the verb." Example 146 above shows that *yiʔdar* is followed by the complementiser *ʔin* heading

a COMP to which the SUBJ pronoun is attached, this should also correspond to the SUBJ of

the lexical verb. This is shown by the ungrammaticality of having a verb with an expletive or

a semantically empty subject in the subordinate clause, as in example 144 above. These facts

provide evidence that *yiʔdar* is an Equi verb which represents obligatory anaphoric control, as

is the case with the English verb 'try' (Dalrymple, 2001).

Having established the analysis of each group of modals and differentiating each according to

the behaviour attested in the data, the following sections aim to represent this analysis in terms

of the lexical entries in section 5.4.4, followed by the f-structures in section 5.4.5 and finally the

c-structures in section 5.4.6.

5.4.4   Lexical Entries

Results of the analysis showed that different groups of modals fall into different kinds of con-

trol relations, where non-inflecting modals are raising predicates, while inflecting modals are

anaphoric control predicates, this information shall be represented in the lexical entry for each

of the modals investigated in the current study. This section represents some examples for

lexical entries of the modals, while the whole list of lexical entries is implemented in the XLE

grammar fragment for ECA, described in chapter 6.

In terms of the syntactic category, modals are treated here as members of I, the functional

category of auxiliaires and modal verbs in English, also defined as: "functional categories are

specialized subclasses of lexical categories which have a syncategorematic role in the grammar (such as marking subordination, clause type, or finiteness)" (Bresnan, 2001, p.101). This was the basis for categorising modals in ECA as members of the I category, despite the fact that most of the modals are morphologically not verbs and they do not carry tense as other members of I, instead they are marked as **MODAL**. Note also that Arabic has nominal sentences where I is not filled, and therefore one might wonder if sentences headed by non-verbal modals can be treated as a verbless sentence, however the current analysis is concerned mainly with the modal property added to the sentences through both verbal and non-verbal modals, and so they are treated as representing the inflectional category I. This also provides a common ground for treating the main predicates throughout the current work as members of I. The lexical verb heading the embedded clause occupies V and marks aspect, as is standard in the current study.

The lexical entry for the modal *yimken* which is a raising predicate that can take both finite and non-finite verbal complements is:

yimken: I

($\uparrow$PRED)= 'may< ($\uparrow$ XCOMP)> ($\uparrow$ SUBJ) '

($\uparrow$ SUBJ) = ($\uparrow$ XCOMP SUBJ)

($\uparrow$ VTYPE)=MODAL

The lexical entry for a non-inflecting modal represents also that of a raising predicate with a non-thematic subject, however they are marked for taking only verbal complements in the bare imperfective form, as shown below for the modal *momken*:

momken: I

($\uparrow$PRED)= 'possible< ($\uparrow$ XCOMP)> ($\uparrow$ SUBJ) '

($\uparrow$ SUBJ) = ($\uparrow$ XCOMP SUBJ)

(↑ VTYPE)=MODAL.

(↑ XCOMP VFORM)=c BARE

The non-inflecting modal *el-mafrūd* is the only non-inflecting modal that takes a complementiser, therefore it shall have a lexical entry specifying these structures, which is:

el-mafruud: I

(↑PRED)= 'the.ought.to< (↑ SUBJ) (↑ COMP)> '

(↑ VTYPE)=MODAL.

Note that it can also behave as a non-inflecting modal, and hence has another lexical entry:

el-mafruud: I

(↑PRED)= 'the.ought.to< (↑ XCOMP)> (↑ SUBJ) '

(↑ SUBJ) = (↑ XCOMP SUBJ)

(↑ VTYPE)=MODAL.

(↑ XCOMP VFORM)=c BARE

On the other hand, inflecting modals are anaphoric control predicates, this applies for both arbitrary control predicates as well as the obligatory control predicate *yiʔdar*, where they have similar syntactic representations in terms of lexical entries and phrase structure rules (Dalrymple, 2001). They only differ in the optionality of taking a PRO as subject of the embedded clause, where it is optional for inflecting modals and obligatory for *yiʔdar*. The lexical entry for the modals *mosammem* and *yiʔdar* are represented below:

mosammem: I

(↑PRED)= 'insist<(↑ SUBJ) (↑ COMP)> '

((↑ COMP SUBJ PRED) = 'PRO')

(↑ VTYPE)=MODAL.

(↑ COMP VFORM)=c BARE.

The brackets defining PRED as PRO are added to show that this represents optional anaphoric control, where the COMP may (or may not) have an independent SUBJ. This is different from how Dalrymple (2001) treats the arbitrary anaphoric control predicate 'gesture' in English, having the same lexical entry as that of obligatory anaphoric control predicates, stated as "Syntactically, obligatory and arbitrary control constructions do not differ; the same phrase structure rule (...) is used in deriving the two, and the syntactic portions of the lexical entries are similar" (Dalrymple, 2001, p.339).

yiʔdar, on the other hand, is an obligatory anaphoric control predicate, where no independent SUBJ can be expressed in its embedded clause, which is shown by the non-optionality of the equation (↑ COMP SUBJ PRED) = 'PRO', represented in its lexical entry:

yiʔdar: I

(↑PRED)= 'able<(↑ SUBJ) (↑ COMP)> '

(↑ COMP SUBJ PRED) = 'PRO'.

(↑ VTYPE)=MODAL.

(↑ COMP VFORM)=c BARE.

The verbal complement of any of the modals occupies V, however for *yimken* it can be either finite or non-finite:

katab: V

(↑PRED)= 'write<(↑ SUBJ)>'

(↑ASPECT) = PERF

While for all of the other modals the verbal complement has to be in the bare imperfective form:

ʔaʃūf: V

(↑PRED)= 'see<(↑ SUBJ)>'

(↑ASPECT) = IMPF
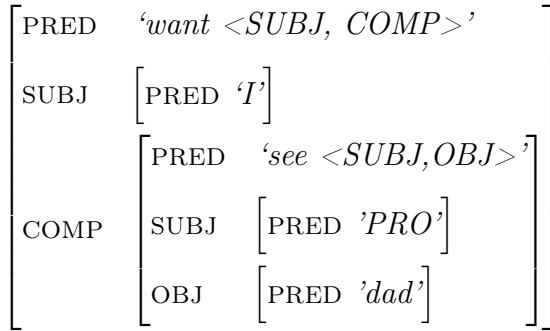
(↑VFORM) = BARE

5.4.5   F-structures

The f-structure for non-inflecting modals, which are raising predicates would be:

(147) momken yinsā
       possible forget.IPFV.3SGM

    He might forget

$$
\begin{bmatrix}
\text{PRED} & \textit{'possible <XCOMP> SUBJ'} \\
\text{SUBJ [1]} & \begin{bmatrix} \text{PRED } \textit{'PRO'} \end{bmatrix} \\
\text{XCOMP} & \begin{bmatrix} \text{PRED} & \textit{'forget <SUBJ>'} \\ \text{SUBJ [1]} & \\ \text{VFORM} & \textit{BARE} \end{bmatrix}
\end{bmatrix}
$$

While inflecting modals which are anaphoric control predicates would have the following f-structure:

(148) ʔana ʕayza   ʔaʃūf   bāba
       I   want.PTCP.SGF see.IPFV.1SG dad

    I want to see dad

$$
\begin{bmatrix}
\text{PRED} & \text{`want <SUBJ, COMP>'} \\[4pt]
\text{SUBJ} & \begin{bmatrix} \text{PRED `I'} \end{bmatrix} \\[4pt]
\text{COMP} & \begin{bmatrix}
\text{PRED} & \text{`see <SUBJ,OBJ>'} \\[4pt]
\text{SUBJ} & \begin{bmatrix} \text{PRED 'PRO'} \end{bmatrix} \\[4pt]
\text{OBJ} & \begin{bmatrix} \text{PRED 'dad'} \end{bmatrix}
\end{bmatrix}
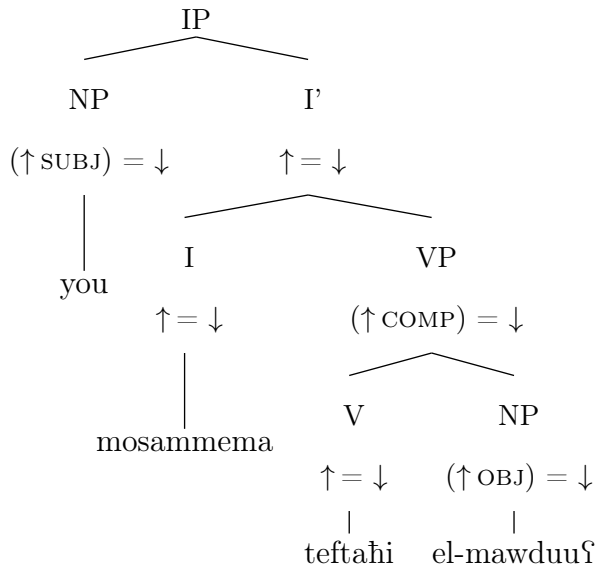\end{bmatrix}
$$

5.4.6   C-structure

The c-structure for non-inflecting modals which take an XCOMP as their complement is:

(149) momken yinsā̄
       possible forget.IPFV.3SGM

   He might forget
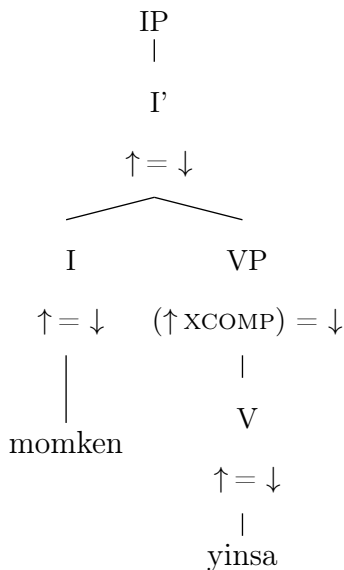
While inflecting modals take a COMP:

(150) ʔenti mosammema   teftaħi   el-mawdūʕ
       you   insist.PTCP.SGF open.IPFV.2SGF the-issue

   You insist on opening the subject

```
                         IP
              ┌───────────┴───────────┐
            NP                        I'
       (↑ SUBJ) = ↓                  ↑ = ↓
            │            ┌────────────┴────────────┐
            │            I                         VP
          you          ↑ = ↓                  (↑ COMP) = ↓
                         │              ┌───────────┴───────────┐
                         │              V                      NP
                    mosammema         ↑ = ↓               (↑ OBJ) = ↓
                                        │                      │
                                     teftaħi             el-mawduuʕ
```

5.5   Conclusions

This chapter was concerned with description and analysis of sentences in which modals are
main predicates in ECA. The data presented here are extracted from a corpus of ECA online
texts, and therefore it represents modals as they are used nowadays. This might be the reason
the structures described here are in some ways different from those in Azer (1980) which was
published more than 35 years ago and therefore represents a different stage in the development
of the dialect. The main difference is the use of *darūri* as an obligation modal in Azer (1980)
which functions similarly to the obligation modal *lāzim*, while in the current study it occured
only having a pragmatic function and therefore was not studied as a modal. The difference
is shown in example 151 below which is acceptable according to Azer (1980), while the same
example is considered unacceptable in the modern use and rather percieved as some 'old' version
of the language[14].

(151) talab-ak   darūri   yiwsal-na   ʔabl   ʔāxir el-ʃahr
       request-your ought.to reach.IPFV.3SGM-us before end   the-month

---

[14]This is how these structures are percieved in my openion and other native speakers' opinion whom I have
consulted regarding this sepcific example. Although I could not find a resource describing what might be an
'old' version of ECA versus a more recent version and the differences between them, this remains one of the
interesting questions about ECA that could be tackled in further research.

Your application should reach us before the end of the month   (Azer, 1980, p.81)

Modals were divided into three sets according to the structures in which they occur, whether inflecting, non-inflecting, and the kind of verbal complements they can have, whether finite non-finite, or both. *yimken* is an non-inflecting modal that can take verbal complements in both finite and non-finite forms, while the other non-inflecting modals could only be followed by non-finite form verbs (the bare imperfective). Inflecting modals could also be followed only by non-finite form verbs, however the modals themselves show agreement with the subject, hence the name 'inflecting modals.' In terms of the syntactic analysis, *yimken* and non-inflecting modals represented functional control (raising) predicates, while inflecting modals are arbitrary anaphoric control predicates, with *yiʔdar* as an obligatory anaphoric control predicate.

In terms of the syntactic category, modals are treated here as members of the I category as they represent a functional category despite the fact that they do not carry tense as is standard for members of **I** in the current study, which poses a problem of having non-tensed sentences. This can be remedied through adding a rule to specify that each member of **I** shall have tense, where it would assign present tense to **I** in the absence of any other tense specification, however I did not deal with this issue in the current study. The lexical entries, f-structures and c-structures for each of these modals are implemented in the XLE grammar fragment for ECA, and presented in more details in chapter 6.

Implementation

6.1 About XLE

XLE stands for Xerox Linguistics Environment, which is a "platform for developing Lexical Functional Grammars and was developed at Palo Alto Research Center (PARC) (..) XLE includes a parser, a generator and a finite state morphological analyser and it can be used both for parsing and generation of natural languages" (Chatsiou, 2010, p.176). Members of the ParGram project used XLE to provide a common set of linguistic principles and grammatical features with the aim of developing broad coverage grammars that can be applied to different languages and a wide range of language phenomena (Chatsiou, 2010).

An LFG grammar of a certain language is represented in the XLE system as a set of phrase structure rules, lexical entries, f-structure annotations and templates which are then used to generate/parse the relevant language structures (Chatsiou, 2010; Kaplan et al., 2002). XLE is used in the current study to build a grammar that is capable of parsing all (and only) the possible grammatical structures in the data as well as representing the c-structures and f-structures for each sentence.

This chapter starts by a description of the verb structures covered throughout the current study as well as the different control relations that are attested with each verb, in section 6.2. This is followed by section 6.3 which describes the structure of the XLE file of the grammar fragment developed for ECA, including the lexicon, grammar rules as well as the templates used. Section 6.4 represents a section of the testfile for the grammar fragment, which contains the list of ECA sentences that the grammar is able to parse. Section 6.5 then discusses the main issues that I have faced during the XLE implementation, and how each of these issues was resolved as well as the ones that remain unresolved. Finally, section 6.6 provides a summary for the process of XLE implementation of the verbal structures in the current study.

6.2   Verb structures

The XLE is used to implement the results of analysis of four verbal complementation patterns in ECA which are included in the current study, these are: auxiliary *kān*, causative *ḫalla*, phasal verbs and modals. Examples for each verbal complementation pattern and the chapter in which it is represented and analysed is listed as follows:

- Auxiliary *kān* 'be' (**Chapter 2**)

  (1) ʔel-barnāmeg kān   bi-yitzāʕ   ʕala el-MBC
      the-program   be.PV.3SGM BI-show.IPFV.PASS.3SGM on   the-MBC
      The program was showing on MBC (a TV channel)

- Causative *ḫalla* 'make' (**Chapter 3**)

  (2) el-eḥterāf   bi-yḫallī-k   teɣanni   kowayyes
      the-professionalism BI-make.IPFV.3SGM-you sing.IPFV.2SGM fine
      Professionalism makes you sing better

- The phasal verbs: *rigiʕ* 'return', *baʔa* 'become', *fidil* 'remain', *ʔaʕad* 'go.on', *badaʔ* 'begin', *liḥiʔ* 'catch', *ʔarrab* 'be.near' and *battal* 'stop' (**Chapter 4**)

(3) battalt   ʔalʕab   baskit
    stop.PV.1SG play.IPFV.1SG basketball
    I stopped playing basketball

- The modals: *yimken* 'may', *yinfaʕ* 'can', *lāzim* 'must', *ʔel-mafrūd* 'the-ought-to', *momken* 'possible', *yiʔdar* 'able', *ʕāwiz* 'want', *mosammem* 'insist', *nāwi* 'intend', as well as *nefs-i* 'wish' (**Chapter 5**)

(4) momken yinsā-ni   baʕd fatra
    possible forget.IPFV.3SGM-me after while
    He might forget me after a while

Each of these verbs is represented by a set of sentences extracted from a corpus of ECA that I have built for these purposes[1], the verbs are then analyzed in terms of their f-structures and c-structures, illustrating the kind of control relation between the verb and its verbal complements. The lexical entry for each verb included information about its PRED value, its arguments, whether or not it has a functional/anaphoric control relation with its verbal complements, its agreement features, any subcategorisation constraints it might have on the verb form of its complements, its temporal/aspectual features (if any), and whether it occupies I or V. Summary for the different verb forms and examples for each and representation of the relevant features are presented in the following subsections, starting by section 6.2.1 which presents a summary of the different tense/aspectual marking on verb forms, and cases where the verb may occupy either I or V nodes in the tree. Followed by section 6.2.2 which presents examples for the difference between functional and anaphoric control relations and how each is represented in the XLE implementation.

6.2.1   Types of verbs

The possible markings of morphosyntactic tense/grammatical aspect on each of the verb forms in ECA are represented as follows:

---

[1]Described in more details in section 1.3.

Form	Tense Marking	Aspect Marking	Example
Perfective	past	perfect	*katab* 'wrote'
bi-prefixed	present	imperfect: progressive/habitual	*bi-yikteb* 'writing/writes'
ħa-prefixed	future	prospective	*ʕa-yikteb* 'will write'
bare form	-	-	*yikteb* 'to write'

Table 6.1: Marking of tense/grammatical aspect

Each verb is assigned as either I or V according to the features it carries in the sentence. Verbs occupying I mark either morphosyntactic tense or are modal, while lexical verbs occupy V and they may or may not mark grammatical aspect. However, some verbs can occupy either I or V according to the structure in which they occur. Each of these cases is illustrated below:

- Any finite form verb that is the main predicate of the sentence and carries tense occupies I, this applies to all finite form verbs (perfective, bi-prefixed, ħa-prefixed) and excludes non-finite verb forms (the bare Imperfective form).

  (5) fehemt   el-serr
      understand.PV.1SG the-secret
      I understood the secret!

- In cases where the main verb (as with the auxiliary, causative, phasal verb or a modal) takes a verbal complement (lexical verb), the main predicate occupies I while the lexical verb occupies V.

  (6) baʔa   bi-yemʃi   zay el-batrīq
      become.PV.3SGM bi-walk.IPFV.3SGM like the-penguin
      Lit: He became (such that) he walks like a penguin

      He walks like a penguin

Here the phasal verb *baʔa* 'become' occupies I and marks past tense, while the lexical verb *biyemʃi* 'walk' occupies V carrying Imperfect aspect and both progressive and habitual features. Therefore this sentence could denote either **past progressive** or **past**

habitual[2].

- Finite verbs can occupy either I or V according to the structure in which they occur, where the finite verb form can represent either tense on a verb in I (as in example 5) or aspect on the verb in V (as in example 6). Example 7 represents the bi-prefixed verb *bi-tɣanni* occupying V where it denotes grammatical aspect, while example 8 represents the same verb form in I as it is the main predicate of the sentence and it denotes tense.

  (7) sondos kānit   bi-tɣanni
      sondos be.PV.3SGF BI-sing.IPFV.3SGF
      Sondos was singing

  (8) sondos bi-tɣanni
      sondos BI-sing.IPFV.3SGF
      Sondos sings/ is singing

  Therefore the lexical entry for this verb should reflect both cases, where it can occur either in I or V:

```
betghanni  V  * (^PRED)='SING<(^SUBJ)>'           @3SGF    (^ASPECT)=IMPF    (^HAB)=+  (^PROG)=+ ;
           I  * (^PRED)='SING<(^SUBJ)>'           @3SGF    (^TENSE)=PRESENT  (^HAB)=+  (^PROG)=+.
```

- Some verbs impose selectional restrictions on the forms of its verbal complements, such restrictions on the forms of the verb's complements should also be represented in its lexical entry. An example is the case where an Imperative form verb shall only be followed by Imperative form complements, this restriction is represented as:

```
(^XCOMP VFORM)=c IMP
```

---

[2]The function of the bi-prefixed verb form in marking grammatical aspect is described in more details in section 2.7.3

which specifies that this verb can only be followed by an imperative verb form, as shown in the lexical entry for the phasal verb *ʔilħaʔ*:

```
2il7a2     I  * @ControlI (^PRED)='CATCH<(^XCOMP)>(^SUBJ)'    @2SGM (^XCOMP VFORM)=c IMP  (^VFORM)=IMP.
```

There are also cases where the main predicate is followed only by non-finite verbal complements, marked as:

```
(^XCOMP VFORM)=c BARE
```

which specifies that the verbal complement should be in the bare imperfective form. An example is the lexical entry for the modal *yinfaʕ*:

```
yinfa3     I  * @ControlI (^PRED)='CAN<(^XCOMP)>(^SUBJ)'       (^XCOMP VFORM)=c BARE (^VTYPE)=MODAL.
```

The XLE grammar fragment should also reflect the kind of control relation between the main predicate and its verbal complement, as it is reflected in the lexical entry of the verb as well as the f-structure for the sentence. This is discussed in more details in the following section, 6.2.2.

6.2.2   Functional/Anaphoric control

Within the domain of the current study, functional control is represented in the predicates: Auxiliary *kān*, Causative *ḫalla*, Phasal verbs, and non-inflecting modals. This relation is represented in the grammar fragment by the template:

```
ControlI = (^SUBJ)=(^XCOMP SUBJ)
```

This specifies that the subject is shared between the main predicate and its XCOMP, an example for a raising predicate is the phasal verb *ʔarrab* 'be.near' in:

(9) ʔana ʔarrabt   ʔamūt   min el-xōf
    I   be.near.PV.1SG die.IPFV.1SG from the-fear
    I'm almost scared to death!

The lexical entry for *ʔarrab* should specify that it subcategorises for an XCOMP, as shown in:

```
2arrabt   I  * @ControlI (^PRED)='BE_NEAR<(^XCOMP)>(^SUBJ)'  @1SG  (^XCOMP VFORM)=c BARE (^TENSE)=PAST.
```

While Anaphoric control is attested in inflecting modals[3], and it is represented by the template:

```
ControlICP = (^COMP SUBJ PRED)= 'PRO'
```

This specifies that the subject of the COMP is a pronoun, this pronoun can optionally refer to the subject of the main predicate. An example for anaphoric control predicate is the modal *nefs-i* 'wish-my' in:

(10) nefs-i   ʔaʃūf-ak   fi masr
     wish-my see.IPFV.1SG-you in Egypt
     I wish I'd see you in Egypt

The lexical entry for *nefs-i* shows that it subcategorises for a COMP:

```
nefsi     I  * @ControlICP (^PRED)='WISH<(^SUBJ)(^COMP)>'     @1SG  (^COMP VFORM)=c BARE (^VTYPE)=MODAL.
```

---

[3]Discussed in section 5.4.3.2.

The list of templates included in the XLE implementation file for each of functional and anaphoric control predicates are shown in section 6.3.2 below.

6.3   Structure of the XLE implementation file

Implementation of the grammar fragment of these structures follows the standard structure for grammar files that are loaded into the XLE parser. The grammar file has the extension .lfg, as mentioned in Chatsiou (2010, p.178) "all XLE grammar files, to be interpreted as such by the parser, need to have an .lfg extension in their filename. An .lfg grammar file may contain c-structure rules with their corresponding f-structure (or other projection) annotations as well as lexical entries (again with their c-structure or other information)." Different sections of the grammar .lfg file which is used to implement the grammar fragment in the current study is represented and described in the following sections, while the whole test file is found in appendix A.

6.3.1   Configuration

This is the first section in the .lfg file, which is concerned with defining various parts of the file that shall be loaded with the grammar. This includes defining the language, the governable and non-governable relations and defining the set of templates. The configuration section of the .lfg file used in the current study is shown below:

```
VERBS      ARABIC     CONFIG (1.0)
  ROOTCAT              S.
  LEXENTRIES           (VERBS ARABIC).
  RULES                (VERBS ARABIC).
  GOVERNABLERELATIONS  SUBJ XCOMP OBJ COMP.
```

```
SEMANTICFUNCTIONS     ADJUNCT.

TEMPLATES             (VERBS ARABIC).
```

6.3.2   Templates

This section includes the different templates used in the lexicon as shortcuts to indicate specific features that are shared among a set of lexical entries. The set of templates used and what they stand for are listed as follows:

- Kind of control relation:

```
ControlI = (^SUBJ)=(^XCOMP SUBJ).

ControlIC = (^OBJ)=(^XCOMP SUBJ).

ControlICP = (^COMP SUBJ PRED)= 'PRO'.
```

Where **ControlI** stands for subject to subject raising, while **ControlIC** stands for subject to object raising attested with the causative _ḫalla_. While **ControlICP** stands for anaphoric control which is attested with inflecting modals. Each of these templates is refered to in the lexical entry of the verb by adding the symbol '@', as shown in the lexical entry of the causative verb _ḫalla_ representing the subject to object raising template **ControlIC**:

```
xalla      I  * @ControlIC (^PRED)='MAKE<(^SUBJ)(^XCOMP)>(^OBJ)' @NoOBJ  @3SGM (^TENSE)=PAST.
```

- Some verbs can have two kinds of complements, either an XCOMP or a COMP. This applies to the modal _ʔel-mafrūd_ and the causative _ḫalla_ which can either be followed by

the XCOMP or by a complement clause headed by the complementiser *ʔinn*. These verbs
have both kinds of complements specified in their lexical entries, as shown in the lexical
entry of the modal *ʔel-mafrūd*:

```
2elmafruud  I  * @ControlI (^PRED)='OUGHT_TO<(^XCOMP)>(^SUBJ)'  (^XCOMP VFORM)=c BARE (^VTYPE)=MODAL ;
            I  * (^PRED)='OUGHT_TO<(^SUBJ)(^COMP)>'                (^VTYPE)=MODAL.
```

- Specifying the presence/absence of a pronominal argument on the verb.

```
OBJ        = (^OBJ PRED)='PRO'.
```

This specifies that the verb has the object as an attached pronoun, as shown in the lexical
entry of the verb *ḫallā-ni* 'make.PV.3SGM-me':

```
xallaani    I  * @ControlIC (^PRED)='MAKE<(^SUBJ)(^XCOMP)>(^OBJ)' @OBJ   @3SGM @1SGOBJ  (^TENSE)=PAST.
```

Specifying this feature stops ungrammatical sentences which have both pro and nominal
Object, such as

(11) *ḫalli-ha   sondos t7ess
     make.IMP.2SG-her Sondos feel.IPFV.3SGF

It is also ungrammatical for the causative verb *ḫalla* to lack any object, hence I needed a
template to specify cases where there is no object pronoun attached to the causative verb,
the template is:

```
NoOBJ   = ~ (^OBJ PRED)='PRO'.
```

This template is represented in the lexical entry of the causative verb *ḫalla*:

```
I  * @ControlIC (^PRED)='MAKE<(^SUBJ)(^XCOMP)>(^OBJ)' @NoOBJ  @3SGM (^TENSE)=PAST
```

Specifying this template stops sentences from having neither a pro or a nominal object for the causative verb, such as:

(12) *fawwaz ḫalla   ynām
      Fawwaz make.PV.3SGM sleep.IPFV.3SGM

- Specifying form of the verbal complement, as in the case where a main verb can not have a verbal complement in the imperative form, specified by the template:

```
NoIMP   = ~ (^XCOMP VFORM)= IMP.
```

This stops ungrammatical sentences such as:

(13) *rigiʕt   ʔiʃtaghal   tāni
      return.PV.2SGM work.IMP.2SGM again

A similar template is used to stop the auxiliary *ḥankūn* 'be.FUT.1PL' from taking a verbal complement in the bare imperfective form, the template is:

```
NoBare  = ~ (^XCOMP VFORM)= BARE.
```

- Specifying agreement features of the verb's arguments, this is represented by a number of templates specifying the person, number, and gender agreement features on the verb.

```
3SGM = (^SUBJ PERS)=3 (^SUBJ NUM)=SG (^SUBJ GEN)=MASC.

3SGF = (^SUBJ PERS)=3 (^SUBJ NUM)=SG (^SUBJ GEN)=FEM.

2SGM = (^SUBJ PERS)=2 (^SUBJ NUM)=SG (^SUBJ GEN)=MASC.

2SGF = (^SUBJ PERS)=2 (^SUBJ NUM)=SG (^SUBJ GEN)=FEM.

1SG =  (^SUBJ PERS)=1 (^SUBJ NUM)=SG.

3PL =  (^SUBJ PERS)=3 (^SUBJ NUM)=PL.

2PL =  (^SUBJ PERS)=2 (^SUBJ NUM)=PL.

1PL =  (^SUBJ PERS)=1 (^SUBJ NUM)=PL.

SGF = (^SUBJ NUM)=SG (^SUBJ GEN)=FEM.

SGM = (^SUBJ NUM)=SG (^SUBJ GEN)=MASC.

3SGFOBJ = (^OBJ PERS)=3 (^OBJ NUM)=SG (^OBJ GEND)=FEM.

3SGMOBJ = (^OBJ PERS)=3 (^OBJ NUM)=SG (^OBJ GEND)=MASC.

2SGFOBJ = (^OBJ PERS)=2 (^OBJ NUM)=SG (^OBJ GEND)=FEM.

2SGMOBJ = (^OBJ PERS)=2 (^OBJ NUM)=SG (^OBJ GEND)=MASC.

1SGOBJ  = (^OBJ PERS)=1 (^OBJ NUM)=SG.
```

These stop sentences having the wrong agreement from occuring, such as the ungrammat-
ical sentence:

(14) *ʔana ħa-tẖalli   el-rāgel   yedrab-ek
       I   make.FUT.2SGF the-man beat.IPFV.3SGM-you

6.3.3   C-structure rules

This is the section of the XLE grammar fragment concerned with specifying the phrase structure
rules which account for the c-structures of the sentences analyzed in the current study, and it
is represented as follows:

```
VERBS  ARABIC  RULES (1.0)
```

```
S  --> IP: ^=!.


IP --> {NP: (^SUBJ)=! | e: (^ SUBJ PRED)='PRO';}

        ({ADVP: (^ADJUNCT)=!})

        I': ^=!.


I' --> ({I: ^=!;})

        ({VP: (^COMP)=!})

        ({NP: (^OBJ)=! | e: (^ OBJ PRED)='PRO';})

        ({CP: (^COMP)=!})

        ({ADVP: (^ADJUNCT)=!})

  ({VP: (^XCOMP)=!})

        ({VP: ^=!})

        ({ADVP: (^ADJUNCT)=!}).



VP --> V: ^=!;

        ({NP: (^OBJ)=!})

        ({VP: (^XCOMP)=!})

        ({PP: (^ADJUNCT)=!})

        ({CP: (^ADJUNCT)=!}).



ADVP --> ADV: ^=!.


PP  ---> P: ^=!;
```

```
        {NP: (^OBJ)=!| e: (^ OBJ PRED)='PRO';}.




  NP ----> N: ^=!;

           ({PP: (^ADJUNCT)=!}).



  CP ----> C': ^=!.



  C' ----> C : ^=!;

           IP: (^ADJUNCT)=!.




----
```

6.3.4   Lexicon

This is the section where the lexical entries for all of the words that comprise the sentences analysed by the grammar fragment are listed. The lexical entries are grouped according to word types, starting by nouns, followed by complementisers, then verbs occupying both I and V, followed by pronouns and finally adverbs. The order of lexical entries does not affect the grammar in any way, however it can facilitate the search for a certain word within the lexicon. Examples for some of the lexical entries is listed as follows, while the whole .lfg file with the full lexicon is represented in appendix A:

```
2ana        NP * (^PRED)='I'         (^PERS)=1 (^NUM)=SG.

homma       NP * (^PRED)='they'      (^PERS)=3 (^NUM)=PL.

howwa       NP * (^PRED)='he'        (^PERS)=3 (^NUM)=SG (^GEN)=MASC.

2enti       NP * (^PRED)='you'       (^PERS)=2 (^NUM)=SG (^GEN)=FEM.

elwalad     NP * (^PRED)='boy'       (^PERS)=3 (^NUM)=SG (^GEN)=MASC (^DEF)=+.

fawwaaz     NP * (^PRED)='Fawaz'.

kintaki     NP * (^PRED)='kfc'.



2inn        C * (^PRED)='that'.



kaan        I  * @ControlI (^PRED)='BE<(^XCOMP)>(^SUBJ)' @3SGM   (^TENSE)=PAST.



xalleet     I  * @ControlIC (^PRED)='MAKE<(^SUBJ)(^XCOMP)>(^OBJ)' @NoOBJ  @2SGM (^TENSE)=PAST.

7atxalliini I  * @ControlIC (^PRED)='MAKE<(^SUBJ)(^XCOMP)>(^OBJ)' @OBJ

@2SGF @1SGOBJ  (^TENSE)=FUT.

rigi3       I  * @ControlI (^PRED)='RETURN<(^XCOMP)>(^SUBJ)' @ NoIMP @3SGM (^TENSE)=PAST.

bada2t      I  * @ControlI (^PRED)='BEGIN<(^XCOMP)>(^SUBJ)'  @1SG  (^XCOMP VFORM)=c BARE (^TENSE)=PAST;

            V  * @ControlI (^PRED)='BEGIN<(^XCOMP)>(^SUBJ)'  @1SG  (^XCOMP VFORM)=c BARE (^ASPECT)=PERF.

battalt     I  * @ControlI (^PRED)='STOP<(^XCOMP)>(^SUBJ)'   @1SG  (^XCOMP VFORM)=c BARE (^TENSE)=PAST.



yimken      I  * @ControlI (^PRED)='MAY<(^XCOMP)>(^SUBJ)'  (^XCOMP VFORM)=c BARE (^VTYPE)=MODAL.

naweet      I  * @ControlICP (^PRED)='INTEND<(^SUBJ)(^COMP)>' @1SG  (^COMP VFORM)=c BARE (^VTYPE)=MODAL (^TENSE)=PAST.

ti2dar      I  * @ControlICP (^PRED)='ABLE<(^SUBJ)(^COMP)>'   @2SGM (^COMP VFORM)=c BARE (^VTYPE)=MODAL.



bnel3ab     V  * (^PRED)='PLAY<(^SUBJ)>'   @1PL    (^ASPECT)=IMPF      (^HAB)=+ (^PROG)=+ ;

            I  * (^PRED)='PLAY<(^SUBJ)>'   @1PL    (^TENSE)=PRESENT    (^HAB)=+ (^PROG)=+.



btet3allim  V  * (^PRED)='LEARN<(^SUBJ)>'  @2SGM   (^ASPECT)=IMPF      (^HAB)=+  (^PROG)=+ ;

            I  * (^PRED)='LEARN<(^SUBJ)>'  @2SGM   (^TENSE)=PRESENT    (^HAB)=+  (^PROG)=+.



ye2fel      V  * (^PRED)='CLOSE<(^SUBJ)(^OBJ)>'   @3SGM  (^VFORM)=BARE.

2asta7mel   V  * (^PRED)='TOELRATE<(^SUBJ)>(^OBJ)' @1SG (^VFORM)=BARE.



fi          P  * (^PRED)='in<(^OBJ)>'.

wara        P  * (^PRED)='after<(^OBJ)>'.



taani       ADV * (^PRED)='again'.

shiwayya    ADV * (^PRED)='a_bit'.
```

These are the different sections of the .lfg file that is capable of parsing the whole set of sentences on which this study is based. One of the most important criteria for testing the efficiency of this grammar fragment is its ability to parse all of the grammatical sentences on which is it based, and none of the ungrammatical sentences. This list of sentences is included in a test file, discussed in details in the following section, 6.4.

6.4   Test files

Another kind of .lfg file that is related to the grammar fragment file is the test file which includes the whole set of sentences to which these rules apply. These are the set of the grammatical sentences that can be correctly parsed using the grammar fragment in which the rules are implemented. This is useful as it provides a clear picture about the range of sentences covered by the grammar fragment, as well as providing an easy way to test the extent to which the grammar is successful through performing a parse test for the whole file. The whole test file of grammatical sentences is presented in appendix B, while a section of this file is presented as follows:

```
# 42


laazim yit7ebes fi 2afas


# 43


momken yinsaani ba3d fatra


# 44
```

```
yinfa3 yit3emel fi sana
```

```
# 45
```

```
2enti 2elmafruud to23odi ma3 elnaas
```

```
# 46
```

```
2e7na momken nettefe2 fi no2ta
```

Sentences are written in transliteration symbols (represented in section 6.5.2), and numbered using the hash symbol so that if an error occurs in one of the sentences it can easily be spotted. The test file contains 81 grammatical sentences that can be parsed using this grammar fragment, and it is represented in appendix B.

Another important test file is that of ungrammatical sentences, which shall not be parsed by the grammar fragment. Parsing of this test file gives errors in all of its sentences, which ensures that the grammar fragment can parse ALL and ONLY grammatical sentences. The ungrammatical sentences test file is presented as follows:

```
# 1
```

```
2irga3 2ishtaghalt taani
```

```
# 2
```

fawwaaz xalla ynaam

3

xalliiha sondos t7ess

4

kaan 2inn mona tsaafer

6.5   Issues in implementation

This section is concerned with some of the issues that I faced during the implementation, including the problems that still need to be formalised in the grammar fragment, as well as the techniques I have used to facilitate the representation of the structures with which the current study is concerned. Section 6.5.1 discusses the techniques I adopted to simplify some of the complex sentences in the data in order to facilitate their analysis without affecting the structure, while the list of transliteration symbols used to represent the sentences in the XLE grammar fragment is presented in section 6.5.2. And finally, section 6.5.3 includes the problems that still need to be formalised in future versions of the grammar fragment.

6.5.1   Simplifying clauses

Since the focus of the current study is on representing complementation patterns of the main predicates such as the auxiliary, causative, phasal verbs and modals, some cases where these predicates occured in embedded clauses were not included in the XLE analysis. Such as in

example 11 repeated below:

(15) Ɂana magnūn l-el-ħad   Ɂilli   yḫallī-ni   Ɂalbes   w
      I   crazy   to-the-extent which make.IPFV.3SGM-me dress.IPFV.1SG and
      anzel   arūħ   el-ʃoɣl maʕ Ɂin-ni ʃoɣl-i   baḫallas-o
      descend.IPFV.1SG go.IPFV.1SG the-work with that-me work-my finish.BI.IPFV.1SG-it
      men ʕala el-Ɂinternet
      from on   the-internet
      I'm crazy to the extent that makes me get dressed and go to work although I finish my

      work on the internet

This example shows the causative verb in the bare imperfective form, where it refers to a generic

case rather than a specific causing event, and the causative verb does not carry any tense or

aspectual features. The causative structure here is embedded within a free-relative clause, which

would require an analysis for free-relative clauses in order to represent the correct c-structure for

the sentence. However the analysis of free relative clauses in Arabic is outside the scope of the

current study, and therefore these sentences were excluded from the implementation. However,

in other cases, ḫalla occured in the embedded clause following Ɂilli but it represents a main

verb, such as example 13 repeated below:

(16) feʕlan   ħa-tkūn   el-ħāga   el-waħīda Ɂilli   ħa-tḫallī-ni
      Actually FUT-be.IPFV.3SGF the-thing the-only   which FUT-make.IPFV.3SGF-me
      Ɂastaħmel   fekret el-dirāsa men Ɂawwil   w   gdīd
      tolerate.IPFV.1SG idea   the-study from beginning and new
      Actually this will be the only thing that will make me tolerate the idea of studying all

      over again

And therefore this sentence was included in the XLE implementation, while simplifying the

structure so the focus would be mainly on the causative structure:

(17) ħa-tḫallī-ni   Ɂastaħmel   el-mozakra
      FUT-make.IPFV.3SGF-me tolerate.IPFV.1SG the-studying
      (She) will make me tolerate studying

Another example for a simplified sentence is:

(18) tarya?a w   hzār   ḫallo   el-walad ynām   nōm   el-wāḥed ?illi
     Irony   and mocking make.PV.3PL the-boy sleep.IPFV.3SGM sleepiness the-one   that
     meʃ lā?-i   ḥad   yesmaʕ-o
     not find-him someone listen.IPFV.3SGM-him

     The irony and mocking made the boy sleep as a person who can't find anyone to listen to

     him (feeling lonely and sad)

Where the subject is formed of a coordinate phrase which was simplified to the pronoun *homma*

'they', the complement of the lexical verb was also removed, the simplified sentence is:

(19) homma ḫallo   el-walad ynām
     They   make.PV.3PL the-boy sleep.IPFV.3SGM

     They made the boy sleep

These changes have no effect on the causative structure itself, but rather aim to reduce the

complexity of the implementation rules.

6.5.2   Transliteration

The sentences included in the .lfg file were transliterated rather than transcribed as in the rest

of the current study, as it is not possible to use IPA symbols within the .lfg file as it produces

an error while running the XLE process. A list of the transliteration symbols that are used in

the .lfg file and their corresponding IPA symbols is presented in table 6.2 [4].

---

[4]All of the other symbols not mentioned in this table are used as they are.

IPA	Transliteration
ʔ	2
ʕ	3
ħ	7
ẖ	x
ʃ	sh
ɦ̄	gh
ī	ii
ā	aa
ū	uu

Table 6.2: Transliteration symbols

6.5.3 Problems

This section discusses the problems that still exist in the implementation, where the rule needs some kind of formalisation in the XLE that I could not execute within the current study. Solutions for these problems could be part of future versions of the grammar fragment.

- Control ICP (specifying anaphoric control) does not allow independent NP Subj in the embedded clause which is possible with arbitrary control predicates, where the template for anaphoric control specifies that the subject of the COMP is always a 'PRO':

```
ControlICP = (^COMP SUBJ PRED)= 'PRO'.
```

And therefore this allows only sentences like example 20 while it does not allow sentences such as example 21 which should be grammatical in ECA:

(20) nefs-i 2asāfer   2amrīka
     wish-I travel.IPFV.1SG America
     I wish I'd travel to America

(21) nefs-i Mona tsāfer   2amrīka
     wish-I Mona travel.IPFV.3SGF America
     I wish Mona would travel to America

This could be resolved by having a rule that specifies that the subject of the COMP can either be an NP or a PRO.

- Lexical verbs occupying $\mathbf{I}$[5] in the main predicates give two results, where in one they are chosen as V which projects from $\mathbf{I}$ directly, and therefore carry no tense. However, this can be remedied through adding a rule to specify that each $\mathbf{I}$ shall have tense, where it would assign present tense to $\mathbf{I}$ in the absence of any other tense specification. This is the case also with modals which occupy $\mathbf{I}$ but have no tense specification in their lexical entries, and with any other V that projects from $\mathbf{I}$ directly.

An example to illustrate the issue of verbs occupying either $\mathbf{I}$ or V is the verb *fehemt* which has the following lexical entry:

```
fehemt     V  * (^PRED)='UNDERSTAND<(^SUBJ)(^OBJ)>'   @1SG       (^ASPECT)=PERF ;

           I  * (^PRED)='UNDERSTAND<(^SUBJ)>(^OBJ)'   @1SG       (^TENSE)=PAST.
```

In sentences where *fehemt* is the main verb such as example 22:

(22) fehemt   el-serr
      understand.PV.1SG   the-secret
      I understood the secret

fehemt can either occupy $\mathbf{I}$ or V according to its lexical entry, since there is no other verb to occupy $\mathbf{I}$ (such as being preceded by a modal, auxiliary or a causative). Therefore parsing of sentence 22 produces two valid results, one where the verb occupies $\mathbf{I}$ and carries tense, presented in figure 6.1. While in the other parsing result the main verb occupies V and carries aspect only, resulting in a tenseless sentence (which is wrong in this case). Figure 6.2 illustrates this latter parsing result:

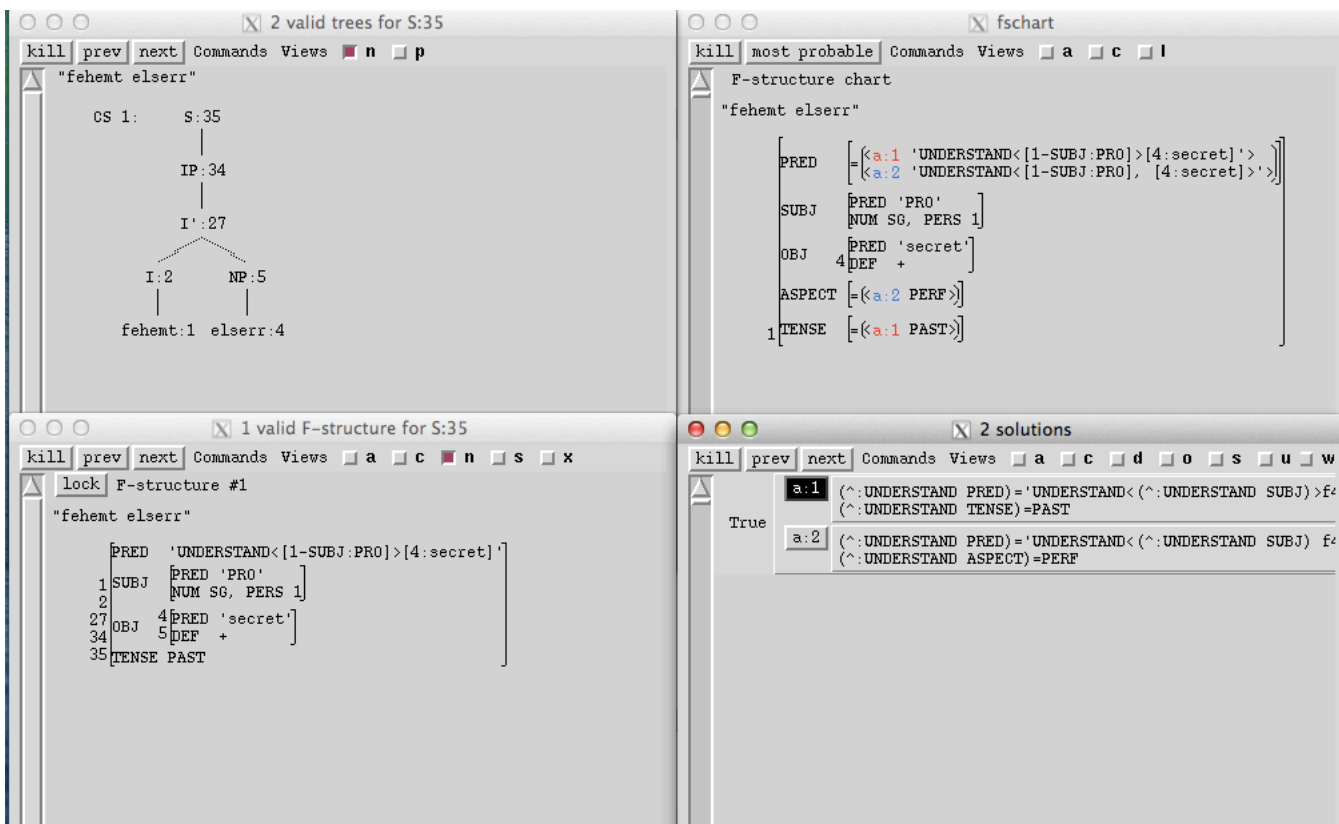[5]The inflectional category $\mathbf{I}$ is in boldface here to avoid the confusion with the pronoun I.
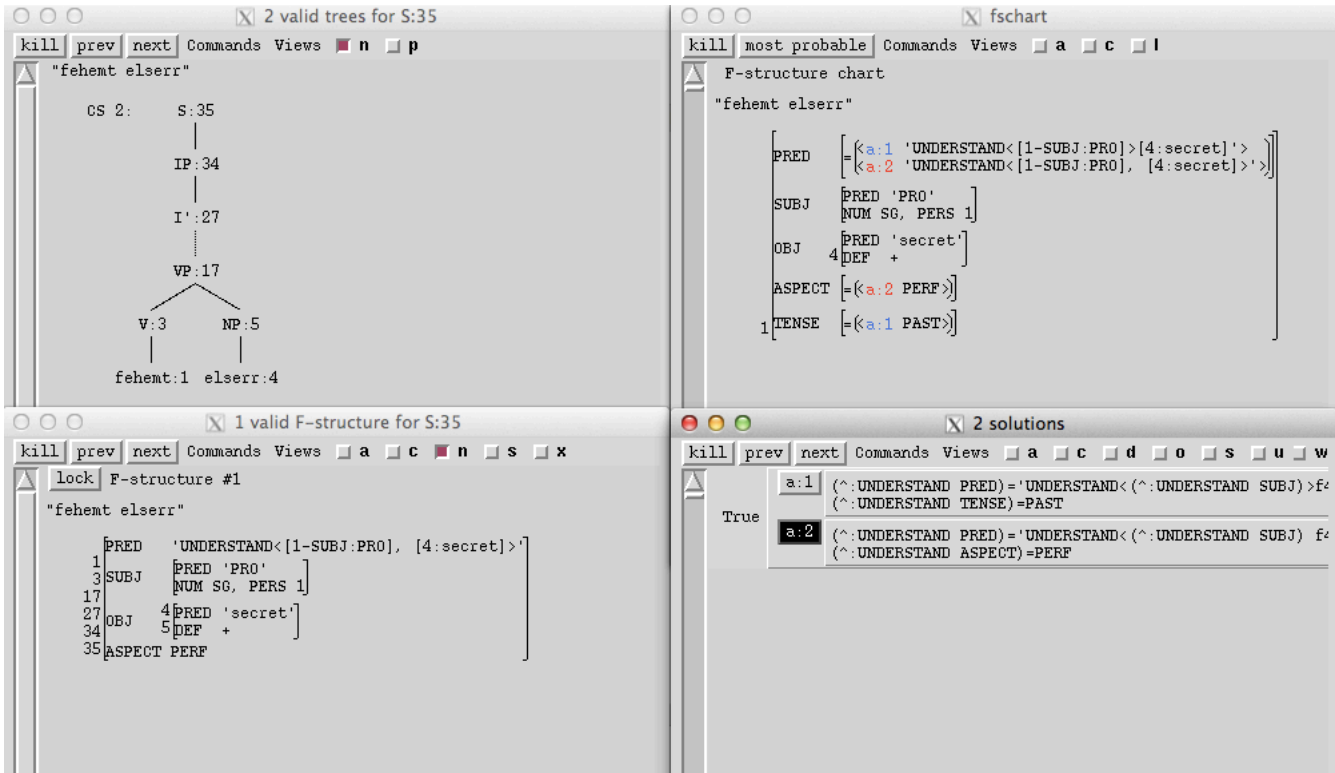
Figure 6.1: Parsing1

Figure 6.2: Parsing2

These problems are not adressed in the current study due to limitations of time, but they are to be considered in future versions of this grammar fragment. Section 6.6 below provides a summary of the findings of the XLE implementation, discussed in the current chapter.

6.6   Summary

The main aim of the current study was to provide a description of a number of verbal complementation patterns in ECA, and develop the suitable syntactic analysis for these structures within the LFG theory. XLE provided a good platform for testing the LFG analysis that I have developed for different ECA verbs throughout the current study, where combining all of the analyses of different verbs together facilitated the comparison between these structures and spotting any contradictions between them. Example for such cases is the introduction of the template **@OBJ** which specifies that the verb has an attached object pronoun, this is used only with causative verbs in the current study as they show subject to object raising, unlike the rest

of the main predicates investigated in the current study (namely the auxiliary, phasal verbs and modals) which can not take object pronouns attached to them.

This grammar fragment can be modified in future versions by increasing the size of the lexicon as well as including different kinds of complementation patterns so as to cover a bigger portion of ECA sentence structures. There were also problems that I could not resolve in the current grammar fragment such as multiple parsing results for verbs occuring in either **I** or V as well as representation of the optional PRO argument in arbitrary control predicates, resolving these issues could be a good starting point for enhancement of this grammar fragment.

Chapter 7

Conclusion

7.1 Summary

The current study aimed at answering some questions about a number of verbal complementation patterns in ECA, namely the auxiliary *kān* in compound tense structures, the causative verb *ḫalla*, phasal verbs as well as modals. Each of these verbs was represented by a set of sentences extracted from a corpus of ECA that is composed of online texts that was built for the purposes of the current study. The kinds of verbal complements were then described in terms of the possible forms of the main predicate, its verbal complement, the ability of the complement phrase to be headed by a complementiser, negation patterns on both verbs as well as shared arguments between the main predicate and the following lexical verb. Results of these descriptions were then used to reach the suitable analysis for the verbal structures in terms of their status as mono- or bi-clausal structures, and the kind of control relation between the main predicate and the following lexical verb in terms of being either functional or anaphoric control according to the principles of LFG syntactic theory. Finally, each of the verbal structures was represented in a unified grammar fragment implemented using XLE, where this grammar is capable of parsing all and only the grammatical sentences in ECA on which the analysis of these verbs was based.

253

The study also tackled a number of controversial issues about ECA, one of which is whether the bare imperfective verb form constitutes the non-finite verb in ECA, which is related to the issue of explaining how tense and aspect are represented in ECA. Another interesting issue is the importance of having corpus-based data for such descriptive studies of the ECA dialect, and the problems encountered during collecting such a corpus using the world wide web as the source of data. Conclusions reached about each of these issues will be represented in the following sections.

7.2   Findings

7.2.1   Tense and Aspect

Both tense and aspect can be expressed on the verb in ECA, however this is exemplified in different ways depending on whether the verb is in a simple or compound tense construction. In simple tense forms where the main verb occupies I in the c-structure, the verb carries tense only, with the addition of the features HAB and/or PROG on *bi*-prefixed Imperfective verbs. However, in compound tense constructions, the main predicate (such as auxiliary *kān*) marks tense and occupies I, while the following lexical verb marks aspect and occupies V in compound tense constructions[1].

This compound tense marking also applies in cases where the main verbal predicate is not the auxiliary *kān*, but rather another verb form that inflects for temporal marking and occupies I, and which can be followed by either a finite or a non-finite form lexical verb. Such verbs which were investigated in the current study are the causative *ḫalla* as well as the phasal verbs *rigiʕ* 'return', *baʔa* 'become', *fiḍil* 'remain' and *ʔaʕad* 'go.on'. This is shown in:

---

[1]Table 2.6 shows the different verb forms for both the auxiliary and lexical verb, as well as the tense or aspectual marking on each in I and V.

(1) baʔa   bi-yemʃi   zay el-batrīq
    become.PV.3SGM BI-walk.IPFV.3SGM   like the-penguin

    Lit: He became (such that) he walks like a penguin

    He walks like a penguin

Where the construction denotes past habitual with the subject getting into a habit of walking like a penguin that he did not have before. The phasal verb denotes past tense and occupies I while the lexical verb denotes imperfect aspect with the habitual feature added and occupies V, therefore expressing a past habitual construction.

While the rest of the phasal verbs had complements of only non-finite verb forms, and modals which do not carry any temporal inflection, and therefore such constructions do not denote compound tense, despite the fact that the modals and these phasal verbs still occupy I. In such cases, the phasal verb in I marks tense on the whole structure, while the lexical verb only denotes the lexical meaning with no expression of grammatical aspect, as shown in example 2 below where the construction denotes past tense only.

(2) ʔalb saʕd badaʔ   yitħarrak
    heart Saad start.PV.3SGM   move.IPFV.3SGM
    Saad's heart started to move (beat)

The *bi*-prefix has an important function in specifying tense and aspect in ECA, where it marks present tense on main verbs in I and marks imperfect aspect on lexical verbs in V, it also adds a HAB feature on stative verbs and both HAB and PROG on non-stative verbs. These properties of the bi-prefix can also account for the status of the bare imperfective form verbs (which lack the *bi*-prefix) as a non-finite verb, where it is prohibited from occupying I and denoting present tense in main clauses, and therefore bare forms occur only in dependent clauses and are marked by (↑VFORM)=BARE as a standard in the XLE analysis of all verbs in the current study. Properties of the *bi*-prefix marking is discussed in more details in section 2.7.3.

7.2.2 Functional and Anaphoric Control

Despite the fact that all of the verbal constructions analysed in the current study represented bi-clausal structures, they differed in the kind of control relation attested between the main verbal predicate and its verbal complement. The definitions of anaphoric and functional control within LFG discussed in section 1.5 were applied to each of the verbal constructions and the result of analysis was based on the behaviour of each of these verbal complementation patterns.

Functional control was attested in constructions where the main predicate is the auxiliary *kān*, the causative verb *ẖalla*, phasal verbs, as well as non-inflecting modals. All of these verbal predicates had XCOMPs as their complements. However, the causative verb *ẖalla* was the only verbal predicate to have the SUBJ as thematic argument, and showed cases of raising to object, where the subject of the following lexical verb was also the object of the causative verb. The rest of the raising verbs had the SUBJ as a non-thematic argument and showed raising to subject.

Anaphoric control was attested only in the case of inflecting modals, with the modal *yiʔdar* 'able' showing a case of obligatory anaphoric control while the rest of inflecting modals showed arbitrary anaphoric control. One of the main motives for reaching this analysis was the ability of anaphoric control predicates to be followed by a complementiser heading the complement clause, which was not possible with the functional control predicates. The issue of the possiblity of having a complementiser presented one of the challenges to the current study, where in some cases I (as a native speaker) was not sure if the complementiser was accpetable or not, and therefore such sentences were disregarded, and the only cases in which the complementiser was totally accetable occured with anaphoric control predicates.

7.2.3 Non-finite Verb

The status of the bare Imperfective verb form as a non-finite verb in ECA was established throughout the current study based on reasons such as:

- The bare Imperfective verb did not occur as the main predicate of indicative sentences in any of the sentences analysed, but rather occured only in dependent clauses or as a complement to other verbal predicates.

- All of the verbal predicates analysed can be followed by the bare Imperfective form verb, while they often have restrictions on other verb forms including the imperative. This leads to the assumption that the bare Imperfective verb represents the basic lexical meaning of the verb in its non-finite form.

- This bare Imperfective form is also considered as the non-finite verb form in a number of studies on Arabic generally, such as Hallman (2015), as well as studies on Egyptian Arabic specifically, such as (Jelinek, 1981; Abdel-Massih et al., 1979).

This led to marking the lexical entry of bare Imperfective form verbs as **VFORM=BARE**, this is to satisfy the constraint on a number of the complementation patterns investigated in the current study, where the main verb can only take verbal complements in the bare Imperfective form, this is the case with the causative verb _ẖalla_ and some phasal verbs as well as some modals. However, in the case where the bare Imperfective form combines with the past tense auxiliary _kān_ to denote past imperfective, the bare Imperfective verb marks imperfective aspect and its lexical entry is marked as **ASPECT=IMPF**. This represents a case of syncretism, where the same **form** (bare Imperfective) is used as a non-finite verb that does not carry any grammatical aspect as well as a marker for imperfective aspect when used in compound tense structure following the auxiliary _kān_. Note however that the only case in which the bare Imperfective verb form was used in compound tense was to denote **past imperfective**, and it occured only once in the corpus (in example 55), which implies that this is an exception to the use of the bare Imperfective as the non-finite form in ECA, and should not affect the assumption that the bare Imperfective verb form represents the non-finite verb in ECA.

7.3   Theoretical Implications

The findings of this study shall have a positive impact on the description as well as syntactic analysis of Egyptian Colloquial Arabic. Where as far as I am aware, there were no previous studies on this dialect within the focus of LFG syntactic theory, which opens the door to further argumentation about these structures and how they shall be represented. Another important contribution of this study is providing an example for a corpus-based study in which the corpus was built using the World Wide Web as a source of data and Sketch Engine as the tool to build and store this corpus. While the corpus remains till now available to use only through my account on Sketch Engine, the steps of building the corpus were explained in section 1.3.1 and can be easily utilised in building a similar corpus.

Establishing the bare Imperfective verb form as the non-finite verb in ECA was also one of the main findings of the thesis. This can lead to further developments in the understanding of finiteness in languages where the verb still inflects to agree with the subject in number, person and gender as finite verbs do while lacking the properties and functions attested with finite verbs, such as functioning as the main predicate in indicative sentences as well as occupying I and specifying tense of the event described by the verb. Last but not least, this is the first study, as far as I am aware, which attempts to develop a grammar for ECA within the scope of the XLE platform. While the grammar developed here was limited to the verbal predicates studied only, and therefore represents a very small section of the language, it provides an insight over the possibility of developing such a grammar as well as the issues correlated with such work. This could also open the door for inclusion of the ECA dialect within the group of languages covered by the ParGram project, in order to help develop broad coverage grammars that can cover a wide range of phenomena in as much bigger a number of languages as possible.

7.4   Future Research

While the current study was limited by the size of the corpus used, the number of main verbal predicates investigated, as well as including only sentences where the main verb was followed by another lexical verb, overcoming these limitations could give rise to interesting future studies. These include investigating different kinds of predicates that are still understudied regarding the dialect, such as the participle and the role it plays as a main predicate in some sentences. Another interesting area to investigate is the sequence of more than two verbs following each other which is possible in ECA, and how these verbs will be represented in the c-structure as well as the tense and aspect marking on each.

Other verbal predicates can also offer an opportunity for interesting research on ECA. These are described in Al-Aqarbeh (2011) as complement taking predicates (CTP) and include utterance predicates (such as *ʔāl* 'tell' in ECA), propositional atitude predicates (such as *ʔiftakar* 'think'- also means 'remember' in ECA) and predicates of knowledge (such as *ʕirif* 'know' in ECA). These verbs can have different forms of complements such as verbal and non-verbal complements, as well as complementisers. Investigating the behaviour of a number of these predicates can provide more insights into the syntactic structure of ECA.

Finally, these theoretical finidings can be then utilised into developing a grammar for ECA that can be used to parse and generate different possible sentences in the language, which would then give rise to useful applications in computational linguistics and its applications for ECA.

Bibliography

Abdel-Massih, E. T., Z. N. Abdel-Malek, and E.-S. M. Badawi (1979). *A Reference Grammar of Egyptian Arabic*. Center for Near Eastern and North African Studies at the University of Michigan.

Abdelmoumene, N. (1988). *Sentential Complementation in French, English, and Arabic*. Ph. D. thesis, University of Essex.

Aboul-Fetouh, H. M. (1969). *A Morphological study of Egyptian Colloquial Arabic*. Mouton & Co. N. V., Publishers, The Hague.

Ademola-Adeoye, F. F. (2011). *A Cross-linguistic Analysis of Finite Raising Constructions*. Ph. D. thesis, University of KwaZulu-Natal.

Aikhenvald, A. Y. and R. Dixon (Eds.) (2006). *Serial Verb Constructions: A Cross-Linguistic Typology*. Oxford University Press.

Al-Aqarbeh, R. (2011). *Finiteness in Jordanian Arabic: A Semantic and Morphosyntactic approach*. Ph. D. thesis, University of Kansas.

Al-Sabbagh, R., R. Girju, and J. Diesner (2014). 3arif: A corpus of modern standard and egyptian arabic tweets annotated for epistemic modality using interactive crowdsourcing. In

Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical papers, pp. 1521–1532.

AlKhawalda, M. I. (1997). *Tense, Aspect, and Time reference with reference to Arabic*. Ph. D. thesis, University of Essex.

Alsina, A. (1992). On the argument structure of causatives. In *Linguistic Inquiry*, Volume 23, pp. 517–555. MIT Press.

Arts, T., Y. Belinkov, N. Habash, A. Kilgarriff, and V. Suchomel (2014). artenten: Arabic corpus and word sketches. *Journal of King Saud University - Computer and Information Sciences 26*(4), 357–371.

Azer, H. A. (1980). *The expression of modality in Egyptian Colloquial Arabic: Its Syntax and Semantics*. Ph. D. thesis, University of London- School of Oriental and African Studies.

Badawi, A. M. (2012). *mosatawayaat al'arabiyya almo'aasera fi masr*. Dar-AlSalam.

Benmamoun, E. (1991). Causatives in moroccan arabic. In B. Comrie and M. Eid (Eds.), *Perspectives on Arabic linguistics III: Papers from the Third annual symposium on Arabic Linguistics*, pp. 173–198. John Benjamins Publishing Company.

Benmamoun, E. (2000). *The Feature Structure of Functional Categories. A comparative study of Arabic Dialects*. Oxford University Press.

Beshr, K. (1998). *dirasaat fi 'elm allogha*. Dar- Ghareeb.

Bhatt, R., T. Bogel, M. Butt, A. Hautli, and S. Sulger (2011). Urdu/hindi modals. In M. Butt and T. H. King (Eds.), *Proceedings of the LFG11 Conference*, pp. 47–67. CSLI Publications.

Binnick, R. I. (1991). *Time and the verb*. Oxford University Press.

Bresnan, J. (Ed.) (1982). *The Mental Representation of Grammatical Relations*. MIT Press.

Bresnan, J. (2001). *Lexical Functional Syntax*. Blackwell Publishing.

Brustad, K. E. (2000). *The Syntax of Spoken Arabic*. Washington, D.C.: GeorgeTown University Press.

Butt, M. (1995). The structure of complex predicates in urdu. In W. J. Joan Bresnan, Sharon Inkelas and P. Sells (Eds.), *Dissertations in Linguistics*. CSLI Publications.

Butt, M., T. H. King, M.-E. Nino, and F. Segond (1999). *A Grammar Writer's Cookbook*. CSLI Publications.

Camilleri, M., S. ElSadek, and L. Sadler (2014). A cross dialectal view of the arabic dative alternation. *Acta Linguistica Hungarica 61*(1), 3–44.

Chatsiou, A. K. (2010). *A Lexical Functional Grammar approach to Modern Relative Clauses*. Ph. D. thesis, University of Essex.

Comrie, B. (1976). *Aspect*. Cambridge University Press.

Comrie, B. (1985). *Tense*. Cambridge University Press.

Comrie, B. (1989). *Language Universals and Linguistic Typology: Syntax and Morphology* (Second Edition ed.). Blackwell Publishing.

Comrie, B. (Ed.) (2009). *The World's Major Languages* (Second Edition ed.). Routledge.

Dalrymple, M. (2001). Lexical functional grammar. In *Syntax and Semantics*, Volume 34. Academic Press.

Dixon, R. (2000). A typology of causatives: form, syntax and meaning. In R. Dixon and A. Y. Aikhenvald (Eds.), *Changing valency Case studies in transitivity*, pp. 30–79. Cambridge University Press.

Dixon, R. and A. Y. Aikhenvald (2000). Introduction. In *Changing valency Case studies in transitivity*, pp. 1–28. Cambridge University Press.

Dowty, D. (1991). Thematic proto-roles and argument selection. In *Language*, Volume 67, pp. 547–619. Linguistic Society of America.

Eisele, J. C. (1990a). Aspectual classification of verbs in cairene arabic. In M. Eid and J. McCarthy (Eds.), *Perspectives on Arabic linguistics, papers from the second annual symposium on Arabic linguistics*, pp. 192–233. Amsterdam and Philadelphia: John Benjamins.

Eisele, J. C. (1990b). Tense reference, tense and formal apsect in cairene arabic. *Current issues in Linguisitc Theory 63*, 173–212.

Eisele, J. C. (1992). Egyptian arabic auxilliaries and the category of aux. In M. E. Ellen Broselow and J. McCarthy (Eds.), *Perspectives on Arabic linguistics IV, papers from the 4th annual symposium on Arabic linguistics*, Volume 85 of *Current Issues in Linguistic Theory*, pp. 143–165. John Benjamins.

ElSadek, S. (2010). Discontinuous morphemes in modern standard arabic: With special reference to verbal nouns of triliteral verbs. Master's thesis, Alexandria University.

Escamilla, R. M. (2012). *An updated Typology of Causative Constructions: Form-Function Mappings in Hupa (California Athabaskan), Chungli Ao (Tibeto-Burman) and Beyond*. Ph. D. thesis, University of California, Berkeley.

Falk, Y. N. (2001). *Lexical-Functional Grammar. An Introduction to Parallel Constraint-Based Syntax*. CSLI Publications.

Falk, Y. N. (2008). Functional relations in the English auxiliary system. *Linguistics 46/4*, 861–889.

Fehri, A. F. (2012). *Key Features and Parameters in Arabic Grammar*. John Benjamins Publishing Company.

Gary, J. O. and S. Gamal-Eldin (1982). Cairene egyptian colloquial arabic. In *Lingua Descriptive Studies (LDS)*, Volume 6. North-Holland.

Hacquard, V. (2011). Modality. In C. Maierborn, K. von Heusinger, and P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning*, Volume 2, Chapter 58. Mouton de Gruyter.

Hallman, P. (2006). Causativity and transitivity in arabic. University of Toronto.

Hallman, P. (2015). The arabic imperfective. In *Brill's Journal of AfroAsiatic Languages and Linguistics*, Volume 7, pp. 103–131.

Hanna, H. M. (1967). *The Phrase structure of Egyptian Colloquial Arabic*. Mouton & Co. N. V., Publishers, The Hague.

Hoffman, M. C. (1991). *The Syntax of Argument-Structure-Changing Morphology*. Ph. D. thesis, Massachusetts Institute of Technology.

Jelinek, M. E. (1981). *On defining categories: AUX and Predicate in Colloquial Egyptian Arabic*. Ph. D. thesis, University of Arizona.

Joshi, S. (1993). *Selection of grammatical and logical functions in Marathi*. Ph. D. thesis, Stanford University.

Kaplan, R. M., T. H. King, and J. T. Maxwell (2002). Adapting existing grammars: the xle experience. In *Proceedings COLING-GEE '02- the 2002 workshop on Grammar engineering and evaluation*, Volume 15, pp. 1–7. Association for Computational Linguistics.

Kearns, K. (2011). *Semantics*. Palgrave Macmillan.

Klein, W. (1992). The present perfect puzzle. *Language 68*, 525–552.

Klein, W. (2009). *The Expression of Time*, Chapter How time is encoded, pp. 39–82. The Expression of Cognitive Categories. Mouton de Gruyter.

Kroeger, P. R. (2004). *Analyzing syntax: A Lexical-Functional approach*. Cambridge University Press.

Landau, I. (2004). The scale of finiteness and the calculus of control. *Natural Language and Linguistic Theory 22*, 811–877.

Lee, K. Y. (2009). *Finite Control in Korean.* Ph. D. thesis, University of Iowa.

Lodrup, H. (2014). How can a verb agree with a verb? reanalysis and pseudocoordination in norwegian. In M. Butt and T. H. King (Eds.), *Proceedings of the LFG14 Conference*, pp. 367–386. CSLI Publications.

Lovestrand, J. (2009). French causatives in lfg: Adverbial compements as evidence of subject-hood properties of the causee.

Maas, U. (2009). Complex predicates in maltese. from a neo-arabic perspective. In B. C. et al. (Ed.), *Introducing Maltese Linguistics*, Volume 113 of *Studies in Language Companion Series (SLCS)*, pp. 113–132. John Benjamins Publishing Company.

Matsumoto, Y. (1992). *On the wordhood of complex predicates in Japanese.* Ph. D. thesis, Stanford University.

Matsumoto, Y. (1998). A reexamination of the cross-linguistic parameterization of causative predicates: Japanese perspectives. In M. Butt and T. H. King (Eds.), *Proceedings of the LFG98 Conference.* CSLI Publications.

McEnery, T. and A. Hardie (2012). *Corpus Linguistics. Method, Theory and Practice.* Cambridge University Press.

Michaelis, L. A. (1998). *Aspectual Grammar and Past-Time Reference.* Routledge studies in Germanic Linguistics. Routledge.

Moreno, J. C. (1993). "make" and the origins of causativity: a typological study. In B. Comrie and M. Polinsky (Eds.), *Causatives and Transitivity*, Volume 23 of *Studies in Language Companion Series (SLCS)*, pp. 155–164. John Benjamins Publishing Company.

Moshref, O. (2012). *Corpus study of tense, aspect, and modality in diglossic speech in Cairene Arabic.* Ph. D. thesis, University of Illinois, Urbana, Chapaign.

Nagy, T. (2009). Aspectual coercion and the complementation of aspectualizers in english. *Acta Universitatis Sapientiae, Philologica 1*, 229–240.

Newmeyer, F. J. (1975). *English Aspectual Verbs.* Mouton, The Hague.

Noonan, M. (2007). Complementation. In T. Shopen (Ed.), *Language Typology and Syntactic Description*, pp. 52–150. Cambridge University Press.

Nuyts, J. (2005). The modal confusion: on terminology and the concepts behind it. In A. Klinge and H. H. Müller (Eds.), *Modality. Studies in Form and Function.*, pp. 5–38. Equinox Publishing Ltd.

Payne, T. E. (1997). *Describing morphosyntax: A guide for field linguists.* Cambridge University Press.

Ritz, M.-E. (2012). *The Oxford Handbook of Tense and Aspect*, Chapter Perfect Tense and Aspect, pp. 881–907. Oxford University Press.

Ryding, K. C. (2005). *A Reference Grammar of Modern Standard Arabic.* Cambridge University Press.

Sells, P. (2007). Finiteness in non-transformational syntactic frameworks. In I. Nikolaeva (Ed.), *Finiteness: Theoretical and empirical foundations*, pp. 59–88. Oxford University Press.

Sharif, A. A. and L. Sadler (2009). Negation in modern standard arabic: An LFG approach. In M. Butt and T. H. King (Eds.), *Proceedings of theLFG09 Conference*, pp. 5–25. CSLI publications.

Shibatani, M. (1973). Lexical versus periphrastic causatives in korean. *Journal of Linguistics 9*(2), 281–297.

Shibatani, M. (2002). *The Grammar of Causation and Interpersonal Manipulation*, Volume 48 of *Typological Studies in Language (TSL)*, Chapter Introduction, pp. 1–22. John Benjamins Publishing Company.

Smith, C. S. (1991). *The Parameter of Aspect*. Studies in Linguistics and Philosophy. Kluwer Academic Publishers.

Stefanowitsch, A. (2001). *Constructing Causation: A Construction Grammar Approach to Analytic Causatives*. Ph. D. thesis, Rice University, Houston, Texas.

Tallerman, M. (2011). *Understanding Syntax*. Hodder Education.

Talmy, L. (2000). *Toward a Cognitive Semantics*, Volume 1: Concept Structuring Systems. MIT Press.

Tawakol, H. (2008). Code mixing in 'ammiyat al-muthaqqafin: A case study of the use of the b-prefix as a tense marker. Master's thesis, The American University in Cairo.

Van Lambalgen, M. and F. Hamm (2005). *The proper treatment of events*. Blackwell Publishing.

Versteegh, K. (2004). From classical arabic to the modern arabic vernaculars. In G. Booij, C. Lehman, J. Mugdan, and S. Skopetas (Eds.), *Morphology: An international handbook on inflection and word-formation*, Volume II, pp. 1740– 1754. Mouton de Gruyter.

Watson, J. C. E. (2002). *The Phonology and Morphology of Arabic*. Oxford University Press.

Wehr, H. (1976). A dictionary of modern written arabic. Spoken Languages Services, Inc.

Woidich, M. (2006). Cairo arabic. In K. V. et al. (Ed.), *Encyclopedia of Arabic Language and Linguistics*. Leiden- The Netherlands: Koninklijke Brill NV.

Yates, N. (2002). French causatives: a biclausal account in lfg. In M. Butt and T. H. King (Eds.), *Proceedings of the LFG02 Conference*, pp. 390–407. CSLI Publications.

Zaghouani, W. (2014). Critical survey of the freely available arabic corpora. In *Proceedings of the Workshop on Free/Open-Source Arabic Corpora and Corpora Processing Tools, LREC 2014*, Reykjavik, Iceland, pp. 26–31.

Zayed, S. H. (1983). *A pragmatic approach to modality and the modals: With applications to Literary Arabic*. Ph. D. thesis, University of Edinburgh.

Appendix A

XLE Grammar File

```
VERBS    ARABIC    CONFIG (1.0)
    ROOTCAT           S.
    LEXENTRIES        (VERBS ARABIC).
    RULES             (VERBS ARABIC).
    GOVERNABLERELATIONS SUBJ XCOMP OBJ COMP.
    SEMANTICFUNCTIONS ADJUNCT.
    TEMPLATES         (VERBS ARABIC).

----

VERBS ARABIC TEMPLATES (1.0)

ControlI = (^SUBJ)=(^XCOMP SUBJ).

ControlIC = (^OBJ)=(^XCOMP SUBJ).

ControlICP = (^COMP SUBJ PRED)= 'PRO'.

OBJ      = (^OBJ PRED)='PRO'.

NoOBJ    = ~ (^OBJ PRED)='PRO'.

rigi3t   = @ControlI * (^PRED)='RETURN<(^XCOMP)>(^SUBJ)'.
```

```
2irga3   = @ControlI * (^PRED)='RETURN<(^XCOMP)>(^SUBJ)'.

NoIMP    = ~ (^XCOMP VFORM) = IMP.

NoBare   = ~ (^XCOMP VFORM) = BARE.

3SGM = (^SUBJ PERS)=3 (^SUBJ NUM)=SG (^SUBJ GEN)=MASC.

3SGF = (^SUBJ PERS)=3 (^SUBJ NUM)=SG (^SUBJ GEN)=FEM.

2SGM = (^SUBJ PERS)=2 (^SUBJ NUM)=SG (^SUBJ GEN)=MASC.

2SGF = (^SUBJ PERS)=2 (^SUBJ NUM)=SG (^SUBJ GEN)=FEM.

1SG  = (^SUBJ PERS)=1 (^SUBJ NUM)=SG.

3PL  = (^SUBJ PERS)=3 (^SUBJ NUM)=PL.

2PL  = (^SUBJ PERS)=2 (^SUBJ NUM)=PL.

1PL  = (^SUBJ PERS)=1 (^SUBJ NUM)=PL.

SGF  = (^SUBJ NUM)=SG (^SUBJ GEN)=FEM.

SGM  = (^SUBJ NUM)=SG (^SUBJ GEN)=MASC.

3SGFOBJ = (^OBJ PERS)=3 (^OBJ NUM)=SG (^OBJ GEND)=FEM.

3SGMOBJ = (^OBJ PERS)=3 (^OBJ NUM)=SG (^OBJ GEND)=MASC.

2SGFOBJ = (^OBJ PERS)=2 (^OBJ NUM)=SG (^OBJ GEND)=FEM.

2SGMOBJ = (^OBJ PERS)=2 (^OBJ NUM)=SG (^OBJ GEND)=MASC.

1SGOBJ  = (^OBJ PERS)=1 (^OBJ NUM)=SG.

----
```

```
VERBS ARABIC RULES (1.0)

S   --> IP: ^=!.

IP --> {NP: (^SUBJ)=! | e: (^ SUBJ PRED)='PRO';}

       ({ADVP: (^ADJUNCT)=!})

       I': ^=!.

I' --> ({I: ^=!;})

       ({VP: (^COMP)=!})

       ({NP: (^OBJ)=! | e: (^ OBJ PRED)='PRO';})

       ({CP: (^COMP)=!})

       ({ADVP: (^ADJUNCT)=!})

       ({VP: (^XCOMP)=!})

       ({VP: ^=!})

       ({ADVP: (^ADJUNCT)=!}).

VP --> V: ^=!;
```

```
          ({NP: (^OBJ)=!}
          ({VP: (^XCOMP)=!}
          ({PP: (^ADJUNCT)=!}
          ({CP: (^ADJUNCT)=!}).

ADVP --> ADV: ^=!.

PP  ---> P: ^=!;
         {NP: (^OBJ)=!| e: (^ OBJ PRED)='PRO';}.

NP ----> N: ^=!;
         ({PP: (^ADJUNCT)=!}).

CP ----> C': ^=!.

C' ---> C : ^=!;
         IP: (^ADJUNCT)=!.
```

```
----

VERBS   ARABIC   LEXICON (1.0)

2ana        NP * (^PRED)='I'               (^PERS)=1  (^NUM)=SG.

2e7na       NP * (^PRED)='we'             (^PERS)=1  (^NUM)=PL.

homma       NP * (^PRED)='they'           (^PERS)=3  (^NUM)=PL.

howwa       NP * (^PRED)='he'             (^PERS)=3  (^NUM)=SG  (^GEN)=MASC.

2enta       NP * (^PRED)='you'            (^PERS)=2  (^NUM)=SG  (^GEN)=MASC.

2enti       NP * (^PRED)='you'            (^PERS)=2  (^NUM)=SG  (^GEN)=FEM.

elwalad     NP * (^PRED)='boy'            (^PERS)=3  (^NUM)=SG  (^GEN)=MASC  (^DEF)=+.

elraagel    NP * (^PRED)='man'            (^PERS)=3  (^NUM)=SG  (^GEN)=MASC  (^DEF)=+.

elmobarrer  NP * (^PRED)='reason'         (^PERS)=3  (^NUM)=SG  (^GEN)=MASC  (^DEF)=+.

bani2adam   NP * (^PRED)='person'         (^PERS)=3  (^NUM)=SG  (^GEN)=MASC.

elnaas      NP * (^PRED)='people'         (^DEF)=+   (^NUM)=PL.

elgama3aat  NP * (^PRED)='groups'         (^DEF)=+   (^NUM)=PL.

ele7teraaf  NP * (^PRED)='professionalism'   (^DEF)=+  (^NUM)=SG.
```

```
elmozakra      NP *  (^PRED)='studying'        (^DEF)=+  (^NUM)=SG.
el2igaar       NP *  (^PRED)='rent'            (^DEF)=+  (^NUM)=SG.
elsiyaasa      NP *  (^PRED)='politics'        (^DEF)=+  (^NUM)=SG.
youm           NP *  (^PRED)='day'             (^NUM)=SG.
youmeen        NP *  (^PRED)='two_days'        (^NUM)=DUAL.
elbatriiq      NP *  (^PRED)='pinguin'         (^DEF)=+.
elfasl         NP *  (^PRED)='class'           (^DEF)=+.
elgalabeyya    NP *  (^PRED)='gown'            (^DEF)=+.
elxoof         NP *  (^PRED)='fear'            (^DEF)=+.
el2imeel       NP *  (^PRED)='email'           (^DEF)=+.
el7ayaa        NP *  (^PRED)='life'            (^DEF)=+.
elshorta       NP *  (^PRED)='police'          (^DEF)=+.
elfacebook     NP *  (^PRED)='Facebook'        (^DEF)=+.
elshakk        NP *  (^PRED)='doubt'           (^DEF)=+.
elmawduu3      NP *  (^PRED)='issue'           (^DEF)=+.
elserr         NP *  (^PRED)='secret'          (^DEF)=+.
2elbarnaameg   NP *  (^PRED)='program'         (^DEF)=+.
elmat3am       NP *  (^PRED)='restaurant'      (^DEF)=+.
elgam3a        NP *  (^PRED)='university'      (^DEF)=+.
elmas2uleyya   NP *  (^PRED)='responsibility'  (^DEF)=+.
```

```
el7uduud    NP * (^PRED)='border'    (^NUM)=PL  (^DEF)=+.
2awlaadu    NP * (^PRED)='kids_his'  (^NUM)=PL.
fawwaaz     NP * (^PRED)='Fawaz'.
kintaki     NP * (^PRED)='kfc'.
billeil     NP * (^PRED)='night'.
kaam        NP * (^PRED)='how_much'.
ma7addesh   NP * (^PRED)='no_one'.
2alb_sa3d   NP * (^PRED)='Saad_heart'.
2usluubu    NP * (^PRED)='his_style'.
2imeel      NP * (^PRED)='email'.
tarii2u     NP * (^PRED)='his_way'.
mobail      NP * (^PRED)='mobile'.
7assan      NP * (^PRED)='Hassan'.
saami       NP * (^PRED)='Sami'.
mark        NP * (^PRED)='Mark'.
sondos      NP * (^PRED)='Sondos'.
soot        NP * (^PRED)='sound'.
goozek      NP * (^PRED)='your_husband'.
baskit      NP * (^PRED)='basket_ball'.
nafsi       NP * (^PRED)='myself'.
```

```
mo7taref    NP  *  (^PRED)='professional'.

mayya       NP  *  (^PRED)='water'.

mona        NP  *  (^PRED)='Mona'.

2afas       NP  *  (^PRED)='cage'.

fatra       NP  *  (^PRED)='time'.

sana        NP  *  (^PRED)='year'.

no2ta       NP  *  (^PRED)='point'.

diini       NP  *  (^PRED)='my_religion'.

2a7mad      NP  *  (^PRED)='Ahmed'.

baaba       NP  *  (^PRED)='dad'.

7aaga       NP  *  (^PRED)='something'.

rasaayel    NP  *  (^PRED)='messages'.

3eenak      NP  *  (^PRED)='your_eyes'.

2ahlek      NP  *  (^PRED)='your_family'.

masr        NP  *  (^PRED)='Egypt'.

koora       NP  *  (^PRED)='ball'.

film        NP  *  (^PRED)='movie'.

huguum      NP  *  (^PRED)='attack'.

elMBC       NP  *  (^PRED)='MBC'.

2amriika    NP  *  (^PRED)='America'.
```

```
ma3had_guuta  NP * (^PRED)='Goethe_Institute'.

2inn      C * (^PRED)='that'.

2innak    C * (^PRED)='that_you'.

3ashaan   C * (^PRED)='such_that'.


kaan      I * @ControlI (^PRED)='BE<(^XCOMP)>(^SUBJ)' @3SGM (^TENSE)=PAST.

kaanit    I * @ControlI (^PRED)='BE<(^XCOMP)>(^SUBJ)' @3SGF (^TENSE)=PAST.

kont      I * @ControlI (^PRED)='BE<(^XCOMP)>(^SUBJ)' @1SG (^TENSE)=PAST.

konna     I * @ControlI (^PRED)='BE<(^XCOMP)>(^SUBJ)' @1PL (^TENSE)=PAST.

7ankuun   I * @ControlI (^PRED)='BE<(^XCOMP)>(^SUBJ)' @1PL @NoBare (^TENSE)=FUTURE.


ykuun     V * (^PRED)='BE<(^SUBJ)>' @3SGM (^VFORM)=BARE.


xallo     I * @ControlIC (^PRED)='MAKE<(^SUBJ)(^XCOMP)>(^OBJ)' @NoOBJ @3PL (^TENSE)=PAST.

xalleet   I * @ControlIC (^PRED)='MAKE<(^SUBJ)(^XCOMP)>(^OBJ)' @NoOBJ @2SGM (^TENSE)=PAST.

7atxalli  I * @ControlIC (^PRED)='MAKE<(^SUBJ)(^XCOMP)>(^OBJ)' @NoOBJ @2SGF (^TENSE)=FUT.

xalli     I * @ControlIC (^PRED)='MAKE<(^SUBJ)(^XCOMP)>(^OBJ)' @NoOBJ @2SGM (^VFORM)=IMP.
```

```
xalla        I  *  @ControlIC (^PRED)='MAKE<(^SUBJ)(^XCOMP)>(^OBJ)' @NoOBJ   @3SGM  (^TENSE)=PAST;
             I  *  (^PRED)='MAKE<(^SUBJ)(^COMP)>'                           @3SGM  (^TENSE)=PAST.
xallah       I  *  @ControlIC (^PRED)='MAKE<(^SUBJ)(^XCOMP)>(^OBJ)' @OBJ    @3SGM  (^TENSE)=PAST.
xallaani     I  *  @ControlIC (^PRED)='MAKE<(^SUBJ)(^XCOMP)>(^OBJ)' @OBJ    @3SGM @1SGOBJ  (^TENSE)=PAST.
7atxalliini  I  *  @ControlIC (^PRED)='MAKE<(^SUBJ)(^XCOMP)>(^OBJ)' @OBJ    @2SGF @1SGOBJ  (^TENSE)=FUT.
7atxalliih   I  *  @ControlIC (^PRED)='MAKE<(^SUBJ)(^XCOMP)>(^OBJ)' @OBJ    @2SGF @3SGMOBJ  (^TENSE)=FUT.
bixalliik    I  *  @ControlIC (^PRED)='MAKE<(^SUBJ)(^XCOMP)>(^OBJ)' @OBJ    @3SGM @2SGMOBJ  (^TENSE)=PRESENT (^HAB)=+.
xalliiha     I  *  @ControlIC (^PRED)='MAKE<(^SUBJ)(^XCOMP)>(^OBJ)' @OBJ    @2SGM @3SGFOBJ  (^VFORM)=IMP.

rigi3        I  *  @ControlI (^PRED)='RETURN<(^XCOMP)>(^SUBJ)' @ NoIMP @3SGM  (^TENSE)=PAST.
rigi3na      I  *  @ControlI (^PRED)='RETURN<(^XCOMP)>(^SUBJ)' @ NoIMP @1PL   (^TENSE)=PAST.
2a3adna      I  *  @ControlI (^PRED)='GO_ON<(^XCOMP)>(^SUBJ)' @ NoIMP @1PL    (^TENSE)=PAST.
ba2eena      I  *  @ControlI (^PRED)='BECOME<(^XCOMP)>(^SUBJ)' @ NoIMP @1PL   (^TENSE)=PAST.
ba2a         I  *  @ControlI (^PRED)='BECOME<(^XCOMP)>(^SUBJ)' @ NoIMP @3SGM  (^TENSE)=PAST.
fidil        I  *  @ControlI (^PRED)='REMAIN<(^XCOMP)>(^SUBJ)' @ NoIMP @3SGM  (^TENSE)=PAST.
fidilna      I  *  @ControlI (^PRED)='REMAIN<(^XCOMP)>(^SUBJ)' @ NoIMP @1PL   (^TENSE)=PAST.
rigi3tu      I  *  @ControlI (^PRED)='RETURN<(^XCOMP)>(^SUBJ)' @ NoIMP @2PL   (^TENSE)=PAST.
rigi3t       I  *  @ControlI (^PRED)='RETURN<(^XCOMP)>(^SUBJ)' @ NoIMP @1SG   (^TENSE)=PAST.
7afdal       I  *  @ControlI (^PRED)='REMAIN<(^XCOMP)>(^SUBJ)' @ NoIMP @1SG   (^TENSE)=FUT.
2irga3       I  *  @ControlI (^PRED)='RETURN<(^XCOMP)>(^SUBJ)'         @2SGM  (^XCOMP VFORM)=c IMP (^VFORM)=IMP.
```

```
2irga3u     I  *  @ControlI  (^PRED)='RETURN<(^XCOMP)>(^SUBJ)'      @2PL   (^XCOMP VFORM)=c IMP  (^VFORM)=IMP.
matirga3sh  I  *  @ControlI  (^PRED)='NOT_RETURN<(^XCOMP)>(^SUBJ)'  @2SGM  (^VFORM)=IMP.

bada2       I  *  @ControlI  (^PRED)='BEGIN<(^XCOMP)>(^SUBJ)'       @3SGM  (^XCOMP VFORM)=c BARE  (^TENSE)=PAST.
bada2t      I  *  @ControlI  (^PRED)='BEGIN<(^XCOMP)>(^SUBJ)'       @1SG   (^XCOMP VFORM)=c BARE  (^TENSE)=PAST;
            V  *  @ControlI  (^PRED)='BEGIN<(^XCOMP)>(^SUBJ)'       @1SG   (^XCOMP VFORM)=c BARE  (^ASPECT)=PERF.
bada2tu     I  *  @ControlI  (^PRED)='BEGIN<(^XCOMP)>(^SUBJ)'       @2PL   (^XCOMP VFORM)=c BARE  (^TENSE)=PAST.
2arrabt     I  *  @ControlI  (^PRED)='BE_NEAR<(^XCOMP)>(^SUBJ)'     @1SG   (^XCOMP VFORM)=c BARE  (^TENSE)=PAST.
l7e2t       I  *  @ControlI  (^PRED)='CATCH<(^XCOMP)>(^SUBJ)'       @1SG   (^XCOMP VFORM)=c BARE  (^TENSE)=PAST.
mal7e2tesh  I  *  @ControlI  (^PRED)='NOT_CATCH<(^XCOMP)>(^SUBJ)'   @1SG   (^XCOMP VFORM)=c BARE  (^TENSE)=PAST.
2il7a2      I  *  @ControlI  (^PRED)='CATCH<(^XCOMP)>(^SUBJ)'       @2SGM  (^XCOMP VFORM)=c IMP   (^VFORM)=IMP.

battalit    I  *  @ControlI  (^PRED)='STOP<(^XCOMP)>(^SUBJ)'        @3SGF  (^XCOMP VFORM)=c BARE  (^TENSE)=PAST.
battal      I  *  @ControlI  (^PRED)='STOP<(^XCOMP)>(^SUBJ)'        @3SGM  (^XCOMP VFORM)=c BARE  (^TENSE)=PAST.
battalu     I  *  @ControlI  (^PRED)='STOP<(^XCOMP)>(^SUBJ)'        @2PL   (^XCOMP VFORM)=c BARE  (^VFORM)=IMP.
battalt     I  *  @ControlI  (^PRED)='STOP<(^XCOMP)>(^SUBJ)'        @1SG   (^XCOMP VFORM)=c BARE  (^TENSE)=PAST.

yimken      I  *  @ControlI  (^PRED)='MAY<(^XCOMP)>(^SUBJ)'         (^VTYPE)=MODAL;
            V  *  @ControlI  (^PRED)='MAY<(^XCOMP)>(^SUBJ)'         (^VTYPE)=MODAL.
laazim      I  *  @ControlI  (^PRED)='MUST<(^XCOMP)>(^SUBJ)'        (^XCOMP VFORM)=c BARE  (^VTYPE)=MODAL;
```

	V	*	@ControlI	(^PRED)='MUST<(^XCOMP)>(^SUBJ)'		(^XCOMP VFORM)=c BARE	(^VTYPE)=MODAL.
momken	I	*	@ControlI	(^PRED)='POSSIBLE<(^XCOMP)>(^SUBJ)'		(^XCOMP VFORM)=c BARE	(^VTYPE)=MODAL;
	V	*	@ControlI	(^PRED)='POSSIBLE<(^XCOMP)>(^SUBJ)'		(^XCOMP VFORM)=c BARE	(^VTYPE)=MODAL.
yinfa3	I	*	@ControlI	(^PRED)='CAN<(^XCOMP)>(^SUBJ)'		(^XCOMP VFORM)=c BARE	(^VTYPE)=MODAL.
2elmafruud	I	*	@ControlI	(^PRED)='OUGHT_TO<(^XCOMP)>(^SUBJ)'		(^XCOMP VFORM)=c BARE	(^VTYPE)=MODAL ;
	I	*		(^PRED)='OUGHT_TO<(^SUBJ)(^COMP)>'			(^VTYPE)=MODAL.
3ayza	I	*	@ControlICP	(^PRED)='WANT<(^SUBJ)(^COMP)>'	@SGF	(^COMP VFORM)=c BARE	(^VTYPE)=MODAL.
3aayez	I	*	@ControlICP	(^PRED)='WANT<(^SUBJ)(^COMP)>'	@SGM	(^COMP VFORM)=c BARE	(^VTYPE)=MODAL.
3aawez	I	*	@ControlICP	(^PRED)='WANT<(^SUBJ)(^COMP)>'	@SGM	(^COMP VFORM)=c BARE	(^VTYPE)=MODAL.
mosammem	I	*	@ControlICP	(^PRED)='INSIST<(^SUBJ)(^COMP)>'	@SGM	(^COMP VFORM)=c BARE	(^VTYPE)=MODAL.
mosammema	I	*	@ControlICP	(^PRED)='INSIST<(^SUBJ)(^COMP)>'	@SGF	(^COMP VFORM)=c BARE	(^VTYPE)=MODAL.
nefsaha	I	*	@ControlICP	(^PRED)='WISH<(^SUBJ)(^COMP)>'	@3SGF	(^COMP VFORM)=c BARE	(^VTYPE)=MODAL.
nefsi	I	*	@ControlICP	(^PRED)='WISH<(^SUBJ)(^COMP)>'	@1SG	(^COMP VFORM)=c BARE	(^VTYPE)=MODAL.
naawi	I	*	@ControlICP	(^PRED)='INTEND<(^SUBJ)(^COMP)>'	@SGM	(^COMP VFORM)=c BARE	(^VTYPE)=MODAL.
nawya	I	*	@ControlICP	(^PRED)='INTEND<(^SUBJ)(^COMP)>'	@SGF	(^COMP VFORM)=c BARE	(^VTYPE)=MODAL.
naweet	I	*	@ControlICP	(^PRED)='INTEND<(^SUBJ)(^COMP)>'	@1SG	(^COMP VFORM)=c BARE	(^VTYPE)=MODAL (^TENSE)=PAS...
ti2dar	I	*	@ControlICP	(^PRED)='ABLE<(^SUBJ)(^COMP)>'	@2SGM	(^COMP VFORM)=c BARE	(^VTYPE)=MODAL.
yi2daru	I	*	@ControlICP	(^PRED)='ABLE<(^SUBJ)(^COMP)>'	@3PL	(^COMP VFORM)=c BARE	(^VTYPE)=MODAL.
7a2dar	I	*	@ControlICP	(^PRED)='ABLE<(^SUBJ)(^COMP)>'	@1SG	(^COMP VFORM)=c BARE	(^VTYPE)=MODAL (^TENSE)=FUT...

```
biti2dar   I  *  @ControlICP (^PRED)='ABLE<(^SUBJ)(^COMP)>'
                 @3SGF (^COMP VFORM)=c BARE
                 (^VTYPE)=MODAL (^TENSE)=PRESENT (^HAB)=+ (^PROG)=+.

biyel3ab   V  *  (^PRED)='PLAY<(^SUBJ)>'    @3SGM   (^ASPECT)=IMPF      (^HAB)=+ (^PROG)=+ ;
           I  *  (^PRED)='PLAY<(^SUBJ)>'    @3SGM   (^TENSE)=PRESENT    (^HAB)=+ (^PROG)=+.

bnel3ab    V  *  (^PRED)='PLAY<(^SUBJ)>'    @1PL    (^ASPECT)=IMPF      (^HAB)=+ (^PROG)=+ ;
           I  *  (^PRED)='PLAY<(^SUBJ)>'    @1PL    (^TENSE)=PRESENT    (^HAB)=+ (^PROG)=+.

btet3allim V  *  (^PRED)='LEARN<(^SUBJ)>'   @2SGM   (^ASPECT)=IMPF      (^HAB)=+ (^PROG)=+ ;
           I  *  (^PRED)='LEARN<(^SUBJ)>'   @2SGM   (^TENSE)=PRESENT    (^HAB)=+ (^PROG)=+.

betghanni  V  *  (^PRED)='SING<(^SUBJ)>'    @3SGF   (^ASPECT)=IMPF      (^HAB)=+ (^PROG)=+ ;
           I  *  (^PRED)='SING<(^SUBJ)>'    @3SGF   (^TENSE)=PRESENT    (^HAB)=+ (^PROG)=+.

katab      V  *  (^PRED)='WRITE<(^SUBJ)(^OBJ)>'        @3SGM   (^ASPECT)=PERF ;
           I  *  (^PRED)='WRITE<(^SUBJ)(^OBJ)>'        @3SGM   (^TENSE)=PAST.

fehemt     V  *  (^PRED)='UNDERSTAND<(^SUBJ)(^OBJ)>'   @1SG    (^ASPECT)=PERF ;
           I  *  (^PRED)='UNDERSTAND<(^SUBJ)>(^OBJ)'   @1SG    (^TENSE)=PAST.
```

```
3amal
          V    *    (^PRED)='DO<(^SUBJ)(^OBJ)>'        @3SGM    (^ASPECT)=PERF  ;
          I    *    (^PRED)='DO<(^SUBJ)(^OBJ)>'        @3SGM    (^TENSE)=PAST.

bixattat
          V    *    (^PRED)='PLAN<(^SUBJ)>(^OBJ)'      @3SGM    (^ASPECT)=IMPF     (^HAB)=+  (^PROG)=+;
          I    *    (^PRED)='PLAN<(^SUBJ)>(^OBJ)'      @3SGM    (^TENSE)=PRESENT   (^HAB)=+  (^PROG)=+.

biyetzaa3
          V    *    (^PRED)='BROADCAST<(^SUBJ)>>'      @3SGM    (^ASPECT)=IMPF     (^HAB)=+  (^PROG)=+;
          I    *    (^PRED)='BROADCAST<(^SUBJ)>>'      @3SGM    (^TENSE)=PRESENT   (^HAB)=+  (^PROG)=+.

bi7ebbeni
          V    *    (^PRED)='LOVE<(^SUBJ)>(^OBJ)'      @3SGM @OBJ @1SGOBJ    (^ASPECT)=IMPF     (^HAB)=+;
          I    *    (^PRED)='LOVE<(^SUBJ)>(^OBJ)'      @3SGM @OBJ @1SGOBJ    (^TENSE)=PRESENT   (^HAB)=+.

2a3od
          V    *    (^PRED)='SIT<(^SUBJ)>>'            @1SG     (^ASPECT)=IMPF  (^VFORM)=BARE.

7anmuut
          V    *    (^PRED)='DIE<(^SUBJ)>>'            @1PL     (^ASPECT)=PROSP;
          I    *    (^PRED)='DIE<(^SUBJ)>>'            @1PL     (^TENSE)=FUTURE.

darabna
          V    *    (^PRED)='HIT<(^SUBJ)(^OBJ)>>'      @1PL     (^ASPECT)=PERF;
          I    *    (^PRED)='HIT<(^SUBJ)(^OBJ)>>'      @1PL     (^TENSE)=PAST.
```

```
ye2fel       V   *   (^PRED)='CLOSE<(^SUBJ)(^OBJ)>'           @3SGM  (^VFORM)=BARE.
ynaam        V   *   (^PRED)='SLEEP<(^SUBJ)>'                 @3SGM  (^VFORM)=BARE.
yeb3ed       V   *   (^PRED)='GET_AWAY<(^SUBJ)>'             @3SGM  (^VFORM)=BARE.
teshtem      V   *   (^PRED)='CURSE<(^SUBJ)>(^OBJ)>'         @2SGM  (^VFORM)=BARE.
t7ess        V   *   (^PRED)='FEEL<(^SUBJ)>'                 @3SGF  (^VFORM)=BARE.
t3ayyat      V   *   (^PRED)='CRY<(^SUBJ)>'                  @3SGF  (^VFORM)=BARE.
teghanni     V   *   (^PRED)='SING<(^SUBJ)>'                 @2SGM  (^VFORM)=BARE.
tedfa3       V   *   (^PRED)='PAY<(^SUBJ)>(^OBJ)>'           @2SGM  (^VFORM)=BARE.
2asta7mel    V   *   (^PRED)='TOELRATE<(^SUBJ)>(^OBJ)>'      @1SG   (^VFORM)=BARE.
2a7terem     V   *   (^PRED)='RESPECT<(^SUBJ)(^OBJ)>'        @1SG   (^VFORM)=BARE.
yedrabek     V   *   (^PRED)='BEAT<(^SUBJ)>(^OBJ)>'          @3SGM @2SGFOBJ  (^VFORM)=BARE.

ba2eet       V   *   (^PRED)='BECOME<(^SUBJ)>(^OBJ)>'        @1SG   (^ASPECT)=PERF.

benet2axxar  V   *   (^PRED)='BE_LATE<(^SUBJ)>'             @1PL   (^ASPECT)=IMPF  (^HAB)=+.
biyemshi     V   *   (^PRED)='WALK<(^SUBJ)>'                @3SGM  (^ASPECT)=IMPF  (^HAB)=+  (^PROG)=+  .
bitetkallif  V   *   (^PRED)='COST<(^SUBJ)(^OBJ)>'         @3SGF  (^ASPECT)=IMPF  (^HAB)=+  (^PROG)=+  .
2ishtaghalt  V   *   (^PRED)='WORK<(^SUBJ)>'               @2SGM  (^ASPECT)=PERF.
```

```
2ishtaghaltu V  * (^PRED)='WORK<(^SUBJ)>'          @2PL   (^ASPECT)=PERF.
2itxane2na   V  * (^PRED)='FIGHT<(^SUBJ)>'         @1PL   (^ASPECT)=PERF.
2ishtaghalu  V  * (^PRED)='WORK<(^SUBJ)>'          @2PL   (^VFORM)=IMP.
2ishtaghal   V  * (^PRED)='WORK<(^SUBJ)>'          @2SGM  (^VFORM)=IMP.
matishtaghalsh V * (^PRED)='NOT_WORK<(^SUBJ)>'     @2SGM  (^VFORM)=IMP.
ruu7         V  * (^PRED)='GO<(^SUBJ)>'            @2SGM  (^VFORM)=IMP.
ruu7u        V  * (^PRED)='GO<(^SUBJ)>'            @2PL   (^VFORM)=IMP.
tishtaghal   V  * (^PRED)='WORK<(^SUBJ)>'          @2SGM  (^VFORM)=BARE.
nitmashsha   V  * (^PRED)='WALK<(^SUBJ)>'          @1PL   (^VFORM)=BARE.
yigri        V  * (^PRED)='RUN<(^SUBJ)>'           @3SGM  (^VFORM)=BARE.
yitmarran    V  * (^PRED)='TRAIN<(^SUBJ)>'         @3SGM  (^VFORM)=BARE.
yisalli      V  * (^PRED)='PRAY<(^SUBJ)>'          @3SGM  (^VFORM)=BARE.
2aakul       V  * (^PRED)='EAT<(^SUBJ)>(^OBJ)>'    @1SG   (^VFORM)=BARE.
yit7arrak    V  * (^PRED)='MOVE<(^SUBJ)>'          @3SGM  (^VFORM)=BARE.
yiktib       V  * (^PRED)='WRITE<(^SUBJ)>(^OBJ)>'  @3SGM  (^VFORM)=BARE.
yitkitib     V  * (^PRED)='BE_WRITTEN<(^SUBJ)>'    @3SGM  (^VFORM)=BARE.
yiktishif    V  * (^PRED)='DISCOVER<(^SUBJ)(^OBJ)>' @3SGM (^VFORM)=BARE.
2amuut       V  * (^PRED)='DIE<(^SUBJ)>'           @1SG   (^VFORM)=BARE.
2ashrab      V  * (^PRED)='DRINK<(^SUBJ)>'         @1SG   (^VFORM)=BARE.
mashrabsh    V  * (^PRED)='NOT_DRINK<(^SUBJ)>'     @1SG   (^VFORM)=BARE.
```

```
2azaakir   V  *  (^PRED)='STUDY<(^SUBJ)>'              @1SG   (^VFORM)=BARE.
laabis     V  *  (^PRED)='WEAR<(^SUBJ)>(^OBJ)'         @3SGM  (^VFORM)=PTCP.
faahim     V  *  (^PRED)='UNDERSTAND<(^SUBJ)>'         @3SGM  (^VFORM)=PTCP.
2a3diin    V  *  (^PRED)='SIT<(^SUBJ)>'                @1PL   (^VFORM)=PTCP.

ti3azzib   V  *  (^PRED)='TORTURE<(^SUBJ)(^OBJ)>'      @3SGF  (^VFORM)=BARE.
yi3mil     V  *  (^PRED)='DO<(^SUBJ)(^OBJ)>'           @3SGM  (^VFORM)=BARE.
2al3ab     V  *  (^PRED)='PLAY<(^SUBJ)>(^OBJ)'         @1SG   (^VFORM)=BARE.
tifakkaru  V  *  (^PRED)='THINK<(^SUBJ)>'              @2PL   (^VFORM)=BARE.

yilaa2i    V  *  (^PRED)='FIND<(^SUBJ)(^OBJ)>'         @3SGM  (^VFORM)=BARE.
tiigi      V  *  (^PRED)='COME<(^SUBJ)>'               @3SGF  (^VFORM)=BARE.
yiigi      V  *  (^PRED)='COME<(^SUBJ)>'               @3SGM  (^VFORM)=BARE.
tib2a      V  *  (^PRED)='BECOME<(^SUBJ)(^OBJ)>'       @2SGM  (^VFORM)=BARE.
yit7ebes   V  *  (^PRED)='IMPRISONED<(^SUBJ)>'         @3SGM  (^VFORM)=BARE.
yit3emel   V  *  (^PRED)='DONE<(^SUBJ)>'               @3SGM  (^VFORM)=BARE.
to23odi    V  *  (^PRED)='SIT<(^SUBJ)>'                @2SGF  (^VFORM)=BARE.
nettefe2   V  *  (^PRED)='AGREE<(^SUBJ)>'              @1PL   (^VFORM)=BARE.
yinsaani   V  *  (^PRED)='FORGET<(^SUBJ)>(^OBJ)'       @3SGM  @OBJ @1SGOBJ  (^VFORM)=BARE.
```

```
mayefre2sh  I  *  (^PRED)='NOT_DIFFERENT<(^SUBJ)>'      @3SGM              (^VFORM)=BARE.

2ashuuf     V  *  (^PRED)='SEE<(^SUBJ)(^OBJ)>'          @1SG               (^VFORM)=BARE.

2axtefi     V  *  (^PRED)='DISSAPPEAR<(^SUBJ)>'         @1SG               (^VFORM)=BARE.

tib3at      V  *  (^PRED)='SEND<(^SUBJ)(^OBJ)>'         @2SGM              (^VFORM)=BARE.

te3mi       V  *  (^PRED)='BLIND<(^SUBJ)(^OBJ)>'        @2SGM              (^VFORM)=BARE.

tit7ammel   V  *  (^PRED)='TOLERATE<(^SUBJ)(^OBJ)>'     @3SGF              (^VFORM)=BARE.

2aruu7      V  *  (^PRED)='GO<(^SUBJ)(^OBJ)>'           @1SG               (^VFORM)=BARE.

2atkallem   V  *  (^PRED)='TALK<(^SUBJ)>'               @1SG               (^VFORM)=BARE.

tsaafer     V  *  (^PRED)='TRAVEL<(^SUBJ)(^OBJ)>'       @3SGF              (^VFORM)=BARE.

2asaafer    V  *  (^PRED)='TRAVEL<(^SUBJ)(^OBJ)>'       @1SG               (^VFORM)=BARE.

yisrifu     V  *  (^PRED)='PAY_EXPENSES<(^SUBJ)>'       @3PL               (^VFORM)=BARE.

tefta7i     V  *  (^PRED)='OPEN<(^SUBJ)(^OBJ)>'         @2SGF              (^VFORM)=BARE.

fata7ti     I  *  (^PRED)='OPEN<(^SUBJ)(^OBJ)>'         @2SGF              (^TENSE)=PAST.

2as2alak    V  *  (^PRED)='ASK<(^SUBJ)>(^OBJ)>'         @1SG  @OBJ @2SGMOBJ  (^VFORM)=BARE.

2a2ollak    V  *  (^PRED)='TELL<(^SUBJ)>(^OBJ)>'        @1SG  @OBJ @2SGMOBJ  (^VFORM)=BARE.

2ashuufak   V  *  (^PRED)='SEE<(^SUBJ)>(^OBJ)>'         @1SG  @OBJ @2SGMOBJ  (^VFORM)=BARE.
```

```
fi         P   *   (^PRED)='in<(^OBJ)>'.

wara       P   *   (^PRED)='after<(^OBJ)>'.

men        P   *   (^PRED)='from<(^OBJ)>'.

zay        P   *   (^PRED)='like<(^OBJ)>'.

kol        P   *   (^PRED)='every<(^OBJ)>'.

le7ad      P   *   (^PRED)='until<(^OBJ)>'.

3annek     P   *   (^PRED)='from_you<(^OBJ)>'.

3anha      P   *   (^PRED)='about_her<(^OBJ)>'.

3an        P   *   (^PRED)='from<(^OBJ)>'.

ba3d       P   *   (^PRED)='after<(^OBJ)>'.

ma3        P   *   (^PRED)='with<(^OBJ)>'.

3ala       P   *   (^PRED)='on<(^OBJ)>'.

3aleeki    P   *   (^PRED)='on_you<(^OBJ)>'.

2ablaha    P   *   (^PRED)='before_it<(^OBJ)>'.


taani      ADV *   (^PRED)='again'.

shiwayya   ADV *   (^PRED)='a_bit'.

dayman     ADV *   (^PRED)='always'.

2imbaari7  ADV *   (^PRED)='yesterday'.
```

```
2innaharda    ADV * (^PRED)='today'.

bokra         ADV * (^PRED)='tomorrow'.

tuul_3omri    ADV * (^PRED)='all_my_life'.

wa7da_wa7da   ADV * (^PRED)='gradually'.

fag2a         ADV * (^PRED)='suddenly'.

kitiir        ADV * (^PRED)='too_much'.

keda          ADV * (^PRED)='like_this'.

kowayyes      ADV * (^PRED)='properly'.

xalaas        ADV * (^PRED)='finally'.

wa2taha       ADV * (^PRED)='by_that_time'.

men_gher_ma_yo2sod    ADV * (^PRED)='unintentionally'.

tuul_elnahaar         ADV * (^PRED)='all_day'.


----
```

Appendix B

Test File

1

7assan xalla sondos t7ess

2

xalliiha t7ess

3

elshakk xalla goozek yeb3ed 3annek

4

2enta xalleet sondos t3ayyat

5

homma xallo elwalad ynaam

6

ele7teraaf bixalliik teghanni kowayyes

7

7atxalliini 2asta7mel elmozakra

8

7atxalli elraagel yedrabek

9

2enti 7atxalli elraagel yedrabek

10

2enti 7atxalliih yedrabek

11

xalli saami yeb3ed 3an elsiyaasa

12

xallaani 2a7terem nafsi

13

xallaani men_gher_ma_yo2sod 2a7terem nafsi

14

2usluubu xallaani ba2eet mo7taref

15

xalliiha t7ess 2innak btet3allim

16

mark xalla 2inn ma7addesh ye2fel elfacebook

17

ba2a biyemshi zay elbatriiq

18

rigi3na 2itxane2na men youmeen

19

rigi3 yisalli fi elfasl

20

2a3adna nitmashsha shiwayya

21

fidil yigri wara elnaas

22

fidilna 2a3diin le7ad billeil

23

ma7addesh ba2a faahim

24

2irga3 2ishtaghal taani

25

2irga3 matishtaghalsh taani

26

matirga3sh tishtaghal taani

27

ba2a biyemshi zay elbatriiq

28

7afdal tuul_3omri 2aakul kintaki

29

ba2eena dayman benet2axxar

30

fawwaaz ba2a dayman laabis elgalabeyya

31

2alb_sa3d bada2 yit7arrak

32

2ana 2arrabt 2amuut men elxoof

33

2ana mal7e2tesh 2ashrab

34

bada2 yiktib 2imeel

35

bada2t 2imbaari7 2azaakir le7ad billeil

36

elshorta battalit ti3azzib elnaas

37

battalt 2al3ab baskit

38

battalu tifakkaru kitiir

39

yimken yilaa2i mayya

40

mona yimken tiigi 2innaharda

41

laazim tib2a mo7taref

42

laazim yit7ebes fi 2afas

43

momken yinsaani ba3d fatra

44

yinfa3 yit3emel fi sana

45

2enti 2elmafruud to23odi ma3 elnaas

46

2e7na momken nettefe2 fi no2ta

47

2elmafruud 2inn diini mayefre2sh

48

2a7mad laazim yiigi 2innaharda

49

3ayza 2ashuuf baaba

50

naawi yi3mil 7aaga

51

ti2dar tib3at rasaayel

52

biti2dar tit7ammel elmas2uleyya

53

7a2dar 2aruu7 elgam3a bokra

54

3ayza 2as2alak 3anha

55

2ana xalaas naweet 2atkallem

56

2enta mosammem te3mi 3eenak

57

2enti mosammema tefta7i elmawduu3

58

mona nefsaha tsaafer masr

59

2ana 3aawez 2a2ollak 3ala 7aaga

60

nefsi 2ashuufak fi masr

61

2ahlek yi2daru yisrifu 3aleeki

62

2ana nawya 2axtefi kitiir

63

fawwaaz biyel3ab ma3 2awlaadu

64

sondos betghanni

65

fehemt elserr

66

kaan 3amal film 2ablaha

67

2a7mad kaan bixattat huguum 3ala el7uduud

68

2elbarnaameg kaan biyetzaa3 3ala elMBC

69

howwa kaan bi7ebbeni

70

kont 2a3od fi elmat3am tuul_elnahaar

71

konna 7anmuut

72

wa2taha 7ankuun darabna 2amriika

73

7ankuun 2innaharda bnel3ab fi ma3had_guuta

74

2a7mad momken ykuun fi masr

75

kont bada2t 2azaakir

76

bada2t 2azaakir

77

sondos kaanit betghanni

78

2elmafruud 2a7terem 7assan

79

2elmafruud 2inn 2ana 2a7terem 7assan

80

kont laazim 2a3od

81

nefsi 2asaafer 2amriika