

University of Alberta

**A corpus study of basic motion verbs
In Modern Standard Arabic**

by

Dana Abdulrahim

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics

©Dana Abdulrahim

Fall 2013

Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

To mom, dad, Sally, and John

Abstract

In this dissertation, I present a corpus-based, constructionist account of Modern Standard Arabic (MSA) GO verbs (*ḍahaba*, *maḍā*, and *rāḥa*) and COME verbs (*atā*, *ḥaḍara*, *ḡā'a*, and *qadima*). These seven *deictic* motion verbs count among the most frequent lexical items in MSA, nevertheless, they are poorly described in contemporary dictionaries, which renders the task of differentiating the use of each verb rather daunting for the learner and the linguist alike.

This study offers a comprehensive and empirically grounded treatment of these verbs in that it closely examines the contextual features that typically associate with each verb in actual usage as compiled in a corpus. Such contextual features include inflectional marking on the verb, syntactic frames hosting each verb, semantic properties of collocating lexical items, as well as the overall properties of the motion event construal.

The quantitative and qualitative analyses I offer in this study rely on annotating a large amount of corpus returns per each verb for a wide range of morphosyntactic and semantic features. These data frames are subjected to selected monivariate and multivariate analyses as a means of identifying exemplary uses per motion verb. A subsequent set of qualitative analyses elaborates on the general statistical findings by scrutinizing individual instances of verb use and examining collocational patterns and less frequent verb uses.

The results obtained from both the quantitative and qualitative analyses highlight the idiosyncratic constructional properties that characterize the use of each verb in various physical and figurative motion event construals. As such, I argue that the availability of multiple GO and COME verbs in MSA is not a reflection of extravagance in the lexical system of MSA. Rather, each verb appears to highlight a particular aspect of the conceptually complex *deictic* motion event.

Finally, I propose a sample of a corpus-based dictionary entry that is more representative of contemporary language usage than entries currently available. I also discuss different directions for future research. This standardized treatment of the lexical and grammatical patterns pertaining to MSA GO and COME verbs highlights the contributions of Arabic to cross-linguistic research on the phenomenon of motion verbs.

Acknowledgments

I am greatly indebted, first and foremost, to my supervisors, Sally Rice and John Newman, for their contributions to all aspects of this project and for their unfailing support throughout the six years of my graduate studies at the University of Alberta. They have been great mentors and educators and have been a source of inspiration and motivation right from the beginning. They taught me what research in Linguistics is all about, and they tirelessly helped me refine my convictions about the study of human languages. John Newman and Sally Rice have always believed in my abilities and challenged me to go that one extra mile. Their exemplary work ethic, both in academic and administrative matters, has inspired me to be resilient, to keep going, and to never compromise the standards that I set for myself. I will always have fond memories of our supervisory meetings that never lacked in knowledge, humor, and parental guidance. I am very honored to have had the opportunity to learn from them in so many ways.

I am also greatly thankful to Dr. Antti Arppe for his help with various aspects of the statistical analysis and for helping me understand a concept as complex as polytomous logistic regression analysis. I would also like to thank Professor Dilworth Parkinson, from Brigham Young University, for introducing ArabiCorpus to me and for helping me tremendously throughout various stages of data collection. Special thanks also go to Professor Eric Atwell from Leeds University, for his valuable feedback and suggestions during and after my dissertation defense exam. A big thank you goes to my colleagues and friends at the Arabic Linguistics Society for supporting my research and for the insightful discussions we had at the various Arabic linguistic Symposia, as well as on the Arabic-List forum.

I gratefully acknowledge the funding sources that made my Ph.D. research in Canada possible. The University of Bahrain has provided substantial funding for five

years. The Department of Linguistics at the University of Alberta has also provided many funding opportunities through teaching and research assistantships as well as teaching opportunities. Special thanks also go to the Department of Modern Languages and Cultural Studies, Faculty of Graduate Studies and Research, and the Graduate Students Association at the University of Alberta for additional funding opportunities.

A large part of my life at the University of Alberta involved having to deal with non-academic issues that, at times, proved to be more challenging than writing a dissertation. Elizabeth French, Diane Mcken, Oliver Rossier, Jana Tomasovic, and Lex Bos have been the most pleasant, supportive and patient administrative assistants. They have always taken care of me and watched my back so that I be able to focus on my research. Their genuine smiles and their goodness of heart are what makes me call the Department of Linguistics ‘my home’.

My friends and colleagues at the Department of Linguistics have contributed immensely to my academic growth. I am grateful to the numerous discussions I had with the phonologists, the typologists, the child language researchers, the historical linguists, the psycholinguists, and the field researchers. Our discussions were, at times, challenging and, other times, a source of relief and a confirmation that we are all working towards the same goal. I am thankful to this wonderful community of friends that I have bonded with over the years, and whose support, within and outside of academia, has kept me on my feet during rough times.

I am grateful to my friends and loved ones in Bahrain, the US, and Canada for always staying in touch and sending me their best wishes. Special thanks go to Joshua Lee and Fatma Bastaki, who have always patiently listened to my complaints about being a Ph.D. student, and to my rants about Linguistics.

Last but not least, I want to thank my family – mom and dad, Hala, Mohammed, Khalid, Yasser, Dunya, Maryam, and my nephews and nieces – for their patience and

constant support all these years. My parents have always been proud of me no matter what unholy adventures I put myself through, and it is their trust and unconditional love that made me who I am today.

Table of Contents

1. Introduction.....	1
1.1 Objectives of this study	1
1.2 The morphology of the Arabic verb	8
1.3 Problems with traditional accounts of the verb in Arabic	10
1.4 Corpus-based, multifactorial approaches to the study of (near-)synonymy	15
1.5 Basic verbs and GO and COME	18
1.6 Modern Standard Arabic GO and COME verbs.....	25
1.5.1 GO verbs in MSA	25
1.5.2 COME verbs in MSA.....	26
1.6 The structure of this study	27
2. Methods.....	31
2.1 Selection of the MSA GO and COME verbs	31
2.1.1 Interchangeability of the motion verbs	33
2.1.2 Morphological simplicity.....	34
2.1.3 High frequency	35
2.1.4 Basic deictic motion events	36
2.2 Data collection.....	37
2.2.1 The corpus	37
2.2.2 Corpus search	38
2.2.3 Forms not included in the corpus search	41
2.2.3.1 Verbal nouns and active participles	41
2.2.3.2 Imperative COME: <i>ta'āla</i>	44
2.3 Constructing data frames for GO and COME verbs	45
2.3.1 Selection of contextual features and the annotation of corpus hits	45
2.4 Statistical analyses	48
2.4.1 Chi-square tests.....	50
2.4.1.1 Chi-square goodness-of-fit	50
2.4.1.2 Chi-square test for independence.....	52
2.4.1.3 Standardized Pearson's residuals.....	53
2.4.2 Hierarchical Agglomerative Cluster Analysis	55
2.4.3 (Hierarchical) Configural Frequency Analysis	58
2.4.4 Polytomous Logistic Regression Analysis.....	61
3. Quantitative analysis of MSA GO verbs <i>ḍahaba</i>, <i>maḍā</i>, and <i>rāḥa</i>.....	65
3.1 Single variable distribution.....	66
3.2 Hierarchical Agglomerative Cluster Analysis	72
3.3 Hierarchical Configural Frequency Analysis.....	74
3.4 Polytomous Logistic Regression Analysis.....	86
3.4.1 Bivariate analysis.....	87
3.4.2 Multivariate analysis.....	90
4. Qualitative analysis of MSA GO verbs <i>ḍahaba</i>, <i>maḍā</i>, and <i>rāḥa</i>	103
4.1 <i>Ḍahaba</i>	104
4.1.1 Physical motion	104
4.1.2 Non-physical motion	112
4.2 <i>Maḍā</i>	119
4.2.1 Physical motion	119
4.2.2 Non-physical motion	121
4.3 <i>Rāḥa</i>	131
4.3.1 Physical and non-physical motion	131
4.3.2 Grammaticalized function	134
4.4 Summary.....	137

5. Quantitative analysis of MSA COME verbs <i>atā</i>, <i>ḥaḍara</i>, <i>ǧā'a</i>, and <i>qadima</i>	140
5.1 Single variable distribution.....	140
5.2 Hierarchical Agglomerative Cluster Analysis	145
5.3 Hierarchical Configural Frequency Analysis.....	147
5.4 Polytomous Logistic Regression Analysis.....	160
5.4.1 Bivariate analysis.....	161
5.4.2 Multivariate analysis.....	162
6. Qualitative analysis of MSA COME verbs <i>atā</i>, <i>ḥaḍara</i>, <i>ǧā'a</i>, and <i>qadima</i>	175
6.1 <i>Atā</i> and <i>ǧā'a</i>	176
6.1.1 Physical motion	177
6.1.2 Non-physical motion	183
6.2 <i>Ḥaḍara</i>	196
6.2.1 Physical motion	197
6.2.2 Non-physical motion	200
6.3 <i>Qadima</i>	202
6.3.1 Physical motion	202
6.4 Summary.....	208
7. Synthesizing the corpus findings	211
7.1 Using corpus data	211
7.2 Using univariate and multivariate analyses	212
7.3 Using informed introspection	220
8. Implications, future research, and practical applications	225
8.1 Typological research on other languages with multiple GO and COME lexemes	225
8.2 Future research	229
8.2.1 Spoken varieties of Arabic.....	229
8.2.2 Non-finite forms related to the GO and COME verbs.....	232
8.2.3 Psycholinguistic experiments	236
8.3 Practical applications: Suggestions for usage-based dictionary entries	240
8.3.1 Corpus-illustrated dictionary entry of <i>atā</i>	243
8.3.2 Sub-sense frequency-based dictionary entry of <i>atā</i>	245
8.3.3 Corpus-based dictionary entry of <i>atā</i>	247
8.4 Conclusion.....	249
References	254

Appendices

Appendix A. Templatic verb forms in Arabic along with their general meaning associations	264
Appendix B. The full inflectional paradigm for a tri-consonantal root (KTB 'to write')	265
Appendix C. Selected monolingual and bilingual dictionary entries of MSA (and CA) GO and COME verbs	266
Appendix D. Examples for annotation per variable (and each level within every variable)	273
Appendix E. Samples of sentence annotation from the GO and COME data frames	278
Appendix F. R commands and results of standardized Pearson's residuals for GO and COME data frames	282
Appendix G. Sample of the entire hierarchical configural frequency analysis table	285

List of Tables

Chapter One

TABLE 1. Examples of derivations of different verbs from the same root, D-R-S, which pertains to learning	9
TABLE 2. Percentages of occurrence of inflected forms per MSA GO verb in the perfective aspect based on inspection of 300 concordance lines	11
TABLE 3. Dictionary entries vs. corpus-based sub-senses of the verb <i>rāḥa</i>	14

Chapter Two

TABLE 1. Sample dictionary entries for GO and COME in three bilingual English-MSA dictionaries.	31
TABLE 2. Ranking of the seven COME and GO verbs as listed in <i>A Frequency Dictionary of Arabic</i>	36
TABLE 3. Regular expressions used to query finite forms of the verb <i>dahaba</i>	41
TABLE 4. A selection of variables GO and COME corpus hits were coded for.	46
TABLE 5. Sample of annotation from the COME data frame for selected variables.	48
TABLE 6. Observed vs. expected frequencies of the levels of the variable TENSE.	51
TABLE 7. Observed frequencies of the levels of the variable TENSE per COME verb.	53
TABLE 8. Expected frequencies of the levels of the variable TENSE per COME verb.	53
TABLE 9. Standardized Pearson's residuals for the distribution of the variable TENSE per COME verb.	54
TABLE 10. Standardized Pearson's residuals for the distribution of the variable TENSE per COME verb (+/-/0).	55
TABLE 11. Sample of a co-occurrence table generated by the BP 1.01 script for the variable TENSE by COME verb.	57
TABLE 12. A sample output hierarchy table generated by the HCFA script	60

Chapter Three

TABLE 1. A list of the variables for which corpus sentences for the three MSA GO verbs were coded	66
TABLE 2. Proportional frequencies of the different levels of ASPECT by GO verb	67
TABLE 3. Proportional frequencies of the different levels of SUBJECT SEMANTIC CATEGORY by GO verb.	68
TABLE 4. Expected vs. observed values for the variable GOAL by GO verb.	70
TABLE 5. Standardized Pearson's residuals for the occurrence of GOAL by GO verb.	70
TABLE 6. Preferences for the distribution of selected logical variables by GO verb.	71
TABLE 7. Sample of co-occurrence table generated by the BP 1.01 script.	72
TABLE 8. Most significant univariate <i>types</i> for TENSE, ASPECT, MORPHOLOGICAL ASPECT AND MOOD, SUBJECT NUMBER, PERSON and GENDER for GO verbs.	75
TABLE 9. Most significant univariate <i>antitypes</i> for TENSE, ASPECT, MORPHOLOGICAL ASPECT AND MOOD, SUBJECT NUMBER, PERSON and GENDER for GO verbs.	76
TABLE 10. Most significant configurations for TENSE, ASPECT, and MORPHOLOGICAL ASPECT and MOOD for all GO verbs	78
TABLE 11. Most significant configurations for SUBJECT NUMBER, PERSON, GENDER, and SEMANTIC CATEGORY for all GO verbs.	80
TABLE 12. Most significant configurations for SUBJECT SEMANTIC CATEGORY and PHRASAL SEMANTIC CATEGORIES for all GO verbs	82
TABLE 13. Excerpt of the list of logical variables	87
TABLE 14. Cross-tabulation of raw frequencies of TENSE.PAST by MORPH_ASP.MOOD.PERF	88
TABLE 15. Bivariate analysis of TENSE.PAST by MORPH_ASP.MOOD.PERF	88
TABLE 16. Bivariate analysis of selected (independent) variables.	89
TABLE 17. A cross-tabulation of the predicted vs. observed verbs for all annotated contexts in the data frame (with the total of 500 contexts per GO verb)	93
TABLE 18. Contextual features that increase the odds in favor of or against the occurrence of GO verbs	95

Chapter Five

TABLE 1. Proportional frequencies of the different levels of TENSE by COME verb.....	140
TABLE 2. Proportional frequencies of the different levels of SUBJECT SEMANTIC CATEGORY by COME verb.	142
TABLE 3. Proportional frequencies of the occurrence of phrases denoting SOURCE and GOAL of motion by COME verb.	143
TABLE 4. Preferences for the distribution of selected logical variables per COME verb	145
TABLE 5. Sample of co-occurrence table generated by the BP 1.01 script.....	145
TABLE 6. Most significant univariate <i>types</i> for TENSE, ASPECT, MORPHOLOGICAL ASPECT AND MOOD, SUBJECT NUMBER, PERSON and GENDER.....	148
TABLE 7. Most significant univariate <i>antitypes</i> for TENSE, ASPECT, MORPHOLOGICAL ASPECT AND MOOD, SUBJECT NUMBER, PERSON and GENDER	148
TABLE 8. Most significant configurations for TENSE, ASPECT, and MORPHOLOGICAL ASPECT AND MOOD for all COME verbs.	150
TABLE 9. Most significant configurations for SUBJECT NUMBER, PERSON, GENDER, and SEMANTIC CATEGORY for all COME verbs.....	151
TABLE 10. SUBJECT SEMANTIC CATEGORY x SEMANTIC PROPOSITIONS configurations for the verbs <i>atā</i> and <i>ǧā'a</i>	154
TABLE 11. SUBJECT SEMANTIC CATEGORY x SEMANTIC PROPOSITIONS configurations for the verbs <i>ḥaḍara</i> and <i>qadima</i>	158
TABLE 12. Bivariate analysis of selected (independent) variables	161
TABLE 13. A cross-tabulation of the predicted vs. observed verbs for all annotated contexts in the data frame (with the total of 500 contexts per COME verb).....	164
TABLE 14. Contextual features that increase the odds in favor of or against the occurrence of COME verbs.....	166

Chapter Eight

TABLE 1. Verbal noun and active participle derivations of the seven motion verbs.....	234
---	-----

List of Figures

Chapter One

- FIGURE 1. Overlapping sub-senses/usages of *dahaba*, *maḍā*, and *rāḥa* based on inspection of 300 corpus hits per verb from ArabiCorpus and dictionary entries from Al-Munjid (monolingual) Dictionary and Al-Mawrid (Arabic-English) Dictionary7

Chapter Three

- FIGURE 1. Relative frequencies of the different levels of ASPECT by GO verb.....67
FIGURE 2. Relative frequencies of the different levels of SUBJECT SEMANTIC CATEGORY by GO verb69
FIGURE 3. Dendrogram based on the of GO multivariate data frame.73

Chapter Four

- FIGURE 1. Schematic representation of *maḍā* as opposed to *dahaba* 124

Chapter Five

- FIGURE 1. Relative frequencies of the different levels of TENSE by COME verb..... 140
FIGURE 2. Relative frequencies of the different levels of SUBJECT SEMANTIC CATEGORY by COME verb 142
FIGURE 3. Proportional frequencies of the occurrence of phrases denoting SOURCE and GOAL of motion by COME verb 144
FIGURE 4. Dendrogram based on the COME multivariate data frame. 147

Chapter Six

- FIGURE 1. Schematic representation of the ‘container’ spatial relations expressed by *atā* and *ḡā’a* in certain contexts of use 194

Chapter Eight

- FIGURE 1. Excerpt from a monolingual Arabic dictionary for the verb *qadima*.....235
FIGURE 2. Sample of corpus-illustrated dictionary entry of *atā*244
FIGURE 3. Sample of sub-sense frequency-based dictionary entry of *atā*246
FIGURE 4. Sample of corpus-based dictionary entry of *atā*247

List of Abbreviations

1	1 ST PERSON
2	2 ND PERSON
3	3 RD PERSON
ABL	ABLATIVE
ACC	ACCUSATIVE
ADJ	ADJECTIVE
ADV	ADVERB
ALL	ALLATIVE
AP	ACTIVE PARTICIPLE
ART	ARTICLE
AUX	AUXILIARY
CL	CLITIC
COM	COMITATIVE
COND	CONDITIONAL
CONJ	CONJUNCTION
DEM	DEMONSTRATIVE
DM	DISCOURSE MARKER
DUAL	DUAL
F	FEMININE
FUT	FUTURE
GEN	GENITIVE
IMPF	IMPERFECTIVE
IMRP	IMPERATIVE
INST	INSTRUMENTAL
INTENS	INTENSIFIER
JUSS	JUSSIVE
LOC	LOCATIVE
M	MASCULINE
MOD	MODAL
NEG	NEGATION
NOM	NOMINATIVE
PASS	PASSIVE
PERF	PERFECTIVE
PL	PLURAL
POSS	POSSESSIVE
PP	PERSONAL PRONOUN
PURP	PURPOSIVE
Q	QUESTION PARTICLE
RP	RELATIVE PRONOUN
SG	SINGULAR
SUBJN	SUBJUNCTIVE
TOP	TOPIC
VN	VERBAL NOUN
VOC	VOCATIVE

Chapter One Introduction

1.1 Objectives of this study

This dissertation aims to provide a construction-based analysis for a set of verbs of motion, those associated with the concepts GO and COME in Modern Standard Arabic (MSA). As such, this study is a departure from previous (overly morphological) research on the properties of the Arabic verb, as well as a departure from current methods commonly employed in the investigation of the syntactic and semantic behavior of basic verbs in any language.

The Arabic-speaking world today is characterized by a strong diglossic situation in which the written varieties, Modern Standard Arabic (MSA) and Classical Arabic (CA), are in sharp contrast with the spoken colloquial dialects on various levels (e.g. phonologically, morphosyntactically, lexically). While the typical Arabic speaker generally refers to the written literary form of Arabic as *al-luġa al-fuṣḥa* ('the most eloquent language'),¹ linguists make the distinction between two basic varieties of literal or written Arabic: CA and MSA. Classical Arabic is the term used to characterize the form of the Arabic language that was documented between the 6th and the 13th century, an era during which Arabic flourished as a poetic language (encompassing both pre-Islamic and Islamic literature) and as a sacralized language, with the Qur'an as the source for a subsequent wealth of religious studies and exegesis as well as theological and grammatical analyses (Ryding, 2005).

The special status of Arabic, reinforced by the belief that it is the *chosen language of the Qur'an* insured, to some extent, that Classical Arabic would remain the literary, if not the written language, of any Arabic-speaking nation. Around the 13th

¹ Throughout this dissertation, I will rely on the DIN 31635 system of transliteration of the Arabic alphabet, which was adopted in 1982.

century, however, the spoken variety of Arabic within the multiple regions of the Arab world gradually started to deviate from the classical variety (Blau, 1961; Versteegh, 1984; Fück, 1955; Miller, 1986; Bateson, 1967). These varieties were not (and are still not) written down, partly as a means of maintaining the higher status of the written form. The spoken varieties of Arabic have been evolving in their own ways and adapting to changes over time, but have never been considered as separate languages within the Arab world. Spoken Arabic is therefore characterized as being “much more flexible and mutable than the written language; it easily coins words, adapts and adopts foreign expressions, and propagates slang, thus producing and reflecting a rich, creative, and constantly changing range of innovation.” (Ryding, 2005:5)

With the different colloquial varieties of spoken Arabic constituting the speech varieties of different regions within the Arab world,² a written form that stems from Classical Arabic serves as the written norm of the modern Arab world. Since the late 18th century, Modern Standard Arabic has been the language of the media encompassing newspapers, journals, books, advertising, literature, as well as the principal medium for public speaking and radio/TV broadcasting (McLaughlin, 1972). It was the development of the periodical press in the Arab world during the 19th century, in particular, that required a heavy reformation and modernization of Arabic in order to accommodate the influx of concepts and notions imported from the West (Versteegh, 1997: 177). Despite the fact that MSA may appear to be an artificial, fabricated language, it is this very practical aspect of MSA as the language of the press that makes it “une langue vivante...qui correspond à un besoin vital”³ (Monteil 1960:28).

While MSA appears to reflect most of the properties of CA, with certain syntactic, stylistic, and mostly lexical differences very much in evidence, the mutated

² For example the Levant, the Arabian/Persian Gulf, the western Arabian peninsula, western north Africa, Egypt, and the Sudan (Ryding, 2005).

³ “A living language that corresponds to vital needs” [translation mine].

vernacular varieties are considerably different from the classical form of the language. The semi-fossilized MSA is more or less the same across the vast geographical region of the Arab world – a fact that contrasts with the situation of the spoken colloquial varieties, some of which (in extreme cases) may not be mutually intelligible; for example, the variety spoken by Bedouins of the deserts of Saudi Arabia vis-à-vis rural Moroccan Arabic. Versteegh (1997) points out, however, that certain lexical distinctions do exist among the Modern Standard Arabic forms used in different regions of the Arab world. The reason for such variation can be ascribed to the fact that the different Arab regions follow different traditions with respect to the creation of new vocabulary. Moreover, this variation could be a reflection of the different colonial histories of each region since the lexicalization of a term in one region may be inspired by the language of the colonizer (1997:184), rather than by cross-dialectal borrowing.

This ideologically motivated preference towards regarding the language of the Qur'an and MSA as one and the same 'higher' form of Arabic (Versteegh, 1997:183) may be the reason that a large number of Arabic dictionaries nowadays still insist on providing lexical entries that are heavily influenced by Classical Arabic dictionaries, sometimes to the extent of ignoring highly frequent uses and collocations in the modern written form (as I will illustrate shortly). In my view, despite the high status that MSA occupies in the Arab world, the tendency to equate it with a more archaic variety has resulted in a misrepresentation of lexical uses and meaning, as far as monolingual Arabic dictionaries are concerned. One of the main objectives of this dissertation is to address these conflicting and inadequate lexical treatments of modern lexicographic practices through a case study of GO and COME verbs as used in Modern Standard Arabic.

Another aim of this study is to introduce an innovative and constructional based treatment of the Arabic verb. The Arabic verb undoubtedly counts as one of the most highly studied aspects of the Arabic language. It has, rightfully, received much linguistic

attention as it represents a perfect example of the peculiarities of the Semitic morphological system, along with the fact that it reflects facets of complex derivational processes on numerous levels: morphological, syntactic, lexical, and semantic. Past and current research on the Arabic verb has mostly been formal in nature, with particular emphasis placed on the morphological and syntactic properties of the verb.

Unfortunately, little research has been dedicated to the semantics of the verb despite the fact that the current situation of Arabic renders the diachronic and synchronic differences in verbal semantics across the three varieties (CA, MSA, and the colloquial varieties) a goldmine for lexical semantic research.

This study represents a departure from the compartmentalized analytical approaches to describing the Arabic verb by adopting a construction-based approach that considers various aspects of language (morphology, syntax, semantics, lexicon, etc.) as equally responsible for defining the behavior of a linguistic item. In this study, I aim to provide a holistic description of two sets of near-synonymous verbs that are related to the concepts of GO and COME in Modern Standard Arabic through quantitative and qualitative analyses based on actual usage. Prior to discussing the rationale behind this particular choice of verbs, a brief note on the theoretical underpinnings of this study is necessary.

Constructionist approaches to language are closely tied to the field of cognitive linguistics, the framework I am adopting for this study of motion verbs in Arabic. The core tenets of constructionist theories of language claim that the basic unit of linguistic organization is a construction. According to Croft and Cruse (2004:257), a construction “consist[s] of pairings of form and meaning that are at least partially arbitrary”, where ‘meaning’ is referred to as the conventionalized function of a construction. This conventionalization of a construction’s meaning/function not only includes the literal meaning of an utterance, but also properties of the discourse situation in which an utterance occurs (e.g. use of spatial deictic terms, such as *here* or *there*, that signal a

reference point in a speech event) as well as the pragmatic implications of an utterance (e.g. use of a yes/no question to request information, as in *Do you have the time?*) (Croft and Cruse, 2004). The term ‘construction’, therefore, covers both the idiomatic portions of language, where the meaning of an utterance is not predictable from the component parts that make up the utterance (e.g. *raining cats and dogs*), as well as any combination of two or more morphemes that reflect general morphosyntactic structures and where the meaning of an utterance is fully predictable from its component parts (e.g. *I want to go*). This view of grammar postulates that “the interaction of syntax and lexicon is much wider and deeper than the associations of certain verbs with certain complements” (Bybee, 2010:77), and that a considerable part of our linguistic knowledge consists of conventionalized expressions, or constructions (Langacker, 1987).

According to any constructionist framework, therefore, the behavior of a lexical item is best understood in its context of use and not in isolation, an idea that stretches back decades (cf. Firth, 1957). The syntactic structures in which it appears, the morphological inflections associated with it, the other lexical elements that co-occur with it in a phrase, etc., all contribute to the composed or conventionalized meaning or function expressed by the linguistic item. Such an approach calls for moving beyond single semantic, morphological, or syntactic properties of a lexical item and scrutinizing the entire lexico-syntactic frame in which it occurs. Increasingly, this is done through examination of corpus data. The availability of corpora caters to this highly contextualized analytical approach since corpora provide a large amount of naturally occurring, contextualized uses that enhance the investigation of the behavior of lexical items or phrases in their natural discourse environment (as opposed to introspective and elicited utterances that may not reflect actual language usage at all). Moreover, corpora provide voluminous amounts of linguistic data that permit a quantitative treatment of the phenomenon under investigation. Corpus data can, therefore, in principle undergo various

statistical analyses that can provide more insight than intuition or casual introspection (or even careful inspection) of data is ever able to achieve.

Verbs related to the concepts of GO and COME in Arabic are an ideal focus for the synchronic constructionist description of near-synonymous lexical elements in MSA. GO and COME verbs are part of a set of basic verbs that correspond to fundamental human activities such as SIT, STAND, LIE, EAT, DRINK, SEE, and HEAR, and are, thus, universal concepts that are represented in all human languages (Newman, 2004). Cross-linguistically, verbs denoting GO/COME event schemas tend to be *polysemous* as well as *heterosemous*,⁴ exhibiting a wide range of sub-senses and grammatical functions. A great deal of research has been dedicated to the investigation of the properties as well as usages of this particular pair of verbs or verbal concepts cross-linguistically (Fillmore, 1966, 1969, 1970, 1971, 1972; Sinha, 1972; Clark, 1974; Gathercole, 1977, 1978; Rauh, 1981; Fleischman, 1982; Bender, 1990; Craig, 1991; Emanation, 1992; Di Meola, 1994; Wilkins and Hill, 1995; Radden, 1996; Goddard, 1997; Newman, 2000, 2004; Botne, 2005, 2006; to name but a few).

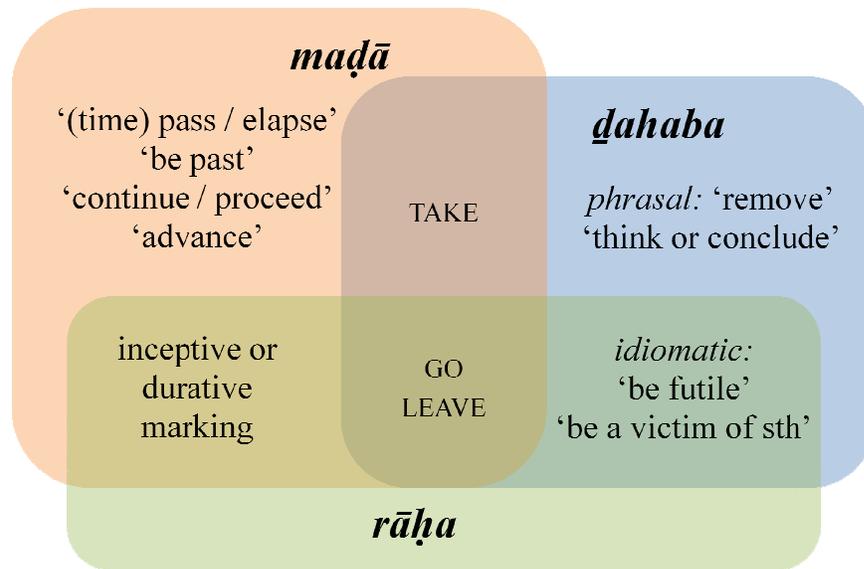
On a more language-specific level, the choice of verbs denoting GO and COME in Arabic is motivated by a number of observations. While the majority of spoken Arabic varieties have a single GO and a single COME verb, MSA (and CA, to some extent) have more than one verb associated with basic GO and COME motion event schemas.⁵ In MSA, the GO set of verbs consists of *dahaba*, *maḍā*, and *rāḥa*, while the COME set consists of *ǧā'a*, *atā*, *ḥaḍara*, and *qadima*, with the suppletive *ta'ala* serving as the imperative COME

⁴ Lichtenberk describes polysemy as “the association of distinct (but related) meanings with one and the same lexeme” (1991:476). He adopts Perrson’s (1988) term *heterosemy* to refer to “cases (within a single language) where two or more meanings or functions that are historically related, in the sense of deriving from the same ultimate source, are borne by reflexes of the common source element that belong in different morphosyntactic categories” (1991:476).

⁵ According to Talmy (1985, 2000, etc.) a motion event schema typically consists of the following components: FIGURE (i.e. the object that is considered moving or is located with respect to another object), MOTION, PATH (of the motion event), and GROUND (i.e. the object with respect to which the FIGURE moves or is located).

in Arabic. Each verb in these sets is associated with a particular cohort of meaning extensions and usages. A pilot study (Abdulrahim, ms.) combining both corpus-based and dictionary-based sub-senses of GO verbs showed that *ḍahaba*, *maḍā*, and *rāḥa* overlap to varying degrees in their usages, as shown in Figure 1.

FIGURE 1. Overlapping sub-senses/usages of *ḍahaba*, *maḍā*, and *rāḥa* based on inspection of 300 corpus hits per verb from ArabiCorpus and dictionary entries from Al-Munjid (monolingual) Dictionary and Al-Mawrid (Arabic-English) Dictionary.



Clearly, these three GO verbs have overlapping uses, although all three have a basic physical motion sub-sense equivalent ‘to go’ or ‘to leave’. One can also notice from Figure 1 that there are particular uses shared between two of the three verbs, as well as uses that are unique to each verb alone. What this inclusion diagram does not show is the distributional information regarding the occurrence of a non-overlapping sub-senses or usages for each individual verb. For instance, *ḍahaba* and *rāḥa* share an idiomatic expression denoting the sub-sense ‘to be a victim of X’, yet, if we inspect the frequency of occurrence of both verbs in this expression, this particular idiomatic construction seems to favor one verb (*rāḥa*) over the other (*ḍahaba*). In addition, inspection of the corpus returns for the verb *rāḥa* reveals that it is used predominantly as a continuous or

durative marker. *Maḍa* also appears as a durative marker, but such uses of this verb are considered minor in comparison with the almost fully grammaticalized *rāḥa*.

Furthermore, there is a difference with regard to the other verbs that collocate with *maḍā* and *rāḥa* in inceptive or durative constructions. Differences in lexical collocations also distinguish the GO and COME sets members in MSA.

This level of usage distinction, e.g. frequency of sub-sense occurrence and collocational patterning, is in most cases absent from traditional lexicographic accounts of these highly frequent verbs in MSA. It is therefore one of the main objectives of this study to provide a quantitative account of the idiosyncratic as well as shared sub-senses and uses of the seven (deictic) motion verbs and to cover the major different morphosyntactic and semantic properties of each verb's behavior. On the one hand, this analysis aims to place Modern Standard Arabic GO and COME verbs within the larger typological literature of the proliferate uses of these lexical items. On the other hand, it also offers a usage-based treatment for highly frequent verbs in MSA that benefits speakers and, most importantly, learners of Arabic.

In the following sections, I briefly sketch out the basic principles of verbal morphology in Arabic and, subsequently, problems associated with traditional treatments of the Arabic verb. This is followed by a brief outline of the methodological approach I adopt, a brief description of MSA GO and COME verbs, and finally a description of the structure of this study.

1.2 The morphology of the Arabic verb

Typically, an Arabic verb word is the outcome of an amalgamation of a consonantal root and a 'grid' or template that serves as a frame that defines the derivations of different verbal forms from one root. Each of these two components comes with its own attendant semantic force. Arabic grammars assert that there are around 10

major templatic verb forms as well as another 5 archaic forms (Ryding, 2005). These templatic verb forms vary in the semantic load that they add to the consonantal root. Form I, for instance, is considered the basic form of a verb and one that is the closest indicator of the meaning of the root, while Form II is a template that turns a Form I verb into a causative verb. An example of such a derivation is provided in Table 1 below. (See Appendix A for the entire list of templatic verb forms in Arabic along with their general meaning associations).

TABLE 1. Examples of derivations of different verbs from the same root, D-R-S, which pertains to learning.

Tri-consonantal root	Form I bare/simple C ₁ aC ₂ V C ₃ -	Form II causative C ₁ aC ₂ C ₂ aC ₃ -	Form VI reciprocal taC ₁ aaC ₂ aC ₃ -
درس DRS (involving learning, studying)	دَرَسَ DaRaSa = to study, to learn	دَرَّسَ DaRRaSa = to teach	تَدَارَسَ taDaaRaSa = to study/learn with one another

Each form, in its turn, appears in a wide range of additional inflected forms, since the Arabic verb can be inflected for aspect and mood (PERFECTIVE / IMPERFECTIVE / JUSSIVE / SUBJUNCTIVE), person (1st / 2nd / 3rd), number (SINGULAR / DUAL / PLURAL), and gender (MASCULINE / FEMININE). This morphological potential naturally generates a sizeable list of different inflected forms in which a certain root + template combination can appear (see Appendix B for a full inflectional paradigm for a tri-consonantal root). Traditional Arabic grammars elaborately spell out the entire inflectional paradigm for every root + template combination (clearly demonstrated by the full inflectional paradigm in Appendix B), and rightfully so, for pedagogical reasons, as not all inflected or derived forms are fully compositional and, thus, readily computable.

1.3 Problems with traditional accounts of the verb in Arabic

While the above description of the verb in Arabic represents an integral component of comprehensive grammars and pedagogical texts, the fact remains that the average verb in Arabic, in actual usage, hardly ever appears in all of its potential inflected forms. Previous inspection of 300 corpus concordance lines of the three MSA GO verbs (Abdulrahim, ms.) *ḍahaba*, *maḍā*, and *rāḥa*, with an additional verb – *inṭalaqa* – that MSA sometimes deploys as a GO verb,⁶ revealed that the total number of verb occurrences in the corpus is by no means evenly distributed across the full paradigm of theoretically possible inflected forms. Instead, corpus data show that each verb seems to be associated with a highly skewed distribution of inflected forms, with one or two particular inflectional patterns being predominant per form (i.e. PERFECTIVE, IMPERFECTIVE, etc.), if not per verb. Table 2 presents a list of the relative frequencies of occurrence of each inflected form per verb for all four GO verbs in the PERFECTIVE aspect.

⁶ The MSA corpus queried in this study is *ArabiCorpus.byu.edu*, which will be introduced at length in Chapter 2.

TABLE 2. Percentages of occurrence of inflected forms per MSA GO verb in the perfective aspect based on inspection of 300 concordance lines. The numbers in bold as well as the gaps in the table reflect the skewed distribution of actually inflected forms.

	ACTIVE/PERFECTIVE			
	<i>dahaba</i>	<i>maḍā</i>	<i>rāḥa</i>	<i>intalaqa</i>
أنا 1 ST .SG	5.0	--	6.3	0.3
أنتَ 2 ND .SG.M	1.3	0.3	0.3	--
أنتِ 2 ND .SG.F	--	--	0.3	--
أنتما 2 ND .DUAL	--	--	--	--
هو 3 RD .SG.M	22.3	44.0	50.7	14.3
هي 3 RD .SG.F	11.0	15.3	26.3	21.0
هما 3 RD .DUAL.M	0.3	0.3	0.3	0.3
هما 3 RD .DUAL.F	--	--	--	--
نحن 1 ST .PL	2.3	0.3	0.3	2.0
أنتم 2 ND .PL.M	0.7	--	--	--
أنتن 2 ND .PL.F	--	--	--	--
هم 3 RD .PL.M	2.7	1.0	9.3	1.0
هن 3 RD .PL.F	0.7	--	0.7	--

We can see that, as far as the perfective aspect is concerned, a full inflectional paradigm is hardly achieved since certain inflected forms did not materialize for any of the four verbs, such as 2DUAL.M, 3DUAL.F, and 2PL.F. Meanwhile, the four verbs appear to favor either 3SG.M or 3SG.F forms or both in their perfective inflections. The preference of an individual verb for either the PERFECTIVE or the IMPERFECTIVE, or the MASCULINE or the FEMININE, is not accidental since certain inflected forms appear to have dedicated meanings and functions in the language. For example, there appear to be very few occurrences of *rāḥa* in the IMPERFECTIVE form (none of which are reported in Table 2), as well as very few occurrences in the PERFECTIVE, all of which underscores its main usage as a simple motion verb, as illustrated in (1).

(1) ويبدو فيها البشر كأنهم موتى تروح أشباحهم وتجيء

<i>wa=yabdu</i>	<i>fī-ha</i>	<i>al=bašar</i>	<i>ka'annahum</i>	<i>mawtā</i>
CONJ= appear.IMPF.3SG.M	LOC-CL.3SG.F	ART=humans	ADV	dead.PL
and appears	in it	the humans	as if they are	dead

<i>taruḥu</i>	<i>ašbāḥu-hum</i>	<i>wa=tağī'</i>
<i>rāḥa</i> .IMPF.3SG.F	ghosts-CL.3PL.M.GEN	CONJ=ğā'a.IMPF.3SG.F
go	their ghosts	and come

'Humans there appear as if they were dead, their ghosts go and come'

However, 77% of *rāḥa* occurrences are in the 3SG PERFECTIVE (MASCULINE and FEMININE) form which is the typical grammaticalized form of the verb in constructions as (2). In this example, the 3SG.M.PERF form of the verb is no longer associated with motion as it purely signals both the early stages of as well as the durativity of the singing event.

(2) راح يغني بصوت جميل

<i>rāḥa</i>	<i>yuganni</i>	<i>bi-šawtin</i>	<i>ğamil</i>
<i>rāḥa</i> .PERF.3SG.M	sing.IMPF.3SG.M	INST-voice	beautiful
went	sing	with voice	beautiful

'He started/went on singing with a beautiful voice'

Along with the grammaticalized usage of *rāḥa*, the 3SG.M.PERF form of the verb is also associated with the idiomatic usage of this verb: راح ضحيتها *rāḥa ḍaḥiyyatuha* ('go/die as a victim of'), as in (3).

(3) بعد مغامرة الحرب العراقية الايرانية التي راح ضحيتها آلاف الايرانيين

<i>ba'da</i>	<i>mugamarat</i>	<i>al=ḥarb</i>	<i>al='iraqiyya</i>	<i>al=iraniyya</i>	<i>allatī</i>
ADV	adventure	ART=war	ART=Iranina	ART=Iraqi	RP
after	adventure	the war	the Iranian	the Iraqi	which

<i>rāḥa</i>	<i>ḍaḥiyyatu-ha</i>	<i>ālaf</i>	<i>al=iraniyyin</i>
<i>rāḥa</i> .PERF.3SG.M	victim.NOM-CL.3SG.F.GEN	thousands	ART=Iranians
went	its victims	thousands	the Iranian

'After the adventure of the Iraqi-Iranian war which reaped the lives of thousands of Iranians...'

Aside from patterns of grammaticalization and idiomaticity, each of the four MSA GO verbs appears to co-occur almost exclusively with particular sets of subjects. As noted earlier, *maḍā* is the GO verb that seems to be most closely associated with 'time passage' (see Figure 1). The majority of 'time passage' construals of the 300 corpus

instances are associated with the 3SG.M.PERF form of *maḍā*, which may explain the 44% incidence rate of this inflected form in Table 2.

The specialized roles of certain inflected forms provide further evidence for claims that the lemmatized form of a lexical item does not constitute a proper basis for understanding its actual behaviour in a language. Rather, the different inflections or morphological constructions of a certain lexical item can each exhibit different idiosyncratic properties and functions and, most importantly, are not equally distributed across the full inflectional paradigm (cf. Newman, 2008; Newman and Rice, 2008). The above examples of *rāḥa* in MSA, in particular, is also a demonstration that “high-frequency instances of constructions undergo grammaticization processes” (Bybee, 2006:711). This brief illustration of the correlation between inflected form and verb usage or meaning provides further motivation for proposing a constructional analysis for basic verbs in Arabic.

As far as current MSA lexicographical practices are concerned, a large number of contemporary MSA dictionaries still tend to report (or repeat) only the lexical entries and meaning extensions found in the prominent dictionaries of Classical Arabic, such as *Lisan Al-Arab* that was compiled in the 13th century by Ibn Manzur. In Table 3, I compare monolingual CA and MSA dictionaries as well as bilingual MSA-English dictionary entries of the verb *rāḥa* with actual verb uses based on my preliminary inspection of 300 concordance lines from an MSA corpus.⁷

⁷ The CA entry for *rāḥa* is part of the headword *rawḥ* which is the tri-consonantal root from which *rāḥa* originally stems. Here, I only cite the information related to Form I *rāḥa*. It has to be noted that the entry for *rawḥ* extends for pages since *Lisan Al-Arab* is considered a kind of illustrated dictionary (20 volumes) with extensive lists of contextualized instances where a lexical item is used in the Qur’an, prophetic traditions, poetry, etc., in addition to etymological information. This makes *Lisan Al-Arab* more of a usage-based dictionary than many of its successors.

TABLE 3. Dictionary entries vs. corpus-based sub-senses of the verb *rāḥa*.

Dictionary/source	<i>rāḥa</i>
CA <i>Lisan Al Arab</i> (13 th century)	راح فلان يروح رواحاً: من ذهابه أو سيره بالعشي قال الأزهرى: وسمعت العرب تستعمل الرواح في السير في كل وقت Used for ‘going’ in the evening. Other opinions say that it’s used for ‘going’ during any time of day.
Monolingual MSA <i>Al Munjid Dictionary</i> (2005)	جاء أو ذهب في الرواح أي العشي وعمل فيه ويستعمل لمطلق الذهاب والمضي ‘Going/coming in the evening’; Used for the “absoluteness” of ‘going’ (telicity?)
Bilingual MSA-English <i>Al Mawrid Oxford</i> <i>Dictionary</i> (2008)	similar to ذهب <i>dahaba</i> , مضى <i>madā</i> : ‘go, go away, leave, depart’ راح يفعل كذا: شرع بدأ <i>rāḥa</i> + imperfective: ‘begin, start, set out’
MSA corpus <i>arabiCorpus.byu.edu</i>	‘begin, start, set out’ 78% idiomatic construction: ‘be a victim of’ 13% ‘go, leave, depart, go away’ 8.7% metaphorical usage: ‘futile, be in vain’ 0.3%

One striking finding in Table 3 is that the grammaticalized function of the verb *rāḥa* is never highlighted in the monolingual *Al-Munjid* dictionary (2005), while the corpus data show that in 78% of 300 concordance lines, *rāḥa* appears as an inchoative marker, denoting the ongoingness of activity or event coded in a second verb in the construction, rather than marking a GO event as the main verb. Despite the predominance of this particular function of *rāḥa* in contemporary MSA, the *Al-Munjid* dictionary only highlights the archaic verb usages of ‘going in the evening’. The *Al-Mawrid Oxford* (2008) bilingual dictionary, on the other hand, appears to acknowledge the durative or inceptive function of *rāḥa*. It is worth noting how the usage of *rāḥa* appears to have evolved from a time-specific motion verb in CA (going in the evening) to, subsequently, losing such associations altogether in many colloquial varieties and functioning as the sole GO verb. The same applies to the very few instances where *rāḥa* is utilized in MSA writing to refer to motion and a basic GO event. Interestingly, however, such usages appear to be condemned by native speakers of Arabic as ‘vernacularizing’ MSA. While many monolingual dictionaries provide an entry for *rāḥa* similar to that in the *Al Munjid*

Dictionary (2005), bilingual English-Arabic dictionaries, on the other hand, do not list *rāḥa* among the MSA verbs that indicate a GO motion event.

Overall, the fact that monolingual MSA dictionaries appear to ignore actual verb usages (as revealed by corpus data) may well be suggestive of a prescriptivist attempt at trying to keep MSA closer to a “purer” form of Arabic (CA), but unfortunately, at the expense of not providing a faithful description of contemporary verb usage. This lack of an adequate description of MSA verbal semantics and usage, aggravated by the conflicting and mismatching accounts of highly frequent verbs in different types of dictionaries, is troublesome—not only because it underrepresents the language to learners of Arabic, but also because Arabic data risk not being appropriately represented in typological studies due to the absence of comprehensive and data-driven treatments of language meaning and use.

1.4 Corpus-based, multifactorial approaches to the study of (near-) synonymy

The brief description of traditional treatments of the Arabic verb given above calls for an alternative and systematic treatment that acknowledges the idiosyncrasies of each of the (near-) synonymous Modern Standard Arabic GO and COME verbs. The availability of corpora for MSA guarantees that a truly usage-based analysis is possible, supported by contextualized and naturally-occurring data. Many studies on lexical and constructional alternations have exploited the large amount of data available in corpora for the investigation of numerous linguistic phenomena, and thus, turning away from purely introspection-based approaches to examining linguistic behavior (Newman, 2010).

A number of studies claim to highlight *the* single most important linguistic factor that determines the difference in behavior among two (or more) synonymous lexical or constructional items. For instance, Church et. al. (1991) identify the *lexical context* as the distinguishing factor between the adjective *strong* and *powerful*; Biber et. al. (1998)

identify the *syntactic argument patterns* as the factor that sets the verbs *start* and *begin* apart; etc. Such monocausal accounts of lexical synonymy have been criticized as short-sighted and insufficient in their representations of lexical behavior (cf. Gries, 2003; Arppe, 2008). In his study on English verb-particle constructions, Gries (2003) emphasized the importance of examining a wide variety of variables that pertain to different levels of linguistic analysis as a means of providing a more holistic account of the phenomenon under study. His multifactorial study of verb-particle placement ([V P NP_{DIRECT_OBJECT}] vs. [V NP_{DIRECT_OBJECT} P]) relied on inspecting the interaction among various morphosyntactic, semantic and lexical features pertaining to the direct object, the determiner of the direct object, as well as the complexity of the direct object. His results provide ample evidence for the necessity of examining a quantitatively wide variety of variables of different linguistic categories as a means of understanding the complexity of lexical or constructional uses.

In particular, as far as the study of lexical synonymy is concerned, more and more studies of (near-) synonymy have explored the benefits of examining multiple contextual features pertaining to the use of a specific linguistic item, in addition to exploiting a number of multivariate statistical analysis that can explain lexical behavior as a function of a combination of linguistic variables. Among these studies are Gries's (2006) account of *run*, Divjak and Gries's (2006) account of Russian verbs of TRYING, and Gries and Otani's (2010) study on size adjectives in English. These studies rely on annotating a large number of corpus concordance lines for an extensive set of diverse linguistic features (morphological, syntactic, and semantic), and exploring the distance between these individual lexical items through clustering techniques. Gries and collaborators have referred to this method as *Behavioral Profiles*, on which I will elaborate more in the next chapter.

Bresnan (2006, 2007) and Arppe (2008) have also argued for a multifactorial approach to the study of lexical synonymy or constructional alternations as reflecting a probabilistic – rather than categorical – view of lexical or constructional choices. This means that “the workings of a linguistic system, represented by a range of variables according to a theory, and its resultant usage would not in practice be categorical, following from exception-less rules, but rather exhibit degrees of potential variation which becomes evident over longer stretches of linguistic usage” (Arppe, 2008:11-12). Both Bresnan (2007) and Arppe (2008) have adopted logistic regression as a statistical method that models lexical or constructional choices as a function of a wide range of contextual features. In his study on Finnish THINK verbs, Arppe (2008) employs *polytomous logistic regression analysis* as a means of identifying exemplary contexts of use that are prototypical to one verb over the rest, as well as identifying contexts of use in which two or more of the four Finnish verbs – *ajatella*, *mieltii*, *pohtia* and *harkita* – are interchangeable.

In this dissertation, I will explore both multi-variate statistical methods – *Behavioral Profiles* and *polytomous logistic regression analysis* – in combination with a third test – *hierarchical configural frequency analysis* – (von Eye, 1990; Gries, 2004) which will help us examine robust clusters of variables in a large body of data. As I will explain in Chapter 2, the quantitative analysis I adopt for the study of the near-synonymous MSA GO and COME verbs relies on the construction of a data frame consisting of a large number of corpus returns of the seven motion verbs under examination and the subsequent markup of the data frame for multiple morphosyntactic and semantic variables, similar to the procedure discussed in Gries and Divjak (2009). This statistical analysis of Arabic GO and COME verbs, therefore, covers a wide range of statistical tests that are both monivariate and multivariate in nature.

1.5 Basic verbs and GO and COME

The field of cognitive linguistics stipulates that language is grounded in our experiential knowledge of the world. This means that whatever we experience, internally or externally, and whatever we interact with, concrete or abstract, all contribute to the ways we conceptualize reality. Language consists of “a vast repository of ready-made conceptualizations made available to its speakers, covering a host of domains”, such as time, space, containment, along with the categorization of humans, animals, plants, places, objects, processes, etc. (Newman, 2000:2). According to Lakoff (1987:266), the ‘experientialist’ approaches to understanding language strive to “characterize meaning in terms of nature and experience of the organisms doing the thinking.”

The study of basic verbs provides evidence and support for such approaches. Basic verbs correspond to concepts that are related to our everyday activities as human beings, such as going, coming, sitting, standing, lying, eating, drinking, seeing, hearing, thinking, etc., and which are represented across all human languages. Basic verbs (or linguistic units related to these fundamental human activities) are, typically, high-frequency linguistic items that tend to be part of the early vocabulary acquired by children. The high frequency of basic verbs is not necessarily tied to their occurrences in their literal senses, but is also reflective of the fact that these verbs tend to grammaticalize across languages, in that they gradually become associated with grammatical functions such as tense or aspect marking, in addition to their wide participation in idiomatic and figurative constructions. Among the many non-literal uses of the English *go*, this verb can be used as a future auxiliary as in *I'm going to be in my office all day long*, where we see that the event does not involve any sort of physical motion, but that *going to* signals a future event. The verb *SIT* in Bahraini Arabic has doubly grammaticalized as a progressive aspect marker (in its active participle form), as shown in (4), and as a

others). Lyons defines deixis as “the location and identification of persons, objects, events, processes and activities being talked about or referred to, in relation to the spatiotemporal context created and sustained by the act of utterance and participation in it” (1977: 637). Across languages, deixis tends to be encoded in highly frequent and grammaticalized lexical items such as demonstratives, tense systems, personal pronouns, some adverbials of time and place, and verbs of motion such as GO and COME (Lenz, 2003). Fillmore, in his lectures on deixis (1971), described verbs of deictic motion (e.g. *go* and *come*) as contextualized motion events, for which the “context is defined in such a way as to identify the participants in the communication act, their location in space, and the time during which the communication act is performed” (1971:38).

Fillmore and other researchers writing on GO and COME (cf. Gathercole, 1978; Rauh, 1981) consider GO and COME as being in deictic opposition, where COME is typically viewed as motion towards the location of the speaker or the addressee (i.e. the deictic centre), at either the time of encoding or the time of reference (Fillmore, 1977: 68), while GO is associated with motion *not* towards the speaker’s location at the time of encoding (Fillmore, 1977: 53). Many such studies have emphasized this deictic aspect of GO and COME verbs, yet crosslinguistic evidence provided by Wilkins and Hill (1995) and Botne (2005) has shown that languages are not necessarily consistent in their assignment of a deictic value to this pair of verbs, especially GO verbs, and that the internal semantics of these verbs differ considerably across languages.

The deictic nature of GO and COME verbs has inspired a number of subsequent studies that have looked beyond basic motion events and examined the role of GO and COME verbs in metaphorical and idiomatic constructions (e.g. Clark, 1974; Radden, 1996). Clark (1974), for example, claims that deixis explains the participation of this pair of verbs in idiomatic expressions in English that denote change of state rather than motion events. In a construction such as *He came round*, Clark suggests that the deictic

centre of *come* corresponds to a normal state of being, that of ‘consciousness’. *Go*, on the other hand, seems to signal departure from the normal state towards the non-normal state, as in *He went mad*.

Deixis in the case of GO and COME often entails *telicity*, or event realization / termination (Radden, 1996; Comrie, 1985:15). The idea that a COME event has a terminal point (GOAL) that is associated with the deictic centre entails that the event is inherently telic. In other words, most COME events are complete events. GO, on the other hand, does not typically imply arrival at a GOAL, which makes it inherently atelic (e.g. *He went away on a long weekend*), unless the destination of the motion event is otherwise specified in the event construction, as in a modifying locative phrase (e.g. *He went to the beach for the long weekend*).

Clark (1974) and Radden (1996) also identify *viewpoint* as an additional property of the semantics of GO and COME verbs. This particular property is not necessarily tied to deixis, as in narrative situations such as *The men came into the house* (Radden, 1996:411), but is also related to what Clark refers to as having ‘evaluative’ connotations, such as the use of *come* and *go* in (7a) and (7b), respectively.

- (7) a. *The hot air balloon came down in the school yard.*
b. *The hot air balloon went down in the school yard.*

In (7a) the outcome is desirable, i.e. that of the hot air balloon landing safely, while in (7b) the outcome is tragic since it describes a crash landing of the hotair balloon. Radden (1996:433) points out that such construals are highly language-specific, and the viewpoint taken by the observer in the English sentences in (7a) and (7b) can be reversed as in the Swedish sentences in (8a) and (8b).

- (8) a. *Planet gick ner.* ‘The plane went down’
b. *Planet kom ner.* ‘The plane came down’

Unlike the English phrase in (7b) The use of *går ner* ‘go’ in (8a) in Swedish is conceptualized from the perspective of inside the plane (and thus being controlled by the pilot), while in (8b) the use of *kommer ner* ‘come’ is seen from the viewpoint of outside the plane where it comes down unexpectedly.

Radden also identifies *locomotion and object motion* as the property inherent to a deictic motion event that is responsible for the moving EGO and moving world metaphors (1996:435). Such metaphors are mostly reflected in construals related to the passage of time, as in the conceptual metaphors (i) TIME PASSING IS MOTION OVER A LANDSCAPE, in which the observer is conceptualized as moving over a landscape; and (ii) TIME PASSING IS MOTION, in which the human observer is stationary, while temporal events are in motion. According to Radden, the former conceptual metaphor is responsible for figurative (and grammaticalized) usage of *go* in *I am going to be a lawyer*, while the latter explains the expression *the days gone by* (1996:435).

Newman (2000) proposes an additional property – *intentionality* – that usually characterizes GO and COME verbs and is responsible for the extra lexical and grammatical functions this pair of verbs assumes in addition to expressing basic motion. According to his classification, while the *atelicity* of GO seems to encourage the process of turning the MSA *rāḥa* into a durative or persistive marker – as in the repeated example in (9) – the inherent *intentionality* of the event of going seems to explain why this same verb (*rāḥa*) also assumes the function of a future tense marker in some colloquial Arabic dialects (Rubin, 2005), e.g. Bahraini and Kuwaiti varieties – as in (10).

(9) راح يغني بصوت جميل

<i>rāḥa</i>	<i>yuḡanni</i>	<i>bi-ṣawtin</i>	<i>ḡamil</i>
<i>rāḥa</i> .PERF.3SG.M	sing.IMPF.3SG.M	INST-voice	beautiful
went	sing	with voice	beautiful
‘He started/went on singing with a beautiful voice’			

(10) احنه راح نكلمها في الموضوع

iḥna	<i>rāḥ</i>	nkallem-ha	fi=l=mawḏūʿ
PP	<i>rāḥa</i> .PERF.3SG.M	talk.IMP.F.1PL-CL.3SG.F.ACC	LOC=ART=topic
we	went	talk her	in the topic

‘We will talk to her about this topic’

The above properties that pertain to the lexical (or constructional) semantics of verbs of GOING and COMING are among the driving forces that motivate their evolution into grammatical particles in any given language and their participation in metaphorical and idiomatic expressions (cf. Lichtenberk, 1991; Newman, 2000; Heine and Kuteva, 2002).

The vast majority of earlier cross-linguistic studies of GO and COME verbs have mostly been concerned with an individual feature or a very small number of features that pertain more often to the lexical semantics of these verbs (e.g. deixis, motion), and have been qualitative in nature, where free data elicitation and experimental elicitation (e.g. Wilkins and Hill, 1995), along with introspection (e.g. Fillmore, 1966, and many others), comprise the main methods for data collection here. Very few studies of GO and COME have attempted to incorporate corpus data as a means of presenting contextualized usages of lexical items, nor have they added a quantitative aspect to the analysis. One such study is Di Meola’s (1994) account of the German KOMMEN and GEHEN verbs, in which he talks about both the deictic and non-deictic uses of this pair of verbs. Another study is Newman and Lin’s (2005) on the purposefulness of *going* in English, where the authors examined 100 instances of usage of the verb *go*, from the BNC, for collocational patterns. The main purpose of their study was to establish the degree to which the encoding of purposefulness in English *go* constructions has been conventionalized. The authors examined the corpus data for instances where *go* co-occurs with other verbs patterns such as *go* and V (e.g. *go and get her*) and *go* to V (e.g. *I mustn’t go to see William*), in addition to expressions where reference to destination (GOAL) is specified which specifically reflects conventionalized purposes (e.g. *go to school / bed / work / the*

library, etc.). Newman and Lin compared these *go* collocations with similar collocations associated with two other verbs, *run* and *walk*, and found that purposefulness is expressed in *go* constructions far more than in *run* and *walk* constructions. For instance, in the sentence *go and make me a cup of tea*, according to Newman and Lin, what appears to be informationally salient is the purpose of making a cup of tea rather than motion away from the speaker. While this utterance does involve an early motion phase, *go* appears to be more associated with purpose here than with motion. In the case of combinations such as *go to school*, this does not necessarily mean motion towards a GOAL per se, but rather motion towards a destination that serves a particular purpose, that of participating in the activities associated with school (teaching, learning, etc.). The authors attribute this finding to the lack of semantic specificity in *go* (compared to *walk* and *run* which encode manner of motion), and so this “relative lack of semantic content with *go* invites a greater elaboration of the associated purpose of the motion” (Newman and Lin, 2005:304).

This study sheds light on the usefulness and reliability of corpus-based methods in grammaticalization research, where an approach as simple as investigating conventionalized co-occurrence patterns in context can reveal much about the diachronic processes of grammaticalization that many highly frequent linguistic items undergo in languages (cf. Newman and Rice (2004) on the co-occurrence of *sit*, *stand*, and *lie* with other verbs in English). The predominance of the purposive use of *go* in English may be symptomatic of a gradual grammaticalization process that has indeed been completed in some world languages, where *go to* serves as a ‘purpose’ marker (Heine and Kuteva, 2002, cited in Newman and Lin, 2005). A study such as Newman and Lin’s provides further support for the usage-based quantitative approach adopted here for the examination the lexico-syntactic behaviour of GO and COME verbs in Modern Standard Arabic.

1.6 Modern Standard Arabic GO and COME verbs

1.6.1 GO verbs in MSA

Bilingual English-Arabic dictionaries are not unanimous about the (Modern Standard) Arabic equivalent to the English verb *go*. I decided to focus on three verbs - *ḡahaba*, *maḡā*, and *rāḡha* – in accordance with the ‘basic motion event’ criteria discussed in detail in the next chapter. One such criterion requires that all selected verbs can, ideally, be used interchangeably in a constructed sentence depicting physical motion event frame, as in (11).

(11) ذهب \ مضى \ راح الأب إلى مركز الشرطة

<i>ḡahaba / maḡā / rāḡha</i>	<i>al= 'ab-u</i>	<i>ilā</i>	<i>markaz</i>	<i>al=šurṡa</i>
<i>ḡahaba / maḡā / rāḡha</i> .PERF.3SG.M	ART=father-NOM	ALL	station	ART=police
went	the father	to	station	the police
‘The father went to the police station’				

Certain lexicographic treatments of these three verbs do, in fact, list some of the individual characterizing features associated with each of these verbs. For instance, a number of monolingual and bilingual dictionaries indicate that *maḡā* is the GO verb that depicts the passage of time, as in (12), while *rāḡha* is predominantly used as a grammaticalized particle marking inceptive and/or durative aspect, as in (13).

(12) مضى وقت طويل على لقائنا

<i>maḡā</i>	<i>waqt-un</i>	<i>ṡawīl</i>	<i>'alā</i>	<i>liqā' i-na</i>
<i>maḡā</i> .PERF.3SG.M	time-ACC	long	LOC	meeting-CL.1PL.GEN
went / pass	time	long	over	our meeting
‘A long time has passed since our meeting’				

(13) وراح ينادي بصوت عالٍ

<i>wa=rāḡha</i>	<i>yunādi</i>	<i>bi-ṡawtin</i>	<i>'alin</i>
CONJ= <i>rāḡha</i> .PERF.3SG.M	call.out.IMPF.3SG.M	INST-voice	loud
and went	call	with voice	loud
‘And he started/went on calling out in a loud voice’			

There is no doubt that *rāḡha* is almost fully grammaticalized in MSA. Nevertheless, the decision to add this verb to the set of GO verbs examined here stems from the fact that a number of the examined corpus returns of this verb still depict deictic motion despite

the stigmatized status of *rāḥa* as a ‘colloquial’ GO verb. Another added benefit of including *rāḥa* in the current quantitative analysis – as well as the qualitative analysis in Chapter 4 – is to compare the partially grammaticalized uses of *maḍā* to the more established grammatical uses of *rāḥa*. The discussion about these three GO verbs, therefore, is intended to shed light on not only the morphosyntactic features of GO verbs used in their literal and figurative senses, but also in their grammaticalized senses. Hence, the corpus-based quantitative as well as qualitative analyses presented in this study aim to identify typical patterns of verb usage and sub-senses per each of the studied GO verbs, as well as any less typical, yet robust, patterns of verb usage.

1.6.2 COME verbs in MSA

As with GO events, more than one verb can signal a COME event in Modern Standard Arabic. In particular, the four verbs *atā*, *ḡā’a*, *ḥaḍara*, and *qadima*, are all translated as COME in English. Some modern and classical dictionaries consider these lexical items synonymous, since they can be used interchangeably in a context such as (14). In this sentence, we have a human agent moving towards a destination (one that is coextensive with the deictic centre), and the event is expressed in a past tense construction with perfective inflection.

(14) أنت / جاءت / حضرت / قدمت جدتي إلى المطار لتودعني

<i>atat / ḡā’at / ḥaḍarat / qadimat</i>	<i>ḡadda-ti</i>	<i>ila</i>
<i>atā / ḡā’a / ḥaḍara / qadima</i> .PERF.3SG.F	grandmother.CL.1SG.GEN	ALL
came	my grandmother	to

<i>al=maḡār</i>	<i>li=tuwaddi’a-ni</i>
ART=airport	PURP=say.goodbye.SUBJN.1SG.ACC
the airport	to say goodbye to me
‘My grandmother came to the airport to say goodbye to me’	

The sentence in (14) would not admit all four verbs when the aspect inflection on the verb is changed. In (15), for instance, if we hold all constructional features constant

and change verb inflection from perfective to jussive, this results in a preference for *atā* and *ḥaḍara* by native speakers of Arabic over *ġā'a* or *qadima*.

(15) لم تأت / أتجيء / تحضر / تقدم جدتي إلى المطار لتودعني

<i>lam</i>	<i>ta'ti / ?taġi' / taḥdur / ?taqdum</i>	<i>ġadda-ti</i>	<i>ila</i>
NEG	<i>atā / ġā'a / ḥaḍara / qadima</i> .JUUS.3SG.F	grandmother.CL.1SG.GEN	ALL
did not	come	my grandmother	to

al=maṭār *li=tuwaddi'a-ni*
 ART=airport PURP=say.goodbye.SUBJN.1SG.ACC
 the airport to say goodbye to me

'My grandmother did not come to the airport to say goodbye to me'

Changing the semantic category of the sentential subject from human to non-human, as in

(16), results in a dispreference for *ḥaḍara* and *qadima*.

(16) أنت / جاءت / حضرت / *قدمت رغبتهم في رؤيتها بعد سماعهم بخبر فوزها بجائزة اليانصيب

<i>atat / ġā'at / *ḥaḍarat / *qadimat</i>	<i>raġbatu-hum</i>	<i>fi</i>	<i>ru'yati-ha</i>	<i>ba'da</i>
<i>atā / ġā'a / *ḥaḍara / *</i>	desire-	LOC	seeing-	ADV
<i>qadima</i> .PERF.3SG.F	CL.3PL.M.GEN		CL.3SG.F.ACC	
came	their desire	in	seeing her	after

samā'i-him *bi-ḥabar* *fawzi-ha* *bi-ġa'izat* *al=yānaṣib*
 hearing-CL.3PL.M.GEN INST-news win.VN-CL.3SG.F.GEN INST-award ART=lottery
 their hearing of news her winning of award the lottery

'Their desire to see her came after hearing the news regarding her winning the lottery'

Clearly, each of these four verbs is associated with a cohort of meaning extensions and usages and, most importantly, specific constructional elements. The constructed examples emphasize the fact that while the four COME verbs share certain contextual features, as in (14), we can see that in (15) and (16) the manipulation of TAM and semantic features reveal sharp selectional distinctions.

1.7 The structure of this study

As mentioned earlier, a more thorough treatment of the highly frequent GO and COME verbs in MSA should take into account the morphosyntactic and lexico-semantic characteristics of the construction hosting the verb. Such a constructionist theoretical

approach is methodologically supported by the availability of MSA corpora, as well as statistical techniques for a quantitative and systematic examination of the behavior of the seven MSA verbs of motion.

In Chapter 2, I will elaborate on the methodological approach adopted for this study of MSA motion verbs. I will discuss the procedure and the criteria followed for the selection of MSA GO and COME verbs. I will then provide a description of the MSA corpus chosen for data collection (ArabiCorpus.byu.edu), as well as a description of the data frame constructed for each GO and COME verb and the variables each verb usage was annotated for. A large portion of this chapter is also dedicated to the explanation of the statistical techniques I have adopted for the examination of single variable distribution, as well as the examination of interaction between multiple variables at a time. These statistical tests have all been conducted in R (www.r-project.org) with the help of statistical scripts and packages written by Stefan Gries and Antti Arppe.

In Chapters 3 and 5, I will present quantitative analyses of GO verbs and COME verbs (respectively). These two chapters follow the same structure, in which the analysis increases gradually in terms of computational complexity, as well as in the number of variables explored. I will therefore begin each of the two chapters with the examination of frequency distributions of individual variables per verb, and then move on to multivariate analyses that investigate the level of interaction between a set of variables. These analyses can eventually help us zero in on exemplary or prototypical constructions that are highly characteristic of the use of each verb.

Chapters 4 and 6, on the other hand, provide a qualitative analysis that takes into account the preceding quantitative discussion of each verb set. More importantly, these analyses aim to discuss aspects of the use of GO and COME verbs in MSA in light of previous cross-linguistic studies on lexical items related to the concept of GO and COME.

Notions such as deixis, telicity, and grammaticalization. are therefore highlighted and discussed in depth in Chapters 4 and 6.

Chapter 7 provides a synthesis of the results obtained and discussed in Chapters 3-6. This chapter deals with multiple themes covered in this dissertation, such as providing an evaluation of the corpus-based quantitative analysis and the extent to which this analysis has succeeded in identifying higher-level and lower-level generalizations regarding the use of each GO and COME verb in MSA.

Finally, in Chapter 8, I discuss future directions of research and practical applications. I will go back to the notion that MSA is a highly literary and media-based language and discuss what my findings may reflect about the usage of motion verbs in such a linguistic variety, in contrast to findings obtained from the investigation of GO and COME verbs in other languages (e.g. Wilkins and Hill, 1995 and Botne, 2005). I will therefore suggest further research on GO and COME verbs in the spoken varieties of Arabic. This is followed by a discussion of experimental approaches, currently underway, that can be regarded as a future extension of the analysis provided in this thesis. I also propose further studies that take into account the non-finite verbal forms related to the three GO and the four COME verbs (e.g. Verbal nouns and Active participles). In terms of practical applications drawn from this research, I will propose more elaborate and more representative dictionary entries for the verbs studied here. I illustrate three kinds of dictionary entries for the COME verb *atā*: (i) corpus-illustrated, which follows the currently available lexicographic accounts of this verb yet provides more exemplary uses per each sub-sense of the verb; (ii) frequency-based, which is a minimal dictionary entry that categorizes verb uses into figurative, non-figurative and phrasal uses and provides information about the frequency of occurrence of each usage; and finally (iii) corpus-based dictionary entries, which draw on the quantitative analysis of *atā* that has been

presented in Chapter 5, and which spells out the constructional features associated with each sub-sense and usage of the verb.

Chapter Two Methods

2.1 Selection of the MSA GO and COME verbs

The MSA verbs that I analyze in this study are *ḍahaba*, *maḍa* and *rāḥa* (the GO verb set) and *atā*, *ḡa'a*, *ḥaḍara* and *qadima* (the COME verb set). Bilingual English-Arabic dictionaries cite other verbs in addition to the ones mentioned above as possible translation equivalents of the English *go* and *come*. Table 1 shows a selection of sub-entries for *go* and *come* verbs in a number of English-Arabic dictionaries.⁹ The seven verbs discussed in this study are shown in boldface. The most obvious observation is that these dictionaries are not unanimous as to what should count as the Arabic (or, strictly speaking, MSA) translation of the English verb.

TABLE 1. Sample dictionary entries for GO and COME in three bilingual English-MSA dictionaries.

Dictionary	Sub-entries for the verb GO	Sub-entries for the verb COME
Concise Oxford English- Arabic Dictionary (1982)	(1) move away: ذهب <i>ḍahaba</i> , مضى <i>maḍa</i> (2) become: تغير <i>tagayyara</i> (3) work, function: اشتغل <i>ištaḡala</i> , عمل <i>'amila</i> (4) be placed: وُضِعَ <i>wuḍi'a</i> (5) extend: امتد <i>imtadda</i> , غطى <i>ḡaṭṭā</i> , وفى <i>waffā</i> بالغرض <i>bil ḡaraḍ</i> etc.	(1) arrive, move, be brought: جاء <i>ḡā'a</i> (2) happen, occur, result: حدث <i>ḥadaṭa</i> , حصل <i>ḥaṣala</i> , نتج <i>nataḡa</i> etc.
Al-Muhit Oxford Study Dictionary English – Arabic (1996)	(1) انتقل من مكان <i>intaḡala</i> ذهب <i>ḍahaba</i> , انتقل من مكان <i>intaḡala</i> من مكان <i>men makān</i> إلى آخر <i>ila 'āḡar</i> 'move from one place to another' (2) امتد <i>imtadda</i> 'extend', قاد <i>qād</i> 'lead', ذهب <i>ḍahaba</i> (3) كان <i>kān</i> 'be: PERFECTIVE – in a certain state', مضى <i>maḍā</i> etc.	(1) أتى <i>atā</i> , جاء <i>ḡā'a</i> , حضر <i>ḥaḍara</i> (2) وصل <i>waṣala</i> , أتى <i>atā</i> , قدم <i>qadima</i> (3) (of an ailment) تطوّر <i>taṭawwara</i> 'develop', بالتطور <i>baltaṭawwara</i> – بدأ <i>badā</i> – بالتطور <i>baltaṭawwara</i> 'started developing' (4) حصل <i>ḥaṣala</i> , حدث <i>ḥadaṭa</i> 'happen' etc.
Google online translator < http://translate.google.com/ >	ذهب <i>ḍahaba</i> : go, leave, be, gild, gang, betake انتقل <i>intaḡala</i> : go, set out, start, dash, tee off, shove off مضى <i>maḍā</i> : go, leave, run out, go on doing خرج <i>ḡaraḡa</i> : go out, party, step out, go, march out, go away مشى <i>maṣā</i> : walk, traipse, go, tread, step, foot سافر <i>sāfara</i> : travel, fly, journey, tour, ride, go غادر <i>ḡāḍara</i> : leave, quit, depart, go, retire, start etc.	جاء <i>ḡā'a</i> : came, come, arrive, bring, turn up أتى <i>atā</i> : came, come, derive وصل <i>waṣala</i> : link, connect, come, arrive, reach, hook up حضر <i>ḥaḍara</i> : present, attend, come, prepare, make, civilize حدث <i>ḥadaṭa</i> : place, happen, occur, come, take place, pass عبّر <i>'abara</i> : cross, express, come, voice, pass, come across قدم <i>qadima</i> : present, offer, submit, show, extend, come etc.

⁹ See Appendix C for a sample of several monolingual and bilingual dictionary entries of the seven verbs in Arabic (MSA and CA)

It is not surprising that dictionaries differ in citing what may count as a translation equivalent for such highly polysemous lexemes as *go* and *come* in English. However, as I have explained in Chapter 1, lexicographic treatments of such basic and highly frequent verbs proved to be inconsistent and impoverished, in addition to being haphazard and random at times. For instance, no literate speaker of Arabic would doubt the fact that *atā* is one of the most widely used COME verbs in MSA. Yet, the Concise Oxford English-Arabic Dictionary (1982) fails to include this important verb in the entry for *come*. It is also quite apparent that the listing of the various sub-senses of the English *go* and *come* verbs along with their MSA equivalents is neither methodical nor based on the frequency in which a certain sub-sense of the *go* or *come* equivalent occurs. In addition, there is a great deal of mismatch between the representation of verbs in monolingual dictionaries versus that in bilingual dictionaries. For instance, the verb *rāḥa* is treated as a GO verb in a number of monolingual dictionaries, mostly to indicate the archaic usage of the verb ‘to go in the evening’; yet it is not mentioned in any of the *go* entries of the bilingual dictionaries given in Table 1. This mismatch between monolingual and bilingual (as well as among dictionaries in the same category) makes the task for a learner of Arabic looking up words in a dictionary quite difficult.

Querying these dictionaries has proved to be an unreliable method for selecting the verbs to include in this study. I decided, therefore, to base my selection of verbs on the following set of four criteria: interchangeability, morphological simplicity, high frequency, and indication of basic deictic motion, which I will discuss in greater detail in the following sections.

2.1.1 Interchangeability of the motion verbs

One of the main criteria upon which the selection of GO and COME verbs for this study was based was the interchangeability of verbs in a *deictic* physical motion event construal. As mentioned in Chapter 1, the constructed sentence in (1) should typically host one of the three GO verbs, and convey physical motion not towards the deictic centre; while the constructed sentence in (2) should typically host one of the four COME verbs and indicate physical motion towards the deictic centre.

(1) Context of use allowing all three GO verbs

ذهب \ مضى \ راح الأب إلى مركز الشرطة
ḡahaba/maḡā/rāḡa *al= 'ab-u* *ilā* *markaz* *al=šurṡa*
ḡahaba/maḡā/rāḡa.PERF.3SG.M ART=father-NOM ALL station ART=police
went the father to station the police
‘The father went to the police station’

(2) Context of use allowing all four COME verbs

أنت / جاءت / حضرت / قدمت جدتي إلى المطار لتودعني

atat / ḡā'at / ḡaḡarat / qaḡimat *ḡadda-ti* *ila*
atā / ḡā'a / ḡaḡara / qaḡima.PERF.3SG.F grandmother.CL.1SG.GEN ALL
came my grandmother to

al=maḡār *li=tuwaddi'a-ni*
ART=airport PURP=say.goodbye.SUBJN.1SG.ACC
the airport to say goodbye to me
‘My grandmother came to the airport to say goodbye to me’

As also mentioned in the introduction, despite its inconsistent lexicographic representations, the usage of the GO verb *rāḡa* is avoided in formal written Arabic due to its strong association with colloquial uses. It is mostly used in MSA as a grammatical marker, signaling inceptive and continuative aspect. I did, however, encounter a number of uses of this verb as a motion verb in the corpus I queried. *Rāḡa* was also found to overlap in grammatical and idiomatic uses with the other two GO verbs – *ḡahaba* and *maḡā* – which constituted a further reason to include this verb in the GO verbs set and examine its usage against that of the other two GO verbs.

2.1.2 Morphological simplicity

Another important criterion for selecting the most ‘basic’ MSA GO and COME verbs for this study is morphological simplicity. Arabic verbs are known for their complex derivational patterns that result from combining certain consonantal lexical roots with a primarily vocalic frame, commonly referred to as ‘template’ or ‘verb form’. According to Ryding “the lexical root of three consonants can theoretically interlock with ten different patterns to produce ten lexical variants on the same root” (2005:434). Traditionally, the combination of lexical root + template results in a lexical item that is semantically related to the root with variations in meaning (e.g. voice, valency).

The simplest form is Form I, for which the conventional consonant-vowel patterning is $C_1aC_2aC_3a$, $C_1aC_2uC_3a$ or $C_1aC_2iC_3a$. Traditional accounts of the Arabic verb refer to this verb form as *fi'l muğarrad* فعل مجرد (i.e. ‘the stripped verb form’), since it represents the simplest form morphologically among all 10 derived verb forms. Arabic grammarians claim that, in most cases, the meaning of verbs in this form reflects the basic meaning of the consonantal root.

The remaining verb forms (II-X) are referred to as *af'aal mazida* أفعال مزيدة (i.e. ‘increased/augmented verb forms’) and involve more complex consonant-vowel patterns. It is assumed that the meaning of the derived verb (formed via the interlocking of consonantal root and template) should be analyzable by pairing the general meaning of the consonantal root with the grammatical meaning of the template. Arabic grammarians have pointed out that the meaning of the derived verb form can sometimes lose its analyzability over time (cf. Ryding, 2005; Holes, 2004). Obvious examples can be taken from the cohort of verbal forms derived from body part terms. For instance, the noun *riğl* ‘leg’ comes from the tri-consonantal root *r-ğ-l*. The combination of this root with the verb form VIII ($iC_1taC_2aC_3a$) results in the word *irtağala* which is used primarily to mean ‘to improvise or to deliver offhandedly, without preparation’ (Al Mawrid Arabic-English

Dictionary, 2008). The COME and GO verbs I selected for this study all have the same morphological make-up in that they all belong to the Form I category of verbal derivation. The purpose of this restriction on the selected verbs is to make sure that no additional semantic information, regarding, for instance, voice and valency, is added to the basic motion sense.

2.1.3 High frequency

The seven GO and COME verbs studied here are among the most frequent words in Arabic, according to *A Frequency Dictionary of Arabic* (Buckwalter and Parkinson, 2010). This frequency dictionary is based on a corpus of 30 million words, 90% of which is compiled from written sources that encompass different genres (e.g. newspapers, literature and non fiction), while the remaining 10% comes from unscripted spoken data from different Arabic dialects.

Table 2 shows the ranking of these seven verbs with respect to the 5000 most frequent words in Arabic according to Buckwalter and Parkinson (2010). Information about raw frequencies per verb as well as dispersion are also provided. The dispersion measure indicates the extent to which a certain lexical item is distributed over the different sub-sections in the corpus. If this figure approaches 100, it means that the lexical item in question is close to being equally represented in the various genres. Table 2 shows that the first 5 verbs have a dispersion of 90 or more, while the calculated range of the distribution of *maḍā* and *qadima* is 89 and 87, respectively. Notice that all seven verbs fall within the top 1000 words in this dictionary, with the exception of *qadima* which ranks as #3121. We can also see that *rāḥa* is most widely used in the spoken sub-section of the corpus, while *maḍā* is more likely to appear in literary writing.

TABLE 2. Ranking of the seven COME and GO verbs as listed in *A Frequency Dictionary of Arabic*.

verb	Rank (/5000 most frequent words)	raw frequency	dispersion
جاء <i>ǧā'a</i>	109	26234	99
راح <i>rāḥa</i>	113	25643	98 (+spoken)
أتى <i>atā</i>	343	12231	90
ذهب <i>ḍahaba</i>	489	8703	90
حضر <i>ḥaḍara</i>	809	4598	99
مضى <i>maḍā</i>	908	4502	89 (+ written literature)
قدم <i>qadima</i>	3121	566	87

The high frequency of these verbs across the different genres of Arabic has been one of the main motivations for selecting GO and COME verbs to study. These verbs count among the main lexical items that learners of Arabic should master and, consequently, there should be an adequate lexicographic treatment of these verbs and the various types of morphosyntactic constructions, lexical collocates and meaning extensions each of these verbs typically associates with.

2.1.4 Basic deictic motion events

Newman (2004) defines ‘basic verbs’ as lexical items encoding basic bodily events and states, such as COME, GO, SIT, STAND, LIE, EAT, DRINK, TAKE, GIVE, etc. The main criterion for selecting the seven MSA GO and COME verbs used in this study is their reference to basic (deictic) motion. The COME verbs *atā*, *ǧā'a*, *ḥaḍara*, and *qadima* signal motion towards a deictic centre (the speaker), while the GO verbs *ḍahaba*, *maḍā*, and *rāḥa* signal motion not towards the deictic centre (the speaker).¹⁰ There are a number of MSA verbs of motion that were also considered for this study but were disregarded due to the fact that they encode additional information about certain aspects of the motion event (e.g. path, manner) besides deixis.

¹⁰ The literature on GO verbs describes these verbs cross-linguistically either as motion *away* from the deictic centre or motion *not towards* the deictic centre (cf. Wilkins and Hill, 1995). I have adopted the latter description since it better captures the properties of MSA GO verb usage. The three GO verbs can either signal motion away from the speaker, or a non-deictic motion event in which the theme (moving entity) is not necessarily moving away from the deictic centre.

One of the excluded verbs, for instance, was the verb *sāra* سار which can sometimes be used interchangeably with GO verbs. This verb, however, seems to conflate path in its basic sense. *Sāra* can refer to the act of walking or movement along the ground, so the movement of people and vehicles can both be depicted by using this GO verb.¹¹ Another verb that some dictionaries would list as a GO verb is *inṭalaqa* انطلق, for which the root is *t-l-q*. This verb basically refers to the act of ‘releasing’, rather than to a deictic motion event. Structurally speaking, this verb is derived from combining the root *t-l-q* with Verb Form VII ‘inC₁aC₂aC₃a’, which is traditionally treated as the verb form that adds a reflexive sub-sense to the general root meaning. *Inṭalaqa* is often used in contexts where it can refer to an entity ‘taking off’ or an event that is just starting. Again, this verb could be used interchangeably with other GO verbs, but it does not fulfill the criteria of being a basic motion verb, with simple morphological structure (as explained in §2.1.2).

Among the verbs that dictionaries commonly refer to as COME verbs is the verb *aqbala* أقبل. Even though this verb has a deictic sense of physical or metaphorical motion towards a deictic centre, it appears to be restricted to contexts of use where the COME event is viewed as a positive, rather than negative, event. For instance, the use of this verb in the context in (2) is quite infelicitous, since ‘coming to the airport to say goodbye’ is not generally considered a positive or a joyous event. If we were to replace the verb *tuwaddi’ani* say.goodbye.SUBJN.ISG.ACC ‘say good bye to me’ with *tastaqbilani* welcome.SUBJN.ISG.ACC ‘welcome me’, the verb *aqbala* would definitely fit in the context in (2).

2.2 Data collection

2.2.1 The corpus

¹¹ This verb can still be used figuratively, especially in the expression *sāra ‘alā mā yurām*, Lit. ‘it went/moved according to what is desired’, which means ‘it went well’.

The Modern Standard Arabic data collected for this study was extracted from ArabiCorpus (arabicorpus.byu.edu), an online corpus developed by Dilworth Parkinson at Brigham Young University. As of October 2012, the corpus contained around 146,000,000 word tokens from different written and spoken genres. At the time of data collection (Fall 2010) the corpus contained around 69,000,000 word tokens. Additional MSA as well as pre-modern texts have been added to the corpus since the beginning of 2011, after the process of data collection had been completed. The written genres covered in ArabiCorpus include newspaper writing, pre-modern writing, modern literature, and nonfiction, in varying proportions with the newspaper writing accounting for over 90% of the total size of the entire corpus (with over 135,000,000 words tokens), and covering issues from 1996 to 2010. ArabiCorpus also includes a small sub-corpus of Egyptian colloquial usage extracted from online chat websites, a play, and an interview. For this study, the MSA sub-corpora that were queried for COME and GO uses are related to newspaper, modern literature, and nonfiction writing. As expected, most examples returned from corpus queries were in fact drawn from the newspaper genre.

ArabiCorpus is not tagged for parts-of-speech (POS) which makes the search for particular grammatical categories a daunting task. It does, however, provide a ‘filter’ function that identifies parts-of-speech based on rough prefix/suffix categories associated with different grammatical categories in Arabic. However, using orthographic regular expressions proved to be a more reliable corpus query method than using these filters, as I will explain shortly.

2.2.2 Corpus search

The absence of POS tagging in this corpus means that every individual inflected form of a verb must be manually queried. It was, therefore, necessary to rely on searches that employ regular expressions as a means of extracting the exact inflected forms for each

verb. The ArabiCorpus provides the user with different ‘filter’ functions: noun, adjective, adverb, verb, and string. Choosing a ‘verb’ filter for the verb *dahaba*, for instance, would return a number of tokens that include the root *d-h-b* attached to a variety of affixes that are generally considered to be related to verb derivation in MSA. In addition to the standard affixes associated with perfective and imperfective derivations of a verb, these affixes also include, for example, conjunction prefixes *wa-* ‘and’ and *fa-* ‘and/so/then’, the purposive and dative prefix *li-*, the future prefix *sa-*, as well as object pronoun suffixes. Unsurprisingly, the combination of these affixes with the stem may return lexical items that are not related to the inflected verb forms needed for this study.

What complicates the corpus querying process even further is the lack of short vowels in written texts of Modern Standard Arabic. Recall that there are three short vowels in Arabic – /ɛ/, /i/ and /u/ – which are not represented by letters from the Arabic alphabet, but by diacritics (commonly referred to as *ḥarakāt*, lit. ‘movements’) which are written over or underneath a letter as an indication of the short vowel sound following the written consonant. The diacritic َ (*fatḥa*) indicates the vowel /ɛ/, ِ (*kasra*) indicates the vowel /i/, and ُ (*ḍamma*) indicates the vowel /u/.¹² Diacritics are mostly found in the Qur’an and in children’s books, but are not present in newspaper writing or any other written form of MSA aimed at fluent readers. The vowel pattern of a word is retrievable from the surrounding context of a lexical item. For instance, the orthographic form ذهب, out of context, can be pronounced either as *dahaba* ‘GO.3SG.M.PERF’, or *dahab* ‘gold’. When supplied with the proper diacritics, these two lexical items look different: *dahaba* ذَهَبَ, and *dahab* دَهَبٌ. Again, these diacritics are redundant when the word is used contextually.

¹² There are other diacritics in the writing system of Arabic such as ّ (*šadda*) which signals gemination as well as ْ (*sukūn*) which indicates lack of a vowel sound following the consonant.

Since the texts fed into ArabiCorpus (or any other MSA corpus for that matter) lack these vowel patterns, we can expect that querying an unvowelized lexical item would return other unrelated forms. The search for the third person masculine perfective form of *ḍahaba* (ذهب) did yield many instances of *ḍahab*, in addition to other words bearing object pronoun clitics, such as *ḍahabuhu* ‘his gold’, *wadḍahabuha* ‘and her gold’, etc. Similarly, the search for most inflected forms of *qadima* yielded numerous instances of the verb *qaddama* ‘to bring forward, to present’, which is a causative verb derived from *qadima*. Obviously, it was necessary that these unrelated forms be manually filtered out.

Furthermore, the ‘filter’ function in ArabiCorpus only goes through prefixed and suffixed forms of the exact string that has been fed into the corpus query. This means that inflected forms of the verbs which include infixation, such as the active participle *ḍāhib* would not be part of the search results. Active participles are lexical forms derived from verbs in Arabic and have not been included in this study (as I will explain in §2.2.3) and therefore the lack of these lexical forms in the returned search was not problematic. A more relevant problem that arises when using the ‘verb’ filter function in ArabiCorpus is that feeding the lemmatized form of a highly frequent verb, like those being examined here, would considerably slow down the querying process and might cause instability in the browser. The use of regular expressions, therefore, proved to be a much reliable (though, time-consuming) method for extracting only the related inflected forms of the verbs under study, as well as breaking down the corpus query of a single verb into several steps.

Table 3 shows an example of the regular expressions (henceforth, ‘regexes’) used for extracting all inflected forms of the verb *ḍahaba* as well as the resulting forms.¹³ The regexes generated for the corpus queries included finite verb forms inflected in the

¹³ For corpus search, there is the option of using letters from the Arabic alphabet or Roman letters for the transliteration of the Arabic letters. Refer to <http://arabiccorpus.byu.edu/dt.html> for the full list of the DT transliteration system.

perfective and imperfective morphological aspects as well as the subjunctive, jussive, and imperative moods. These moods of the verb also had to be inflected for person, number and gender. Regexes for these four moods of the verb were paired with object pronoun suffixes (for the motion verbs that can be used transitively, such as some COME verbs) in addition certain prefixes such as the future tense *sa-* prefix; the purpose/intention-marking *li-* preposition; the permissive or hortative imperative *l-*; and the *la-* particle that precedes the result clause in a contrary-to-fact conditional (*law ... la-* ‘if... then’). As stated earlier, ArabiCorpus also automatically supplies conjunction clitics, such as *wa-* ‘and’ and *fa-* ‘and/then/so’. In the case of *ḍahaba*, as in Table 3, I ended up with over 60 inflected forms for the verb. Needless to say, the use of regular expressions still yielded certain unrelated forms that had to be discarded through manual inspection, though the number of unrelated forms was substantially lower than that returned by using search filters.

TABLE 3. Regular expressions used to query finite forms of the verb *ḍahaba*.

queried forms	regex	returned forms
PERFECTIVE	<code>vh(btmAlbtm btl btAlbt bAlb wAlbnAlbn b)</code>	<i>ḍahabtuma, ḍahabtum, ḍahabtunna, ḍahabatā, ḍahabat, ḍahabā, ḍahabū, ḍahabnā, ḍahabana, ḍahaba</i>
<i>la-</i> prefixed PERFECTIVE	<code>lvh(btmAlbtm btl btAlbt bAlbwAlbnAlbn b)</code>	<i>laḍahabtuma, laḍahabtum, laḍahabtunna, laḍahabatā, laḍahabat, laḍahabā, laḍahabū, laḍahabnā, laḍahabana, laḍahaba</i>
IMPERFECTIVE, JUSSIVE, SUBJUNCTIVE AND IMPERATIVE	<code>[ALEnty]vh(byn bAn bwn bn nlbAlbwAlby b)</code>	<i>taḍhabīn, taḍhabān, yaḍhabān, taḍhabūn, yaḍhabūn, taḍhabna, yaḍhabna, iḍhabna, taḍhabā, yaḍhabā, iḍhabā, iḍhabū, taḍhabū, yaḍhabū, iḍhabī, taḍhabī, aḍhab, iḍhab, naḍhab, taḍhab, yaḍhab</i>
<i>l-/li-</i> prefixed JUSSIVE and SUBJUNCTIVE	<code>l[ALnty]vh(byn bAn bwn bn bAlbwAlby b)</code>	<i>liyaḍhabna, litaḍhabna, litaḍhabā, liyaḍhabā, litaḍhabū, liyaḍhabū, litaḍhabī, liaḍhab, linaḍhab, litaḍhab, liyaḍhab</i>
<i>sa-</i> prefixed IMPERFECTIVE	<code>s[ALnty]vh(byn bAn bwn bn lbAlbwAlby b)</code>	<i>sataḍhabīn, sataḍhabān, sayaḍhabān, sataḍhabūn, sayaḍhabūn, sataḍhabna, sayaḍhabna, saāḍhab, sanaḍhab, sataḍhab, sayaḍhab</i>

2.2.3 Forms not included in the corpus search

2.2.3.1 Verbal nouns and active participles

The inflected forms I extracted from ArabiCorpus for the seven GO and COME verbs are limited to the morphological aspects/moods mentioned in §2.2.2: PERFECTIVE and IMPERFECTIVE aspects, and SUBJUNCTIVE, JUSSIVE and IMPERATIVE moods. Traditional treatments of the Arabic verb indicate three deverbal substantives that relate to the verb root in Arabic: (1) the verbal noun (المصدر *maṣḍar*), (2) the active participle (اسم فاعل *ism al-fā'il*), and (3) the passive participle (اسم مفعول *ism maf'ul*). The corpus-based analysis of GO and COME verbs in MSA does not include these three forms for practical reasons. The quantitative analysis (and the subsequent qualitative analysis) was restricted to a set of inflected *verb forms* that are comparable in their morphosyntactic and semantic features: PERFECTIVE and IMPERFECTIVE aspects, and SUBJUNCTIVE, JUSSIVE and IMPERATIVE moods. For the purposes of building a data frame – which I will elaborate on in §2.3 – introducing additional non-verbal forms would result in a substantially larger variable set (i.e. more variables to code each verb usage for) and would consequently require the examination of an even larger amount of data. Since the coding process for each of the corpus returns for each verb was entirely manual, I had to restrict myself to a limited set of variables, and hence, a limited set of forms to examine.

That being said, verbal nouns and active participles in particular, though excluded from further analysis here, are interesting forms and can associate with sub-senses and collocational patterns that diverge from those associated with the fully inflected verb they are derived from, as I will explain in Chapter 8. Generally speaking, the active participle can have different uses and meanings. According to Beeton, “[m]any words which have the pattern of a participle contain highly specialized senses within their semantic spectrum, in addition to the fundamental value” (1970:35). One example of the many uses of active participles is signaling two events that happen simultaneously, e.g. *walking around* and *coming from the opera*, as in the *qadima* sentence in (3).

(3) كنت أتجول مشياً قادماً من الأوبرا

<i>kuntu</i>	<i>ataḡawwalu</i>	<i>mašyan</i>	<i>qādiman</i>	<i>min</i>	<i>al=ūpera</i>
be.PERF.1SG	stroll.IMPF.1SG	walk.VN	qadima .AP.1SG	ABL	ART=opera
I was	stroll	walking	coming	from	the opera

‘I was strolling around on foot coming from the opera’

The usage of verbal nouns in MSA, for instance, can correspond to the use of non-finite forms of the verb in English as in (4) and (5).

(4) لم يستطع المجئ

<i>lam</i>	<i>yastaṭiʿ</i>	<i>al=maḡīʿ</i>
NEG	be.able.to.JUSS.3SG.M	ART= ḡāʿa .VN
did not	be able to	the coming

‘He couldn’t come’

(5) إذا أراد المجئ إلى هنا

<i>iḏa</i>	<i>arāda</i>	<i>al=maḡīʿ-a</i>	<i>ila</i>	<i>huna</i>
COND	want.PERF.3SG.M	ART= ḡāʿa .VN-ACC	ALL	ADV
if	wanted	the coming	to	here

‘If he wanted to come here’

Such (non-finite) nominal forms were not included in the corpus data examined in this study. However, the sentence in (5) can be paraphrased using a SUBJUNCTIVE form of a COME verb, as in (6).¹⁴

(6) إذا أراد أن يأتي إلى هنا

<i>iḏa</i>	<i>arāda</i>	<i>an</i>	<i>yaʿtiya</i>	<i>ila</i>	<i>huna</i>
COND	want.PERF.3SG.M	TOP	atā .SUBJN.3SG.M	ALL	ADV
if	wanted	to	come	to	here

‘If he wanted to come here’

Even though for (5) and (6), the English translation equivalent is roughly the same (‘if he wanted to come here’), sentences including the usage of the verbal noun in (5) are absent in the corpus data discussed in the following chapters.

¹⁴ According to Ryding (2005), the construction hosting *atā* in (6) is referred to as a ‘matrix verb’ in which the verb *arāda* ‘wanted’ is followed by the (TOPIC) particle *an* and a subjunctive form of the verb.

2.2.3.2 Imperative COME: ta'āla

The four COME verbs discussed in this study – *atā*, *ğā'a*, *ḥaḍara*, and *qadima* – appear very rarely in their imperative forms: '*i'ti*, *ği*', *iḥḍar* and *iqdim*, respectively. The coded 2000 lines of COME did not include any such imperative forms for either verb.

There is another verb in Arabic – تعال *ta'ala* – which stands in a suppletive relation with the other COME verbs and which is exclusively used as an imperative 'come'.

Structurally, this verb does not share a tri-consonantal root with any of the other COME verbs and is associated with Form VI of the root '-l-w 'be exalted'. Testen (1997) has argued that “[a]t some point in the past this verb, the literal meaning of which was presumably '*be exalted', seems to have been pressed into service in the encoding of politeness and deference on the part of the speaker” (1997:186), and eventually acquired the function of calling the addressee to approach the speaker.

As is the case with imperative COME forms in many languages, the imperative *ta'ala* can be used in physical as well as non-physical settings, as in (7) and (8).

(7) تعال بسرعة لدينا مفاجأة جميلة

<i>ta'āla</i>	<i>bi-sur'a</i>	<i>ladaynā</i>	<i>mufāğā'a</i>	<i>ğamīla</i>
come.IMPR	INST-speed	POSS	surprise	beautiful
come	quickly	we have	surprise	beautiful

'Come quickly! We have a beautiful surprise'

(8) إذن تعال بنا نتمنى السعادة

<i>iḍan</i>	<i>ta'āla</i>	<i>bi-nā</i>	<i>natamannā</i>	<i>al=sa'āda</i>
ADV	come.IMPR	COM-CL.1PL	wish.IMPF.1PL	ART=happiness
then	come	with us	we wish	the happiness

'Then come along! Let's hope for happiness'

In addition to the above deverbal substantives excluded from the quantitative analysis, I decided not to include the verb *ta'āl*, as well, since the main purpose of the analysis presented in this dissertation is to compare and contrast different potentially interchangeable GO and different COME verbs. Adding a suppletive imperative verb would not fulfill these objectives. Moreover, *ta'ala* does not count among the 5000 most

frequent lexical items in Arabic, as per Bukwalter and Parkinson's (2010) frequency dictionary.

2.3 Constructing data frames for GO and COME verbs

The analysis presented here for MSA GO and COME verbs is both quantitative and qualitative in nature. The quantitative analysis relies on constructing a data frame for every verb under investigation. Each data frame is typically composed of a large number of corpus concordance lines (500 concordance lines in this study), where a certain verb appears in its natural context of use. Subsequently, every concordance line is examined and marked up for a wide range of morphosyntactic and semantic features. This includes the syntactic structure, or construction, that hosts the verb, the patterns of verbal inflections for every instance of verb use (e.g. subject number, person, and gender, as well as other morphosyntactic aspects for the Arabic verb), the semantic properties of other elements of the construction (e.g. semantic properties of the subject), as well as the inclusion/exclusion of, for example, phrases denoting a starting point of the event (SOURCE), a terminal point of the event (GOAL), as well as specification of the PATH of motion, if present. Such a heavily annotated dataset can therefore be statistically explored by various tools including both monivariate and multivariate analyses. The quantitative approach to such a dataset will undoubtedly help define the specific characteristics of the constructions associated with the various meanings and functions of each MSA GO and COME verb involved in this study.

2.3.1 Selection of contextual features and the annotation of corpus hits

In order to construct a data frame where contextualized verb occurrence is annotated for a host of morphosyntactic and semantic features, the first step is to generate a list of features or variables that are relevant for the verbs in questions and that reflect

the morphosyntax of Modern Standard Arabic. Along the lines of Gries’s study on the polysemy of the English verb *run* (2006), Gries and Divjak’s (2006) investigation of Russian verbs of TRY, as well as Gries and Otani’s (2010) analysis of the synonymy and polysemy of adjectives of size in English, I developed a large set of morphological, syntactic, and semantic features that are relevant to the phenomenon at hand.

TABLE 4. A selection of variables GO and COME corpus hits were coded for.

category of variable	variable	levels
<i>morphological</i>	TENSE	PRESENT, PAST, FUTURE, IRREALIS (non-finite forms)
	ASPECT	SIMPLE, HABITUAL, PROGRESSIVE, PERFECT, INCHOATIVE, NON-FIN (non-finite forms)
	MORPHOLOGICAL ASPECT AND MOOD OF THE VERB	IMPERFECTIVE, PERFECTIVE, SUBJUNCTIVE, JUSSIVE, IMPERATIVE
	SUBJECT PERSON	1 ST , 2 ND , 3 RD
	SUBJECT NUMBER	SINGULAR, DUAL, PLURAL
	SUBJECT GENDER	FEMININE, MASCULINE, NIL (for 1 st person inflections)
	<i>syntactic</i>	TRANSITIVITY
INTERROGATIVE		YES, NO
NEGATIVE		YES, NO
PREPOSITIONAL PHRASE		YES, NO
LOCATIVE ADVERB PHRASE		YES, NO
ADVERBIAL PHRASE		YES, NO
SERIAL VERB CONSTRUCTION		YES, NO
<i>semantic</i>		SUBJECT CATEGORY
	GOAL PHRASE	YES, NO
	SOURCE PHRASE	YES, NO
	MANNER PHRASE	YES, NO
	SETTING PHRASE	YES, NO
	PATH PHRASE	YES, NO
	PURPOSIVE PHRASE	YES, NO
	COMITATIVE PHRASE	YES, NO
	TEMPORAL PHRASE	YES, NO
	DEGREE PHRASE	YES, NO

Table 4 shows the dichotomous and non-dichotomous nominal variables that MSA GO and COME verbs were coded for, and the ‘levels’ within each variable (e.g. YES/NO for GOAL, or SINGULAR/DUAL/PLURAL for NUMBER).¹⁵ In Appendix D I provide examples and illustrations of the different annotations of levels within each variable. This set of 22 linguistic features, or variables, was motivated primarily by certain lexico-syntactic properties that pertain to a deictic motion event schema, such as phrases specifying a GOAL and/or a SOURCE of the motion event, in addition to MANNER of motion and the inclusion of a COMITATIVE phrase (i.e. accompaniment by an object/individual in the GO or COME event). Each verb usage was also coded for the semantic category of the subject or the moving entity involved in the motion event, such as HUMAN, OBJECT/ARTIFACT, or more abstract/non-physical entities such as EVENT, COMMUNICATION (i.e. a statement), COGNITION (i.e. an idea), etc. The morphosyntactic component of the list of features in Table 4 reflects the inflectional properties of the MSA verb (MORPHOLOGICAL ASPECT AND MOOD, NUMBER, PERSON, and GENDER) as well as the TENSE and ASPECT of the construction hosting GO and COME verbs. The variable labeled TRANSITIVITY, only pertains to certain uses of COME verbs in MSA where COME verbs can appear in transitive constructions in which the direct object is the GOAL of the motion event as in (9).

(9) كانت البداية صعبة اذ لم يأتها أي زبون

<i>kānat</i>	<i>al=bidaya</i>	<i>ša'ba</i>	<i>id</i>	<i>lam</i>	<i>ya'ti-ha</i>
be.PERF.3SG.F	ART=beginning	hard	ADV	NEG	<i>atā</i> .JUSS.3SG.M-CL.3SG.F.ACC
was	the beginning	hard	since	did not	come to her

ayyu zabun
any customer
any customer
‘the beginning was hard, since no customer came to her’

¹⁵ The data frame was, in fact, coded for more variables than the set laid out in Table 4, such as the different morphosyntactic realizations of GOAL, SOURCE, MANNER, etc., as well as certain recurring lexical elements (e.g. adverbs, adverbial uses, and other lexical items). These additional variables do not form part of the quantitative analysis. Nevertheless, they are of some interest and will be referred to occasionally later in the qualitative analysis.

Text genre was not considered a variable since the majority of the annotated 3500 corpus hits belong to the genre of newspaper writing. The following results should consequently be considered as mostly reflective of the usage of GO and COME verbs in newspaper writing. Sentences (10) and (11) are sample uses that feature the COME verbs *atā* and *qadima*, respectively. Table 5 shows a sub-set of the variables which the two usages in (10) and (11) were coded for. Appendix E provides more extensive annotation examples for a selection of sentences from the GO and COME data frame.

(10) ويأتي الرد سريعا وبنبيرة عالية

wa=ya'ti *al=radd* *sari'an* *wa=bi=nabra* *'āliya*
 CONJ=*atā*.IMPF.3SG.M ART=response quickly CONJ=INST=pitch high
 and comes the response quickly and with pitch high
 'and the response **comes** quickly in a high pitch'

(11) وكذلك تحدثت عن مدرسته ومكتبها التي عمل فيها علماء قدموا من أماكن مختلفة

wa=kaḍalik *tahaddaṭat* *'an* *madrasati-h* *wa=maktabati-ha* *allati*
 CONJ=also talk.PERF.3SG.F about school-CL.3SG.M CONJ=library.CL.3SG.F RP
 and also talked about his school and its library that

'amila *fī-ha* *'ulamā'* *qadimu* *min* *amākin* *muḥtalifa*
 WORK.PERF.3SG.M LOC-CL.3SG.F scholars *qadima*.PERF.3PL.M ABL places different
 worked in it scholars came from places different
 'it also talked about its school and its library where scholars who **came** from different places have worked'

TABLE 5. Sample of annotation from the COME data frame for selected variables.

	VERB	TENSE	MORPH_ASP. MOOD	SUBJ_ NUM	SUBJ_ CAT	ADVER- PP	SOU- BIAL	MAN- RCE	NER
10	<i>atā</i>	PRES- ENT	IMPER- PECTIVE	SING- ULAR	COMMUNI- CATION	NO	YES	NO	YES
11	<i>qadima</i>	PAST	PERFE	PL	HUMAN	YES	NO	YES	NO

2.4 Statistical analyses

As explained in the preceding section, 500 random sentences per verb were annotated resulting in a three-verb GO data frame consisting of 1500 lines and a four-verb COME data frame consisting of 2000 lines. In most of the following statistical analyses, the independent variable is the motion VERB, while the dependent variables are the

contextual features each verb usage was coded for. The dependent variables are mostly categorical (binary) in nature, for which the presence or absence of a feature is indicated either as YES or NO; while the remaining variables are nominal (e.g. TENSE, ASPECT, SUBJECT SEMANTIC CATEGORY, etc.) with three or more levels (e.g. TENSE has four levels: PRESENT, PAST, FUTURE, IRREALIS). These data frames were then loaded into R (www.r-project.org) for the purpose of statistical analysis. The command `attach(COME);str(COME)` yields the summary of the entire COME data frame, with regard to the number of variables and the number of levels within each variable. Such a summary is provided in (12).

(12) COME data frame loaded into R

```
> attach(COME); str(COME)
'data.frame': 2000 obs. of 23 variables:
 $ VERB      : Factor w/ 4 levels "ATE","HDR","JAC",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ TENSE     : Factor w/ 4 levels "FUT","IRR","PAST",...: 3 1 2 2 4 3 2 4 2 1 ...
 $ ASPECT    : Factor w/ 6 levels "HAB","NON-FIN",...: 6 6 3 3 2 6 3 2 3 6 ...
 $ MORPH_ASP.MOOD: Factor w/ 4 levels "IMPF","JUSS",...: 3 1 1 1 1 2 4 1 4 1 ...
 $ TRANSITIVITY : Factor w/ 2 levels "NO","YES": 1 2 1 1 1 1 1 1 1 1 ...
 $ SUBJ_NUM  : Factor w/ 3 levels "DUAL","PL","SING": 3 3 3 3 3 3 3 3 3 3 ...
 $ SUBJ_PER  : Factor w/ 3 levels "1ST","2ND","3RD": 3 3 3 3 3 3 3 3 3 3 ...
 $ SUBJ_GEN  : Factor w/ 3 levels "FEM","MASC","NIL": 2 2 1 2 1 1 2 2 2 2 ...
 $ SUBJ_CAT  : Factor w/ 20 levels "ACTIVITY","ANIMAL",...: 20 20 20 20 20 20 20 ...
 $ INTEROG   : Factor w/ 2 levels "NO","YES": 1 1 1 1 1 1 1 1 1 1 ...
 $ NEGATION  : Factor w/ 2 levels "NO","YES": 1 1 1 1 1 2 1 1 1 1 ...
 $ PP        : Factor w/ 2 levels "NO","YES": 1 1 2 1 1 2 1 1 1 1 ...
 $ LOC_ADV   : Factor w/ 2 levels "NO","YES": 2 1 1 1 2 1 1 2 1 2 ...
 $ ADVERBIAL : Factor w/ 2 levels "NO","YES": 1 1 1 1 2 1 1 1 1 1 ...
 $ GOAL      : Factor w/ 2 levels "NO","YES": 1 2 1 1 1 2 1 1 1 1 ...
 $ SOURCE    : Factor w/ 2 levels "NO","YES": 1 1 1 1 1 1 1 1 1 1 ...
 $ MANNER    : Factor w/ 2 levels "NO","YES": 1 1 1 1 2 1 1 1 1 1 ...
 $ SETTING   : Factor w/ 2 levels "NO","YES": 1 1 1 1 1 1 1 1 1 1 ...
 $ PATH      : Factor w/ 2 levels "NO","YES": 1 1 1 1 1 1 1 1 1 1 ...
 $ PURPOSIVE : Factor w/ 2 levels "NO","YES": 1 1 2 1 1 1 1 1 1 1 ...
 $ COMITATIVE : Factor w/ 2 levels "NO","YES": 2 1 1 1 1 2 1 1 1 1 ...
 $ TEMPORAL  : Factor w/ 2 levels "NO","YES": 1 1 1 1 2 1 1 2 1 2 ...
 $ DEGREE    : Factor w/ 2 levels "NO","YES": 1 1 1 1 1 1 1 1 1 1 ...
```

A wide variety of statistical analyses are available in order to explore such multifactorial data frames for different purposes. The statistical analyses I propose here should help serve the following objectives in the examination of at least three statistical properties. Firstly, the analysis should help us better understand the distribution of contextual elements in the overall data frame. This will provide a preliminary glimpse into the skewed distributional patterns within, for instance, inflectional paradigms. Secondly, the analyses should reveal the distribution of contextual elements per each GO

and COME verb as a first step towards identifying divergence in usage patterns associated with each MSA motion verb. This will eventually lead us to the examination of a third property – the interaction patterns among the contextual features and the identification of clusters of features that are closely tied to certain verb uses. This important step facilitates the identification of prototypical uses of each verb as well as the less prototypical uses.

The statistical analyses I will introduce below start with basic univariate *chi-square tests* and then move on to multivariate statistical approaches that vary in their complexity and highlight different patterns and aspects of interaction between variables. Respectively, these statistical tests include: hierarchical agglomerative cluster analysis, hierarchical configural frequency analysis, and polytomous logistic regression analysis.

2.4.1 *Chi-square tests*

2.4.1.1 *Chi-square goodness-of-fit test*

The most basic statistical approach for exploring the COME and GO data frames is to examine the distribution of the contextual features across the entire data frame for each verb in a set. That is to say, if we examine the distribution of the variable TENSE, for instance, across the four COME verbs, we would want to know whether the different levels of TENSE – PRESENT, PAST, FUTURE, IRREALIS – have the same frequencies across the entire data frame or whether the distribution of the different tenses would be skewed.

If we consider that all 2000 lines of coded COME verbs might provide a representative sampling of COME verb usage in MSA, and similarly that all 1500 lines of coded GO verbs represent GO verb usage more generally in MSA, we can assume that the data (for COME and GO data frames separately) only consists of dependent variables and no independent variable. The null hypothesis and (non-directional) alternative hypothesis – H_0 and H_1 , respectively – that our distributional exploration is tied to could be formulated in the following:

H₀: *The frequencies of the different levels of all variables across a verb set (GO or COME) are identical and any difference in frequencies in the data frame would be the result of random variation.*

H₁: *The frequencies of the different levels of all variables are not identical across a verb set.*

A *chi*-square goodness-of-fit test would require, first of all, a tabulation of the observed frequencies and a computation of the expected frequencies based on the null hypothesis. For instance, if we want to conduct a *chi*-square analysis on TENSE for the COME data frame (for all COME verbs combined), the observed versus expected frequencies for each level within TENSE are listed in Table 6.

TABLE 6. Observed vs. expected frequencies of the levels of the variable TENSE.

TENSE	Observed freq.	Expected freq.
PRESENT	372	500
PAST	1396	500
FUTURE	53	500
IRREALIS	179	500

Given the above hypotheses, the null hypothesis postulates that observed frequencies should be equal to expected frequencies. As we can see in Table 6, this is far from being the case. In order to conduct a *chi*-square test, 80% of the expected frequencies should be larger than or equal to 5 and all expected frequencies should be larger than 1 (Gries, 2009). Since the frequencies reported in Table 6 fulfill the minimum requirements, we can now conduct a *chi*-square test. If conducted manually, the procedure will involve computing the contributions to *chi*-square for each observed frequency, then summing these values to get the test statistic *chi*-square and determining the degrees of freedom and the probability of error. However, in R, A `chisq.test()` command can calculate these values automatically for the data reported in Table 6, as shown in (13).

(13) *Chi*-square goodness-of- fit test for the variable TENSE across all COME verbs.

```
Chi-squared test for given probabilities
data: TENSE
X-squared = 2244.1, df = 3, p-value < 2.2e-16
```

According to this *chi*-square goodness-of-fit test, the distribution of the four levels of TENSE deviates significantly from the expected distribution ($X^2 = 2244.1$, $df = 3$, $p_{\text{two-tailed}} < 0.001$). In other words, there is statistical evidence that the variable TENSE interacts with the lexical choice of COME verb in MSA.

2.4.1.2 Chi-square test for independence

In addition to examining single variable distribution as per the goodness-of-fit *chi*-square test explained in §2.4.1.1, a number of statistical analyses presented in this dissertation will be exploring the distribution of the dependent variables (i.e. different constructional elements) per independent variable (i.e. MSA motion verb), as a means of highlighting the differences in usage across the four COME verbs, and the three GO verbs under study here. The null and alternative hypotheses for this kind of analysis, therefore, postulate the following:

- H₀: *The frequencies of the different levels of the dependent variables do not vary as a function of the different VERBs.*
- H₁: *The frequencies of the different levels of the dependent variables vary as a function of the different VERBs.*

These hypotheses are based on the examination of, for instance, the distribution of TENSE (PRESENT, PAST, FUTURE, IRREALIS) per VERB (*atā*, *ǧa'a*, *ḥaḍara*, *qadima*, in the case of COME verbs). The question such a statistical test attempts to answer is, “do all COME verbs share a similar behavior with respect to the frequencies of TENSE inflections on the verb or do they differ among each other in that respect?”. To test this set of hypotheses –where we have an independent variable and a dependent variable – the procedure is similar to that conducted in §2.4.1.1 for the goodness-of-fit *chi*-square test.

First of all, we need to cross-tabulate the distribution of different levels of the dependent variable for each COME verb (observed frequencies only), as shown in Table 7.

TABLE 7. Observed frequencies of the levels of the variable TENSE per COME verb.

TENSE VERB	FUTURE	IRREALIS	PAST	PRESENT
<i>atā</i>	14	94	81	311
<i>ḥaḍara</i>	38	63	347	52
<i>ḡā'a</i>	0	11	485	4
<i>qadima</i>	1	11	483	5

In R, we can run the function `chisq.test()` on the cross-tabulation of observed frequencies (as given in Table 7), which yields the results in (14).

(14) *Chi-square test for independence for the variable TENSE per COME verb.*

```
Pearson's Chi-squared test
data: verb.tense
X-squared = 1191.187, df = 9, p-value < 2.2e-16
```

This *chi-square* test for independence indicates that the distribution of the four levels of TENSE for each COME verb deviates highly significantly from the expected distribution ($X^2 = 1191.187$, $df = 9$, $p\text{-value} < 2.2e\ 16$). We may also retrieve the expected cell-wise frequencies for this cross-tabulation, as shown in Table 8, through the function `chisq.test()$expected` in R.

TABLE 8. Expected frequencies of the levels of the variable TENSE per COME verb.

TENSE VERB	FUTURE	IRREALIS	PAST	PRESENT
<i>atā</i>	13.25	44.75	349	93
<i>ḥaḍara</i>	13.25	44.75	349	93
<i>ḡā'a</i>	13.25	44.75	349	93
<i>qadima</i>	13.25	44.75	349	93

2.4.1.3 Standardized Pearson's residuals

The two *chi-square* tests discussed in the two previous sections do not provide information about the statistical significance of the individual cell-wise contributions to *chi-square* or whether a certain cell-wise observed frequency is significantly higher or lower than expected. A variety of methods can be employed in order to assess both

significance and direction of individual cell contributions to *chi*-square (i.e. higher or lower than expected). One way is to calculate the cell-wise $p_{\text{corrected}}$ -values for individual cells with $df=1$, a procedure which characterizes *configural frequency analysis* tests. I will return to this point in §2.4.3. Another way on which I will rely in the univariate analysis of GO and COME verbs also builds on *chi*-square tests as a means of calculating standardized Pearson's residuals for each cell in order to assess its significance. The formula for calculating these residuals is given in (15) (from Arppe, 2008:83, quoting Agresti, 2002, formula 3.13):

(15)
$$e_{ij}/\text{standardized Pearson residual} = (O_{ij} - E_{ij}) / [E_{ij} \cdot (1 - R_i/N) \cdot (1 - C_j/N)]^{1/2}$$
 Where i and j are the row and column indices, I and J are the number of rows and columns,
 R_i and C_j are the row and column marginal totals, respectively, and N is the overall total.

Generally, when the *standardized Pearson's residual* for a certain cell exceeds 2 or is less than -2, the cell-wise deviation is considered to be statistically significant. The threshold of 2 (in absolute value) may be increased for larger contingency tables. However, this particular threshold (of ≥ 2 or ≤ -2) has not been challenged in the relevant literature (Arppe, 2008). The R function `chisq.test()$std` can help us retrieve these cell-wise values, as shown in Table 9.

TABLE 9. Standardized Pearson's residuals for the distribution of the variable TENSE per COME verb.

TENSE VERB	FUTURE	IRREALIS	PAST	PRESENT
<i>atā</i>	0.2411319	8.909208	-30.1431044	28.931582
<i>ḥadara</i>	7.9573536	3.301382	-0.2249485	-5.441261
<i>ḡā'a</i>	-4.2599974	-6.105295	15.2965007	-11.811518
<i>qadima</i>	-3.9384881	-6.105295	15.0715522	-11.678804

In §2.4.4, I will discuss the *polytomous logistic regressions analysis* (Arppe, 2008), which is one of the multivariate statistical methods I have adopted for the study of GO and COME verbs. Arppe (2011) has developed an R statistical package, {polytomous},

that combines a number of functions that render the analysis easier and reduces the number of steps in order to obtain the results. One of the functions added to the {polytomous} package pertains to conducting post-hoc *chi*-square tests. The R function `chisq.posthoc()` can yield several test statistics including the standardized Pearson's residuals as we have seen in Table 14. Furthermore, it provides a simplified version of Table 9 in which residuals higher than 2 (i.e. significantly higher than expected) are represented by a '+' (plus) sign; residuals lower than -2 (i.e. significantly less than expected) are represented by a '-' (minus) sign; and the residuals falling between -2 and 2 (which are considered insignificant values) are represented by a '0' (zero), as shown in Table 10. This table can also be retrieved by the R function `chisq.posthoc()$cells$std.pearson.residuals.sign`.

TABLE 10. Standardized Pearson's residuals for the distribution of the variable TENSE per COME verb.

TENSE VERB	FUTURE	IRREALIS	PAST	PRESENT
<i>atā</i>	0	+	-	+
<i>ḥaḍara</i>	+	+	0	-
<i>ḡā'a</i>	-	-	+	-
<i>qadima</i>	-	-	+	-

In Chapters 3 and 5, I will discuss the quantitative analysis of GO and COME verbs with *chi*-square tests of goodness-of-fit or independence, as well as the standardized Pearson's residuals. These tests are a first attempt at understanding the distributional patterns of selected variables among the different verbs. Such univariate analyses will undoubtedly set the stage for the more complex multifactorial analyses that will follow and to which I turn next.

2.4.2 Hierarchical Agglomerative Cluster Analysis

Before proceeding to discuss multivariate methods for examining specific interactions of variables (and levels of variables), we can examine the joint effect on the

overall verbal behavior for each verb in the GO and COME verb set. The method I will rely on in the subsequent analysis is referred to as Hierarchical agglomerative cluster analysis. Generally speaking, this clustering method groups together the lexical elements that are most similar to one another and, at the same time, the ones that are highly dissimilar to other elements in other clusters. Therefore, what we expect to see from this statistical method is a clustering dendrogram that shows us which COME verbs, for instance, overlap in their usage as opposed to the ones with which they hardly share any characteristics.

The Hierarchical agglomerative cluster analysis has been advanced by Stefan Gries and colleagues (e.g. Gries, 2006; Diviaj and Gries, 2006; Gries and Otani, 2010) in what he labeled the *Behavioral Profiles* (BP) method for studying synonymous, antonymous, and polysemous lexical items. The term *Behavioral Profiles* was introduced by Hanks (1996) in his investigation of *urge*, looking at patterns of collocations (the co-occurrence of lexical items with other lexical items) and colligations (the co-occurrence of lexical items with grammatical elements). His claim was that “the semantics of a verb are determined by the totality of its complementation patterns” (1996:77). The ‘behavioral profile’ of a lexical item is, therefore, determined by such co-occurrence patterns. Gries and colleagues elaborated on this analytical approach in their corpus-based examination of the various semantic phenomena, such as the polysemy of the verb *run* in English (Gries, 2006), the near-synonymy of Russian *try* verbs (Divjak and Gries, 2006), the synonymy and antonymy of size adjectives in English (Gries and Otani, 2010), among many other studies in lexical semantics.

This method requires generating a table that lists relative frequencies (or proportions) of co-occurrence values of dependent variables per independent variable (the GO and COME verbs under study). A similarity/dissimilarity matrix is first computed followed by

computing a cluster structure based on a specific amalgamation rule.¹⁶ The resulting cluster structure can then be visually represented in a dendrogram. The calculations involved in the different stages of Hierarchical agglomerative cluster analysis have been made easier to conduct using BP 1.01 script, a program written by Stefan Gries (2009) for R.

This R-based script uses a host of statistical methods required in the stages mentioned above. It initially generates a co-occurrence table of relative frequencies of the different levels (IDTAG-LEVELS) within variables (IDTAGS).¹⁷ Table 11 shows a sample of such output table generated by BP 1.01 regarding the distribution of TENSE by COME verb. The behavioral profile of a verb in this table is, therefore, the vector of co-occurrence proportions of ID tags per verb.

TABLE 11. Sample of a co-occurrence table generated by the BP 1.01 script for the variable TENSE by COME verb.

IDTAG	IDTAG-LEVEL	<i>atā</i>	<i>ḥaḍara</i>	<i>ǧā'a</i>	<i>qadima</i>	
TENSE	FUT	0.028	0.076	0	0.002	} columns sum to 1.0
	IRREALIS	0.188	0.126	0.022	0.022	
	PAST	0.162	0.694	0.97	0.966	
	PRES	0.622	0.104	0.008	0.01	

The BP 1.01 script returns a comprehensive table with similar values for all dependent by independent variable co-occurrences that have been fed into the script. This particular table can be subjected to a number of tests including the Hierarchical agglomerative cluster analysis as well as post-hoc pair-wise analysis on items that have been grouped together in one cluster or in different clusters. I will elaborate on the former method later Chapters 3 and 5.

¹⁶ An amalgamation rule is what determines whether or not two items are sufficiently similar in order to be linked or clustered together.

¹⁷ The idea of an ID tag was introduced by Atkins (1987) in her work on *danger*, where she examined collocates, colligations, POS, as well as other characteristics of the key word. An ID tag was therefore used to refer to the individual contextual features co-occurring with the keyword.

2.4.3 (Hierarchical) Configural Frequency Analysis

I mentioned in §2.4.1.3 that cell-wise contributions to *chi*-square can be tested for significance, either by calculating standardized Pearson's residuals, or by calculating cell-wise $p_{\text{corrected}}$ -values for individual cells with $df=1$. This latter method comprises the main procedure in conducting a *Configural Frequency Analysis* or CFA (von Eye, 1990).

Unlike univariate standardized Pearson's residuals, however, CFA examines configurations of (or interactions between) variables and assigns significance values for the co-occurrence of two or more (levels of) variables. This test allows us to examine combinations of variables such as:

VERB x TENSE *or* VERB x ASPECT *or* VERB x MORPHOLOGICAL ASPECT AND MOOD
VERB x TENSE x ASPECT
VERB x TENSE x ASPECT x MORPHOLOGICAL ASPECT AND MOOD
VERB x TENSE x ASPECT x MORPHOLOGICAL ASPECT AND MOOD x SUBJECT
NUMBER, etc.

For instance, if we decide to examine the significant interactions between the variables VERB x TENSE x ASPECT, a CFA test would go through all possible interactions between the different levels within each variable, such as:

atā x PRESENT x SIMPLE
atā x PRESENT x PROGRESSIVE
atā x PAST x SIMPLE
atā x PAST x PROGRESSIVE
qadima x PRESENT x PERFECT
qadima x PRESENT x SIMPLE
ğā'a x PAST x SIMPLE
hađara x FUTURE x PROGRESSIVE, etc.

For this particular interaction between COME verbs and TENSE and ASPECT, we would end up with 80 pairwise configurations: 4 (levels of VERB) x 4 (levels of TENSE) x 5 (levels of ASPECT). A CFA test tells us which of these 'configurations' of variables would occur more or fewer times than expected. This kind of analysis can help us zero in on various kinds of constructional elements that tend to co-occur frequently in the usage of a certain verb.

A CFA test is relatively simpler conceptually than the more complex multivariate statistical analyses, such as regression modeling. The basic steps for running a CFA test involve: (1) tabulating the observed frequencies, (2) calculating contributions to *chi*-square, and (3) calculating $p_{\text{corrected}}$ -values for the contribution to *chi*-square for $df = 1$ (Gries, 2009). For data frames such as the GO and COME ones constructed for this study, running individual CFA tests on each possible combination of variables would turn into a tedious and time-consuming job. As a way of automating the procedure, Gries (2004) has created an interactive R-based script – HCFA 2.3– that can run through every conceivable combination of variables. That is to say, if we feed the variable set VERB x TENSE x ASPECT x MORPHOLOGICAL ASPECT AND MOOD x SUBJECT NUMBER in an HCFA test, HCFA 2.3 will run through all possible combinations of variables, such as

VERB x TENSE
TENSE x MORPHOLOGICAL ASPECT AND MOOD
VERB x SUBJECT NUMBER
TENSE x MORPHOLOGICAL ASPECT AND MOOD x SUBJECT NUMBER
VERB x TENSE x ASPECT x MORPHOLOGICAL ASPECT AND MOOD x SUBJECT
NUMBER
and so on.

HCFA is, therefore, basically an exhaustive, “hierarchical” variant of *Configural Frequency Analysis*. Ideally, there is no limit on the number of combinations of variables HCFA 2.3 can process at a time. Yet, for practical reasons, the user of the script needs to limit herself/himself to a small subset of variables to run in order to avoid technical problems (because the more variables added, the more configuration tables need to be generated), as well as to come up with easily interpretable results. Table 12 is an excerpt from the overall output table generated by HCFA 2.3 for the variables COME VERB x TENSE x ASPECT x MORPHOLOGICAL ASPECT AND MOOD.

TABLE 12. A sample output hierarchy table generated by the HCFA script.

VERB	TENSE	ASPECT	MORPH	Freq	Exp	Cont.chisq	Obs-exp	P.adj.Holm	Dec	Q
			ASP.MOOD							
<i>ǧa'a</i>	PAST	SIMPLE	PERF	484	201.6032	395.569	>	2.91E-71	***	0.157
<i>atā</i>	PRES	HAB	IMPF	105	1.5722	6804.1949	>	2.16E-147	***	0.052
<i>qadima</i>	PAST	SIMPLE	IMPF	0	70.7126	70.7126	<	2.02E-29	***	0.037
<i>haḍara</i>	PAST	HAB	PERF	1	16.8206	14.88	<	0.000288	***	0.008

Typically, the first few columns in an HCFA output table contain the variables for which interactions between their different levels are evaluated. The remaining columns report the actual frequency of occurrence for each configuration (Freq), as opposed to the expected frequency (Exp). In addition, values of contributions to *chi*-square are reported (Cont.chisq). The adjusted Holm *p*-value is reported under ‘P.adj.Holm’ and the additional column ‘Dec’ indicates the level of statistical significance for the occurrence of each configuration (which is based on the adjusted Holm *p*-value)¹⁸. The ‘Dec’ column, paired with the column ‘Obs(erved)-exp(ected)’ gives us an indication of whether the observed frequencies are significantly higher or lower than expected. We can see in Table 12, for instance, that the configuration *ǧa'a* x PAST x SIMPLE x PERFECTIVE occurs significantly more than expected, and is referred to in CFA tests as a *type*. The configuration *haḍara* x PAST x HABITUAL x PERFECTIVE, on the other hand, occurs significantly fewer times than expected and is referred to as an *antitype*. The *Q* value reported in the HCFA output table expresses the “degree of pronouncedness” or significance of the configurations of values. It is independent of sample size (note that the contribution to *chi*-square value changes as *N* changes) and is calculated using the following simple formula in (16)¹⁹.

$$(16) \quad Q = (\text{observed frequency} - \text{expected frequency}) / (N - \text{expected frequency})$$

¹⁸ The adjusted Holm *p*-value is the *p*-value of the observed frequency given the expected frequency, according to (adjusted) *p*-values based on the binomial distribution. For a detailed account of adjusted Holm *p*-value, see von Eye, 1990.

¹⁹ This is a slightly simplified version of the original formula.

In sum, to calculate the Q value for the configuration $\check{g}a'a$ X PAST X SIMPLE X PERFECTIVE, we have $(484 - 201.60) / (2000 - 201.60) = 0.157$, where $N = 2000$ annotated corpus hits. Note that the Q value lies between 0 and 1, with 0 being lack of pronouncedness and 1 being perfect pronouncedness (Krauth and Lienert, 1995:33-34). In the subsequent HCFA analysis of MSA GO and COME verbs, I will be reporting mainly on significant *types* and will therefore be reporting the ‘Dec’ and ‘Q’ columns as assessment measures of the significance and entrenchment of configurations.

2.4.4 Polytomous Logistic Regression Analysis

Moving to more complex multivariate statistical analyses, the method known as “Polytomous Logistic Regression” has also been referred to in the literature as ‘multinomial’, ‘polychotomous’, ‘multiple-class’, and ‘discrete-choice’ logistic regression. It specifically applies to cases of linguistic alternation in which the possible alternatives are ‘multiple’ (e.g. the four COME verbs and three GO verbs) rather than binary. Logistic regression is a kind of regression analysis that employs a binary logistic function (which takes a nominal data set with binary variables, i.e. TRUE/FALSE or 0/1 values) and considers all outcomes as proportions for all observations with the same context, rather than considering each outcome as having a dichotomous distribution (i.e. either occurring in a context or not occurring). The outcome of such an analysis is a set of probability estimates that aim to predict the possible outcome of single trials which are being modeled as a function of a set of explanatory variables.²⁰ Polytomous logistic regression analysis is, therefore, compatible with a probabilistic view of language (Bresnan, 2006, 2007; Arppe, 2008).

²⁰ For a detailed description of this method, see Arppe, 2008 and Han, Arppe, and Newman (in press).

Applying this method to the GO and COME verbs means that the probability of the occurrence of each verb is calculated for each sentence in the data frame. For instance, each COME verb receives a probability value of occurrence in every one of the 2000 annotated corpus returns of the four COME verbs, given the explanatory variables included in the model. Polytomous logistic regression analysis aims to build upon the preceding multivariate analyses in that it allows us to systematically examine the actual contexts of use in which the usage of two or more COME or GO verbs are interchangeable, and the contexts where verb selection is almost categorical (i.e. only one verb is allowed per context). This analysis would, therefore, contribute to our goal of identifying prototypical uses of each of the seven verbs examined here.

The main focus of this part of the quantitative analysis of GO and COME verbs is to arrive at a sound model which estimates variable parameters that can be interpreted as ‘odds’ (Harrell, 2001:218). That is to say, for each predictor variable (i.e. constructional element), the model assigns a value that indicates the extent to which the existence of that constructional element (e.g. PRESENT TENSE) increases or decreases the chances of the occurrence of the outcome variable (i.e. verb), with all other variables being equal.

Following the procedure discussed in detail in Arppe (2008:187-248), this kind of analysis would require several steps of monivariate and bivariate analyses in order to select the appropriate predictor variables to include in the polytomous logistic regression model. As mentioned earlier, logistic regression analysis requires a binomial data set. Consequently, the current nominal form of the GO and COME data frames would not be suitable for this analysis and should, therefore, be converted into a logical form where every level of variable is considered a variable in its own right, with the binary values of TRUE/FALSE indicating its presence or absence from the annotated context. For instance, in a logical data frame, instead of having a single variable TENSE with four different levels (PRESENT, PAST, FUTURE and IRREALIS), we would have four variables:

TENSE.PRESENT, TENSE.PAST, TENSE.FUTURE, and TENSE.IRREALIS. Naturally, the original 20+ variable set in a nominal data frame is expanded into an 80+ variable set in a logical data frame. Of course, not all 80+ variables can be used as predictors in the logistic regression model and so a sub-set of variables need to be selected in order to be incorporated in the *polytomous logistic regression model*. The larger the amount of predictor variables, the higher the risk is of over-fitting the data (i.e. not being able to arrive at a general model that can account for the majority of data points). I will elaborate more on the selection on variables in Chapters 3 and 5. Generally speaking, however, the procedure for determining the set of predictors (or independent variables) to be included in the model involves the following steps:

- i. Inspect the distribution of variables across all GO and COME verbs using standardized Pearson's residuals. This monivariate analysis gives us a clear indication of which variables seem to have explanatory potential as opposed to those that do not. For instance, variables with a standardized Pearson's residuals value approaching 0 would not be included in the polytomous logistic regression model.
- ii. Inspect pair-wise association patterns between variables. That is to say, examine the extent to which certain variables have a high rate of co-occurrence (e.g. TENSE.PAST and MORPH_ASPECT.MOOD.PERFECTIVE). Only one of the two highly co-occurring pair of variables would therefore be selected for the model. This would be one way of reducing collinearity in the regression model which can be triggered by high association (either positive or negative) between variables (Arppe, 2008). Another more straightforward method to reduce collinearity is to identify variables that are symmetrically complementary and avoid including both in the model. An example of such variables would be GOAL.YES and GOAL.NO, each of which is indicated by the values TRUE/FALSE. For such originally binary variables – e.g. SOURCE, MANNER, NEGATION, PP, etc. – the logical complementary variable bearing the '.NO' suffix have been avoided in the selection of variables. In other words, we only need to keep the variable GOAL.YES with the values TRUE/FALSE, and avoid adding the complementary variable GOAL.NO (also with the values TRUE/FALSE).

As I will explain in Chapters 3 and 5, the resulting variable set is run through the polytomous logistic regression model. For both sets of GO and COME verbs, I will examine the calculated odds for the variables that increase or decrease that chances of the occurrence of one verb over the others. The aggregate effect of these odds is interpreted

as the probability estimates for the occurrence of each verb in an annotated context. I will then turn to discuss a selection of sentences for which the verb choice is either categorical (i.e. potentially prototypical uses of a verb) or equiprobable among two or more verbs (i.e. potential contexts where more than one verb can be used interchangeably).

In the following chapters, I will report on the quantitative and qualitative analyses of motion verbs in MSA starting with GO verbs in Chapters 3 and 4, respectively, then I will move on to discussing the quantitative and qualitative analyses of COME verbs in Chapters 5 and 6. As mentioned earlier, the findings reported in Chapters 3 and 5 are primarily obtained from exploring the GO and COME data frames through various statistical analyses. The qualitative analysis presented in Chapters 4 and 6 also makes reference to the data frames and draws on observations obtained from the quantitative chapters.

Chapter Three

Quantitative analysis of MSA GO verbs *ḍahaba*, *maḍā*, and *rāḥa*

The quantitative analysis presented in this chapter is based on an examination of the constructed GO data frame where 500 random concordance lines for each of the three MSA GO verbs have been annotated for a wide range of morphological, syntactic, and semantic variables. The present analysis is concerned with identifying patterns of verbal behaviour through the inspection of (i) single variable distribution, as in §3.2.1, as well as (ii) interactions between multiple variables and the emerging prototypical patterns of verb usage, as in §3.2.2-4 where different multivariate statistical analyses will be discussed. Hierarchical agglomerative cluster analysis constitutes the first step towards examining the joint effect of multiple variables on the clustering of the three verbs as a means of visualizing the ‘closeness’ between the three verbs in terms of the constructional features for which they were annotated. The subsequent hierarchical configural frequency analysis shifts attention towards particular clusters of variables and the robust patterns of interaction found among different levels of different variables. This analysis can help us identify typical sets of features that characterize the use of each verb, such as morphosyntactic and semantic features. In §3.2.4, I will present a polytomous logistic regression analysis where the three GO verbs (outcomes) are modeled as a function of a subset of constructional features (predictors). Among the results obtained from this analysis are probability estimates of the occurrence of the three GO verbs in each of the 1,500 annotated corpus hits. The benefits of this analysis go beyond extracting exemplar sentences per verb into possibly predicting the contexts of use where more than one verb is likely to occur. The variables for which GO verb uses were coded are summarized again in Table 1. The statistical analysis which follows is based on these 23 nominal and binary variables.

TABLE 1. A list of the variables for which corpus sentences for the three MSA GO verbs were coded.

Variable	Levels
VERB (independent variable)	<i>ḡahaba, maḡā, rāḡha</i>
TENSE	PRESENT, PAST, FUTURE, IRREALIS (non-finite forms)
ASPECT	SIMPLE, HABITUAL, PROGRESSIVE, PERFECT, INCEPTIVE, NON-FINITE
MORPHOLOGICAL ASPECT AND MOOD	IMPERFECTIVE, PERFECTIVE, SUBJUNCTIVE, JUSSIVE, IMPERATIVE
SUBJECT NUMBER	1 ST , 2 ND , 3 RD
SUBJECT PERSON	SINGULAR, DUAL, PLURAL
SUBJECT GENDER	FEMININE, MASCULINE, NIL (for 1 st person inflections)
SUBJECT CATEGORY	ACTIVITY, ANIMAL, ATTRIBUTE, BODY, COGNITION, COMMUNICATION, CONTENT (of a document/speech), DEMONSTRATIVE, DUMMY SUBJECT, EVENT, GROUP, HUMAN, LOCATION, NOTION, OBJECT, SENSE, STATE, SUBSTANCE, TIME
INTERROGATION	YES, NO
NEGATION	YES, NO
SERIAL VERB CONSTRUCTION	YES, NO
PREPOSITIONAL PHRASE/COMPLEMENT	YES, NO
LOCATIVE ADVERB	YES, NO
ADVERBIAL	YES, NO
GOAL	YES, NO
SOURCE	YES, NO
MANNER	YES, NO
SETTING	YES, NO
PATH	YES, NO
PURPOSIVE	YES, NO
COMITATIVE	YES, NO
TEMPORAL	YES, NO
DEGREE	YES, NO

3.1 Single variable distribution

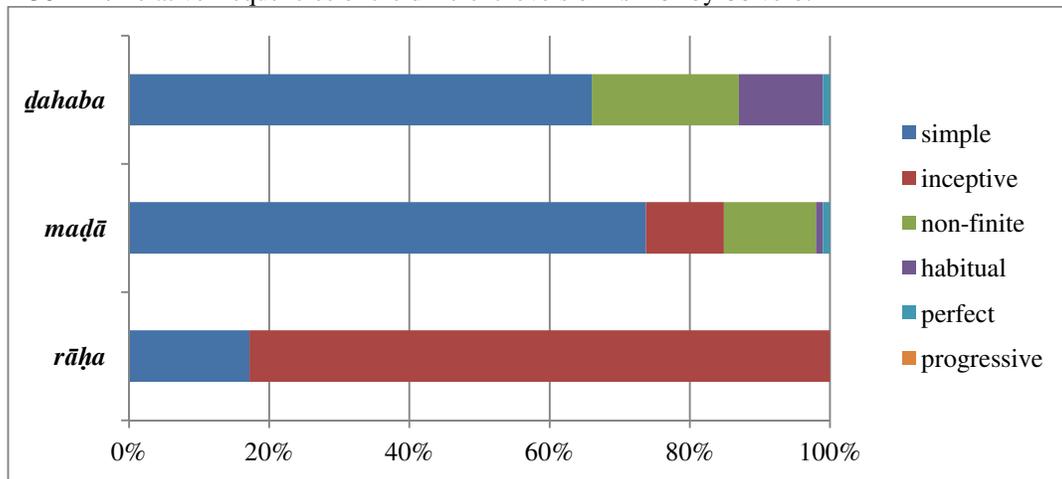
Prior to examining the degree of interaction between different sets of variables, an examination of the proportional frequencies of selected variables can give us a first glimpse into just how the heterogeneous (i.e. skewed) the distribution of a single variable across the three verbs is. We can start with the variable ASPECT as an example. Table 2 is a cross-tabulation of the different levels of the variable ASPECT X VERB, with proportional frequencies of occurrence adding up to 1.0 per verb.

TABLE 2. Proportional frequencies of the different levels of ASPECT by GO verb.

ASPECT	<i>ḍahaba</i>	<i>maḍā</i>	<i>rāḥa</i>
SIMPLE	0.66	0.73	0.17
INCEPTIVE	0.00	0.11	0.82
NON-FINITE	0.21	0.13	0.00
HABITUAL	0.12	0.01	0.00
PERFECT	0.01	0.01	0.00
PROGRESSIVE	0.00	0.00	0.00
<i>total</i>	1	1	1

Figure 1 plots the relative frequencies presented in Table 1.

FIGURE 1. Relative frequencies of the different levels of ASPECT by GO verb.



We can see that the distribution of these levels of variables per verb is in no way homogenous (i.e. evenly distributed). It is clear from both Table 2 and Figure 1, for instance, that both *ḍahaba* and *maḍā* are more likely to occur in the SIMPLE aspect. *Rāḥa*, on the other hand, occurs more than 80% of the time in INCEPTIVE constructions. We can also see that around 10% of the use of *maḍā* is also in the INCEPTIVE, which will be an interesting point for discussion later in the next chapter when I compare the grammaticalized functions of *maḍā* and *rāḥa*. In addition, Table 2 shows that HABITUAL aspect is only a characteristic of the use of the verb *ḍahaba*, while PERFECT and PROGRESSIVE aspects do not seem to have an interestingly skewed distributional pattern

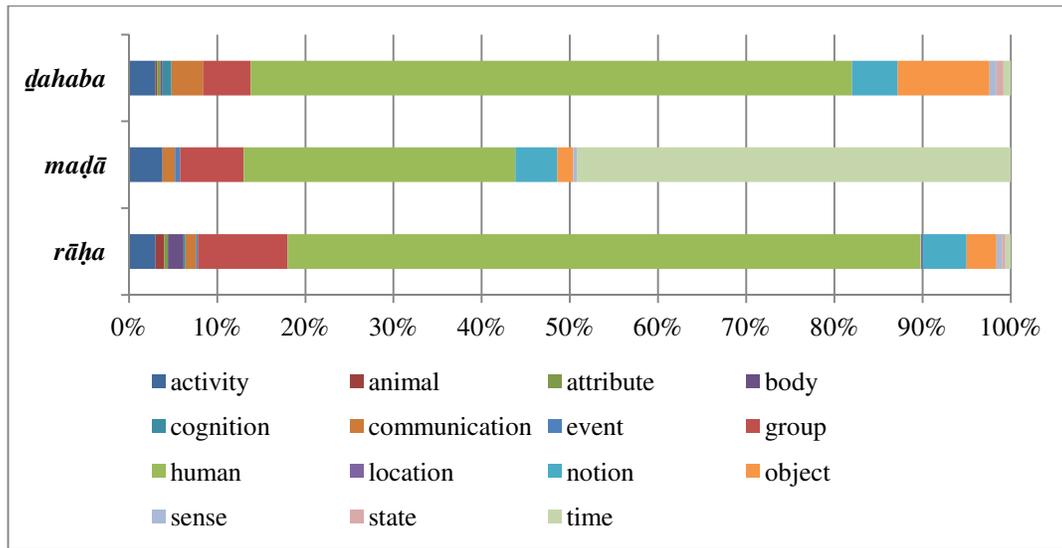
across the three verbs, most likely due to their very low frequency of occurrence in the overall GO data frame (PERFECT: 11 times, PROGRESSIVE: 4 times out of the 1500 GO sentences).

Another example of variable distribution is that of SUBJECT SEMANTIC CATEGORY, where the nature of the sentential subject is indicated as being either HUMAN, PHYSICAL OBJECT/ARTIFACT, EVENT, TIME, etc. Table 3 lists the proportional frequencies while Figure 2 shows the relative frequencies for the different levels of this variable. The distribution of SUBJECT SEMANTIC CATEGORY, like ASPECT, is also far from homogenous. *Ḍahaba* and *rāḥa* both appear to favor HUMAN subjects, while *maḍā* associates mostly with TIME-related subjects and, to a lesser extent, with HUMAN subjects. As far as the remaining categories of this variable are concerned, it appears that each verb can collocate with different types of sentential subjects, such as GROUP, NOTION, OBJECT and ACTIVITY, but to varying degrees.

TABLE 3. Proportional frequencies of the different levels of SUBJECT SEMANTIC CATEGORY by GO verb.

SUBJECT SEMANTIC CATEGORY	<i>ḍahaba</i>	<i>maḍā</i>	<i>rāḥa</i>
ACTIVITY	0.03	0.04	0.03
ANIMAL	0.00	0.00	0.01
ATTRIBUTE	0.00	0.00	0.00
BODY	0.00	0.00	0.02
COGNITION	0.01	0.00	0.00
COMMUNICATION	0.04	0.01	0.01
EVENT	0.00	0.01	0.00
GROUP	0.05	0.07	0.10
HUMAN	0.68	0.31	0.72
LOCATION	0.00	0.00	0.00
NOTION	0.05	0.05	0.05
OBJECT	0.10	0.02	0.03
SENSE	0.01	0.00	0.01
STATE	0.01	0.00	0.00
TIME	0.01	0.49	0.01
<i>total</i>	1	1	1

FIGURE 2. Relative frequencies of the different levels of SUBJECT SEMANTIC CATEGORY by GO verb.



An additional measure to test for divergence in cell-wise values is a *chi*-square test, which indicates whether the observed cell-wise values significantly diverge from the expected values. Recall that in order to run such a test one of the basic conditions that should be met is that the expected values for each cell should not be less than 5 occurrences (Cocheran, 1954). The distributions in Table 3 (given as percentages of 1) do not meet that condition to run a *chi*-square test of significance. Nevertheless, looking at the proportional frequencies, it is quite clear that these examples of single variable distribution show a great deal of divergence among the individual cells across the different levels of variables, as well as across the three GO verbs. We can, of course, run a *chi*-square test on a variable distribution that meets the above condition, such as the occurrence of a GOAL phrase in the GO motion event. Table 4 shows the observed versus expected values for the occurrence of GOAL (YES) as opposed to the lack thereof (NO) for each GO verb.

TABLE 4. Expected vs. observed values for the variable GOAL by GO verb.

VERB	YES (GOAL)		NO (GOAL)	
	EXP. FREQ.	OBS. FREQ.	EXP. FREQ.	OBS. FREQ.
<i>ḡahaba</i>	110.3333	298	389.6667	202
<i>maḡā</i>	110.3333	32	389.6667	468
<i>rāḡa</i>	110.3333	1	389.6667	499

The calculated Pearson’s *chi*-square test for the distribution given in Table 4 proved to be quite significant: $X^2 = 277.1034$, $df = 6$, $p\text{-value} < 2.2e-16$.²¹ We can also examine the cell-wise divergences from a uniform distribution for this particular contingency table by conducting a standardized Pearson’s residual (discussed in Agresti 2002: 81; Arppe, 2008: 83-84). These test statistics can either be retrieved in R by using the command `chisq.test()$std` or by running the function `chisq.posthoc()`, which is part of the statistical package {*polytomous*} developed by Antti Arppe (2012). Table 5 contains the calculated values, which indicate whether the observed co-occurrence frequency reported in each individual cell is significantly more or less than expected.²² The `chisq.posthoc()` function presents an easier way to interpret these figures, in that it assigns $+/-/0$ values for each cell, which can be interpreted as insignificant (0), significantly more than expected (+), or significantly less than expected (–).

TABLE 5. Standardized Pearson’s residuals for the occurrence of GOAL by GO verb.

VERB	YES (GOAL)	NO (GOAL)
<i>ḡahaba</i>	24.78665 (+)	-24.78665 (–)
<i>maḡā</i>	-10.34611 (–)	10.34611 (+)
<i>rāḡa</i>	-14.44053 (–)	14.44053 (+)

The polytomous logistic regression analysis discussed in §3.2.4 includes options for a “univariate” analysis based on these standardized Pearson’s residuals. A series of useful functions available at the {*polytomous*} R package (Arppe, 2012) can display

²¹ Note that the *chi*-square test was run on absolute rather than relative frequencies in this test and the following tests in both Chapters 3 and 5.

²² Typically, the standardized Pearson’s residual value is significantly higher than what is expected when it is > 2.0 , and significantly lower than expected when the value is < -2.0 .

standardized Pearson’s residuals for levels of all variables in the GO data frame in one table. Such a procedure calls for converting the current nominal format of the data frame into a logical format, whereby every level of a variable is turned into a variable in its own right with the categorical values of TRUE and FALSE. For instance, the variable TENSE with its four levels (PRESENT, PAST, FUTURE and IRR) is turned into four separate variables (TENSE.PRESENT, TENSE.PAST, TENSE.FUTURE, and TENSE.IRR) and each has the binary values of TRUE/FALSE indicating its presence or absence from the context. Table 6 shows the preferences for the distribution (0/+/- values) for a selection of logical variables (with the value TRUE) based on their calculated standardized Pearson’s residuals. The full table generated is presented in Appendix F. Note that the listed variables also conform to the condition required in a *chi*-square test, where the expected values of occurrence are not less than 5.

TABLE 6. Preferences for the distribution of selected logical variables by GO verb.

FEATURE	<i>dahaba</i>	<i>maḍā</i>	<i>rāḥa</i>
TENSE.PAST	–	–	+
TENSE.PRES	+	+	–
ASPECT.HABITUAL	+	–	–
ASPECT.INCP	–	–	+
MORPH_ASP.MOOD.IMPF	+	+	–
MORPH_ASP.MOOD.PERF	+	+	–
SUBJ_CAT.HUMAN	+	–	+
SUBJ_CAT.TIME	–	+	–
PP.YES	+	+	–
NEGATION.YES	+	+	–
MANNER.YES	0	0	0
PATH.YES	0	+	–

Table 6 provides more evidence that, unsurprisingly, the distribution of constructional elements across the three GO verbs is far from homogeneous. The subsequent analysis, however, goes beyond examining single variable distribution to, rather, determining the cumulative effect of all variables or subsets of variables.

3.2 Hierarchical agglomerative cluster analysis

Moving from single variable distribution, we can now investigate the overall effect resulting from the heterogeneous distribution of variables discussed in the previous sections on the clustering of the three GO verbs. In Chapter 2, I discussed the Behavioral Profiles method for clustering synonymous and polysemous lexical items (Gries, 2006; Gries and Divjak, 2006; Gries and Otani, 2010, among others) that employs hierarchical agglomerative cluster analysis as a means of calculating and visualizing the overall conceptual “distance” between the three GO verbs. This method of analysis takes into account all the different (levels of) variables that each verb was coded for.

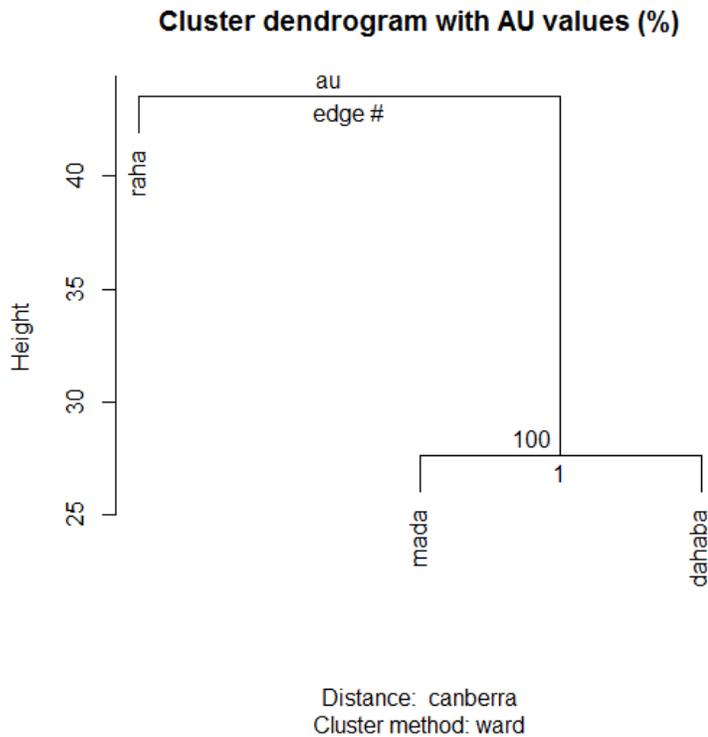
TABLE 7. Sample of co-occurrence table generated by the BP 1.01 script.

IDTAG	IDTAG-LEVEL	<i>dahaba</i>	<i>maḍā</i>	<i>rāḥa</i>	
MORPH_ASP/MOOD	IMPF	0.36	0.266	0.048	} column subset sums to 1.0
	IMPR	0.014	0.008	0	
	JUSS	0.052	0.05	0.002	
	PERF	0.46	0.612	0.95	
	SUBJN	0.114	0.064	0	
SUBJ_NUM	DUAL	0.008	0.002	0.012	} column subset sums to 1.0
	PL	0.126	0.064	0.098	
	SING	0.866	0.934	0.89	
GOAL	NO	0.404	0.936	0.998	} column subset sums to 1.0
	YES	0.596	0.064	0.002	

As discussed in Chapter 2, the cluster analysis can be conducted with the help of the BP 1.01 script developed by Stefan Gries (2009) for R. This particular statistical script takes a multivariate data frame with multiple vectors of annotation and, as a first step, generates a list of proportions of co-occurrences between a dependent variable (here, the three GO verbs) and a set of independent variables (the various constructional elements), similar to the cross-tabulations in Tables 2 and 3 that we saw earlier. The BP 1.01 script refers to these variables as “ID tags” and to the different levels of these nominal or categorical variables as “ID tag level”. Table 7 shows an excerpt of the overall co-occurrence table generated by this script. The hierarchical agglomerative cluster analysis I am reporting on here applies the (dis)similarity metric of ‘Canberra’ and the

amalgamation rule ‘Ward’ for the computation of cluster structure. The result is a dendrogram as in Figure 3.

FIGURE 3. Dendrogram based on the of GO multivariate data frame.



This dendrogram shows a clear divide between the verb pair *mada* and *dahaba* and the verb *raha*. As discussed earlier, this clustering technique groups together elements that are more similar to one another and at the same time dissimilar to other elements in other clusters. The BP 1.01 script employs the pvclust statistical package (Suzuki and Shimodaira, 2006) which assigns Approximately Unbiased (AU) values to each cluster as a kind of measure of reliability of a cluster. This particular measure is based on performing multiscale bootstrap resampling in order to calculate the *p*-values found for each cluster in a dendrogram. Here we find that the AU *p*-value for the *mada* and *dahaba* cluster is 100%. The implication here is that these two verbs are very similar to one another in terms of usage (as reflected in the set of variables selected for coding) and at the same time they differ dramatically from the usage of the verb *raha*. The

following sections on hierarchical configural frequency analysis and polytomous logistic regression analysis will further show (i) the extent to which these verbs overlap and differ in their constructional characteristics from one another, as well as (ii) the type of constructional elements each individual verb typically associates with.

3.3 Hierarchical configural frequency analysis

So far, we found ample evidence to reject the null hypothesis which was based on the assumption that the different levels of variables are evenly distributed within each variable per verb, and that the three GO verbs show similar behaviour with respect to every (level of) variable. The following hierarchical configural frequency analysis examines significant combinations of variables that are characteristic of the use of each GO verb.

As alluded to earlier, even though there is no limit on the number of variables run through the interactive R script hcfa 3.2 (Gries, 2004), we still need to limit ourselves to examining a restricted amount of variables at a time for practical and conceptual reasons. Since the HCFA analysis examines every possible combination of variables, the number of generated tables consequently increases dramatically the more variables are included in the test. Trying to run all variables at once would definitely cause certain technical problems, let alone generating hundreds of thousands of results that may not necessarily be easy to interpret. Therefore, the procedure I followed in conducting HCFA tests on the GO data frame involved breaking down the entire variable set into different sub-groups and subsequently regrouping certain variables from one set with other variables. This method of reporting on HCFA findings proved to be consistent with the assumption that the different constructional elements associated with each verb are interlinked rather than working individually, and that in order to understand the distributional pattern of one variable per verb, we have to explain its distribution with regards to other variables.

Morphological types and antitypes

Prior to examining different interactions between variables, I decided to investigate the most significant morphological features that characterize the use of MSA GO verbs (mostly in newspaper writing) as well as the morphological features that GO verbs in an MSA corpus hardly associate with. The idea behind this test is to provide evidence that elements within an inflectional paradigm are not evenly distributed in actual usage (Newman and Rice, 2004, 2006a; among others). More specifically, I am interested in examining which particular elements belonging to the inflectional categories of TENSE, ASPECT, MORPHOLOGICAL ASPECT AND MOOD, SUBJECT NUMBER, PERSON and GENDER are mostly represented in the usage of MSA GO verbs, irrespective of verb at this point. An HCFA test including all 6 variables was conducted, resulting in the list of *types* in Table 8 and *antitypes* in Table 9.

TABLE 8. Most significant univariate *types* for TENSE, ASPECT, MORPHOLOGICAL ASPECT AND MOOD, SUBJECT NUMBER, PERSON and GENDER for GO verbs.

TENSE	ASPECT	MORPH_ASP. MOOD	SUBJ_ NUM	SUBJ_ PER	SUBJ_ GEN	Freq	Exp	Ob-Ex	Dec	Q
.	.	.	SING	.	.	1345	500	>	***	0.845
.	.	.	.	3RD	.	1323	500	>	***	0.823
.	.	PERF	.	.	.	1011	300	>	***	0.592
PAST	1016	375	>	***	0.57
.	MASC	934	500	>	***	0.434
.	SIMPLE	777	250	>	***	0.422
.	INCP	464	250	>	***	0.171

TABLE 9. Most significant univariate *antitypes* for TENSE, ASPECT, MORPHOLOGICAL ASPECT AND MOOD, SUBJECT NUMBER, PERSON and GENDER for GO verbs.

TENSE	ASPECT	MORPH_ASP. MOOD	SUBJ_ NUM	SUBJ_ PER	SUBJ_ GEN	Freq	Exp	Ob-Ex	Dec	Q
.	.	.	DUAL	.	.	11	500	<	***	0.489
.	.	.	.	2 ND	.	30	500	<	***	0.47
.	.	.	PL	.	.	144	500	<	***	0.356
.	.	.	.	1 ST	.	147	500	<	***	0.353
.	NIL	150	500	<	***	0.35
FUT	40	375	<	***	0.298
.	.	IMPR	.	.	.	13	300	<	***	0.239
.	.	JUSS	.	.	.	50	300	<	***	0.208
.	PROG	4	250	<	***	0.197
.	PERT	11	250	<	***	0.191
IRR	176	375	<	***	0.177
.	.	SUBJN	.	.	.	89	300	<	***	0.176
.	HAB	69	250	<	***	0.145

In Tables 8 and 9, the first few columns contain the variables for which the (monovariate) distribution is reported. The remaining columns contain the following statistics: the actual frequency of occurrence as opposed to the expected frequency and whether the observed frequencies are higher or lower than expected (in the column labeled ‘Ob-Ex’).²³ In addition, the column ‘Dec’ provides us with decisions regarding the statistical significance of each configuration and, finally, the *Q* value reported in the above HCFA output tables expresses the “degree of pronouncedness” or significance of the configurations of values. It is independent of sample size (unlike the contribution to *chi*-square value that changes as *N* changes)

As far as the results of morphological *types* and *antitypes* are concerned, there is a noticeably skewed distribution of (i) SUBJECT NUMBER, where the subject marking on the verbs is predominantly SINGULAR, and much less so DUAL and PLURAL; (ii) 3RD

²³ Two additional columns that have been omitted from these two tables and the following HCFA output tables pertain to the calculated contribution to *chi*-square per configuration, as well as another column that reports on the adjusted Holm *p*-value, which the “Dec” measure is based on, and which were discussed in Chapter 2. I decided to report only on the “Dec” measure of significance of configuration in these tables as well as the *Q* value as measures of assessing the robustness of each configuration.

PERSON is most likely to be the person inflection on GO verbs with very low frequency of 1ST and 2ND; and (iii) a bias towards MASCULINE gender inflection. As far as TENSE and MORPHOLOGICAL ASPECT AND MOOD marking are concerned, we find the same general pattern of frequent use of PAST SIMPLE tense realized with PERFECTIVE inflection. On the other hand, FUTURE TENSE, PROGRESSIVE, PERFECT, HABITUAL ASPECTS, as well AS IMPERFECTIVE, JUSSIVE and SUBJUNCTIVE MORPHOLOGICAL ASPECTS/MOODS count as *antitypes* – occurring significantly fewer times than expected.

As discussed before, in the following multivariate HCFA analysis, I will be examining significant combinations between these morphological variables as well as other kinds of variables the corpus data was coded for. The overall objective of the following analysis is to show that the examination of a combination of variables, rather than one variable at a time, can provide us with a better understanding of the larger morphosyntactic frames or *constructions* hosting the lexical items under study. This analysis will also show that the three GO verbs have different preferences with regards to the morphosyntactic features they typically associate with.

VERB x TENSE x ASPECT x MORPHOLOGICAL ASPECT AND MOOD

The single variable distribution discussed in the previous section manifested a generally skewed preference towards certain individual features in the Arabic verb inflectional paradigm (e.g. PAST, SIMPLE/INCEPTIVE, PERFECTIVE). It is safe to assume, however, that each individual GO verb would show its own profile of TENSE/ASPECT/MORPHOLOGICAL ASPECT or MOOD marking (TAM) that sets the usage of each of the three verbs apart. Since the categories of TENSE, ASPECT, and MOOD are often discussed in relation to one another, Table 10 reports on the significant configurations regarding these three variables per each GO verb.

TABLE 10. Most significant configurations for TENSE, ASPECT, and MORPHOLOGICAL ASPECT and MOOD for all GO verbs.

VERB	TENSE	ASPECT	MORPH_ASP.		Freq	Exp	Ob-Ex	Dec	Q
			MOOD						
<i>ḍahaba</i>	PAST	SIMPLE	PERF		217	71.5171	>	***	0.34
<i>ḍahaba</i>	PRES	SIMPLE	IMPF		92	32.5901	>	***	0.127
<i>ḍahaba</i>	IRR	NON-FIN	SUBJN		57	2.5618	>	***	0.109
<i>ḍahaba</i>	PRES	HAB	IMPF		46	5.9616	>	***	0.081
<i>ḍahaba</i>	IRR	NON-FIN	JUSS		25	1.1236	>	***	0.048
<i>ḍahaba</i>	FUT	SIMPLE	IMPF		19	4.487	>	***	0.029
<i>ḍahaba</i>	IRR	NON-FIN	IMPR		8	0.3596	>	***	0.015
<i>maḍā</i>	PAST	SIMPLE	PERF		248	135.9595	>	***	0.308
<i>maḍā</i>	PRES	SIMPLE	IMPF		92	20.2772	>	***	0.15
<i>maḍā</i>	IRR	NON-FIN	SUBJN		32	0.5832	>	***	0.063
<i>maḍā</i>	PAST	INCP	PERF		49	20.5999	>	***	0.059
<i>maḍā</i>	IRR	NON-FIN	JUSS		25	0.4556	>	***	0.049
<i>maḍā</i>	FUT	SIMPLE	IMPF		21	4.0554	>	***	0.034
<i>maḍā</i>	PRES	HAB	IMPF		5	0.391	>	**	0.009
<i>maḍā</i>	IRR	NON-FIN	IMPR		4	0.0729	>	***	0.008
<i>rāḥa</i>	PAST	INCP	PERF		397	366.6696	>	*	0.083
<i>rāḥa</i>	PRES	SIMPLE	IMPF		10	0.2064	>	***	0.02
<i>rāḥa</i>	PRES	INCP	IMPF		10	0.9792	>	***	0.018

In general, it seems that the verbs *ḍahaba* and *maḍā* associate with a wider range of TAM marking than the verb *rāḥa*, in addition to the fact that the former two verbs overlap to a great extent in their preferred TAM patterns. For instance, both *ḍahaba* and *maḍā* are most likely to occur in the SIMPLE PAST, with PERFECTIVE inflection. Both verbs are also likely to appear in SIMPLE PRESENT (sometimes also HABITUAL) constructions with IMPERFECTIVE inflection, as well as non-finite constructions where the verb takes SUBJUNCTIVE, JUSSIVE, or IMPERATIVE inflection. One noticeable difference is the inclination of *maḍā* to function as an inceptive/durative marker, which seems to be an overlapping pattern of use with that of *rāḥa*.

As far as *rāḥa* is concerned, it is no surprise that the vast majority of this verb's coded corpus hits relate to the grammaticalized use of *rāḥa* as an inceptive marker, since this verb is almost fully grammaticalized in MSA. Interestingly, we can see that 397

(79.4%) of the inceptive constructions involving *rāḥa* are in the PAST tense with IMPERFECTIVE inflection, while only 10 (2%) of the hits are in the PRESENT tense, with IMPERFECTIVE marking, which seems to indicate a preference towards PAST events. The uses of *rāḥa*, however, do not seem to be restricted to marking aspect, which explains the 10 instances of verb use in the SIMPLE PRESENT IMPERFECTIVE. As will be discussed in the following chapter, most uses of *rāḥa* uses are idiomatic, with the exception of a few physical deictic motion uses.

VERB x SUBJECT NUMBER x PERSON x GENDER x SEMANTIC CATEGORY

Another interaction worth examining is one concerning the properties of the sentential subject that collocates with the MSA GO verbs in the annotated contextual uses. Such properties are reflected in the SUBJECT NUMBER, PERSON, and GENDER agreement marking on the verb, in addition to the SEMANTIC CATEGORY of the subject, i.e. whether the subject denotes a human agent or a non-human agent such as OBJECT, COMMUNICATION, NOTION, etc. The results of the HCFA analysis involving all five variables including VERB are reported in Table 11.

TABLE 11. Most significant configurations for SUBJECT NUMBER, PERSON, GENDER, and SEMANTIC CATEGORY for all GO verbs.

VERB	SUBJ_ NUM	SUBJ_ PER	SUBJ_ GEN	SUBJ_ CAT	Freq	Exp	Ob- Ex	Dec	Q
<i>ḍahaba</i>	SING	3RD	MASC	HUMAN	189	140.1819	>	*	0.036
<i>ḍahaba</i>	SING	1ST	NIL	HUMAN	47	2.5015	>	***	0.03
<i>ḍahaba</i>	SING	3RD	FEM	OBJECT	33	5.7026	>	***	0.018
<i>ḍahaba</i>	PL	1ST	NIL	HUMAN	26	0.2678	>	***	0.017
<i>ḍahaba</i>	SING	3RD	FEM	NOTION	18	5.4833	>	*	0.008
<i>ḍahaba</i>	SING	3RD	FEM	COMMU- NICATION	13	2.2664	>	***	0.007
<i>ḍahaba</i>	SING	3RD	FEM	HUMAN	27	62.4365	<	***	0.025
<i>maḍā</i>	SING	3RD	MASC	TIME	178	41.5293	>	***	0.094
<i>maḍā</i>	SING	3RD	FEM	TIME	67	18.497	>	***	0.033
<i>maḍā</i>	SING	3RD	FEM	GROUP	26	8.3346	>	***	0.012
<i>maḍā</i>	PL	1ST	NIL	HUMAN	18	0.2678	>	***	0.012
<i>maḍā</i>	SING	1ST	NIL	HUMAN	11	2.5015	>	ms	0.006
<i>maḍā</i>	SING	3RD	FEM	HUMAN	12	62.4365	<	***	0.035
<i>maḍā</i>	SING	3RD	MASC	HUMAN	97	140.1819	<	*	0.032
<i>rāḥa</i>	SING	3RD	MASC	HUMAN	233	140.1819	>	***	0.068
<i>rāḥa</i>	SING	3RD	FEM	GROUP	43	8.3346	>	***	0.023
<i>rāḥa</i>	SING	1ST	NIL	HUMAN	34	2.5015	>	***	0.021
<i>rāḥa</i>	PL	3RD	MASC	HUMAN	38	15.0083	>	***	0.015
<i>rāḥa</i>	PL	1ST	NIL	HUMAN	9	0.2678	>	***	0.006

Our single variable-based analysis in §4.2.1 showed that each individual verb collocates with numerous categories of sentential subject, but to varying degrees. Table 11, however, shows that for larger constructional units involving not only SUBJECT SEMANTIC CATEGORY, but also SUBJECT NUMBER, PERSON and GENDER, it appears that *ḍahaba* associates with a wider range of sentential subjects than do *maḍā* and *rāḥa*. That is to say, Table 10 shows that *maḍā* typically associates with subjects related to TIME and, to a lesser degree, HUMAN agents; whereas *rāḥa* typically associates with HUMAN agents and subjects denoting GROUP (e.g. organization, country, news agency, etc.). On the other hand, *ḍahaba* appears to be the multi-purpose GO verb in this set. In addition to human agents, *ḍahaba* also collocates with abstract entities such as NOTION and COMMUNICATION as well as non-human OBJECTS. In general, with regard to the most

typical and frequent configurations involving SUBJECT PERSON, NUMBER, and GENDER agreement on these verbs, all three verbs seem to typically bear a 3RD SINGULAR MASCULINE agreement (as was previously observed in Table 7).

As far as SINGULAR vs. PLURAL marking is concerned, the abundance of SINGULAR inflection seen here could be symptomatic of word order properties, most particularly in the case of HUMAN subjects: the Arabic verb bears plural agreement only when the sentential subject is human and the word order in the sentence is SV. As for GENDER agreement on the verb, we see for instance that one of the significant configurations involving *dahaba* (27 occurrences) relate to HUMAN agents that are females. The other *dahaba* configurations that also involve FEMININE agreement on the verb relate to non-human subjects such as NOTION, OBJECT, and COMMUNICATION, etc. The gender agreement here may indicate either a subject that is singular and grammatically feminine or a plural non-human noun (masculine or feminine) which in MSA is treated syntactically as a singular feminine noun.

Another interesting pattern worth noting is the lower frequency of PLURAL subjects. Regardless of the effect of word order and human vs. non-human agent properties, it seems that a lot of the PLURAL uses across the three GO verbs seem to be related to 1ST PERSON. As I mentioned previously, the vast majority of corpus returns coded in the data frame belong to newspaper writing. Within these newspaper uses, however, there are narrative texts in which we are likely to find these 1ST PERSON uses. As far as the DUAL marking is concerned, it still does not feature as one of the prominent characteristic of the use of either verb according to Table 11.

VERB x SUBJECT SEMANTIC CATEGORY x PHRASAL SEMANTIC CATEGORY

So far, HCFA analysis has shown that each MSA GO verb associates with a set of distinct morpho-syntactic features in addition to the semantic properties of the sentential

subjects that collocate with each verb. Granted, two or all three of the GO verbs may overlap with regard to certain configurations of the variables examined above. Such frequent and significant configurations can better be explained in the context of the larger conceptual frames in which the three GO verbs occur. The following analysis reports on significant interactions between the following semantic variables: SUBJECT SEMANTIC CATEGORY as well as a set of phrasal semantic categories which indicate the presence (or lack) of GOAL, MANNER, SETTING, PURPOSE, COMITATIVE, and TEMPORAL information.²⁴ Table 12 shows the most significant configurations involving these variables per each GO verb.

TABLE 12. Most significant configurations for SUBJECT SEMANTIC CATEGORY and PHRASAL SEMANTIC CATEGORIES for all GO verbs.

VERB	SUBJ_ CAT	GO- AL	MAN- NER	SET- TING	PATH	PURPO- SIVE	COMIT- ATIVE	TEM- PO- RAL	Freq	Exp	Ob- Ex	Dec	Q
<i>ḡahaba</i>	HUMAN	YES	NO	NO	NO	NO	NO	NO	129	40.25	>	***	0.061
<i>ḡahaba</i>	OBJECT	YES	NO	NO	NO	NO	NO	NO	29	3.68	>	***	0.017
<i>ḡahaba</i>	HUMAN	YES	NO	NO	NO	YES	NO	NO	27	2.24	>	***	0.017
<i>ḡahaba</i>	HUMAN	YES	NO	NO	NO	NO	NO	YES	15	2.39	>	***	0.008
<i>ḡahaba</i>	COMMUNI- CATION	YES	NO	NO	NO	NO	NO	NO	11	1.46	>	**	0.006
<i>ḡahaba</i>	HUMAN	YES	NO	NO	NO	YES	NO	YES	6	0.13	>	***	0.004
<i>maḡā</i>	TIME	NO	NO	NO	NO	NO	NO	NO	155	42.12	>	***	0.077
<i>maḡā</i>	TIME	NO	NO	NO	YES	NO	NO	NO	49	4.40	>	***	0.03
<i>maḡā</i>	TIME	NO	NO	NO	NO	NO	NO	YES	18	2.50	>	***	0.01
<i>maḡā</i>	GROUP	NO	NO	YES	NO	NO	NO	NO	11	1.20	>	***	0.007
<i>maḡā</i>	HUMAN	NO	YES	YES	NO	NO	NO	NO	12	1.43	>	***	0.007
<i>maḡā</i>	TIME	NO	NO	NO	YES	NO	NO	YES	10	0.26	>	***	0.006
<i>maḡā</i>	GROUP	NO	YES	YES	NO	NO	NO	NO	6	0.19	>	***	0.004
<i>rāḡa</i>	HUMAN	NO	NO	NO	NO	NO	NO	NO	286	142.16	>	***	0.106
<i>rāḡa</i>	HUMAN	NO	YES	NO	NO	NO	NO	NO	67	22.63	>	***	0.03
<i>rāḡa</i>	GROUP	NO	NO	NO	NO	NO	NO	NO	50	18.98	>	***	0.021

²⁴ SOURCE and DEGREE specifications were not included in this analysis due to their low frequency of occurrence for all three verbs combined. SOURCE occurs 10/1500 times while DEGREE occurs 12/1500 times. Although the HCFA analysis can handle any number of variables at a time, for practical reasons I decided to exclude these two variables from the present analysis. The fact that a SOURCE phrase was found to be of a lower frequency of occurrence than GOAL phrases has also been noted by Stefanowitsch and Rodhe (2004) in their investigation of GOAL bias across motion verbs in English. Verspoor et al. (1999) refers to this as the “goal-over-source” principle.

The configurations listed in Table 12 pertain to the most significant interactions among all the variables for all three verbs. That is to say, if we run an HCFA test on each verb individually the emerging significant patterns may be more varied since we would be assessing the robustness of different clusters of variables per verb. Since our interest here primarily lies in teasing apart the differences in usage for all three GO verbs combined, the HCFA results reported in this table aim towards highlighting such differences. Appendix G contains tables of significant results obtained from running an HCFA test on individual verbs one at a time.

Moving from the specific morphosyntactic preferences of each verb to a broader view of verb use, Table 12 gives us a glimpse into the extent to which the three verbs differ with respect to their coding of the basic (deictic) motion event construal. In other words, we would like to determine which verb is typically used in a motion event that involves an end-point and which verb is used in a context where the path of motion is highlighted, etc.

Dahaba

One of the obvious differences among the use of these three verbs is the fact that a physical or figurative motion event featuring *dahaba*, as opposed to *madā* and *rāha*, is most likely to involve a GOAL or a destination of the motion event.²⁵ Table 12 lists the most significant configurations of these *dahaba* hits that involve a GOAL phrase. Among the robust interactions listed here is the motion of HUMANS towards GOALS that may or may not be accompanied by a phrase denoting the purpose of the motion event, as in (1), or a phrase locating that motion event in time. Similarly, the occurrence of a non-human entity such as PHYSICAL OBJECT/ARTIFACT or COMMUNICATION (e.g. a statement, a letter,

²⁵ 298/500 (59.6%) of the coded *dahaba* hits involve motion towards an expressed GOAL.

an announcement) with a GOAL phrase seems to be quite significant, as in sentences (2) and (3).

(1) ولم يذهب إلى النادي ليمارس الرياضة ويبني عضلات ذراعيه وصدره

<i>wa=lam</i>	<i>yaḏhab</i>	<i>ilā</i>	<i>al=nādi</i>	<i>li=yumāris</i>	<i>al=riyaḏa</i>
CONJ=NEG	ḏahaba .JUSS.3SG.M	ALL	ART=gym	PURP=practice.SUBJN.3SG.M	ART=sports
and did not	go	to	the gym	to practice	the sports

<i>wa=yabni</i>	<i>'aḏalāt</i>	<i>ḏirā'ay-h</i>	<i>wa=ṣadri-h</i>
CONJ=build.SUBJN.3SG.M	muscles	arms-CL.3SG.M	CONJ=chest.CL.3SG.M
and build	muscles	his arms	and his chest

'And he didn't go to the gym to work out and build his arm and chest muscles'

(2) ومساعداتنا تذهب إلى الشيشان من خلال جمعيات غير حكومية

<i>wa=musā'adatu-na</i>	<i>taḏhab</i>	<i>ilā</i>	<i>al=šišān</i>	<i>min</i>	<i>ḥilāl</i>
CONJ=aid-CL.1PL	ḏahaba .IMPF.3SG.F	ALL	ART=Chechnya	ABL	ADV
and our aid	goes	to	the Chechnya	from	through

<i>ḡam'iyyāt</i>	<i>gayr</i>	<i>ḥukūmiyya</i>
organizations	NEG	governmental
organizations	non-	governmental

'And our financial aid goes to Chechnya through non-government organizations'

(3) وذهبت الاراء المؤيدة لهذا الاتجاه الى ان توظيف قدرة البناء الذاتي

<i>wa=ḏahabat</i>	<i>al='ārā'</i>	<i>al=mu'ayyida</i>	<i>li=hāḏa</i>	<i>al='ittiḡāi</i>
CONJ= ḏahaba .PERF.3SG.F	ART=opinions	ART=supporting	ALL=DEM	ART=direction
and went	the opinions	the supporting	to this	the direction

<i>ilā</i>	<i>anna</i>	<i>tawzīf</i>	<i>qudrat</i>	<i>al=binā'</i>	<i>al=ḏāti</i>
ALL	TOP	employ.VN	ability	ART=building	ART=self.ADJ
to	that	employing	ability	the building	the self

'And the opinions supporting this direction claim that employing the ability of self-development ...'

Maḏā

I noted earlier that the most frequent uses in the corpus of *maḏā* relate to the figurative motion of time. In fact, *maḏa* appears to be the only MSA GO verb that can be used to talk about the passage of time. Nevertheless, we saw earlier in Table 11 that this verb can also collocate with HUMAN agents, as well as sentential subjects coded as GROUP. In terms of the phrasal semantic categories, we can see that *ḏahaba* and *maḏā* radically differ in that sentences containing *ḏahaba* usually specify GOAL, PURPOSIVE

and/or TEMPORAL information, while sentences containing *maḍa*, instead, would specify features of the MANNER, SETTING, PATH, as well as TIME of motion. As far as subjects denoting TIME are concerned, the significant configurations listed above seem to indicate that a TIME “motion” event can involve a statement about the PATH and/or TEMPORAL information, as in (4). On the other hand, a *maḍā* motion event including HUMANS or GROUPS (e.g. organization, institution, which are viewed metonymically as HUMANS) is likely to involve a SETTING and/or MANNER phrase, as in (5), where SETTING refers to a plan or a speech.

(4) ولم يمض اسبوعان على تصديق القانون الجديد حتى أعلنت شركة «بيل اتلانتيك» أنها

<i>wa=lam</i>	<i>yamḍi</i>	<i>'usbū'ān</i>	<i>'alā</i>	<i>taṣḍīq</i>	<i>al=qānūn</i>	<i>al=ḡadīd</i>	<i>ḥattā</i>
CONJ=NEG	<i>maḍā</i> .JUSS.3SG.M	two weeks	LOC	sign.VN	ART=bill	ART=new	ADV
and did not	go	two weeks	over	signing	the bill	the new	until

<i>a'lanat</i>	<i>šarikat</i>	<i>bil</i>	<i>atlantik</i>	<i>annahā</i>
announce.PERF.3SG.F	company	Bill	Atlantic	TOP.CL.3SG.F
announced	company	Bill	Atlantic	that it

‘And it was hardly two weeks after the new bill was signed when Bill Atlantic Company announced that it...’

(5) وهي تمضي بسرعة في مؤامراتها

<i>wa=hiya</i>	<i>tamḍi</i>	<i>bi=sur'a</i>	<i>fī</i>	<i>mu'āmarāti-ha</i>
CONJ=PP	<i>maḍā</i> .IMPF.3SG.F	INST=speed	LOC	conspiracies-CL.3SG.F.GEN
and she	goes	quickly	in	her conspiracies

‘And it’s [i.e. Israel is] quickly going ahead with its conspiracies’

Rāḥa

Table 12 shows that 286 of the 500 uses of *rāḥa* involve HUMANS or GROUPS but no additional phrasal categories, and we can safely assume that the vast majority of these uses signal the grammaticalized function of this verb. However, I will discuss in Chapter 4 that despite the fact that *rāḥa* is heavily grammaticalized in MSA, there are still a number of deictic motion-related uses of *rāḥa* that fall in this category and which diverge from the aspect-marking pattern. In addition to deictic motion uses, a number of *rāḥa*

corpus returns (67/500) signal an idiomatic use of the verb, in which we have a MANNER phrase indicated in the sentence. In this particular idiomatic usage, a HUMAN agent typically “becomes a victim (of X)” (lit. ‘go as a victim (for X)’), as in (6). This counts as one of the overlapping uses between *rāḥa* and *ḍahaba*.

(6) وهي المحاولة التي راحت ضحيتها الطفلة البريئة شيماء

<i>wa=hiya</i>	<i>al=muḥawala</i>	<i>allatī</i>	<i>rāḥat</i>	<i>ḍaḥiyyatu-ha</i>
CONJ=PP	ART=attempt	RP	<i>rāḥa</i> .PERF.3SG.F	victim.CL.3SG.F.GEN
and she	the attempt	that	went	its victim

<i>al=ṭifla</i>	<i>al=barī'a</i>	<i>šaymā'</i>
ART=child.F	ART=innocent.F	Shayma
the child	the innocent	Shayma

‘And it was the attempt that took the life of the innocent child “Shayma”’

3.4 Polytomous logistic regression analysis

So far the univariate analysis conducted on the MSA GO data frame helped us zero in on the distribution of single variables per verb or levels of variables within a single variable, while the multivariate analysis in §3.2.3 helped us figure out prominent patterns of interaction between variables as a means of better understanding the constructional properties of contextualized verb uses. The current polytomous logistic regression analysis applies more advanced algorithms for estimating the relative impact of the various variables and their values on whether an outcome – i.e. each of the three GO verbs – occurs or not. It also estimates the joint effect of multiple predictor variables (i.e. constructional features), and provides us with expected probability estimates for outcome variables (i.e. GO verbs) as a means of evaluating the likelihood of a certain verb in a specific context of use. Such analysis is another step towards identifying not only prototypical uses of each GO verb, but also contexts of use where more than one verb can be used interchangeably.

As mentioned earlier in Chapter 2, in order to conduct a polytomous logistic regression analysis on the MSA GO data frame we need to convert the current data frame

in its multinomial form into a logical form where every level of variable is turned into an individual dummy variable with the levels TRUE/FALSE indicating whether this variable has or has not been observed in the context of use. With the aid of multinomial2logical function incorporated within the {polytomous} package developed for R (Arppe, 2012), all contextual features in the GO data frame can be converted into logical variables, as we can see from the sample of binary variables in Table 13.

TABLE 13. Excerpt of the list of logical variables.

LOGICAL VARIABLES/PREDICTORS	LEVELS
TENSE.PRESENT	TRUE, FALSE
TENSE.PAST	TRUE, FALSE
TENSE.FUTURE	TRUE, FALSE
TENSE.IRR	TRUE, FALSE
SUBJ_PERSON.SINGULAR	TRUE, FALSE
SUBJ_PERSON.PLURAL	TRUE, FALSE
SUBJ_PERSON.DUAL	TRUE, FALSE
SUBJ_CATEGORY.HUMAN	TRUE, FALSE
SUBJ_CATEGORY.OBJECT	TRUE, FALSE
MANNER.YES	TRUE, FALSE
MANNER.NO	TRUE, FALSE
NEGATION.YES	TRUE, FALSE
NEGATION.NO	TRUE, FALSE

The results of conducting standardized Pearson’s residuals discussed in §3.2.1. (Table 5) were based on this logical form of the GO data frame. The following analysis takes advantage of the functions available in the {polytomous} package, in order to arrive at a reasonable model that explains verbal behaviour. The selection of the predictor variables for the subsequent polytomous logistic regression model should therefore be based on the results from the univariate analysis discussed in §3.2.1, as well as the following bivariate analysis.

3.4.1 Bivariate analysis

Prior to establishing a model to fit the data, we need to examine the degree of association between the existing *independent* predictor variables, two variables at a time. Testing for association is a necessary step to identify the pairs of variables that highly co-

occur and, consequently, avoid the selection of both in the polytomous logistic regression model as a means of avoiding excessive co-linearity. To illustrate, if we cross-tabulate the variables TENSE.PAST and MORPH_ASP.MOOD.PERF, we can see in Table 14 that the majority of verb uses coded as appearing in the PAST tense (TENSE.PAST: TRUE) bear a PERFECTIVE morphological aspect marking (MORPH.ASP.MOOD.PERF: TRUE).

TABLE 14. Cross-tabulation of raw frequencies of TENSE.PAST by MORPH_ASP.MOOD.PERF.

TENSE.PAST \ MORPH_ASP.MOOD.PERF	FALSE	TRUE
	FALSE	470
TRUE	19	997

The statistical model constructed in §3.2.4.2 will have to include either but not both of these logical variables. The statistical measure employed for determining the level of association or dependence among pairs of variables is called *Theil's Uncertainty Coefficient* (UC), described in more detail in Arppe (2008: 90-91) and Theil (1970). The concept behind such a test is to find out the extent to which our uncertainty about the occurrence of a certain categorical variable would decrease given another variable. This relationship is not necessarily symmetric, as we will see in Table 16 below. The `associations()` function in the `{polytomous}` package calculates these values, and we can apply this function to a pair of variables at a time, as in 15.

TABLE 15. Bivariate analysis of TENSE.PAST by MORPH_ASP.MOOD.PERF.

VARIABLE 1	VARIABLE 2	UC 1 2	UC 2 1
TENSE.PAST	MORPH_ASP.MOOD.PERF	0.8333372	0.8365791

Table 15 is interpreted thus: the uncertainty of predicting VARIABLE 1 (i.e. TENSE.PAST) is reduced by 83.33% given VARIABLE 2 (i.e. MORPH.AS.MOOD.PERF); and the uncertainty of predicting VARIABLE 2 is reduced 83.66% given VARIABLE 1. Thinking in terms of predictor variables to be included in the logistic regression model, this strong association between the two variables shows that either of these variables is redundant

and, hence, our selection of predictors should include one but not the other member of this pair of variables. A combination of functions incorporated within the {polytomous} package can yield a single table that lists all possible pairs of logical variables and their two-way Uncertainty Coefficients, as demonstrated by the selected pairs of variables in Table 16. This table also lists the raw frequencies of the occurrence of each independent variable (N.1 and N.2) as well as the raw frequency of cases where both variables co-occur in the same context (N.common).

TABLE 16. Bivariate analysis of selected (independent) variables.

VARIABLE 1	VARIABLE 2	UC 1 2	UC 2 1	N.1	N.2	N.COMMON
TENSE.PRES	MORPH_ASP.MOOD.IMPF	0.70225565	0.61870277	268	337	263
TENSE.IRR	ASPECT.NON-FIN	0.98490174	0.98858574	176	175	175
ASPECT.INCP	SVC.YES	0.76453295	0.73622157	464	513	457
MORPH_ASP.MOOD.IMPR	SUBJ_PER.2ND	0.72510137	0.36818003	13	30	13
SUBJ_PER.1ST	SUBJ_GEN.NIL	0.96942640	0.95625510	147	150	147
PP.YES	GOAL.YES	0.34811038	0.43991165	579	331	323
ADVERBIAL.YES	MANNER.YES	0.43792761	0.47760062	237	206	168

Note that the overall generated list of all possible pairs of variables would unavoidably include pairs of complementary variables. For instance, the overall table contained the pair TENSE.FUTURE and TENSE.PAST. Since no contextualized verb use could be inflected for the past tense and future tense at the same time, the co-occurrence frequency (N. common) is zero. These complementary cases are therefore not interesting for our current purposes. The following observations based on Table 16 are meant to (i) further **illustrate** the merits of examining Theil's *Uncertainty Coefficients*, as well as (ii) **guide** the following multivariate analysis:²⁶

1. Our uncertainty of predicting the tense to be PRESENT is reduced by 70.23% (UC 1|2 = 0.7023) given that the morphological marking on the verb is in the IMPERFECTIVE. On the other hand, given that the tense is PRESENT, our uncertainty of predicting whether the aspect marking on the verb is IMPERFECTIVE is reduced 61.87% (UC 2|1 = 0.6187). This shows strong association between TENSE.PRESENT and MORPH_ASP.MOOD.IMPF. An even stronger association exists between

²⁶ Note that according to Arppe (2008), the rule of thumb is to search for pairwise associations in which the UC value is approximately > 0.5 , since such associations are most likely to result in a badly fitted model.

TENSE.IRR and ASPECT.NON-FIN, which is closely tied to non-finite uses of the GO verbs.

2. Having a GO verb appearing in an INCEPTIVE construction strongly implies that this verb also appears in a SERIAL VERB CONSTRUCTION (SVC.YES).
3. Knowing that the subject person inflection on the verb is 2ND PERSON, our uncertainty about whether or not the verb appears in the IMPERATIVE is reduced 72.51% (UC 1|2 = 0.7251). The opposite is true 36.82% of the time (UC 2|1 = 0.3682).
4. Another strong association we have in Table 16 is motivated by subject person inflection properties of MSA where 1ST PERSON inflection on the verb does not necessitate gender inflection (as opposed to 2ND AND 3RD PERSON inflection).
5. As far as phrasal semantic categories are concerned, we can see that having a GOAL phrase in the context of verb use reduces our uncertainty of predicting the inclusion of a PREPOSITIONAL PHRASE in the sentence 34.81% (UC 1|2 = 0.3481), while having a specified MANNER of motion reduces our uncertainty about the use of an ADVERBIAL down to 43.79% (UC 2|1 = 0.4379). This particular observation does not necessarily entail excluding either of these variables in the subsequent logistic regression analysis.

3.4.2 Multivariate analysis

Out of the entire set of 80+ logical variables available, I decided to exclude a number of them from the polytomous logistic regression primarily for a methodological reason. The number of predictors selected for a model should not exceed 1/10 of the frequency of data points for the least frequent outcome (cf. Arppe, 2008). Therefore, in the case of the GO data frame where we have 500 occurrences per verb, we need to limit ourselves to 50 or less predictors. One criterion for variable selection was based on high association values as was shown in the bivariate analysis discussed in the previous section. The second criterion was based on the frequency of occurrence of a variable. In order to be selected for the model, a feature had to occur twice with all three verbs (a way of making sure that the co-occurrence is not a fluke), and at least 20 times with two of the verbs. The minimum frequency required for any feature to be included was therefore 42. The

resulting set consisted of 28 explanatory variables – or predictors – that I selected to include in the polytomous logistic regression model, based on the desire to minimize collinearity among variables.

Following Arppe (2008) the heuristic chosen for implementing the polytomous logistic regression is *one-vs.-rest*, and a model concerning the three MSA GO verbs was fitted using the `polytomous()` function available in the `{polytomous}` package (Arppe, 2012). The complete summary output returned by the `polytomous()` function on the selected predictors is shown in (7).

- (7) Summary for results from running the `polytomous()` function in R on a selection of 28 predictors. The Odds section lists the estimated odds for explanatory variables in favor of or against the occurrence of each verb, while (typically) the non-significant odds ($P < 0.05$) are shown in parentheses.

Formula:

```
VERB ~ ASPECT.HAB + ASPECT.SIMPLE + INTEROG.YES + LOC_ADV.YES +
MANNER.YES + MORPH_ASP.MOOD.IMPF + MORPH_ASP.MOOD.PERF +
NEGATION.YES + PATH.YES + PP.YES + SETTING.YES +
SUBJ_CAT.ACTIVITY+ SUBJ_CAT.COMMUNICATION + SUBJ_CAT.GROUP +
SUBJ_CAT.HUMAN + SUBJ_CAT.OBJECT + SUBJ_CAT.NOTION + SUBJ_CAT.TIME
+ SUBJ_GEN.FEM + SUBJ_GEN.MASC + SUBJ_NUM.PL + SUBJ_NUM.SING +
SUBJ_PER.1ST + SUBJ_PER.2ND + SVC.YES + TENSE.IRR + TEMPORAL.YES
```

Heuristic:

one.vs.rest

Odds:

	maḍā	rāḥa	ḍahaba
(Intercept)	(0.02921)	(0.06303)	(0.03679)
ASPECT.HAB	0.2342	0.02449	489.8
ASPECT.SIMPLE	1.994	0.06234	50.9
INTEROG.YES	(0.3028)	(2.161)	(2.466)
LOC_ADV.YES	(1.03)	(0.2187)	(1.901)
MANNER.YES	1.934	15.49	0.1734
MORPH_ASP.MOOD.IMPF	(1.071)	67.09	(0.532)
MORPH_ASP.MOOD.PERF	0.2784	215.5	(1.22)
NEGATION.YES	(0.5394)	(6.494)	(1.279)
PATH.YES	3.002	(0.8023)	0.3975
PP.YES	(0.7703)	0.02733	3.996
SETTING.YES	6.087	(3.543)	0.1468
SUBJ_CAT.ACTIVITY	4.293	(0.4106)	(0.3806)
SUBJ_CAT.COMMUNICATION	(1.576)	(0.3082)	(1.265)
SUBJ_CAT.GROUP	(2.819)	(0.6279)	(0.3108)
SUBJ_CAT.HUMAN	(1.422)	(0.6576)	(0.8291)
SUBJ_CAT.NOTION	3.901	(1.508)	0.2253
SUBJ_CAT.OBJEC	(0.9131)	(0.5317)	(1.732)
SUBJ_CAT.TIME	388.8	0.05562	0.002903
SUBJ_GEN.FEM	(0.3181)	(1.098)	(7.607)
SUBJ_GEN.MASC	(0.4553)	(0.7341)	(5.785)
SUBJ_NUM.PL	(14.3)	(0.5541)	(0.1371)
SUBJ_NUM.SING	(12.47)	(0.2674)	(0.1972)
SUBJ_PER.1ST	(0.487)	(0.8533)	(5.654)
SUBJ_PER.2ND	(0.5197)	(2.643)	(2.009)
SVC.YES	(1.616)	4.126	0.2798
TEMPORAL.YES	(0.8081)	(0.2591)	(1.827)
TENSE.IRR	(1.053)	0.09549	91.43

Null deviance: 3296 on 4500 degrees of freedom
Residual (model) deviance: 1343 on 4416 degrees of freedom

R2.likelihood: 0.5926
AIC: 1511
BIC: 1957

The value R2.likelihood (R_L^2) indicates the extent to which the constructed model fits the actual occurrences of verbs in the data frame. That is to say, the calculation of R_L^2 is

based on comparing the estimated probabilities for the originally occurring verb per context and the associated variable cluster against the baseline probability for each verb (which is its overall proportion in the data). The R_L^2 calculated for the above model is 0.59 which may be considered relatively high for logistic regression models in general. This may be due to fact that we have quite a large number of variables some of which still correlate with one another to a certain extent.

We can also examine the extent to which the model could accurately predict the use of a certain GO verb in a given context. The measure *accuracy* (Menard, 1995: 28-30; Arppe, 2008: 129-132) can be retrieved from the model statistics in R. This particular measure indicates the number of times that a particular GO verb that was observed in a certain context had received the highest probability estimate for that annotated context. In the case of the model in (7), *accuracy* was estimated to be 0.837. In other words, the model was very successful in predicting which GO verb belongs to which context of use 83.7% of the time. This particular measure is an aggregate of the accuracy value per each GO verb: *ḍahaba* 87.4%, *maḍā* 68.8%, and *rāḥa* 94.8%. It is possible to examine the number of instances in which the model managed to accurately predict the verb observed in context, as in Table 17.

TABLE 17. A cross-tabulation of the predicted vs. observed verbs for all annotated contexts in the data frame (with the total of 500 contexts per GO verb).

PREDICTED OBSERVED	<i>ḍahaba</i>	<i>maḍā</i>	<i>rāḥa</i>
<i>ḍahaba</i>	437	41	22
<i>maḍā</i>	96	344	60
<i>rāḥa</i>	18	8	474

The numbers in bold in Table 16 correspond to the number of times the model accurately guessed which verb is supposed to go into which context. We find, for instance, that out of 500 instances of each GO verb, the model is best at predicting the cases where the verb *rāḥa* originally appears (474/500 times) followed by the verb

ḍahaba (437/500), and less so the verb *maḍā* (344/500). The remaining values listed in this Table show the number of times the model mis-predicted the verb corresponding to the context in question. I hasten to add that these are not uninteresting cases. For instance, 96 instances of the use of *maḍā* were predicted as *ḍahaba*, and 18 cases of *rāḥa* were also predicted as *ḍahaba*. What would be interesting is to examine these individual sentences (contexts of use) and scrutinize the explanatory variables such verb uses were coded for.

Moving back to the individual estimated odds listed in (10) above, the values given per predictor X verb indicate the impact of a given predictor on the selection of each GO verb. These values can be interpreted either predictor-wise or verb-wise. Predictor-wise interpretation of odds means that we can examine the impact each predictor can have on the occurrence of the three possible outcomes (i.e. verbs). For instance, if we take the variable ASPECT.HABITUAL we can see that the estimated odds for the occurrence of *ḍahaba* = 489.8, *maḍā* = 0.2342 and *rāḥa* = 0.02449. Note that in (10), the odds that increase the chances of the occurrence of one verb in a given context need to be >1.0, while the odds that decrease the chances of the occurrence of a verb per contexts are <1.0. Odds indicated in parentheses in (10) are not significant because the model did not have enough data to reliably estimate the impact of those particular predictors on the occurrence of a verb (i.e. the *p*-value of the odds is >0.5). The above odds for the variable ASPECT.HABITUAL, therefore, indicate that the occurrence of this particular predictor increases the chances of observing *ḍahaba* in that context, and would decrease the chances of the occurrence of *maḍā* and *rāḥa*.

On the other hand, if we decide to interpret these odds verb-wise, this would involve examining which of the listed predictors is in favor of the occurrence of a verb in context. That is to say, if we examine the odds calculated for the verb *maḍā*, we can see which of the variables is likely to increase the chances of the occurrence of this verb and, in contrast, which is likely to decrease the chances of the occurrence of this verb. Table 18

summarizes the significant estimated odds for predictors – contextual features – that are significantly in favor of or against the occurrence of each GO verb, as listed in (7).

TABLE 18. Contextual features that increase the odds in favor of or against the occurrence of GO verbs.

VERB	ODDS IN FAVOR OF THE VERB		ODDS AGAINST THE VERB	
<i>ḍahaba</i>	ASPECT.HABITUAL	489.8	SUBJ_CAT.TIME	0.003
	TENSE.IRR	91.43	SETTING.YES	0.147
	ASPECT.SIMPLE	50.9	MANNER.YES	0.173
	PP.YES	3.996	SUBJ_CAT.NOTION	0.225
			SVC.YES	0.28
			PATH.YES	0.398
<i>maḍā</i>	SUBJ_CAT.TIME	388.8	ASPECT.HAB	0.234
	SETTING.YES	6.087	MORPH_ASP.MOOD.PERF	0.278
	SUBJ_CAT.ACTIVITY	4.293		
	SUBJ_CAT.NOTION	3.901		
	PATH.YES	3.002		
	ASPECT.SIMPLE	1.994		
	MANNER.YES	1.934		
<i>rāḥa</i>	MORPH_ASP.MOOD.PERF	215.5	ASPECT.HAB	0.024
	MORPH_ASP.MOOD.IMPF	67.09	PP.YES	0.027
	MANNER.YES	15.49	SUBJ_CAT.TIME	0.056
	SVC.YES	4.126	ASPECT.SIMPLE	0.062
			TENSE.IRR	0.095

Table 18 shows, for instance, that SIMPLE and IRREALIS tenses, HABITUAL aspect, as well as the use of a prepositional phrase/complement in the overall frame, increase the odds in favor of *ḍahaba*. On the other hand, variables such as the subject semantic categories of TIME and NOTION decrease the odds in favor of *ḍahaba*, so does the inclusion of a MANNER, SETTING and PATH phrases and the use of the verb in a serial verb construction. As far as the verb *maḍā* goes, the use of sentential subjects denoting TIME, ACTIVITY and NOTION appear to increase the odds in favor of the occurrence of this verb, in addition to using the verb in the SIMPLE aspect and the inclusion of phrases denoting SETTING, PATH and MANNER. PERFECTIVE morphological aspect as well as HABITUAL aspect appear to dis-favor the use of that verb, and favor the other two verbs. Finally it appears that the chances of the verb *rāḥa* occurring increase when the verb is inflected in either the PERFECTIVE or the IMPERFECTIVE, in addition to the occurrence in a SERIAL VERB CONSTRUCTION and the use of phrases denoting MANNER. The occurrence

of HABITUAL and SIMPLE aspects, SIMPLE tense, prepositional phrases/complements, as well as subjects denoting TIME all appear to increase the odds *against* the occurrence of the verb *rāḥa*.

Probability estimates

As mentioned earlier, the polytomous logistic regression analysis predicts the outcome of the polytomous dependent variable (i.e. VERB) based on the independent predictor variables selected for the model (i.e. the 28 contextual features). The model therefore returns probability estimates of the occurrence of each of the three GO verb, per annotated contextualized verb usage (from the original data frame). The calculation of these probability estimates is based on the joint effect of the explanatory variables included in the fitted model. Examining these probability estimates allows us to identify exemplary (and potentially, prototypical) sentences for each of the studied GO verbs, where the accurately predicted verb (in a given context) receives a very high probability estimate. Sentences (8)-(10) are exemplary sentences for *ḍahaba*, *maḍā*, and *rāḥa* extracted from the original data frame, where the actually observed verb in context received a very high – and almost categorical – probability estimate. I also decided to examine the contextual features (selected to be included in the fitted model) that characterize verb use in each of these contexts.

(8) Sentence #317

<p><i>ḍahaba</i> = 0.983 (observed) <i>maḍā</i> = 0.004 <i>rāḥa</i> = 0.013</p>	<p>contextual features used (in the model): ASPECT.SIMPLE, MORPH_ASP.MOOD.PERF, SUBJ_GEN.FEM, SUBJ_CAT.OBJECT, PP.YES</p>
---	--

أما جائزنا التمثيل فقد ذهبنا للعروض الأفريقية

<i>ammā</i>	<i>ḡā'izatā</i>	<i>al=tamīl</i>	<i>faqad</i>	<i>ḍahabatā</i>
ADV	two awards	ART=acting	DM	<i>ḍahaba</i> .PERF.3DUAL.F
as for	two awards	the acting	already	went

li=l= 'urūd al=afriqiyya
 ALL=ART=performances ART=African
 to the performances the African
 'As for the two acting awards, they went to African performances'

(9) Sentence #662

<i>ḍahaba</i> = 0.001 <i>maḍā</i> = 0.998 (observed) <i>rāḥa</i> = 0.001	contextual features used (in the model): MORPH_ASP.MOOD.PERF, SUBJ_NUM.SING, SUBJ_GEN.MASC, SUBJ_CAT.TIME, NEGATION.YES, PP.YES, PATH.YES, TEMPORAL.YES
--	---

لم يكن مضى وقت طويل على الاعلان رسميا [...] عن رفع العلاقات الدبلوماسية بين فنزويلا وايران إلى مستوى
 السفراء عندما اعتقل [...] ثلاثة لبنانيين

lam *yakun* *maḍā* *waqt* *ṭawīl* 'ala *al=i'lān*
 NEG be.JUSS.3SG.M **maḍā**.PERF.3SG.M time long LOC ART=announcement
 did not be went time long over the announcement

rasmiyyan 'an *raf'* *al= 'ilaqāt* *al=diblomasiyya* *bayna*
 officially about elevate.VN ART=relations ART=diplomatic LOC
 officially about elevating the relations the diplomatic between

fenezuela *wa=irān* *ilā* *mustawā* *sufarā'* 'indamā
 Venezuela CONJ=Iran ALL level ambassadors ADV
 Venezuela and Iran to level ambassadors when

i'tuqila *ṭalāṭat* *lubnāniyyīn*
 arrest.PASS.3SG.M three Lebanese.PL.M
 was arrested three Lebanese

'Hardly a long time has passed since the official declaration... about elevating
 the status of diplomatic relations between Venezuela and Iran to the level of
 exchanging ambassadors ... when three Lebanese were arrested...'

(10) Sentence #1129

<i>ḍahaba</i> = 0.055 <i>maḍā</i> = 0.008 <i>rāḥa</i> = 0.937 (observed)	contextual features used (in the model): MORPH_ASP.MOOD.PERF, SUBJ_NUM.SING, SUBJ_GEN.MASC, SUBJ_CAT.TIME, NEGATION.YES, PP.YES, PATH.YES, TEMPORAL.YES
--	---

وقد راحا يتجولان في أنحاء روسيا

wa=qad *rāḥā* *yataḡawwalān* *fī* *anḥā'* *rūsya*
 CONJ=DM **rāḥā**.PERF.3DUAL.M stroll.around.IMPF.3DUAL.M LOC different.parts Russia
 and went stroll around in different parts Russia
 'And they kept strolling around the different parts of Russia'

The returned probability estimates also include contexts where two or all three
 verbs received somewhat equal probability estimates, possibly indicating that some or all
 of the verbs maybe be used interchangeably in such contexts. We can retrieve the indices
 for a specific set of annotated sentences where the standard deviation of the estimated

probabilities (as a measure of dispersion) is small which, therefore, indicates that the probability estimates for all three verbs are close. In sentences (11)-(15), two or more verbs were deemed to be interchangeable. The probability estimates returned for the GO verbs for the sentences in (11) and (12) appear to be quite close, with *ḡahaba* scoring highest and also being the actually observed verb in the two contexts. Relying on my native speaker's intuition, it is not improbable for either *maḡḡā* or *rāḡḡa* to appear in either context due to the fact that these two sentences do depict (deictic) physical motion to a certain extent.

(11) Sentence #467

<i>ḡahaba</i> = 0.437 (observed) <i>maḡḡā</i> = 0.203 <i>rāḡḡa</i> = 0.360	contextual features used (in the model): ASPECT.SIMPLE, MORPH_ASP.MOOD.PERF, SUBJ_NUM.SING, SUBJ_GEN.MASC, SUBJ_CAT.HUMAN, SVC.YES
--	---

فذهب شقيق المنشق وأسس مطعماً آخر

fa=ḡahaba *šaḡīq* *al=munšaḡ* *wa='assas*
CONJ=*ḡahaba*.PERF.3SG.M brother ART=dissident CONJ=establish.PERF.3SG.M
and went brother the dissident and established

maḡ'aman *'āḡḡar*
restaurant other
restaurant other

'So the dissident's brother left and started another restaurant'

(12) Sentence #5

<i>ḡahaba</i> = 0.437 (observed) <i>maḡḡā</i> = 0.203 <i>rāḡḡa</i> = 0.360	contextual features used (in the model): ASPECT.SIMPLE, MORPH_ASP.MOOD.PERF, SUBJ_NUM.SING, SUBJ_GEN.MASC, SUBJ_CAT.HUMAN, SVC.YES
--	---

قال لها زوجها عندما تركها وذهب ليتزوج بامرأة أخرى

qāl *la-hā* *zawḡu-ha* *'indamā*
say.PERF.3SG.M ALL-CL.3SG.F husband-CL.3SG.F.GEN ADV
and said to her her husband when

taraka-hā *wa=ḡahaba* *li=yatazzawaḡa*
leave.PERF.3SG.M-CL.3SG.F.ACC CONJ=*ḡahaba*.PERF.3SG.M ALL=marry.SUBJN.3SG.M
left her and went to marry

bi=imra'atin *uḡrā*
INST=woman other
with woman other

'Her husband said to her when he left her and went to marry another woman'

Unfortunately, the sentences in (14) and (15) may be among the few contexts of use where the model almost accurately predicted a level of interchangeability among two or all three verbs. The contexts in (16)-(18) also considered two or three verbs to be interchangeable when, in actual usage, that may not be the case.

(13) Sentence #258

<i>ḍahaba</i> = 0.199 (observed) <i>maḍā</i> = 0.308 <i>rāḥa</i> = 0.492	contextual features used (in the model): MORPH_ASP.MOOD.PERF, SUBJ_NUM.SING, SUBJ_GEN.FEM, SUBJ_CAT.NOTION, PP.YES
--	---

وهي الفكرة التي كانت ذهبت تقريباً من عقول الناس
wa=hiya al=fikra allatī kānat ḍahabat taqrīban
 CONJ=PP ART=idea RP be.PERF.3SG.F *ḍahaba*.PERF.3SG.F ADV
 and she the idea that was went almost

min 'uqūl al=nās
 ABL minds ART=people
 from minds the people

‘And it’s the idea that had almost disappeared from people’s minds’

(14) Sentence #19

<i>ḍahaba</i> = 0.448 (observed) <i>maḍā</i> = 0.333 <i>rāḥa</i> = 0.219	contextual features used (in the model): TENSE.IRR, SUBJ_NUM.PL, SUBJ_PER.1ST, SUBJ_CAT.HUMAN, NEGATION.YES, MANNER.YES
--	--

لن نذهب مذهب الروائي فواز حداد في تقديمه للرواية فنقول...
lan naḍhab maḍhab al=riwā'i fawwāz ḥaddād
 NEG *ḍahaba*.SUBJN.1PL path ART=novelist Fawaz Haddad
 will not go path the novelist Fawaz Haddad

fi taqdīm-ih li=l=riwāya fa=naqūl
 LOC preface-CL.3SG.M.GEN ALL=ART=novel CONJ=say.SUBJN.1PL
 in his preface to the novel and say

‘We will not go the same path as the novelist Fawaz Haddad in his preface to the novel and say...’

(15) Sentence #1331

<i>ḍahaba</i> = 0.137 <i>maḍā</i> = 0.396 <i>rāḥa</i> = 0.431 (observed)	contextual features used (in the model): TENSE.SIMPLE, MORPH_ASP.MOOD.IMPF, SUBJ_NUM.SING, SUBJ_GEN.FEM, SUBJ_CAT.HUMAN, MANNER.YES
--	--

وتفجيره بالشكل البشع الذي تروح ضحيته فتيات صغيرات

wa=tafḡīru-h bi=l=šakl al=baši' allaḍi tarūḥ
 CONJ=blowing-CL.3SG.M.ACC INST=ART=way ART=horrible RP *rāḥa*.IMPF.3SG.F
 and blowing it with the way the horrible which goes

<i>dahīyyatu-hu</i>	<i>fatayāt</i>	<i>ṣaġīrāt</i>
victim-CL.3SG.M.GEN	girls	young.PL.FEM
its victim	girls	young

‘and blowing it [i.e. hatred] in this horrible way and cause the death of little girls’

For instance, the verb observed in the sentence in (16) is *dahaba*, yet this verb has received a much lower probability estimate of occurrence than have *maḍa* and *rāḥa*. As mentioned earlier, the estimated probabilities of occurrence basically depend on the joint effect of the calculated odds for all variables included in the model. If we compare the predictor variables (or contextual features) included in the logistic regression model (for which the sentence in (16) was coded) with the list of variables that increase/decrease the odds for each verb in Table 17, we find that the odds for SUBJECT_NUMBER.SING and SUBJECT_GENDER.FEM across the three verbs seem to be quite neutral. This leaves MORPH_ASP.MOOD.PERF, SUBJ_CAT.NOTION and PP.YES as the predictors playing a role in determining which each GO verb applies in the context. In Table 17, the variable MORPH_ASP.MOOD.PERF seems to drastically increase the odds in favor of *rāḥa* while at the same time it decreases the odds in favor of the occurrence of *maḍā*; SUBJ_CAT.NOTION increases the odds for *maḍā* but decreases the odds for *dahaba*; and finally PP.YES increases the odds for *dahaba* but decreases the odds for *rāḥa*. The complex interplay of variables in a polytomous logistic regression, therefore, yields the probability estimates for each verb in (13).

In addition to this case, the sentences in (14) and (15) are examples of idiomatic use of *dahaba* and *rāḥa* (respectively). That is, in (14) the expression *dahaba maḍhab* ‘go/walk the path’ can only host *dahaba* in this context in which the verb is paired with its derived *ism makān* ‘noun of place’ – *maḍhab*. Even though the model assigns the highest probability estimate to this verb, still it seems that *maḍā* with a 0.333 probability estimate could be interpreted as another verb that can fill in the verb slot in this sentence. In the case of the construction in (15), as I will discuss in the next chapter, there is a large

number of idiomatic uses (*rāḥa/yaruḥ ḍaḥiyyatuhu*) among the non-grammaticalized uses of *rāḥa* that express the meaning of “to be a victim of”. This particular idiomatic usage is strongly tied to the verb *rāḥa* and, to a lesser degree, to the verb *ḍahaba*. It can never, however, host the verb *maḍā*. Surprisingly, even though the model assigns the highest probability estimate to *rāḥa*, it nevertheless indicates that it is *maḍā* that is interchangeable with *rāḥa* in that context.

There are a number of implications that we can draw from these above sentences (11)-(15). We need to be reminded that the returned probability estimates rely on how the model was trained – through the set of predictor variables fed into it – to predict a certain outcome, i.e. GO verb, in a given context. Consequently, the accuracy of prediction relies primarily on the predictor variables selected for inclusion in the polytomous logistic regression model.²⁷ We will see in Chapter 5 that the equi-probable cases of occurrences of two or three verbs in a single context actually can reflect interchangeable contexts of use for all four MSA COME verbs. This was not the case with GO verbs. I speculate that one reason for such results pertains more specifically the lack of lexical collocates as variables included in the annotated GO data frame. The sentences in (14) and (15), for instance, were specific examples of idiomatic uses of two GO verbs and the model could have been more successful at assigning probability estimates for such cases had the data frame included lexical items considered as collocates of the GO verbs. This may be considered one of the limitations of the data frame constructed for the current purpose and, therefore, further modifications of the current data frame should probably include adding lexical collocates to the variable set.

²⁷ That being the case I ran the model a number of times with different variable sets based on altered mathematically-based selection of variables (e.g. at least occurring 10 times for two verbs, and at least 20 times overall). The returned results increased the R_L^2 yet upon examination of particular cases of equi-probable verb occurrences, I ran into the same problem as above.

Generally speaking, however, the quantitative analysis of MSA GO verbs – presented in this chapter – has succeeded in teasing apart the different morphosyntactic and semantic properties of the use of each of the three verbs. In the following qualitative analysis, I will zoom in on individual instances of verb use, including patterns of use that have not been highlighted in the preceding statistical analysis. This case-by-case analysis, therefore, constitutes a fundamental part of our examination of the behavior of GO verbs in MSA.

Chapter Four

Qualitative analysis of MSA GO verbs *ḍahaba*, *maḍā*, and *rāḥa*

The quantitative analysis of MSA GO verbs presented in the previous chapter has shed light on the behavioral differences and similarities across the three verbs studied here, *ḍahaba*, *maḍā*, and *rāḥa*, with respect to a wide variety of morphosyntactic and semantic variables. More specifically, the statistical tests (monovariate as well as multivariate) were mostly concerned with identifying the highly frequent morphosyntactic and semantic patterns that associate with each individual verb, leading us to identify the most prototypical and exemplary uses of each GO verb.

That being said, the statistical analyses of GO verbs have not necessarily succeeded in drawing our attention to the less frequent uses of each verb. In addition, the lack of coding for specific lexical collocates in the data frame may have resulted in over-generalizing the results in some cases. In order to provide a more comprehensive and a more descriptively useful analysis of the three GO verbs under study, we need to go back to the 500 corpus returns included in the data frame and examine individual instances of actual verb usage. The present qualitative analysis, therefore, intends to build upon the statistical methods discussed in Chapter 3 by pairing the quantitative results with morphosyntactic and lexico-semantic observations obtained from manual inspection of the annotated corpus hits

More importantly, the current analysis is an attempt at examining GO verb behavior in MSA in light of the previous treatment of GO verbs cross-linguistically. As discussed in Chapter 1, the vast majority of cross-linguistic research on GO and COME verbs has explored (i) properties of the motion event frame – e.g. SOURCE-PATH-GOAL as per Talmy (1975, 1985), (ii) the deictic properties of these verbs (e.g. Fillmore, 1966, 1969, 1970; among others), (iii) the metaphorically extended uses based on these properties (e.g. Clark, 1974, Radden, 1996), as well as (iv) paths of grammaticalization of GO and COME

verbs (e.g. Newman, 2000; Heine, 2002; Botne 2006). The following analysis attempts to address similar properties that specifically pertain to the three MSA GO verbs against the background of the existing research. The discussion of each verb in this chapter starts with examining properties of the physical motion event depicted by each of the three verbs in its basic usage profile, followed by examining the different figurative motion construals, as well as the grammaticalized uses.

4.1 *Dahaba*

4.1.1 Physical motion

Recall from the previous chapter and, more specifically, from the HCFA analysis of *dahaba*, that a large number of the examined corpus hits of this verb involve GOAL-oriented physical motion. In fact, it appears that *dahaba* is *the* MSA GO verb that literate Arabic speakers would use to encode a bounded physical motion (i.e. terminates at a specific end point) away from (or, not towards) the deictic centre rather than *maḍā* or *rāḥa*. Consistent with findings of the monivariate and multivariate analyses discussed in the previous chapter, *dahaba* seems to collocate with a wide range of sentential subjects that belong to the categories of HUMAN, PHYSICAL OBJECT/ARTIFACT, GROUP, NOTION, COMMUNICATION, ACTIVITY, etc. We can expect that motion event schemas involving a HUMAN agent as well as a PHYSICAL OBJECT/ARTIFACT theme are much more likely to depict physical motion than, for instance, subjects that are NOTIONS, COMMUNICATIONS, or ACTIVITIES.

Stefanowitsch and Rodhe (2004), in their corpus-based investigation of GOAL-bias among English motion verbs, found that GOAL-indicating prepositional phrases co-occur with the verb *go* more often than prepositional phrases indicating other aspects of the motion event (e.g. source and trajectory); and that this bias applies to both physical and non-physical motion events. So far, the Arabic data for *dahaba* seems to follow a

similar pattern. As noted earlier, a large number of uses of *ḍahaba* involve an explicitly mentioned endpoint, or a GOAL. As I will explain in §4.1.2, this is not necessarily exclusive to physical motion uses but also extends to the figurative uses of *ḍahaba* as well. Instances of such basic motion event schemas typically involve a prepositional phrase headed by *ilā/li-* ‘to’, as shown in (1) and (2):

(1) ثم ذهب إلى غرفته في هدوء

<i>tumma</i>	<i>ḍahaba</i>	<i>ilā</i>	<i>ḡurfat-ih</i>	<i>fī</i>	<i>hudū'</i>
ADV	<i>ḍahaba</i> .PERF.3SG.M	ALL	room-CL.3SG.M.GEN	LOC	silence
then	went	to	his room	in	silence

‘And then he went to his room silently’

(2) وقالت الفتيات الثلاث [...] إنهن ذهبن للمحكمة سبع مرات لحضور محاكمة أمير

<i>wa=qālat</i>	<i>al=fatayāt</i>	<i>al=ṭalāt</i>	<i>inna-hunna</i>	<i>ḍahabna</i>
CONJ=say.PERF.3SG.F	ART=girls	ART=three	TOP-PP	<i>ḍahaba</i> .PERF.3PL.F
and said	the girls	the three	that they	went

<i>sab'</i>	<i>marrāt</i>	<i>li=ḥuḍūr</i>	<i>muḥākamat</i>	<i>amīr</i>
seven	times	PURP=attend	trial	Amir
seven	times	to attend	trial	Amir

‘And the three girls [...] said that they went to court seven times to attend Amir’s trial’

Examining the different types of goals a human agent arrives at in such a construction reveals a lot about the different motion construals depicted by the verb *ḍahaba*. Sentences (3) - (6) indicate physical motion towards a destination that is either an event/activity, a location, or another human being.

GOAL: event/activity

(3) ويذهب للدروس حتي العاشرة مساء

<i>wa=yadāhaba</i>	<i>li=l=durūs</i>	<i>ḥatta</i>	<i>al='āšira</i>	<i>masā'an</i>
CONJ= <i>ḍahaba</i> .SUBJN.3SG.M	ALL=ART=lessons	ADV	ART=ten	evening
and goes	to the lessons	until	the ten	evening

‘And attends/goes to lessons until 10 in the evening’

(4) ذهبت إلى الإختبار

<i>ḍahabat</i>	<i>ilā</i>	<i>al='iḥtibār</i>
<i>ḍahaba</i> .PERF.3SG.F	ALL	ART=exam
went	to	the exam

‘She went to the exam’

GOAL: location

(5) يكفي أن تذهب مرة واحدة إلى متحف اللوفر وترى الأجنحة المصرية

<i>yakfī</i>	<i>an</i>	<i>tadhāba</i>	<i>marra</i>	<i>wāhida</i>	<i>ilā</i>	<i>muthaf</i>
suffice.IMPF.3SG.M	TOP	dhāba .SUBJN.2SG.M	time	one	ALL	museum
suffices	that	you go	time	one	to	museum

<i>al=lūfer</i>	<i>wa=tarā</i>	<i>al='aġniha</i>	<i>al=mašriyya</i>
ART=Louvre	CONJ=see.SUBJN.2SG.M	ART=wings	ART=Egyptian
the Louvre	and see	the wings	the Egyptian

'You just need to go one time to the Louvre and see the Egyptian sections'

GOAL: human

(6) فذهب لمراقب المباراة ليسأله

<i>fa=dahaba</i>	<i>li=murāqib</i>	<i>al=mubārā</i>	<i>li=yas'ala-hu</i>
CONJ= dhāba .PERF.3SG.M	ALL=observer	ART=match	PURP=ask.SUBJN.3SG.M-CL.3SG.M.ACC
so he went	to observer	the match	to ask him

'so he went to the referee to ask him'

Notice the statement of purpose in sentences (2) and (6) – *to attend the trial of Amir* and *to ask him* – which is a common characteristic of *go to X* sentences and which was also found among the most robust configurations associated with *dhāba* in the HCFA discussion. Newman and Lin (2007) have previously discussed the notion of the purposefulness of *going* in English that is exhibited in utterances they referred to as expressing “conventional purpose” such as *go home, go to bed, go to school, go to work*, as well as utterances like *go and VERB, go (to) VERB, go because, and go for*, all of which seem to render the *going* event more purposeful than a statement of pure motion. Similarly, a large number of the Arabic *dhāba ila/li-* event schemas also contain a statement about the purpose of GOING expressed in a number of constructions involving prepositional phrases or adverbials, as exemplified in sentences (7)-(11)²⁸.

²⁸ Even though the data frame did not include specific lexical collocates, I did in fact code phrasal semantic categories (e.g. GOAL, MANNER, etc.) for the type of construction expressing these settings as well as the specific particles or lexical elements that characterize such constructions. These data were not included in the quantitative analysis. The purpose of this coding was to facilitate manual inspection of such constructions for the present qualitative analysis.

(7) ولم يذهب إلى النادي ليمارس الرياضة ويبنى عضلات ذراعيه وصدره

wa=*lam* yaḡhab ilā al=*nādi* li=*yumāris* al=*riyaḡa*
 CONJ=NEG ḡahaba.JUSS.3SG.M ALL ART=gym PURP=practice.SUBJN.3SG.M ART=sports
 and did not go to the gym to practice the sports

wa=*yabni* 'aḡalāt ḡirā'ay-h wa=*ṡadri-h*
 CONJ=build.SUBJN.3SG.M muscles arms-CL.3SG.M CONJ=chest.CL.3SG.M
 and build muscles his arms and his chest

'And he didn't go to them gym to work out and build his arm and chest muscles'

(8) كأن أذهب لتناول آيس كريم في أي وقت حتي ولو كان متأخرا

ka'an aḡhab li=*tanawul* 'ays krīm fī ay waqt ḡattā
 ADV ḡahaba.IMPF.1SG PURP=take.VN ice cream LOC any time ADV
 such as I go to have ice cream in any time even

wa=*law* kāna muta'aḡḡiran
 CONJ=COND be.PERF.3SG.M late
 and if was late

'As in go to have ice-cream any time even if it were late'

(9) والطلاب قد انتهوا من المقررات وسيذهبون إلى المدارس لكي يلعبوا

wa=*l=ṡullāb* qad intahaw min al=*muḡarrarāt*
 CONJ=ART=students DM finish.PERF.1PL.M ABL ART=courses
 and the students had finished from the courses

wa=*sa-yaḡhabūn* ilā al=*madāris* li=*kay* yal'abū
 CONJ=ART=FUT-ḡahaba.IMPF.3PL.M ALL ART=schools PURP=PURP play.SUBJN.3PL.M
 and will go to the schools in order to play

'And the students had finished courses and will go to school to play'

(10) فذهبوا الى تركيا ايضا من اجل تسجيل المزيد من الالحن

fa=*ḡahabū* ilā turkiyā ayḡan min aḡl tasḡīl
 CONJ=ART=ḡahaba.IMPF.3PL.M ALL Turkey ADV PURP record.VN
 so they went to Turkey also in order to recording

al=*mazīd* min al=*alḡān*
 ART=more ABL ART=tunes
 the more of the tunes

'And so they went to Turkey as well to record more tunes'

(11) يسبق لي أن زرت المنصورة الا في الاسبوع الماضي عندما ذهبت اليها تلبية للقاء مع شبابها

yasbiḡu lī an zurtu al=*manṡūra* illā fī
 precede.IMPF.3SG.M ALL-CL.1SG TOP visit.PERF.1SG Al Mansura CONJ LOC
 precedes to me that I visited Al Mansoura except in

al=*'usbū'* al=*māḡī* 'indamā ḡahabtu ilay-hā talbiyatan
 ART=week ART=last ADV ḡahaba.PERF.1SG ALL-CL.3SG.F answer.VN
 the week the last when went to it answering

li=liqāa' *ma'a* *šabābi-ha*
 ALL=meet.VN COM youth-CL.3SG.F.GEN
 to meeting with its youth
 'I have visited Mansoura before, however last week when I went there to meet
 with its youth as they have invited me'

Newman and Lin (2007) also discussed the English construction *go and VERB*, such as *Go and get her* and *They want me to go and do my shopping*. They remarked that in such constructions “the semantic contribution of GO is variable. While movement away from a deictic centre is present in all the examples listed, the GO sense can be relatively weak, compared with the informational salience of the purpose clause” (2007:298). Among the constructions that were annotated as containing purpose of the motion event are instances in which *dahaba* serializes with another verb (as in *dahaba wa-ffa-*) in order to express purpose. Sentences (12)-(15) are examples of such purposeful usage as found in the data frame. In (12) and (13) we have an imperative *iḏhab* followed by a coordinated clause containing another imperative verb that is the focus of the given command here – ‘finish your work first’ and ‘take it’ – rather than commanding the interlocutor to move away from the deictic centre (i.e. where the speaker is). In (14) and (15), the deictic motion sense is also weakened since the focus on these two utterances is mostly on the event represented by the conjoined verb – ‘study or even finish higher studies’ and ‘[he] established another restaurant’.

(12) قال اذهب وانجز عملك أولا

<i>qāla</i>	<i>iḏhab</i>	<i>wa=angiz</i>
say.PERF.3SG.M	<i>dahaba</i> .IMPR.2SG.M	CONJ=finish.IMPR.2SG.M
said	go	and finish

<i>'amal-ak</i>	<i>awwalan</i>
work-CL.2SG.M.GEN	first
your work	first

'He said, go and finish your work first'

(13) اذهب وخذها

idhab *wa=ḥuḍ-hā*
ḍahaba.IMPR.2SG.M CONJ=take.IMPR.2SG.M-CL.3SG.F
 go and take it
 ‘Go and take it’

(14) هكذا انتهت أحلامي بأن أذهب وأدرس أو حتى أكمل دراستي العليا

hākaḍa *intahat* *aḥlām-i* *bi='an* *aḍhaba*
 ADV end.PERF.3SG.F dreams-CL.1SG INST=TOP *ḍahaba*.SUBJN.1SG
 this way ended my dreams of that I go

wa=adrusa *aw* *ḥattā* *ukmila* *dirāsāt-i* *al='ulya*
 CONJ=study.SUBJN.1SG CONJ ADV finish.SUBJN.1SG studies-CL.1SG ART=high
 and study or even I finish my studies the high
 ‘Therefore, my dreams to go and study or even finish higher studies ended’

(15) فذهب شقيق المنشق وأسس مطعماً آخر

fa=ḍahaba *šaḳīq* *al=munšaḳ* *wa='assas*
 CONJ=*ḍahaba*.PERF.3SG.M brother ART=dissident CONJ=establish.PERF.3SG.M
 so went brother the dissident and established

maḳ'aman 'āḥar
 restaurant other
 restaurant other
 ‘So the dissident’s brother left and started another restaurant’

To a lesser extent, a number of *ḍahaba*-related constructions may denote the simple sense of ‘leave’, as in (16), where the goal of the motion event is not made explicit. Note that one point of contrast between sentence (16) and sentences (17) and (18) is that although the two latter cases may not contain an explicit GOAL-like end point of the motion event, it may still be implicit.

(16) الله يسامحه. قتل الجبلاوى، ثم أعطى الناظر سحره وذهب

Allah *yisamḥ-uh* *qatal* *alḡablawi* *ṭumma*
 God forgive.IMP.3SG.M-CL.3SG.M.ACC kill.PERF.3SG.M Al Gablawi ADV
 God forgive him killed Al Gablawi then

a'ṭā *al=nāzīr* *siḥr-ah* *wa=ḍahab*
 give.PERF.3SGMF ART=viewer magic-CL.3SG.M CONJ=*ḍahaba*.PERF.3SG.M
 gave the viewer his magic and went
 ‘May God forgive him. He killed Al-Gablawi, then he gave the viewer his magic and left’

(17) كلا يا أمى، لن أذهب

<i>kallā</i>	<i>yā</i>	<i>umm-ī</i>	<i>lan</i>	<i>aḡhab</i>
no	VOC	mother-CL.1SG	NEG	<i>aḡhaba</i> .SUBJN.1SING
no	oh	mother	will not	go

‘No, mother, I won’t go’

(18) قلت له: سوف أذهب معك، مادمت سأري الأستاذ هيكل

<i>qultu</i>	<i>la-hu</i>	<i>sawfa</i>	<i>aḡhabu</i>	<i>ma’a-ka</i>
say.PERF.1SG	ALL-CL.3SG.M	FUT	<i>aḡhaba</i> .IMPF.1SG	COM-CL.2 ND .M
I said	to him	will	go	with you

<i>ma-dumtu</i>	<i>sa-’arā</i>	<i>al=ustād</i>	<i>haykal</i>
NEG-remain.PERF.1SG	FUT-see.IMPF.1SG	ART=professor/Mr.	Haykal
as long as I	will see	the professor	Haykal

‘I told him “I’ll go with you, as long as I will see professor/Mr. Haykal”’

The GO motion event in (17) may be interpreted as either having a GOAL, i.e. ‘I will not go [there]’, or as the speaker refusing the concept of leaving, i.e. moving away from the deictic centre that involves the speaker and the interlocutor. In (18), however, we have a typical case of an elided goal of the motion event – ‘I will go with you, as long as I will be able to see Mr. Haykal’ – where the endpoint is retrievable from the context. We know that the speaker and the interlocutor are discussing going to a particular place/event where the speaker specifically expects to meet another person. The GOAL-less phrase in (16), however, lacks even an implicit goal and therefore the interpretation of the motion event is more focused on the “motion away from the deictic centre” aspect of a GO verb. Stefanowitsch and Rohde (2004) found similar GOAL-less constructions where the GOAL of the motion event is not retrievable from the context. Since their corpus-based analysis mainly attempted to test the conceptual motivation behind GOAL-bias as far as motion events are concerned (and particularly *go* in their first study), they remarked that for such *go* uses “the [GOAL] is not mentioned because it is not necessary for a conceptualization of the motion event described” (2004:254). This seems to strongly apply to a number of *aḡhaba* as well as *maḡā* uses in MSA.

Another prominent use of *aḡhaba* pertains to the phrasal use of the verb: *aḡhaba bi-* which denotes ‘to take something/someone (somewhere)’, as in (19) and (20). This

construction (*ḍahaba* X COMITATIVE) did not feature among the significant interactions in the HCFA test, yet there are still a considerable number of corpus hits that are related to this phrasal usage of the verb and which are worth examining. This particular construction is not exclusively tied to the physical motion construals involving *ḍahaba*; it also extends to the figurative domain as we will see in §4.1.2.

(19) هل من المعقول ان يضرم أحد النار في الطابق الأول من المبنى ويذهب بنفسه إلى الطابق العلوي

<i>hal</i>	<i>min</i>	<i>al=ma'qūl</i>	<i>an</i>	<i>yuḍrīma</i>	<i>aḥadun</i>	<i>al=nār</i>	<i>fi</i>
Q	ABL	ART=believable	TOP	light.SUBJN.3SG.M	someone	ART=fire	LOC
is	of	the believable	that	lights	someone	the fire	in

<i>al=ṭābiq</i>	<i>al=awwal</i>	<i>wa=yaḍhaba</i>	<i>bi=naḥsihi</i>
ART=floor	ART=first	CONJ= <i>ḍahaba</i> .SUBJN.3SG.M	COM=himself
the floor	the first	and goes	with himself

<i>ilā</i>	<i>al=ṭābiq</i>	<i>al='ulwi</i>
ALL	ART=floor	ART=upper
to	the floor	the upper

‘Is it believable that someone would start fire on the first floor of the building and then takes himself to the upper floor

(20) بعد أن طلبت أمي أن أذهب بالغداء لأحد أقاربنا

<i>ba'da</i>	<i>an</i>	<i>ṭalabat</i>	<i>umm-ī</i>	<i>an</i>	<i>aḍhaba</i>
ADV	TOP	ask.PERF.3SG.F	mother-CL.1SG	TOP	<i>ḍahaba</i> .SUBJN.1SG
after	that	asked	my mother	that	I go

<i>bi=l=ḡadā'</i>	<i>li='aḥad</i>	<i>aqāribina</i>
COM=ART=lunch	ALL=one	relative
with the lunch	to one	relative

‘After my mother had asked that I take lunch to a relative’

The preposition *bi-* appears to indicate a wide range of uses including spatio-temporal, instrumental, comitative, manner adverbial, abstract or figurative uses, in addition to other uses (Ryding, 2005:367). As far as the deictic motion verbs discussed are concerned, it appears that pairing *bi-* with *atā*, *ḡā'a*, or *ḍahaba*, for instance, involves a comitative usage: ‘come with’ or ‘go with’. Cross-linguistically, it is common to see ‘come with’ constructions giving rise to the meaning ‘to bring’ (Newman, 2000).²⁹

²⁹ I will discuss the MSA ‘come with’ = ‘bring’ in more depth in Chapter 6.

Accompaniment can also be expressed in Arabic by using the locative adverb *ma'a* 'with'.

4.1.2 Non-physical motion

A wide variety of constructions that characterize the use of *dahaba*, which involve both human and non-human agents, pertain to the **fictive** rather than physical motion of entities (Talmy, 1996, 2000). Upon inspection of the 500 coded corpus hits, I found a variety of figurative motion construals each highlighting a particular aspect of the GO motion event. We saw earlier in the monivariate analysis of subject semantic category that the second most frequent category of sentential subjects, following HUMAN, was PHYSICAL OBJECT/ARTIFACT.

Since most of the coded sentences were extracted from newspaper articles, the majority of sentential subjects coded as PHYSICAL OBJECT/ARTIFACT included entities such as '(financial) aid', 'surpluses', 'money', 'stock', 'produce', 'award', and so forth. Sentences (21)-(23) are examples of this particular fictive motion construal, in which a certain coveted object "goes to" a particular recipient or beneficiary. Recall that PHYSICAL OBJECT/ARTIFACT X GOAL was considered among the significant and frequent uses of *dahaba* in the HCFA analysis. In light of the above discussion regarding the purposefulness of going, even though the moving entity in the event schemas in (21)-(23) is non-human (e.g. 'most of the money went to Russia') the 'going' in such sentences is not haphazard. Instead, these event *dahaba* event schemas are better interpreted as purposeful transfer events, where the money/aid/award is an entity that is intentionally moved from the possession of one party to that of another. Notice that in the following instances of verb usage, *dahaba* is being used in a deictic sense again.

(21) وذهب معظم هذا المال إلى روسيا

wa=dahaba *mu'zam* *hāda* *al=māl* *ilā* *rūsya*

CONJ=*dahaba*.PERF.3SG.M majority DEM ART=money ALL Russia
 and went majority this the money to Russia
 ‘And most of this money went to Russia’

(22) ومساعداتنا تذهب الى الشيشان من خلال جمعيات غير حكومية

*wa=*musā'adatu-na *taḡhab* *ilā* *al=*šišān *min* *ḥilāl*
 CONJ=aid-CL.1PL.GEN *dahaba*.IMPF.3SG.F ALL ART=Chechnya ABL LOC
 and our aid goes to the Chechnya from through

ḡam'iyyāt *ḡayr* *ḥukūmiyya*
 organizations NEG governmental
 organizations non- governmental
 ‘And our financial aid goes to Chechnya through non-government organizations’

(23) وذهبت معظم جوائز المهرجان الى الاعمال المصرية والسورية

*wa=*dahabat *mu*'ẓam *ḡawā*'iz *al=*mahraḡān *ilā* *al=*a'māl
 CONJ=*dahaba*.PERF.3SG.F majority awards ART=festival ALL ART=productions
 and went majority awards the festival to the productions

*al=*maṣriyya *wa=l=*suriyya
 ART=Egyptian CONJ=ART=Syrian
 the Egyptian and the Syrian
 ‘And most of the awards at the festival went to Egyptian and Syrian productions’

Another aspect of figurative motion that is highly characteristic of the use of *dahaba* in MSA – and most particularly newspaper writing – is generally concerned with the conceptual domain of cognition. Human agents, as well as what was coded as GROUP (e.g. newspaper, organization) and COMMUNICATION (e.g. speech, document, opinion), often appear in a set of fixed phrases where the construction involving *dahaba* refers to ‘making a claim’, as can be seen in (24) - (26). These sentences contain a prepositional phrase denoting a GOAL, signaling the mental arrival at a conclusion: *dahaba ilā anna X*, ‘go to that X’, or *dahaba ilā al qawl bi'anna X*, ‘go to the saying that X’. These constructions are part of the general and highly frequent pattern of HUMAN or COMMUNICATION X GOAL that was deemed significant by the HCFA test.

(24) ويذهب البعض الى ان هذا التحالف برز قبل مرحلة الاستقلال

*wa=*yadhabu *al=*ba'd *ilā* *anna* *hāḡa* *al=*taḡāluḡ
 CONJ=*dahaba*.IMPF.3SG.M ART=some ALL TOP DEM ART=alliance
 and went the some (people) to that this the alliance

<i>baraža</i>	<i>qabla</i>	<i>marḥalat</i>	<i>al=istiqlāl</i>
emerge.PERF.3SG.M	ADV	stage	ART=independence
emerged	before	stage	the independence

‘And some claim that this alliance emerged prior to the stage of independence’

(25) وذهب البعض الى القول بأنها ليست بديلا عن الجمعيات الفلاحية

<i>wa=dahaba</i>	<i>al=ba'd</i>	<i>ilā</i>	<i>al=qawl</i>	<i>bi=anna-hā</i>
CONJ= <i>dahaba</i> .PERF.3SG.M	ART=some	ALL	ART=say.VN	INST=TOP-CL.3SG.F
and went	the some (people)	to	the saying	of that it

<i>laysat</i>	<i>badīlan</i>	<i>'an</i>	<i>al=ḡam'iyyāt</i>	<i>al=fallāḥiyya</i>
NEG	alternative	about	ART=societies	ART=farmer.ADJ
not	alternative	of	the societies	the farmers-based

‘And some went to say that it is not an alternative to farmers’ societies’

(26) وذهبت بعض التحاليل إلى أن الإختيار الذي تم التوافق حوله بخصوص نظام الإقتراع، بدأت تأثيراته المباشرة تبرز [...] بشكل واضح

<i>wa=dahabat</i>	<i>ba'd</i>	<i>al=taḥālil</i>	<i>ilā</i>	<i>anna</i>	<i>al='iḥtiyār</i>	<i>allaḍī</i>
CONJ= <i>dahaba</i> .PERF.3SG.F	some	ART=analyses	ALL	TOP	ART=choice	RP
and went	some	the analyses	to	that	the choice	that

<i>tamma</i>	<i>al=tawāfiq</i>	<i>ḥawla-h</i>	<i>bi=ḥuṣūṣ</i>	<i>nizām</i>	<i>al='iqtirā'</i>
PASS	ART=agreement	ADV-CL.3SG.M	INST=concerning	system	ART=vote.VN
happened	the agreement	around it	with concerning	system	the voting

<i>bada'at</i>	<i>ta'īrātu-h</i>	<i>al=mubāšira</i>	<i>tabruz</i>
start.PERF.3SG.F	influences-CL.3SG.M.GEN	ART=direct	appear.IMP.F.3SG.F
started	its influences	the direct	appears

<i>bi=šakl</i>	<i>wāḍiḥ</i>
INST=shape	clear
with shape	clear

‘And some analyses claim that the selected voting system started to have obvious influences’

Along these lines where *dahaba* expresses the cognitive ‘motion’ towards a particular end point, other verb uses appear to highlight a different aspect of the motion event frame. In (27)-(30) it is the path/trajectory of the fictive motion event that is now more important than the goal. In these sentences, a sentient being is conceptualized as ‘going further’, ‘going too far’ and ‘going against X’, in referring to cognitive activities. Statistically speaking, this particular construction, however, is considered among more marginal uses of *dahaba*.

(27) ذهب أبعد مما ذهب إليه البنا

<i>ḍahaba</i>	<i>ab'ad</i>	<i>mi-mmā</i>	<i>ḍahaba</i>	<i>ilay-hi</i>	<i>albannā</i>
ḍahaba .PERF.3SG.M	ADV	ABL-RP	ḍahaba .PERF.3SG.M	ALL-CL.3SG.M	Al Banna
went	further	from what	went	to it	Al Banna

‘He went further than Al-Banna did’

(28) يذهب بعيدا في الحلم

<i>yadhāb</i>	<i>ba'īdan</i>	<i>fi</i>	<i>al=ḥulm</i>
yadhāb .IMPF.3SG.M	ADV	LOC	ART=dream
went	far	in	the dream

‘He goes too far in the dream’

(29) من المستحيل أن يذهب السياسيون ضد نصيحة اثنين من أكبر ضباط العمليات خبرة في الجيش البريطاني

<i>min</i>	<i>al=muṣtaḥīl</i>	<i>an</i>	<i>yadhāba</i>	<i>al=siyasiyyūn</i>	<i>ḍid</i>	<i>naṣīḥat</i>
ABL	ART=impossible	TOP	yadhāba .SUBJN.3SG.M	ART=politicians	ADV	advice
of	the impossible	that	goes	the politicians	against	advice

<i>iḥmayn</i>	<i>min</i>	<i>akbar</i>	<i>ḍubbāṭ</i>	<i>al='amaliyyāt</i>	<i>ḥibra</i>
two	ABL	biggist	officers	ART=operations	experience
two	of	biggist	officers	the operations	experience

<i>fi</i>	<i>al=ḡayṣ</i>	<i>al=braṭāni</i>
LOC	ART=army	ART=British
in	the army	the British

‘It is impossible for politicians to go against the advice of two of the most experienced operations officers in the British army’

(30) لن نذهب مذهب الروائي فواز حداد في تقديمه للرواية فنقول...

<i>lan</i>	<i>naḍhab</i>	<i>maḍhab</i>	<i>al=riwā'i</i>	<i>fawwāz ḥaddād</i>	<i>fi</i>
NEG	naḍhab .SUBJN.1PL	path	ART=novelist	Fawaz Haddad	LOC
will not	go	path	the novelist	Fawaz Haddad	in

<i>taqdīm-ih</i>	<i>li=l=riwāya</i>	<i>fa=naqūl</i>
preface-CL.3SG.M.GEN	ALL=ART=novel	CONJ=say.SUBJN.1PL
his preface	to the novel	and say

‘We will not go the same path as the novelist Fawaz Haddad in his preface to the novel and say...’

The construction in (30), *ḍahaba maḍhab*, is an interesting idiomatic usage of *ḍahaba* that generally translates as ‘go the same path as [person X]’ or ‘walk the walk’. This is the only instance of a transitive use of *ḍahaba* that is still used to a certain extent in contemporary MSA. The object of the transitive verb, *maḍhab*, is derived from *ḍahaba* and this particular pattern of morphological marking is referred to in traditional Arabic

grammars as *ism makān* ‘name of place’. In this collocation *maḏhab* mainly refers to ‘path’.

The phrasal use of *ḏahaba bi-*, ‘take someone/something somewhere’, discussed in §4.1.1. can also map onto the cognitive domain as illustrated in (31). In some rare occurrences, this ‘take’ sense can also extend to mean ‘to take away’, i.e. ‘to cause the loss’ of something, as in (32).

(31) حتى ذهب الخيال الشعبي ببعض العوام فاعتبروه نهاية الكون

<i>ḥattā</i>	<i>ḏahaba</i>	<i>al=ḥayāl</i>	<i>al=ša'bi</i>	<i>bi=ba'ḏ</i>	<i>al='awām</i>
ADV	ḏahaba .PERF.3SG.M	ART=imagination	ART=folk.ADJ	COM=some	ART=folks
even	went	the imagination	the folk	with some	the people

<i>fa='tabarū-hu</i>	<i>nihāyat</i>	<i>al=kawn</i>
CONJ=consider.PERF.3PL.M-CL.3SG.M.ACC	end	ART=universe
and considered it	end	the universe

‘Even people got carried away in their collective imagination and considered it [i.e. that year] the end of the universe’

(32) ولا أقول نعم يوماً وأتبعها بلا ولو ذهبت بالمال والولد

<i>wa=lā</i>	<i>aqūlu</i>	<i>na'am</i>	<i>yawman</i>	<i>wa='utbi'uha</i>	<i>bi=lā</i>
CONJ=NEG	say.IMPF.1SG	yes	one day	CONJ=make.follow.IMPF.1SG	INST=NEG
and not	say	yes	one day	and I follow it	with no

<i>wa=law</i>	<i>ḏahabat</i>	<i>bi=l=māl</i>	<i>wa=l=walad</i>
CONJ=COND	ḏahaba .PERF.3SG.F	COM=ART=money	CONJ=ART=children
and if	it went (took)	with the money	and the children

‘And I don’t say ‘yes’ one day then follow it with a ‘no’ if that cost me money and children’

Related to the above, a very common set of idiomatic uses of *ḏahaba* convey the concepts of ‘futility’ and ‘waste’, or ‘going out of existence’, which Newman (2000) has pointed out as one of the possible sub-senses that may be associated with a GO verb cross-linguistically. Sentences (33)-(37) present a number of such idiomatic uses where *ḏahaba* collocates with adverbs such as *hadran/habā'an/suda* that generally denote ‘waste’ and ‘non-existence’. The expressions in (36) and (37) also express a similar notion: *ḏahabat adrāğ alriyāḥ*, ‘go along the path of the winds’, and *ḏahabat ilā ġayr rağ'a*, ‘went to a point of no-return’.

(33) الى الاعتراف بأن عملية عناقيد الغضب ذهبت هدرأ وأنه خدع الشعب

ilā al='i'tirāf bi-'anna 'amaliyyat 'anāqīd al=ḡaḏab ḏahabat
 ALL ART=confession INST-TOP operation grapes ART=wrath **ḏahaba**.PERF.3SG.F
 to the confession of that operation grapes the wrath went

hadran wa=anna-hu ḥada'a al=ša'b
 waste.ADV CONJ=that.CL.3SG.M deceive.PERF.3SG.M ART=people
 as waste and that he deceived the people

'to the confession that "Grapes of Wrath" operation was in vain and that he deceived the people'

(34) ولم تفد هذه الدروس الأب ولا الابن فذهبت الأموال هباءً

wa=lam tufid hāḏihi al=durūs al='ab wa=lā al='ibn
 CONJ=NEG benefit.JUSS.3SG.F dem ART=lessons ART=father CONJ=NEG ART=son
 and did not benefit these the lessons the father and not the son

fa=ḏahabat al=amwāl habā'an
 CONJ=**ḏahaba**.PERF.3SG.F ART=money dust
 so went the money dust

'And neither the father nor the son benefited from these lessons and the money went down the drain'

(35) الأموال التي دفعت من ثروة الرئيس الحريري لم تبدد أو تذهب سدى

al=amwāl allatī dufi'at min ḡarwat al=ra'īs al ḡarīri lam
 ART=money RP pay.PASS.3SG.F ABL fortune ART=president Al Hariri NEG
 the money that was paid from fortune the president Al Hariri did not

tubbaddad aw taḏhab suda
 waste.PASSIVE.3SG.F CONJ **ḏahaba**.JUSS.3SG.F vain.ADV
 was wasted or went vain

'The money that was paid from President Hariri's fortune was not squandered or was not all for nothing'

(36) لكن جهوده ذهبت ادراج الريح

lākin ḡuhūdu-hu ḏahabat adrāḡ al=riyāḥ
 CONJ efforts-CL.3SG.M.GEN **ḏahaba**.PERF.3SG.F traces ART=wind
 but his efforts went traces the wind

'But his efforts were wasted'

(37) الذي ظن كثيرون انه ذهب إلى غير رجعة

allaḏī ḡanna al=kaḡīrūn anna-hu ḏahaba
 RP think.PERF.3SG.M ART=many TOP-CL.3SG.M **ḏahaba**.PERF.3SG.M
 that thought the many (people) that it went

ilā ḡayr raḡ'a
 ALL NEG returning
 to no returning

'Many thought that he was gone once and for all'

Radden refers to this kind of construal as representing 'non-deictic object motion'

(1996:444). He attributes an utterance such as 'Their plans went awry' to a 'diversion

schema' for which the underlying conceptual metaphor is UNEXPECTED CHANGE OF

OUTCOME IS DIVERSION TO A DIFFERENT DESTINATION. In such a construal, the highlighted aspect of a basic motion event is not deixis or locomotion of entities, but rather force vectors (Johnson, 1987) or the notion that “SOURCE-PATH-GOAL schema typically involve a force which imparts the vectors, i.e. a directionality, on the moving object” and that animate agents such as humans and animals as well as intentions or physical causes constitute the origin of force vectors (Radden, 1996:436).

The idea of perishing or disappearing from existence is one of the very few uses that unite the three GO verbs studied here. Among the marginal uses of *ḡahaba* is the idiomatic expression in (38), which more typically associates with *rāḡa* but can also feature *ḡahaba* instead: *ḡahaba ḡaḡiyyatuha X* ‘X goes as/is a victim of...’. Similarly, the same simple construction we discussed above, denoting ‘leave’ in (16), can also extend to mean ‘perish’ or ‘die’, as in (39) and (40).

(38) استهدفها عبر العملية التي ذهب ضحيتها السياح اليونانيون

<i>istahdafa-hā</i> target.PERF.3SG.M- CL.3SG.F.ACC targeted it	<i>'abra</i> LOC through	<i>al='amaliyya</i> ART=operation the operation	<i>allatī</i> RP that	<i>ḡahaba</i> <i>ḡahaba</i> .PERF.3SG.M went	<i>ḡaḡiyyatu-hā</i> victims- CL.3SG.F.GEN its victims
--	--------------------------------	---	-----------------------------	---	--

<i>al=suyyāḡ</i> ART=tourists the tourists	<i>al=yunāniyyūn</i> ART=Greek the Greek
--	--

‘targeted it through the operation that cost the lives of the Greek tourists’

(39) لكن اذ ذهب ما له كله يكون لديه وهم فارغ لا معنى له

<i>lākin</i> CONJ but	<i>iḡ</i> COND if	<i>ḡahaba</i> <i>ḡahaba</i> .PERF.3SG.M went	<i>mālu-hu</i> money-CL.3SG.M.GEN his money	<i>kullah</i> all all	<i>yakūn</i> be.IMPF.3SG.M is
-----------------------------	-------------------------	---	---	-----------------------------	-------------------------------------

<i>ladayhi</i> POSS with him	<i>wahm</i> illusion illusion	<i>fāriḡ</i> empty empty	<i>lā</i> NEG no	<i>ma'nā</i> meaning meaning	<i>la-h</i> ALL-CL.3SG.M to it
------------------------------------	-------------------------------------	--------------------------------	------------------------	------------------------------------	--------------------------------------

‘But if all his money is gone all he will have left is an empty and meaningless illusion’

(40) لماذا يذهب الطيبون؟! لماذا يبقى المجرمون!؟

<i>limāḡa</i> Q why	<i>yaḡḡab</i> <i>ḡahaba</i> .IMPF.3SG.M go	<i>al=ḡayyibūn</i> ART=good.people the good people	<i>limāḡa</i> Q why	<i>yabqā</i> stay.IMPF.3SG.M stay	<i>al=murimūn</i> ART=criminals the criminals
---------------------------	---	--	---------------------------	---	---

‘Why do good people go?! Why do criminals stay?!’

There is no doubt that the concepts of ‘leaving’ and ‘perishing’ are conceptually related; that is to say both departing and dying entities are moving away from the deictic centre, albeit with the former (‘leaving’) involving physical motion while the latter (‘perishing/dying’) not necessarily so. The negative associations with ‘leaving’ in the above sentences can be assumed to highlight the deictic nature of GO, where moving *away* from the deictic centre is considered negative, while that which lies closer to the deictic centre is considered positive (Lakoff and Johnson, 1979, 1980).

4.2 *Maḍā*

It is not surprising that the monivariate as well as multivariate analyses have shown a strong association between the verb *maḍā* and the notion of ‘the passage of time’. As a consequence, most of this verb’s uses relate to figuratively construed motion events. The HCFA analysis has also yielded some robust interactions between this verb and HUMAN or GROUPS in motion event frames including, e.g. MANNER or SETTING. Such instances of verb use also largely depict figurative uses of the verb. I will, nevertheless, follow the structure set for this chapter and begin the discussion of *maḍā* with the physical and less frequent aspects of verb use. The following discussion aims to highlight the grammaticalized uses of this verb which foreshadow the discussion regarding the almost fully grammaticalized *rāḥa* in §4.3.

4.2.1 Physical motion

Inspection of the coded *maḍā* corpus returns has revealed a number of verb uses that belong to the physical domain and which can be considered contexts of use where *maḍā* and *ḍahaba* are interchangeable. For instance, a number of *maḍā* utterances involving physical motion of human agents or concrete objects can contain an endpoint, as in (41), or no endpoint at all as in (42). To a literate native speaker of Arabic, both sentences can

host either *maḍā* or *ḍahaba* and the choice of verb may be only reflective of a stylistic preference.

(41) ويهفو قلبي كلما مضيت إلى المدينة

<i>wa=yahfū</i>	<i>qalb-ī</i>	<i>kullamā</i>
CONJ=flutter.IMPF.3SG.M	heart-CL.1SG.GEN	ADV
and flutters	my heart	every time

<i>maḍaytu</i>	<i>ilā</i>	<i>al=madīna</i>
<i>maḍā</i> .PERF.1SG	ALL	ART=Medina
I went	to	the Medina

‘and my heart flutters every time I go to Medina [of the Prophet]’

(42) حان الوقت يجب أن أمضي

<i>hāna</i>	<i>al=waqt</i>	<i>yaḡibu</i>	<i>an</i>	<i>amḍī</i>
arrive.PERF.3SG.M	ART=time	make.necessary.IMPF.3SG.M	RP	<i>maḍā</i> .SUBJN.1SG
arrived	the time	it is necessary	that	I go

‘It is time, I must go’

A small number of *maḍā* sentences include a statement of purpose of the motion event, much like what we saw earlier with *ḍahaba*, as in (43).

(43) ومضيت لزيارة عليّة محمد وأسرتها

<i>wa=maḍaytu</i>	<i>li=ziyārat</i>	<i>‘aliyya muḥammad</i>	<i>wa=‘usrati-ha</i>
CONJ= <i>maḍā</i> .PERF.1SG	PURP=visiting	Aliyya Muhammad	CONJ=family-CL.3SG.F.GEN
and I went	to visit	Aliyya Muhammad	and her family

‘So I went to visit Aliyya Muhammad and her family’

The overwhelming GOAL-bias we saw in the physical motion events hosting *ḍahaba* does not seem to be a characterizing feature of *maḍā*. Numerous instances of the latter verb’s use in which a physical motion event is depicted may instead include a MANNER of motion description of the event, or a PATH of motion description, as in (44) and (45).

(44) ودعاه الجرسون إلى التليفون فمضى مسرعاً ملهوفاً

<i>wa=da‘ā-hu</i>	<i>al=ḡarson</i>	<i>ilā</i>	<i>al=tilifūn</i>
CONJ=call.PERF.3SG.M-CL.3SG.M.ACC	ART=waiter	ALL	ART=phone
and called him	the waiter	to	the phone

<i>fa=maḍā</i>	<i>musri‘an</i>	<i>malhufan</i>
CONJ= <i>maḍā</i> .PERF.3SG.M	quickly	anxiously
so he went	quickly	anxiously

‘And the waiter asked him to get the phone so he went quickly and anxiously’

(45) ومتي اقتربت الطائرة من مجالنا الجوي تعلن عن نفسها وتمضي في طريقها بغير مشاكل الي المطار

<i>wa=matā</i>	<i>iqtarabat</i>	<i>al=īā'ira</i>	<i>min</i>	<i>mağālina</i>	<i>al=ğawwi</i>
CONJ=RP	approach.PERF.3SG.F	ART=aircraft	ABL	space	ART=air-ADJ
and when	approached	the aircraft	from	space	the air

<i>tu'lin</i>	<i>'an</i>	<i>nafsi-ha</i>	<i>wa=tamḏī</i>	<i>fī</i>
declare.PERF.3SG.F	about	self-CL.3SG.F.GEN	CONJ= maḏā .IMPF.3SG.F	LOC
declares	about	itself	ang goes	in

<i>īariqi-ha</i>	<i>beğayr</i>	<i>mašākil</i>	<i>ilā</i>	<i>al=maḏār</i>
way.CL.3SG.F.GEN	ADV	problems	ALL	ART=airport
its way	without	problems	to	the airport

‘And whenever an aircraft approaches our airspace it declares itself then goes on its way to the airport without problems’

In (44), the adverb, *sari'an* ‘quickly’, describes the *pace* of motion, while in (45), we see the *path* of a physical motion event highlighted. As we will see in the following discussion about more figurative uses of this verb, one of the defining aspects of the use of *maḏā* is a focus on the path/trajectory of motion, rather than on the GOAL per se.

4.2.2 Non-physical motion

The literal motion event construals involving *maḏā* (in a physical domain) can also extend to the figurative domain. As an example, the data frame contained instances of verb use in which the figurative or fictive motion of an entity involves motion away from the deictic centre and, more specifically, the notion of ‘leaving’, as we also saw with *ḏahaba*.

(46) فتحية لجيل مضى كان همه [...] الجدية في العمل

<i>fa=taḥiyya</i>	<i>li-ğīl</i>	<i>maḏā</i>	<i>kāna</i>	<i>hammu-hu</i>
CONJ=salutation	ALL=generation	maḏā .PERF.3SG.M	be.PERF.3SG.M	concern-CL.3SG.M.GEN
so salutations	to generation	went	was	his concern

<i>al=ğiddiya</i>	<i>fī</i>	<i>al='amal</i>
ART=earnestness	LOC	ART=work
the earnestness	in	the work

‘Salutations to a past generation whose concern was earnestness in the workplace’

In the discussion of the figurative uses of *ḏahaba* involving goal constructions, we found that the two most prominent verb uses in such contexts can denote either

“transfer of property” or “going to a conclusion”. Goal constructions associated with the use of *maḍā* express different types of relations. In (47)-(49) the use of *maḍā* represents the metaphorical motion of an entity (‘globalization’, ‘novel’, ‘human’) towards a certain objective (‘crushing freedoms’, ‘illustrating a complex world’, ‘success’). Note, however, that these goal-oriented uses were not deemed significantly frequent in the HCFA or the other statistical measures.³⁰

(47) عولمة القرن 21 وهي في يد العنصرية اليهودية تمضي إلى سحق الحريات

<i>'awlamat</i>	<i>al=qarn</i>	<i>21</i>	<i>wa=hiya</i>	<i>fī</i>	<i>yad</i>	<i>al='unṣuriyya</i>
globalization	ART=century	21	CONJ=PP	LOC	hand	ART=racism
globalization	the century	21	and she	in	hand	the racism

<i>al=yahūdiyya</i>	<i>tamḍī</i>	<i>ilā</i>	<i>saḥq</i>	<i>al=ḥurriyāt</i>
ART=Jewish	<i>maḍā</i> .IMPF.3SG.F	ALL	crushing	ART=freedoms
the Jewish	goes	to	crushing	the freedoms

‘The globalization of the 21st century, controlled by Jewish racism, moves towards crushing freedoms’

(48) في حين تمضي رواية الخطيب الى رسم عالم مركب

<i>fi</i>	<i>ḥīn</i>	<i>tamḍī</i>	<i>riwāyat</i>	<i>alḥaṭīb</i>	<i>ilā</i>	<i>rasm</i>	<i>'ālam</i>	<i>murakkab</i>
LOC	ADV	<i>maḍā</i> .IMPF.3SG.F	novel	Al Khatib	ALL	draw.VN	world	complex
in	while	go	novel	Al Khatib	to	drawing	world	complex

‘Meanwhile, Al-Khatib’s novel goes towards illustrating a complex world’

(49) نحن نمضي من نجاح إلى نجاح

<i>naḥnu</i>	<i>namḍī</i>	<i>min</i>	<i>naḡāḥ</i>	<i>ilā</i>	<i>naḡāḥ</i>
PP	<i>maḍā</i> .IMPF.3PL	ABL	success	ALL	success
we	go	from	success	to	success

‘We are moving from success to success’

Interestingly, relying on my native speaker intuition, the sentences in (47)-(49) may not sound as felicitous had the verb slot been filled with *ḍahaba* instead. The translation equivalent of *maḍā* in these sentences can either be ‘go’ or ‘move’, and the fact that *ḍahaba* may not be interchangeable with *maḍā* in these particular contexts tells us

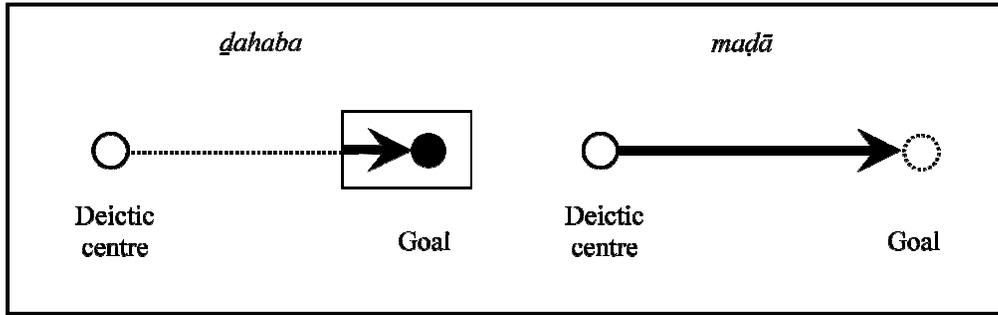
³⁰ Recall that the HCFA analysis, as laid out in Chapter 3, Table 11, was meant to differentiate between the three GO verbs. That is to say, even if many corpus returns of *maḍā* did contain motion towards a goal, in contrast to the extremely goal-biased *ḍahaba*, *maḍā* goal-oriented events are not as frequent, and hence not significant.

something about lexical semantics of the two verbs that may trigger such lexical restrictions.

Wilkins and Hill (1995) suggested that languages differ with respect to the lexical semantics of COME and GO verbs, and that GO verbs are not necessarily inherently deictic. Along these lines the non-deictic uses of GO verbs discussed here may well represent the primitive TRAVEL notion suggested by Miller and Johnson-Laird (1976). As far as *dahaba* and *maḏā* are concerned, the data shows that while most instances of *dahaba* usage involve deictic motion (to a goal), the majority of *maḏā* uses do not and instead they highlight the locomotion aspect of the motion event. The lexical restrictions on the use of *dahaba* in (47)-(49) may point out to the fact that a motion event including *dahaba* is more purposeful in nature than a *maḏā* motion event, an assumption that might be validated by the GOAL-bias associated more predominantly with the former rather than the latter verb.

The quantitatively and qualitatively analyzed corpus returns for both *maḏā* and *dahaba* suggest that one very salient difference among the two verbs, which seems to explain their lexical preferences and restrictions, pertains to the particular aspect of deictic motion each verb seems to emphasize. Figure 1 is a schematic representation – à la Wilkins and Hill (1995) – that marks the difference in general focus in *maḏā* versus *dahaba* motion events. Thicker lines represent the more salient aspects of the motion event (i.e. either the ‘goal’ of the motion event in the schema on the left, or ‘locomotion’ in the schema on the right).

FIGURE 1. Schematic representation of *maḍā* as opposed to *ḍahaba*.



This schematic representation in Figure 1 could explain verbal behaviour to a large extent. Focus on (fictive) locomotion or path of motion of entities motivates the following constructions, shown in (50) and (51), that are exclusive to *maḍā*.

(50) أصبح عملية حقيقية تمضي في ارض الواقع

<i>aṣbaḥa</i>	<i>'amaliyya</i>	<i>ḥaқиiqiyya</i>	<i>tamḍī</i>	<i>fī</i>	<i>'arḍ</i>	<i>al=wāқи'</i>
become.PERF.3SG.M	procedure	true	<i>maḍā</i> .IMPF.3SG.F	LOC	land	ART=reality
become	procedure	true	goes on	in	land	the reality

'It has become a real process going on in reality'

(51) فلا يمضي الفلسطيني في دفع ثمن ما ارتكبه غيره

<i>CONJ=lā</i>	<i>yamḍā</i>	<i>al=falasṭīni</i>	<i>fī</i>	<i>daf'</i>	<i>taman</i>	<i>ma</i>
CONJ=NEG	<i>maḍā</i> .IMPF.3SG.M	ART=Palestinian	LOC	pay.VN	price	RP
so not	go on	the Palestinian	in	paying	price	what

<i>irtakaba-hu</i>	<i>ḡayruh</i>
commit.PERF.3SG.M-CL.3SG.F.ACC	other
committed	other

'So that the Palestinian would not have to go on paying for the mistakes made by others'

Other corpus uses of *maḍā* may highlight information about the pace of motion, as in (52) and (53), or the direction of motion, as in (54). In fact, *maḍā quduman* 'go/move forward', counts as one of the most frequent collocational uses of this verb.

(52) فقد مضت اليابان الرأسمالية سريعا، سريعا جدا، إلى الأمام

<i>faqad</i>	<i>maḍat</i>	<i>al=yābān</i>	<i>al=ra'smāliyya</i>	<i>sarī'an</i>	<i>sarī'an</i>
DM	<i>maḍā</i> .PERF.3SG.F	ART=Japan	ART=capitalist	ADV	ADV
already	went	the Japan	the capitalist	quickly	quickly

<i>ḡiddan</i>	<i>ilā</i>	<i>al='amām</i>
INTENS	ALL	ART=ahead
very	to	the ahead

‘Capitalist Japan moved ahead quickly, very quickly’

(53) نمضي بخطا ثابتة، إلى الهيكل

namḍī *bi=ḥuṭan* *tābitatin* *ilā* *al=haykal*
maḍā.IMPF.1PL INST=steps solid ALL ART=temple
we go with steps solid to the temple
‘We move with solid steps towards the temple’

(54) نيناتياهو الذي اعلن مرارا انه سيمضي قدما في عملية السلام

nitinyāhu *allaḍī* *a'lana* *mirāran* *anna-hu*
Netanyahu RP announce.PERF.3SG.M ADV TOP-CL.3SG.M
Netanyahu who announced several times that he

sa-yamḍī *quduman* *fī* *'amaliyyat* *al=salām*
FUT-*maḍā*.IMPF.3SG.M forward LOC process ART=peace
will go forward in process ART=peace

‘Netanyahu who announced time and again that he will go ahead with the peace process’

Newman (2000) noted that while GO event frames are atelic in nature – as opposed to the telic COME event frames – GO is therefore “more appropriate as a source for conceptualizing the persistence of an event than [COME]” (2000:79). This atelic nature of GO therefore gives rise to aspectual markers such as durative, continuative, and persistent. The atelic nature of *maḍā* is manifested again in utterances where the verb can indicate a continuative sense ‘go on’ (Ryding, 2005), as in (55) and (56). Notice that this particular event construal would not admit *ḍahaba*, presumably due to its more telic GOAL-bias that sets its usage apart from the most felicitous uses with *maḍā*.

(55) فعملية الديمقراطية ستمضي وتشق طريقها الى قلب بكنغهام

fa= 'amaliyyat *al=damaqrata* *sa-tamḍī* *wa=tašūq* *ṭarīqa-ha*
CONJ=process ART=democratization FUT- CONJ=cut.IMPF.3SG.F way-
so process the democratization will go on and cut CL.3SG.F.GEN
its way

ilā *qalb* *bakinghām*
ALL heart Buckingham
to heart Buckingham

‘The democratization process will go on and make its way into the heart of Buckingham’

(56) وهكذا يمضي البروفسور مزاوي مفندا وكاشفا مواد «اعلان المبادئ»

<i>wa=hākḍa</i>	<i>yamḍī</i>	<i>al=profesor</i>	<i>mazāwi</i>	<i>mufannidan</i>
CONJ=ADV	<i>maḍā</i> .IMPF.3SG.M	ART=professor	Mazawi	refute.AP.3SG.M
and thus	goes on	the professor	Mazawi	refuting

<i>wa=kāšifan</i>	<i>mawād</i>	<i>i'lān</i>	<i>al=mabādi'</i>
CONJ=expose.AP.3SG.M	articles	declaration	ART=principles
and exposing	articles	declaration	the principles

‘Professor Mazawi, thus, goes on calling in question and exposing the articles of “declaration of principles”’

The construction in (56), *yamḍī mufannidan* ‘goes on calling into question’, in which *maḍā* is typically followed by an active participle is also another recurring structure or constructional type that characterizes the figurative uses of this verb. This construction is based upon using *maḍā* in conjunction with an active participle form (*fā’ilan*). In the 500 instances of *maḍā*, there were 23 such instances of this construction, which should not be considered as an infrequent pattern.³¹ Certain classical and modern dictionaries would list the collocation *maḍā qā’ilan* ‘went on saying/to say’ as one of the main uses of *maḍā*, as in (57). The verb *qāla* ‘say’ appears to be the most frequent verb to collocate with *maḍā* in this construction (18 out of the 23 instances of this construction). Nevertheless, a subset of verbs related to *speaking*, as in (56) - (58), or *sound making* (e.g. singing), as in (59), was also found to collocate with *maḍā* in this particular construction.

(57) ومضى الرئيس الإيراني قائلا

<i>wa=maḍā</i>	<i>al=raīs</i>	<i>al=’irani</i>	<i>qā’ilan</i>
CONJ= <i>maḍā</i> .PERF.3SG.M	ART=president	ART=Iranian	say.AP.3SG.M
and went	the president	the Iranian	saying

‘The Iranian President went on to say...’

(58) ويمضي الكتاب ساردا سيرة طريق الحرير

³¹ Such construction was coded as an SVC (serial verb construction) usage of *maḍā*. This variable was not discussed in the HCFA analysis, but was included in the polytomous logistic regression analysis as one of the predictor variables added in the model. The calculated odds for this variable, reported in the summary of results of the regression model, were not deemed as either significantly in favor of or against the occurrence of *maḍā*. It should nevertheless be treated as one characteristic features of the use of this verb.

<i>wa=yamḏī</i>	<i>al=kitāb</i>	<i>sāridan</i>	<i>sīrat</i>	<i>ṭarīq</i>	<i>al=ḥarīr</i>
CONJ= <i>maḏā</i> .IMPF.3SG.M	ART=book	narrate.AP.3SG.M	history	road	ART=silk
and goes	the book	narrating	history	road	the silk

‘And the book goes on narrating the history of the Silk Road’

(59) ثم يمضي صوت القصيدة مغنيا مصفرا ضاحكا منتشيا باحساسه بتحقق الذات

<i>tumma</i>	<i>yamḏī</i>	<i>ṣawt</i>	<i>al=qaṣīda</i>	<i>muḡanniyan</i>	<i>muṣaffīran</i>
CONJ	<i>maḏā</i> .IMPF.3SG.M	voice	ART=poem	sing.AP.3SG.M	whistle.AP.3SG.M
then	go	voice	the poem	singing	whistling

<i>ḏāḥikan</i>	<i>muntaṣīyan</i>	<i>bi=’iḥsāsīh</i>	<i>bi=taḥaqquq</i>	<i>al=ḏāt</i>
laugh.AP.3SG.M	be.intoxicated.AP.3SG.M	INST=feeling	INST=fulfillment	ART=self
laughing	being intoxicated	with feeling	the fulfillment	the self

‘Then the voice of the poem goes on singing, whistling, laughing, and being intoxicated by the sense of self-fulfillment’

Notice that in (60), where *maḏā* collocates with the active participle *tārikan* (*ba’ḏa alnudūb*) ‘leaving (some scars)’, the usage of *maḏā* no longer indicates a continuative sense but simply motion away from the speaker/deictic centre; while the active participle is strictly describing the manner in which the ‘leaving’ event takes place.

(60) ثم يعود أحيانا ليخفق في عروقي ثم يمضي تاركا بعض الندوب على يدي

<i>tumma</i>	<i>ya’ūd</i>	<i>aḥyānan</i>	<i>li=yahfiqa</i>	<i>fī</i>	<i>’urūq-i</i>
CONJ	return.IMPF.3SG.M	ADV	PURP=beat.SUBJN.3SG.M	LOC	veins.CL.3SG.M.GEN
then	returns	sometimes	to beat	in	my veins

<i>tumma</i>	<i>yamḏī</i>	<i>tārikan</i>	<i>ba’ḏa</i>	<i>al=nudūb</i>	<i>’alā</i>	<i>yad-i</i>
CONJ	<i>maḏā</i> .IMPF.3SG.M	leave.AP.3SG.M	some	ART=scars	LOC	hand-CL.1SG.GEN
then	goes	leaving	some	the scars	on	my hand

‘And then it comes back to beat in my veins then goes away leaving some scars on my hand’

The continuative use of *maḏā* given above in (60), which may also be interpreted as an inceptive usage, seems to be restricted to contexts where *maḏā* collocates with a very specific set of semantically-related verbs. The same functional aspect of *maḏā* usage is also apparent (61) and (62) in which *maḏā* is typically inflected in the perfective and is followed by a verb in the imperfective. The same construction is characteristic of the grammaticalized use of *rāḥa*. Again, in this construction, the verb *qāla* ‘say’ counts among the most frequent verbs that follow the grammaticalized *maḏā*, as in (61).

(61) و مضى الحريري يقول بعد اجتماعه مع وفد برلماني فرنسي

<i>wa=maḍā</i>	<i>alḥarīri</i>	<i>yaqūl</i>	<i>ba'da</i>	<i>iğtima'-ih</i>	<i>ma'a</i>
wa= maḍā .PERF.3SG.M	Al Hariri	say.IMP.3SG.M	ADV	meeting-CL.3SG.M.GEN	COM
and went	Al Hariri	says	after	his meeting	with

<i>wafd</i>	<i>barlamani</i>	<i>faransi</i>
delegate	parliamentary	French
delegate	parliamentary	French

‘And Al Hariri went on saying after his meeting with a French parliamentary delegate’

(62) ومضى الاثنان ينشدان معا

<i>wa=maḍā</i>	<i>al='iṭnān</i>	<i>ynšidān</i>	<i>ma'an</i>
CONJ= maḍā .PERF.3SG.M	ART=two	sing.AP.3DUAL.M	together
and went	the two	singing	together

‘And the two went on singing together’

In addition to verbs denoting ‘talking’ or ‘sound making’, a small number of instances – related to this grammaticalized usage of *maḍā* in the 500-line data frame – involve verbs of perception, such as *naẓara* ‘see’ in (63), and *ta'ammala* ‘stare/contemplate’ in (64).

(63) ومضيت أنظر إليهما من وراء وهما يشربان ويبتادلان الحديث

<i>wa=maḍaytu</i>	<i>anzuru</i>	<i>il=ayhimā</i>	<i>min</i>	<i>warā'</i>	<i>wa=humā</i>
CONJ= maḍā .PERF.1SG	look.IMP.1SG	ALL=CL.3DUAL.M	ABL	behind	CONJ=PP
and went	looking	at them	from	behind	and they

<i>yašrabān</i>	<i>wa=yatabādālān</i>	<i>al=ḥadīṭ</i>
drink.IMP.3DUAL.M	CONJ=exchange.IMP.3DUAL.M	ART=conversation
drinking	and exchanging	the conversation

‘And I went on looking at them as they were drinking and having a conversation’

(64) ثم تربع على أريكته ومضى يتأمل (زبائنه) بعينه نصف المغمضة

<i>tumma</i>	<i>tarabba'a</i>	<i>'alā</i>	<i>arīkat-ih</i>	<i>wa=maḍā</i>
CONJ	sit.cross-legged.PERF.3SG.M	LOC	couch-CL.3SG.M.GEN	CONJ= maḍā .PERF.3SG.M
then	sat cross-legged	on	his couch	and went

<i>yata'ammal</i>	<i>zabā'in-a-h</i>	<i>bi='ayn-ih</i>	<i>niṣf</i>	<i>al=muğmaḍa</i>
gaze.IMP.3SG.M	customers-ACC-CL.3SG.M.GEN	INST=eye-CL.3SG.M.GEN	half	ART=closed
gaze	his customers	with eye	half	the closed

‘Then he sat cross-legged on his couch and went on gazing at his customers with a half-closed eye’

As mentioned in the previous chapter, the HCFA analysis showed that the most robust configurations of variables involving the verb *maḍā* are likely to include a sentential subject denoting TIME. These instances of verb use seem to fall in line with the observations made so far with regards to the atelic nature of *maḍā* and the emphasis on the locomotion aspect of the verb rather than arrival at an endpoint. In the following constructions in which *maḍā* collocates with a subject denoting TIME, the atelic motion event is construed as unfolding along a stretched out path. The coded corpus hits I have inspected contained various distinct constructions in which *maḍā* signals the passage of time, most of which are idiomatic expressions as in (65): *akṭar min ay waqtin maḍā* ‘more than before/any time in the past’, (66): *munḍu [X time period] maḍā/maḍat* ‘since [X amount of time]’, and (67): *fīma maḍā* ‘in the past’.

(65) أكثر من أي وقت مضى

akṭar min ‘ay waqtin maḍā
 more ABL any time **maḍā**.PERF.3SG.M
 more from any time went
 ‘more than any time in the past’

(66) منذ ثلاثة أشهر مضت

munḍu talāṭat aṣḥur maḍat
 ADV three months **maḍā**.PERF.3SG.F
 since three months went
 ‘since 3 months’

(67) فيما مضى

fī=mā maḍā
 LOC=RP **maḍā**.PERF.3SG.M
 in what went
 ‘in the past’

More general constructions related to the passage of time (in which TIME is the mover) are included in sentences like (68) - (70).

(68) مضت سنوات وأنا أنتظر الولادة

<i>maḍat</i>	<i>sanawāt</i>	<i>wa=anā</i>	<i>antaḥiru</i>	<i>al=wilāda</i>
maḍā .PERF.3SG.F	years	CONJ=PP	wait.IMPF.1SG	ART=birth
went	years	and I	wait	the birth

‘Years have passed as I waited for the child birth’

(69) لكن الأيام تمضي

<i>lākin</i>	<i>al=’ayyām</i>	<i>tamḍī</i>
CONJ	ART=days	maḍā .IMPF.3SG.F
but	the days	go

‘But days go by’

(70) ولم تمض فترة طويلة حتى كان الوزير يرسل لي خطابا ثانيا

<i>wa=lam</i>	<i>maḍā</i>	<i>fatra</i>	<i>ṭawīla</i>	<i>ḥattā</i>	<i>kāna</i>	<i>al=wazīr</i>
CONJ=NEG	maḍā .JUSS.3SG.M	period	long	ADV	AUX	ART=minister
and did not	go	period	long	until	was	the minister

<i>yursil</i>	<i>l=ī</i>	<i>ḥiṭāban</i>	<i>tāniyan</i>
send.IMPF.3SG.M	ALL=CL.1SG	letter	second
send	to me	letter	second

‘shortly after, the minister sent me a second letter’

In some of the frequent uses of *maḍā*, the motion verb collocates with the preposition ‘*alā*’ ‘over/on’ to express temporal relations. In (71), for instance, a certain amount of time – ‘more than five years’ – passes over a continuous state – ‘my presence abroad’, and therefore indicating that this particular state has been going on for 5 years. In (72), however, a certain amount of time – ‘long time’ – passes over a specific (telic) event – ‘launching it’ – to indicate the amount of time that has gone by since the occurrence of that event.

(70) مضى على وجودي في الخارج أكثر من خمس سنوات

<i>maḍā</i>	<i>’lā</i>	<i>wujūd-i</i>	<i>fi</i>	<i>al=ḥāriġ</i>	<i>aḳtar</i>
maḍā .PERF.3SG.M	LOC	presence-CL.1SG.GEN	LOC	ART=outside	more
went	over	my presence	in	the outside	more

<i>min</i>	<i>ḥams</i>	<i>sanawāt</i>
ABL	five	years
from	five	years

‘I have been living abroad for more than five years’

(71) التي لم يمض وقت طويل على اطلاقها

<i>allatī</i>	<i>lam</i>	<i>yamḍi</i>	<i>waqtun</i>	<i>ṭawīl</i>	<i>'alā</i>	<i>iṭlāqi-ha</i>
RP	NEG	maḍā .JU.SS.3SG.M	time	long	LOC	launch.VN-CL.3SG.F.ACC
that	did not	go	time	long	over	launching it

'It wasn't launched too long ago'

4.3 *Rāḥa*

I previously mentioned that *rāḥa* tends to be treated formally as a fully grammaticalized verb, marking the inceptive and/or continuous aspect. I also stated that the reason I included this verb in the set of GO verbs is the fact that, regardless of its overwhelmingly grammaticalized uses, *rāḥa* can still feature in sentences where either physical or figurative motion is conveyed. Even though the non-grammaticalized uses count as a more marginal use of this verb – as we saw in the quantitative analysis of the verb – *rāḥa* can, in some cases, be used interchangeably with *ḍahaba* or *maḍā*. The following discussion introduces the less frequent, non-grammaticalized uses of this verb (both physical and metaphorical) in MSA, followed by more grammaticalized uses.

4.3.1 Physical and non-physical motion

Out of the entire set of 500 corpus returns for *rāḥa*, 90 instances were found to involve either physical or figurative motion. The majority of these returns involve the sub-sense 'to die'. This particular sub-sense is mostly demonstrated in the strongly collocational usage of the verb *rāḥa ḍaḥiyyata* 'went (died) as a victim of X', as in (73). In this construction, either *ḍahaba* or *rāḥa* can fill the GO verb slot, although *rāḥa* is more common.³²

(73) فضلا عن المليون قتيل الذين راحوا ضحية مباشرة للحرب

<i>faḍlan</i>	<i>'an</i>	<i>al=milyūn</i>	<i>qatīl</i>	<i>allaḍīn</i>	<i>rāḥū</i>	<i>ḍaḥiyya-tan</i>
ADV	about	ART=million	dead	RP	rāḥa .PERF.3PL.M	victim-ADV
beside	of	the million	dead	who	went	victim

³² This explains the robust configuration found in the previous chapter, in Table 11, in which HUMAN X MANNER featured around 67 times in the 500-line data frame of *rāḥa*.

mubāšira *li=l=ħarb*
 direct ALL=ART=war
 direct to the war

‘Beside the one million dead people who count as direct war casualties’

The collocational usage in (73) is not the only construction where *rāħa* expresses the notion of ‘dying’ or ‘perishing’. Other constructions, as in (74), express a similar sub-sense of *rāħa*.

(74) جميعهم راحوا بين قتلى وجرحى او تاهو في الغابات المجاورة

ġamī’u-hum *rāħū* *bayna* *qatlā* *wa=ġarħā* *aw* *tāiū*
 all-CL.3PL.M *rāħa*.PERF.3PL.M LOC dead CONJ=injured CONJ get.lost.PERF.3PL.M
 all of them went between dead and injured or got lost

fi *al=ġābāt* *al=muġāwira*
 LOC ART=forests ART=nearby
 in the forests the nearby

‘All of them are gone as some of them died while others were injured or got lost in the nearby woods’

Another common usage of *rāħa* involves motion away from the deictic centre or, generally, ‘leaving’, as in (75).

(75) مساء حياني... وراح!!

masā’ *ħayyā-ni* *wa=rāħ*
 evening greet.PERF.3SG.M-CL.1SG.ACC CONJ=*rāħa*.PERF.3SG.M
 evening greeted me and went

‘An evening greeted me and left!’

In all of the previous sentence the deictic facet of the fictive motion event plays an important role. Moving *away* from the deictic centre can be viewed as negative, as I mentioned earlier, and therefore can be equated with ‘going out of existence’ (Newman, 2000).

In terms of *rāħa* examples that highlight physical locomotion, the corpus returns contained a handful of physical motion events as in (76) and (77). Prescriptive Arabic textbooks and grammars advise against the use of this verb in such a sense, due to its association with colloquial usage.

(76) وراحوا وفتحوا باب الغرفة التي تنام فيها أسرة أبي صباح

<i>wa=rāḥū</i>	<i>wa=fataḥū</i>	<i>bāb</i>	<i>al=ḡurfa</i>	<i>allati</i>
CONJ= <i>rāḥa</i> .PERF.3PL.M	CONJ=open.PERF.3PL.M	door	ART=room	RP
and they went	and they opened	door	the room	which

<i>tanām</i>	<i>fī=ha</i>	<i>'usrat</i>	<i>abī ṣayyāḥ</i>
sleep.IMPF.3SG.F	LOC=CL.3SG.F	family	Abu Sayyah
sleep	in it	family	Abu Sayyah

‘And they went and opened the door to the room where Abu Sayyah’s family sleeps’

(77) وفي غيبة الوزير راح الملك الى الجارية في بيتها وراودها عن نفسها

<i>wa=fī</i>	<i>ḡaybat</i>	<i>al=wazīr</i>	<i>rāḥa</i>	<i>al=malik</i>	<i>ilā</i>	<i>al=ḡāriya</i>
CONJ=LOC	absence	ART=vizier	<i>rāḥa</i> .PERF.3SG.M	ART=king	ALL	ART=slave.girl
and in	absence	the vizier	went	the king	to	the slave girl

<i>fī</i>	<i>bayti-hā</i>	<i>wa=rāwada-hā</i>	<i>'an</i>	<i>naḡsihā</i>
LOC	house-CL.3SG.F.GEN	CONJ=seduce.PERF.3SG.M-CL.3SG.F.ACC	about	herself
in	her room	and seduced her	of	herself

‘And during the vizier’s absence, the king went to the slave girl’s house and tried to seduce her’

Even though these uses are relatively marginal they nevertheless are quite pervasive in a contemporary MSA corpus. Both verb uses in (76) and (77) depict purposeful motion, similar to what we saw with *ḡahaba*. In addition, the sentence in (77) involves motion towards a GOAL.

An even smaller sub-set of uses of *rāḥa* involves the collocational use of ‘GO and COME’ – as a means of expressing busy movement (Newman, 2000). The collocating COME verb is almost exclusively *ḡā'a* and both verbs are always inflected in the imperfective, as exemplified in (78) and (79).³³

(78) وتظل المسودات تروح وتجيء الى ان يتوصل الطرفان الى نص مقبول

<i>wa=tazal</i>	<i>al=miswaddāat</i>	<i>tarūḥ</i>	<i>wa=taḡī'</i>	<i>ilā</i>	<i>'an</i>
CONJ=stay.IMPF.3SG.F	ART=drafts	<i>rāḥa</i> .IMPF.3SG.F	CONJ= <i>ḡā'a</i> .IMPF.3SG.F	ALL	TOP
and remain	the drafts	go	and come	to	that

<i>yatawaṣṣal</i>	<i>al=ṡarafān</i>	<i>ilā</i>	<i>naṣṣ</i>	<i>maqbul</i>
reach.IMPF.3SG.M	ART=two.sides	ALL	text	acceptable

³³ Upon examination of all possible collocations of COME and GO and GO and COME verbs in ArabiCorpus, I found that as far as ‘GO and COME’ is concerned, the most frequent collocation is *rāḥa*:IMPF *wa ḡā'a*:IMPF, and to a much lesser extent *ḡahaba*:IMPF *wa ḡā'a*:IMPF. As for ‘COME and GO’ collocations, the most frequent collocates are *atā*:IMPF *wa ḡahaba*:IMPF, and only a couple of instances of *ḡā'a*:IMPF *wa ḡahaba*:IMPF and *atā*:IMPF *wa rāḥa*:IMPF. When the nominal form of the verb is used, as in ‘COMING and GOING’ or ‘GOING and COMING’, the collocational patterns are event different and other motion verbs not discussed here are used.

reach the two sides to text acceptable
 ‘And the drafts keep going and coming until the two sides agree on an acceptable text’

(79) كان يروح ويجيء ويردد أمام من يراه

<i>kāna</i>	<i>yarūḥ</i>	<i>wa=yagī'</i>	<i>wa=yraddid</i>
AUX	<i>rāḥa</i> .IMPF.3SG.M	CONJ= <i>gā'a</i> .IMPF.3SG.M	CONJ=repeat.IMPF.3SG.M
he was	go	and come	and repeat

<i>amām</i>	<i>man</i>	<i>yarā-hu</i>
LOC	RP	see.IMPF.3SG.M-CL.3SG.M.ACC
in front	who	sees him

‘He used to go and come and repeat in front of whoever he meets’

4.3.2 Grammaticalized function

Out of the 500 coded *rāḥa* corpus returns, 410 represent instances of *rāḥa* being used as an aspectual marker. As stated earlier, the inherently atelic nature of a GO motion event makes this verb an ideal source for motivating durative, continuative, and persistent markers across many world languages (Newman, 2000).

It is not always simple to tease apart the particular aspectual force that *rāḥa* lends to a construction since the internal event structure of the verb collocating with *rāḥa*, as well as the presence of adverbials that shape the event structure in a particular usage, can both contribute to the interpretation of the grammaticalized function of the MSA *rāḥa* verb. For instance, in (80) the atelic verb *tanḥani* ‘lean over’ collocates with *rāḥa* to indicate the beginning stages of leaning over and, possibly, the elongation of the duration of that event. The added adverbial *bi buṭ'* ‘slowly’ undoubtedly strengthens the durative aspect of the event.

(80) وتروح تنحني في بطة ناحية الحجر

<i>wa=tarūḥ</i>	<i>tanḥanī</i>	<i>fī</i>	<i>buṭ'</i>	<i>nāḥiyat</i>	<i>al=ḥaḡar</i>
CONJ= <i>rāḥa</i> .IMPF.3SG.F.	lean.over.IMPF.3SG.F	LOC	slowness	ADV	ART=rock
and goes	lean over	in	slowness	toward	the rock

‘And she goes on leaning over slowly towards the rock’

In (81), another atelic verb, *yanḥasir* ‘decrease/decline’ collocates with *rāḥa*, again indicating a combination of inceptive aspect and durative or continuous aspect. In this sentence, as well, the durative sense is heightened by the adverbial *šay’an fašay* ‘little by little’.

(81) وراح الفارق الزمني ينحسر شيئاً فشيئاً

<i>wa=rāḥā</i>	<i>al=fāriq</i>	<i>al=zamāni</i>	<i>yanḥasir</i>	<i>šay’an</i>	<i>fa=šay’an</i>
CONJ= <i>rāḥa</i> .PERF.3S	ART=difference	ART=time.	decrease.IMPF.3S	somethi	CONJ=someth
G.M	nce	ADJ	G.M	ng	ing
and went	the	the time	decrease	somethi	and
	difference			ng	something

‘And the time difference went on decreasing little by little’

A temporal adverbial as in (82) *munḍu ḡālīka alḥīn* ‘since then’ seems to also add to the inceptive marking function of *rāḥa*, yet the presence of the adverbial *akṭar wa akṭar* ‘more and more’ seems to bias our interpretation of the event, ‘tightening the grip’, as being extended in duration.

(82) وراحت منذ ذلك الحين تضيق الخناق عليه أكثر وأكثر

<i>wa=rāḥat</i>	<i>munḍu</i>	<i>ḡālīka</i>	<i>al=ḥīn</i>	<i>tudayyiq</i>
CONJ= <i>rāḥa</i> .PERF.3SG.F	ADV	DEM	ART=time	tighten.IMPF.3SG.F
and went	since	that	the time	tightening

<i>al=ḥināq</i>	<i>al=ayh</i>	<i>akṭar</i>	<i>wa=akṭar</i>
ART=grip	LOC=CL.3S.M	more	CONJ=more
the grip	on him	more	and more

‘And since then it [i.e. Washington] kept on/went on tightening the grip on him more and more’

The atelic *rāḥa* can also strengthen an iterative reading when combined with iterative or telic verbs. For instance, in (83) *rāḥa* collocates with *raddada* ‘to repeat’, a verb indicating a repetitive event (saying the same thing over and over). The resulting effect is an intensified repetitiveness of the event ‘kept on repeating it’. Note also that the phrase ‘in more than one language’ adds to this iterative construal.

(83) ورحت أردده على نفسي بأكثر من لغة

<i>wa=ruḥtu</i>	<i>'uraddidu-hu</i>	<i>'alā</i>	<i>nafs-ī</i>
CONJ= <i>rāḥa</i> .PERF.1SG	repeat.IMPF.1SG-CL.3SG.M.ACC	LOC	self-CL.1SG
and I went	repeat it	on	myself

<i>bi=akṭar</i>	<i>min</i>	<i>luḡa</i>
INST=more	ABL	language
with more	from	language

'And I went on repeating it to myself in more than one language'

On the other hand, *rāḥa* can also combine with a punctual event, such as *našaba* 'to break out/erupt' in (84), to add a sense of iterativity. That is to say, the punctual event of a battle erupting took place more than once. Note that in both examples (83) and (84), the inceptive sense indicated by *rāḥa* is still present to a certain degree.

(84) وراحت المعارك المتفرقة تنشب عند حدودها مع سورية والأردن

<i>wa=rāḥat</i>	<i>al=ma'ārik</i>	<i>al=mutafarriqa</i>	<i>tanšab</i>	<i>'inda</i>
CONJ= <i>rāḥa</i> .PERF.3SG.F	ART=battles	ART=separate	erupt.IMPF.3SG.F	LOC
and went	the battles	the separate	erupting	at

<i>ḥudūdi-ha</i>	<i>ma'a</i>	<i>sūriya</i>	<i>wa=l='urdun</i>
borders-CL.3SG.F	COM	Syria	CONJ=ART=Jordan
its borders	with	Syria	and the Jordan

'And separate battles kept erupting at the borders it [i.e. Israel] shares with Syria and Lebanon'

Interestingly, the verb *qāla* 'say', seems to collocate almost exclusively with *maḏā*, as discussed in §4.2.2, rather than with *rāḥa* to indicate continuative or durative aspect. I did, however, find a few instances among the *rāḥa* coded sentences where the latter collocates with *qāla* 'say', as in (85) and (86).

(85) وراح يقول مدافعا عن جيله

<i>wa=rāḥa</i>	<i>yaqūl</i>	<i>mudāfi'an</i>	<i>'an</i>	<i>ḡīl-ih</i>
CONJ= <i>rāḥa</i> .PERF.3SG.M	say.IMPF.3SG.M	defend.AP.3SG.M	about	generation-CL.3SG.M.GEN
and went	say	defending	of	his generation

'And he went on saying, defending his generation ...'

(86) راحوا يقولون للشبان بمرارة يصعب كتمانها

rāhū	yaqūlūn	li=l=šabāb	bi=marārah
<i>rāḥa</i> .PERF.3PL.M	say.IMP.F.3PL.M	ALL=ART=youth	INST=bitterness
they went	say	to the young men	with bitterness

yaṣ'ub	kutmānu-ha
be.hard.IMP.F.3SG.M	hide.VN-CL.3SG.F.GEN
it is hard	hiding it

'They went on saying to the young people with such a bitterness that was hard to hide'

Both (85) and (86) are felicitous if we substitute *rāḥa* with *maḍa* and would have similar interpretation ('went on saying...'). As stated previously, *maḍā*, as an aspectual marker, does collocate with other verbs besides 'say', yet such uses are marginal. Such patterns of lexical restrictions, per *maḍā* and *rāḥa*, call for further psycholinguistic investigation in order to arrive at a better description of the aspectual function of the two verbs.

4.4 Summary

The previous quantitative and qualitative analyses of *ḍahaba*, *maḍā*, and *rāḥa* have highlighted the morphosyntactic and lexico-semantic properties of each of these three GO verbs in MSA. Pairing a qualitative, case-by-case analysis with the statistical, quantitative analysis proved to be a rather comprehensive method that helped identify the general patterns and the larger constructions attracting each verb, in addition to identifying the less frequent patterns of verb use. In sum, each of the three MSA GO verbs was found to be associated with a specific set of uses and sub-senses, which I summarize in the following:

The verb *ḍahaba* strongly associates with a wide range of morphosyntactic features (e.g. TAM inflection patterns). Similarly, this verb collocates with sentential subjects belonging to a variety of semantic categories (e.g. HUMAN, EVENT, NOTION, PHYSICAL OBJECT/ARTIFACT). It is no surprise, then, that this verb is associated with numerous sub-senses and uses in Modern Standard Arabic. The quantitative analysis also showed that a motion event frame (both physical and figurative) involving the verb

ḍahaba is most likely to include an end point and/or a statement of purpose, something which was not found with the other two GO verbs. *Ḍahaba* can, therefore, be regarded as the MSA GO verb used when talking about a motion event that is purposeful and telic. In addition, the qualitative analysis highlighted less statistically robust, yet common uses of the verb such as the phrasal *ḍahaba bi-* ‘take something somewhere’ or Lit. ‘go with X’, and a number of collocational uses including *ḍahaba ḍaḥiyyata* ‘be a victim of X’ or Lit. ‘go as a victim of X’, which counts as one of the few instances in which this verb overlaps in usage with the verb *rāḥa*.

Maḍā, on the other hand, was found to associate with a less diverse set of constructional features. For instance, this verb was found to collocate most predominantly with a sentential subject denoting TIME. None of the other GO verbs seems to overlap with *maḍā* in that respect. Physical motion events construed with *maḍā* are considered among the common uses of the verb but are less frequent than the figurative uses. For instance, even though this verb collocates strongly with sentential subjects referring to HUMAN or GROUP participants, a large number of these constructions express a ‘go on’ or a ‘go ahead’ sub-sense. Generally speaking, as mentioned in this chapter, *maḍā* can be considered the atelic GO verb in MSA. In the majority of the corpus returns of this verb it is the locomotion aspect of a GO event that is emphasized rather than arrival at a GOAL, which appears to motivate the use of *maḍā* for expressing the passage of time. Moreover, this atelic property may explain the grammaticalized uses of this verb as a durative/continuous aspect marker, which I discussed earlier in this chapter.

Finally, the majority of corpus returns of *rāḥa*, as expected, pertained to the grammatical function that this verb has in MSA – marking an inceptive and/or continuative aspect. Nevertheless, the motion sense of this verb is not entirely bleached out in contemporary usage. For instance, I pointed out that the collocational pattern *rāḥa:IMPF wa ḡā’a:IMPF* ‘go and come’ counts among the common uses of this verb

(though statistically not robust). In addition, *rāḥa* does appear in constructions that denote physical motion towards a GOAL or, simply, away from the speaker, as well as figurative motion such as in *rāḥa ḍaḥiyyata* ‘go as a victim of X’, Lit. ‘be a victim of X’.

In sum, I can safely state that there is a small window of overlap among these three MSA GO verbs. In the most part, however, each verb has a highly specialized lexico-syntactic profile in the language. It would be unrealistic and dishonest to refer to this set of verbs as ‘synonyms’. The following quantitative and qualitative analyses on MSA COME verbs in Chapters 5 and 6, respectively, will further emphasize the notion that the even though more than one lexeme may express a deictic motion event in MSA, these verbs are by no means identical in their usage.

Chapter Five
Quantitative analysis of MSA COME verbs *atā*, *haḍara*, *ḡā'a*, and *qadima*

The quantitative analysis conducted on COME verbs in this chapter mirrors the analysis of GO verbs presented in Chapter 3, and is therefore based on examining the annotated COME data frame (2,000 lines of annotated corpus hits) through multiple monivariate and multivariate statistical analyses. The variable set examined in this chapter is identical to that reported in Chapter 3.

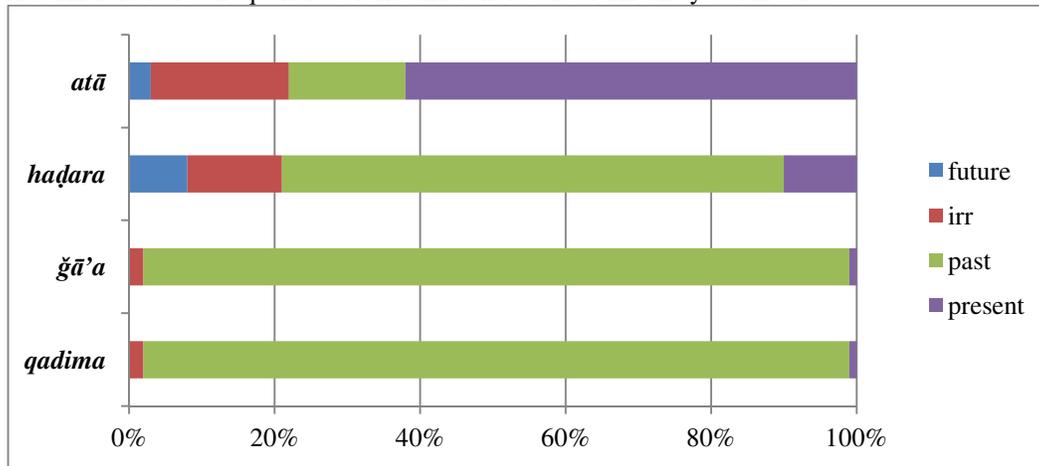
5.1 Single variable distribution

As a first step towards discerning constructional patterns particular to each of the targeted COME verbs, we can examine the raw or proportional frequencies of the occurrence of various variables across the four verbs. To exemplify, Table 1 profiles a cross-tabulation of the proportional frequencies of occurrence for each of four levels of the variable TENSE (PRESENT, PAST, FUTURE, and IRR) for each of the four COME verbs. As expected, the observed cell-wise values in this table diverge across the different tenses as well as across the different verbs. A *chi*-square test provides further evidence that this distribution is significantly heterogeneous: $X^2 = 1191.187$, $df = 9$, $p\text{-value} < 2.2e-16$. The distribution of the variable TENSE can also be represented visually in Figure 1.

TABLE 1. Proportional frequencies of the different levels of TENSE by COME verb.

TENSE	<i>atā</i>	<i>haḍara</i>	<i>ḡā'a</i>	<i>qadima</i>
FUTURE	0.03	0.08	0.00	0.00
IRR	0.19	0.13	0.02	0.02
PAST	0.16	0.69	0.97	0.97
PRESENT	0.62	0.10	0.01	0.01
<i>total</i>	1	1	1	1

FIGURE 1. Relative frequencies of the different levels of TENSE by COME verb.



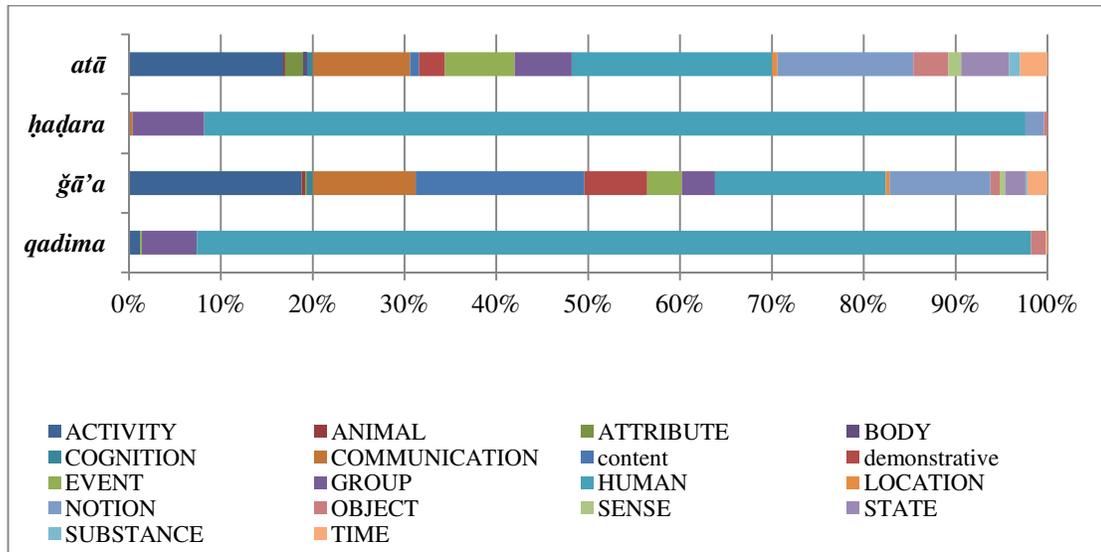
In Chapter 3, I examined the distribution of subject collocates that distinguish the use of the three GO verbs and found that each GO verb seems to collocate with a different set of sentential subjects. Similarly, Table 2 shows the proportional frequencies for the four COME verbs the different levels of the variable SUBJECT SEMANTIC CATEGORY, and Figure 2 provides a bar graph representation of this same distribution.³⁴

³⁴ The distribution in Table 1 does not meet the required conditions to conduct a *chi*-square test, since we have a large number of cells in this table and therefore the expected cell-wise values for a number of cells were less than 5 occurrences. It is nevertheless possible to discern, from looking at both Table 3 and Figure 2 that this distribution is heterogeneous.

TABLE 2. Proportional frequencies of the different levels of SUBJECT SEMANTIC CATEGORY by COME verb.

SUBJECT SEMANTIC CATEGORY	<i>atā</i>	<i>ḥaḍara</i>	<i>ḡā'a</i>	<i>qadima</i>
ACTIVITY	0.17	0.00	0.19	0.01
ANIMAL	0.00	0.00	0.00	0.00
ATTRIBUTE	0.02	0.00	0.00	0.00
BODY	0.00	0.00	0.00	0.00
COGNITION	0.01	0.00	0.01	0.00
COMMUNICATION	0.11	0.00	0.11	0.00
content	0.01	0.00	0.18	0.00
demonstrative	0.03	0.00	0.07	0.00
EVENT	0.08	0.00	0.04	0.00
GROUP	0.06	0.08	0.04	0.06
HUMAN	0.22	0.89	0.19	0.91
LOCATION	0.01	0.00	0.00	0.00
NOTION	0.15	0.02	0.11	0.00
OBJECT	0.04	0.00	0.01	0.02
SENSE	0.01	0.00	0.01	0.00
STATE	0.05	0.00	0.02	0.00
SUBSTANCE	0.01	0.00	0.00	0.00
TIME	0.03	0.00	0.02	0.00
<i>total</i>	1	1	1	1

FIGURE 2. Relative frequencies of the different levels of SUBJECT SEMANTIC CATEGORY by COME verb.



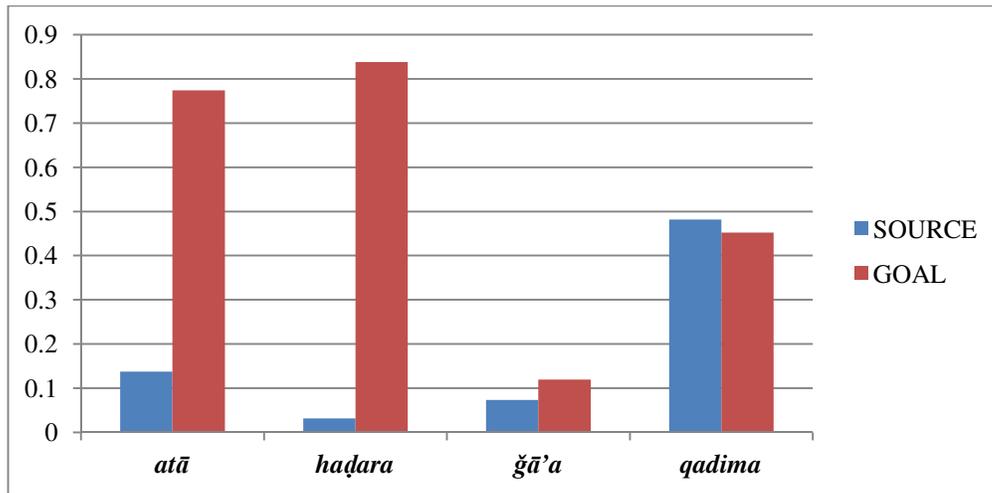
It is obvious from both Table 2 and Figure 2 that each COME verb co-occurs with a different set of sentential subjects. For instance, both *atā* and *ǧā'a* seem to collocate with a wide variety of subjects (e.g. HUMAN, ACTIVITY, COMMUNICATION, EVENT). *Haḍara* and *qadima*, on the other hand, seem to both favor sentential subjects denoting HUMAN agents and, to a lesser degree, GROUPS, which is the category denoting metonymical representation of collective humans such as organizations, countries, newspapers, etc. Looking back at Figure 2 in Chapter 3 that plots the distribution of the same variable for the three GO verbs, it is interesting to see the wider variety of subject collocates that would co-occur with COME as opposed to those associated with GO verbs in general.

Another interesting pair of variables involves the occurrence of phrases that indicate the GOAL or the SOURCE of the motion event, across the four COME verbs. Table 3 lists the raw proportions of occurrence of each variable (SOURCE and GOAL) per verb, and this distribution is represented in Figure 3. We can see that, as far as the GOAL of the motion event is concerned, sentences featuring the verbs *atā*, *haḍara* and *qadima* to a large extent seem to include a phrase that signals the destination or the GOAL of the motion event. The majority of *ǧā'a* sentences, however, do not seem to involve a GOAL. On the other hand, while over 70% of *atā* and *haḍara* motion events seem to have an end point, very few sentences involving these two verbs would include a SOURCE of the motion event (*come from*). In contrast, almost 50% of *qadima* uses do indicate the SOURCE of the motion event.

TABLE 3. Proportional frequencies of the occurrence of phrases denoting SOURCE and GOAL of motion by COME verb.

VERB	<i>atā</i>	<i>haḍara</i>	<i>ǧā'a</i>	<i>qadima</i>
SOURCE	0.14	0.03	0.07	0.48
GOAL	0.77	0.84	0.12	0.45

FIGURE 3. Proportional frequencies of the occurrence of phrases denoting SOURCE and GOAL of motion by COME verb.



As mentioned earlier, running a *chi*-square test on a cross-tabulation of two variables as in Tables 1 and 2 helps us determine whether the variable distribution is significantly heterogeneous or not. In Chapter 3, I discussed an additional measure – standardized Pearson’s residuals – that tests the extent to which the observed values diverge from the expected values in each cell in a pair-wise interaction (VERB X variable). Instead of examining these cell-wise values of divergence, the `chisq.posthoc()$cells$std.pearson.residuals.sign` function – which is part of the `{polytomous}` package (Arppe, 2012) – can be used to retrieve the + / - / 0 values that are representative of the standardized Pearson’s residuals: ‘+’ refers to cell-wise co-occurrence values significantly higher than expected; ‘-’ refers to co-occurrence values significantly lower than expected; while ‘0’ refers to co-occurrence values that are close to expected values. Table 4 lists a selection of variables and the estimated standardized Pearson’s residuals represented by the signs + / - / 0. This table provides further evidence of the heterogeneous distributions of variables in the COME data frame.

TABLE 4. Preferences for the distribution of selected logical variables per COME verb.

FEATURE	<i>atā</i>	<i>hadara</i>	<i>ǧā'a</i>	<i>qadima</i>
MORPH_ASP.MOOD.IMPF.TRUE	+	–	–	–
TENSE.PAST.TRUE	–	0	+	+
TRANSITIVITY.YES.TRUE	–	+	–	–
SUBJ_NUM.PL.TRUE	–	–	–	+
SUBJ_NUM.SING.TRUE	+	+	+	–
SUBJ_CAT.HUMAN.TRUE	–	+	–	+
SUBJ_CAT.ACTIVITY.TRUE	+	–	+	–
GOAL.YES.TRUE	–	+	–	+
SOURCE.YES.TRUE	–	–	–	+

5.2 Hierarchical agglomerative cluster analysis

We can now examine the effect of the distributions of all variables combined on the clustering of the four COME verbs, and for these purposes I will be discussing the Behavioral Profiles method advanced by Gries and colleagues (e.g. Gries, 2006; Gries and Divjak, 2006; Gries and Otani, 2010). As discussed in Chapters 2 and 3, the cluster method employed in the quantitative analysis of COME and GO verbs relies on generating a table that lists relative frequencies, or proportions, of co-occurrence between the lexical items under investigation and every dependent variable included in the cluster analysis. Table 5 shows an excerpt of the co-occurrence table generated by BP 1.01 script which also resembles the relative frequency tables in saw in §5.2.1. This table displays co-occurrence proportions of the different levels of ID tags (dependent variables) with each COME verb, such that the total sum of proportions within a certain ID tag for each verb is 1.0. The Behavioral Profile of a verb in this table is, therefore, the vector of co-occurrence proportions of ID tags per verb.

TABLE 5. Sample of co-occurrence table generated by the BP 1.01 script.

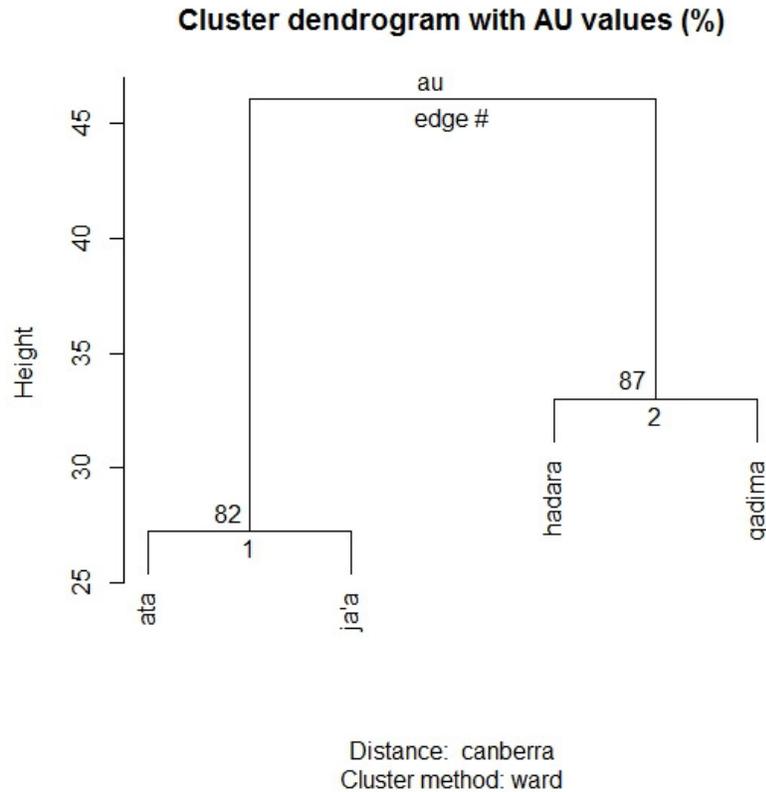
IDTAG	IDTAG-LEVEL	<i>ata</i>	<i>hadara</i>	<i>ga'a</i>	<i>qadima</i>	
MORPH_ASP/MOOD	IMPF	0.752	0.208	0.006	0.014	} columns sum to 1.0
	JUSS	0.042	0.042	0	0	
	PERF	0.126	0.692	0.992	0.984	
	SUBJN	0.08	0.058	0.002	0.002	
SUBJ_NUM	DUAL	0.002	0.004	0.002	0.04	} columns sum to 1.0
	PL	0.062	0.07	0.03	0.486	
	SING	0.936	0.926	0.968	0.474	
GOAL	NO	0.782	0.162	0.888	0.548	} columns sum to 1.0
	YES	0.218	0.838	0.112	0.452	

This table can now be subjected to hierarchical agglomerative cluster analysis. This method of analysis can handle a large number of variables, and the resulting clusters group together items that are similar to one another and dissimilar to other items in other clusters. In the case of the COME data set, it would be interesting to see which verbs cluster together, based on all 20+ variables and their 70+ levels. As I mentioned in Chapter 3, I decided to follow earlier Behavioral Profiles studies in which the (dis)similarity metric used is ‘Canberra’, and the amalgamation rule that computes a cluster structure is ‘Ward’.³⁵

The dendrogram in Figure 1 shows two major divides between the four verbs that the hierarchical agglomerative cluster analysis deemed significant. The first cluster formed in this analysis appears to group the verbs *atā* and *ǧā’a* together, while the other cluster groups *ḥaḍara* and *qadima* together. Here, we find that the AU *p*-value (Approximately Unbiased) – which is a probability measure computed through multi-scale bootstrap resampling – for the cluster containing *ḥaḍara* and *qadima* is calculated to approximate 87%, while the AU *p*-value for the cluster of *atā* and *ǧā’a* is 82%. Again, this does not necessarily imply that *ḥaḍara* and *qadima* are highly similar, but that they are very dissimilar from *atā* and *ǧā’a*. The following hierarchical configural frequency analysis will help us identify the constructional differences across the four verbs.

³⁵ For a detailed description of this clustering method see Gries (2009), pp 306-319.

FIGURE 4. Dendrogram based on the COME multivariate data frame.



5.3 Hierarchical configural frequency analysis

Recall from Chapter 3 that HCFA helped us identify robust patterns of variable interaction within the large GO data frame and, thus, zero in on specific constructional (morphosyntactic and semantic) properties that characterized the use of each GO verb. Similar to the procedure discussed in §3.2.3 I will examine various patterns of co-occurrence among different types of variables, which will involve running an HCFA analysis on particular subsets of variables, each at a time. The following analysis starts with a monivariate investigation of general morphological types and antitypes, moves on to examining interactions between morphological variables, subject-related variables, and finally to event construal-related variables per verb.

Morphological types and antitypes

Prior to examining significant and highly frequent configurations of variables that characterize the use of each COME verb, we can employ HCFA as an additional monovariate measure for investigating the distribution of morphological and TAM inflectional variables: ASPECT, MORPHOLOGICAL ASPECT AND MOOD, SUBJECT NUMBER, PERSON and GENDER. The analysis yielded the following *types* listed in Table 6 and *antitypes* listed in Table 7.³⁶ Note that the singular variable distributions reported in Tables 6 and 7 pertain to the entire COME data frame as a whole and not bound to a particular COME verb.

TABLE 6. Most significant univariate *types* for TENSE, ASPECT, MORPHOLOGICAL ASPECT AND MOOD, SUBJECT NUMBER, PERSON and GENDER.

TENSE	ASPECT	MORPH_ASP. MOOD	SUBJ_ NUM	SUBJ_ PER	SUBJ_ GEN	Freq	Exp	Obs-exp	Dec	Q
.	.	.	.	3RD	.	1926	666.6667	>	***	0.944
.	SIMPLE	1654	333.3333	>	***	0.792
.	.	.	SING	.	.	1652	666.6667	>	***	0.739
.	MASC	1501	666.6667	>	***	0.626
.	.	PERF	.	.	.	1397	500	>	***	0.598
PAST	1396	500	>	***	0.597

TABLE 7. Most significant univariate *antitypes* for TENSE, ASPECT, MORPHOLOGICAL ASPECT AND MOOD, SUBJECT NUMBER, PERSON and GENDER.

TENSE	ASPECT	MORPH_ASP. MOOD	SUBJ_ NUM	SUBJ_ PER	SUBJ_ GEN	Freq	Exp	Obs-exp	Dec	Q
.	.	.	.	2 ND	.	15	666.6667	<	***	0.489
.	.	.	DUAL	.	.	24	666.6667	<	***	0.482
.	NIL	49	666.6667	<	***	0.463
.	.	.	.	1 ST	.	59	666.6667	<	***	0.456
.	.	JUSS	.	.	.	42	500	<	***	0.305
FUT	53	500	<	***	0.298
.	.	SUBJN	.	.	.	71	500	<	***	0.286
.	.	.	PL	.	.	324	666.6667	<	***	0.257
IRR	179	500	<	***	0.214

³⁶ Recall from Chapter 3 that *types* is the technical term employed by von Eye (1990) in CFA analysis to refer to configurations of variables that occur significantly more times than expected, while *antitypes* refers to configurations of variables that occur significantly less times than expected.

The above patterns of inflectional variable distribution certainly look familiar. We can see a resemblance to the patterns of morphological inflections associated with MSA GO verbs discussed earlier. Irrespective of verb, and for all the 2,000 annotated corpus hits of this data frame, it seems that the major trend with respect to TENSE, ASPECT, MORPHOLOGICAL ASPECT AND MOOD marking is PAST, SIMPLE, PERFECTIVE. We can also see that the verbs seem to be typically inflected for 3RD PERSON, SINGULAR, MASCULINE. Again, this may well be a property of COME verbs usage exhibited in newspaper writing.

Table 7, on the other hand, lists a number of inflectional elements that MSA COME verbs typically do *not* associate with. Combining results from both tables, we can see that, for instance, PERSON inflection on the verb is highly skewed towards 3RD PERSON, while 1ST and 2ND person inflections are quite infrequent in this data frame. The same applies to NUMBER inflection, where SINGULAR accounts for a large proportion of NUMBER agreement on COME verbs. However, PLURAL and, especially, DUAL inflections occur significantly less often than SINGULAR does. As expected, these findings show strong evidence for the notion of “inflectional islands” (Rice and Newman, 2004), that inflected forms in a paradigm are distributionally skewed and not equally represented in the usage of a certain lexical item.

VERB x TENSE x ASPECT x MORPHOLOGICAL ASPECT AND MOOD

Table 8 shows the most significant configurations for VERB x TAM – or TENSE, ASPECT (grammatical) and MOOD (or in the case of Arabic morphosyntax: morphological ASPECT and MOOD). In this table I only report on *types*, since these would constitute the most prototypical TAM markings that characterize the inflected forms of COME verbs.

TABLE 8. Most significant configurations for TENSE, ASPECT, and MORPHOLOGICAL ASPECT AND MOOD for all COME verbs.

VERB	TENSE	ASPECT	MORPH_ASP.MOOD	Freq	Exp	Obs-exp	Dec	Q
<i>atā</i>	PRES	SIMPLE	IMPF	206	18.8432	>	***	0.094
<i>atā</i>	PRES	HAB	IMPF	105	1.5722	>	***	0.052
<i>atā</i>	IRR	NON-FIN	SUBJN	40	0.1239	>	***	0.02
<i>atā</i>	IRR	NON-FIN	IMPF	32	0.8552	>	***	0.016
<i>atā</i>	IRR	NON-FIN	JUSS	17	0.0733	>	***	0.008
<i>atā</i>	FUT	SIMPLE	IMPF	14	2.6846	>	***	0.006
<i>ḥaḍara</i>	PAST	SIMPLE	PERF	339	201.6032	>	***	0.076
<i>ḥaḍara</i>	FUT	SIMPLE	IMPF	38	2.6846	>	***	0.018
<i>ḥaḍara</i>	IRR	NON-FIN	SUBJN	29	0.1239	>	***	0.014
<i>ḥaḍara</i>	IRR	NON-FIN	JUSS	20	0.0733	>	***	0.01
<i>ḥaḍara</i>	PRES	HAB	IMPF	11	1.5722	>	***	0.005
<i>ǧā'a</i>	PAST	SIMPLE	PERF	484	201.6032	>	***	0.157
<i>qadima</i>	PAST	SIMPLE	PERF	467	201.6032	>	***	0.148
<i>qadima</i>	PAST	PERT	PERF	15	2.0721	>	***	0.006

We can see in this table that the verb *atā* is most likely to be used in the SIMPLE, PRESENT, IMPERFECTIVE, and to a lesser degree in the HABITUAL. *ǧā'a*, *ḥaḍara*, and *qadima*, on the other hand, all seem to appear almost exclusively in the SIMPLE, PAST, PERFECTIVE. *Atā* and *ḥaḍara* seem to associate with a wider range of TAM marking than *ǧā'a* and *qadima* do. For instance, both *atā* and *ḥaḍara* may appear in a non-finite form (cases where TENSE and ASPECT marking is indicated by IRR). Non-finite uses of these two verbs paired with the SUBJUNCTIVE or the JUSSIVE mood can be reflective of a negative or an infinitive construction. In addition, we can see in this table that only *atā* and *ḥaḍara* would appear in FUTURE constructions.

VERB x SUBJECT NUMBER x PERSON x GENDER x SEMANTIC CATEGORY

I grouped SUBJECT NUMBER, PERSON and GENDER together with TAM variables in Tables 6 and 7 to examine the general patterns of morphological marking on all annotated uses of the four COME verbs in the data frame. However, the explanation of every individual verb's preferences with respect to SUBJECT NUMBER, PERSON and

GENDER variables is strongly dependent on including the semantic properties of the sentential subject in the analysis. I therefore re-ran those three variables in the HCFA script together with SUBJECT SEMANTIC CATEGORY. Table 9 shows the most significant configurations of values found for the interaction between VERB and these four variables.

TABLE 9. Most significant configurations for SUBJECT NUMBER, PERSON, GENDER, and SEMANTIC CATEGORY for all COME verbs.

VERB	SUBJ_NUM	SUBJ_PER	SUBJ_GEN	SUBJ_CAT	Freq	Exp	Obs-exp	Dec	Q
<i>atā</i>	SING	3RD	FEM	ACTIVITY	43	8.2328	>	***	0.017
<i>atā</i>	SING	3RD	FEM	EVENT	32	2.5951	>	***	0.015
<i>atā</i>	SING	3RD	FEM	NOTION	33	6.1298	>	***	0.013
<i>atā</i>	SING	3RD	FEM	COMMUNI- CATION	24	4.9665	>	***	0.01
<i>atā</i>	SING	3RD	FEM	OBJECT	14	1.1633	>	***	0.006
<i>ḥaḍara</i>	SING	3RD	MASC	HUMAN	369	164.6162	>	***	0.111
<i>ḥaḍara</i>	SING	3RD	FEM	GROUP	25	5.2797	>	***	0.01
<i>ǧa'a</i>	SING	3RD	MASC	CONTENT	92	14.4767	>	***	0.039
<i>ǧa'a</i>	SING	3RD	FEM	ACTIVITY	42	8.2328	>	***	0.017
<i>ǧa'a</i>	SING	3RD	MASC	DEMONSTR- ATIVE	34	7.1637	>	***	0.013
<i>ǧa'a</i>	SING	3RD	MASC	ACTIVITY	52	27.4609	>	*	0.012
<i>ǧa'a</i>	SING	3RD	FEM	COMMUNI- CATION	24	4.9665	>	***	0.01
<i>qadima</i>	PL	3RD	MASC	HUMAN	238	32.2855	>	***	0.105
<i>qadima</i>	DUAL	3RD	MASC	HUMAN	19	2.3915	>	***	0.008
<i>qadima</i>	SING	3RD	FEM	GROUP	18	5.2797	>	*	0.006

We can see here that the most prominent subject-related marking on the verb is 3RD SINGULAR. Interestingly, only the verb *qadima* deviates from this pattern in that it is mostly marked in the PLURAL, in addition to it being marked in the DUAL form at least 19 out of the 24 times this NUMBER inflection appears in the entire COME data frame. This may well be a by-product of the fact that *qadima* is mostly used to talk about the physical motion of HUMAN agents. As far as GENDER marking goes, there does not appear to be a striking pattern of distribution except that the significant *atā* configurations reported in this table include FEMININE marking on the verb. The most frequent *ḥaḍara* and *qadima* configurations appear to involve MASCULINE gender marking.

As far as SUBJECT NUMBER goes, the predominance of SINGULAR marking on all verbs – with the exception of *qadima* – may superficially suggest that the other three verbs only take SINGULAR subjects. However, this particular finding is more reflective of the fact that *qadima* associates with HUMAN subjects more often than the other verbs do. One of the rules of number agreement in MSA is that the verb is allowed to be inflected in the plural only when the subject argument precedes the predicate, and the plural subject argument refers to a (feminine or masculine) HUMAN agent (e.g. *men*, *women*, etc.). Otherwise, non-human plural subjects – and even human plural subjects in a VS construction – would cause the verb to be marked for the singular. Since I did not code the verb uses for subject-verb order, this observation should also suggest that these prototypical *qadima* constructions also involve an SV ordering.

Along the same lines, we may also assume that the patterns of GENDER marking we see in Table 9 (e.g. *atā* being mostly marked in the FEMININE) could be attributable to the plural system of Arabic. In Arabic, a plural noun denoting an inanimate entity is always treated morphosyntactically as a singular feminine noun (whether its singular form bears a feminine or a masculine syntactic gender). The current set of morphological features coded for in this data frame does not include the grammatical gender of the subject noun in the singular form nor the number of the sentential subject, as opposed to number agreement on the verb. It is, therefore, not possible to make any conclusive comments about the NUMBER and GENDER patterns in configurations where the verb is marked for SINGULAR FEMININE, at this point. Nevertheless, we may have a more transparent interaction between the variables in the most frequent *ḥaḍara* and *qadima* configurations, where we see a predominance of HUMAN subject collocates with MASCULINE gender marking on the verb.

The kinds of sentential subjects each verb associates with seem to considerably vary among the four verbs. Entities such as ACTIVITY, EVENT, NOTION and

COMMUNICATION seem to be the most frequent types of subjects in an *atā* sentence (with a typical 3RD, SINGULAR, FEMININE agreement on the verb). Again we need to be aware that this reflects newspaper writing genre more specifically and therefore this pattern might slightly change if we had had more modern literature verb uses in the annotated data frame. These instances of *atā* usage show that this verb is mostly used to talk about figurative motion of inanimate, abstract entities.

Similar to *atā*, *ǧā'a* also seems to appear with a wider range of sentential subjects than *haḍara* and *qadima* do. More specifically, though, it seems to be mostly used in “prefabs” that are very characteristic of newspaper writing, such as the constructions involving the use of a demonstrative as a sentential subject or what I systematically coded as ‘content’ of a document or a speech, etc. Generally speaking, *ǧā'a* also appears to be mostly used to talk about figurative rather than physical motion of entities. The relative lower frequencies of the occurrence of *atā* and *ǧā'a* with SUBJECT NUMBER x PERSON x GENDER x SEMANTIC CATEGORY as opposed to the higher frequencies observed for *haḍara* and *qadima* are indicative of the wider range of constructions *atā* and *ǧā'a* appear in. *Haḍara* and *qadima*, on the other hand, seem to be more restricted to HUMAN and GROUP subjects.

VERB x SUBJECT SEMANTIC CATEGORY x PHRASAL SEMANTIC CATEGORY

The semantic categories of the sentential subject can also be interpreted within a larger conceptual event frame of a COME event. As mentioned earlier, verb uses were coded for the inclusion of semantic characteristics of additional (non-subject) phrasal constituents that specify SOURCE or GOAL of the motion event, MANNER and PATH of motion, larger context where the motion event takes place (SETTING), as well as phrases signaling the involvement of other participants in the motion event (COMITATIVE), a PURPOSE of the motion event, in addition to time (TEMPORAL) and frequency (DEGREE)

specifying complements. For the following analysis, I had to exclude DEGREE from the set of variables included in HCFA since the overall frequency of this variable in the 2,000-line data frame did not exceed 20.

I decided to combine the following HCFA analysis with the previous BP analysis, in terms of providing a further investigation of the pairs of verbs clustered together in Figure 1. In other words, the following analysis will be comparing the uses of *atā* and *ǧā'a* on the one hand and *ḥaḍara* and *qadima* on the other. For that purpose I had to split the data frame into two sets each containing 1,000 hits of two verbs. These data frames were then subjected to an HCFA test for the following set of variables: VERB x SUBJECT SEMANTIC CATEGORY x GOAL x SOURCE x MANNER x SETTING x PURPOSIVE x PATH x COMITATIVE x TEMPORAL.

Atā* and *ǧā'a

The verbs *atā* and *ǧā'a* seem to share a wide range of the larger conceptual frames hosting a COME event, as well as the semantic category of the sentential subjects they collocate with. It is therefore useful to run an HCFA test that involves only those two verbs in order to distinguish their different uses. Table 10 shows configurations considered to be *types* – with the observed frequency significantly higher than the expected frequency – which involve the two COME verbs and the other 9 variables.

TABLE 10. SUBJECT SEMANTIC CATEGORY x SEMANTIC PROPOSITIONS configurations for the verbs *atā* and *ǧā'a*.

VERB	SUBJ_ CAT	GOAL	SOUR- CE	MAN- NER	SET- TING	PATH	PURP- OSIVE	COMIT- ATIVE	TEMP- ORAL	FREQ	EXP	DEC	Q
<i>atā</i>	EVENT	NO	NO	NO	YES	NO	NO	NO	NO	20	2.4983	***	0.018
<i>atā</i>	NOTION	NO	YES	NO	NO	NO	NO	NO	NO	15	2.4979	***	0.013
<i>atā</i>	HUMAN	YES	NO	NO	NO	NO	NO	NO	NO	24	6.6215	**	0.017
<i>atā</i>	GROUP	YES	NO	NO	NO	NO	NO	NO	NO	11	1.6062	*	0.009
<i>atā</i>	ACTIVITY	NO	NO	NO	YES	NO	NO	NO	NO	23	7.8018	ms	0.015
<i>ǧā'a</i>	CONTENT	NO	NO	NO	YES	NO	NO	NO	NO	88	4.2516	***	0.084
<i>ǧā'a</i>	ACTIVITY	NO	NO	YES	NO	NO	NO	NO	NO	30	5.9623	***	0.024
<i>ǧā'a</i>	DEM	NO	NO	NO	YES	NO	NO	NO	NO	22	2.1039	***	0.02
<i>ǧā'a</i>	ACTIVITY	NO	NO	NO	NO	NO	NO	NO	YES	22	6.0478	**	0.016
<i>ǧā'a</i>	DEM	NO	NO	NO	NO	NO	NO	NO	YES	11	1.6309	*	0.009

We can see that the verb *atā* takes on a relatively wider range of sentential subjects than *ǧā'a* appears to do. *Atā* can appear with subjects denoting EVENT, NOTION, HUMAN, GROUP, and ACTIVITY, while *ǧā'a* seems to have more specialized uses (at least as far as newspaper writing is concerned). In addition to ACTIVITY-related collocates, and as I mentioned above, *ǧā'a* features mostly as part of expressions that involve the use of a demonstrative (DEM) as a subject or would collocate with what I coded as 'content' of a documentation, speech, etc.

In terms of the overall event frames associated with the two COME verbs, we find that an ACTIVITY collocating with *atā* most of the time comes in a certain setting/context, as in (1), while an ACTIVITY collocating with *ǧā'a* comes mostly in a certain manner, as in (2), and at a certain time frame, as in (3).

(1) وقال المحافظ أن المشروع يأتي ضمن عدة قرارات لتحقيق سيولة مرورية

<i>wa=qāla</i>	<i>al=muḥafez</i>	<i>anna</i>	<i>al=mašru'</i>	<i>ya'ti</i>	<i>ḍimna</i>	<i>'iddat</i>
CONJ=say.PERF.3SG.M	ART=mayor	TOP	ART=project	<i>atā</i> .IMPF.3SG.M	ADV	several
and said	the mayor	that	the project	came	within	several

<i>qararāt</i>	<i>li=taḥqiq</i>	<i>suyula</i>	<i>mururiyya</i>
decisions	PURP=acheive	fluidity	traffic.ADJ
decisions	to achieve	flow	traffic

'And the mayor said that the project comes as part of a number of decisions taken to achieve traffic flow in the city of Giza and its streets'

(2) وجاء الفوز الهلالي مستحقا بعد ان قدم مباراة اذهلت المتابعين

<i>wa=ǧā'a</i>	<i>al=fawz</i>	<i>al=hilali</i>	<i>mustahaqqan</i>	<i>ba'da</i>	<i>an</i>
CONJ= <i>ǧā'a</i> .PERF.3SG.M	ART=succcess	CONJ=Hilal.ADV	well earned	LOC	TOP
and came	the success	the Hilali	well earned	after	that

<i>qaddama</i>	<i>mubāratan</i>	<i>'aḍhalat</i>	<i>al=mutabe'in</i>
present.PERF.3SG.M	game	astonish.PERF.3SG.F	ART=followers
presented	game	astonished	the followers

'And the Hilal's [soccer team] victory came well-earned after having played a match that astonished the audience'

(3) وجاء هذا التوضيح بعد مقال نشرته أول من امس جريدة «الجمهورية»

<i>wa=ǧā'a</i>	<i>haḍa</i>	<i>al=tawḍiḥ</i>	<i>ba'da</i>	<i>maqal</i>	<i>našarat-hu</i>
CONJ= <i>ǧā'a</i> .PERF.3SG.M	DEM	ART=clarification	LOC	article	publish.PERF.3SG.F-CL.3SG.M.ACC
and came	DEM	the clarification	after	article	published it

'awwal min 'ams ġaridat al ġumhuriyya
 first ABL yesterday newspaper Al Jumhuriyya
 first from yesterday newspaper Al Jumhuriyya
 'And this clarification came after an article that the Jumyoriya newspaper
 published the day before yesterday'

When the verb *ġā'a* collocates with a demonstrative (as the sentential subject),
 most of the time this construction is accompanied by clauses that specify the context (4)
 or the time frame (5) in which the figurative motion event takes place.

(4) جاء هذا في رسالة وجهها ميچور الى اعضاء الحملة

ġā'a *haḍa fi risala waġġaha-ha* *meġor*
ġā'a.PERF.3SG.M DEM LOC letter direct.PERF.3SG.M-CL.3SG.F.ACC Major
 came this in letter directed it Major

ila 'a'ḍā' al=ḥamla
 ALL members ART=campaign
 to members the campaign
 'This came in a letter that Major addressed to campaign members'

(5) وجاء ذلك في وقت استمرت المعارك في شمال كابل

wa=ġā'a *dālika fi waqt istamarrat* *al=ma'arek*
 CONJ=*ġā'a*.PERF.3SG.M DEM LOC time continue.PERF.3SG.F ART=battles
 and came that at time continued the battles

fi šamāl kābul
 LOC north Kabul
 in north Kabul
 'And that came at a time when the battles continued in northern Kabul'

We can see, however, in Table 10 that the larger bulk of *ġā'a* uses is where the verb
 collocates with subjects that denote some 'content' of a document or a speech, which also
 "comes" or appears in a certain setting (6).

(6) فقد قرأت بعناية ما جاء في مقالكم تحت عنوان رسالة الى شيخ الأزهر

faqad qara'tu bi=inaye mā ġā'a *fi maqāli-kum*
 DM read.PERF.1SG INST=care RP *ġā'a*.PERF.3SG.M LOC article-CL.2PL.GEN
 already I read with care what came in your article

taḥta 'inwāan risala ila šayḥ al=azhar
 LOC title letter ALL sheikh ART=Azhar
 under title letter to sheikh the Azhar
 'I have carefully read what appeared in your article under the title "a letter to the
 sheikh of Azhar"'

Atā on the other hand associates with a more diverse set of event frames. In addition to the highly frequent constructions involving ACTIVITY X SETTING, it is mostly used to talk about an EVENT in a certain setting (7).

(7) مشيراً إلى أن الزيارة تأتي ضمن الجهود الفرنسية لدعم مكافحة الإرهاب

<i>muširan</i>	<i>ila</i>	<i>anna</i>	<i>al=ziyara</i>	<i>ta'ti</i>	<i>ḡimna</i>
point.out.AP.3SG.M	ALL	TOP	ART=visit	atā .IMPF.3SG.F	ADV
pointing out	to	that	the visit	comes	among

<i>al=ḡuhud</i>	<i>al=faransiyya</i>	<i>li=mukāfaḡat</i>	<i>al='irhāb</i>
ART=efforts	ART=French	PURP=fighting	ART=terrorism
the efforts	the French	to fighting	the terrorism

‘pointing out that the visit comes as part of the French efforts to support fighting terrorism’

We can also see from Table 10 that *atā* is more likely than *ḡā'a* to collocate with subjects denoting HUMAN agents. Most typically, HUMANS come to a destination, as in (8).

Similarly, GROUPS, such as organizations and institutions (which is a metonymic representation of a humans), also come to a (less physical) destination (9).

(8) وكنت أحضر الحفلات التي كان يقيمها، كما كان، من جهته، يأتي إلى حفلاتي

<i>wa=kuntu</i>	<i>aḡḡuru</i>	<i>al=ḡafalāt-a</i>	<i>allati</i>	<i>kāna</i>
CONJ=AUX	ḡaḡara .IMPF.1SG	ART=parties-ACC	RP	be.PERF.3SG.M
and I was	attend	the parties	that	he was

<i>yuḡimu-ha</i>	<i>kama</i>	<i>kāna</i>	<i>min</i>	<i>ḡihati-h</i>
make.stand.IMPF.3SG.M-CL.3SG.F.ACC	CONJ	be.PERF.3SG.M	ABL	side-CL.3SG.M.GEN
make them stand	also	he was	from	his side

<i>ya'ti</i>	<i>ila</i>	<i>ḡafalāt-i</i>
atā .IMPF.3SG.M	ALL	parties-CL.1SG.GEN
comes	to	my parties

‘And I used to attend the parties that he threw, just as he, on his part, used to come to my parties’

(9) وتأتي الولايات المتحدة في مقدم الدول الأجنبية غير العربية استثماراً في السعودية

<i>wa=ta'ti</i>	<i>al=wilayāt</i>	<i>al=muttaḡida</i>	<i>fi</i>	<i>muḡaddam</i>	<i>al=duwal</i>
CONJ= atā .IMPF.3SG.F	ART=States	ART=United	LOC	forefront	ART=countries
and comes	the States	The United	in	forefront	the countries

al=’aġnabiyya *ġayr* *al=’arabiyya* *’istiṭmaran* *fi* *al=sa’udiyya*
 ART=foreign NEG ART=arab investing-ADV LOC ART=Saudi
 the foreign non- the Arab investing-wise in the Saudi
 ‘And the United States comes at the forefront of foreign, non-Arab countries that
 invest in Saudi Arabia’

On a more abstract level, a NOTION collocating with *atā* most of the time comes from a certain SOURCE, as in (10).

(10) ولعل الأمل باحياء روسيا يأتي من قدرتها على الغفران

wa=la’alla *al=’amal* *bi=’ihyā’* *rusya* *ya’ti* *min*
 CONJ=MOD ART=hope INST=revive.VN Russia *atā*.IMPF.3SG.M ABL
 and maybe the hope with reviving Russia comes from

qudrati-ha *’ala* *al=ġufrān*
 ability-CL.3SG.F LOC ART=forgiveness
 its ability on the forgiveness
 ‘The hope to revive Russia might come from its ability to forgive’

Ḥaḍara and qadima

An HCFA test was run again for the verbs *ḥaḍara* and *qadima*, with a similar set of variables explored for *atā* and *ġā’a*. For this test I decided to exclude PATH from the previous set of variables since path of motion was not coded for in the usage of either *ḥaḍara* or *qadima* and therefore adding this variable would only be a burden on the process of running the script. Table 11 shows the most significant *type* configurations found for those two verbs.

TABLE 11. SUBJECT SEMANTIC CATEGORY X SEMANTIC PROPOSITIONS configurations for the verbs *ḥaḍara* and *qadima*.

VERB	SUBJ_ CAT	GOAL	SOU- RCE	MAN- NER	SET- TING	PURP- OSIVE	COMIT- ATIVE	TEMP- ORAL	Freq	Exp	Dec	Q
<i>ḥaḍara</i>	HUMAN	YES	NO	NO	NO	NO	NO	NO	321	142.6842	***	0.208
<i>qadima</i>	HUMAN	NO	YES	NO	NO	NO	NO	NO	136	27.1637	***	0.112
<i>qadima</i>	HUMAN	NO	NO	NO	NO	YES	NO	NO	41	15.2935	***	0.026
<i>qadima</i>	HUMAN	NO	YES	NO	NO	YES	NO	NO	30	5.2899	***	0.025
<i>qadima</i>	GROUP	NO	YES	NO	NO	NO	NO	NO	12	2.0802	**	0.01

Table 11 shows that the most predominant use of *ḥaḍara* involves a HUMAN agent arriving at a GOAL. That particular construction accounts for more than 60% of the

annotated uses of this verb. As I will explain in the next chapter, an EVENT is by far the most frequent type of destination of the motion event. Sentence (11) below is an example of a typical usage of the verb *ḥaḍara*, where we have the verb appearing in a past simple perfective transitive construction, and a HUMAN agent arriving or being present at a destination (EVENT).

(11) وحضر الاجتماع ايضا السفير الاميركي في دمشق كريستوفر روس

<i>wa=ḥaḍara</i>	<i>al=iḡtima'-a</i>	<i>ayḍan</i>	<i>al=safir-u</i>
CONJ= <i>ḥaḍara</i> .PERF.3SG.M	ART=meeting-ACC	also	ART=ambassador-NOM
and attended	the meeting	also	the ambassador

<i>al=amriki</i>	<i>fi</i>	<i>dimašq</i>	<i>Christopher Ross</i>
ART=American	LOC	Damascus	Christopher Ross
the American	in	Damascus	Christopher Ross

‘And the American ambassador in Damascus, Christopher Ross, also attended the meeting’

Similar to *ḥaḍara*, the verb *qadima* also collocates mostly with HUMAN agents (and to a lesser degree subjects denoting GROUPS, i.e. organizations and institutions). As far as the dominant conceptual frames hosting the COME event in *qadima* constructions, we can see that a lot of the configurations reported in Table 11 include specifying the SOURCE as well as the PURPOSE of the motion event. The most frequent and robust configuration we find in Table 11 involves HUMANS coming from a certain SOURCE, as in (12), and to a much lesser degree GROUPS coming from a SOURCE as well. What also appears to highlight the use of the verb *qadima* is that the motion of HUMAN agents, in many cases, involves a PURPOSE for coming (to or from a certain location), as we can see (13).

(12) الأمر الذي لم يحدث اللا في دولة قطر التي قدم منها

<i>al='amr</i>	<i>allaḍi</i>	<i>lam</i>	<i>yaḥḍuṭ</i>	<i>illa</i>	<i>fi</i>	<i>dawlat</i>	<i>qaṭar</i>
ART=matter	RP	NEG	happen.JUSS.3SG.M	ADV	LOC	country	Qatar
the matter	which	did not	happen	except	in	country	Qatar

<i>allati</i>	<i>qadima</i>	<i>min=ha</i>
RP	<i>qadima</i> .PERF.3SG.M	ABL=CL.3SG.F
which	came	from it

‘which did not happen except for in Qatar, the country where he came from’

(13) استقطبت عددا كبيرا من الزوار قدموا خصيصا من مختلف أنحاء العالم لشراء الذهب

<i>qadimu</i>	<i>hiṣṣiṣan</i>	<i>min</i>	<i>muxtalaf</i>	‘ <i>anḥā</i> ’
<i>qadima</i> .PERF.3PL.M	ADV	ABL	various	parts
came	especially	from	various	parts

<i>al=‘ālam</i>	<i>li=širā’</i>	<i>al=ḡahab</i>
ART=world	PURP=buy.VN	ART=gold
the world	to buying	the gold

‘they came from different parts of the world especially to buy gold’

What this table does not show is that *qadima* also associates with arrival at a destination (GOAL) to a large extent, but since the aim of the HCFA analysis here is to highlight the constructional and conceptual characteristics that tease the two verbs apart, it appears that *ḥaḍara* appears in more GOAL constructions than *qadima* does. I will be elaborating on this point in Chapter 6.

5.4 Polytomous logistic regression analysis

As discussed earlier, the polytomous logistic regression analysis (Arppe, 2008) applies more advanced algorithms in order to determine the relative effects of multiple predictor variables (the constructional features) on the choice of outcome variables (COME verbs). We saw in Chapter 3 that the regression analysis also calculated probability estimates for the occurrence of each GO verb per each annotated context. In order to arrive at a reasonable model that can more or less predict the occurrence of a verb in a context of use, we need to select a set of predictor variables. The process of selection, as I mentioned previously, is based on (i) monivariate analysis, as represented in the standardized Pearson’s residuals, and already discussed in §5.2.1 (Table 4); and (ii) a bivariate analysis that examines the level of association between pairs of variables at a time. In §5.2.4.2 I will present the results obtained from running the polytomous logistic regression analysis, as well as the probabilities of occurrence calculated for each COME

verb for a number of original sentences from the data frame. Some of these sentences may be treated as exemplary cases of verb usage, while others may prove to be contexts of use in which two or more COME verbs can be used interchangeably.

Recall that, for polytomous logistic regression analysis, it is better to convert the nominal form of the data frame to a logical, whereby every level of each variables is turned into a variable in its own right, with the binary values TRUE/FALSE. For instance, the binary levels YES/NO for the variable GOAL, are converted into two variables: GOAL.YES and GOAL.NO.

5.4.1 Bivariate analysis

The current bivariate analysis is a fundamental step towards the selection of variables for the logistic regression model and it aims at highlighting pairs of logical variables with high association values, i.e. variables that seem to co-occur very frequently. The subsequent logistic regression model should, therefore, avoid the inclusion of such pairs of variables as a means of avoiding excessive collinearity. As I mentioned previously, the {polytomous} statistical package, developed for R by Arppe (2012), contains the function `associations()` that enables us to examine the degree of association between all logical variables. Table 12 reports the pairs of variables that were deemed to have high levels of association.

TABLE 12. Bivariate analysis of selected (independent) variables.

VARIABLE 1	VARIABLE	UC 1 2	UC 2 1	N.1	N.2	N.COMMON
TENSE.IRR	ASPECT.NON-FIN	0.989730257	0.985940813	179	180	179
TENSE.PAST	MORPH_ASP.MOOD.PERF	0.803724735	0.804275523	1396	1397	1370
TENSE.PRES	MORPH_ASP.MOOD.IMPF	0.708608854	0.611368742	372	490	371
SUBJ_PER.1ST	SUBJ_GEN.NIL	0.934035147	0.877138025	59	64	59

In Table 12, we find the number of times each of the two variables was observed in the entire COME data frame (N.1 and N.2), and the number of times both variables co-occur in the same context of use (N.COMMON). As discussed before, the UC values

(Theil's Uncertainty Coefficient) reported here can be interpreted as the following: UC1|2 is the uncertainty coefficient of VARIABLE1 given VARIABLE2. That is to say, our uncertainty about the occurrence of VARIABLE1 is decreased x% of the time given that VARIABLE2 is also present in the context. For instance, knowing that the verb is inflected in the PERFECTIVE decreases our uncertainty that the tense is PAST 80.37% (UC1|2). The reverse is true 80.43% of the time. The pairs of variables listed in this table are very similar to what we noticed with the results obtained in the bivariate analysis of GO: TENSE.IRR highly co-occurs with ASPECT.NON-FIN; TENSE.PRESENT and MORPH_ASP.MOOD.IMPF; and SUBJ_PER.1ST and SUBJ_GEN.NIL. In the selection of predictors for the following polytomous logistic regression analysis, I therefore decided to eliminate one member of each pair, instead of including both.

5.4.2 Multivariate analysis

In addition to the bivariate analysis above, the frequency of occurrence of a variable was also an added measure in the process of selecting of variables for the regression model. That is to say, the set of predictor variables have to have an overall frequency of 20, with at least 10 occurrences for two verbs; thus allowing for 2 verbs to not have any occurrence of a certain variable at all.³⁷ The resulting set variables consisted of 31 independent predictor variables to be included in the polytomous logistic regression model. I chose *one-vs-rest* heuristic for implementing the polytomous logistic regression model (explained in Arppe, 2008), and fitted a model using the `polytomous()` function available in the `{polytomous}` package (Arppe, 2012), for the four COME verbs. The complete summary output returned by the model is shown in (14).

³⁷ Note that this mathematical criterion for selection of predictor variables is slightly different from that opted for earlier for the logistic regression model concerning GO verbs.

- (14) Summary for results from running the `polytomous()` function in R on a selection of 31 predictors. The Odds section lists the estimated odds for explanatory variables in favor of or against the occurrence of each verb, while (typically) the non-significant odds ($P < 0.05$) are shown in parentheses.

Formula:

```
VERB ~ ADVERBIAL.YES + ASPECT.HAB + ASPECT.SIMPLE + COMITATIVE.YES +
  GOAL.YES + LOC_ADV.YES + MANNER.YES + MORPH_ASP.MOOD.SUBJN +
  NEGATION.YES + PP.YES + PURPOSIVE.YES + SETTING.YES + SOURCE.YES +
  SUBJ_CAT.ACTIVITY + SUBJ_CAT.GROUP + SUBJ_CAT.COMMUNICATION +
  SUBJ_CAT.demonstrative + SUBJ_CAT.EVENT + SUBJ_CAT.HUMAN +
  SUBJ_CAT.STATE + SUBJ_CAT.TIME + SUBJ_GEN.FEM + SUBJ_NUM.PL +
  SUBJ_PER.1ST + SUBJ_PER.3RD + TEMPORAL.YES + TENSE.FUT +
  TENSE.PAST + TENSE.PRES + TRANSITIVITY.YES
```

Heuristic:

one.vs.rest

Odds:

	atā	ḥaḍara	ǧā'a	qadima
(Intercept)	14.96	0.003285	(0.8435)	0.001214
ADVERBIAL.YES	(1.573)	(0.6025)	(2.018)	(0.817)
ASPECT.HAB	(0.7544)	(0.4431)	(1.76)	(0.2339)
ASPECT.SIMPLE	0.1608	(1.073)	18.26	(0.8508)
COMITATIVE.YES	2.314	(1.277)	(0.9449)	0.3218
GOAL.YES	0.486	(1.461)	0.3431	4.094
LOC_ADV.YES	2.912	(0.4666)	(0.8031)	(2.163)
MANNER.YES	(0.4991)	(1.931)	(0.5484)	(1.648)
MORPH_ASP.MOOD.SUBJN	2.262	(1.917)	0.06885	0.07908
NEGATION.YES	2.387	2.83	(1/Inf)	0.04173
PP.YES	3.717	0.145	(0.9599)	3.559
PURPOSIVE.YES	(0.5119)	(1.115)	(0.9116)	(1.555)
SETTING.YES	0.1232	4.741	3.699	0.2415
SOURCE.YES	(0.717)	(0.469)	0.1812	9.262
SUBJ_CAT.ACTIVITY	(0.7845)	(1/Inf)	(1.417)	(1.423)
SUBJ_CAT.COMMUNICATION	(1.228)	(0.274)	(1.947)	(1/Inf)
SUBJ_CAT.demonstrative	(0.7276)	(1/Inf)	(3.262)	(1/Inf)
SUBJ_CAT.EVENT	(1.242)	(1/Inf)	(1.225)	(1.497)
SUBJ_CAT.GROUP	0.2149	25.9	0.1341	9.171
SUBJ_CAT.HUMAN	0.1742	20.8	0.07039	20.84
SUBJ_CAT.STATE	(2.273)	(1/Inf)	(0.8897)	(1/Inf)
SUBJ_CAT.TIME	(1.242)	(1/Inf)	(0.7262)	(4.991)
SUBJ_GEN.FEM	1.845	(0.5457)	(0.6684)	(1.487)
SUBJ_NUM.PL	0.3999	0.3929	0.274	6.214
SUBJ_PER.1ST	(0.6644)	9.778	(2.016)	(0.5814)
SUBJ_PER.3RD	0.1176	11.23	(1.115)	(2.065)
TEMPORAL.YES	0.4369	(1.212)	(1.004)	(1.14)
TENSE.FUT	11.99	(3.354)	(1/Inf)	(0.1307)
TENSE.PAST	(0.3052)	(0.8872)	(0.5049)	(2.717)
TENSE.PRES	32.55	(1.692)	0.0006418	0.105
TRANSITIVITY.YES	0.1903	18.84	0.1852	0.03933

Null deviance: 5545 on 8000 degrees of freedom
 Residual (model) deviance: 1914 on 7876 degrees of freedom

R2.likelihood: 0.6548
 AIC: 2162
 BIC: 2856

The first value to examine here is the value R^2 likelihood (R_L^2) which indicates the fitness of the model to the actual occurrences of COME verbs in the annotated data frame. The value of 0.65 is quite high for a polytomous logistic regression model, as was the case with the R_L^2 resulting from modelling the GO verbs. Again this may be due to the large number of independent predictor variables, and hence there is always a certain level of association between different combinations of variables. However, this high value may also be an indication of the fact that this large number of variables was indeed able to describe verbal behavior with a high degree of confidence. The *accuracy* rate calculated for this model was found to be 0.845. As explained previously, the *accuracy* measure (Menard, 1995: 28-30; Arppe, 2008: 129-132) corresponds to the number of times the model assigned the highest probability estimate to the actually observed verb in a given annotated context. This figure is also the aggregate value of all the individual *accuracy* rates calculated per COME verb: *atā* 0.82, *ḥaḍara* 0.83, *ḡā'a* 0.83, and *qadima* 0.90. Table 13 lists the number of instances in which the model accurately predicted the verb observed in context. More specifically, the numbers in bold indicate the number of times the model assigned the highest probability estimate to the observed verb.

TABLE 13. A cross-tabulation of the predicted vs. observed verbs for all annotated contexts in the data frame (with the total of 500 contexts per COME verb).

OBSERVED \ PREDICTED	PREDICTED			
	<i>atā</i>	<i>ḥaḍara</i>	<i>ḡā'a</i>	<i>qadima</i>
<i>atā</i>	409	27	37	27
<i>ḥaḍara</i>	26	415	16	43
<i>ḡā'a</i>	10	24	414	52
<i>qadima</i>	9	10	29	452

As we saw with GO verb previously, this model was also successful at predicting the verb that actually appears in context. Most of all, the model seemed to be successful at identifying the contexts in which the verb *qadima* appears (452/500 annotated sentences). The remaining values listed in this table correspond to the number of times

the model “mis-predicted” the verb in context. For instance, 52 out of the 500 *ğā’a* hits were predicted to be *qadima* hits, and so forth. As mentioned earlier, these are not uninteresting cases and they merit some attention, since they might signal some contexts in which two or more verbs could be used interchangeably. The fact that these accuracy rates are rather high may be indicative of either that (i) the model was successful in teasing apart the different uses of COME verbs, or that (ii) these verbs, as used in MSA newspaper writing, are associated with very specialized constructions with little window of overlap among the four verbs. I will come back to this point in Chapter 7.

The summary of the model reported in (14) lists the odds estimated by the model for each independent predictor variable that was included in the logistic regression model. Recall that these odds indicate whether the presence of a (logical) variable increases the chances of the occurrence of a particular verb in a given context (estimated odds > 1.0); or, on the contrary, whether the presence of a variable decreases the chances that a verb would occur in a given context (estimated odds < 1.0). A large number of the estimated odds are listed in parentheses, which indicates that the model was not confident in calculating the given values, possibly due to some levels of association between these variables and other. These odds are therefore unreliable and can be deemed insignificant.

As mentioned in Chapter 3, these odds can be interpreted either verb-wise or feature-wise. That is to say, if we take each verb on its own, we can identify the contextual feature that seems to highly favor the occurrence of that verb. For instance, as far as the verb *atā* is concerned, PRESENT tense seems to be the contextual feature with the highest value (32.55); for *ğā’a* it is SIMPLE aspect; for *ḥaḍara* it is the subject semantic category of GROUP; and for *qadima* it is the subject semantic category of HUMAN. On the other hand, examining one individual contextual feature across the four verbs at a time, we can see, for instance, that while SIMPLE aspect increases the chances of the occurrence of *ğā’a*, it instead seems to decrease the chances of the occurrence of *atā*, and has a neutral

(i.e. non-significant) impact on *ḥaḍara* and *qadima*. Table 14 summarized the contextual features that were considered to be in favor of or against the occurrence of each COME verb.

TABLE 14. Contextual features that increase the odds in favor of or against the occurrence of COME verbs.

VERB	ODDS IN FAVOR OF THE VERB		ODDS AGAINST THE VERB	
<i>atā</i>	TENSE.PRES	32.55	SUBJ_PER.3 RD	0.1176
	TENSE.FUT	11.99	SETTING.YES	0.1232
	PP.YES	3.717	ASPECT.SIMPLE	0.1608
	LOC_ADV.YES	2.912	SUBJ_CAT.HUMAN	0.1742
	NEGATION.YES	2.387	TRANSITIVITY.YES	0.1903
	MORPH_ASP.MOOD.SUBJN	2.262	SUBJ_CAT.GROUP	0.2149
	COMITATIVE.YES	2.314	SUBJ_NUM.PL	0.3999
	SUBJ_GEN.FEM	1.845	TEMPORAL.YES	0.4369
			GOAL.YES	0.486
<i>ḥaḍara</i>	SUBJ_CAT.GROUP	25.9	PP.YES	0.145
	SUBJ_CAT.HUMAN	20.8	SUBJ_NUM.PL	0.3929
	TRANSITIVITY.YES	18.84		
	SUBJ_PER.3 RD	11.23		
	SUBJ_PER.1 ST	9.778		
	SETTING.YES	4.741		
	NEGATION.YES	2.83		
<i>ḡā'a</i>	ASPECT.SIMPLE	18.26	TENSE.PRES	0.00064
	SETTING.YES	3.699	MORPH_ASP.MOOD.SUBJN	0.06885
			SUBJ_CAT.HUMAN	0.07039
			SUBJ_CAT.GROUP	0.1341
			SOURCE.YES	0.1812
			TRANSITIVITY.YES	0.1852
			SUBJ_NUM.PL	0.274
			GOAL.YES	0.3431
<i>qadima</i>	SUBJ_CAT.HUMAN	20.84	TRANSITIVITY.YES	0.03933
	SOURCE.YES	9.262	NEGATION.YES	0.04173
	SUBJ_CAT.GROUP	9.171	MORPH_ASP.MOOD.SUBJN	0.07908
	SUBJ_NUM.PL	6.214	TENSE.PRES	0.105
	GOAL.YES	4.094	SETTING.YES	0.2415
	PP.YES	3.559	COMITATIVE.YES	0.3218

Generally speaking, the findings reported in Table 14 are congruent with findings discussed in the previous statistical analyses. For instance, the odds listed in the table confirm the fact that PRESENT tense and NEGATIVE polarity highly associates with *atā*, in addition to LOCATIVE ADVERB phrases, COMITATIVE phrases, among others. We also saw in the previous analyses that *qadima* appears in the PLURAL more frequently than any other COME verb, and that it associates almost exclusively with HUMAN or GROUP agents.

We can also turn our attention to the variables that decrease the chances of the occurrence of each verb in a given context. For example, the presence of a PREPOSITIONAL PHRASE as well as PLURAL number inflection on the verb seem to highly disassociate with *ḥaḍara*; while PRESENT tense very strongly disfavors the presence of *ġā'a* in a context of use (0.00064).

Probability estimates

We can now examine the probability estimates that the polytomous logistic regression analysis assigns to each of the COME verbs per annotated context (4 verbs X 2,000 sentences). As I discussed earlier, the calculation of these probabilities is based on the collective effect of the estimated odds – listed in (14) – for each verb, per annotated sentence. As we saw with GO verbs, we can expect the estimated probabilities to range from very high values (approaching 1.00) to very low values (approaching 0.00) and any values in between, depending on the set of predictors present in a particular context of use. The following, (15) – (18), are sentences extracted from the original data frame for which the verb in its context received an almost categorical probability estimate.

(15) Sentence #435

<p><i>atā</i> ≈ 1 (observed) <i>ḥaḍara</i> = 0 <i>ġā'a</i> = 0 <i>qadima</i> = 0</p>	<p>contextual features used (in the model): SUBJ_PER.3RD + SUBJ_CAT.STATE + NEGATION.YES + PP.YES + SOURCE.YES</p>
--	---

الموقف الفرنسي التقليدي المناصر لصدام لم يأت من منطق الحب له
al=mawqif *al=faransi* *al=taqlidi* *al=munāṣir* *li=ṣaddām*
ART=position ART=French ART=traditional ART=supporting ALL=Saddam
the position the French the traditional the supporting to Saddam

lam *ya'ti* *min* *manṭiq* *al=ḥub* *la=h*
NEG ***atā***.JUSS.3SG.M ABL logic ART=love ALL=CL.3SG.M
did not come from logic the love to him

‘The traditional French position supporting Saddam is not out of love for him’

(16) Sentence #1288

<i>atā</i> = 0.020 <i>ḥaḍara</i> = 0.973 (observed) <i>ǧā'a</i> = 0.000 <i>qadima</i> = 0.006	contextual features used (in the model): TENSE.FUT + ASPECT.SIMPLE + TRANSITIVITY.YES + SUBJ_NUM.PL + SUBJ_PER.3 RD + SUBJ_CAT.HUMAN + GOAL.YES
--	--

الذي سيقدم كبري المفاجآت يوميا للجمهور الذين سيحضر المباريات
allaḍi sa=yuqaddim kubrā al=mufaǧǧa'āt yawmiyyan li=l=ǧamāhīr
RP FUT=present.IMPF.3SG.M biggest ART=surprises ADV ALL=ART=audiences
who will present biggest the surprises daily to the audiences

allaḍīna sa=yaḥḍuruna al=mubārayāt-i
RP FUT=*ḥaḍara*.IMPF.3PL.M ART=games-ACC
who will attend the games

'Who is presenting big surprises daily to the audiences who are attending the games'

(17) Sentence #694

<i>atā</i> = 0.022 <i>ḥaḍara</i> = 0.000 <i>ǧā'a</i> = 0.978 (observed) <i>qadima</i> = 0.000	contextual features used (in the model): TENSE.PAST + ASPECT.SIMPLE + SUBJ_PER.3 RD + SUBJ_CAT.DEM + LOC_ADV.YES + SETTING.YES
--	--

جاء ذلك خلال تصريحات أدلى بها الوزير خورشيد
ǧā'a dālika ḥilāl taṣrīḥāt adlā
ǧā'a.PERF.3SG.M DEM ADV statements declare.PERF.3SG.M
came that during statements declared

bi=hā al=wazīr ḥūršid
INST=CL.3SG.F ART=minister Khurshid
by it the minister Khurshid

'This came during statements that the minister Khurshid made'

(18) Sentence #1736

<i>atā</i> = 0.010 <i>ḥaḍara</i> = 0.015 <i>ǧā'a</i> = 0.008 <i>qadima</i> = 0.967 (observed)	contextual features used (in the model): TENSE.PAST + ASPECT.SIMPLE + SUBJ_NUM.PL + SUBJ_PER.3 RD + SUBJ_CAT.HUMAN + PP.YES + LOC_ADV.YES + GOAL.YES + SOURCE.YES + TEMPORAL.YES
--	--

اجدادنا الذين قدموا الى هذه الشعاب البركانية منذ ثلاثمائة سنة تقريبا من جبل لبنان وحلب وادلب وفلسطين
aǧḍādu-na allaḍīna qadimū ilā ḥāḍīhi al=šī'āb
grandfathers-CL.1PL.GEN RP *qadima*.PERF.3PL.M ALL DEM ART=regions
our grandfathers who came to this the regions

al=burkāniyya munḍu ṭalātmā'at sana taqrīban min ḡabal lubnān
ART=volcanic ADV three hundred year ADV ABL mount Lebanon
the volcanic since three hundred year almost from Mount Lebanon

wa=ḥalab wa=idlib wa=falasṭīn
CONJ=Aleppo CONJ=Idlib CONJ=Palestine
and Aleppo and Idlib and Palestine

‘Our forefathers who came to these volcanic regions almost 300 years ago from Mount Lebanon, Aleppo, Idlib and Palestine’

These sentences can be considered as prototypical uses of each verb. It is also possible to examine the set of contextual features that each sentence was coded for and which were used as predictor variables in the logistic regression model. For instance, one can see that, in (15), *atā* received a categorical probability estimate. If we interpret the set of contextual variables responsible for assigning such high probability estimate to *atā* against the list of odds in favor or against the occurrence of each verb, we can theorize that this probability estimate is the result of the following: both NEGATION and PP count as odds in favor of *atā*, while none of the remaining variables (not deemed by the model as in favor of *atā*) are against the occurrence of this verb; on the other hand, NEGATION is a variable that is highly against the occurrence of *qadima*, and similarly PP is against the occurrence of *ḥaḍara*, while the presence of a SOURCE phrase is against the occurrence of *ḡā’a*. This may not be an adequate description of the exact statistical calculation resulting in these probability estimates, yet, this is just one way of showing that these results do make sense in light of the estimated odds.

Of course, not all predictions made by the model were accurate. Among the sentences for which a single verb received a very high probability estimate, I found a number of instances in which the predicted verb was not the observed verb. The logistic regression model, being blind to the actual verb observed in the 2000 contexts of use, relies fundamentally on the set of independent predictor variables in assigning probability estimates. It is therefore quite expected to come across ‘mis-predictions’, as in (19) and (20).

(19) Sentence #848

<i>atā</i> = 0.014 <i>ḥaḍara</i> = 0.015 <i>ǧā'a</i> = 0.006 (observed) <i>qadima</i> = 0.965 (predicted)	contextual features used (in the model): TENSE.PAST + ASPECT.SIMPLE + SUBJ_NUM.PL + SUBJ_PER.3 RD + SUBJ_GEN.FEM + SUBJ_CAT.HUMAN + PP.YES + GOAL.YES + SOURCE.YES
--	---

جئن للتكريم من مختلف المناطق
ǧi'na *li=l=takrām* *min* *muḥtalaf* *al=manāṭiq*
ǧā'a.PERF.3PL.F ALL=ART=honor.VN ABL different ART=places
they came to the honoring from different places
‘They came for the honoring ceremony from different places’

(20) Sentence #1742

<i>atā</i> = 0.022 <i>ḥaḍara</i> = 0.962 (predicted) <i>ǧā'a</i> = 0.005 <i>qadima</i> = 0.011 (observed)	contextual features used (in the model): SUBJ_PER.3 RD + SUBJ_CAT.HUMAN + ADVERBIAL.YES + GOAL.YES + MANNER.YES + TRANSITIVITY.YES
--	--

وكان علي بن عبد الله إذا قدم مكة حاجا أو معتمرا عطلت قريش مجالسها
wa=kāna *'ali bin 'abdillah* *iḍā* *qadima* *makka-ta* *ḥāǧǧan*
CONJ=be.PERF.3SG.M Ali Bin Abdullah COND *qadima*.PERF.3SG.M Mecca-ACC pilgrim
and was Ali Bin Abdullah if he came Mecca pilgrim

aw *mu'tamiran* *'aṭṭalat* *qurayš* *maǧālisaha*
CONJ pilgrim suspend.PERF.3SG.F Quraysh meetings-CL.3SG.F
or minor.pilgrim suspended Quraysh its meetings
‘When Ali bin Abdullah used to come to Mecca on a pilgrimage Quraysh would suspend its meetings’

In (19), the predicted verb was *qadima*. However, the actual verb that occurred in that sentence in the corpus was *ǧā'a*. If we examine the nine contextual features characterizing the sentence in (19), we see that five of them are considered among the contextual features that increase the odds in favor of *qadima*: SUBJ_CAT.HUMAN + SOURCE.YES + SUBJ_NUM.PL + GOAL.YES + PP.YES. In addition, 3 of these contextual features are considered among the variables that act against the occurrence of *ǧā'a*: SUBJ_NUM.PL, SUBJ_CAT.HUMAN, GOAL.YES. This sentence involves the use of *ǧā'a* to construe a physical motion event, which is not one of the most frequent uses of this verb. As for the example in (20), the particular usage of *qadima* in this sentence can be found in a specific genre, that of historical narrative. As I will explain in the qualitative analysis,

in Chapter 6, while *atā*, *ġā'a*, and *ḥaḍara* all appear in transitive constructions in MSA, *qadima* is used transitively to signal a shift in register. Since such pattern of use occurs less frequently than the general overall usage of *qadima*, the model assigns *ḥaḍara* instead as the most plausible verb choice for such context.

I pointed out earlier that another set of probability estimates worthy of scrutiny are the cases in which the estimated values are equi-probable among two, three, or all four COME verbs. It is highly likely that a number of such contexts would in fact allow two or three or all four COME verbs to be used interchangeably. Even though this proved not to be the case with GO verbs receiving equal probability estimates, the analysis conducted on COME verbs yielded more satisfactory results.

For sentences (21) - (24) COME verbs have been assigned less categorical probability values than what we saw in the previous examples.

(21) Sentence #182

<i>atā</i> = 0.076 <i>ḥaḍara</i> = 0.169 <i>ġā'a</i> = 0.322 <i>qadima</i> = 0.434	(observed)	contextual features used (in the model): TESNE.PAST + ASPECT.SIMPLE + SUBJ_PER.1 ST + SUBJ_CAT.HUMAN + PP.YES + GOAL.YES + TEMPORAL.YES
---	------------	--

فأنا لم أصل صلاة الجمعة ولا صلاة الجماعة منذ أتيت الى السجن في تشرين الاول 1995
fa='ana lam uṣalli ṣalāt al=ġum'a wa=lā ṣalāt
 CONJ=PP NEG pray.JUSS.1SG prayer ART=Friday CONJ=NEG prayer
 and I did not pray prayer the Friday and not prayer

al=ġamā'a munḍu ataytu ilā al=siġn fī tešrīn al awwal 1995
 ART=congregation ADV *atā*.PERF.1SG. ALL ART=prison LOC October 1995
 the congregation since I came to the prison in October 1995

'I have not prayed a Friday prayer or a congregation prayer since I was sent to jail in October 1995'

(22) Sentence #1183

<i>atā</i> = 0.199 <i>ḥaḍara</i> = 0.137 (observed) <i>ġā'a</i> = 0.247 <i>qadima</i> = 0.416	contextual features used (in the model): TESNE.PAST + ASPECT.SIMPLE + SUBJ_PER.3 RD + SUBJ_CAT.HUMAN + PP.YES + LOC.ADV.YES + MANNER.YES + COMITATIVE.YES
--	--

وقد حضر الأب علي الفور ومعه عددا من زملائه الأطباء
wa=qad haqara al='ab 'alā al=fawr
 CONJ=DM *haqara*.PERF.3SG.M ART=father LOC ART=immediately
 and already came the father on the immediately

wa=ma'a-hu 'adadun min zumalā'-ih al='aṭibbā'
 CONJ=COM-CL.3SG.M number ABL colleagues-CL.3SG.M.GEN ART=doctors
 and with him number of his colleagues the doctors
 'And the father came immediately with a number of his physician colleagues'

(23) Sentence #970

<i>atā = 0.231</i> <i>haqara = 0.132</i> <i>ḡā'a = 0.484</i> (observed) <i>qadima = 0.153</i>	contextual features used (in the model): TESNE.PAST + ASPECT.SIMPLE + SUBJ_PER.1 ST + SUBJ_CAT.HUMAN + PP.YES + ADVERBIAL.YES + GOAL.YES + TEMPORAL.YES + COMITATIVE.YES
--	---

جئت برفقة زوجتي صباح يوم الخميس 2006/9/7 الى مدينة حمص
ḡi'tu bi=rifqat zawḡat-i ṣabāḥ yawm al=ḥamīs
ḡā'a.PERF.1SG INST=company wife-CL.1SG.GEN morning day ART=Thursday
 I came with company my wife morning day the Thursday

7/9/2006 ilā madinat ḥimṣ
 7/9/2006 ALL city Hims
 7/9/2006 to city Hims

'I came accompanied by my wife to the city of Hams on the morning of Thursday 7/9/2006'

(24) Sentence #1844

<i>atā = 0.157</i> <i>haqara = 0.140</i> <i>ḡā'a = 0.314</i> <i>qadima = 0.390</i> (observed)	contextual features used (in the model): TESNE.PAST + ASPECT.SIMPLE + SUBJ_PER.1 ST + SUBJ_CAT.HUMAN + PP.YES + GOAL.YES
--	--

وبعد ذلك قدمت إلى المملكة العربية السعودية والتقيت ببعض الاخوة
wa=ba'da dālika qadimtu ilā al=mamlaka al='arabiyya al=su'udiyya
 CONJ=ADV DEM *qadima*.PERF.1SG. ALL ART=kingdom ART=Arab ART=Saudi
 and after that I came to the Kingdom the Arab the Saudi

wa=iltaqaytu bi=ba'd al='iḥwa
 CONJ=meet.PERF.1SG COM=some ART=brothers
 and I met with some the brother

'And after that I came to Saudi Arabia and I met with some brothers'

In these sentences the probability estimates for the four verbs (per context) mostly range between 0.1 and 0.5. The contextual features (used in the model) that characterize the use of verb in each context seem to be similar across the four sentences: (i) all four verb uses

are in the PAST SIMPLE, (ii) person inflection is either 1st or 3rd, (iii) the subject semantic category is HUMAN, and (4) all four sentences include a prepositional phrase that indicates mostly a GOAL of the motion event, or provides COMITATIVE or TEMPORAL information about the event.

Note that the verb which has received the highest probability estimate in these four sentences is not necessarily the observed verb. Statistically, all four verbs seem to have a shot at being the observed verb in these four sentences. Relying on my native speaker intuition, substituting the observed verb with the other COME verbs in these contexts does not raise any red flags, since all four of these usage contexts of use indicate physical motion of a HUMAN agent. Such results confirm the fact that these four COME verbs can be interchangeable in literal motion event frames.

This is another evidence of the usefulness and effectiveness of polytomous logistic regression analysis, at least as far as COME verb analysis is concerned. I pointed out earlier (in Chapter 3) that the probability estimates calculated for the GO data frame did not necessarily agree with my native speaker's intuition, and I attributed that to the set of variables that GO verbs were coded for in the data frame.³⁸ More specifically I suggested that the data frame should include more lexical or collocational variables. The COME data frame was coded for a comparable set of variables with the same level of details, yet the present logistic regression analysis of COME proved to be more successful at reflecting the distribution of the four verbs across a wide variety of contexts than did the analysis of GO verbs.

In conclusion, the preceding quantitative analysis of MSA COME verbs has shed light on the idiosyncratic behavior of each verb, as well as features shared among two or more verbs, which echos the findings obtained in Chapter 3 regarding GO verbs. I will

³⁸ Native speaker intuition as a validation of the logistic regression model analysis is one of the future (and current) research directions that I will discuss in Chapter 8.

now turn to a case-by-case examination of the four COME verbs, which aims to consolidate the findings reported in this chapter, as well as discuss additional common uses of each verb than have not been highlighted by the previous statistical analyses.

Chapter Six

Qualitative analysis of MSA COME verbs *atā*, *ḥaḍara*, *ǧā'a*, and *qadima*

The discussion on MSA GO verbs in Chapter 4 provided more evidence to support Wilkins and Hill's (1995) claim about the non-deictic nature of GO verbs cross-linguistically. We saw that what Sally Rice (p.c.) calls the "four Cs" (context, construal, construction, and conventions of uses) all contribute to our interpretation of GO verbs as either being deictic or simply representing the basic TRAVEL notion (Miller and Johnson-Laird, 1976). COME verbs, on the other hand, are inherently deictic in nature even though they differ in their base semantics cross-linguistically (Wilkins and Hill, 1995) and therefore motion *to* or *towards* the deictic centre is a central aspect of event construals involving a COME verb.

The deictic centre can either correspond to the speaker's location – what Fillmore (1966) referred to as *speaker deixis* – or it can be a specified endpoint, introduced in the utterance as an adpositional phrase describing a GOAL, that stands in a metonymic relationship with the speaker. For instance, in an utterance such as *Is John coming to the cinema?* (Gathercole, 1978), it is presupposed that the speaker would be present at the destination (*the cinema*), but is not necessarily present at the moment of speaking. According to Radden, the deictic sense here "results from an extension by metonymic transfer from person to place" or that it could be the result of "the speaker's or hearer's identification or involvement with the goal" (1996:430). Radden (ibid.) also indicates that the choice of a deictic motion verb in an utterance is conditioned by 'viewpoint'. In a narrative situation, such as *The King of Bahrain went to The Netherlands* the motion event is being viewed from outside of The Netherlands, while in the utterance *The King of Bahrain came to the Netherlands* the speaker is construing the event from a viewpoint inside The Netherlands. Radden refers to the notion of 'viewpoint' in such a situation as "[being] seen as a more general concept which covers both the 'deictic' viewpoint taken

by one of the speech act participants and the ‘mental’ viewpoint taken by the narrator” (1996:431). Lyons (1977:579) has referred to such a situation – in which the deictic centre corresponds to the protagonist in a narrative setting – as *deictic projection*. As far as the usage of MSA COME verbs is concerned, the data from the four verbs discussed in the following sections aims to shed light on the conceptualization of deictic motion in formal written Arabic.

The quantitative analysis in the previous chapter highlighted general patterns of each verb’s behavior which need to be discussed in more detail. The current analysis, by contrast, provides a case-by-case examination of the typical patterns of use of the four MSA COME verbs, in addition to identifying less frequent patterns of use that still constitute part of each verb’s behaviour. The present discussion also aims to situate the usage of MSA COME verbs on the map of cross-linguistic studies on basic motion verbs. Following the structure of the qualitative analysis of GO verbs, I address here both physical and figurative motion events of the four studied verbs, as well as examine metaphorically extended uses, idiomatic uses and collocates involving these verbs. COME verbs in MSA have not undergone any grammaticalization processes but as we will see in the subsequent sections, a large bulk of uses pertaining especially to *atā* and *ǧā’a* are highly idiomatized.

6.1 *Atā and ǧā’a*

I decided to group the two verbs *atā* and *ǧā’a* together in this qualitative, case-by-case analysis due to the fact that, lexically, the two verbs can be considered quite synonymous, since they are more likely to share the same lexico-syntactic frame than any other pair of verbs studied here. Indeed, the quantitative analysis has shown a great deal of overlap between the morphosyntactic and semantic properties of these two verbs. We saw, for instance, that the two verbs were grouped together according to the hierarchical

agglomerative cluster analysis. Nevertheless, the quantitative analysis has shown different distributional patterns for the two verbs with regard to subject collocates, morphosyntactic properties, as well as the larger conceptual event frames which attract *atā* and *ǧā'a*.

The following analysis further highlights the collocational, morphosyntactic, and lexico-semantic properties of the two verbs through the examination of actual verb uses, in addition to focusing on what constructional properties set the two verbs apart. However, the discussion presented in the following will occasionally refer to the quantitative findings obtained in the previous chapter.

In general, the quantitative analysis has shown that both verbs collocate with a wide range of sentential subjects belonging to different semantic categories. The majority of subject collocates denote non-human entities such as EVENT, ACTIVITY, NOTION, etc., which is an indication of the highly metaphorized usage of both *atā* and *ǧā'a*. To a lesser extent, subjects denoting HUMAN agents collocate with the two verbs in both physical and figurative motion event frames. In the following analysis, I will first examine the properties of physical motion event construals involving *atā* and *ǧā'a* followed by the figurative uses of these verbs.

6.1.1 Physical motion

Recall from the discussion in §1.4 that an event construal centred around the use of a COME verb inherently involves motion to or towards the deictic centre (Fillmore, 1966, 1972; among others). Arabic COME verbs are no exception. Uses of *atā* and *ǧā'a* include physical motion of a human agent towards the speaker, as in (1) and (2). The endpoint of the motion event can be introduced by a prepositional phrase, as in (3)-(6), where the deictic centre corresponds to the speaker's location, as in (3) and (4), or the speaker is metonymically – but not physically – present at the endpoint of the motion

event, as in (5) and (6). In terms of statistical distribution of this construction per verb, we saw in Table 10 in Chapter 5, *atā* is more likely than *ḡā'a* to express such physical motion of a HUMAN agent.

(1) قالت لي متى يأتي بابا

<i>qālat</i>	<i>l=ī</i>	<i>matā</i>	<i>ya'tī</i>	<i>bāba</i>
say.PERF.3SG.F	ALL=CL.1SG	Q	<i>atā</i> .IMPF.3SG.M	dad
she said	to me	when	comes	dad

'she said to me when will daddy come'

(2) الأربعاء جاءنا زبون الاثنين ورد البضاعة التي اشتراها

<i>al=arbi'ā'</i>	<i>ḡā'a-nā</i>	<i>zabūn</i>	<i>al=aṭṭayn</i>	<i>w=radda</i>
ART=Wednesday	<i>ḡā'a</i> .IMPF.3SG.M- CL.1PL.ACC	costumer	ART=Monday	CONJ= return.PERF.3SG.M
the Wednesday	came to us	the customer	the Monday	and he returned

<i>al=biḏā'a</i>	<i>allatī</i>	<i>ištarā-ha</i>
ART=merchandise	RP	purchase.PERF.3SG.M-CL.3SG.F.ACC
the merchandise	that	he purchased it

'on Wednesday the customer from Monday came to us and returned the merchandise he's purchased'

(3) كما كان، من جهته، يأتي إلى حفلاتي

<i>kamā</i>	<i>kāna</i>	<i>min</i>	<i>ḡihati-h</i>	<i>ya'tī</i>	<i>ilā</i>	<i>ḥaflāt-ī</i>
CONJ	AUX	ABL	side-CL.3SG.M.GEN	<i>atā</i> .IMPF.3SG.M	ALL	parties-CL.1SG.GEN
also	was	from	his side	comes	to	my parties

'he also, on his part, used to come to my parties'

(4) ولا بد ان الحكومة تستبعد سياحا يجيئون الى وطننا الجميل للتمتع بحظر التجول

<i>wa=labudda</i>	<i>anna</i>	<i>al=ḥukūma</i>	<i>tastab'ed</i>	<i>suyyāḥ</i>	<i>yaḡṭ'ūn</i>
CONJ=MOD	TOP	ART=government	exclude.IMPF.3SG.F	tourists	<i>ḡā'a</i> .IMPF.3PL.M
and it must	that	the government	excludes	tourists	come

<i>ilā</i>	<i>waṭani-na</i>	<i>al=ḡamīl</i>	<i>li=l=tamattu'</i>	<i>bi=ḥadr-i</i>	<i>al=taḡawwul</i>
ALL	home-CL.1PL.GEN	ART=beautiful	PURP=ART=enjoyment	INST=ban	ART=wandering
to	our home	the beautiful	to the enjoyment	of the ban of	the wandering

'the government must deport tourists who come to our beautiful country to enjoy the curfew'

(5) وكان الأوروبيون يأتون إلى الهند

<i>wa=kāna</i>	<i>al=oroppiyūn</i>	<i>ya'tūn</i>	<i>ilā</i>	<i>al=hind</i>
CONJ=AUX	ART=Europeans	<i>atā</i> .IMPF.3PL.M	ALL	ART=India
and was	the Europeans	come	to	India

'and the Europeans used to come to India'

(6) عبقرية باراك أنه جاء إلى واشنطن وهو يعرف بالتحديد من أين توكل الكتف

<i>'abqariyyat</i>	<i>barak</i>	<i>anna-hu</i>	<i>ğā'a</i>	<i>ilā</i>	<i>wašinṭun</i>	<i>wa=huwa</i>
brilliance	Barack	TOP-CL.3SG.M	ğā'a .PERF.3SG.M	ALL	Washington	CONJ=PP
brilliance	Barack	that he	came	to	Washington	and he
<i>ya'ref</i>	<i>bi=l=tahdād</i>	<i>min</i>	<i>ayna</i>	<i>tu'kal</i>	<i>al=katef</i>	
know.IMP.3SG.M	INST=ART=specification	ABL	Q	eat.PASS.3SG.M	ART=shoulder	
knows	exactly	from	where	be eaten	the shoulder	

'Barack's brilliance lies in the fact that he came to Washington knowing exactly
'where to bite a shoulder'

In cases such as the above, where a goal of the COME motion event is made explicit – 'COME to X' – the nature of the destination can vary. Both *atā* and *ğā'a* physical motion events can have a location as their end point, as in (4) - (6), or an ACTIVITY/EVENT as in the sentence in (3). In some utterances where the GOAL of the COME event corresponds with *speaker deixis*, 'COME to me/us', or when the deictic centre is seen from a 3rd person viewpoint 'COME to him/her/them', *atā* and *ğā'a* are usually used transitively. That is to say, the goal of the motion event is marked by attaching an object pronoun suffix to the COME verb, as in (2) as well as (7). As far as *atā* and *ğā'a* are concerned, such transitive uses are not necessarily associated with physical motion, as I will discuss in §6.1.2.

(7) وجاءه أحد مندوبيه المتجولين الباحثين عن الجميلات

<i>wa=ğā'a-hu</i>	<i>aḥad</i>	<i>mandūbī-hi</i>	<i>al=mutağawwilīn</i>
CONJ= ğā'a .PERF.3SG.M-CL.3SG.M.ACC	one	representative-CL.3SG.M.GEN	ART=travelling
and came to him	one	of his representatives	the travelling

'And one of his travelling representatives who look for beautiful women came to him'

The source of the COME physical motion event can also be made explicit, as in (8) - (10), though such construal is considered rather marginal in contrast to the more frequent goal specifications. Source phrases usually indicate the country, city or geographical region of origin of the COMING human agents. The purpose of motion is either to engage in an activity, as in (8) and (9), or migration, as in (10) and (11). As we saw with GO verbs, the purposive serial verb construction consists of the verb – *atā* or

ğā'a – followed by the purposive preposition *li-* prefixed to a verb in the subjunctive form.

(8) وكان طلاب العلم من أنحاء الدنيا يأتون الى جامعاتنا ليتعلموا الطب والهندسة والفلك

wa=kāna	ṭullāb	al='ilm	min	anḥā'	al=dunyā	ya'tun	ilā
CONJ=AUX	students	ART=knowledge	ABL	parts	ART=world	atā .IMPF.3PL.M	ALL
and was	students	the knowledge	from	parts	the world	come	to

<i>ğāmi'āti-na</i>	<i>li=yata'allamū</i>	<i>al=ṭib</i>
universities-CL.1PL	PURP=learn.IMPF.3PL.M	ART=medicine
our universities	to learn	the medicine

<i>wa=l=handasa</i>	<i>wa=l=falak</i>
CONJ=ART=geometry	CONJ=ART=astronomy
and the geometry	and the astronomy

'and seekers of knowledge came to our universities from all over the world to learn medicine, geometry and astronomy'

(9) جمهور لبناني جاء من المناطق القريبة والبعيدة ليشاهد نجوم فرقة البولشوي الروسية

<i>ğumhūr</i>	<i>lubnāni</i>	<i>ğā'a</i>	<i>min</i>	<i>al=manāiq</i>	<i>al=qarība</i>	<i>wa=l=ba'īda</i>
audience	Lebanese	ğā'a .PERF.3SG.M	ABL	ART=areas	ART=near	CONJ=ART=far
audience	Lebanese	came	from	the areas	the near	and the far

<i>li=yuṣāhid</i>	<i>nuğūm</i>	<i>firqat</i>	<i>al=bulṣuy</i>	<i>al=rūsiyya</i>
PURP= watch.SUBJN.3PL.M	stars	band	ART=Bolshoi	ART=Russian
to watch	stars	the band	the Bolshoi	the Russian

'Lebanese audience that came from near and far to watch the stars of the Russian Bolshoi band'

(10) وهم يأتون الى هذا المكان من سائر انحاء اسرائيل

<i>wa=hum</i>	<i>ya'tūn</i>	<i>ilā</i>	<i>hāḏa</i>	<i>al=makān</i>	<i>min</i>	<i>sā'ir</i>	<i>anḥā'-i</i>	<i>isrā'īl</i>
CONJ=PP	atā .IMPF.3PL.M	ALL	DEM	ART=place	ABL	remaining	parts-GEN	Israel
and they	come	to	this	the place	from	the rest	of parts	Israel

'And they come to this place from all over Israel'

(11) غالبيتهم [...] جاء من مناطق معينة من السعودية

<i>ğālibiyyatu-hum</i>	<i>ğā'a</i>	<i>min</i>	<i>manāṭiq</i>
majority-CL.3PL.M.GEN	ğā'a .PERF.3SG.M	ABL	places
most of them	came	from	places

<i>mu'ayyana</i>	<i>min</i>	<i>al=su'ūdiyya</i>
specific	ABL	ART=Saudi
specific	from	Saudi

'Most of them come from specific parts of Saudi'

Among the common uses of both *atā* and *ğā'a* is the phrasal usage of *atā bi-* and *ğā'a bi-* where the preposition *bi-*, as I mentioned in §4.1.1, can have multiple uses such as marking comitative or instrumental senses, among many others. While *ḏahaba bi-*

seems to indicate the sub-sense ‘to take’, *atā bi-* and *ğā’a bi-* are associated with the sub-sense ‘to bring’, as shown in (12) and (13).

(12) كولمبوس السفاح الذي [...] أتى بأمراض الزهري والسيلان إلي أمريكا من أوروبا

<i>kulumbus</i>	<i>al=saffāḥ</i>	<i>allaḍī</i>	<i>atā</i>	<i>bi=’amrāḍ</i>	<i>al=zuhri</i>
Columbus	ART=assassin	RP	atā .PERF.3SG.M	COM=diseases	ART=syphilis
Columbus	the assassin	who	came	with diseases	the syphilis

<i>wa=l=saylān</i>	<i>ilā</i>	<i>amrika</i>	<i>min</i>	<i>oroppa</i>
CONJ=ART=gonorrhea	ALL	America	ABL	Europe
and the gonorrhea	to	America	from	Europe

‘Columbus the assassin who [...] brought syphilis and gonorrhea to America from Europe’

(13) فجاءته بإحدى القريبات من اللاني هينن ليكن ربات بيوت

<i>fa=ğā’at-hu</i>	<i>bi=’ihdā</i>	<i>al=qarībāt</i>	<i>min</i>	<i>allā’ī</i>
CONJ= ğā’a .PERF.3SG.F-CL.3SG.M.ACC	COM=one	ART=relative	ABL	RP
and she came to him	with one	the relatives	from	who

<i>huyyi’na</i>	<i>li=yakunna</i>	<i>rabbāt</i>	<i>buyūt</i>
prepare.PASS.3PL.F	PURP=be.SUBJN.3PL.F	mistresses	houses
were prepared	to be/become	mistresses	houses

‘She brought him one of the relatives who were prepared to become housewives’

The notion that COME and GO verbs can take part in lexicalization processes involving instrumental and comitative purposes has been attested in Rama (Craig, 1991), where *siik* ‘come’ combines with what she calls the relational preverb *yu-* to give rise to the partially lexicalized usage ‘to bring’ (1991:467). In many Arabic dialects, the only COME verb used is *ğeh* or *yeh*, which is clearly a phonologically modified form of *ğā’a*. The verb meaning ‘to bring’ in these dialects is *ğāb*. Ferguson (1959) pointed out that the lexicalization of the middle weak verb (i.e. in which the second root consonant is a semi-consonant, e.g. /w/ or /y/) /ğyb/ ‘to bring’ “has arisen from a fusion, at some early date, of [ğā’a] and bi-” (1959:62), and according to him this seems to be an evidence of the koineisation process among classical and vernacular varieties of Arabic.³⁹

³⁹ “The modern reflex of *ja’a* is not used with *bi-* to mean ‘bring’. The regular word for ‘bring’ in the dialects is a new verb *jab* (imperfect *yjib*), which clearly has arisen from a fusion, at some early date, of *ja’a* and *bi-*. This verb behaves like a middle-weak verb (*j y b*) with full regularity of form and no evidence of any morphemic boundary remaining between the original *ja’a* part and

As far as the motion of non-humans is concerned, such event construals are highly infrequent in the physical domain usage of *atā* and *ǧā'a*. I did, however, find an instance of *atā* usage, as in (14) that depicts “motion” of inanimate entities. What seems to characterize the event schema in (14) is the lack of voluntary motion, as opposed to the previous examples involving motion of human agents.

(14) رسوم البضائع التي تأتي مباشرة وبشكل غير قانوني الى سورية

<i>rusūm</i>	<i>al=baḡā'i'</i>	<i>allatī</i>	<i>ta'tī</i>	<i>mubašara</i>	<i>wa=bi=šakl</i>
tariffs	ART=merchandise	RP	atā .IMPF.3SG.F	directly	CONJ=INST=shape
tariffs	the merchandise	which	come	directly	and in shape

<i>ǧayr</i>	<i>qānūni</i>	<i>ilā</i>	<i>sūrya</i>
NEG	legal	ALL	Syria
non-	legal	to	Syria

‘The tariffs on merchandise that comes directly and illegally to Syria’

We could possibly consider the travelling sound waves in space towards a hearer’s ear as another type of physical motion that is less salient. Both *atā* and *ǧā'a* can take part in utterances that involve the motion of sound, as in (15) – (17), which also count among the marginal uses of both verbs. Interestingly, in the few instances found where the moving theme is ‘sound’, when a verb of perception is used in such utterances, it would be *sami'a* ‘hear’, rather than *istama'a* ‘listen’, as in (15) and (16), which emphasizes the role of the experiencer, being the recipient of ‘sound’ or the deictic centre towards which the sound is travelling.

(15) سمعت ذلك الصوت الذي أتاني في البداية يقول لي

<i>sami'tu</i>	<i>ḡālika</i>	<i>al=šawt</i>	<i>alladī</i>	<i>atā-ni</i>	<i>fi</i>
hear.PERF.1SG	DEM	ART=voice	RP	atā .PERF.3SG.M-CL.1SG.ACC	LOC
I heard	that	the voice	which	came to me	in

<i>al=bidāya</i>	<i>yaqūl</i>	<i>l=i</i>
ART=beginning	say.IMPF.3SG.M	ALL=CL.1SG
the beginning	says	to me

‘I heard that voice that came to me at first saying to me’

the original *bi-* part. In the Classical language there is no trace of the fused verb. [...] To explain the persistence of this pattern throughout the Arab world one would have to assume that this unparalleled fusion was made at many times and places and always outlived the other forms. The common origin of the dialects is a much simpler explanation.” Ferguson (1959:629).

(16) فسمع صوتا يأتي من ناحية الصالون

<i>fa=sami'a</i>	<i>ṣawtan</i>	<i>ya'tī</i>	<i>min</i>	<i>nāḥiyat</i>	<i>al=ṣālon</i>
CONJ=hear.PERF.3SG.M	voice	atā .IMPF.3SG.M	ABL	side	ART=living.room
and he heard	voice	comes	from	side	the living room

'And he heard a voice coming from where the living room is'

(17) جاءني صوت أقدس عبد الحميد على الهاتف لتقول انها لا تنام الليل

<i>ḡā'a-ni</i>	<i>ṣawt</i>	<i>aqdas</i>	<i>'abdelḥamīd</i>	<i>'alā</i>	<i>al=hātef</i>
ḡā'a .PERF.3SG.M-CL.1SG.ACC	voice	Aqdas.GEN	Abdulhamid	LOC	ART=phone
came to me	voice	of Aqdas	Abdulhamid	over	the phone

<i>li=taqūl</i>	<i>anna=hā</i>	<i>lā</i>	<i>tanām</i>	<i>al=layl</i>
PURP=say.IMPF.3SG.M	TOP=CL.3SG.F	NEG	sleep.IMPF.3SG.F	ART=night
to say	that she	not	sleeps	the night

'The voice of Aqdas Abdulhamid came to me over the phone to say that she can't sleep at night'

The motion of 'fire' can also count as an instance of physical motion, as in (18).

In this particular construal, the flames are moving over a PATH or a particular area, i.e. the machines in the factory, therefore leaving nothing behind but ashes. The phrasal use of *atā 'alā* 'come over/on' here is therefore associated with the sub-sense 'to destroy' or 'to demolish'. This construal can also map on to the figurative domain, as we will see in §6.1.2. This sub-sense is only associated with uses of *atā*. I could not find such corpus instances involving *ḡā'a*.

(18) وأنت ألسنة اللهب على جميع الأجهزة والآلات والماكينات

<i>wa=atat</i>	<i>alsinat</i>	<i>al=lahab</i>	<i>'alā</i>	<i>ḡamī'</i>	<i>al='aḡhiza</i>
CONJ= atā .PERF.3SG.F	flames	ART=fire	LOC	entire	ART=appliances
and came	flames	the fire	over	entire	the appliances

<i>wa=l='ālāt</i>	<i>wa=l=makināt</i>
CONJ=ART=instrument	CONJ=ART=machines
and the instrument	and the machines

'And the flames destroyed all the appliances and instruments and machines'

6.1.2 Non-physical motion

Unlike *ḥaḍara* and *qadima*, the majority of the annotated corpus returns of *atā* and *ḡā'a* belong to the domain of figurative motion event construals. In my discussion on both the physical and figurative uses of GO verbs, I mentioned that the deictic aspect of

the motion event does not necessarily play a role in event construals revolving around a GO verb. This is not the same with COME verbs, since a deictic reading of a COME event is inescapable. This undoubtedly gives rise to various metaphorically extended uses based on the notion of motion towards the speaker or the deictic centre, as will be apparent in the following discussion.

Fictive motion towards the speaker – *speaker deixis* (Fillmore, 1966) – counts among the most prototypical uses of both *atā* and *ḡā'a*, as shown in (19) – (21), where the theme is an abstract notion, such as ‘justice’, ‘knowledge’, and ‘opportunity’.

(19) بما لا يعنى أن من بعدهم ستأتينا العدالة

<i>bi=mā</i>	<i>lā</i>	<i>ya'nī</i>	<i>anna</i>	<i>min</i>	<i>ba'di-him</i>
INST=RP	NEG	mean.IMPF.3SG.M	TOP	ABL	after-CL.3PL.M
of what	not	means	that	from	after them

<i>sa-ta'tī-na</i>	<i>al= 'adāla</i>
FUT- <i>atā</i> .IMPF.3SG.F-CL.1PL	ART=justice
will come to us	the justice

‘Which does not mean that after them justice will come to us’

(20) فماذا يقول لربه بعد أن جاءه البيان

<i>fa=māqā</i>	<i>yaqūl</i>	<i>li=rabbi-hi</i>	<i>ba'da</i>	<i>an</i>
CONJ=Q	say.IMPF.3SG.M	PURP=god-CL.3SG.M	after	TOP
and what	says	to his god	after	that

<i>ḡā'a-hu</i>	<i>al=bayān</i>
<i>ḡā'a</i> .PERF.3SG.M-CL.3SG.M	ART=knowledge
came to him	the knowledge

‘What would he say to his god after knowledge has come to him’

(21) وقد جاءتهم الفرصة

<i>wa=qad</i>	<i>ḡā'at-hum</i>	<i>al=furṣa</i>
CONJ=DM	<i>ḡā'a</i> .PERF.3SG.F-CL.3PL.M	ART=chance
and had	came to them	the chance

‘And the chance has come to them’

Again, very few utterances include the source of the fictive motion event, as we can see in (22) and (23).

(22) اذ ثمة ايضا الأخطار الخارجية التي تأتينا من الدول القادرة و«المتقدمة»

id *tammata* *aydan* *al=aḥṭār* *al=ḥārġiyya* *allatī* *ta'tī-na*
 ADV ADV ADV ART=dangers ART=external RP **atā**.IMPF.3SG.F-CL.1PL
 since there is also the dangers the external that come to us

min *al=duwal* *al=qādira* *wa=l=mutaqaddima*
 ABL ART=countries ART=powerful CONJ=ART=advanced
 from the countries the powerful and the advanced

‘There are also the external dangers that come to us from the powerful and
 “advanced” countries’

(23) تلقت السياسة الاميركية الخارجية مؤخرا العديد من الإنتقادات جاءت أشدها من الوزيرة السابقة مادلين أولبرايت

tallaqat *al=siyāsa* *al=amrikiyya* *al=ḥārġiyya* *mo'axxaran* *al='adīd*
 receive.PERF.3SG.F ART=policy ART=American ART=external ADV ART=many
 received the policy the America the external lately the many

min *al='intiḳādāt* *ġā'at* *ašadda-hā* *min* *al=wazīra*
 ABL ART=criticism **ġā'a**.PERF.3SG.F harshest-CL.3SG.F ABL ART=minister
 of the criticism came harshest of it from the minister

al=sābiqa *madlin olbrayt*
 ART=late Madeline Albright
 the late Madeline Albright

‘American Foreign policy lately received lot of criticism, the harshest of which
 came from the ex-minister, Madeline Albright’

The added fictive motion aspect seems to strengthen the ablative ‘from’ sense:

‘external dangers from advanced countries’ and ‘the harshest criticism from the ex-

minister’. Both sentences can be paraphrased, as in (24) and (25), to either exclude the

deictic motion verb or replace it with a copula.

(24) اذ ثمة ايضا الأخطار الخارجية من الدول القادرة و«المتقدمة»

id *tammata* *aydan* *al=aḥṭār* *al=xārġiyya* *min*
 ADV ADV ADV ART=dangers ART=external ABL
 since there is also the dangers the external from

al=duwal *al=qādira* *wa=l=mutaqaddima*
 ART=countries ART=powerful CONJ=ART=advanced
 the countries the powerful and the advanced

‘There are also the external dangers from the powerful and “advanced” countries’

(25) تلقت السياسة الاميركية الخارجية مؤخرا العديد من الإنتقادات كانت أشدها من الوزيرة السابقة مادلين أولبرايت

tallaqat *al=siyāsa* *al=amrikiyya* *al=ḥārġiyya* *mo'aḥḥaran* *al='adīd*
 receive.PERF.3SG.F ART=policy ART=American ART=external ADV ART=many
 received the policy the America the external lately the many

<i>min</i>	<i>al=’intiḡādāt</i>	<i>kānat</i>	<i>ašadda-hā</i>	<i>min</i>	<i>al=wazīra</i>
ABL	ART=criticism	be.PERF.3SG.F	the.harshesht-CL.3SG.F	ABL	ART=minister
of	the criticism	came	harshesht of it	from	the minister

<i>al=sābiqa</i>	<i>madlin olbrayt</i>
ART=late	Madeline Albright
the late	Madeline Albright

‘American Foreign policy lately received lot of criticism, the harshesht of which was from the ex-minister, Madeline Albright’

It appears that, while the preposition *min* ‘from’ already introduces an ablative sense, the use of *atā* and *ḡā’a* in conjunction with this preposition further highlights the spatio-temporal dimension to the ablative ‘from’ sense. While the use of COME verbs may be a stylistic choice in these MSA sentences, in other languages, COME verbs may fully grammaticalize into ablative markers (Heine and Kuteva, 2002).

The uses of *atā* and *ḡā’a* exemplified in (19) - (25) are rather marginal in contrast to the following more idiomatic expressions. One of the frequent uses of *atā* and *ḡā’a* that is characteristic of newspaper writing depicts an entity such as human, vehicle, country or an abstract notion, achieving a certain ranking. The construal of the deictic centre, to which the theme “moves”, can vary: ‘in first place’, ‘at the forefront’, ‘at the last step’, ‘at the centre’, ‘at the top’ of a certain listing, as we can see in (26)-(30). The use of *atā* in such context is nevertheless more frequent than that of *ḡā’a*.

(26) اعتبر أن الحب يأتي في الدرجة الأولى

<i>a’tabir</i>	<i>anna</i>	<i>al=ḡub</i>	<i>ya’tī</i>	<i>fi</i>	<i>al=daraḡa</i>	<i>al=’ūla</i>
consider.IMPF.1SG	TOP	ART=love	<i>atā</i> .IMPF.3SG.M	LOC	ART=degree	ART=first
I consider	that	the love	comes	in	the first	the place

‘To me, loves comes in first place’

(27) واعترف بأن السياحة العربية لليمن تأتي في السلم الأخير

<i>wa=a’taref</i>	<i>bi=anna</i>	<i>al=siyaḡa</i>	<i>al=’arabiyya</i>	<i>ta’tī</i>
CONJ=admit.IMPF.1SG	INST=TOP	ART=tourism	ART=Arabic	<i>atā</i> .IMPF.3SG.F
and I admit	of that	the tourism	the Arabic	comes

<i>fi</i>	<i>al=sullam</i>	<i>al=aḡīr</i>
LOC	ART=step	ART=last
in	the step	the last

‘And I admit that Arabic tourism in Yemen comes in last place’

(28) ويأتي على على رأس هذه المحاصيل القمح والقطن والشمندر السكري والتفاح والحمضيات

wa=ya'tī *'ala ra's hāḍihi* *al=maḥāṣīl* *al=qamḥ* *wa=l=quṭn*
 CONJ=*atā*.IMPF.3SG.M LOC head DEM ART=crops ART=wheat CONJ=ART=cotton
 and comes on top of these the crops the wheat and the cotton

wa=l=ṣamandar *al=sukkari* *wa=l=tuffaḥ* *wa=l=ḥimdiyyāt*
 CONJ=ART=beets ART=sweet CONJ=ART=apples CONJ=ART=citruses
 and the beets the sweet and the apples and the citruses
 'And on top of these crops comes wheat, cotton, beets, apples and citruses'

(29) وجاء في المركز السادس الأخير الزورق «فيكتوري»

wa=ḡā'a *fi* *al=markaz* *al=sādis* *al=aḥīr* *al=zawraq* *victory 4*
 CONJ=*ḡā'a*.PERF.3SG.M LOC ART=place ART=sixth ART=last ART=boat Victory 4
 and came in the place the sixth the last the boat Victory 4
 'and in sixth and final place came the boat "Victory 4"'

(30) وجاءت في المرتبة الثالثة دولة الإمارات العربية

wa=ḡā'at *fi* *al=martaba* *al=tālīṭa* *dawlat*
 CONJ=*ḡā'a*.PERF.3SG.F LOC ART=place ART=third country
 and came in the place the third country

al='imārāt-i *al='arabiyya*
 ART=Emirates-GEN ART=Arab
 of the Emirates the Arab
 'And in third place came the United Arab Emirates'

The examples of comparable usage of *atā* and *ḡā'a* discussed so far show an interesting pattern. While both verbs can express the same event construal, the morphological aspect marking on the verb differs among the two verbs. *Atā* is mostly inflected in the IMPERFECTIVE, whereas *ḡā'a* is almost exclusively inflected in the PERFECTIVE. The above sentences in (26)-(30) are no exception. This preference, as has already been shown in the quantitative analysis, may be symptomatic of a suppletive alternation between these two highly synonymous verbs. This may be further illustrated with a constructional 'minimal pair'. The sentence in (31) is a widely-cited line from the poetry of Al-Mutanabbi, a famous poet from the 10th century:

(31) تجري الرياح بما لا تشتهي السفن

<i>tağrī</i>	<i>al=riyāḥ</i>	<i>bi=mā</i>	<i>lā</i>	<i>taštahī</i>	<i>al=sufun</i>
run.IMPF.3SG.F	ART=wind	INST=RP	NEG	desire.IMPF.3SG.F	ART=ships
runs	the wind	with what	not	desire	the ships

‘The wind runs contrary to what the ships desire’

This particular line became an idiomatic expression used in both formal and colloquial Arabic to express undesirable outcomes. The original motion verb in this line is related to the verb ‘to run’, *ğarā*, as in ‘the wind runs contrary to what the ship desires’. In modern usage, the RUN verb is sometimes substituted with a COME verb. I looked up this particular quote in the newspaper sub-section of arabiCorpus (www.arabicorpus.byu.edu) for the following inflected forms: *ta’ti* ‘*atā*.IMPF’, *atat* ‘*atā*.PERF’, *ğā’at* ‘*ğā’a*.PERF’, and *tağī* ‘*ğā’a*.IMPF’.⁴⁰ I found 15 instances of *atā*.IMPF and 3 of *atā*.PERF, as well as 9 instances of *ğā’a*.PERF and none of *ğā’a*.IMPF. Sentences (32) and (33) are examples of the corpus hits for *atā*.IMPF and *ğā’a*.PERF in this construction. The selectional restriction here – of whether the verb should be *atā* or *ğā’a* – seems to be mostly related to the TAM marking of the motion verb. If the verb slot requires a COME verb in the past, inflected in the perfective, the selected verb is *ğā’a*; if it is inflected in the imperfective, for a present tense, the verb would be *atā*.

(32) وكاد حماسهم ينجح ولكن تأتي الرياح بما لا تشتهي السفن

<i>wa=kāda</i>	<i>ḥamāsu-hum</i>	<i>yanğah</i>	<i>wa=lākin</i>	<i>ta’īl</i>
CONJ=MOD	enthusiasm-CL.3PL.M	succeed.IMPF.3SG.M	CONJ=CONJ	<i>atā</i> .IMPF.3SG.F
and almost	their enthusiasm	succeed	and but	comes

<i>al=riyāḥ</i>	<i>bi=mā</i>	<i>lā</i>	<i>taštahī</i>	<i>al=sufun</i>
ART=wind	INST=RP	NEG	desire.IMPF.3SG.F	ART=ships
the wind	with what	not	desire	the ships

‘Their enthusiasm could have worked but the wind comes contrary to what the ships prefer’

⁴⁰ Needless to say, a widely-cited poetry line like this one is subject to modifications to fit the specific situation described by Al-Mutanabbi’s verse. I restricted my corpus search of this quote to COME *alriyāḥ bi-* ‘COME the wind INST-’.

(33) ولكن جاءت الرياح بما لا تشتهي السفن

<i>wa=lākin</i>	<i>ġā'at</i>	<i>al=riyāḥ</i>	<i>bi-mā</i>	<i>lā</i>	<i>taštahī</i>	<i>al=sufun</i>
CONJ=CONJ	ġā'a .PERF.3SG.F	ART=wind	INST-RP	NEG	desire.IMPF.3SG.F	ART=ships
and but	came	the wind	with what	not	desire	the ships

‘But the wind came contrary to what the ships desire’

Further evidence to support this claim can be seen in sentence (34), taken directly from ArabiCorpus. Here we have an ideal example in which COME is used both in the perfective and the imperfective in the same sentence (‘came and [still] comes’) to indicate the continuous event of ‘pistachios coming from Iran’. The perfective ‘came’ is encoded by *ġā'a* while the imperfective ‘comes’ is encoded by *atā*.

(34) المستوردات الإسرائيلية من الفستق جاءت وتأتي من إيران

<i>al=mustawradāt</i>	<i>al='isrā'iliyya</i>	<i>min</i>	<i>al=fustuq</i>	<i>ġā'at</i>
ART=imports	ART=Israeli	ABL	ART=pistachios	ġā'a .PERF.3SG.F
the imports	the Israeli	of	the pistachios	came

<i>wa=ta'ī</i>	<i>min</i>	<i>irān</i>
CONJ= atā .PERF.3SG.F	ABL	Iran
and comes	from	Iran

‘The Israeli pistachio imports came and still come from Iran’

Related to the construction in (32) and (33), and to the construction discussed in §6.1.1, corpus data showed that *atā bi-* and *ġā'a bi-* (associated with the sub-sense ‘to bring’) can also extend to the figurative domain as in (35) and (36). Unlike in sentences (12) and (13), the motion aspect in (35) and (36) is bleached out. This figurative use of COME is similar to the English *come up with*. The sentences in (37) and (38) demonstrate the same phrasal usage. However, the sub-sense implied here may also be interpreted as ‘cause’, which seems to be a concept that is closely related to ‘bring’. Recall again that such usage counts among the marginal, yet, common uses of *atā* and *ġā'a*.

(35) ومن قائل بأن رئيس الوزراء الإسرائيلي المنتخب لن يأتي بجديد

<i>wa=man</i>	<i>qā'il</i>	<i>bi=anna</i>	<i>ra'īs</i>	<i>al=wuzarā'-i</i>	<i>al='isrā'ili</i>	<i>al=muntaḥab</i>
CONJ=Q	say.AP.3SG.M	INST=TOP	head	ART=ministers-GEN	ART=Israeli	ART=elected
and who	sayer	of that	head	the ministers	the Israeli	the elected

lan *ya'ti* *bi-ḡadīd*
 NEG.FUT **atā**.SUBJN.3SG.M COM=new
 will not come with new
 ‘And who said that the newly elected Israeli prime minister will not bring/come up with anything new’

(36) فلا ندري من أين جاء الكاتب بهذا الكلام

fa=lā *nadrī* *min* *ayna* *ḡā'a* *al=kātib*
 CONJ=NEG know.IMP.F.1PL ABL Q **ḡā'a**.IMP.F.3SG.M ART=writer
 and not we know from where came the writer

bi=hāda *al=kalām*
 COM=DEM ART=talk
 with that the talk

‘We don’t know where the writer came up with this’

(37) سياسات التحرير الاقتصادي التي جاءت بمردود ايجابي

siyāsāt *al=taḥrīr-i* *al='iqtisādī* *allatī* *ḡā'at*
 policies ART=liberation-GEN ART=economic RP **ḡā'a**.PERF.3SG.F
 policies of the liberation the economic which came

bi=mardūd *īḡābiy*
 COM=outcome positive
 with outcome positive

‘The policies of economic liberation that resulted in good outcomes’

(38) للقطاع الخاص المستعجل والذي قلما يأتي بهموم وهو اجس سوى الهاجس المادي

li=l=qitā' *al=hāṣ* *wa=l=musta'ḡal* *wa=allaḡī* *qallamā* *ya'tī*
 OBL=ART=sector ART=private CONJ=ART=rushed CONJ=RP rarely **atā**.IMP.F.3SG.M
 to the sector the private and the rushed and which rarely comes

bi=humūm *wa=hawāḡis* *siwā* *al=hāḡis* *al=māddiy*
 COM=concerns CONJ=obsessions except ART=obsession ART=materialistic
 with concerns and obsessions except the obsession the materialistic

‘To the rapidly-growing private sector that rarely causes any concerns or obsessions besides materialistic obsessions’

In §6.1.1, I mentioned that the phrasal use of *atā 'alā* indicates the meaning ‘to destroy/demolish’. Sentence (39) involves a metaphorically extended usage of this construction. In (18) the physical motion of flames over a certain area resulted in complete destruction, while in (39) the deficit is conceptualized as the destructive force that *comes over* the bank’s capital, and hence causing it to be lost.

(39) وقد أعلن مؤخرا عن خسارة تفوق 120 مليار سنتيم أنت على رأسمال البنك

<i>wa=qad</i>	<i>a'lana</i>	<i>mu'ahḥaran</i>	<i>'an</i>	<i>ḥasāra</i>	<i>tafūq</i>
CONJ=DM	announce.PERF.3SG.M	lately	about	loss	exceed.IMPF.3SG.F
and had	announced	lately	about	loss	exceeds

<i>120 millyār</i>	<i>sentīm</i>	<i>atat</i>	<i>'alā</i>	<i>ra'smāl</i>	<i>al=bank-i</i>
120 billion	Moroccan dirham	<i>atā</i> .PERF.3SG.F	LOC	capital	ART=bank-GEN
120 billion	Moroccan dirham	came	over	capital	of the bank

'And [the bank] has announced lately a deficit that exceeds 120 billion Moroccan dirham which consumed the bank's capital'

There is no doubt that the construal in (39) includes a path of the motion event: the bank's capital, as well as the machines in sentence (18), over which the cause of destruction is 'moving'. This counts as one of the few instances of verb use where the path of the motion event frame is highlighted. Recall from Chapter 5 that path of motion was not found to be a distinguishing element in the usage of either verb. The above phrasal usage as well as the following idiomatic expressions count among the few instances in which *atā* or *ḡā'a* event frames would include information about the trajectory.

The idiomatic expressions *atā/ḡā'a 'alā dīkr* 'come over the mention', as in (40) and (41), and *atā/ḡā'a 'alā lisān* 'come over the tongue [of x]', as in (42) and (43), also count among common, but infrequent uses of the two verbs. These two expressions provide different ways to construe the sub-sense 'to mention' using COME motion events figuratively. In (40) and (41) the theme (i.e. moving entity) is the person mentioning a certain topic, while the path of motion is the mention of the topic itself, such as 'the mention of questions' in (40) and 'the mention of Kissinger' in (41). In (42) and (43), on the other hand, the topic that is being mentioned is construed as the theme, while the 'tongue' of the person mentioning the topic constitutes the path '[topic] comes over/on the tongue of [somebody]'.

(40) الأسئلة التي أتينا على ذكرها تحمل بين طياتها بعض الإجابات

<i>al=as'ila</i>	<i>allatī</i>	<i>ataynā</i>	<i>'alā</i>	<i>ḍikri-hā</i>	<i>taḥmil</i>	<i>bayna</i>
ART=questions	RP	atā .PERF.1PL	LOC	mention-CL.3SG.F	carry.IMPF.3SG.F	LOC
the questions	that	we came	over	its mention	carry	between

<i>ṭayyāti-ha</i>	<i>ba'ḍa</i>	<i>al='iḡabāt-i</i>
folds-CL.3SG.F	some	ART=answers-GEN
its folds	some	of the answers

'The questions that we mentioned carry some answers between the lines'

(41) ليته ما جاء على ذكر كيسنجر

<i>layta-hu</i>	<i>mā</i>	<i>ḡā'a</i>	<i>'alā</i>	<i>ḍikr</i>	<i>Kissinger</i>
MOD-CL.3SG.M	NEG	ḡā'a .PERF.3SG.M	LOC	mention	Kissinger-GEN
if only he	not	came	over	mentin	of Kissinger

'If only he didn't mention Kissinger'

(42) صيغة الجمع تأتي على لسان الفرد لتقرير أمر عام

<i>ṣīḡat</i>	<i>al=ḡam'-i</i>	<i>ta'tī</i>	<i>'alā</i>	<i>lisān</i>	<i>al=fard-i</i>
formulation	ART=plural-GEN	atā .IMPF.3SG.F	LOC	tongue	ART=individual-GEN
formulation	of the plural	comes	over	tongue	of the individual

<i>li=taqrīr</i>	<i>amr</i>	<i>'ām</i>
PURP=establishment	case	general
to establish	case	general

'the plural form can be used to refer to generic types by the individual'

(43) تأكيدات سورية لانتفك تتواصل أبرزها ما جاء على لسان رئيس الوزراء

<i>ta'kīdāt</i>	<i>suriyya</i>	<i>lā tanfakku</i>	<i>tatawāṣal</i>	<i>abrazu-hā</i>	<i>mā</i>
confirmations	Syrian	MOD	continue.IMPF.3SG.F	most.prominent-CL.3SG.F	RP
confirmations	Syrian	do not stop	continue	the most prominent of it	what

<i>ḡā'a</i>	<i>'alā</i>	<i>lisān</i>	<i>ra'īs-i</i>	<i>al=wuzarā'-i</i>
ḡā'a .PERF.3SG.M	LOC	tongue	head-GEN	ART=ministers
came	over	tongue	of head	of the ministers

'Syrian confirmations that are continuous, the most prominent of which was what the prime minister mentioned'

Returning to the more frequent and statistically significant uses of the two verbs,

I commented earlier that a large number of *atā* and *ḡā'a* uses in the corpus are reflective of the genre of newspaper writing. Some examples of frequent expressions involving both verbs are related to 'reporting speech or a statement' are shown in (44) – (47).

(44) فيما يأتي نص المقابلة مع مهديّة بن بلة

fī=mā ya'tī naṣ al=muqābala-ti ma'a mahdiyya bin balla
LOC=RP **atā**.IMPF.3SG.M text ART=interview-GEN COM Mahdiyya Bin Balla
in what comes text of the interview with Mahdiyya Bin Balla
'The text of the interview with Madiyya Bin Balla is in the following'

(45) وجاء في رد رسمي للجيش الاسرائيلي... أن

wa=ḡā'a fī radd rasmeṣ li=l=ḡayṣ al='isrā'īli anna
CONJ=**ḡā'a**.PERF.3SG.M LOC reponse official ALL=ART=military ART=Israeli TOP
and came in response official of the military the Israeli that
'In an official response from the Israeli military ... was...'

(46) ورد الرئيس الهراوي بكلمة جاء فيها

wa=radda al=ra'īs alhrāwi bi=kalmia
ART=reply.PERF.3SG.M ART=president Al Hrawi INST=statement
and replied the president Al Hrawi with statement

ḡā'a fī=hā
ḡā'a.PERF.3SG.F LOC=CL.3SG.F
came in it

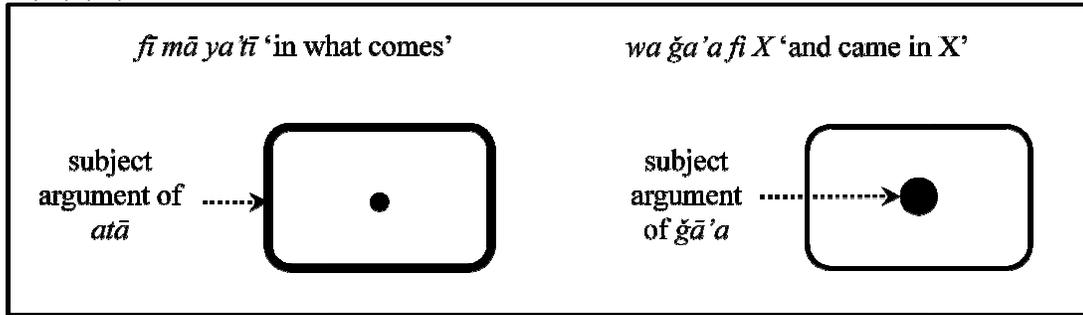
'And President Hrawi responded with a statement in which he said'

(47) وكان نص اليمين كما يأتي

wa=kāna naṣ al=yamīn-i kamā ya'tī
ART=be.PERF.3SG.M text ART=oath-GEN ADV **atā**.IMPF.3SG.M
and was text of the oath as comes
'The verbatim text of the oath was as follows'

Most of these frequently recycled expressions involve the use of a locative preposition *fī* 'in/within' and are, therefore, reflective of a 'container' spatial relation. For instance, in (44), the subject argument of the predicate *atā* is conceptualized as the container or the space (physical or figurative) that contains the reported interview verbatim. The subject argument in such a construction is typically signalled by a relative pronoun *mā* 'what' and basically refers to 'the following lines/text'. On the other hand, in (45) and (46), the subject argument of the verb *ḡā'a*, represents the reported speech/statement verbatim, rather than representing the figurative space in which the reported speech is contained.

FIGURE 1. Schematic representation of the ‘container’ spatial relations expressed by *atā* and *ġā’a* in (44)-(46).



According to Radden (1996), one of the metaphorical extensions that draw on the deictic property of COME verbs is the notion of *coming into existence*, which can be interpreted according to the conceptual metaphors EXISTENCE IS LOCATION HERE and COMING INTO EXISTENCE IS MOVING HERE (1996:442). This conceptual metaphor seems to motivate numerous *atā* and *ġā’a* uses. In (48) and (49) a ‘visit’ is conceptualized as coming into existence, or happening, within a certain setting, e.g. political context. Recall that EVENT/ACTIVITY X SETTING was one of the most robust configurations related to the use of *atā*.

(48) مشيرا الى أن الزيارة تأتي ضمن الجهود الفرنسية لدعم مكافحة الإرهاب

<i>muširan</i>	<i>ilā</i>	<i>anna</i>	<i>al=ziyara</i>	<i>ta'ti</i>	<i>dimna</i>	<i>al=ġuhud</i>
point.out.AP.3SG.M	ALL	TOP	ART=visit	<i>atā</i> .IMPF.3SG.F	ADV	ART=efforts
point out	to	that	the visit	comes	among	the efforts

<i>al=faransiyya</i>	<i>li=mukāfaḥat</i>	<i>al='irhāb-i</i>
ART=French	PURP=fighting	ART=terrorism-GEN
the French	to fighting	of the terrorism

‘pointing out that the visit comes as part of the French efforts to support fighting terrorism’

(49) إن الزيارة جاءت في اطار التحضير والتنسيق لاجاد موقف عربي موحد

<i>inna</i>	<i>al=ziyāra</i>	<i>ġā'at</i>	<i>fī</i>	<i>iṭār</i>	<i>al=taḥḍīr-i</i>
TOP	ART=visit	<i>ġā'a</i> .PERF.3SG.F	LOC	frame	ART=preparation-GEN
that	the visit	came	in	frame	of the preparation

<i>wa=l=tansīq-i</i>	<i>li='iġād</i>	<i>mawqif</i>	<i>'arabi</i>	<i>muwaḥḥad</i>
CONJ=ART=coordination-GEN	PURP=creating	position	Arabic	united
and of the coordination	to creating	position	Arabic	united

‘The visit came within the context of the preparations made to create a unanimous Arabic position’

The corpus returns that depict the sub-sense ‘happen’ or ‘come into existence’ have also been found to include a phrase that locates a particular event in time, as in (50) and (51), which is a very frequent usage of *ḡā’a* in particular. In addition, an entity coming into existence may be fulfilling a certain purpose, as in (52) and (53), or it can be the result of a particular cause, as in (54) and (55).

(50) يأتي ذلك في وقت بدأت فيه بوادر خلاف بين الولايات المتحدة والإتحاد الأوروبي

<i>ya’rī</i>	<i>dālika</i>	<i>fī</i>	<i>waqt</i>	<i>bada’at</i>	<i>fī=h</i>	<i>bawādir</i>
atā .IMPF.3SG.M	DEM	LOC	time	start.PERF.3SG.M	LOC-CL.3SG.M	signs
comes	that	at	time	started	in it	signs

<i>ḥilāf-in</i>	<i>bayna</i>	<i>al=wilāyāt</i>	<i>al=muttaḥida</i>	<i>wa=l=’ittiḥād</i>	<i>al=’oroppiy</i>
conflict-GEN	LOC	ART=states	ART=united	CONJ=ART=union	ART=European
of conflict	between	the states	the united	and the union	the European

‘This happens/comes at a time when conflict between the USA and the EU started to emerge’

(51) وجاء اغلب هذا الارتفاع في الأشهر الثلاثة الماضية

<i>wa=ḡā’a</i>	<i>aḡlab</i>	<i>hāḡā</i>	<i>al=’irtifā’</i>	<i>fī</i>	<i>al=ašhur</i>
CONJ= ḡā’a .PERF.3SG.M	most	DEM	ART=rise	LOC	ART=months
and came	most	this	rise	in	the months

<i>al=’alāya</i>	<i>al=māḡiya</i>
ART=three	ART=past
the three	the past

‘Most of this rise [of the value of the US dollar] happened during the last three months’

(52) وأنت زيارة شيراك لتطمئن اللبنانيين

<i>wa=atat</i>	<i>ziyāra</i>	<i>šīrāk-in</i>	<i>li=tuṭam’ina</i>	<i>al=lubnaniyyīn</i>
CONJ= atā .PERF.3SG.F	visit	Chirac-GEN	PURP=reassure.SUBJN.3SG.F	ART=Lebanese
and came	visit	of Chirac	to reassure	the Lebanese

‘And Chirac’s visit came to reassure the Lebanese’

(53) وجاءت حرب أكتوبر 73 المجيدة لتعيد لمصر والأمة العربية كرامتها

<i>wa=ḡā’at</i>	<i>ḡarb</i>	<i>uktobar 73</i>	<i>al=maḡīda</i>	<i>li=tu’īd</i>
CONJ= ḡā’a .PERF.3SG.F	war	October 73	ART=glorious	PURP=restore.SUBJN.3SG.F
and came	war	October 73	the glorious	to restore

<i>li=mašr</i>	<i>wa=l=’umma</i>	<i>al=’arabiyya</i>	<i>karāmata-hā</i>
ALL=Egypt	CONJ=ART=nation	ART=Arab	dignity-CL.3SG.F
to Egypt	and the nation	the Arab	its dignity

‘The glorious war of October ’73 came to restore the dignity of Egypt and the Arab nation’

(54) بأن نقد التاريخ يأتي بدافع الحرص على إحيائه

bi=anna naqd al=tārīḥ-i ya'tī bi=dāfi'
 INST=TOP criticism ART=history-GEN *atā*.IMPF.3SG.M INST=motivation
 of that criticism of the history comes of motivation

al=ḥirṣ-i 'alā 'ihyā'-ih
 ART=desire-GEN LOC reviving-CL.3SG.M
 of the desire on reviving it

'That criticism of history comes from the desire to revive it'

(55) وجاء اختيار جعفر بعد اتهام كثيرين للمدرب الحالي بالطاعة العمياء لكرول

wa=ḡā'a iḥtiyār ḡa'far-in ba'da ittihām kaṭirīn li=l=mudarreb
 CONJ=*ḡā'a*.PERF.3SG.M choosing Jaffar-GEN LOC accusing many ALL=ART=coach
 and came choosing of Jaffar after accusing many to the coach

al=ḥāli bi=l=tā'a al='amyā' li=krol
 ART=current INST=ART=obedience ART=blind ALL=Krol
 the current of the obedience the blind to Krol

'Choosing Jafaar came after having accused the current coach of blindly obeying Krol'

6.2 *Ḥaḍāra*

The verb *ḥaḍāra* may sometimes be interchangeable with *atā* and *ḡā'a*, to the extent that even some classical and modern Arabic dictionaries list *ḥaḍāra* as a synonym of these two verbs.⁴¹ According to the analyzed data frame, the interchangeability between *ḥaḍāra* and the two verbs discussed above seems to be restricted to physical deictic motion events, as I will discuss in §6.2.1. The vast majority of *ḥaḍāra* uses, however, pertain to the sub-sense 'to attend' or 'to be present', in which the motion aspect of COME is downplayed to a large extent, and the focus of attention is shifted towards presence at the endpoint instead. What may bias such interpretation is the transitive use of the verb, as in (56), and the lack of a GOAL prepositional phrase as we saw with *atā* and *ḡā'a* and as I will explain shortly.⁴²

⁴¹ e.g. *Takmilat Al-Ma'āḡim Al-'arabiyya* (1871-1877-1927-1978), *Al-Munḡid fi Al-Luḡa wa Al-A'lam* (2005), and *Mu'ḡam Al-Luḡa Al-'arabiyya Al-Mu'āšira* (2008).

⁴² Both *atā* and *ḡā'a* can also be used transitively. However, the transitive construction associated with these two verbs is restricted to instances in which the syntactic object is represented as an object pronoun suffixed to the verb, as in 'COME to me/him/them/it, etc.' where the object argument is referred to anaphorically.

(56) مصر تحضر اجتماع وزراء الخارجية العرب

<i>maṣr</i>	<i>taḥḍur</i>	<i>'iḡtimā'-a</i>	<i>wuzarā'-i</i>	<i>al=ḥāriḡiyya</i>	<i>al='arab</i>
Egypt	ḥaḍara .PERF.3SG.F	meeting-ACC	ministers-GEN	ART=external	ART=Arabs
Egypt	attends	meeting	of ministers	the external	the Arabs

'Egypt attends the meeting for Arab Foreign Affairs ministers'

Having said that, the conceptualization of the event in (56) may still imply motion to a certain degree. That is, a motion event is presupposed prior to 'attending' the meeting. Therefore, physical motion – though not salient in such usage of the verb – should not be entirely ruled out. Following this assumption, I decided to include instances of the sub-sense 'to attend'/'to be present' as part of the following discussion on physical and non-physical motion, instead of introducing it in a separate section.

6.2.1 Physical motion

Corpus returns of *ḥaḍara* include instances of verb usage in which the motion event is quite salient, similar to that observed in physical motion events expressed with *atā* and *ḡā'a*. More typically, the goal of motion in such constructions is indicated by a prepositional phrase, *li-/ilā* 'to', as in (57) and (58) 'came with me to Aswan' and 'they come to my office'.

(57) لقد حضر معي لاسوان الدكتور ستيف سيفل

<i>laqad</i>	<i>ḥaḍara</i>	<i>ma'-ī</i>	<i>li=aswān</i>	<i>al=doktor</i>	<i>stīv sevil</i>
DM	ḥaḍara .PERF.3SG.M	COM-CL.1SG	ALL=Aswan	ART=doctor	Steve Seville
had	came	with me	to Aswan	the doctor	Steve Seville

'Doctor Steve Seville came with me to Aswan'

(58) انهم يحضرون الى مكنتي كما يذهبون الى الطبيب البشري

<i>inna=hum</i>	<i>yaḥḍurūn</i>	<i>ilā</i>	<i>maktab-i</i>	<i>kamā</i>	<i>yaḍhabūn</i>
TOP=CL.3PL.M	ḥaḍara .IMPF.3PL.M	ALL	office-CL.1SG	ADV	ḡahaba .IMPF.3PL.M
they	come	to	my office	also	go

<i>ilā</i>	<i>al=ṭabīb</i>	<i>al=bašari</i>
ALL	ART=doctor	ART=human.ADJ
to	the doctor	the human-related

'They come to my office as much as they go to a doctor for humans'

In (59) and (60) the endpoint is not explicitly stated. It may be elided, as in ‘the security forces came *to the protest location* last Monday night’ and ‘and then Shams and his driver came *to his house* and received one bag’. These are instances of use, as I discussed earlier, where the speaker of the utterance is metonymically present at the deictic centre (Radden, 1996).

(59) إن السلطات الأمنية قد حضرت ليلة الإثنين الماضي ومزقت خيمة المعتصمات

<i>inna</i>	<i>al=suluṭāt</i>	<i>al=amniyya</i>	<i>qad</i>	<i>ḥaḍarat</i>	<i>laylat</i>
TOP	ART=authorities	ART=security.ADJ	DM	ḥaḍara .PERF.3SG.F	night
that	the authorities	the security-related	had	come	night

<i>al='iṅṅayn-i</i>	<i>al=māḍi</i>	<i>wa=mazzaqat</i>	<i>ḥaymat</i>	<i>al=mu'taṣimāt</i>
ART=Monday-GEN	ART=past	CONJ=tear.down.PERF.3SG.F	tent	ART=protestors.FEM
of the Monday	the last	and tore down	tent	the female protestors

‘That security forces had come last Monday night and tore down the female protestors tent’

(60) وحضر بعدها شمس وسائقه واستلما كيسا واحدا فيه 51 كلغ من الهيرويين الأسمر

<i>wa=ḥaḍara</i>	<i>ba'da-hā</i>	<i>šamṣ</i>	<i>wa=sā'iqu-h</i>
CONJ= ḥaḍara .PERF.3SG.M	LOC-CL.3SG.FEM	Shams	CONJ=driver-CL.3SG.M
and came	after it	Shams	and his driver

<i>wa=istalamā</i>	<i>kīs-an</i>	<i>wāḥdan</i>	<i>fī-h</i>	<i>51 klġ</i>	<i>min</i>
CONJ=receive.PERF.3DL.M	bag-ACC	one	LOC-CL.3SG.M	51 kg	ABL
and the two received	bag	one	in it	51 kg	of

<i>al=hirowīn</i>	<i>al=asmar</i>
ART=heroin	ART=dark
the heroin	the dark

‘And afterwards Shams and his driver came and received one bag of 51 kg of dark heroin’

Structurally, corpus returns such as the above may either refer to deictic, physical motion event, or they can refer to the intransitive use of ‘to attend’ as in (61). It is necessary to examine the larger context in which the event is taking place to determine the intended sub-sense of *ḥaḍara*. As we can see in (61), the interpretation of the verb meaning in this sentence, ‘to attend’, is contingent on the preceding sentence (retrievable from ArabiCorpus) in which ‘a sports tournament’ is mentioned.

(61) كما سيحضر امين عام الجامعة العربية الدكتور عصمت عبدالمجيد

<i>kamā</i>	<i>sa=yahḏur</i>	<i>amīn 'ām</i>	<i>al=ḡāmi'a-i</i>
ADV	FUT= <i>ḥaḏara</i> .IMPF.3SG.M	Secretary-General	ART=league-GEN
also	will attend/come	Secretary-General	of the league

<i>al='arabiyya</i>	<i>al=doktor</i>	<i>'iṣmat 'abdelmaḡīd</i>
ART=Arab	ART=doctor	Ismat Abdulmajid
the Arab	the doctor	Ismat Abdulmajid

'The Secretary-General of the Arab League, Dr. Ismat Abdulmajeed, will also attend [the tournament]'

As I mentioned earlier, the sub-sense 'to attend', shifts the focus towards 'presence at the endpoint' rather than the process of getting there. It is no surprise then that the nature of the 'GOAL' of *ḥaḏara* in such uses mostly indicates an event such as a meeting, protest, lecture, summit, conference, as in the repeated example in (62) and in (63).

(62) مصر تحضر اجتماع وزراء الخارجية العرب

<i>maṣr</i>	<i>taḥḏur</i>	<i>'iḡtimā'-a</i>	<i>wuẓarā'-i</i>	<i>al=ḥāriḡiyya</i>	<i>al='arab</i>
Egypt	<i>ḥaḏara</i> .PERF.3SG.F	meeting-ACC	ministers-GEN	ART=external	ART=Arabs
Egypt	attends	meeting	of ministers	the external	the Arabs

'Egypt attends the meeting for Arab Foreign Affairs ministers'

(63) وقال إنه حضر هذه التظاهرة للتعبير عن إدانته لانتهاك حقوق الإنسان بالمغرب

<i>wa=qāla</i>	<i>inna=hu</i>	<i>ḥaḏara</i>	<i>hāḏihi</i>	<i>al=tazāhura-ta</i>
CONJ=say.PRF.3SG.M	TOP=CL.3SG.M	<i>ḥaḏara</i> .PERF.3SG.M	DEM	ART=protest-ACC
and said	that he	attended	this	the protest

<i>li=l=ta'bīr</i>	<i>'an</i>	<i>idānati-hi</i>	<i>li='intihāk</i>	<i>ḥuqūq-i</i>
PURP=ART=expressing	about	condemnation-CL.3SG.M	ALL=violating	rights-GEN
for the expressing	about	his condemnation	to violating	the rights

<i>al='insān-i</i>	<i>bi=l=maḡrib</i>
ART=human-GEN	LOC=ART=Morocco
of the human	in the Morocco

'He said that he attended this protest to express his condemnation of the violation of human rights in Morocco'

The above sentences can be considered prototypical uses of *ḥaḏara* due to the fact that this particular usage accounts for over 50% of the overall coded verb returns. Notice that in the transitive constructions in (62) and (63), the object argument is expressed nominally, which is a morphosyntactic property that distinguishes the use of

ḥaḍara from that of the remaining COME verbs in modern usage. The object argument can nevertheless be expressed pronominally, where the pronoun anaphorically refers to the event that is being attended by the subject argument, as in (64).

(64) واقِمت في المناسبة حفلة حضرها الأهل والأصدقاء

<i>wa=uqīmat</i>	<i>fi</i>	<i>al=munāsaba</i>	<i>ḥafla</i>	<i>ḥaḍara-hā</i>
CONJ=make,stand.PASS.3SG.F	LOC	ART=occasion	party	<i>ḥaḍara</i> .PERF.3SG.M-CL.3SG.F
and was raised	in	the occasion	party	attended it

<i>al=ahl</i>	<i>wa=l=aṣḍiqā'</i>
ART=family	CONJ=ART=friends
the family	and the friends

'A party was held for this occasion and was attended by family and friends'

As far as the deictic property of *ḥaḍara* in the above sentences is concerned, I mentioned that in a sentence such as (59) or (60) – and possibly (63) – the deictic centre may be collapsible with *speaker deixis* (Fillmore, 1966) if we consider the speaker's metonymic presence at the goal of the COME event (Radden, 1996). However, in the 'attend' examples, and relying on my native speaker intuition, it does not seem to be possible to get a deictic interpretation of the verb in (61)-(63) the same way we could with *ḥaḍara* in (57) – (60). In other words, it seems more feasible to imagine the speaker of the utterances in (57)-(60) at or identifying with the GOAL, but not easily so with (61)-(63). Psycholinguistic experimentation may be required to further investigate this distinction.

6.2.2 Non-physical motion

The sub-sense of *ḥaḍara* meaning 'to recall' is listed in a number of monolingual and bilingual dictionaries as one of its many uses. Only a handful of *ḥaḍara* corpus returns related to this usage were found, two of which are exemplified in (65) and (66). These two sentences also feature *ḥaḍara* in a transitive construction, with the object argument – being an object pronoun – suffixed on the verb. The object pronoun here

(most likely in the 1st person) refers to the speaker who is recalling a certain memory. It can also refer to ‘thinking’ or ‘being inspired’, as in (67).

(65) يحضرنني كلام قاله زياد مرة في زمن بعيد

<i>yaḥḍuru-ni</i>	<i>kalām-un</i>	<i>qāla-hu</i>	<i>ziyād</i>	<i>marra</i>
ḥaḍara .IMPF.3SG.M-CL.1SG	talk-NOM	say.PERF.3SG.M-CL.3SG.M	Ziyad	once
comes to me	talk	said it	Ziyad	once

<i>fi</i>	<i>zaman</i>	<i>baʿīd</i>
LOC	time	far
in	time	far

‘I recall something that Ziyad had once said long time ago’

(66) ولم يعد يحضرنني اسم الطبيب

<i>wa=lam yaʿud</i>	<i>yaḥḍuru-ni</i>	<i>ism-u</i>	<i>al=ṭabīb-i</i>
CONJ=NEG return.JUSS.3SG.M	ḥaḍara .IMPF.3SG.M-CL.1SG	name-NOM	ART=doctor-GEN
and no longer	comes to me	name	of the doctor

‘I could no longer recall the name of the doctor’

(67) لبث الجواهري منتقلا ما بين بيروت ودمشق لا يحضره شيء

<i>labiṭa</i>	<i>alḡawāheri</i>	<i>mutanaqqilan</i>	<i>mā</i>	<i>bayna</i>	<i>bayrūt</i>
remain.PERF.3SG.M	Al Jawaheri	travel.AP.3SG.M	RP	LOC	Beirut
remained	Al Jawaheri	travelling	what	between	Beirut

<i>wa=dimašq</i>	<i>lā</i>	<i>yaḥḍuru-hu</i>	<i>šayʿ-un</i>
CONJ=Damascus	NEG	ḥaḍara .IMPF.3SG.M-CL.3SG.M	something-NOM
and Damascus	not	comes to him	something

‘The Jawaheri kept travelling back and forth between Beirut and Damascus without getting a single bit of inspiration’

The sub-sense related to ‘attending’ discussed in §6.2.1. can also extend to metaphorical presence of non-human entities, such as ‘pain’, ‘sarcasm’ and ‘flowers’ as in (68) and (69).

(68) وفي المقالات التي قرأتها حضر الألم كما حضرت السخرية

<i>wa=fi</i>	<i>al=maqālāt</i>	<i>allati</i>	<i>qaraʿtu-ha</i>	<i>ḥaḍara</i>	<i>al=alam</i>
CONJ=LOC	ART=articles	RP	read.PERF.1SG-CL.3SG.F	ḥaḍara .PERF.3SG.M	ART=pain
and in	the articles	that	I read it	came/was present	the pain

<i>kamā</i>	<i>ḥaḍarat</i>	<i>al=suḥriya</i>
ADV	ḥaḍara .PERF.3SG.F	ART=sarcasm
also	came/was present	the sarcasm

‘In the articles I read, pain was present as well as sarcasm’

(69) فالورد يحضر في الحب الحزين أيضا

<i>fa=l=ward</i>	<i>yaḥḍur</i>	<i>fi</i>	<i>al=ḥub</i>	<i>al=ḥazīn</i>	<i>ayḍan</i>
CONJ=ART=flowers	ḥaḍara .IMPF.3SG.M	LOC	ART=love	ART=sad	ADV
and the flowers	come/ is present	in	the love	the sad	too

‘Flowers are also present in sorrowful love’

6.3 *Qadima*

Inspection of the coded corpus returns of *qadima* showed that this verb is exclusively used to refer to physical motion events. More typically, *qadima* can be labeled as the ‘migration’ verb in the set of COME verbs studied here, due to the fact that almost half of the verb uses are inflected in the plural, in addition to showing statistically robust preferences towards specifying the SOURCE and the GOAL of the motion event. In addition, *qadima* mostly collocates with sentential subjects referring to human agents. In other words, this verb is more likely used to talk about large groups of humans moving from one location to another. The sentence in (70) may count as a prototypical example of the use of *qadima*, which depicts the migration of large groups of people from one geographical area to another.

(70) الاميركيون اللاتينيون الذين قدموا الى تكساس من الجنوب

<i>al=amrikiyyūn</i>	<i>al=lātiniyyūn</i>	<i>allaḍina</i>	<i>qadimū</i>	<i>ilā</i>	<i>teksās</i>
ART=Americans	ART=Latin	RP	qadima .PERF.3PL.M	ALL	Texas
the Americans	The Latin	who	came	to	Texas

<i>min</i>	<i>al=ḡanūb</i>
ABL	ART=south
from	the South

‘The Latin Americans who came to Texas from the south’

6.3.1 Physical motion

In my qualitative analysis of GO verbs, I discussed the GOAL-bias that GO motion events seem to show, in that more verb uses encoded the endpoint rather than the source or the path of the motion event. We may assume that this also holds true for the three COME verbs discussed in the previous sections. In the *atā* data frame, 113 sentences included GOAL, while 69 included SOURCE; for *ḡā’a*, 60 sentences included GOAL, and 37

sentences included SOURCE; for *ḥaḍara*, 419 sentences included GOAL, while only 16 sentences included SOURCE. As for *qadima*, the amount of sentences containing SOURCE of the motion event surpasses that of the sentences including GOAL: 226 GOAL phrases and 241 SOURCE phrases. This in itself is an interesting observation and, as the following analysis will reveal, is not coincidental.

In §6.1.1, I mentioned that the inclusion of the SOURCE of the *atā* or *ḡā'a* motion event (mostly a geographical area) is to indicate the purpose of the COME event, e.g. to engage in a specific activity, or for migration purposes. Such usage is marginal as far as the overall usage of *atā* and *ḡā'a* is concerned. This purposeful motion *from* one (geographical) location *to* another, nevertheless, constitutes the largest part of *qadima* usage, as in (70) as well as (71) and (72).

(71) ان عدد السياح الذين قدموا الى الأردن من البلدان الخليجية مجتمعة أكثر من 192 ألف سائح

<i>anna</i>	<i>'adad</i>	<i>al=suyyāḥ-i</i>	<i>allaḍina</i>	<i>qadimū</i>	<i>ilā</i>	<i>al='urdn</i>
TOP	number	ART=tourists-GEN	RP	<i>qadima</i> .PERF.3PL.M	ALL	ART=Jordan
that	number	of the tourists	who	came	to	the Jordan

<i>min</i>	<i>al=buldān</i>	<i>al=ḥalīḡiyya</i>	<i>muḡtami'a</i>	<i>aḡtar</i>	<i>min</i>	<i>192 alf</i>	<i>sā'iḥ</i>
ABL	ART=countries	ART=Gulf.ADJ	combined	more	ABL	192 thousand	tourist
from	the countries	the Gulf	combined	more	from	192 thousand	tourist

'That the number of tourists who came to Jordan from the Gulf countries combined was more than 192 thousand tourist'

(72) اجدادنا الذين قدموا الى هذه الشعاب البركانية منذ ثلاثمائة سنة تقريبا من جبل لبنان و حلب و ادلب و فلسطين

<i>aḡdādu.na</i>	<i>allaḍina</i>	<i>qadimū</i>	<i>ilā</i>	<i>ḥāḍihi</i>	<i>al=ši'āb</i>
grandfathers.CL.1PL	RP	<i>qadima</i> .PERF.3PL.M	ALL	DEM	ART=regions
our grandfathers	who	came	to	these	the regions

<i>al=burkaniyya</i>	<i>munḡu</i>	<i>ṭalāṭmi'at</i>	<i>sana</i>	<i>taqrīban</i>	<i>min</i>	<i>ḡabal</i>	<i>lubnān</i>
ART=volcanic	ADV	three hundred	year	ADV	ABL	mountain	Lebanon
the volcanic	since	three hundred	year	almost	from	mountain	Lebanon

<i>wa=ḥalab</i>	<i>wa=idlib</i>	<i>wa=falasīn</i>
CONJ=Aleppo	CONJ=Idlib	CONJ=Palestine
and Aleppo	and Idlib	and Palestine

'Our forefathers who came to these volcanic regions from Mount Lebanon, Aleppo, Idlib, and Palestine around 300 years ago'

A number of *qadima* instances can also include motion to/towards a non-geographical location, or another human being, as in (73) and (74), or no SOURCE or GOAL as in (75).

(73) الجمهور الذي لبي الدعوة... قدم الى القاعة من أمكنة مختلفة في لبنان ومن أزمنة مختلفة

<i>al=ḡumhūr</i>	<i>allaḡī</i>	<i>labbā</i>	<i>al=da'wa</i>	<i>qadima</i>	<i>ilā</i>
ART=audience	RP	answer.PERF.3SG.M	ART=invitation	<i>qadima</i> .PERF.3SG.M	ALL
the audience	who	answered	the invitation	came	to

<i>al=qā'a</i>	<i>min</i>	<i>amkina</i>	<i>muḡḡalifa</i>	<i>fi</i>	<i>lubnān</i>	<i>wa=min</i>	<i>azmina</i>	<i>muḡḡalifa</i>
ART=hall	ABL	places	different	LOC	Lebanon	CONJ=ABL	times	different
the hall	from	places	different	in	Lebanon	and from	times	different

'The audience who accepted the invitation [...] came to the hall from different parts of Lebanon and from different generations'

(74) قد أخفى سروره حين قدم اليه موفدو الكسي كومنين

<i>qad</i>	<i>aḡfā</i>	<i>surūra-hu</i>	<i>ḡīna</i>	<i>qadima</i>
DM	hide.PERF.3SG.M	happiness-CL.3SG.M	ADV	<i>qadima</i> .PERF.3SG.M
had	hid	his happiness	when	came

<i>'ila=yhi</i>	<i>mufadū</i>	<i>aleksī komnīn</i>
LOC=CL.3SG.M	delegates	Alexios-GEN Komnenos
on him	delegates	of Alexios Komnenos

'He had suppressed his joy when the delegate of Alexios Komnenos came to him'

(75) وكانت أنوثة المرأة التي قدمت من نوع آخر

<i>wa=kānat</i>	<i>'unūḡat</i>	<i>al=mar'a-ti</i>	<i>allati</i>	<i>qadimat</i>
ART=be.PERF.3SG.F	femininity	ART=woman-GEN	RP	<i>qadima</i> .PERF.3SG.F
and was	femininity	of the woman	who	came

<i>min</i>	<i>naw'in</i>	<i>āḡar</i>
ABL	type	different
from	type	different

'The femininity of the woman who came was of a different kind'

Recall that *atā* and *ḡā'a* can be used transitively in a sentence such as (74) where the GOAL of the motion event is a human being. This particular construction is not allowed with *qadima*. However, as I will explain later, *qadima* may still participate in a transitive construction.

A handful of the coded corpus hits include a non-human subject argument, as in (76)-(79). Similar to the examples in (71) and (72), in (76) and (77) the subject argument metonymically refers to large groups of people – ‘migrations’ and ‘French colonialism’.

(76) مع الهجرات الجنوبية التي قدمت من الهند وإيران والعراق وآسيا الصغرى

ma‘a	al=hiġrāt	al=ġanūbiyya	allatī	qadimat	min	al=hind
COM	ART=migrations	ART=southern	RP	<i>qadima</i> .PERF.3SG.M	ABL	ART=India
with	the migrations	the southern	that	came	from	the India

wa=’irān	wa=l=’irāq	wa=āsya	al=ṣuġrā
CONJ=Iran	CONJ=ART=Iraq	CONJ=Asia	ART=small
and Iran	and the Iraq	and the Anatolia	Eyalet

‘With the Southern waves of immigration that came from India, Iran, Iraq, and Anatolia Eyalet’

(77) فالإستعمار الفرنسي قدم الى بلادنا في ظل اتفاقية سايكس بيكو الاستعمارية

fa=l=’isti‘mār	al=faransi	qadima	ilā	bilādi-na
CONJ=ART=colonialism	ART=French	<i>qadima</i> .PERF.3SG.M	ALL	country-CL.1PL
and the colonialism	the French	came	to	our country

fī	zil	ittifāqiyat-i	sayks pīko	al=isti‘mariyya
LOC	shadow	agreement-GEN	Sykes Picot-GEN	ART=colonialist
in	shadow	of agreement	of Sykes-Picot	the colonialist

‘French colonialism came to our country as a result of the colonialist Sykes-Picot agreement’

A number of sentences also included an inanimate subject argument denoting a vehicle, as in ‘airplanes’ in (78), as well as other airborne projectiles in (79).

(78) وأشار الى أن الطائرات التي قدمت من الأجواء السعودية والكويتية

wa=ašāra	ilā	anna	al=ṭā’irāt	allatī	qadimat
CONJ=point.PERF.3SG.M	ALL	TOP	ART=airplanes	RP	<i>qadima</i> .PERF.3SG.F
and pointed	to	that	the airplanes	that	came

min	al=aġwā’	al=su’ūdiyya	wa=l=kuwaytiyya
ABL	ART=airspaces	ART=Saudi	CONJ=ART=Kuwaiti
from	the airspaces	the Saudi	and the Kuwaiti

‘And he pointed out that the airplanes that came from the Saudi and Kuwaiti airspace’

(79) ويقدر ان بعض هذه المقذوفات قدم الى الارض من القمر والمريخ

wa=yuqaddar	anna	ba‘d	hāḍihi	al=maqḍūfāt
CONJ=estimate.PASS.3SG.M	TOP	some	DEM	ART=projectiles
and is estimated	that	some	these	the projectiles

qadima ilā al=arḍ min al=qamar wa=l=marrīḥ
qadima.PERF.3SG.M ALL ART=earth ABL ART=moon CONJ=ART=Mars
 came to the earth from the moon and the Mars
 ‘And it is estimated that some of these projectiles came to earth from the moon or Mars’

Purposeful motion is another distinguishing characteristic of the use of *qadima*.

As we saw in the HCFA analysis in Table 11 (Chapter 5), the configurations HUMAN X PURPOSIVE and HUMAN X GOAL X PURPOSIVE were quite robust. Purposeful motion can be expressed by a phrase indicating the objective of COMING, as in (80) and (81). In these sentences, an individual or a group of individuals move to a certain part of the world to fulfill a certain mission, e.g. ‘to carry out an investigation’ or ‘to preserve peace’.

(81) وقد قدم خصيصا من بلده المانيا ليجري تحقيقا عن معمرنا

wa=qad *qadima* *ḥiṣṣān* *min* *baladi-hi* *almanya*
 CONJ=DM *qadima*.PERF.3SG.M especially ABL country-CL.3SG.M Germany
 and had came especially from his country Germany

li=yuḡriya *taḥqīqan* ‘*an* *mu‘ammiri-na*
 PURP=conduct.SUBJN.3SG.M investigation about old.person-CL.1PL
 to conduct investigation about our oldest person
 ‘And he had especially come from his country, Germany, to conduct an interview with our oldest senior citizen’

(81) أن الجنود الأمريكيين الذين قدموا الى الصومال لحفظ السلام

anna *al=ḡunūd* *al=amrikiyyīn* *allaḡina* *qadimū* *ilā*
 TOP ART=soldiers ART=American RP *qadima*.PERF.3PL.M ALL
 that the soldiers the America who came to

al=ṣūmāl *li=ḥifz* *al=salām-i*
 ART=Somalia PURP=preserving ART=peace-GEN
 the Somalia to preserve the peace
 ‘That the American soldiers who came to Somalia to preserve peace’

In a large subset of *qadima* corpus returns, the lexical semantics of the sentential subject can also convey the purpose of the motion event. *Qadima* collocates to a large extent with nouns referring to an individual or a group of individuals moving from one place to another for a specific reason. For instance it could be migration due to war or economic reasons, as in ‘refugees’, ‘immigrants’, ‘expatriates’, ‘settlers’; moving for

leisure purposes, as in ‘visitors’, ‘tourists’, ‘travellers’, ‘audience’; or for political purposes, e.g. ‘colonialists’, ‘delegate’, ‘army’, etc. Sentences (82) and (83) are examples of such usage.

(82) فهي تحكي عن المستوطنين الاوروبيين الاوائل الذين قدموا الى استراليا

<i>fa=hiya</i>	<i>taḥki</i>	<i>'an</i>	<i>al=mustawinīn</i>	<i>al=oroppiyīn</i>	<i>al=awā'il</i>	<i>allaḍīna</i>
CONJ=PP	tell.PERF.3SG.F	about	ART=settlers	ART=European	ART=first	RP
and she	tells	about	the settlers	the European	the first	who

<i>qadimū</i>	<i>ilā</i>	<i>ustrāliya</i>
<i>qadima</i> .PERF.3PL.M	ALL	Australia
came	to	Australia

‘It talks about the first European settlers who came to Australia’

(83) عدد السياح الذين قدموا الى كندا في الفترة المذكورة

<i>'adad</i>	<i>al=suhhāḥ-i</i>	<i>allaḍīna</i>	<i>qadimū</i>	<i>ilā</i>	<i>kanada</i>	<i>fi</i>
number	ART=tourists-GEN	RP	<i>qadima</i> .PERF.3PL.M	ALL	Canada	LOC
number	of the tourists	who	came	to	Canada	in

<i>al=fatra</i>	<i>al=maḍkūra</i>
ART=period	ART=mentioned
the period	the mentioned

‘The number of tourists who came to Canada in the aforementioned time period’

Qadima can also feature in transitive constructions, as in (84) and (85), in which the direct object of *qadima* would be the endpoint of the motion event. Interestingly, all 10 hits that involve such a construction are sentences extracted from narrative texts (from the newspapers sub-corpus) that specifically narrate stories about Prophet Muhammad or other individuals who lived around that historical period. This construction seems to be a remnant of Classical Arabic that has persisted throughout MSA, through such narratives.

(84) لكون الرسول صلى الله عليه وسلم لما قدم المدينة وجد اليهود يصومونه

<i>li=kawn</i>	<i>al=rasūl</i>	<i>ṣalla allahu 'alayhi wasallam</i>	<i>lammā</i>
PURP=be.VN	ART=prophet	peace and blessings of Allah be on him	ADV
for being	the prophet	peace and blessings of Allah be on him	when

<i>qadima</i>	<i>al=madīna-ta</i>	<i>waḡada</i>
<i>qadima</i> .PERF.3SG.M	ART=Medina-ACC	find.PERF.3SG.M
he came	the Medina	he found

<i>al=yahūd</i>	<i>yaṣūmūna-hu</i>
ART=Jews	fast.IMPF.3PL.M-CL.3SG.M
the Jews	fast it

‘Being that when the prophet, peace and blessings be upon him, came to Medina he found the Jews fasting [that day]’

(85) وكان علي بن عبد الله إذا قدم مكة حاجا أو معتمرا عطلت قريش مجالسها

<i>wa=kāna</i>	<i>'ali bin 'abdillah</i>	<i>iḏā</i>	<i>qadima</i>	<i>makka-ta</i>	<i>ḥāḡḡan</i>
CONJ=be.PERF.3SG.M	Ali Bin Abdullah	COND	<i>qadima</i> .PERF.3SG.M	Mecca-ACC	pilgrim
and was	Ali Bin Abdullah	if	he came	Mecca	pilgrim

<i>aw</i>	<i>mu'tamiran</i>	<i>'aṭṭalat</i>	<i>qurayš</i>	<i>maḡālisa-ha</i>
CONJ	pilgrim	suspend.PERF.3SG.F	Quraysh	meetings-CL.3SG.F
or	minor.pilgrim	suspended	Quraysh	its meetings

'When Ali bin Abdullah used to come to Mecca on a pilgrimage Quraysh would suspend its meetings'

Another usage of *qadima* found mostly in such narratives is the use of the preposition '*alā*' 'on/over' instead of *ilā*' 'ALLATIVE' in the GOAL prepositional phrase, as in (86). In this construction, the GOAL of the COME motion event is always a human being.

(86) وقدم وفد من كندة على رسول الله صلى الله عليه وسلم

<i>wa=qadima</i>	<i>wafd</i>	<i>min</i>	<i>kinda</i>	<i>'alā</i>
CONJ= <i>qadima</i> .PERF.3SG.M	delegate	ABL	Kindah	LOC
and came	delegate	from	Kindah	on

<i>rasūl</i>	<i>allāh-i</i>	<i>ṣalla allahu 'alayhi wasallam</i>
prophet	Allah-GEN	peace and blessings of Allah be on him
prophet	of Allah	peace and blessings of Allah be on him

'And a delegate from Kindah came to the prophet, peace and blessings be upon him'

6.4 Summary

The preceding quantitative and qualitative analysis of MSA COME verbs *atā*, *ḥaḍara*, *ḡā'a*, and *qadima* has provided a rather comprehensive account of the constructional preferences of each of the four COME verbs. I started the analysis of these verbs with the assumption that all four verbs can be used interchangeably used in a physical deictic motion event. The data discussed in the last two chapters confirmed these assumptions and also highlighted other contexts of use that may attract two or more COME verbs. Generally speaking, however, each verb has a distinct usage profile in MSA, which I summarize in the following:

I examined the verbs *atā* and *ǧā'a* side-by-side due to the fact that these two verbs share more features than they do with the remaining COME verbs. For instance, the analysis in this present chapter showed that either verb can be used in a number of constructions that denote physical and figurative motion, as well as the phrasal uses *atā bi-* and *ǧā'a bi-* 'bring' (or Lit. 'come with'). However, this is where the importance of pairing a qualitative analysis with a quantitative analysis is most apparent. *Atā* and *ǧā'a* may share a number of features, yet it is the distributional properties of these verbs that set them apart. For instance, an *atā* motion event is more likely to include a GOAL than a *ǧā'a* motion event is. Furthermore, I mentioned that while *atā* is most likely to be inflected for SIMPLE PRESENT IMPERFECTIVE, *ǧā'a*, on the other hand, almost exclusively appears in SIMPLE PAST PERFECTIVE constructions. In addition, the quantitative analysis showed that *atā* and *ǧā'a* differ to a great extent with respect to the most syntactically robust interactions between SUBJECT SEMANTIC CATEGORY and SEMANTIC PROPOSITIONS (i.e. inclusion of GOAL, SOURCE, MANNER, etc.). In other words, each verb highly associates with different motion event construals, which is partly reflected in the different collocational profiles of the two verbs.

As far as the verb *ḥaḍara* is concerned, I argued that events involving this verb tend to highlight 'presence' at the end point rather than the motion process itself. The sub-sense 'to attend', therefore, constitutes one of the main uses of this verb and explains the numerous instances of transitive uses of *ḥaḍara* in the corpus data. Unlike *atā* and *ǧā'a*, *ḥaḍara* collocates mostly with subjects denoting HUMAN agents or GROUPS of humans (e.g. organization, country, newspaper). This strong collocational pattern reflects the fact that this verb is mostly used to depict a physical (motion) event. A few instances of figurative motion events were, nevertheless, found among the coded *ḥaḍara* returns which indicated the event of 'recalling' or the event of 'being present' when talking, for instance, about the presence of a certain emotion in a poem.

Finally, I discussed that the verb *qadima* is only used to refer to physical motion events. This verb can be regarded as *the* migration verb in MSA. Similar to the pattern observed with *ḥaḍara*, the quantitative analysis in the previous chapter showed that the subject argument of *qadima* is most likely to be either HUMAN or GROUP. Moreover, the subject argument of *qadima* is more likely than any other COME verb to be marked in the plural. Another unique distributional property of *qadima* is the fact that the motion event tends to include a phrase specifying the SOURCE of motion, in addition to specifying the GOAL and the PURPOSE of the motion event. Therefore, many of the corpus returns of *qadima* that were examined in the present chapter depicted large groups of HUMANS moving from a certain location to another, for a specific purpose.

As stated earlier, each of the four COME verbs and three GO verbs has a specific profile in MSA. A combination of statistical and case-by-case examination of verb uses needed to be undertaken in order to zero in on the idiosyncratic behavior of each verb. In the following chapter I will present a synthesis of the results obtained in Chapters 3-6 and an evaluation of the methodological approach adopted for the study of MSA deictic motion verbs.

Chapter Seven Synthesizing the corpus findings

7.1 Using corpus data

Newman (2011) has commented on the role of corpora in five of the core topic areas in the field of cognitive linguistics: metaphor, synonymy, polysemy, prototypes, and constructional analysis. I would like to think that the analysis presented in the previous chapters touched upon most, if not all, of these core topics in relation to the use of GO and COME verbs in Modern Standard Arabic. The statistical investigation of a large number of corpus returns (coded for a wide variety of morphosyntactic and semantic variables) proved to be a quite comprehensive method for understanding the complexity of a (deictic) motion event, and providing a more holistic account of verb usage that goes beyond examining deictic properties of these verbs as well as the reliance on introspection alone.

Let me begin this discussion by briefly re-addressing the role of using corpus data in the context of my analysis of MSA GO and COME verbs. As I mentioned in the introduction to this dissertation, the growing tendency to use corpora in cognitive and usage-based accounts of lexical items and constructions is a reflection of the fundamental premise of this field: Exploring naturally occurring linguistic data, rather than relying mainly on elicited language uses or ones provided by the researcher's own intuition (cf. Sandra and Rice (1995) for a critique of the latter). Not to discredit introspection as a source for linguistic data, but relying on corpora in this kind of research takes advantage of the large-scale resources – such as the 50,000+ corpus returns for verbs such as *ḍahaba* or *atā* – in multiple ways. These corpus-based sets of data allow us to examine not only specific constructions exhibited in individual sentences, but also distributional information regarding various aspects such as inflectional preferences, collocates, or even semantic properties strongly associated with a certain lexical item. Such distributional

facts are an integral part of the speaker's internalized information about the usage of that lexical or constructional element (Bybee 1985, 2007, among others). This level of linguistic knowledge cannot be examined solely through reliance on the methods of introspection and elicitation. This limitation motivates an alternative, usage-based methodology for the examination of GO and COME verbs in MSA, as presented in Chapters 3 and 5.

7.2 Using univariate and multivariate analyses

The creation of a 500-row data frame per verb allowed me to probe into the frequency and distribution facts regarding the usage of the seven (deictic) motion verbs in MSA. Moreover, the annotation of each corpus return for a wide range of constructional and semantic features offered the possibility of foregrounding the most prototypical aspects of use for each verb. Thankfully, there is a range of statistical tests now that can be utilized for examining a large amount of data from different angles and for different purposes. The statistical analyses adopted for this study vary from simple univariate tests of significance to more complex analyses that investigate the effects of interaction among sets of variables.

The univariate analysis – in the form of *chi*-square tests, standardized Pearson's residuals, and even contingency tables – already indicated skewed distributional pattern within (i) the members of a particular inflectional paradigm (e.g. morphological aspect and subject number), as well as (ii) across the verbs in a verb set. Calculating the standardized Pearson's residuals, in particular, was a useful heuristic for examining single variable distribution, as well as identifying the variables that do not seem to distinguish the use of one verb from another (e.g. due to low frequency of occurrence). The challenge of investigating contingency tables and frequency counts in this kind of analysis is not a trivial one, especially when it comes to discussing a language that is considered very rich in its verb inflectional paradigms, but I will elaborate more on this later.

The methods I opted for in the multivariate analysis each contribute in a unique way to our understanding of the synonymous relations among verbs in a set (GO or COME) as well as the kinds of constructions that typically associate with each member of the set. First of all, the hierarchical agglomerative cluster analysis, which I adapted from Gries and colleagues' *Behavioural Profiles* quantitative method (e.g. Divjak and Gries, 2006; Gries and Otani, 2010; etc.), helped identify the overall joint effect of the constructional features that each contextual verb use was coded for through measuring the behavioural distance between verbs and grouping them in clusters.

A second multivariate method, hierarchical configural frequency analysis provided a more detailed look at which constructional elements frequently co-occur to a statistically significant degree. The hcfa 3.2 script provides a number of test statistics such as the contribution to *chi*-square, as well as the significance level per configuration ("Dec") which was signalled by, for instance, '****' or 'ns', and which are determined by calculating the adjusted Holm *p*-value. In my analysis of the robust configurations associated with GO and COME verbs, I decided to report on these significance levels in combination with the "Q" value, which is also another measure of the significance of the configuration value. In principle, there is no limit on the number of variables that can be considered using hcfa 3.2. However, as I addressed earlier, for practical purposes I had to break down the entire variable set into sub-groups of constructional elements. This can be regarded more of an advantage, rather than a limitation, since it forces the researcher to focus on smaller constructions by selecting particular sub-sets of morphological, syntactic, and semantic variables at a time.

In the hierarchical configural frequency analysis in Chapters 3 and 5, I started with the examination of the most common TAM marking tendencies of each verb in contrast to the other GO or COME verbs, then moved on to examining morphological and semantic properties pertaining to the syntactic subject. Through these two separate steps,

I managed to identify the most frequent and robust morphosyntactic features of each verb in addition to the characteristics of the sentential subject it most prototypically associates with. Needless to say, the verbs showed huge discrepancies in terms of the clusters of variables they each associated with the most. An additional step I followed in this analysis was to examine the larger conceptual frames that characterize the use of each GO and COME verb. For example, I examined the semantic category of the subject combined with phrases indicating the occurrence of SOURCE, GOAL, MANNER, PATH, etc. of the motion event underlying each verb's usage. The main objective of grouping these variables together was to zero in on the various construals that involve GO and COME verbs. Indeed, I managed to find robust interactions among these variables, some of which overlapped between verbs, such as SUBJECT_HUMAN x GOAL, for the COME verbs *atā*, *ḥaḍara* and *qadima*, while others were unique to a particular verb, such as SUBJECT_NOTION x SOURCE for *atā* or SUBJECT_ACTIVITY x MANNER for *ḡā'a*.

The third type of multivariate analysis adopted in this dissertation, polytomous logistic regression analysis, emphasizes the notion that lexical choices (e.g. in a near-synonymy situation) are probabilistic rather than categorical (Bresnan, 2006, 2007; Arppe, 2007, 2008, 2009, etc.). First, a set of variables had to be selected to be included in the regression model (based on univariate and bivariate analyses). The logistic regression model then calculated the “odds” per explanatory variable (i.e. the extent to which the existence of a variable would increase or decrease the chances of the occurrence of each verb in a particular context – with all other variables being equal). The model, as we saw, also calculated the probability of occurrence of each verb per context. I did find some sentences where one verb is most likely to occur than all the others (with probability estimate almost equal to 1.0), while in other cases I found two or more verbs having almost equal probability estimates. By testing these probability estimates against my native speaker intuition, I found that many of the categorical as well as equi-probable

estimates were intuitively satisfying, especially as far as COME verbs are concerned.⁴³ These findings motivated a psycholinguistic experiment aimed at comparing the probability estimates calculated by the logistic regression model with lexical choices made by literate Bahraini speakers of Arabic – a topic which I will talk more about in the next chapter. As far as GO verbs are concerned, and as I discussed in Chapter 3, the probability estimates that were calculated did not seem as intuitively correct as the ones calculated for the COME verbs. I theorized that this may be caused by lack of a certain defining variable that the data frame was not coded for, e.g. collocations, or, more likely, the idea that the MSA GO verbs studied here have less in common between each other than the MSA COME verbs do. In my opinion, it is a combination of both factors that has led to such results with the MSA GO verbs.

As I mentioned at the beginning of this discussion, Newman (2011) identified the study of ‘sense determination’ as one of the most researched topics in the field of cognitive linguistics, and there is no doubt that both the quantitative and qualitative analyses conducted in this study have shed light on the (near-) synonymous relationships, or lack thereof, between the MSA GO verbs on the one hand and the COME verbs on the other. As evident in the qualitative analysis, of all the verbs studied here, the pair of verbs that seem to have a more or less synonymous relationship are the COME verbs *atā* and *ǧā’a*. It was even assumed that the relationship between those two verbs is almost suppletive: while both may appear in similar constructions, *atā* shows a preference to be inflected in the imperfective, whereas *ǧā’a* appears most of the time in the perfective. Nevertheless, the qualitative analysis only tells one side of the story. What we also need to consider is the frequency of the various constructions that attract both verbs and determine whether or not the distribution of these constructions are similar across the two

⁴³ I am using the term “native speaker of MSA” quite loosely here, since there are no real native speakers of Modern Standard Arabic. It is a variety of Arabic that is formally learned rather than acquired.

verbs. Of course, the quantitative analysis showed that they are not. For instance, the HCFA analysis supported the fact that the two verbs differ with respect to the most frequent and robust construals. Typically, *atā* would denote the sub-sense of ‘to happen / to take place’ (SUBJECT_EVENT or ACTIVITY X SETTING.YES) or the physical or figurative motion of humans or groups of humans towards a goal. The verb *ǵā’a*, on the other hand, tends to feature mostly in highly recurrent expressions that are typical of the MSA newspaper or even news broadcast genre. Among the most robust configurations related to the use of *ǵā’a* is one in which the verb is used in a phrase such as “this/that *came* in the president’s speech”, where *ǵā’a* denotes the occurrence of a particular statement in a document or speech or any other mode of communication. This might also be a reflection on the functionality of imperfective and perfective morphological aspects; that is to say, given the fact that *atā* and *ǵā’a* have different aspectual preferences, this distributional data might provide further evidence that the two aspects (perfective vs. imperfective) have different functions rather than the naive idea that one is related to present/irrealis events while the other is related to events that are completed. This aspectual split is also why we need to combine findings from both quantitative and qualitative analyses in our interpretation of the usage of a synonymous set of lexical items. As far as the entire set of MSA COME verbs is concerned, the polytomous logistic regression analysis, in particular, helped us understand the degree of overlap in the usage of the COME verbs. More specifically, certain contexts can admit two or all four verbs, while other contexts demand a categorical choice. What I found particularly useful about examining the individual sentences – for which each verb receives a certain probability estimate – is the fact that it is also possible to examine the constructional features these sentences were coded for and which were included in the logistic regression model.

As for the MSA GO verbs, I pointed out that the polytomous logistic regression analysis did not yield the same level of satisfactory results as it did with COME verbs.

This may reinforce the idea that *dahaba*, *maḍā*, and *rāḥa* have much less in common than COME verbs do. Indeed, one of these verbs is highly grammaticalized, with a lower count of motion uses, while *dahaba* and *maḍā* each seem to have strong and idiosyncratic collocational and constructional affinities. This may still not be the only reason for not acquiring more distinguishing characteristics and significant results with the logistic regression model. It is possible that the GO data frame needed to be coded for additional variables – such as collocates – or that some of the existing variables needed to be further broken down into more specific categories. Overall, the statistical analyses conducted on the GO data frame did succeed in teasing apart the differences between the three verbs and zeroing in on the most prototypical uses of these verbs. Based on both the quantitative and qualitative analyses of GO verbs, I can still conclude that the expression of a GO event and the extended uses, be they metaphorical or grammaticalized, are encompassed by those three verbs rather than one verb, which also applies to the four COME verbs. The existence of more than one COME or GO verb in MSA does not signal redundancy and extravagance in the lexical system. Rather, it emphasizes the fact that these (deictic) motion events are undoubtedly complex and that the different COME and GO verbs in MSA capture different aspects and angles of these basic motion event frames.

It is important, at this point, to reflect on the selection of variables used to annotate the corpus returns. In Chapter 2, I listed the morphological, syntactic, and semantic variables each contextualized verb use was coded for. The rich inflectional paradigm of the MSA verb naturally took up a large proportion of the overall variable set. As for the syntactic variables that I opted for, these included transitive uses of the verb (in the case of COME), interrogation, polarity, and the inclusion of PP, locative adverb phrases, and adverbial phrases. The semantic variables varied from subject semantic category to the inclusion of constructions indicating goal, source, manner, path, purposive, comitative, temporal extent, and degree (frequency). The beauty of

constructing such a data frame is that it is always possible to expand on the number of variables selected for coding the verbs. We could, for instance, include SV order, clause type (main vs. subordinate) as well as specific collocates. Needless to say, the 1,500 GO + 2,000 COME corpus hits had to be manually coded for each of these variables (and some other variables that were not included in the quantitative analysis).⁴⁴ The existing variable set is in itself an expansion on a pilot study examining the *Behavioral profiles* of MSA GO verbs (Abdulrahim, ms.), which had already yielded satisfactory results. Indeed the current variable set proved to be quite efficient for both the quantitative and qualitative analysis.

Nevertheless, the variable set has its limitations in that it cannot account for all the variability and contextual richness of constructions that host such highly frequent verbs. This is another rationale for pairing the quantitative analysis with case-by-case inspection of actual verb uses. One of the limitations of the variable set pertains to the fact that certain variables represent general categories, such as SUBJECT SEMANTIC CATEGORY. The different levels within this variable referred to generic concepts such as ‘human’, ‘artifact/object’, ‘activity’, ‘notion’, etc. Without examining individual corpus returns, I would not have been able to identify the specialized semantic category of numerous subject collocations of the verb *qadima* – such as ‘visitors’, ‘immigrants’, ‘refugees’, ‘delegate’, etc – which all have been dubbed rather schematically as ‘human’ in the data frame. This may raise the question of how much detail should be reflected in the different types and levels of variables. For instance, if we further narrow down a

⁴⁴ It is important to remind the reader again that the larger the variable set the more data points need to be included for statistical power. Some statisticians have claimed that for each predictor variables there has to be a minimum of 10 observations, i.e. data points, (Howel, 2002). According to Sheskin, however, these assumptions have not yet been supported by empirical evidence (2007:1439). Counting all levels of variables as individual predictor variables, I ended up with an overall of 80+ predictor variables (for 500 data points per verb).

certain category (e.g. SUBJECT SEMANTIC CATEGORY), we run the risk of over-specifying and overfitting the data and, hence, not being able to identify high-level general patterns.

The variable set did, however, help us zero in on low-level generalizations regarding the usage of MSA GO and COME verbs. One such generalization pertains to the inflectional patterns of the verb and the skewed distribution of verb forms across the many levels of the verb inflectional paradigm. Going back to the notion of a ‘lemma’, Newman and Rice (2004, 2006a, 2006b) and Rice and Newman (2005, 2008) have extensively explored the ‘inflected form’ as opposed to the abstract ‘lemma’ with regard to collocational distribution and semantic pattern. The main premise of this series of papers is to emphasize the importance of studying the behaviour of the inflected form. Their manifesto regarding what they refer to as ‘inflectional islands’ is that “syntactic (constructional), semantic, and collocational properties tend to inhere in individual inflections of a lexical items in a register specific manner. These properties may not extend across all inflections (the paradigm) to characterize the lemma as a whole” (Newman and Rice, 2006b). In the following discussion, I argue that the analysis of GO and COME verbs, presented in the previous chapters, lends support to the notion of inflectional islands.

In light of the discussion of GO and COME verbs in Chapters 3 and 5, the first part of the quantitative analysis, as well as the hierarchical configural frequency analysis, revealed a dramatically skewed distribution of inflectional properties in the MSA verb conjugation paradigm. The first univariate analysis in the HCFA discussion showed that the most recurring inflected forms for the three GO verbs and the four COME verbs regarding morphological aspect, number, person and gender were PERFECTIVE, SINGULAR, 3RD, MASCULINE. In contrast, the least recurring inflectional forms include IMPERATIVE, JUSSIVE, SUBJUNCTIVE, DUAL, PLURAL, 1ST, 2ND, and NIL gender (related to 1st person) for GO verbs, and a very similar set for COME verbs. Following these

inflectional categories was an examination of robust patterns of TAM marking related to each verb. I found among COME verbs, for instance, that both *ǧā'a* and *qadima* have strong preferences towards appearing in the PERFECTIVE. *Ḥaḍara* also showed preference towards such TAM marking but was also inflected numerous times for the IMPERFECTIVE, SUBJUNCTIVE and JUSSIVE. *Atā* on the other hand had an overwhelming preference for appearing in the IMPERFECTIVE, in addition to multiple instances of being inflected in the SUBJUNCTIVE and JUSSIVE. Such item-specific distributional patterns for morphological aspect as well as the other inflectional categories offer insight into verb usage in the (mostly) newspaper writing in Modern Standard Arabic. As I have mentioned earlier, traditional grammars of Arabic place the verb conjugational paradigm in the spotlight and emphasize the need for learners of Arabic to master the entirety or full potentiality of these inflected forms. Little attention is paid to the most frequent (vs. the least frequent) inflected forms of a lexical item within a paradigm or the particular uses associated with an individual inflected form. These distributional skewes, in my opinion, should be a fundamental part of learning a new verb in MSA. In Chapter 8, I suggest a *corpus-based* (as opposed to *corpus-illustrated*) dictionary entry for the verb *atā*, in which the different verb-related uses are listed starting with the most frequent constructions hosting *atā*, as has been observed in the data retrieved from ArabiCorpus.

7.3 Using informed introspection

As stated earlier, in order to fully understand lexical behavior, both general distribution patterns as well as individual instances of usage need to be examined. The qualitative analyses presented in Chapters 4 and 4 have provided further insight into verb usage that univariate and multivariate analyses alone could not have achieved.

Going back to the notion of the *inflected construction*, the qualitative analysis highlighted expressions or sub-senses of the verb that would admit only one inflected

form of a particular verb. To illustrate, I will re-introduce some of the GO and COME examples I discussed in Chapters 4 and 6. Starting with the GO verb *maḍā*, I pointed out that this verb tends to express the passage of time in most of the cases studied. The expression in (1) would only admit the verb in the 3rd singular masculine perfective to mean ‘in the past’. In (2), the same form (admitting either the masculine or feminine form) would be the translation equivalent of the English adverb ‘ago’.⁴⁵

(1) فيما مضى

fī=mā maḍā
 LOC=RP *maḍā*.PERF.3SG.M
 in what went
 ‘in the past’

(2) منذ ثلاثة أشهر مضت

munḍu ʔalāʔat ašhur maḍat
 ADV three months *maḍā*.PERF.3SG.F
 since three months went
 ‘since 3 months’

Another obvious example of strong associations between usage and an inflected form is found in the grammaticalization of *rāḥa*. In the vast majority of uses in which the verb acts as an aspectual marker the verb is typically inflected in the imperfective, as in (3). On the other hand, the collocational pattern denoting ‘go and come’ and which *rāḥa* can be part of, only hosts the verb in the imperfective form, as in (4).

(3) وراحت منذ ذلك الحين تضيق الخناق عليه اكثر واكثر

wa=rāḥat munḍu ḍālīka al=ḥīn tuḍayyīq
 CONJ=*rāḥa*.PERF.3SG.F ADV DEM ART=time tighten.IMPF.3SG.F
 and went since that the time tightening

al=ḥināq ʔal=ayh akṭar wa=akṭar
 ART=grip LOC=CL.3S.M more CONJ=more
 the grip on him more and more
 ‘And since then it [i.e. Washington] kept tightening the grip on him more and more’

⁴⁵ Note that the English *ago* is etymologically related to the verb *go*: “ago (adj.), early 14c., shortened form of Old English *agan*, *agone* “departed, passed away,” past participle of an obsolete verb *ago* “to go forth,” formed from *a-* “away.”” (www.etymoline.com).

(4) يروح ويجيء

yarūḥ *wa=yağī'*
rāḥa.IMPF.3SG.M CONJ=*ğā'a*.PERF.3SG.M
 goes and comes
 'goes and comes'

As far as COME verbs are concerned, I discussed the semi-suppletive relationship between *atā* and *ğā'a* which I further illustrated with the sentence in (5). In this sentence, I pointed out to the fact that two COME verbs are used in the same sentence in which one COME verb refers to a past event (*ğā'a*) while the other (*atā*) refers to a habitual event. Such example supports the claim I made earlier that the existence of several verbs corresponding to a COME event in Modern Standard Arabic is not a mere reflection of stylistic options available to the speaker, but that each verb has its specified role in the lexico-syntactic system of Arabic.

(5) المستوردات الاسرائيلية من الفستق جاءت وتأتي من ايران

al=mustawradāt *al='isrā'iliyya* *min* *al=fustuq* *ğā'at*
 ART=imports ART=Israeli ABL ART=pistachios *ğā'a*.PERF.3SG.F
 the imports the Israeli of the pistachios came

wa=ta'tī *min* *irān*
 CONJ=*atā*.PERF.3SG.F ABL Iran
 and comes from Iran
 'The Israeli pistachio imports came and still come from Iran'

Similarly, a 3rd singular masculine imperfective is the only form of the verb admitted in the collocational phrase in (6), in which *fīmā ya'tī* can be roughly translated as 'in the following'.

(6) فيما يأتي نص المقابلة مع مهديّة بن بلة

fi=mā *ya'tī* *naṣ* *al=muqābala-ti* *ma'a* *mahdiyya bin balla*
 LOC=RP *atā*.IMPF.3SG.M text ART=interview-GEN COM Mahdiyya Bin Balla
 in what comes text of the interview with Mahdiyya Bin Balla
 'The text of the interview with Madiyya Bin Balla is in the following'

Finally, I discussed the highly transitive use of *ḥaḍara* to express the sub-sense 'to attend'. The syntactic object in these instances of verb use is always expressed lexically.

The only instance of an object clitic attached to *ḥaḍara* in the come data frame was when the verb is being used to mean ‘to recall/to get inspired’, as in (7).

(7) يحضرني كلام زياد مرة في زمن بعيد

yaḥḍuru-ni	kalām-un	qāla-hu	ziyād	marra
<i>ḥaḍara</i> .IMPF.3SG.M-CL.1SG	talk-NOM	say.PERF.3SG.M-CL.3SG.M	Ziyad	once
comes to me	talk	said it	Ziyad	once

fi	zaman	baʿīd
LOC	time	far
in	time	far

‘I recall something that Ziyad had once said long time ago’

In addition to specific inflected forms, low-level generalizations (as per Newman and Rice) can also include subject-verb collocations. The univariate analyses in Chapters 3 and 5, as well as the subsequent multivariate analyses, have explored this particular aspect of verb use for the seven verbs under study. Recall from the quantitative and qualitative analyses of GO and COME verbs that each verb had a different profile in terms of the subject semantic category each verb typically collocated with. For instance, I pointed out in my discussion on COME verbs that *ḥaḍara* and *qadima* mostly collocate with subjects denoting HUMAN or a GROUP of humans. *Atā* and *ḡāʿa*, however, have a completely different profile in terms of the semantic category of the syntactic subject. Both verbs seem to collocate with subjects denoting HUMAN, ACTIVITY, NOTION, COMMUNICATION, GROUP, EVENT, as well as the use of a demonstrative as in ‘this/that came. Nevertheless, we can still find undeniable differences in the relative frequencies of the subject collocates of *atā* and *ḡāʿa*.

Lastly, an additional low-level pattern that I referred to in my discussion of the grammaticalized uses of GO verbs concerns the semantic category of the main verb in the inceptive/continuous constructions with *maḍā* and *rāḥa*. I argued that the imperfective verb in the main verb position, following *maḍā*, tends to be filled by a verb that denotes a speech or noise-making event or a verb of perception, as in the repeated example in (8).

These semantic preferences do not hold for *rāḥa* when used as an inceptive/continuous marker.

(9) ومضى الاثنان ينشدان معا

wa=maḍā	al='iṭnān	ynšidān	ma'an
CONJ= <i>maḍā</i> .PERF.3SG.M	ART=two	sing.AP.3DUAL.M	together
and went	the two	singing	together

'And the two went on singing together'

The examples given in (1)-(9) of skewed inflectional distributions, subject collocational preferences, and serial verb construction properties are in line with Newman and Rice's suggestions that a true examination of lexical behaviour requires going beyond the abstract and overly idealized 'lemma' form to examining actual inflected forms in context and patterns of collocation. It is through these lower-level patterns that we can actually make accurate generalizations. To use Ronald Langacker's words, we need to "find the hierarchy of lower-level structures... [that] specify the actual array of subcases and specific instances that support and give rise to the higher-level generalizations" (1991: 281-282).

In general, this study has offered a comprehensive method for the examination of basic verbs in Modern Standard Arabic, which considered defining aspects of lexical usage: frequency, construction, convention, and so forth. In the following final chapter of this dissertation, I will move on to discussing the different implications of this study on typological research on basic verbs, as well as the practical applications of the analysis presented in the earlier chapters. I will also discuss different directions for future research related to motion verbs or other basic verbs in the different varieties of Arabic.

Chapter Eight

Implications, future research, and practical applications

This study presented extensive corpus-based quantitative and qualitative analyses of seven motion verbs in a highly literary and formal variety of Arabic, MSA. In §8.1, I will sketch out the implications of this study and contributions it offers to the general typological research on motion verbs and, particularly, on languages with more than one GO and one COME lexeme. The discussion on the use of GO and COME verbs in MSA unavoidably brings into question the role of motion verbs in the vernacular dialects of Arabic. In §8.2, I will, therefore, discuss the necessity of conducting similar large-scale, corpus-based research on motion verbs in the spoken varieties of Arabic. I will also discuss the importance of examining non-finite forms of the seven motion verbs, namely the verbal noun forms and the active participle forms in §8.3. In §8.4, I will talk about current and future research that combines corpus findings with psycholinguistic methods as a means of tapping into native speakers' intuition about lexical usage. Finally, in §8.5, I propose three types of bilingual dictionary entries – corpus-illustrated, general sub-sense frequency-based, and corpus-based.

8.1 Typological research on other languages with multiple GO and COME lexemes

The literature on GO and COME lexemes cross-linguistically is extensive, yet very few studies have focused on languages in which GO and/or COME motion events are expressed by more than one lexical item. Among such studies is Wilkins and Hill's (1995) work on Mparntwe Arrernte – a Pama-Nyungan language – and Botne's (2005)

work on Chindali – an Eastern Bantu language.⁴⁶ Following Talmy (1985), Wilkins and Hill proposed a four-way semantic feature system for the analysis of the two GO verbs and the two COME verbs in this language which includes: (i) an anchoring reference point (i.e. source or goal of the motion event), (ii) a directional component related to the orientation of the path of motion, (iii) a description of the path, and lastly (iv) the deictic aspect of the motion event. Unsurprisingly, their analysis showed that GO and COME verbs behave in different ways based on these semantic features. Wilkins and Hill observed that “[t]he verbs that depict COME and GO scenes cross-linguistically vary in their base semantics to such a degree that there is no useful sense in which they may be considered universal notions or lexical universals” (1995:214), and that “[t]here are languages in which the GO verb is not inherently deictic” (ibid.:215). The authors, therefore, concluded that GO and COME verbs do not typically express a two-element sub-system in a language.

Along these lines, Botne (2005) examined a dozen Chindali GO and COME verbs with the main objective of providing further evidence to support Wilkins and Hill’s claims. In addition to the four semantic features mentioned above, Botne observed that two additional semantic features are needed in order to fully comprehend the (deictic) motion scene: (i) salience, i.e. “prominence of some element encoded by verb”, and (ii) coincidence, i.e. “prominence of contact between motion figure and place” (2005:45). Needless to say, Botne found that the Chindali GO and COME verbs tend to differ on the basis of different combinations of these six semantic features, for the purpose of achieving a specific pragmatic effect.

It is important to note that both Wilkins and Hill’s (1995) and Botne’s (2005) studies used elicitation tasks as the source of data. Botne, however, added spoken data

⁴⁶ Wilkins and Hill also examined Longgu – an Oceanic language – in this paper which has a general TRAVEL verb, to which a ‘hither/thither’ morpheme is added to indicate a ‘go thither’ and ‘come’ motion event.

from two speakers of Chindali and a number of contextualized written instances of the verbs under study extracted from narratives that were produced by native speakers of the language. Consequently, the motion events discussed in these papers were restricted to event construals involving motion of human agents. In a way, the analysis I presented in the previous chapters on the one hand confirms the claims of Wilkins and Hill, yet on the other hand it is incommensurate with findings reported in these two studies. For one thing, the language I am reporting on is highly formal and literary and the vast majority of verb uses in the MSA sub-section of ArabiCorpus are far from being representative of everyday motion event construals. Modern Standard Arabic, therefore, stands in sharp contrast with Mparntwe Arrernte and Chindali in terms of the functional and pragmatic aspects of language use. It is therefore necessary to turn our attention next to the colloquial register of Arabic. There is no doubt that the usage of GO and COME verbs across the different spoken dialects will further contribute to our understanding of how this pair of motion verbs behaves cross-linguistically (e.g. in terms of grammaticalization, metaphorical and idiomatic uses). I will further discuss the importance of examining colloquial uses of GO and COME in the following sub-section.

Moreover, Wilkins and Hill and Botne reported on languages in which a basic (deictic), physical motion event can be encoded in more than one lexeme. While this is also true of the GO and COME verbs in MSA, motion event scenes constitute a marginal aspect of the use of some of these verbs, such as the highly prolific *atā* and *ǧā'a* as well as the specialized *maḍa* verb that mostly refers to the passage of time. In the case of the physical motion usage of these verbs, I concluded that physical motion event construals are the only cases in which all verbs can be used interchangeably. Granted, the particular lexical semantics of a verb may highlight one aspect of the motion event over another, for which Botne (2005) refers to as the notion of 'saliency'. In (1), for instance, I pointed out that due to the fact that the COME verb *ḥaḍara* is most likely to indicate 'to attend' or

express the state of ‘being present’, the sentence in (11) may highlight presence at the end point rather than the motion event itself.

(1) إن السلطات الأمنية قد حضرت ليلة الإثنين الماضي ومزقت خيمة المعتصمات

<i>inna</i>	<i>al=suluṭāt</i>	<i>al=amniyya</i>	<i>qad</i>	<i>ḥaḍarat</i>	<i>laylat</i>
TOP	ART=authorities	ART=security.ADJ	DM	ḥaḍara .PERF.3SG.F	night
that	the authorities	the security-related	had	come	night

<i>al=’iṭnayn-i</i>	<i>al=māḍi</i>	<i>wa=mazzaqat</i>	<i>ḥaymat</i>	<i>al=mu’taṣimāt</i>
ART=Monday-GEN	ART=past	CONJ=tear.down.PERF.3SG.F	tent	ART=protestors.FEM
of the Monday	the last	and tore down	tent	the female protestors

‘That security forces had come last Monday night and tore down the female protestors tent’

For the most part, the differences that exist among the GO and COME verbs in MSA are more strongly manifested in the idiomatic and metaphorical uses of these verbs. Similar to Wilkins and Hill’s (1995) findings, the Arabic motion verbs do differ in their “base semantics”, but mostly in their usage with GO and COME verbs in other languages. In the case of MSA GO and COME verbs, as opposed to *go* and *come* in English, this was most apparent, as we can expect, in situations in which the verbs are used idiomatically or in particular collocational phrases. For instance, the English phrasal usage *come off* or the collocational usage *go back* are not translatable into MSA with the help of a COME or GO verb. Similarly, the phrasal use of *dahaba bi-* ‘go with’ (or Lit. ‘to take something somewhere’) or the idiomatic usage *atā ’ala ḍikr* ‘come over the mention of’ (or Lit. ‘to mention’) certainly are not construed by employing a (deictic) motion verb in English. These language-specific extensions that encompass collocational patterns, idiomatic and metaphorical uses, and grammaticalized functions all point to the fact that the GO and COME event scenes are conceptually complex and therefore, to echo Wilkins and Hill, should not be regarded as universal concepts.

I limited myself in this study to (deictic) motion verbs that are ‘basic’ in the sense that they do not encode information about the path of motion, the manner of the motion, or other semantic prosodic information. This process of selection resulted in excluding

other possible MSA COME and GO verbs. Among these verbs were the (COME) verb *aqbala*, which I left out on the basis of its encoding of positive semantic prosody, and the (GO) verb *wallā*, on the basis of encoding negative semantic prosody (as well as the overall low frequency of both verbs in the corpus). In addition to these two verbs, I had to leave out the motion verb *sāra* due to the fact that it encodes manner and/or path (mostly referring to ‘walking’, or motion along the ground). The existence of such lexical items in the language, again, confirms the notion that a motion event – deictic or not – is multifaceted and complex and is heavily grounded in the collocational context that includes not only the nature of the theme (moving entity), or the starting or ending points, but also the manner of the motion (e.g. on foot), the temporal properties of the event (e.g. punctual or durative), the nature of the path (e.g. ground or non-specific), the attitude of the speaker (e.g. negative or positive), etc. It is by means of such factors that languages vary greatly in the amount of information they pack into or leave out of a single lexical item.

In retrospect, it is rather naive to subsume a simple lexical item or a number of different terms under an ill-defined conceptual label such as GO or COME. It is not a question of how many lexical items a language uses to express deictic motion events. Rather, it is a question of (i) how the semantic load of a deictic motion event is distributed lexically (e.g. *dahaba* ‘to go’ and *maḍā* ‘to go by’); (ii) what aspects of this event are encoded for within a single lexical item (e.g. *atā* ‘to come’ and *haḍara* ‘to come/to attend/be present’); (iii) what kind of construals (e.g. physical or metaphorical) are most likely to associate with the lexical items; and (iv) how this lexical item is used conventionally in the language (e.g. constructional patterns, idioms, collocates, etc).

8.2 Future research

8.2.1 Spoken varieties of Arabic

The discussion in §8.1 emphasized the need to shift our attention from the modern formal variety of Arabic to the vernacular dialects spoken across the Arab world. As far as motion verbs are concerned, I mentioned in Chapter 1 that the spoken dialects of Arabic tend to rely on only a single GO and a single COME lexeme. In most dialects, the verb denoting the COME event is most likely to be a phonologically modified form of *ǧā'a*, e.g. *yeh* in certain Arabian Gulf dialects, *ža* in Moroccan Arabic, *geh* in Egyptian Arabic, and *iža* in Levantine dialects. As for GO verbs, most dialects may use the verb *rāh*, while others may employ a different verb, such as *mša* (from *mašā* 'to walk') in Moroccan Arabic.

I have already discussed the fact that MSA motion verbs are mostly used metaphorically, which is a reflection of the functionality of MSA – that of being a formal register used to communicate news, knowledge, intellectual and literary discourse, etc., and as such is a language that is learned systematically and is only used by literate speakers of Arabic. Vernacular dialects, on the other hand, are the acquired first languages of Arabic speakers and they are the varieties Arab speakers use to communicate with one another in their everyday lives. The examination of (deictic) motion verbs across the spoken Arabic dialects, therefore, would undoubtedly shed light on verb uses that pertain to everyday life experiences. In a specialized Bahraini Arabic mini-corpus of COME verb uses, I found that around 80% of the verb uses (out of 174 contextualized uses of *yeh*) depict some form of physical motion, as in (2) and (3). The remaining sentences included a more figurative and idiomatic use of the verb, as in the examples given in (4)-(6).

(2) انتي لو يالغ أحد جدي

<i>enti</i>	<i>law</i>	<i>yā=la=č</i>	<i>aḥḥad</i>	<i>čidī</i>
PP	COND	<i>ǧā'a</i> .PERF.3SG.M=ALL=CL.2SG.F	someone	ADV
you	if	came to you	someone	like this

'If someone like that comes to you...'

(3) قال شوفي لين ياج الثعلب

<i>gāl</i>	<i>šūfī</i>	<i>len</i>	<i>yā-č</i>	<i>el=ta'lab</i>
say.PERF.3SG.M	see.IMPR.2SG.F	ADV	<i>ḡā'a</i> .PERF.3SG.M-CL.2SG.F	ART=fox
said	see	when	come to you	the fox

'He said "Look! When the fox comes to you..."'

(4) الفكرة يات على بالي

<i>el=fikra</i>	<i>yāt</i>	<i>'ala</i>	<i>bāl-ey</i>
ART=idea	<i>atā</i> .PERF.3SG.F	LOC	mind-CL.1SG.GEN
the idea	came	on	my mind

'The idea came to my mind'

(5) جفتي الزولية اللي تبيح كلش فلات

<i>čifti</i>	<i>el=zuliyye</i>	<i>elli</i>	<i>tyī-č</i>	<i>killiš</i>	<i>flāt</i>
see.PERF.2SG.F	ART=carpet	RP	<i>atā</i> .IMPF.3SG.F-CL.2SG.F	INTENS	flat
see	the carpet	that	comes to you	very	flat

'See that carpet that comes very flat...'

(6) شيبيلج أينشتاين؟

<i>ši=iyyi=l=ač</i>	<i>'ayneštayn</i>
Q= <i>atā</i> .IMPF.3SG.M=ALL=CL.2SG.F	Einstein
what comes to you	Einstein

'Is Einstein a relative of yours?'

With the exception of (4) – that depicts the motion of 'ideation' – most of the metaphorical uses of *yeh*, such as the ones in (5) and (6), do not correspond to the metaphorical uses of its MSA cognate *ḡā'a*.

A more fulsome study of GO and COME verbs should, therefore, aim to contrast the uses and sub-senses of all (deictic) motion verbs not only between a certain spoken dialect and MSA, but also among different spoken dialects of Arabic. A multivariate analysis, such as this one conducted on MSA verbs, should aim to differentiate between (i) physical and figurative uses of these verbs per dialect, (ii) constructional, collocational, and idiomatic features related to the use of these verbs, as well as (iii) highlight grammatical functions of these verbs in order to arrive at a better understanding of the use of motion verbs in the different varieties of Arabic.⁴⁷ The examination of colloquial GO and COME verbs should also shed light on other lexical items in each dialect that may assume the role of a deictic motion event verb in particular constructions, e.g.

⁴⁷ See Brustad (2000) for a description of the syntactic functions and properties of verbs of motion across Moroccan, Egyptian, Kuwaiti and Syrian dialects of Arabic.

the verb *meša* ‘to walk (away)’ in some dialects, or the verb *rawwah* ‘to leave/to go home’ in Egyptian Arabic.

Needless to say, a proper quantitative study of the Arabic dialects requires the existence of sizeable and readily accessible corpora of spoken Arabic. Fortunately, there is a growing interest among Arabic linguists nowadays to construct corpora of specific dialects, such as Tunisian Arabic (developed by Karen McNeil at Georgetown University and Miled Faiza at University of Virginia), the Egyptian Colloquial component of ArabiCorpus (mostly from online sources), and the Gulf and Iraqi Arabic conversational telephone speech corpora available on the Linguistic Data Consortium website.

8.2.2 Non-finite forms related to the GO and COME verbs

As stated earlier, the analysis presented in this dissertation focuses uniquely on finite uses of selected MSA verb and excludes other forms systematically related to the verb such as verbal nouns (VN) and active participles (AP).⁴⁸ I should add that the suppletive imperative *ta'āl* ‘come!’ has also not been part of the COME data discussed throughout Chapters 3-6. The decision to exclude these forms was based on practical reasons, i.e. minimizing the amount of variability in the constructional elements that were selected for the manual annotation of the corpus data. Further explorations of GO and COME in MSA would benefit from examining the morphosyntactic and collocational behavior of VN and AP forms in particular.

Almašdar or verbal noun, as the label suggests, is the nominalized form of the verb it corresponds to. According to Ryding (2005), the VN can assume different functions in MSA such as acting as the gerund or the infinitive, as in (7), or taking part in *iḍāfa* or what is referred to as the genitive construction, as in (8).

⁴⁸ Being ‘systematically’ related to the MSA verbs means that they share the same trilateral (or quadrilateral) root as the verb, but have a different vocalic pattern and exhibit different morphosyntactic and functional features.

- (7) لم أستطع الذهاب
lam astaṭi' al=ḍahāb
 NEG be.able.JUSS.1SG ART=*ḍahaba*.VN
 did not be able the going
 'I could not go'
- (8) قبل مجيء الرئيس
qabla maḡṭ' al=ra'īs
 ADV *ḡā'a*.VN ART=president
 before coming the president
 'before the arrival of the president'

Ism alfā'il or the active participle according to Ryding can either signal the doer of the action (similar to English in *-er* or *-or* suffixes), as in (9). It may also be used as an adjective, as in (10), or as a predicate adjective, as in (11).

- (9) وصل المسافرون
waṣala al=musāfirūn
 arrive.PERF.3SG.M ART=travel.AP.3PL.M
 arrived the travellers
 'the travelers have arrived'
- (10) الأسبوع القادم
al=usbū' al=qādim
 ART=week ART=*qadima*.AP.3SG.M
 the week the coming
 'next week'
- (11) صرخ قائلاً
ṣaraḥa qā'ilan
 scream.PERF.3SG.M say.AP.3SG.M
 screamed saying
 'he screamed saying...'

Table 3 lists the VN and AP forms related to the seven motion verbs discussed in this study. Notice that the AP forms of *rāḥa* and *ḡā'a* do not exist in MSA, which is an interesting fact since these two verbs (as I have mentioned in §8.2) are the only GO and COME verbs found in numerous vernacular dialects of Arabic and the AP forms of these two verbs are highly frequent words.

TABLE 1. Verbal noun and Active participle derivations of the seven motion verbs.

	verb root	verbal noun (VN)	active participle (AP)
GO	<i>ḡahaba</i>	ḡahāb	ḡāhib
	<i>maḡā</i>	muḡiy	māḡin
	<i>rāḡa</i>	rawāḡ	
COME	<i>atā</i>	ityān	‘ātin
	<i>ḡā’a</i>	maḡī’	
	<i>ḡaḡara</i>	ḡuḡūr	ḡāḡir
	<i>qadima</i>	quḡūm	qāḡim

An examination of the morphosyntactic, semantic, and collocational properties of the VN and AP forms, listed in Table 1, provides further motivation for treating these derivations as constructions in their own right rather than merely being forms that are morphologically related to the verb. For instance, the previous discussion highlighted the fact that *atā* ‘come’ has multiple deictic and non-deictic uses that relate to physical or figurative motion. The VN form of this verb, however, has a very restricted set of uses and is considered a low frequency lexical item (occurring around 630 times throughout the 146,000,000-word ArabiCorpus in all its subsections). According to the contextualized uses of *ityān* or *atā.VN* ‘coming’ in ArabiCorpus, the sub-senses of this nominal form include: (i) having intercourse, (ii) doing a good or a bad deed, (iii) consulting with a fortune teller or a sorcerer, and (iv) motion to location (very archaic usage). On the other hand, *maḡī’* ‘coming’, the nominal form of *ḡā’a* ‘come’, features in numerous physical and figurative construals similar to the variety of constructions the verb *ḡā’a* ‘come’ can participate in.

As far as the APs are concerned, the AP form of *ḡā’a* is not used in MSA at all. Instead, we have the AP derivations of *atā* (‘*ātin*) and *qadima* (*qāḡim*) that can be used to indicate gerundive ‘coming’. What is interesting is that both the VN and AP forms of *qadima* collocate most frequently with a subject denoting TIME, while the verb itself was not found to collocate with a temporal subject throughout the 500 corpus returns that were examined. More specifically, the AP form *qāḡim* is commonly used as an adjective

modifying a noun such as *usbū* ‘week’ or *sana* ‘year’ to mean ‘next week’ and ‘next year’ as in (12) and (13).

- (12) الأسبوع القادم
al=usbū *al=qādim*
 ART=week ART=*qadima*.AP.3SG.M
 the week the coming
 ‘next week’

- (13) السنة القادمة

al=sana *al=qādira*
 ART=year ART=*qadima*.AP.3SG.F
 the year the coming
 ‘next year’

While a number of bilingual dictionaries, such as the Hans Weir Arabic-English dictionary (1994), acknowledge the different functional and collocational properties of VNs and APs with the verb stem, many monolingual Arabic dictionaries follow the traditional lexicographic system of listing all possible derivations related to a verb form under the head word (being the trilateral or quadrilateral root) and not paying any extra attention to these forms. Figure 1 is an excerpt of the dictionary entry for the verb *qadima* from the monolingual dictionary *Mu’ğam Al-Luğa Al-‘arabiyya Al-Mu‘āšira* (2008), accompanied by my transliteration/translation of certain parts.

FIGURE 1. Excerpt from a monolingual Arabic dictionary for the verb *qadima*

٣٩٦٥ - ق د م

قَدِمَ يَقْدُمُ، قَدَمًا وَقُدُومًا، فَهُوَ قَادِمٌ، وَالْمَفْعُولُ مَقْدُومٌ

• قَدِمَ الْقَوْمَ: سَبَقَهُمْ فَصَارَ قَدَامَهُمْ، أَي أَمَامَهُمْ "هُوَ يَقْدُمُ الطَّابُورَ - «يَقْدُمُ قَوْمَهُ، يَوْمَ الْعَيْتَةِ فَأَزْرَدَهُمُ الْقَارَ» .

قَدِمٌ يَقْدُمُ، قَدِيمًا وَقَدَامَةً، فَهُوَ قَدِيمٌ

• قَدِمَ الثَّوْبُ: مَضَى عَلَى وُجُودِهِ زَمَنٌ طَوِيلٌ، عَكْسَهُ حَدَثَ "قَدِمَ بِهِ الْعَهْدُ - اللهُ هُوَ الْمَعْبُودُ مِنْذُ الْقَدَمِ".

قَدِمَ / قَدِمَ إِلَى / قَدِمَ مِنْ يَقْدُمُ، قُدُومًا، فَهُوَ قَادِمٌ، وَالْمَفْعُولُ مَقْدُومٌ

• قَدِمَ فَلَانٌ الْمَدِينَةَ / قَدِمَ فَلَانٌ إِلَى الْمَدِينَةِ: دَخَلَهَا، جَاءَ إِلَيْهَا، حَلَّ بِهَا "تَفْتَحُ الْأَزْهَارُ مَعَ قَدُومِ الرَّبِيعِ".

qadima yaqḍum [qadima.IMPF], **quḍuman** [qadima.ADV] **quḍūman** [qadima.VN], so he is coming [*qādim*, *qadima*.AP], and the passive participle is *maqḍūm*.

qadima / qadima to / qadima from yaqḍum [qadima.IMPF], **quḍūman** [qadima.VN], so he is coming [*qādim*, *qadima*.AP], and the passive participle is *maqḍūm*

qadima sb ART.city.ACC / qadima sb to the city: He entered it, came (*ğā'a* to it), arrived at it, "flowers blossom with the arrival [*quḍūm*, *qadima*.VN] of spring"

Note the listing of verbal, nominal, and participial forms together under the verb head word. Needless to say, there is no mention of the specific uses of the AP form throughout the entire dictionary entry. Note, as well, the mismatch between the general constructional meaning “*qadima* sb to the city”, and the example sentence given – “flowers blossom with the arrival [*qadima.VN*] of spring”. In this sentence, the subject is not HUMAN as implied by the construction entry, neither is the verb form similar to that in the example, since *qadima* is being used nominally in the sample sentence.

Clearly, there is an intricate relationship between the meanings and use of the finite verb forms and their non-finite VN and AP counterparts, as well as among the individual VN and AP forms for GO and COME verbs. What we have observed as a typical behavior of, for instance, the verb *atā* regarding its sub-senses and its collocational patterns hardly applies to its nominal form. This same discrepancy between finite and non-finite forms applies to the remaining motion verbs examined in this dissertation. Such mismatch between the uses of a finite verbal form and its nominal or participial derivations provides even more evidence for the claims made by Newman and Rice (2004, 2006a; among others) that each inflected form has a life of its own that does not necessarily resemble that of the other members in an inflectional or derivational paradigm.

8.2.3 Psycholinguistic experiments

There has been a growing tendency in recent years to compare or combine results obtained from corpus-based analyses with those obtained from psycholinguistic experiments that tap into native speakers’ intuition about linguistic uses (e.g. Gries, 2002; Rosenbauch, 2003; Featherston, 2005; Bresnan, 2006; Arppe and Järvikivi, 2007). It is true that corpora give us insight into natural language use and frequency distributions of, for instance, lexical items, collocates, and constructions, more than would be achieved by

relying on introspection alone. Nevertheless, corpora cannot “[account] accurately for rare but possible linguistic phenomena, and therefore, corpus data cannot be our only source of empirical evidence” (Arppe and Järvikivi, 2007:3).

The studies listed above combined corpus data with experimental evidence obtained from forced-choice and/or acceptability ratings tasks. Bresnan (2006), for instance, employs a forced-choice scalar rating experiment that roughly combines both forced-choice and acceptability rating tasks as a means of comparing native speakers’ intuition about the use of the English dative alternations with probability estimates drawn from *logistic regression analysis* of corpus data. Bresnan found that the probability estimates calculated by the model did indeed correlate with native speakers’ judgments. Similarly, Arppe and Järvikivi (2007) examined correlations between quantitative accounts of two THINK verbs in Finnish – *miettä* and *pohtia* – and native speakers’ intuitions about the contexts of use that permit these two verbs. They hypothesized that a forced-choice task and an acceptability ratings task would tap into different linguistic processes: production and introspection. The authors conducted a Fisher’s exact test to measure the significance of differences in feature distribution among the two verbs.⁴⁹ As far as the forced-choice task is concerned, Arppe and Järvikivi found that the participants’ preferences in the forced-choice task correlated to a great deal with results obtained from the quantitative analysis. That is, the presence or the absence of a feature highly associated with a certain verb strongly correlated with the forced-choice preferences made by the participants. As for the acceptability ratings task, the results supported both findings obtained from their monivariate analysis as well as the forced-choice task. However, it also showed that some infrequent cases of lexeme + feature combinations were still considered acceptable to a certain degree by the Finnish participants. Based on

⁴⁹ The constructional features examined in this study included PERSON, NUMBER, and SUBJECT SEMANTIC CATEGORY.

these results, the authors presented a fine-tuned general hypothesis that concerns the relationship between corpus-based results and native speakers' preferences and judgments which they summarized as follows:

“frequency (N.B. in relative terms) entails acceptability ..., and unacceptability entails infrequency ... On the other hand, acceptability can entail either frequency or infrequency ... Therefore, most importantly we cannot state that infrequency correlates, without exception, with unacceptability ... nor that acceptability correlates with frequency ... Furthermore, with regards to choice in corpora or in experimental judgments, frequency correlates with preference ..., as does infrequency with dispreference ...” (2007:25)

The corpus-based quantitative results presented in this dissertation on GO and COME verbs in MSA can also be subjected to experimental validation via forced-choice and acceptability rating tasks. What would be of most interest is the extent to which the observed usage patterns of GO and COME verbs overlap with the intuitions of native speakers of a modern Arabic dialect about MSA. Following in the spirit of Bresnan (2006), the probabilities estimated by the *polytomous logistic regression* model can be compared with selectional preferences made by literate native speakers of Arabic and/or their judgments of the plausibility of placing a verb in a certain context of use. Recall that the overall accuracy rate of the model constructed for COME verbs was 84.5% (i.e. the rate at which the actual observed verb receives the highest probability estimate). I conducted a pilot study (Abdulrahim and Arppe, in preparation) in which we compared the probability estimates of COME verbs per a selection of contexts (from the original data frame) with literate Bahraini speakers' preferences of verb usage in a forced-choice task. The 30 speakers of Bahraini Arabic had to read 100 sentences that included 50 COME sentences and 50 fillers and choose the most appropriate verb that fit the context. In the selection of the stimuli for the task, the 50 COME sentences represented a wide range of the contextual richness that is apparent in the corpus data and the diverse probability distributions: (i) near-categorical preferences of only one COME verb, (ii) high probability

estimates for the occurrence of two particular verbs, and (iii) approximately equal probability estimates for all four COME verbs. The calculated correlation between probability estimates and verb selection was very strong ($r_{pearson} = 0.747$, $p < 2.2e-16$). In (1) and (2), I list a couple of stimulus sentences paired with the individual probability estimates calculated per COME verb, as well as the overall selection proportions made by all 30 participants (in the column ‘choice %’). This is a sample of sentences for which we found a correspondence between the distribution of probability estimates across the four verbs and the overall rates of verb selection by the Bahraini speakers. We can see, for instance, that the choice of *ġā’a* by the native speakers of Arabic, for the context in (14), mirrors the probability of occurrence of this verb estimated by the model. Similarly, the non-categorical choices made by the speakers for the verb in (15) also correspond to the equi-probable estimates calculated by the model for this particular context.

(14) **وجاء** تعيين أبو رحمة في وقت توجه اتهامات بالفساد إلى السلطة الفلسطينية

wa=ġā’a *taīn’y* *abū raĥma* *fī* *waqt* *tuwaġġah*
 CONJ=*ġā’a*.PERF.3SG.M appoint.VN Abu Rahma LOC time direct.PASS.IMP.3SG.F
 and came appointment Abu Rahma in time is directd

fī=h *ittihamāt* *bi=l=fasād* *ilā* *al=sulṭa* *al=falastīniyya*
 LOC=CL.3SG.M accusations INST=ART=corruption ALL ART=authority ART=Palestinian
 in it accusations of the corruption to the authority the Palestinian
 ‘The appointment of Abu Rahma took place at a time when the Palestinian authority is facing accusations of corruption’

	<i>atā</i>	<i>ġā’a</i>	<i>ḥaḍara</i>	<i>qadima</i>
probability estimate	0.074	0.922	0.003	0.000
choice %	0.000	0.933	0.000	0.067

(15) وبعد ذلك **قدمت** إلى المملكة العربية السعودية والتقيت ببعض الاخوة

wa=ba’da *ḍālīka* *qadimtu* *ilā* *al=mamlaka* *al=’arabīyya* *al=su’udīyya*
 CONJ=ADV DEM *qadima*.PERF.1SG. ALL ART=kingdom ART=Arab ART=Saudi
 and after that I came to the Kingdom the Arab the Saudi

wa=iltaqaytu *bi=ba’ḍ* *al=’ihwa*
 CONJ=meet.PERF.1SG COM=some ART=brothers
 and I met with some the brother
 ‘And after that I came to Saudi Arabia and I met with some brothers’

	<i>atā</i>	<i>ǧā'a</i>	<i>ḥaḍara</i>	<i>qadima</i>
probability estimate	0.15	0.33	0.14	0.38
choice %	0.167	0.433	0.067	0.333

On the one hand, this cross-evidential comparison between corpus, quantitative analysis and forced-choice results sheds light on the extent to which literate Arabic speakers (from a certain dialectal background) have internalized the rules and conventions of lexical uses in the written standard register, MSA. On the other hand it also lends support to the selection of constructional elements that were included in the *polytomous logistic regression* model, as I have alluded to in the previous chapter.

A follow-up study might examine literate native speakers' judgments of the acceptability of using COME verbs in a wide variety of contexts that have been deemed by the model as either exclusively admitting one verb or allowing two or more verbs to be used in a particular construction. As per Arppe and Järvikivi (2007), such experimental technique would help us identify the contexts of use that may typically favor one verb over the rest, but which may still be deemed felicitous by speakers of the language when another verb is used. It is uncertain, however, that one would get the same results as Arppe and Järvikivi since, as mentioned earlier, there appears to be only a small window of overlap within the uses COME and GO verbs in MSA. Therefore, it is likely that the acceptability ratings would still reflect the probability estimates and the forced-choice selections.

The same experiment could also be replicated with learners of MSA – possibly of higher levels of language acquisition if we restricted ourselves to stimulus items found in corpora. The objectives of conducting research with learners are twofold: (i) gain knowledge about their acquisition of the use of these verbs (which can be compared to that of the native speakers); and (ii) the results would, more or less, provide an evaluation

of the teaching material. That is to say, one can find out the extent to which the students are learning the constructional and idiomatic properties of a lexical item.

8.3 Practical applications: Suggestions for usage-based dictionary entries

The lexicographic treatments of the highly frequent motion verbs studied here as exhibited in bilingual and, mostly, monolingual dictionaries range from almost adequate to completely mis-representative descriptions of the major and minor senses of these verbs. Many monolingual dictionaries, such as *Al-Munğid fi Al-Luğa wa Al-A'lam* (2005), follow a traditional and highly ideological system of lexical representation whereby archaic uses of a lexical item are foregrounded and little attention is paid to more contemporary uses. In fact, it is the expressed intention of the authors of this dictionary to follow this lexicographic tradition:

“tamma iğtahadnā fi đikr al-luğati al-‘um allatī yantamī ilayha al-kalām al-daxīl ... fawuffiqnā ilā ɧarīqa taħfaz rūh al-lluğa wa turā’ī taqālīd al-ma’āğim” [then we exercised some efforts in describing the mother tongue, which foreign manner of speaking (i.e. borrowed words and contemporary usage) associates with ... then we succeeded in finding a way in which we could preserve the essence of the language and observe the lexicographic conventions (2005:i).] [translation mine]

It is worth noting that such a dictionary does not differentiate between Classical Arabic and Modern Standard Arabic, which is also an expression of the prescriptivist view of the Arabic language as being one timeless, pure form – *al-luğa al-fuṣṣḩā* ‘the most eloquent language’ – rather than being divided into old versus modern Arabic, or written versus spoken.

On the other hand, the dictionaries that do list the different contemporary sub-senses of the verbs (in addition to archaic one) and some collocational patterns, such as the Hans Weir Arabic-English Dictionary (1995) are impoverished in terms of example sentences that show how a lexical item is actually used. Alternatively, they may

exemplify a certain word's usage by providing an irrelevant, constructed sentence (as we saw with *Mu'ğam Al-Luġa Al-'arabiyya Al-Mu'āšira* (2008) in Figure 1). For a learner of Arabic, such inconsistent and abstracted accounts of a lexical item render the task of consulting a dictionary rather frustrating.

There have been many attempts in the past to provide a frequency dictionary of Arabic, listing the most common words in mostly (MSA), such as Abduh's (1979) *Al-mufradat al-šā'i'a fi al-luġa al-'arabiyya* (Frequently Used Arabic Vocabulary) and Kholoughli's (1991) *Lexique fondamentale de l'Arabe standard modern* (Basic Lexicon on Modern Standard Arabic), among others. According to Buckwalter and Parkinson (2011:2), these dictionaries are “seriously outdated and are based on corpora that are considered very small by today's standards”. Buckwalter and Parkinson's (2011) *A Frequency Dictionary of Arabic: Core Vocabulary for Learners* provides a fresh and modern account of 5,000 of the most frequent lexical items found in a corpus of 30 million words. Roughly 90% of this corpus comes from written sources encompassing different registers (e.g. newspaper, literature), while the remaining 10% consists of spoken Arabic produced by speakers from different regions in the Arab world. There are so many benefits from consulting with this dictionary, either as a learner or as a native-speaking researcher. One of the procedures followed in constructing this dictionary and which, in my opinion, deserves the highest praise, is the separation between a verb form lemma (V) and its nominal and participial derivations – verbal nouns and active and passive participles – since “the lexicographic description warranted such distinctions” (2011:5). This dictionary nevertheless falls short in terms of providing stereotypical examples of lexical uses, in addition to the fact that all the inflected verb forms (PERFECTIVE, IMPERFECTIVE, JUSSIVE, SUBJUNCTIVE, etc.) are all subsumed under one

lemmatized form – 3rd singular masculine.⁵⁰ The previous quantitative and qualitative accounts of MSA GO and COME verbs have repeatedly shown that different sub-senses and uses cluster around individual members of a verb's inflectional paradigm.

In the following sub-sections I will illustrate with the use of the COME verb *atā* different types of dictionary entries that can serve different purposes and which can fill the gaps formed in modern lexicographic accounts. In 8.2.1, I present a *corpus-illustrated* dictionary entry that elaborates on the existing (bilingual) dictionary entries of the verb by supplementing relevant corpus examples for each verb sub-sense or usage. In 8.2.2, I present a minimalist sub-sense frequency-based dictionary entry that orders the verb entries according to the frequency of occurrence of the overall general usage (physical, metaphorical, etc.). In 8.2.3, I suggest a *usage-based* dictionary entry for *atā* that is directly based on the quantitative analysis of the verb presented in Chapter 5. All three sample entries are restricted to finite verb forms, and do not include derived non-finite forms such as VNs and APs.

8.3.1 Corpus-illustrated dictionary entry of *atā*

The following sample dictionary entry for *atā* elaborates on the typical main dictionary entries specified across different bilingual dictionaries (Arabic-English) and is modeled to some extent after a usage-based dictionary such as the Collins COBUILD English Dictionary (Sinclair et al., 1995). Note that the different senses and uses of the verbs are ordered from the most to the least literal uses of the verb.

Such dictionary entry is quite elaborate in terms of the instances of verb uses exemplified in each entry. It is also quite extensive with respect to teasing apart the different uses of a lexical item. Unlike the entry for *atā* that I suggest in 8.3.3, this corpus-illustrated entry does not take sub-sense and usage frequency into account, which

⁵⁰ See examples of GO and COME entries from this dictionary in Appendix C.

is considered useful information for both the language learner and the researcher. It also does not provide a gloss for each individual instance of verb use which makes it rather inaccessible for a non-Arabic speaking researcher.

FIGURE 2. Sample of corpus-illustrated dictionary entry of *atā*

أتى 1 to come, to arrive			
أتى، يأتي / ata/ أتى			
1. أتى، يأتي إلى مكان / ~ من مكان			
atā (PERFECTIVE), ya'tī (IMPERFECTIVE)			
When a person or a thing comes to a particular place, especially to a place where you are, they move there.			
<i>When will daddy come?...</i>	متى يأتي بابا؟		
<i>He used to come to my parties...</i>	كان يأتي إلى حفلاتي		
<i>The merchandise that comes directly to Syria...</i>	البضائع التي تأتي مباشرة إلى سوريا		
<i>And they come to this place from all over the world...</i>	وهم يأتون إلى هذا المكان من سائر أنحاء العالم		
<i>Justice will come to us...</i>	ستأتينا العدالة		
<i>I heard that voice that came to me...</i>	سمعت ذلك الصوت الذي أتاني		
<i>Didn't come out of nowhere...</i>	لم يأت من فراغ		
<i>Then comes the role of organizations...</i>	ثم يأتي دور المؤسسات		
2. أتى، يأتي لـ / ~ لكي			
atā (PERFECTIVE), ya'tī (IMPERFECTIVE)			
To come or to move to a specific location for a purpose.			
<i>Seekers of knowledge used to come to our universities to learn...</i>	وكان طلاب العلم يأتون إلى جامعاتنا ليتعلموا		
<i>He only came to argue...</i>	انما أتى كي يناقش		
3. أتى، يأتي الوقت			
atā (PERFECTIVE), ya'tī (IMPERFECTIVE)			
Used to talk about when a certain period of time comes or arrives.			
<i>Ramadan comes/starts in a few hours...</i>	ساعات ويأتينا رمضان		
4. يأتي في المركز / ~ على رأس			
ya'tī (IMPERFECTIVE)			
Used to talk about achieving a certain ranking.			
		<i>The most important crops are wheat and cotton...</i>	يأتي على رأس هذه المحاصيل القمح والقطن
		<i>Saudi Arabia comes in first place...</i>	المملكة العربية السعودية تأتي في المركز الأول
		5. يأتي ضمن / ~ في إطار / ~ لـ / ~ بـ	
		ya'tī (IMPERFECTIVE)	
		Used to talk about an event or an activity happening or taking place in a particular context, and/or for a particular purpose.	
		<i>The project takes place fulfilling one of several decisions...</i>	المشروع يأتي ضمن عدة قرارات
		<i>The symposium takes place as part of a series of meetings...</i>	الندوة تأتي في إطار سلسلة اللقاءات
		<i>Chirac's visit took place to reassure the Lebanese...</i>	أنت زيارة شيراك لتطمئن اللبنانيين
		<i>Criticizing history comes out of the desire to revive it...</i>	نقد التاريخ يأتي بدافع الحرص على أحياءه
		<i>Our participations in this tournament comes as an expression of our solidarity and support for Lebanon...</i>	مشاركتنا في هذه الدورة تأتي تعبيراً عن تضامننا ودعمنا للبنان
		6. يأتي إلى الشيء / ~ إلى الموضوع	
		ya'tī (IMPERFECTIVE)	
		Used to talk about tackling or approaching a certain topic or issue.	
		<i>When we approach the issue of 'quality' we find that it's a matter of administration...</i>	عندما نأتي إلى الجودة نجد أنها مشكلة إدارة
		<i>And now let's discuss the law decreed two months ago...</i>	ونأتي للقانون الصادر منذ شهرين
		7. فيما يأتي / كما يأتي	
		ya'tī (IMPERFECTIVE)	
		In what comes; (as) in the following.	

In the following is the script of the interview...
 The script of the speech is as following...
 فيما يأتي نص المقابلة
 وكان نص الخطاب كما يأتي

And that governments come and go...
 وأن الحكومات تأتي وتذهب

أتى 2 to do something
 أتى، يأتي /ata/ أتى

1. أتى، يأتي الشيء: فعله

atā (PERFECTIVE), ya'tī (IMPERFECTIVE)
 Lit: to come something. This is a transitive usage of the verb that refers to the act of doing something or committing a crime, sin, etc.

And the heinous sins he has committed...
 والمعاصي الفظيعة التي أتاها
 The women did something fascinating...
 أتت النساء حركة عجيبة

أتى 3 Phrasal uses
 أتى، يأتي /ata/ أتى

1. أتى بـ

atā (PERFECTIVE), ya'tī (IMPERFECTIVE)
 Lit: to come with something. Used to mean 'to bring' something or 'come up with' something.

Where did he come up with all this money?...
 من أين أتى بكل هذه الأموال؟
 He will not come up with anything new...
 لن يأتي بجديد

2. أتى على

atā (PERFECTIVE)
 Lit: to come over something. Used to mean 'to destroy', 'to demolish', 'to consume', etc.

The flames destroyed all the machines...
 أتت ألسنة اللهب على جميع الأجهزة
 This loss consumed the bank's capital...
 هذه الخسارة أتت على رأسمال البنك

8. أتى على ذكر / ~ على لسان

atā (PERFECTIVE)
 Lit: to come over the mention of something/someone; to come over someone's tongue. Used to mean 'to mention' or 'to talk about'.

They were surprised that an old writer who say something like that...
 استغربوا أن هذا الكلام يأتي على لسان كاتب عجوز
 The Turkish president did not mention the Middle East...
 لم يأت الرئيس التركي على ذكر الشرق الأوسط

10. أتى يفعل / أتى وفعل / يأتي ويفعل

atā (PERFECTIVE), ya'tī (IMPERFECTIVE)
 To come doing or in order to do something; to come and do something.

And he came to offer his services...
 وأتى يعرض خدماته
 All we hope for is that you come and knock on our door...
 كل أملنا أن تأتي وتطرق باب منزلنا
 Tourist groups come and buy Egyptian cottons...
 الأفواج السياحية تأتي وتشترى الأقطان المصرية

11. يأتي ويذهب

ya'tī (IMPERFECTIVE)
 To come and go; used to talk about frequent movement.

All these protests come and go for nothing...
 كل هذه التظاهرات تأتي وتذهب في أدرج الرياح

8.3.2 Sub-sense frequency-based dictionary entry of atā

In the following sample entry, only the main categories of atā sub-senses are highlighted: (i) COME: figurative motion, (ii) COME: physical motion, (iii) phrasal uses, and (iv) the transitive use: to do something. The frequency of the occurrence of these sub-senses (as per a sample of 500 corpus uses of the verb) are indicated by the filled circles (●). That is, if 3 out of the 10 circles are filled, then this means the frequency of this sub-

sense is $\approx 30\%$ for the 500 corpus returns. Examples from ArabiCorpus are used to illustrate the different sub-senses and uses of *atā*. Notice that even this dictionary entry is highly minimalist since it lumps together many constructional properties of the verb in one category. Nevertheless, it gives a general overview of the distribution of the different sub-senses/uses of the verb, and would in my opinion count as a useful entry for a frequency dictionary of the language.

FIGURE 3. Sample of sub-sense frequency-based dictionary entry of *atā*

<p>.....</p>	
<p>أتى 1 to come: figurative motion ●●●●●●●●○○</p>	
<p>أتى، يأتي /ata/ أتي atā (PERFECTIVE), ya'tī (IMPERFECTIVE) The verb can be used to talk about the figurative motion or the 'coming' of an entity (mostly non-human). It can refer to achieving a rank, taking place in a particular context, or mean 'following'. <i>Saudi Arabia comes in first place...</i></p>	<p><i>Tourist groups come and buy Egyptian cottons...</i> <i>الأفواج السياحية تأتي وتشترى الأقطان المصرية</i></p> <p><i>And that governments come and go...</i> <i>وأن الحكومات تأتي وتذهب</i></p>
<p>أتى 2 to come: physical motion ●●○○○○○○○○</p>	
<p>أتى، يأتي /ata/ أتي atā (PERFECTIVE), ya'tī (IMPERFECTIVE) The verb can be used to talk about the physical motion or the 'coming' of mostly humans or concrete objects. The coming event can also be purposeful. <i>And they come to this place from all over the world...</i> <i>Seekers of knowledge used to come to our universities to learn...</i></p>	<p><i>المملكة العربية السعودية تأتي في المركز الأول</i> <i>المشروع يأتي ضمن عدة قرارات</i> <i>فيما يأتي نص المقابلة</i></p> <p><i>Where did he come up with all this money?...</i> <i>The flames destroyed all the machines...</i></p> <p><i>من أين أتى بكل هذه الأموال؟</i> <i>أنت السنة الالهب على جميع الأجهزة</i></p>
<p>أتى 3 phrasal uses ●○○○○○○○○○○</p>	
<p>أتى، يأتي /ata/ أتي atā (PERFECTIVE), ya'tī (IMPERFECTIVE) The verb can be used in combination with the preposition <i>bi-</i> to mean 'to bring', or with the preposition '<i>alā</i>' to mean 'to destroy/demolish'. <i>Where did he come up with all this money?...</i> <i>The flames destroyed all the machines...</i></p>	<p><i>أتى 4</i> to do something ●○○○○○○○○○○</p> <p>أتى، يأتي /ata/ أتي atā (PERFECTIVE), ya'tī (IMPERFECTIVE) <i>Lit.</i>: to come something. This is a transitive usage of the verb that refers to the act of doing something or committing a crime, sin, etc. <i>And the heinous sins he has committed...</i></p> <p><i>والمعاصي الفظيعة التي أتاها</i></p>
<p>.....</p>	

8.3.3 Corpus-based dictionary entry of *atā*

The following dictionary entry draws specifically on the HCFA analysis of *atā* in terms of the larger constructional patterns associated with the verbs (i.e. subject collocates and phrasal semantic categories, such as GOAL, SOURCE, etc.), paired with the morphosyntactic features that characterize each construction. The verb sub-entries are listed starting with the most statistically robust constructions (indicated with ‘***’, ‘**’ and ‘*’). The examples used to illustrate each usage can be provided with a Romanized transliteration, a morphological gloss, and a literal gloss (lit.) to accompany the free or figurative gloss (fig.) to draw attention to the particular use of *atā* in a certain context. A linguistically glossed representation of the construction is also provided in bold.

This type of dictionary entry may be of some use to a language learner, but is mostly intended for researchers. Note that each verb entry refers to a specific lexico-syntactic frame, and the verb is fully inflected in each frame. The four-tier glosses make the dictionary entry more accessible to the non-Arabic speaking linguist by providing standardized linguistic glosses for multiple samples of verb usage. In addition, the distributional information provided in these entries sets the expectations for the language researcher as to how frequently s/he will encounter a specific construction in an MSA corpus.

FIGURE 4. Sample of corpus-based dictionary entry of *atā*

.....	
أتى	takes place in a particular context.
<p>1. يأتي في إطار/ يأتي في سياق/ يأتي في ظل / يأتي ضمن ***</p> <p>atā.IMPF in NOUN; atā.IMPF among</p> <p><i>Lit:</i> comes in the frame of, comes in the context of, comes in the shadow of; comes among. This usage of the verb is metaphorical. The subject argument is typically related to an ‘activity’ or an ‘event’ that happens or</p>	<p>المشروع يأتي ضمن عدة قرارات</p> <p><i>Al=mašrū' ya'tī ḍimna 'iddat qarārāt</i></p> <p>ART=project atā.IMPF.3SG.M ADV several decisions</p> <p><i>lit.</i> The project comes among several decisions</p> <p><i>fig.</i> The project takes place as part of several decisions</p>

الندوة تأتي في اطار سلسلة اللقاءات
Al=nadwa ta'tī fī iṭār silsilat al=liqā'āt
 ART=symposium **atā.IMPF.3SG.F** LOC frame
 chain ART=meetings

*** 2. يأتي من

atā.IMPF from

This usage is specific to an abstract notion coming from a certain source.

لعل الأمل باحياء روسيا يأتي من قدرتها على الغفران
La'alla al='amal bi='iḥyā' rūsya ya'tī min qudrati-hā 'alā al=ḡufrān

MOD ART=hope INST=reviving Russia
ata. IMPF.3SG.M LOC ability-CL.3SG.F.GEN
 LOC ART=forgiveness

lit. Maybe the hope with reviving Russian comes from its ability on the forgiveness
fig. It is possible that the hope for reviving Russia comes from its ability to forgive

3. أتى إلى، يأتي إلى / أتاه، يأتيه

atā.IMPF/PERF to; atā.IMPF/PERF-CL.ACC
Lit: comes or came to; comes or came it (intransitive). Used to talk about the physical motion of humans towards a destination. The destination can be a location, an activity, a notion, another human being, etc.

كان يأتي الى حفلاتي
Kāna ya'tī ilā ḥaflā-tī
 AUX **atā.IMPF.3SG.M** ALL parties-CL.1SG.GEN
lit. He was comes to my parties
fig. He used to come to my parties

كانت البداية صعبة اذ لم يأتيها أي زبون
Kānat al=bidāya ṣa'ba id lam ya'tī-hā ayyu zabūn
 Be.PERF.3SG.F ART=beginning hard ADV NEG
atā.JUSS.3SG.M-CL.3SG.F.ACC any customer
lit. Was the beginning hard since did not come to her any customer
fig. The beginning was hard since no customer came to her

4. يأتي في المركز / يأتي في المقدمة

atā.IMPF in the place; atā.IMPF in the forefront
 This expression is used to talk about a country, an organization achieving a certain ranking.

المملكة العربية السعودية تأتي في المركز الأول
Al=mamlaka al='arabiyya al=su'udiyya ta'tī fī al=markaz al='awwal
 ART=kingdom ART=Arab ART=Saudi
atā.IMPF.3SG.F LOC ART=place ART=first

lit. The kingdom the Arab the Saudi comes in the place the first
fig. Saudi Arabia comes in first place

5. يأتي بعد / يأتي قبل / يأتي في وقت / يأتي ذلك في الوقت الذي

atā.IMPF/PERF locative adverb of time; atā.IMPF/PERF time adverbial, e.g. at a time, at the time in which, etc.

Used to talk about a certain activity taking place at a certain period of time.

وتأتي الحملة في وقت تزداد الإنتقادات البريطانية لسياسة
 الإستيطان الإسرائيلية

Wa=ta'tī al=ḥamla fī waqt tazdād fi-tazdād fi-h al='intiḡādāt al=brīṭaniyya li=siyāsāt al='istītān al='isrā'iliyya

CONJ=atā.IMPF.3SG.F ART=campaign LOC time increase.IMPF.3SG.F LOC-CL.3SG.M
 ART=criticism ALL=policy ART=occupation
 *ART=Israeli

lit. And comes the campaign in time increases the criticism the British to policy the occupation the Israeli
fig. The campaign comes at a time when the British criticism of Israel's occupation policies has increased

وقد أتى هذا الهجوم بعد ساعات فقط من مقتل أربعة
 أمريكيين

Wa=qad atā hāḍa al=huḡūm ba'da sā'āt faḡaṭ min maḡtal arba 'at amrikiyyīn

CONJ=DM atā.PERF.3SG.F DEM ART=attack ADV hours only ABL murdering four Americans

lit. And already came this the attack after hours only from murdering four Americans
fig. This attack came only hours after the murder of four Americans

6. يأتي نتيجة / يأتي + حال

atā.IMPF as a result of; atā.IMPF manner adverbial

This expression is used to talk about an activity or a notion that happens as a result of a prior event or state.

هذا الإنجاز يأتي نتيجة دعم سمو أمير البلاد
Hāḍa al='inḡāz ya'tī natīḡat da'm summuw amīr al=bilād

DEM ART=achievement **atā.IMPF.3SG.M** result.ADV support.VN his.highness Emir ART=country

lit. This accomplishment comes resulting support his highness Emir the country

fig. This accomplishment is the result of the support shown by his highness the Emir

صحة اجراءات التحكيم التي يجب أن تأتي تطبيقاً للقانون ذاته

ṣiḥḥat iḡrā'āt al=taḥkīm allatī yaḡib an ta'tiya taḡbīqan li=l=qānūn dātih

Validity procedures ART=arbitration RP MOD TOP **atā.SUBJN.3SG.F** apply.ADV

ALL=ART=law itself

fig. The validity of the arbitration procedures that should come as an application to the same law

7. أتى، يأتي + حال

atā.IMPF/PERF manner adverbial

Used to talk about discourse in any medium – statement, letter, prose, response, etc. – which ‘comes’ in a certain manner, to fulfill a certain objective, or as a result of a previous state or event.

يأتي الرد سريعاً

Ya'tī al=radd sarī'an

atā.IMPF.3SG.M ART=response fast.ADV

lit. Comes the response quickly

fig. The response comes fast

العبارة أتت توكيداً لخطاب انتخابي سابق

Al= 'ibāra atat tawkīdan li=āiṭāb intiḡābi sābiq

ART=expression **atā.PERF.3SG.F** confirming

ALL=speech election.ADJ previous

lit. The expression came confirming to

speech electoral previous

fig. The expression came as a confirmation of a previous election speech

8. يأتي ليفعل، أتى ليفعل / يأتي لفعل، أتى لفعل / يأتي للشيء، أتى للشيء

atā.IMPF/PERF to do/ doing something;

atā.IMPF/PERF for something

This expression is used to talk about the physical motion of humans (to or towards the speaker or a certain location) to fulfill a particular purpose.

وكان طلاب العلم يأتون إلى جامعاتنا ليتعلموا

Wa=kāna ṭullāb al='ilm ya'tūn ilā ḡāmi 'ātina li=yata 'allamū

CONJ=AUX students ART=knowledge

atā.IMPF.3PL.M ALL universities-CL.1PL.GEN

PURP=learn.SUBJN.3PL.M

lit. And was students the knowledge come to our universities to learn

fig. Seekers of knowledge used to come to our universities to learn

أتى لإعطائنا دروساً

Atā li='i'tā'i-nā durus-an

atā.PERF.3SG.M PURP=give.VN-CL.1PL.ACC lessons-ACC

lit. he came to giving us lessons

fig. He came to give us lessons

هل أتى للاختبار أم لا داعي؟

Hal atī li=l='iḡtibār 'am lā dā'ī?

Q **atā.IMPF.1SG** ALL=ART=exam CONJ NEG need

lit. Should I come to the exam or no need?

fig. Should I come for the exam, or there is no need to?

9. أتى بـ، يأتي بـ

atā.IMPF/PERF with

This is a phrasal use of the verb atā that means ‘to bring’ or ‘to come up with’.

من أين أتى بكل هذه الأموال؟

Min ayna atā bi=kull hādīhi al=amwāl?

ABL Q **atā.PERF.3SG.M** COM=all DEM

ART=money

lit. From where he came with all this the money?

fig. Where did he come up with all this money from?

لن يأتي بجديد

Lan ya'tiya bi=ḡadīd

NEG **atā.SUBJN.3SG.M** COM=new

lit. He will not come with new

fig. He will not come up with anything new

8.4 Conclusion

This case study of seven motion verbs in Arabic has touched upon multiple major themes related to linguistic analysis and research on the Arabic language. In this final

section, I reflect on general implications that can be drawn from the corpus-based quantitative and qualitative analyses of GO and COME verbs in MSA and take stock of what we can learn from examining the behavior of highly frequent lexical items in context. The theoretical and methodological approaches I promote here apply to research on lexicosyntactic data in any language. In these concluding remarks, however, I would like to emphasize the implications of this type of research on the linguistic analysis of Arabic in particular.

The present analysis adheres to the assumption that there should be no separation between grammar and lexicon. This theoretical premise has been one of the primary motivations for my selection of a very specific set of lexical items in MSA. A great deal of linguistic analysis has largely focused on the syntactic structure (e.g. subject-verb order, agreement patterns, scope of negation, modification) without paying much attention to the lexical items occupying specific positions in a construction or to recurring collocations. In contrast, most previous analyses of MSA verbs have marginalized the role of the surrounding lexico-syntactic context.

The study I described in this dissertation moves away from such compartmentalized approaches to linguistic description by examining language at the level of *inflected construction*. As I have shown in the quantitative analysis of GO and COME verbs, each of the seven verbs showed different preferences for TAM marking; person, number, and gender agreement; collocational patterns; in addition to other lexico-semantic properties of the arguments and modifying phrases. Moreover, the overly discussed Arabic verb inflection paradigm was not fully realized in the numerous corpus returns inspected in this study. Verb-specific preferences provide further evidence for the need to examine individual lexical items in their respective contexts of use and to focus on *fully inflected forms* rather than idealize the lemmatized form and segregate it from its natural morphosyntactic profile.

Deictic motion verbs in MSA, therefore, proved to be an ideal case study to the constantly growing literature on constructionist approaches. Note that the idiosyncratic behavior of each motion verb was better detected through balancing the mega- and micro-analyses. The statistical analyses presented here revealed information about verb usage that manual inspection of a few sentences alone could not have achieved. As stated previously, frequency or distributional data provides us with a window into psycholinguistic processes and the structure of language. There is no surprise that the experimental data I referred to in §8.2.3 confirms the quantitative results on MSA COME verbs.

One of the main requirements for conducting the kind of study presented here is to recognize the legitimacy of language produced by native speakers and its reliability for linguistic analysis. Adhering to the notion of ‘grammaticality’ results in misrepresenting linguistic uses (in grammars and lexicographic descriptions) by dismissing a wide range of constructions as being ‘ungrammatical’ or rather ‘idiomatic’. Holding on to such biases makes corpus work on a language outside the mainstream. Unfortunately, many Arabic linguists are still resistant to the insights an Arabic corpus has to offer as they insist on continuing to examine constructed sentences or even theoretically possible sentences and structures that no speaker of Arabic actually says.

The misrepresentations of verb uses found in current dictionaries of Arabic are a direct reflection of such prescriptivist biases. Even though Modern Standard Arabic is considered the ‘higher’ variety of Arabic and the more prestigious variety for speakers compared to their respective vernacular dialects, it is rather ironic that even MSA is a target for linguistic prejudices. A number of current dictionaries (such as *Al-Munğid fi al-luğa wa al-a’lām*, referred to earlier) pride themselves in ‘preserving the purity of the language’, which implies that current uses that deviate from the Classical standard are marginalized or, rather, ignored. It is for these ideological ends that the larger Arabic-

speaking community does not make the distinction between a Classical variety and a Modern variety of standard Arabic. There is a single term to describe this ‘one’ language: *al-luġa al-fuṣḥā* ‘the most eloquent language’. Recall that in the analysis I presented in this dissertation, I have pointed out a number of times Classical Arabic structures currently used in specific registers (e.g. religious discourse or historical narratives), that deviate from the Modern Standard norms. Any accurate description of the language needs to acknowledge the evolutionary process a language has undergone. Luckily, many Arabists or Arab linguists acknowledge such discrepancies between archaic and contemporary usage and, therefore, we can now find specialized corpora that cater to researchers interested in examining, for example, Classical Arabic usage. The present study of MSA usage could be seen as a contribution on the way to a higher goal of adequately documenting usage patterns in all the written and spoken vernaculars.

Even though the Arabic variety I opted to describe in this study is Modern Standard Arabic, I have, nevertheless, made it clear throughout that I am a strong proponent of linguistic research on the vernacular dialects spoken across the Arab world. I believe that the level of analysis presented in this dissertation should also be applied to the study of motion verbs in the vernacular dialects, which can only be made possible by the availability of comprehensive spoken corpora.

Finally, I tried to present here a “standardized” treatment of the MSA verb by situating the description of Arabic motion verbs within the general linguistics literature on motion verbs cross-linguistically. By providing detailed, multi-tiered glosses for each sentence under scrutiny, I intend to make the data discussed here accessible to the wider linguistic community. This includes, for instance, language typologists interested in lexical and grammatical patterns, cognitive linguists interested in language-specific motivations for using motion verbs; corpus linguists taking on the challenge of exploring untapped corpora of non-European languages; quantitative linguists interested in the

application of statistical methods on various types of data; computational linguists using annotated data for machine learning purposes; and, last but not least, sociolinguists interested in describing formal and vernacular registers of a language. I hope that this study has succeeded in showing that Arabic, in its different varieties, has a lot to offer to all of these fields of linguistic analysis.

References

- Abduh, D. (1979). *Al-mufradat al-šā'i'a fi al-luġa al-'arabiyya [Frequently used Arabic vocabulary]*. Riyadh: University of Riyadh.
- Abdulrahim, D. (in preparation). Quantitative approaches to analyzing COME constructions in Modern Standard Arabic. To appear in A. Hardie, A. T. McEnrey (eds.), *Arabic Corpus Linguistics*.
- Abdulrahim, D., & Arppe, A. (in preparation). Converging linguistic evidence on two flavors of production: The synonymy of Arabic COME verbs.
- Abdulrahim, D., & A. Arppe. (2013) The different ways to COME in Modern Standard Arabic. *American Association for Corpus Linguistics (AACL)*. San Diego State University. San Diego, 18-20 January 2013.
- Abdulrahim, D., Newman, J., & Rice, S. (2011) Multifactorial methods for exploring contextual factors in the usage of Modern Standard Arabic COME verbs. *Workshop on Arabic Corpus Linguistics*, University of Lancaster, UK, 11-12 April, 2011.
- Abdulrahim, D., Newman, J., & S. Rice. (2011). COME verbs in Modern Standard Arabic: A corpus-based constructionist account of *atā*, *ġā'a*, *ḥadara*, and *qadima*. *25th Annual Symposium on Arabic Linguistics (ALS)*. University of Arizona, Tucson, 3-5 March 2011.
- Abdulrahim, D., & Newman, J. (2010). How many ways to GO in Arabic? A corpus-based approach to determining polysemy and synonymy of the verbs *dahaba*, *maḍā*, *rāha*, and *intalaqa*. *24th Annual Symposium on Arabic Linguistics (ALS): Arabic Linguistics Across Traditions*. University of Texas at Austin, 9-11 April 2010.
- Abu-Chacra, F. *Arabic: An essential grammar*. New York: Taylor & Francis.
- Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Hoboken: John Wiley & Sons, Hoboken.
- Arppe, A. (2007). Multi-variate methods in corpus-based lexicography: A study of synonymy in Finnish. *Proceedings from the Corpus Linguistics Conference (CL2007)*, Birmingham, United Kingdom.
- Arppe, A. (2008). Univariate, bivariate and multivariate methods in corpus-based lexicography - a study of synonyms. (PhD, University of Helsinki). *Publications of the Department of General Linguistics*, (44).
- Arppe, A. (2009). Linguistic choices vs. probabilities - how much and what can linguistic theory explain? *The Fruits of Empirical Linguistics (Proceedings of the International Conference on Linguistic Evidence, Tübingen, Germany, 31.1-2.2.2008)*, 1. (pp. 1-24).

- Arppe, A., & Järvikivi, J. (2007). Every method counts - combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory*, 3(2), 131-159.
- Baalbaki, R. (2008). *Al-Mawrid: A modern Arabic-English dictionary*. Beirut: Dar El-Ilm.
- Badawi, E. M. (1985). Educated spoken Arabic: A problem in teaching Arabic as a foreign language. In K. R. Jankowsky (Ed.), *Scientific and humanistic dimensions of language*. Washington: Georgetown University Press.
- Bateson, M. C. (1967; 2003). *Arabic language handbook*. Washington: Georgetown University Press.
- Beetson, A. F. L. (1970). *The Arabic language today*. London: Hutchinson University Library.
- Bender, L. (1990). Coming and going in Afrasian. *AAP*, 22, 19-49.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Binnick, R. (1971). Bring and come. *Linguistic Inquiry*, 2, 260:265.
- Blau, J. (1961). The importance of middle Arabic dialects for the history of Arabic. In U. Heyd (Ed.), *Scripta hierosolymitana IX: Studies in Islamic history and civilization*. Jerusalem: Magnes Press.
- Botne, R. (2005). Cognitive schemas and motion verbs: Coming and going in Chindali (Eastern Bantu). *Cognitive Linguistics*, 16(1), 43-80.
- Botne, R. (2006). Motion, time, and tense: On the grammaticalization of come and go to future markers in Bantu. *Studies in African Linguistics*, 35(2), 127-188.
- Bresnan, J. (2006). Is syntactic knowledge probabilistic? Experiments with the English dative alternation. *Pre-Proceedings of the International Conference on Linguistic Evidence*, Tübingen, Germany. (pp. 2-4).
- Bresnan, J. (2007). Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In S. Featherston, & W. Sternefeld (Eds.), *Roots: Linguistics in search of its evidential base*. Berlin: Mouton de Gruyter.
- Bybee, J. (1985). *Morphology: A study of the relation between meaning and form*. Amsterdam/Philadelphia: John Benjamins.
- Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford: Oxford University Press.

- Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82(4), 711-33.
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Church, K., Gale, W., Hanks, P., & Hindle, D. (1991). Using statistics in lexical analysis. In U. Zernik (Ed.), *Lexical acquisition: Exploiting on-line resources to build a lexicon* (pp. 115-164). Hillsdale: Lawrence Erlbaum Associates.
- Clark, E. (1974). Normal states and evaluative viewpoints. *Language*, 60(2), 316-332.
- Comrie, B. (1985). *Tense*. Cambridge: Cambridge University Press.
- Craig, C. (1991). Ways to go in Rama: A case study in polygrammaticalization. In C. Traugott, & B. Heine (Eds.), *Approaches to grammaticalization: Vol. 2 Focus on types of grammatical markers* (pp. 456-492). Amsterdam/Philadelphia: John Benjamins.
- Croft, W., & Cruse, A. D. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Di Meola, C. (2003). Non-deictic uses of the deictic motion verbs *kommen* and *gehen* in German. In F. Lenz (Ed.), *Deictic conceptualization of space, time and person* (pp. 41-68). Amsterdam/Philadelphia: John Benjamins.
- Di Meola, C. (1994). *'Kommen' und 'gehen'. eine kognitiv-linguistische untersuchung der polysemie deiktischer bewegungsverben*. Tübingen: Max Niemeyer Verlag.
- Divjak, D. S., & Gries, S. T. (2006). Ways of trying in Russian: Clustering and comparing behavioral profiles. *Corpus Linguistics and Linguistic Theory*, 2(1), 23-60.
- Elgibali, A. (1993). Stability and language variation in Arabic: Cairene and Kuwaiti dialects. In M. Eid, & C. Holes (Eds.), *Perspectives on Arabic linguistics V*. Amsterdam/Philadelphia: John Benjamins.
- El-Hassan, S. A. (1978). Educated spoken Arabic in Egypt and the Levant: A critical review of diglossia and related concepts. *Archivum Linguisticum*, 8, 112-32.
- Emanation, M. (1992). Chagga 'come' and 'go': Metaphor and the development of tense-aspect. *Studies in Language*, 16, 1-33.
- Featherston, S. (2005). The decathlon model of empirical syntax. In M. Reis, & S. Kepser (Eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives* (pp. 187-208). Berlin: Mouton de Gruyter.
- Ferguson, C. A. (1959). The Arabic koine. *Language*, 35(4), 616-630.

- Fillmore, C. (1966). Deictic categories in the semantic of 'come'. *Foundations of Language*, 2, 219-227.
- Fillmore, C. (1969). Types of lexical information. In F. Keifer (Ed.), *Studies in syntax and semantics* (pp. 109-137). Dordrecht: Reidel.
- Fillmore, C. (1970). Subjects, speakers and roles. *Synthese*, 21, 251-274.
- Fillmore, C. (1971). Toward a theory of deixis. *The PCCLLU Papers*, 3(4), 218:242.
- Fillmore, C. (1972). How to know whether you are coming or going. *Studies in Descriptive and Applied Linguistics*, 5, 3-17.
- Fillmore, C. (1977). Scenes-and-frames semantics. In A. Zampolli (Ed.), *Linguistic structures processing* (pp. 55-81). Amsterdam: North-Holland.
- Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. In J. R. Firth. 1968. *Selected Papers of J. R. Firth 1952-1959* (pp. 168-205). London: Longman.
- Fleischman, S. (1982). The past and the future: Are they coming or going? *Proceedings of the eighth annual meeting of the Berkeley linguistics society* (pp. 322-334). Berkeley, Cal: Berkeley Linguistics Society.
- Fück, J. (1955). *Arabiyya*. Paris: Marcel Didier.
- Gathercole, V. (1977). A study of the comings and goings of the speakers of four languages: Spanish, Japanese, English, and Turkish. *Kansas Working Papers in Linguistics*, 2, 61-94.
- Gathercole, V. (1978). Towards a universal for deictic verbs of motion. *Kansas Working Papers in Linguistics*, 3, 72-88.
- Goddard, C. (1997). The semantics of coming and going. *Pragmatics*, 7(2), 147-162.
- Gries, S. T. (2002). Evidence in linguistics: Three approaches to genitives in English. In R. M. Brend, W. J. Sullivan & A. R. Lommel (Eds.), *LACUS forum XXVIII: What constitutes evidence in linguistics?* (pp. 17-31). Fullerton: LACUS.
- Gries, S. T. (2003). *Multifactorial analysis in corpus linguistics: A study of particle placement*. London: Continuum.
- Gries, S. T. (2006). Corpus-based methods and cognitive semantics: The many meanings of to run. In S. T. Gries, & A. Stefanowitsch (Eds.), *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis*. Berlin/New York: Mouton de Gruyter.
- Gries, S. T. (2008). *Statistik für sprachwissenschaftler*. Göttingen: Vandenhoeck und Ruprecht.

- Gries, S. T. (2009). *Statistics for linguistics with R: A practical introduction*. Berlin: Mouton de Gruyter.
- Gries, S. T. (reprint of 2010h). Behavioral profiles: A fine-grained and quantitative approach in corpus-based lexical semantics. In G. Jarema, G. Libben & C. Westbury (Eds.), *Methodological and analytic frontiers in lexical research*. Amsterdam & Philadelphia: John Benjamins.
- Gries, S. T., & Berez, A. L. (2009a). In defense of corpus-based methods: A behavioral profile analysis of polysemous *get* in English. *Proceedings of the 24th Northwest Linguistics Conference*, 27. (pp. 157-166).
- Gries, S. T., & Divjak, D. S. (2009). Behavioral profiles: A corpus-based approach towards cognitive semantic analysis. In Vyvyan Evans, & S. S. Pourcel (Eds.), *New directions in cognitive linguistics* (pp. 57-75). Amsterdam/Philadelphia: John Benjamins.
- Gries, S. T., & Otani, N. (2010). Behavioral profiles: A corpus-based perspective on synonymy and antonymy. *ICAME Journal*, 34. 121-150.
- Hanks, P. (1996). Contextual dependency and lexical sets. *International Journal of Corpus Linguistics*, 1(1), 75-98.
- Harrell, F. E. (2001). *Regression modeling strategies: With applications to linear models, logistic regression and survival analysis*. New York: Springer-Verlag.
- Hary, B. (1996). The importance of the language continuum in Arabic multiglossia. In A. Elgibali (Ed.), *Understanding Arabic*. Cairo: American University in Cairo Press.
- Heine, B., & Kuteva, T. (2002). *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Holes, C. (2004). *Modern Arabic, revised edition: Modern Arabic: Structures, functions, and varieties*. Washington: Georgetown University Press.
- Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, Cal.: Duxbury Press.
- Ibn Manzur, Muammad ibn Mukarram. (1883). *Lisan al-arab*. Bullag, Egypt: Al-Matb'ah Al-Kubra Al-'Amiriyah.
- Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago/London: University of Chicago Press.
- Khouloughli, D. (1991). *Lexique fondamentale de l'Arabe standard moderne*. Paris: L'Harmattan.

- Krauth, J., & Lienert, G. A. (1995). *Die konfigurationsfrequenzanalyse (KFA) und ihre anwendung in Psychologie und medizin*. Weinheim: Beltz Psychologie Verlagsunion.
- Lakoff, G., & Johnson, M. (1979). *Toward an experientialist philosophy: The case from literal metaphor*. Unpublished MA,
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Lakoff, G. (1987). *Woman, fire and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Langacker, R. (1987). *Foundations of cognitive grammar, Vol. I: Theoretical prerequisites*. Stanford, Calif.: Stanford University Press.
- Langacker, R. (1991). *Foundations of cognitive grammar*. Stanford, Cal.: Stanford University Press.
- Lenz, F. (2003). Deictic conceptualization of space, time and person. In F. Lenz (Ed.), *Deictic conceptualization of space, time and person* (pp. vii-xiv). Amsterdam/Philadelphia: John Benjamins.
- Lichtenberk, F. (1991). Semantic change and heterosemy in grammaticalization. *Language*, 67(3), 475-509.
- Liu, M. (2000). Reciprocal marking with deictic verbs come and go in Mandarin. In Z. Frajzyngier (Ed.), *Reciprocals: Forms and functions* (pp. 123-132). Amsterdam/Philadelphia: John Benjamins.
- Lyons, J. (1977). *Semantics*. Cambridge: Cambridge University Press.
- McLoughlin, L. J. (1972). Towards a definition of modern standard Arabic. *Archivum Linguisticum: New Series*, 3, 57-73.
- Menard, S. (1995). *Applied logistic regression analysis*. Thousand Oaks: Sage Publications.
- Miller, A. M. (1986). The origin of the modern Arabic sedentary dialects: An evaluation of several theories. *Al-'Arabiyya*, 19, 47-74.
- Miller, G., & Johnson-Laird, P. N. (1977). *Language and perception*. Cambridge, Mass.: Harvard University Press.
- Mitchell, T. F. (1986). What is educated spoken Arabic? *International Journal of the Sociology of Language*, 61, 7-32.
- Monteil, V. (1960). *L'Arabe moderne*. Paris: Klincksieck.

- Newman, J. (2004). Motivating the uses of basic verbs: Linguistic and extralinguistic considerations. In G. Radden, & K. U. Panther (Eds.), *Studies in linguistic motivation* (pp. 193-218). Berlin/New York: Mouton de Gruyter.
- Newman, J. (2008). Aiming low in linguistics: Low-level generalizations in corpus-based research. *Keynote Address Presented to the 11th International Symposium on Chinese Languages and Linguistics (IsCLL-11)*, National Chiao Tung University, Hsinchu, Taiwan.
- Newman, J. (2011). Corpora and cognitive linguistics. *Brazilian Journal of Applied Linguistics*, 11(2), 521-559.
- Newman, J., & Lin, J. (2007). The purposefulness of going: A corpus-linguistic study. In J. Walinski, K. Kredens & S. Gozdz-Roszkowski (Eds.), *Corpora and ICT in language studies* (pp. 293-308). Lodz Studies in Language, Vol. 13. Frankfurt am Main: Peter Lang.
- Newman, J., & Rice, S. (2004). Patterns of usage for English sit, stand, and lie: A cognitively-inspired exploration in corpus linguistics. *Cognitive Linguistics*, 15, 351-396.
- Newman, J., & Rice, S. (2006a). Transitivity schemas of English EAT and DRINK in the BNC. In S. T. Gries, & A. Stefanowitsch (Eds.), *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis* (pp. 225-260). Berlin/New York: Mouton de Gruyter.
- Newman, J., & Rice, S. (2006b). English adjectival inflection: A radical Radical Construction Grammar approach. *Conceptual Structure, Discourse, and Language Conference*, University of California, San Diego.
- Newman, J., & Rice, S. (2008). Asymmetry in English multi-verb sequences: A corpus-based approach. In B. Lewandowska-Tomaszczyk (Ed.), *Asymmetric events: An interpretation* (pp. 3-22). Amsterdam/New York: John Benjamins.
- Parkinson, D. (1993). Knowing standard Arabic: Testing Egyptians' MSA abilities. In M. Eid, & C. Holes (Eds.), *Perspectives on Arabic linguistics V*. Amsterdam/Philadelphia: John Benjamins.
- Perrson, G. (1988). Homonymy, polysemy and heterosemy: The types of lexical ambiguity in English. *Symposium on Lexicography III: Proceedings of the Third International Symposium on Lexicography*, University of Copenhagen.
- Radden, G. (1996). Motion metaphorized: The case of coming and going. In E. H. Casad (Ed.), *Cognitive linguistics in the redwoods: The expansion of a new paradigm in linguistics* (pp. 423-458). Berlin/NY: Mouton de Gruyter.
- Rice, S., & Newman, J. (2004). Aspect in the making: A corpus-analysis of English aspect-marking prepositions. In S. Kemmer, & M. Achard (Eds.), *Language, culture and mind* (pp. 313-327). Stanford: CSLI Publications.

- Rice, S., & Newman, J. (2005). Inflectional islands. *Presentation at the 9th International Cognitive Linguistics Conference*, Yonsei University, Seoul, Korea.
- Rice, S., & Newman, J. (2008). Beyond the lemma: Inflection-specific constructions in English. *Presentation to American Association for Corpus Linguistics 2008*, Provo, Utah.
- Rauh, G. (1981). On coming and going in English and German. *Papers and Studies in Contrastive Linguistics*, 13, 53-68.
- Rosenbach, A. (2003). Aspects of iconicity and economy in the choice between the s-genitive and the of genitive in English. In G. Rohdenburg, & B. Mondorf (Eds.), *Determinants of grammatical variation in English* (pp. 379-411). Berlin/New York: Mouton de Gruyter.
- Rubin, A. (2005). *Studies in Semitic grammaticalization*. Cambridge, MA: Harvard Semitic Museum / Eisenbrauns.
- Ryding, K. C. (1990). *Formal spoken Arabic: Basic course*. Washington: Georgetown University Press.
- Ryding, K. C. (1991). Proficiency despite diglossia: A new approach for Arabic. *Modern Language Journal*, 75(2), 212-18.
- Ryding, K. C. (2005). *A reference grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.
- Sheskin, D. J. (2007). *Handbook of parametric and nonparametric statistical procedures* (4th ed.). Boca Raton, FL: Chapman and Hall/CRC.
- Sinha, A. K. (1972). On the deictic use of 'coming' and 'going' in Hindi. *Papers from the Eighth Regional Meeting of the Chicago Linguistic Society*,
- Slobin, D. I. (1996). Two ways to travel: Verbs of motion in English and Spanish. In M. S. Shibatani, & S. A. Thompson (Eds.), *Grammatical constructions: Their form and meaning* (pp. 195-220). Oxford: Clarendon Press.
- Stefanowitsch, A., & Rodhe, A. (2004). The goal bias in the encoding of motion events. In G. Radden, & K. U. Panther (Eds.), *Studies in linguistic motivation* (pp. 249-268). Berlin/New York: Mouton de Gruyter.
- Suzuki, R., & Shimodaira, H. (2006). Pvelust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22(12), 1540-1542.
- Talmy, L. (1975). Semantics and syntax of motion. In J. P. Kimball (Ed.), *Syntax and semantics* (pp. 181-238). New York: Academic Press.

- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description* (pp. 57-149). Cambridge: Cambridge University Press.
- Talmy, L. (1990). Fictive motion and change in language and cognition. Plenary address of the Conference of the International Pragmatics Association. Barcelona, Spain.
- Talmy, L. (1996). The windowing of attention in language. In S. A. Thompson (Ed.), *Grammatical constructions* (pp. 235-287). Oxford: Clarendon Press.
- Talmy, L. (2000). *Toward a cognitive semantics*. Cambridge: MIT Press.
- Testen, D. (1997). The suppletive imperative of Arabic "come". In M. Eid, & R. Ratcliffe (Eds.), *Perspectives on Arabic linguistics X: Papers from the tenth annual symposium on Arabic linguistics* (pp. 173-187). Amsterdam: John Benjamins.
- Theil, H. (1970). On the estimation of relationships involving qualitative variables. *The American Journal of Sociology*, 76(1), 103-154.
- Versteegh, C. H. M. (1997). *The Arabic language*. Edinburgh: Edinburgh University Press.
- Versteegh, C. H. M. (1984). *Pidginization and creolization: The case of Arabic*. Amsterdam/Philadelphia: John Benjamins.
- von Eye, A. (1990). *Introduction to configural frequency analysis: The search for types and antitypes in cross-classification*. Cambridge: Cambridge University Press.
- Wilkins, D. P., & Hill, D. (1995). When "go" means "come": Questioning the basicness of basic motion verbs. *Cognitive Linguistics*, 6(2/3), 209-260.

Dictionaries

- Al-Muhit Oxford Study Dictionary (English-Arabic)*. (1996). Beirut: Academia Lebanon.
- Al-Munjid fi al-lughah wa-al-alam* (2005). Beirut: Dar al-Mashriq.
- Arabic-English Dictionary: The Hans Wehr Dictionary of Modern Written Arabic* (1994). J. M. Cowan (ed.). Itaca, N.Y.: Spoken Language Services.
- Buckwalter, T., & Parkinson, D. (2011). *A frequency dictionary of Arabic: Core vocabulary for learners*. New York: Routledge.
- Collins COBUILD English Dictionary*. (1995). J. Sinclair (ed.). London: Harper Collins.
- Concise Oxford English-Arabic Dictionary* (1982). Oxford: Oxford University Press.

Omar, A. M. (2008). *Mu'ğam Al-Luġa Al-'arabiyya Al-Mu'āšira [Dictionary of modern contemporary Arabic]*. Cairo: 'ālam al-kutub.

R Statistical Script

Arppe, A. (2012). *Polytomous*.

Gries, S. Th. (2004). *HCFA 3.2 – A Program for hierarchical configural frequency analysis for R for windows*.

Gries, S. Th. (2009). *BehavioralProfiles 1.01. A program for R 2.7.1 and higher*.

Appendix A

Templatic verb forms in Arabic along with their general meaning associations
From Ryding (2005:429-437)

Form	Pattern	Meaning(s)
I	(PERF) C ₁ aC ₂ VC ₃ (IMPF) ya-C ₁ C ₂ VC ₃	The closest indicator of the meaning of the lexical root
II	(PERF) C ₁ aC ₂ C ₂ aC ₃ (IMPF) yu-C ₁ aC ₂ C ₂ iC ₃	Usually the causative of transitive FormI verbs, or adding a transitive meaning to non transitive FormI verbs Intensive or repeated action Denominative, derive verbs out of nouns
III	(PERF) C ₁ aaC ₂ aC ₃ (IMPF) yu-C ₁ aaC ₂ iC ₃	“Associative”: involves another person in the action Reciprocal/repeated/attempted actions
IV	(PERF) ‘aC ₁ C ₂ aC ₃ (IMPF) yu-C ₁ C ₂ iC ₃	Often causatives of FormI Transitivizes the intransitive FormI, and ditransitivizes the transitive FormI May have meanings similar to FormIV
V	(PERF) taC ₁ aC ₂ C ₂ aC ₃ (IMPF) ya-taC ₁ aC ₂ C ₂ aC ₃	May be the reflexive (medio-passive) or resultative form of FormII verb Gradual progress in activity or state and acquisition or imitation of a quality
VI	(PERF) taC ₁ aaC ₂ aC ₃ (IMPF) ya-taC ₁ aaC ₂ aC ₃	Usually the reciprocal of FormIII Gradual, continuous movement or increase in a quality Pretending of feigning something
VII	(PERF) inC ₁ aC ₂ aC ₃ (IMPF) ya-nC ₁ aC ₂ iC ₃	May be reflexive, resultative, passive or medio-passive Are claimed to express the ergative and unaccusative in Arabic
VIII	(PERF) iC ₁ taC ₂ aC ₃ (IMPF) ya-C ₁ taC ₂ iC ₃	May be reflexive or medio-passive, plus a wide range of meanings that are difficult to predict
IX	(PERF) iC ₁ C ₂ aC ₃ C ₃ (IMPF) ya-C ₁ C ₂ aC ₃ C ₃	Acquisition of color or physical trait (infrequent in MSA)
X	(PERF) istaC ₁ C ₂ aC ₃ (IMPF) ya-staC ₁ C ₂ iC ₃	May be requisitive or estimative The reflexive of FormIV
XI	(PERF) iC ₁ C ₂ aaC ₃ C ₃ (IMPF) ya-C ₁ C ₂ aaC ₃ C ₃	These forms are chiefly archaic or poetic in use
XII	(PERF) iC ₁ C ₂ awC ₂ aC ₃ (IMPF) ya-C ₁ C ₂ awC ₂ iC ₃	
XIII	(PERF) iC ₁ C ₂ awwaC ₃ (IMPF) ya-C ₁ C ₂ awwiC ₃	
XIV	(PERF) iC ₁ C ₂ anC ₃ aC ₃ (IMPF) ya-C ₁ C ₂ anC ₃ iC ₃	
XV	(PERF) iC ₁ C ₂ anC ₃ aa (IMPF) ya-C ₁ C ₂ anC ₃ ii	

Appendix B

The inflectional paradigm for a tri-consonantal root (KTB ‘to write’)

Form I *Sound Root*: كَتَبَ KaTaBa, يَكْتُبُ yaKTuBu

Adapted from Ryding (2005:475)

	Active	Active	Active	Active	Active	Passive	Passive	Passive	Passive
	Perfect	Imperfect	Imperfect	Imperfect	Imperfect	Perfect	Imperfect	Imperfect	Imperfect
		Indicative	Subjunctive	Jussive	Imperative			Subjunctive	Jussive
أنا 1 st .SG	كَتَبْتُ KaTaBtu	اكتُبُ aKTuBu	اكتُبْ aKTuBa	اكتُبْ aKTuB		كُتِبْتُ KuTiBtu	اكتُبْ uKTaBu	اكتُبْ uKTaBa	اكتُبْ uKTaB
أنت 2 nd .SG.M	كَتَبْتَ KaTaBta	تكتُبْ taKTubu	تكتُبْ taKTuba	تكتُبْ taKTuB	اكتُبْ uKTuB	كُتِبْتَ KuTiBta	تكتُبْ nuKTaBu	تكتُبْ tuKTaBa	تكتُبْ tuKTaB
أنت 2 nd .SG.F	كَتَبْتِ KaTaBti	تكتُبِينَ taKTubina	تكتُبِي taKTuBi	تكتُبِي taKTuBi	اكتُبِي uKTuBi	كُتِبْتِ KuTiBti	تكتُبِينَ tuKTaBina	تكتُبِي tuKTaBi	تكتُبِي tuKTaBi
انتما 2 nd .DUAL	كُتِبْتُمَا KaTaBtuma	تكتُبَانِ taKTuBaani	تكتُبَا taKTuBaa	تكتُبَا taKTuBa	اكتُبَا uKTuBaa	كُتِبْتُمَا KuTiBtuma	تكتُبَانِ tuKTaBaani	تكتُبَا tuKTaBaa	تكتُبَا tuKTaBaa
هو 3 rd .SG.M	كَتَبَ KaTaBa	يكتُبْ yaKTuBu	يكتُبْ yaKTuBa	يكتُبْ yaKTuB		كُتِبَ KuTiBa	يكتُبْ yuKTabu	يكتُبْ yuKTaBa	يكتُبْ yuKTaB
هي 3 rd .SG.F	كَتَبَتْ KaTaBat	تكتُبْ taKTuBu	تكتُبْ taKTuBa	تكتُبْ taKTuB		كُتِبَتْ KuTiBat	تكتُبْ tuKTaBu	تكتُبْ tuKTaBa	تكتُبْ tuKTaB
هما 3 rd .DUAL.M	كُتِبَا KaTaBaa	يكتُبَانِ yaKTuBaani	يكتُبَا yaKTuBa	يكتُبَا yaKTuBaa		كُتِبَا KuTiBaa	يكتُبَانِ yuKTaBaani	يكتُبَا yuKTaBaa	يكتُبَا yuKTaBaa
هما 3 rd .DUAL.F	كُتِبَتَا KaTaBataa	تكتُبَانِ taKTuBaani	تكتُبَا taKTuBaa	تكتُبَا taKTuBaa		كُتِبَتَا KuTiBataa	تكتُبَانِ tuKTaBaani	تكتُبَا tuKTaBaa	تكتُبَا tuKTaBaa
نحن 1 st .PL	كُتِبْنَا KaTaBnaa	نكتُبْ naKTuBu	نكتُبْ naKTuBa	نكتُبْ naKTuB		كُتِبْنَا KuTiBnaa	نكتُبْ nuKTaBu	نكتُبْ nuKTaBa	نكتُبْ nuKTaB
انتم 2 nd .PL.M	كُتِبْتُمْ KaTaBtum	تكتُبُونَ taKTuBuuna	تكتُبُوا taKTuBuu	تكتُبُوا taKTuBuu	اكتُبُوا uKTuBuu	كُتِبْتُمْ KuTiBtum	تكتُبُونَ tuKTaBuuna	تكتُبُوا tuKTaBuu	تكتُبُوا tuKTaBuu
انتن 2 nd .PL.F	كُتِبْتُنَّ KaTaBtunna	تكتُبْنَ taKTuBna	تكتُبْنَ taKTuBna	تكتُبْنَ taKTuBna	اكتُبْنَ uKTuBna	كُتِبْتُنَّ KuTiBtunna	تكتُبْنَ tuKTaBna	تكتُبْنَ tuKTaBna	تكتُبْنَ tuKTaBna
هم 3 rd .PL.M	كُتِبُوا KaTabu	يكتُبُونَ yaKTuBuuna	يكتُبُوا yaKTuBuu	يكتُبُوا yaKTuBuu		كُتِبُوا KuTiBuu	يكتُبُونَ yuKTaBunna	يكتُبُوا yuKtaBuu	يكتُبُوا yuKtaBuu
هن 3 rd .PL.F	كُتِبْنَ KaTaBna	يكتُبْنَ yaKTuBna	يكتُبْنَ yaKTuBna	يكتُبْنَ yaKTuBna		كُتِبْنَ KuTiBna	يكتُبْنَ yuKTaBna	يكتُبْنَ yuKTaBna	يكتُبْنَ yuKTaBna

Appendix C

Selected monolingual and bilingual dictionary entries of MSA (and CA) GO and COME verbs

Dahaba

DICTIONARY	DICTIONARY ENTRY
Al-Mawrid (2008) <i>Arabic – English</i> pp. 564.565	<p>الذهابُ: السَّيرُ والمُروءُ؛ ذَهَبَ يَذْهَبُ ذَهَابًا وَذُهُوبًا فَهُوَ ذَاهِبٌ وَذُهُوبٌ والمَذْهَبُ مصدر، كالذَّهَابِ وَذَهَبَ بِهِ وَأَذْهَبَهُ غَيْرُهُ: أزاله ويقال: أَذْهَبَ بِهِ، قال أبو إسحق: وهو قليل. فأما قراءة بعضهم: يَكادُ سَنًا بَرَقَهُ يَذْهَبُ بالأُصْبارِ، فنادرٌ وقالوا: ذَهَبْتُ الشَّامَ، فَعُدُّهُ بِغَيْرِ حَرْفٍ، وإن كان الشَّامُ طَرَفًا مَخْصُوصًا شَبَّهَهُ بِالْمَكَانِ الْمُنِيِّمِ، إذ كان يَقَعُ عليه المكانُ والمَذْهَبُ وحكى اللحياني: إنَّ اللَّيْلَ طَوِيلٌ، ولا يَذْهَبُ بِنَفْسِ أَحَدٍ مَنًا، أي لا ذَهَبَ والمَذْهَبُ المُتَوَصِّطُ، لأنَّهُ يَذْهَبُ إليه وفي الحديث: أنَّ النَّبِيَّ، صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ، كان إذا أراد الغائطَ أَيْعَدَ في المَذْهَبِ، وهو مَفْعَلٌ مِنَ الذَّهَابِ. الكسائي: يقال لموضع الغائط: الخلاء، والمَذْهَبُ، والمرْفَقُ، والمرْحاضُ والمَذْهَبُ المُتَعَدِّ الذي يَذْهَبُ إليه؛ وَذَهَبَ فَلانٌ لِدَهْبِهِ أي لِمَذْهَبِهِ الذي يَذْهَبُ فيه وحكى اللحياني عن الكسائي: ما يُدْرَى له أين مَذْهَبُ، ولا يُدْرَى له ما مَذْهَبُ أي لا يُدْرَى أين أصلُهُ ويقال: ذَهَبَ فَلانٌ مَذْهَبًا حَسَنًا</p>
Lisān Al-‘arab <i>Classical Arabic</i> (http://www.baheth.info/)	<p>ذَهَبٌ، كَمَنَعٌ، ذَهَابًا وَذُهُوبًا وَمَذْهَبًا، فَهُوَ ذَاهِبٌ وَذُهُوبٌ: سارَ، أو مَرَّ، ~ به: أزاله، كَأَذْهَبُهُ، وبه والمَذْهَبُ: المُتَوَصِّطُ، والمُعْتَدُّ الذي يَذْهَبُ إليه، والطَّرِيقَةُ، والأصلُ وَبَضَمَ الميم: الكَعْبَةُ، وَقَرَسَ أَبْرَهَةَ بِنِ عُمَيْرٍ، وَغَيَّ بِنِ أَعْصَرَ، وَشَيْطَانُ الوُضوءِ، وَكَسُرَ هاتِهِ الصوابُ، وَوَهَمَ الجوهريُّ</p>
Al-Qāmūs Al-Muḥīṭ <i>Classical Arabic</i> (http://www.baheth.info/)	<p>ذَهَبٌ: مصدره ذَهَبٌ [بمعنى سار ومضى وزال وامحى هو ذهاب وذهوب ومذهب]، [...] اذهب فعل الأمر يستعمل للتحريض والزرع مثل مقابله الفرنسي (معجم الماوردي). وذهب: هلك. فعند ابن القوطية (ص 7 و): فدارت بينهما حرب عظيمة ذهب فيها كلثوم وعشرة آلاف من الجيش. وفي النويري (الأندلس ص 457): مجاعة ذهب فيها خلقٌ كثير. وفي معجم البيان (ص 15): مما يذهب فيه الوصف بمعنى مما لا يمكن وصفه (تاريخ البربر 2: 45). ذهب عنه: تزكاه وأقلت منه. ففي المقرئ (241: 1): فما للصنيعة مذهب عنه. وذهب: خرج من المعسكر ليقضي حاجته، ومصدره مَذْهَبٌ (أنظر لين) ونجد عند كوسج (الطرائف ص 141): وكان جميل إذا أراد الحاجة أبعده في المذهب. وفي تاريخ البربر (607: 1): أبعده المذهب. وذهب: ذاع وانتشر. ففي الأغاني (ص 44) في كلامه عما حصل عليه امرؤ من ذبوع الصيت: قال معبد غنيت فأعجبني غناتي وأعجب الناس به وذهب لي به صوت وذكر. ذهب في: دخل، تغلغل، ففي ابن العوام (194: 1) في كلامه عن نباتات: ما لا يذهب عروقها في الأرض. وفي (290: 1) عليك أن تقرأ وفقا لما جاء في مخطوطتنا: لأنه ليس له أصل ذاهب في الأرض. ويقال أيضا: ذاهب في الهواء أو ذاهب في السماء أي مرتفع جدا. وذاهب في العرض أي عريض جدا (معجم الإدريسي) وكلمة قاطع معناها قوي، يقال نبيذ قاطع وخميرة قاطعة إلى غير ذلك غير أن صاحب محيط المحيط يفسر "دواء قاطع" (أي دواء قوي) بقوله: ذهبت قوته. ذهب عليه: لا يعني نسيه فقط (لين، دي يونج) بل يعني أيضا: لم ينتبه إليه وانصرف عنه، ففي التنوي (ص 81) وقد نقله دي يونج: وأي علم كان يذهب على الشافعي، يريد أن الشافعي درس كل العلوم. وذهب يليها فعل مضارع: من أفعال الشروع بمعنى أخذ يفعل وبدأ يفعل (معجم الطرائف، تاريخ الأغالية ص 16). وذهب: صمم، عزم، نوى، قصد، يقال: ذهب أن، ففي كتاب محمد بن الحارث (249): فذهب صاحب المدينة أن يأمر بزجره. كما يقال: ذهب إلى، ففي حين (ص 57 ق): وذهب إلى ادخال المسجد الجامع معه في قصيته. وفيه أيضا: اجتمع بنو خلدون – لأفكار ماذهب إليه من ذلك. ويجب إضافة إلى، إلى العبارة في حيان – بسام (1: 46 ق): فعرّفناه من كره من ورائنا لاجتياز ذهابهم (إلى) التمرس به (المقدمة 2: 44، ملر ص 8، أماري ديب ص 244). ويقال: ذهب إلى أن أيضا، ففي حيان (ص 57 ق): وذهب أمية بن عبد الغافر إلى أن يأخذ بالجزم في حراسة نفسه ودولته. وذهب إلى: فكر، رأى، ففي رحلة ابن بطوطة (2: 368): ذهب الأمير إلى راحتي.</p>
Takmilat Al-Ma‘āgim Al-‘arabiyya (1871-1877-1927-1978) <i>Monolingual</i> pp. 29-30	<p>ذَهَبٌ: مصدره ذَهَبٌ [بمعنى سار ومضى وزال وامحى هو ذهاب وذهوب ومذهب]، [...] اذهب فعل الأمر يستعمل للتحريض والزرع مثل مقابله الفرنسي (معجم الماوردي). وذهب: هلك. فعند ابن القوطية (ص 7 و): فدارت بينهما حرب عظيمة ذهب فيها كلثوم وعشرة آلاف من الجيش. وفي النويري (الأندلس ص 457): مجاعة ذهب فيها خلقٌ كثير. وفي معجم البيان (ص 15): مما يذهب فيه الوصف بمعنى مما لا يمكن وصفه (تاريخ البربر 2: 45). ذهب عنه: تزكاه وأقلت منه. ففي المقرئ (241: 1): فما للصنيعة مذهب عنه. وذهب: خرج من المعسكر ليقضي حاجته، ومصدره مَذْهَبٌ (أنظر لين) ونجد عند كوسج (الطرائف ص 141): وكان جميل إذا أراد الحاجة أبعده في المذهب. وفي تاريخ البربر (607: 1): أبعده المذهب. وذهب: ذاع وانتشر. ففي الأغاني (ص 44) في كلامه عما حصل عليه امرؤ من ذبوع الصيت: قال معبد غنيت فأعجبني غناتي وأعجب الناس به وذهب لي به صوت وذكر. ذهب في: دخل، تغلغل، ففي ابن العوام (194: 1) في كلامه عن نباتات: ما لا يذهب عروقها في الأرض. وفي (290: 1) عليك أن تقرأ وفقا لما جاء في مخطوطتنا: لأنه ليس له أصل ذاهب في الأرض. ويقال أيضا: ذاهب في الهواء أو ذاهب في السماء أي مرتفع جدا. وذاهب في العرض أي عريض جدا (معجم الإدريسي) وكلمة قاطع معناها قوي، يقال نبيذ قاطع وخميرة قاطعة إلى غير ذلك غير أن صاحب محيط المحيط يفسر "دواء قاطع" (أي دواء قوي) بقوله: ذهبت قوته. ذهب عليه: لا يعني نسيه فقط (لين، دي يونج) بل يعني أيضا: لم ينتبه إليه وانصرف عنه، ففي التنوي (ص 81) وقد نقله دي يونج: وأي علم كان يذهب على الشافعي، يريد أن الشافعي درس كل العلوم. وذهب يليها فعل مضارع: من أفعال الشروع بمعنى أخذ يفعل وبدأ يفعل (معجم الطرائف، تاريخ الأغالية ص 16). وذهب: صمم، عزم، نوى، قصد، يقال: ذهب أن، ففي كتاب محمد بن الحارث (249): فذهب صاحب المدينة أن يأمر بزجره. كما يقال: ذهب إلى، ففي حين (ص 57 ق): وذهب إلى ادخال المسجد الجامع معه في قصيته. وفيه أيضا: اجتمع بنو خلدون – لأفكار ماذهب إليه من ذلك. ويجب إضافة إلى، إلى العبارة في حيان – بسام (1: 46 ق): فعرّفناه من كره من ورائنا لاجتياز ذهابهم (إلى) التمرس به (المقدمة 2: 44، ملر ص 8، أماري ديب ص 244). ويقال: ذهب إلى أن أيضا، ففي حيان (ص 57 ق): وذهب أمية بن عبد الغافر إلى أن يأخذ بالجزم في حراسة نفسه ودولته. وذهب إلى: فكر، رأى، ففي رحلة ابن بطوطة (2: 368): ذهب الأمير إلى راحتي.</p>

<p>Mu'ḡam Al-Luġa Al-'arabiyya Al-Mu'āšira (2008) <i>Monolingual</i> pp. 823</p>	<p>وذهب مع: وافق، طواع، جارى (تاريخ البربر 1:608، 2:165). ذهب الشخص: 1 انصرف، غادر المكان "ذهب فلان على عجل - ذهب ولم يعد - ذهبوا أيدي سبأ [مثل]: تفرقوا كما تفرقت قبائل اليمن في البلاد عندما غرقت أرضهم وذهبت جناتهم - (يحبسون الأجزاء لم يذهبوا) ". ذهب بخياله بعيداً: شطح - ذهب جهده سدى: لم يأت عمله بأي نتيجة، وانتهى بدون جدوى - ذهب عمله أدراج الرياح: ضاع جهده عبثاً ودون فائدة وبلا نتيجة. 2 سار، مضى ومر "الوقت من ذهب لا يعود منه ما ذهب - (إنا ذهبنا نستيق وتركنا يوسف عند متاعنا)". ذهب وجاء: كرر الذهاب والحضور - ذهبت بهم الرياح كل مذهب: تشتتوا وتفرقوا في كل اتجاه. 3 ابتعد وقد يفيد التهديد حين يرد بصيغة الأمر (أذهب فمن تبعك منهم فإن جهنم جزاؤكم). أذهب إلى الجحيم: عبارة تهديد وقد تدل على الإستهاء الشديد. 4 مات، هلك "ذهب الطيبون - (فلا تذهب نفسك عليهم حسرت) ". ذهاب الروح: الموت - ذهب إلى العالم الآخر: مات - ذهب حسرة: هلك- ذهبوا تحت كل كوكب: تفرقوا في كل اتجاه في الأرض. ذهب الخير: ذاع وانتشر ذهب الأثر: زال وأمحي "ذاهب اللون - ذهب مع الرياح: تلاشى - وانما الأمم الأخلاق ما بقيت... فإن هم ذهبت أخلاقهم ذهبوا - (ولا تنازعوا فتفشلوا وتذهب ريحكم)". ذهب كأمس الذابن: زال من دون أن يترك أثراً - ذهبت ريحه: زالت دولته، غلب، ضعف أمره. ذهب مع فلان: وافقه، طواعه، جراه. ذهب مذهب فلان: قصد قصده وسلك طريقه. ذهب في الدين مذهباً: رأى فيه رأياً، وأحدث فيه بدعة. ذهب إلى عمله: قصده، توجه إليه "ذهب إلى الجامعة / بيروت - أذهب إلى أبيك والتمس منه الصفح - ذهب إلى قول فلان: أخذ به - (أذهب إلى فرعون انه طغى)". ذهب رأساً إليه. ذهب به: 1 أزاله وأضاعه "ذهب الله بنورهم وتركهم في ظلمات لا يبصرون) - (يكاد سنا برقه يذهب بالأبصار) : يخطفها". ذهب برشده - ذهبت به الخيلاء: أزالته عن وقاره فتماذى في الكبرياء والعجب. 2 انفرد واستقل (إذا لذهب كل إله بما خلق). 3 فاز (ولا تعضلوهن لتذهبوا ببعض ما آتينهون). 4 صاحبه في المضى "ذهب الشرطي باللص إلى قسم الشرطة - (فلما ذهبوا به وأجمعوا أن يجعلوه في غيبت الجب). ذهب بخياله مذهباً بعيداً: شطح بخياله - ذهب فلان بالفخر: انفرد به. 5 أنبله "ذهب الزمن بنضرته - فلان ذاهب اللون". 6 أماته، أهلكه "ذهبت به الحمى". ذهب عليه كذا: نسبه ولم ينتبه إليه، وانصرف عنه "ذهب علي الموعد فأرجو المعذرة". ذهب عنه: تركه، وأفلت منه "ذهب عنه الغضب - أذهب عني، فلا أريد سماعك - تجنّب الطمع يذهب عنك الفقر". ذهب الشيء في الشيء: اختلط، دخل، تغلغل. ذهبت النفس فيه كل مذهب: تحيرت في فهمه.</p>
<p>Al-Munġid fī Al-Luġa wa Al-A'lam (2005) <i>Monolingual</i> pp. 239-240</p>	<p>1. سار مضى مات و - الأمر: انقضى و - علي الشيء: نسبته و - به: استصحبه وذهب معه أزاله من مكانه ويقال "ذهبت به الخيلاء" أي تهادى في الكبرياء والعجب. 2. ذهب في المسألة إلى كذا: رأى فيها ذلك الرأي</p>
<p>A Frequency Dictionary of Arabic (2011), pp. 53</p>	<p>489 ذهب v. I (a) to go, leave, depart لماذا لا تذهب إلى التحقيق وتخبرهم بالحقيقة؟ Why don't you go to the investigator's office and tell them the truth? range count/dispersion = 90% of the corpus raw frequency = 8703 </p>

Maḍā

DICTIONARY	DICTIONARY ENTRY
<p>Al-Mawrid (2008) <i>Arabic – English</i> pp. 1055-1056</p>	<p>ذهب to go, go away, leave, depart مضى to pass, elapse, go by, slip by, expire, run out, be past, be over مضى في: to continue (to do), go on doing; to proceed in or with مضى على: to conclude, wind up, accomplish, carry out, execute, perform مضى to advance, proceed, go forward, go ahead, go on; to progress, make progress, advance, make head; to be underway, in progress, in process, afoot, on foot مضى: كان حاداً ماضياً مضى سبيله أو لسبيله: مات مضى قاتلاً، مضى يقول، مضى فقال he went on to say, he added, he continued saying مضى على ذلك أعوام years have passed since then مضى من أسبوع مضى one week ago, last week فيما مضى - راجع في</p>
<p>Al-Qāmūs Al-Muḥīṭ <i>Classical Arabic</i> (http://www.baheth.info/)</p>	<p>مَضَى يَمْضِي مَضِيًّا وَمُضَوًّا: خَلَا، و- فِي الْأَمْرِ مَضَاءً وَمُضَوًّا: نَقَدَ، وَأَمْرٌ مَمْضُوٌّ عَلَيْهِ، و- سَنِيْبُهُ: مَاتَ، و- السَّنِيْفُ مَضَاءٌ: قَطَعَ وَأَمْضَاءٌ: أَنْفَدَهُ</p>

<p>Takmilat Al-Ma'āgim Al-'arabiyya (1871-1877-1927-1978) Monolingual pp. 81</p>	<p>أتى به الى موضع كذا: أوصله ففي معجم أبي الفداء: أتى بالخليج الى موضع كذا وأتى فلان (بالبناء للمجهول): استولع عليه العدو و غلبه، ففي المختار من تاريخ العرب: لست أوتى من قلة الرجال، أي لست أغلب وأتيت: فعلت به الفحشاء (معجم الإدريسي والقرني 2:461) وأتى عليه: أتممه وأنهاه، يقال مثلاً: أتى على ذكر فلان: أنهى أو أتم ذكر تاريخه (معجم أبي الفداء) وأتى عليه: أهلكه وأفناه (معجم بديون ومعجم البلاذري)، (أتى في معجم بديون ليس معناه أهلكه وأفناه بل معناه أشرف عليه العدو ودنا منه)، وهو المعنى الذي ذكره لين 16</p>
<p>Al-Munğid fi Al-Luğa wa Al-A'lam (2005) Monolingual pp. 3</p>	<p>1. أتى: جاء و – المكان: حضره و – الرجل: مر به 2. – الشيء: فعله و – على الشيء: أتمه، أنفذه وبلغ آخره. ومنها "أتى عليه الدهر": أهلكه</p>
<p>Mu'ğam Al-Luğa Al-'arabiyya Al-Mu'āšira (2008) Monolingual pp. 59</p>	<p>أتى الشخص: جاء وحضر، وصل، نزل وحل "سوف يأتي يوم الحساب بيننا أجلاً أو عاجلاً" - (فسوف يأتي الله بقوم بحبيهم ويحبونه) كما يأتي: كما يلي أتى الأمر: حان، قرب ودنا " (أتى أمر الله فلا تستعجلوه) ". أنت ساعته دنا أجله أتى الشيء تاماً: صار وأصبح تاماً " (فألقوه على وجه أبي يأت بصيراً) " أتى المكان والرجل: جاء، قصده، مر به "إذا أتاكم من ترضون دينه وخلقه فزوجه (حديث) - (فلما أتتها نودي بيموسى)" أتى المرأة: باشرها وجامعها " (فاذا تطهرن فأتوهن من حيث أمركم الله)" أتى الأمر: فعله، أنجزه وحققه "أتى جرماً - * لا تنته عن خلق وتأت مثله* (أتأتون الفاحشة) - أتى البيوت من أبوابها: توصل إلى الأمور من مدخلها الطبيعي، ذهب مباشرة إلى الهدف أتى البيان من قواعد: هدمه " (فأتى الله بنيينهم من القواعد)" أتى به: جاء به وجلبه، أحضره، أوصله "لم يأت بجديد في بحثه - أتى بخطة جديدة - أتى بملاحظة غير متوقعة" أتى عليه: 1 مر به " (ما تذر من شيء أتت عليه إلا جعلته كالرميم) ". 2 أتمه وأنهاه، نفذه وحققه "أتى على المشروع" - أتى على آخره: أتمه، 3 أشرف عليه " (حتى إذا أتوا على واد النمل) ". 4 أنفذه وأفناه، أهلكه وقضى عليه "أتى عليه الدهر - أتت العاصفة على المحصول - أتت النيران على المنزل - (ما تذر من شيء أتت عليه إلا جعلته كالرميم) " - أتى على الأخضر واليابس: دمر كل شيء</p>
<p>A Frequency Dictionary of Arabic (2011), pp. 39</p>	<p>343 - أتى v.I (i) to come إلى to sth; to come to sb; to bring ب sth بهذا مشاهدنا الكرام نأتي إلى ختام حلقة اليوم من السلطة الرابعة with this, dear viewers, we come to the conclusion of today's episode of "The Forth Estate" range count/dispersion = 90% of the corpus raw frequency = 12231 </p>

Gā'a

DICTIONARY	DICTIONARY ENTRY
<p>Al-Mawrid (2008) Arabic – English pp. 405</p>	<p>أتى to come, arrive, show up,; to reach, get to أحضر to bring, bring forward, fetch, get, produce, advance, present, introduce فعله to do, perform; to make ارتكبه to commit, perpetrate to be mentioned, stated, reported, said (في): ورد، ذكر the newspaper reports that جاء في الصحيفة أن...</p>
<p>Lisān Al-'arab Classical Arabic (http://www.baheth.info/)</p>	<p>المجيء: الإتيان. جاء حَيِّئًا وَمَجِيئًا. وحكى سيبويه عن بعض العرب هو يَجِيكُ بحذف الهمزة. وأجاءه إلى الشيء: جاء به وأجأه واضطره إليه؛ قال زهير بن أبي سلمى: وجارٍ، سارٍ مُعْتَمِدًا اليُكْمِ، * أجاءتُهُ المخافة والرَّجاء قال الفراء: أصله من جئت، وقد جعلته العرب الرجاء.</p>
<p>Takmilat Al-Ma'āgim Al-'arabiyya (1871-1877-1927-1978) Monolingual pp. 355-356</p>	<p>جاء، يقال: جاء من مثل ما يقال. دخل من ففي ألف ليلة وليلة (86:1): اطلع من المكان الذي جئت منه وجاء النبات والشجر: نمت جيدا ونجحت زراعة (ابن العوام 1:320) وجاء: بلغه ووصل إليه (معجم هابيشنت في الجزء الرابع من طبيعته لالف ليلة وجاء: شغل، ملأ المكان، يقال مثلاً: جاء الصندوق قياس الحاصل سوا يسوا (هابيشنت معجم جاء في بطنه: جرحه في بطنه (كرتاس 67) جاء الحديث عليه: صار دوره للتحدث (كوسج مختارات ص 61) الآن جاء الجد في قطع حباتي: الآن عليك أن تبذل كل جهد وتجد في قطع حباتي (كليلة ودمنة من 224) جاءت طريقهم على تلك الدار: أوصلتهم الطريق إلى تلك الدار (ألف ليلة 1:67) مهما جاء عليه أنا أوزنه عنه: مهما صارت حصته من النقطة فانا أؤدبها عنه (ألف ليلة 1:60) جاء عليه: طابقه، ناسبه، لاق عليه، كان على قدمه، يقال مثلاً: ماتجىء عليك هذه البذلة، أي أنها ليست مطابقة ومناسبة ولائقة لجسمك (بوشر) جاء على ميله: كان موافقا لذوقه، وقع عنده موقع الرضا (بوشر) وجاء عليه وبه: كلفه، يقال مثلاً: هذا الشيء جاء علي بكذا. أي كلفني كذا، بلغ ثمنه كذا (فوك)</p>

	<p>جاء له من: كسب من، استفاد من، انتفع من. يقال مثلاً: أيش قد يجيك من وظيفتك، أي كم تكسب؟ ويقال: يجي لك من دا ايه بمعنى أي نفع لك في هذا (بوشر) جاءت نفسه: عاد إلى رشده، استفاق (الأغاني 52) جاء من قدرك أن تتكلم بهذا الكلام: أصار من قدرك أن تتكلم بهذا الكلام، كيف جرؤت أن تتكلم بهذا الكلام (بوشر) خُذ مني على ما يجيك: لن أنساها لك وسأنتقم منك (بوشر) جاء. من اليوم وجاي: أي من اليوم إلى ما يليه (دي ساس ديب 471:9)</p> <p>جيب: جابه، في لغة العامة مختصر جاء به وهو بمعناه أي أتى به، يقال: جابت الشجرة: أتت بالثمر، أثمرت. وجاب شهودا: أتى بشهود</p>
Al-Munğid fi Al-Luğa wa Al-A'lam (2005) <i>Monolingual</i> pp. 112	<p>1. جاء: أتى و – به: آجاءه و – هُ وإليه: آتاه 2. جاء الشيء: فعله</p>
Mu'ğam Al-Luğa Al-'arabiyya Al-Mu'āšira (2008) <i>Monolingual</i> pp. 426	<p>جاء الأمر: حدث، تحقق "جاءت البشرية - (إذا جاء نصر الله والفتح)". جاء على هواه/جاء على ميله: كان موافقاً لذوقه، وقع عنده موقع الرضا - جاء له من حيث لا يدري: كسبه، استفاد منه، انتفع منه جاء الشخص/جاءني الشخص/جاء الي الشخص: حضر، أتى، أقبل "جاء من طلوع الشمس - جاء من السفر: عاد - جاء من السجن: خرج - (حتى إذا جاءوها وفتحت أبوابها): دخلوها". جاء في حينه: في الوقت المناسب - جاء في صحبته - جاء في عقبه/جاء عقبه - جاء من ذي نفسه: طوعاً غير مكره، من تلقاء نفسه - جاءوا على بكرة أبيهم: جميعاً - ذهب، وجاء جاء الأمر/جاء بالأمر: فعله، صنعه "جاء بالحسنة - جاء رجال الشرطة بالمتهم: أحضروه - جاء بالخير: بلّغه - (لقد جنتم شيئاً أداً)" جاء في الصحف/جاء في المقال: ورد "جاء ذكره في الكتاب"</p>
A Frequency Dictionary of Arabic (2011), pp. 18	<p>109 جاء v.I (i) to come إلى to; to come to sb: e.g. جاءني صديق. a friend came to (see) me; I received a letter; to appear, show up (في in sth in written or spoken) أوضحوا أن الإرتفاعات في الأسعار جاءت كنتيجة حتمية لارتفاع أسعار الأعلاف والمياه وأجور النقل The clarified that the rise in prices came as an inevitable result of the rise in prices of feed and water and transportation costs range count/dispersion = 99% of the corpus raw frequency = 26234 </p>

Ḥaḍara

DICTIONARY	DICTIONARY ENTRY
Al-Mawrid (2008) <i>Arabic – English</i> pp. 475	<p>أتى to attend, be present, be there, to report (for duty, to a certain place), present oneself; to come, show up, appear, arrive; to reach, get to; to visit, go to شاهد، شهد to attend, go to; to view, see, watch; to witness تذكره، حضره to be recalled by, come to someone's mind, occur to حضره الموت - راجع احتضر حضر: تحضر، تمّدن - راجع تحضر</p>
Lisān Al-'arab <i>Classical Arabic</i> (http://www.baheth.info/)	<p>حضر (لسان العرب) الحضور: نقيض المغيب والغيبية؛ حضر يحضر حضوراً وحضارة؛ ويُعدى فيقال: حضر هو حضره * قوله: «فيقال حضره وحضره الخ» أي فهو من بابي نصر وعلم كما في القاموس). يحضره، وهو شاذ، والمصدر كالمصدر.</p>
Takmilat Al-Ma'āğim Al-'arabiyya (1871-1877-1927-1978) <i>Monolingual</i> pp. 224-225	<p>حضر: أتى. ففي تاريخ بني زيان (ص95): حضر من فاس إلى تلمسان وحضر الكتاب: ذهب إلى الكتاب أي موضع تعليم الصبيان. ففي رياض النفوس (ص70): فسأل أبي عني إن كنت أحضر الكتاب فقال له أبي نعم أي فسأل أبي إن كنت أذهب إلى الكتاب. وحضر على فلان: شهد الدرس الذي يلقيه (أنظر سمع على) (المقري 1:842). ويقال أيضاً: حضر عند فلان (ميرسنج ص21) ويقال: حضر على فلان كتاباً (طنطاوي في زبشر كند 51:7) ولم أحضر نحواً: لم أشهد درس النحو (نفس المصدر 7:1) وحضرت في النحو والفقهِ (نفس المصدر 3:1) ويذكر بوشر: حضر له وحضرتي كذا. وعند لين: أتأذنين في ذكر شيء حضر، أي أتأذنين في ذكر شيء خطر ببالي؟ (معجم برون) حضره شيء، يعني أيضاً: رغب في عمل شيء. ففي رياض النفوس (ص48): ثم نهض للقيام وقال من حضره (كذا) الزيارة لو اصل (اسم شخص) فليقم ثم خرج من فوره وخرج مع أصحابه. وحضر فلانا وحضر به: أتى إليه بشيء (أخبار ص19) وحضر فيه: تكلم فيه. يقال: ونحضر فيهم كل يوم محاضرة أي نتكلم فيهم في كل اجتماع (ماري ديب ص2)</p>

	وحضر: ازدهر، عمر (المعجم الإدريسي)
Al-Munğid fi Al-Luğa wa Al-A'lam (2005) <i>Monolingual</i> pp. 139	1. حضر: ضد غاب و - حضورا المجلس شهده يقال " حضرت الأمر بخير " أي رأيت فيه رأيا صوابا و - : جعله حاضرا 2. حضر الوقت: حان. يقال " حضرت الصلاة " أي جاء وقتها و - : الموت: جاءه و - : الأمر: خطر بباله و - : إليه: أتى
Mu'ğam Al-Luğa Al-'arabiyya Al-Mu'āšira (2008) <i>Monolingual</i> pp. 512	حضر الشخص ونحوه: قدم، ضد غاب " حضرت الشرطة بعد تلقيها البلاغ - ذو حضور مؤثر - حضور الذهن: سرعة الإدراك - من خافك حاضرا أبغضك غائبا (مثل) - (وأعوذ بك رب أن يحضرون): أن تصيبني الشياطين بسوء - (ووجدوا ما عملوا حاضرا): مسجلا محفوظا حضر الشيء أو الأمر: أتى، جاء ونهياً " حضر الكتاب " حضرت الصلاة: حل وقتها " حضر الوقت: أرف، حان، وافى " حضر المجلس/حضر المكان: شهده، ذهب إليه " حضر الوزير الجلسة الختامية - حضر حرب أكتوبر: أدركها - (وإذا حضر القسمة أولوا القربى واليتامى والمساكين فارزقوهم) " حضر الأمر فلانا: 1 حل نزل به " (كُنْتُ عَلَيْكُمْ إِذَا حَضَرَ أَحَدَكُمُ الْمَوْتَ إِنْ تَرَكَ خَيْرًا الْوَصِيَّةُ لِلْوَالِدَيْنِ وَالْأَقْرَبِينَ) ". 2. خطر بباله " حضره ذلك الحادث الذي رآه بعينه " حضر عن فلان: قام مقامه في الحضور، ناب عنه " حضر الوزير الاحتفال نائباً عن رئيس الجمهورية "
A Frequency Dictionary of Arabic (2011), pp. 82	809 - حضر v. I (u) to come, show up; to attend, be present at (meeting, party, concert); to view (film, TV show) ذياب لم يحضر الى الفندق منذ شهرين تقريبا ولعله لن يعود Diyab has not come to the hotel for about two months; and maybe he will not return range count/dispersion = 99% of the corpus raw frequency = 4598

Qadima

DICTIONARY	DICTIONARY ENTRY
Al-Mawrid (2008) <i>Arabic - English</i> pp. 852	to come, arrive, show up, reach; to reach, get to to return, come up
Al-Şahhāh fi Al-Luğa <i>Monolingual</i> (http://www.baheth.info/)	قَدِمَ من سفره قَدوماً ومَقْدَمًا بفتح الدال. يقال: وَرَدْتُ مَقْدَمَ الْحَاجِّ، تجعله ظرفاً وهو مصدرٌ، أي وقت مَقْدَمِ الْحَاجِّ. وقَدِمَ بالفتح يَقْدِمُ قَدَمًا، أي تَقْدِمُ، قال الله تعالى: "يَقْدُمُ قَوْمَهُ يَوْمَ الْقِيَامَةِ فَأُورِدُهُمُ النَّارَ. الشيء بالضم قَدِمًا فهو قَدِيمٌ، وتَقَادِمٌ مثله
Takmilat Al-Ma'āğim Al-'arabiyya (1871-1877-1927-1978) <i>Monolingual</i> pp. 199-200	قدم الى: وصل (بوشر، كلبلة ودمنة ص 280)، ويقال أيضا: قدم ل (معجم أبي الفداء) قدم على، أو قدم أن: اجترأ على مهاجمته، أو قتله (معجم الطرائف، معجم بدرن)
Al-Munğid fi Al-Luğa wa Al-A'lam (2005) <i>Monolingual</i> pp. 613	قَدِمَ المدينة: أتاها و - من سفره: عاد و - الى الأمر: قصد له
Mu'ğam Al-Luğa Al-'arabiyya Al-Mu'āšira (2008) <i>Monolingual</i> pp. 1783	قدم / قدم الى / قدم من، يقدم، قنوما، فهو قادم، والمفعول مقدم قدم فلان المدينة / قدم فلان الى المدينة: دخلها، جاء إليها، حل بها "تنتفتح الأزهار مع قدوم الربيع" قدم الى الأمر: قصد له وعمد إليه (وقدمنا الى ما عملوا من عمل فجعلناه هباء منثورا) قدم من السفر: عاد، رجع "قدم من الخارج - عاد محملا بأفكار جديدة بعد قدومه من أمريكا"
A Frequency Dictionary of Arabic (2011), pp. 288	3121 - قدم v. I (a) to arrive, come to; approach على sth غالبية الأطباء العاملين في الولايات المتحدة قدموا من الهند والفلبين وباكستان Most doctors working in the United States came from India, the Philippines, and Pakistan range count/dispersion = 87% of the corpus raw frequency = 566

Appendix D

Examples for annotation per variable (and each level within every variable)

i. Morphological variables

categories	levels	sample of annotation
TENSE	PRESENT	مساعداتنا <u>تذهب</u> إلى الشيشان aids.CL.1PL.GEN <u>dahaba.IMPF.3SG.F</u> ALL ART=Chechnya 'Our aid goes to Chechnya' تأتي الإعانات الرئيسية من المتبرعين <u>atā.IMPF.3SG.F</u> ART=aids ART=main ABL ART=donors 'The main financial aids come from the donors'
	PAST	<u>ذهب</u> الأصدقاء و <u>ذهب</u> زمانهم <u>dahaba.PERF.3SG.M</u> ART=friends <u>CONJ=dahaba.PERF.3SG.M</u> time.CL.3PL.M.GEN 'friends went away and so did their time' و <u>جاء</u> حفل الإفتتاح بسيطا وجميلا <u>CONJ=gā'a.PERF.3SG.M</u> party ART=opening simple.ADV CONJ=beautiful.ADV 'And the opening ceremony was simple and beautiful'
	FUTURE	<u>سيذهب</u> الحزن <u>FUT-dahaba.IMPF.3SG.M</u> pain 'pain will go away' غدا ستأتيهم سيارة عند الفجر tomorrow <u>FUT-atā.IMPF.3SG.F-CL.3PL.M.ACC</u> car LOC ART=sunrise 'Tomorrow a car will come to them at sunrise'
	IRREALIS (non-finite forms)	من الظلم أن <u>تذهب</u> البطولة لغيره ABL unfairness TOP <u>dahaba.SUBJ.3SG.F</u> championship ALL=other.CL.3SG.M 'It is unfair that the championship goes to someone other than him' ولم <u>يذهب</u> القلق مع رد المضيفة <u>CONJ=NEG dahaba.IMPF.3SG.M</u> ART=anxiety COM response ART=hostess 'The anxiety did not go away with the response of the hostess'
ASPECT	SIMPLE	مساعداتنا <u>تذهب</u> إلى الشيشان aid.CL.1PL.GEN <u>dahaba.IMPF.3SG.F</u> ALL ART=Chechnya 'Our aid goes to Chechnya' و <u>جاء</u> حفل الإفتتاح بسيطا وجميلا <u>CONJ=gā'a.PERF.3SG.M</u> party ART=opening simple.ADV CONJ=beautiful.ADV 'And the opening ceremony was simple and beautiful'
	HABITUAL	<u>كان يذهب</u> إلى المقابر كل يوم <u>AUX dahaba.IMPF.3SG.M</u> ALL ART=graveyards everyday 'He used to go to graveyards everyday' تأتي الإعانات الرئيسية من المتبرعين <u>atā.IMPF.3SG.F</u> ART=aids ART=main ABL ART=donors 'The main financial aids come from the donors'
	PROGRESSIVE	رأوا عددا كبيرا من الرجال <u>يأتون</u> مسرعين see.PERF.3PL.M number big ABL men <u>atā.IMPF.3PL.M</u> quickly 'They saw a large number of men approaching very quickly'

	PERFECT	وكان قد ذهب إلى ألمانيا CONJ= <u>AUX DM dahaba.SUBJ.3SG.F</u> ALL Germany 'And he had gone to Germany'
	DURATIVE / CONTINUOUS	وراح البابا بيارك الحضور CONJ= <u>rāha.PERF.3SG.M</u> ART=Pope <u>bless.IMPF.3SG.M</u> ART=audience 'And the Pope went on blessing the audience' ومضى الرئيس الإيراني يقول CONJ= <u>madā.PERF.3SG.M</u> ART=president ART=Iranian <u>say.IMPF.3SG.M</u>
	NON-FIN (non-finite forms)	من الظلم أن تذهب البطولة لغيره ABL unfairness TOP <u>dahaba.SUBJ.3SG.F</u> championship ALL=other.CL.3SG.M 'It is unfair that the championship goes to someone other than him' ولم يذهب القلق مع رد المضيفة <u>CONJ=NEG dahaba.IMPF.3SG.M</u> ART=anxiety COM response ART=hostess 'The anxiety did not go away with the response of the hostess'
MORPHOLOGICAL ASPECT OR MOOD OF THE VERB	IMPERFECTIVE	يذهب yaḏhabu <u>dahaba.IMPF</u> , يأتي ya'tī <u>atā.IMPF</u> , يقدم yaqdumu <u>qadima.IMPF</u> , etc.
	PERFECTIVE	ذهب <u>dahaba dahaba.PERF</u> , أتى <u>atā atā.PERF</u> , قدم <u>qadima qadima.PERF</u> , etc.
	SUBJUNCTIVE	يذهب yaḏhaba <u>dahaba.IMPF</u> , يأتي ya'tia <u>atā.IMPF</u> , يقدم yaqduma <u>qadima.IMPF</u> , etc.
	JUSSIVE	يذهب yaḏhab <u>dahaba.IMPF</u> , يأتي ya'ti <u>atā.IMPF</u> , يقدم yaqdu <u>qadima.IMPF</u> , etc.
	IMPERATIVE	اذهب iḏhab <u>dahaba.IMPR</u> , امض imḏi <u>maḏā.IMPR</u> , etc.
SUBJECT PERSON	1 ST	ذهبت <u>dahabtu dahaba.1SG</u> , أتيت <u>ataytu atā.1SG</u> , قدمت <u>qadimtu qadima.1SG</u> , etc.
	2 ND	ذهبت <u>dahabata dahaba.2SG</u> , أتيت <u>atayta atā.2SG</u> , قدمت <u>qadimta qadima.2SG</u> , etc.
	3 RD	ذهب <u>dahaba dahaba.3SG</u> , أتى <u>atā atā.3SG</u> , قدم <u>qadima qadima.3SG</u> , etc.
SUBJECT NUMBER	SINGULAR	ذهب <u>dahaba dahaba.3SG</u> , أتى <u>atā atā.3SG</u> , قدم <u>qadima qadima.3SG</u> , etc.
	DUAL	ذهبا <u>dahabā dahaba.3DUAL</u> , أتيا <u>atayā atā.3DUAL</u> , قدما <u>qadimā qadima.3DUAL</u> , etc.
	PLURAL	ذهبوا <u>dahabū dahaba.3PL</u> , أتوا <u>ataw atā.3PL</u> , قدموا <u>qadimū qadima.3PL</u> , etc.
SUBJECT GENDER	FEMININE	ذهبت <u>dahabat dahaba.3SG.F</u> , أتيت <u>atayna atā.3PL.F</u> , قدمت <u>qadimtunna qadima.2PL.F</u> , etc.
	MASCULINE	ذهبوا <u>dahabū dahaba.3PL.M</u> , أتى <u>atā atā.3SG.M</u> , قدمت <u>qadimta qadima.2SG.M</u> , etc.
	NIL (for 1 st person inflections)	ذهبت <u>dahabtu dahaba.1SG</u> , أتينا <u>ataynā atā.1PL</u> , etc.

ii. Syntactic variables

categories	levels	sample of annotation
TRANSITIVITY	YES	كان يفترض أن يحضروا المؤتمر AUX suppose.PASS.3SG.M TOP <u>ḥaḏara.SUBJN.3PL.M</u> ART= <u>conference.ACC</u>

		<p>‘They were supposed to attend the conference’ غدا ستأتيهم سيارة عند الفجر tomorrow FUT-atā.IMPF.3SG.F-CL.3PL.M.ACC car LOC ART=sunrise ‘Tomorrow a car will come to them at sunrise’</p>
INTERROGATION	YES	<p>من أين جاء؟ ABL Q ḡā’a.PERF.3SG.M ‘Where did he come from?’ لماذا يذهب الطيبون؟ Q dahaba.IMPF.3SG.M ART=good.ones ‘Why do the good people go/die?’</p>
NEGATION	YES	<p>لما مضى الأمر على خير PURP=NEG maḡā.PERF.3SG.M ART=issue LOC good ‘It would not have gone well’ ولم يذهب القلق مع رد المضيفة CONJ=NEG dahaba.IMPF.3SG.M ART=anxiety COM response ART=hostess ‘The anxiety did not go away with the response of the hostess’</p>
SERIAL VERB CONSTRUCTION (also covers auxiliary and main verb patterns)	YES	<p>وراح البابا بيارك الحضور CONJ=rāha.PERF.3SG.M ART=Pope bless.IMPF.3SG.M ART=audience ‘And the Pope went on blessing the audience’ وجاء يحييك كأنه افتقدك سنوات CONJ=ḡā’a.PERF.3SG.M greet.IMPF.3SG.M-CL.2SG.M.ACC ADV miss.PERF.3SG.M-CL.2SG.M.ACC years ‘And he came greeting you as if he missed you for years’</p>
PREPOSITIONAL PHRASE	YES	<p>كان يذهب إلى المقابر كل يوم AUX dahaba.IMPF.3SG.M ALL ART=graveyards everyday ‘He used to go to graveyards everyday’ تأتي الإعانات الرئيسية من المتبرعين atā.IMPF.3SG.F ART=aids ART=main ABL ART=donors ‘The main financial aids come from the donors’</p>
LOCATIVE ADVERB PHRASE	YES	<p>ولم يذهب القلق مع رد المضيفة CONJ=NEG dahaba.IMPF.3SG.M ART=anxiety COM response ART=hostess ‘The anxiety did not go away with the response of the hostess’ تأتي دائما عبر عمليات السطو المنتظم atā.IMPF.3SG.F ADV LOC operations burglary ART=organized ‘Comes always through operations of organized burglary’</p>
ADVERBIAL PHRASE	YES	<p>وجاء حفل الإفتتاح بسيطا وجميلا CONJ=ḡā’a.PERF.3SG.M party ART=opening simple.ADV CONJ=beautiful.ADV ‘And the opening ceremony was simple and beautiful’ هذه الجهود لم تذهب هدرا DEM ART=efforts NEG dahaba.JUSS.3SG.F vain.ADV ‘These efforts weren’t in vain’</p>

iii. Semantic variables

categories	levels	sample of annotation
SUBJECT CATEGORY	ACTIVITY	ارتفاع، ‘voting’، تصويت، ‘operations’، عمليات، ‘attack’، هجوم، ‘increase’، تأجيل، ‘postponing’، etc.

	ANIMAL	جواد 'horse', كلب 'dog', etc.
	ATTRIBUTE	كرم 'generosity', شجاعة 'bravery', شهرة 'fame', etc.
	BODY	عيون 'eyes', قدم 'foot', رؤوس 'heads', etc.
	COGNITION	تفكير 'thought', خيال 'imagination', توجس 'apprehension', etc.
	COMMUNICATION	رد 'question', تقرير 'report', تصريح 'statement', 'response', كلمات 'words', etc.
	CONTENT (of a document/speech)	البيان جاء في <i>gā'a.PERF.3SG.M LOC ART=statement</i> 'came in the statement...' الرسالة جاء في <i>gā'a.PERF.3SG.M LOC ART=letter</i> 'came in the statement...' التقرير جاء في <i>gā'a.PERF.3SG.M LOC ART=report</i> 'came in the statement...', etc.
	DEMONSTRATIVE	ذلك جاء <i>gā'a.PERF.3SG.M LOC DEM</i> 'that came...' هذا جاء <i>gā'a.PERF.3SG.M LOC DEM</i> 'this came...', etc.
	EVENT	اجتماع 'meeting', ندوة 'symposium', قمة 'summit', حفل 'party', زيارة 'visit', محاضرة 'lecture', etc.
	GROUP (representing humans collectively)	اليابان 'Japan', المنتخب 'varsity', الحكومة 'the government', الصحيفة 'the newspaper', اللجنة 'the committee', etc.
	HUMAN	الأولاد 'the boys', البابا 'the Pope', الرئيس التركي 'the Turkish president', etc.
	LOCATION	موقع 'location', المدن 'the cities', etc.
	NOTION	الأذية 'harm', مصدر 'source', قانون 'law', الإسلام 'Islam', الحل 'solution', حضور 'presence', etc.
	PHYSICAL OBJECT/ARTIFACT	الأموال 'money', منح 'grants', القمح 'wheat', البضائع 'merchandise', مروحية 'helicopter', etc.
	SENSE	صوت 'voice/sound'
	STATE	التوازن 'balance', انتماء 'belonging', مرحلة 'phase', الموت 'the death', etc.
	SUBSTANCE	حرائق 'fires', مطر 'rain', رياح 'winds', etc.
	TIME	يوم 'day', السنة 'the year', موسم 'season', الغد 'tomorrow', دقيقة 'minute', وقت 'time', ثلاثة شهور 'three months', etc.
GOAL PHRASE	YES	كان يذهب إلى المقابر كل يوم AUX <i>dahaba</i> .IMPF.3SG.M ALL ART=graveyards everyday 'He used to go to graveyards everyday' مساعداتنا تذهب إلى الشيشان aid.CL.1PL.GEN <i>dahaba</i> .IMPF.3SG.F ALL ART=Chechnya 'Our aid goes to Chechnya'
SOURCE PHRASE	YES	تأتي الإعانات الرئيسية من المتبرعين <i>atā</i> .IMPF.3SG.F ART=aids ART=main ABL ART=donors 'The main financial aids come from the donors' الهجرات الجنوبية التي قدمت من الهند ART=immigrations ART=southern RP <i>qadima</i> .PERF.3SG.F ABL ART=India 'The southern immigrations that came from India...'
MANNER PHRASE	YES	هذه الجهود لم تذهب هدرا DEM ART=efforts NEG <i>dahaba</i> .JUSS.3SG.F vain.ADV 'These efforts weren't in vain' وجاء حفل الافتتاح بسيطا وجميلا CONJ= <i>gā'a</i> .PERF.3SG.M party ART=opening simple.ADV CONJ=beautiful.ADV 'And the opening ceremony was simple and beautiful'

SETTING PHRASE	YES	وتأتي هذه العمليات الهجومية في ظل زيارة نائب الرئيس CONJ= <i>atā</i> .IMPF.3SG.F DEM ART=operations ART=attack.ADJ LOC shadow visit vice ART=president 'These attacking operations coincide with the visit of the vice president' بل تأتي في اطار مخطط شامل CONJ <i>atā</i> .IMPF.3SG.F LOC frame plan comprehensive 'It, however, comes as part of a comprehensive plan'
PATH PHRASE	YES	فتمضي في طريقك حامدا ربك CONJ= <i>maḏā</i> .2SG.M LOC path.CL.2SG.M.GEN thank.AP.SG.M lord.CL.2SG.M 'Going in your path, thanking your God' خسارة أنت على رأسمال البنك deficit <i>atā</i> .PERF.3SG.F LOC capital ART=bank 'A deficit that destroyed the bank's capital'
PURPOSIVE PHRASE	YES	ذهبت لزيارته وسألته <i>ḏahaba</i> .PERF.1SG PURP=visit.CL.3SG.M.ACC CONJ=ask.CL.3SG.M.ACC 'I went to visit him and asked him' الذين قدموا من بيروت للمشاركة في هذه المناسبة RP <i>qadima</i> .PERF.3PL.M ABL Beirut PURP=participate.VN LOC DEM ART=occasion 'Who came from Beirut to participate in this occasion'
COMITATIVE PHRASE	YES	امرأة حضرت مع اولادها لمشاهدة العمل woman <i>hāḏara</i> .PERF.3SG.F COM sons.CL.3SG.F.GEN PURP=watch.VN ART=show 'A woman who came with her kids to watch the show' برنامجكم لم يأتي بجديد show.CL.3PL.M.GEN NEG <i>atā</i> .PERF.3SG.M COM=new 'Your show did not come up with anything new'
TEMPORAL PHRASE	YES	أذهب لتناول آيس كريم في أي وقت <i>ḏahaba</i> .IMPF.1SG PURP=have.VN ice cream LOC any time 'I go to have ice cream at any time' ومنذ سنتين قدمت لزيارة بلدي CONJ=ADV two years <i>qadima</i> .PERF.1SG PURP=visit.VN country.CL.1SG.GEN 'And two years ago I came to visit my country'
DEGREE PHRASE	YES	تأتي دائما عبر عمليات السطو المنتظم <i>atā</i> .IMPF.3SG.F ADV LOC operations burglary ART=organized 'Comes always through operations of organized burglary' يكفي أن تذهب مرة واحدة إلى متحف اللوفر suffice.IMPF.3SG.M TOP <i>ḏahaba</i> .SUBJN.2SG.M time one ALL museum ART=Louvre 'If you go to the Louvre museum only one time, it would be enough to...'

Appendix E

Samples of sentence annotation from the GO and COME data frames

Dahaba

- (1) ولم يذهب إلى النادي ليمارس الرياضة وبينى عضلات ذراعيه وصدره
wa=lam yaḏhab ilā al=nādi li=yumāris al=riyaḏa
 CONJ=NEG *ḏahaba*.JUSS.3SG.M ALL ART=gym PUR=practice.SUBJN.3SG.M the=sports
 and did not go to the gym to practice the sports
- wa=yabni 'aḏalāt ḏirā'ay-h wa=ṣadri-h*
 CONJ=build.SUBJN.3SG.M muscles arms-CL.3SG.M CONJ=chest.CL.3SG.M
 and build muscles his arms and his chest
 'And he didn't go to them gym to work out and build his arm and chest muscles'

GENRE MODERN_LIT	VERB ḏahaba	TENSE IRREALIS	ASPECT NON-FINITE
MORPH_ASP/MOOD JUSSIVE	SUBJ_NUM SINGULAR	SUBJ_PER 3RD	SUBJ_GEN MASCULINE
SUBJ_CAT HUMAN	INTEROG NO	NEGATION NO	SVC YES
SOURCE NO	MANNER NO	SETTING NO	PATH NO
PP NO	LOC_ADV NO	ADVERBIAL NO	GOAL YES
PURPOSIVE YES	COMITATIVE NO	TEMPORAL NO	DEGREE NO

Maḏā

- (2) وهي تمضي بسرعة في مؤامراتها
wa=hiya tamḏī bi=sur'a fī mu'amarāti-ha
 CONJ=PP *maḏā*.IMPF.3SG.F INST=speed LOC conspiracies-CL.3SG.F.GEN
 and she goes quickly in her conspiracies
 'And it's [i.e. Israel] quickly going ahead with its conspiracies'

GENRE NEWS	VERB maḏā	TENSE PRESENT	ASPECT SIMPLE
MORPH_ASP/MOOD IMPERFECTIVE	SUBJ_NUM SINGULAR	SUBJ_PER 3RD	SUBJ_GEN FEMININE
SUBJ_CAT GROUP	INTEROG NO	NEGATION NO	SVC NO
PP YES	LOC_ADV NO	ADVERBIAL YES	GOAL NO
SOURCE NO	MANNER YES	SETTING YES	PATH NO
PURPOSIVE NO	COMITATIVE NO	TEMPORAL NO	DEGREE NO

Rāḥa

(3) وراحت منذ ذلك الحين تضيق الخناق عليه أكثر وأكثر

wa=rāḥat *munḍu* *ḍālika* *al=ḥīn* *tuḍayyiq*
 CONJ=**rāḥa**.PERF.3SG.F ADV DEM ART=time tighten.IMPF.3SG.F
 and went since that the time tightening

al=ḥināq *'al=ayh* *aḳtar* *wa=aḳtar*
 ART=grip LOC=CL.3S.M more CONJ=more
 the grip on him more and more

‘And since then it [i.e. Washington] kept tightening the grip on him more and more’

GENRE NEWS	VERB rāḥa	TENSE PAST	ASPECT INCEPTIVE
MORPH_ASP/MOOD PERFECTIVE	SUBJ_NUM SINGULAR	SUBJ_PER 3RD	SUBJ_GEN FEMININE
SUBJ_CAT GROUP	INTEROG NO	NEGATION NO	SVC YES
PP NO	LOC_ADV NO	ADVERBIAL NO	GOAL NO
SOURCE NO	MANNER NO	SETTING NO	PATH NO
PURPOSIVE NO	COMITATIVE NO	TEMPORAL NO	DEGREE NO

Atā

(4) وكان طلاب العلم من أنحاء الدنيا يأتون الى جامعاتنا ليتعلموا الطب والهندسة والفلك

wa=kāna *ṭullāb* *al='ilm* *min* *anḥā'* *al=dunyā*
 CONJ=be.PERF.3SG.M students ART=knowledge ABL parts ART=world
 and was students the knowledge from parts the world

ya'tun *ilā* *ya'tun* *ilā* *ḡāmi 'āti-na* *li=yata'allamū*
atā.IMPF.3PL.M ALL **atā**.IMPF.3PL.M ALL universities-CL.1PL PURP=learn.IMPF.3PL.M
 come to come to our universities to learn

al=ṭib *wa=l=handasa* *wa=l=falak*
 ART=medicine CONJ=ART=geometry CONJ=ART=astronomy
 the medicine and the geometry and the astronomy

‘and seekers of knowledge came to our universities from all over the world to learn medicine, geometry and astronomy’

GENRE NEWS	VERB atā	TENSE PAST	ASPECT HABITUAL	MORPH_ASP/ MOOD IMPERFECTIVE
TRANSITIVITY NO	SUBJ_NUM PLURAL	SUBJ_PER 3RD	SUBJ_GEN MASCULINE	SUBJ_CAT HUMAN
INTEROG NO	NEGATION NO	SVC YES	PP YES	LOC_ADV NO
ADVERBIAL NO	GOAL YES	SOURCE NO	MANNER NO	SETTING NO
PATH NO	PURPOSIVE YES	COMITATIVE NO	TEMPORAL NO	DEGREE NO

Ġā'a

(5) وجاءت في المرتبة الثالثة دولة الإمارات العربية

wa=ġā'at fi al=martaba al=tālīta dawlat
 CONJ=ġā'a.PERF.3SG.F LOC ART=place ART=third country
 and came in the place the third country

al='imārāt-i al='arabiyya
 ART=Emirates-GEN ART=Arab
 of the Emirates the Arab
 'And in third place came the United Arab Emirates'

GENRE NEWS	VERB ġā'a	TENSE PAST	ASPECT SIMPLE	MORPH_ASP/ MOOD PERFECTIVE
TRANSITIVITY NO	SUBJ_NUM SINGULAR	SUBJ_PER 3RD	SUBJ_GEN FEMININE	SUBJ_CAT GROUP
INTEROG NO	NEGATION NO	SVC NO	PP YES	LOC_ADV NO
ADVERBIAL NO	GOAL YES	SOURCE NO	MANNER NO	SETTING NO
PATH NO	PURPOSIVE NO	COMITATIVE NO	TEMPORAL NO	DEGREE NO

Ḥaḍara

(6) إن السلطات الأمنية قد حضرت ليلة الإثنين الماضي ومزقت خيمة المعتصمات

inna al=suluḡāt al=amniyya qad ḥaḍarat laylat
 TOP ART=authorities ART=security.ADJ DM ḥaḍara.PERF.3SG.F night
 that the authorities the security-related had come night

al='iṅayn-i al=māḍi wa=mazzaqat ḥaymat al=mu'taṣimāt
 ART=Monday-GEN ART=past CONJ=tear.down.PERF.3SG.F tent ART=protestors.FEM
 of the Monday the last and tore down tent the female protestors

‘That security forces had come last Monday night and tore down the female protestors tent’

GENRE NEWS	VERB ḥaḍara	TENSE PAST	ASPECT SIMPLE	MORPH_ASP/ MOOD PERF
TRANSITIVITY NO	SUBJ_NUM SING	SUBJ_PER 3RD	SUBJ_GEN FEM	SUBJ_CAT GROUP
INTEROG NO	NEGATION NO	SVC NO	PP NO	LOC_ADV NO
ADVERBIAL YES	GOAL NO	SOURCE NO	MANNER NO	SETTING NO
PATH NO	PURPOSIVE NO	COMITATIVE NO	TEMPORAL YES	DEGREE NO

Qadima

(7) الجمهور الذي لبى الدعوة [...] قدم الى القاعة من أمكنة مختلفة في لبنان ومن أزمنة مختلفة

al=ḡumhūr allaḍī labbā al=da'wa qadima ilā
 ART=audience RP answer.PERF.3SG.M ART=invitation *qadima*.PERF.3SG.M ALL
 the audience who answered the invitation came to

al=qā'a min amkina muxtalifa fi lubnān wa=min azmina muxtalifa
 ART=hall ABL places different LOC Lebanon CONJ=ABL times different
 the hall from places different in Lebanon and from times different
 ‘The audience who accepted the invitation [...] came to the hall from different parts of Lebanon and from different generations’

GENRE NEWS	VERB qadima	TENSE PAST	ASPECT SIMPLE	MORPH_ASP/ MOOD PERFECTIVE
TRANSITIVITY NO	SUBJ_NUM SINGULAR	SUBJ_PER 3RD	SUBJ_GEN MASCULINE	SUBJ_CAT GROUP
INTEROG NO	NEGATION NO	SVC NO	PP YES	LOC_ADV NO
ADVERBIAL NO	GOAL YES	SOURCE YES	MANNER NO	SETTING NO
PATH NO	PURPOSIVE NO	COMITATIVE NO	TEMPORAL YES	DEGREE NO

Appendix F

R commands and results of *standardized Pearson's residuals* for GO and COME data frames

The following are the direct results obtained from running a number of commands on a logical form of GO and COME data frames in R. The objective of this series of commands is to calculate the *standardized Pearson's residuals* for each logical variable per verb. The results indicate that the variable occurs significantly higher than expected (+), significantly lower than expected (-), or that the observed frequency is more or less close to the expected frequency (0). Note that these commands are part of the {polytomous} package (Arppe, 2012), and therefore this package needs to be installed prior to running the following commands. Note, also, that the verbs have been dubbed as *dahaba* = VHB, *maḍā* = MDE, *rāḥa* = RAH, *atā* = ATE, *ḥaḍara* = HDR, *ḡā'a* = JAC, *qadima* = QDM (for ease of coding).

GO verbs

```

> library(polytomous)
> GO.univariate <- NULL
> for(i in 2:97) { y <- table(GO.logical[,i],GO.logical$VERB); GO.univariate <-
rbind(GO.univariate,unlist(c(colnames(GO.logical)[i],length(which(GO.logical[,i])),
chisq.test(y)$p.value,associations(y)[c("uc.CR","uc.RC")],as.character(unlist(chisq.posthoc(y)
$cells$std.pearson.residuals.sign[2,])))))}
There were 39 warnings (use warnings() to see them)
> GO.univariate <- data.frame(GO.univariate,stringsAsFactors=FALSE)
> colnames(GO.univariate) <- c("feature","N","X2","uc.CR","uc.RC","MDE","RAH","VHB")
> GO.univariate$N <- as.numeric(GO.univariate$N)
> GO.univariate$X2 <- as.numeric(GO.univariate$X2)
> GO.univariate$uc.RC <- as.numeric(GO.univariate$uc.RC)
> GO.univariate$uc.CR <- as.numeric(GO.univariate$uc.CR)
> GO.univariate

```

	feature	N	X2	uc.CR	uc.RC	MDE	RAH	VHB
1	TENSE.FUT	40	3.195994e-05	1.003900e-02	8.969774e-02	+	-	0
2	TENSE.IRR	176	5.945129e-24	4.382080e-02	1.331444e-01	0	-	+
3	TENSE.PAST	1016	3.030202e-59	9.612773e-02	1.679333e-01	+	-	+
4	TENSE.PRES	268	1.035967e-20	3.231027e-02	7.562646e-02	+	-	+
6	ASPECT.HAB	69	3.580939e-21	2.822584e-02	1.662119e-01	-	-	+
7	ASPECT.INCP	464	2.227615e-199	3.086776e-01	5.482308e-01	-	+	-
8	ASPECT.NON-FIN	164	2.193731e-23	4.337238e-02	1.380666e-01	0	-	+
9	ASPECT.PERT	11	5.848405e-02	2.746751e-03	6.960712e-02	0	-	0
10	ASPECT.PROG	4	3.668971e-01	9.858046e-04	5.864212e-02	0	0	0
11	ASPECT.SIMPLE	778	3.948404e-80	1.179540e-01	1.871408e-01	-	+	-
12	MORPH_ASP.MOOD.IMPF	337	1.162586e-32	5.249418e-02	1.082508e-01	+	-	+
13	MORPH_ASP.MOOD.IMPR	11	3.375978e-02	2.973004e-03	7.534074e-02	0	-	+
14	MORPH_ASP.MOOD.JUSS	52	6.313960e-06	1.047282e-02	7.639454e-02	+	-	+
15	MORPH_ASP.MOOD.PERF	1011	6.756753e-63	1.024398e-01	1.782669e-01	+	-	+
16	MORPH_ASP.MOOD.SUBJN	89	1.976058e-13	2.510910e-02	1.225304e-01	0	-	+
17	SUBJ_NUM.DUAL	11	1.755146e-01	1.223591e-03	3.100777e-02	0	0	0
18	SUBJ_NUM.PL	144	3.871864e-03	3.458902e-03	1.201753e-02	-	0	+
19	SUBJ_NUM.SING	1345	1.631858e-03	4.076341e-03	1.347487e-02	-	0	+
20	SUBJ_PER.1ST	147	2.519879e-05	6.271934e-03	2.148784e-02	-	0	+
21	SUBJ_PER.2ND	30	4.285505e-04	4.422814e-03	4.956142e-02	0	-	+
22	SUBJ_PER.3RD	1323	3.594342e-08	9.961586e-03	3.015510e-02	-	-	+
23	SUBJ_GEN.FEM	417	6.916466e-01	2.249041e-04	4.180366e-04	0	0	0
24	SUBJ_GEN.MASC	945	3.194543e-01	6.920460e-04	1.153780e-03	0	0	0
25	SUBJ_GEN.NIL	138	9.424232e-04	4.118536e-03	1.473160e-02	-	0	+
26	SUBJ_CAT.ACTIVITY	49	7.135110e-01	2.000653e-04	1.527484e-03	0	0	0
27	SUBJ_CAT.ANIMAL	6	2.977589e-02	2.368058e-03	9.976205e-02	0	+	0
28	SUBJ_CAT.ATTRIBUTE	4	3.668971e-01	9.858046e-04	5.864212e-02	0	0	0
29	SUBJ_CAT.BODY	10	6.432396e-04	4.723758e-03	1.295819e-01	-	+	0
30	SUBJ_CAT.COGNITION	6	2.977589e-02	2.368058e-03	9.976205e-02	0	0	+
31	SUBJ_CAT.COMMUNICATION	31	1.251459e-02	2.483599e-03	2.711634e-02	0	0	+
32	SUBJ_CAT.EVENT	4	1.729627e-01	1.304544e-03	7.760281e-02	0	0	0
33	SUBJ_CAT.GROUP	115	2.124086e-02	2.306418e-03	9.365525e-03	0	+	-
34	SUBJ_CAT.HUMAN	853	3.680865e-46	6.428365e-02	1.032970e-01	+	-	-
35	SUBJ_CAT.LOCATION	1	3.676341e-01	6.670632e-04	1.322363e-01	0	0	0
36	SUBJ_CAT.NOTION	75	9.587689e-01	2.555705e-05	1.414365e-04	0	0	0
37	SUBJ_CAT.OBJECT	75	6.424430e-09	1.102812e-02	6.103122e-02	-	-	+
38	SUBJ_CAT.SENSE	9	7.150910e-01	2.074193e-04	6.212820e-03	0	0	0
39	SUBJ_CAT.STATE	6	1.342526e-01	1.687349e-03	7.108497e-02	0	0	0
40	SUBJ_CAT.SUBSTANCE	3	4.948865e-02	2.003642e-03	1.525745e-01	0	0	+

41	SUBJ_CAT.TIME	253	3.735438e-122	1.774984e-01	4.297413e-01	+	-	-
42	INTEROG.NO	1476	6.413813e-05	5.586770e-03	7.481893e-02	0	-	+
43	INTEROG.YES	24	6.413813e-05	5.586770e-03	7.481893e-02	0	-	+
44	NEGATION.NO	1427	8.703777e-08	1.375494e-02	7.766515e-02	+	-	+
45	NEGATION.YES	73	8.703777e-08	1.375494e-02	7.766515e-02	+	-	+
46	SV_ORDER.NILL	1437	2.095265e-12	2.039644e-02	1.285963e-01	0	-	+
50	SVC.NO	987	4.026507e-179	2.655280e-01	4.541308e-01	-	+	-
51	SVC.YES	513	4.026507e-179	2.655280e-01	4.541308e-01	-	+	-
52	PP.NO	921	1.528593e-113	1.961267e-01	3.230755e-01	+	-	+
53	PP.YES	579	1.528593e-113	1.961267e-01	3.230755e-01	+	-	+
54	LOC_ADV.NO	1440	1.385315e-06	1.003120e-02	6.561941e-02	0	-	+
55	LOC_ADV.YES	60	1.385315e-06	1.003120e-02	6.561941e-02	0	-	+
77	ADVERBIAL.NO	1263	3.353650e-03	3.376133e-03	8.500421e-03	-	0	+
78	ADVERBIAL.YES	237	3.353650e-03	3.376133e-03	8.500421e-03	-	0	+
79	GOAL.NO	1169	2.377427e-135	1.991571e-01	4.145831e-01	-	-	+
80	GOAL.YES	331	2.377427e-135	1.991571e-01	4.145831e-01	-	-	+
81	SOURCE.NO	1490	5.967799e-02	2.594047e-03	7.115979e-02	0	-	0
82	SOURCE.YES	10	5.967799e-02	2.594047e-03	7.115979e-02	0	-	0
83	MANNER.NO	1294	6.631307e-01	2.522317e-04	6.926011e-04	0	0	0
84	MANNER.YES	206	6.631307e-01	2.522317e-04	6.926011e-04	0	0	0
85	SETTING.NO	1411	6.220140e-13	2.093623e-02	1.021671e-01	+	-	0
86	SETTING.YES	89	6.220140e-13	2.093623e-02	1.021671e-01	+	-	0
87	PATH.NO	1358	2.713707e-24	3.953322e-02	1.386691e-01	+	-	0
88	PATH.YES	142	2.713707e-24	3.953322e-02	1.386691e-01	+	-	0
89	PURPOSIVE.NO	1421	7.110483e-21	2.958772e-02	1.575689e-01	-	-	+
90	PURPOSIVE.YES	79	7.110483e-21	2.958772e-02	1.575689e-01	-	-	+
91	COMITATIVE.NO	1463	5.661614e-06	8.612008e-03	8.178580e-02	0	-	+
92	COMITATIVE.YES	37	5.661614e-06	8.612008e-03	8.178580e-02	0	-	+
93	TEMPORAL.NO	1416	4.017415e-09	1.648445e-02	8.391397e-02	+	-	+
94	TEMPORAL.YES	84	4.017415e-09	1.648445e-02	8.391397e-02	+	-	+
95	DEGREE.NO	1488	4.859700e-02	1.697180e-03	4.001645e-02	0	0	+
96	DEGREE.YES	12	4.859700e-02	1.697180e-03	4.001645e-02	0	0	+

COME verbs

```

> library(polytomous)
> COME.univariate <- NULL
> for(i in 2:63) { y <- table(COME.logical[,i],COME.logical$VERB); COME.univariate <-
  rbind(COME.univariate,unlist(c(colnames(COME.logical)[i],length(which(COME.logical[,i])),
  chisq.test(y)$p.value,associations(y)[c("uc.CR","uc.RC")],as.character(unlist(chisq.posthoc(y)
  $cells$std.pearson.residuals.sign[2,])))))}
> COME.univariate <- data.frame(COME.univariate,stringsAsFactors=FALSE)
> colnames(COME.univariate) <- c("feature","N","X2","uc.CR","uc.RC","ATE","HDR","JAC","QDM")
> COME.univariate$N <- as.numeric(COME.univariate$N)
> COME.univariate$X2 <- as.numeric(COME.univariate$X2)
> COME.univariate$uc.RC <- as.numeric(COME.univariate$uc.RC)
> COME.univariate$uc.CR <- as.numeric(COME.univariate$uc.CR)
> COME.univariate

```

	feature	N	X2	uc.CR	uc.RC	ATE	HDR	JAC	QDM
1	TENSE.FUT	53	1.084723e-15	0.0141368096	0.160168962	0	+	-	-
2	TENSE.IRR	179	1.283539e-26	0.0238132520	0.109536911	+	+	-	-
3	TENSE.PAST	1396	2.105286e-222	0.1998562585	0.452305741	+	0	-	-
4	TENSE.PRES	372	1.962008e-185	0.1482416276	0.427808252	+	-	-	-
5	ASPECT.HAB	138	5.579550e-60	0.0439475640	0.242683069	+	-	-	-
6	ASPECT.NON-FIN	180	3.916553e-26	0.0232965454	0.106749857	+	+	-	-
7	ASPECT.PERT	17	4.177145e-08	0.0063364066	0.179307626	-	0	-	+
8	ASPECT.PROG	11	1.047871e-02	0.0027775668	0.112914030	0	+	0	0
9	ASPECT.SIMPLE	1654	1.177354e-74	0.0608979409	0.183283525	+	0	-	-
10	MORPH_ASP.MOOD.IMPF	490	7.427262e-216	0.1885128578	0.469371392	+	-	-	-
11	MORPH_ASP.MOOD.IMPR	2	1.113184e-01	0.0010010794	0.175508475	0	+	0	0
12	MORPH_ASP.MOOD.JUSS	40	6.474353e-09	0.0101660131	0.143749633	+	+	-	-
13	MORPH_ASP.MOOD.PERF	1397	1.018406e-254	0.2387098744	0.540607728	+	0	-	-
14	MORPH_ASP.MOOD.SUBJN	71	6.696949e-15	0.0152248189	0.137616175	+	+	-	-
15	TRANSITIVITY.NO	1569	1.652291e-253	0.1977193575	0.525943710	-	+	-	-
16	TRANSITIVITY.YES	431	1.652291e-253	0.1977193575	0.525943710	-	+	-	-
17	SUBJ_NUM.DUAL	24	1.370522e-09	0.0066961184	0.142808012	-	0	-	+
18	SUBJ_NUM.PL	324	3.994229e-112	0.0826509671	0.258659007	-	-	-	+
19	SUBJ_NUM.SING	1652	5.690638e-125	0.0926152976	0.277801253	-	-	-	+
20	SUBJ_PER.1ST	59	9.048347e-02	0.0012232833	0.012750626	0	0	-	0
21	SUBJ_PER.2ND	15	6.616337e-02	0.0012363385	0.038804598	+	0	0	0
22	SUBJ_PER.3RD	1926	2.347946e-02	0.0018545069	0.016241673	0	0	-	0
23	SUBJ_GEN.FEM	447	2.033947e-41	0.0348505265	0.090933343	+	-	+	-
24	SUBJ_GEN.MASC	1489	2.651641e-37	0.0308959115	0.075366947	+	-	+	-
25	SUBJ_GEN.NIL	64	5.782448e-02	0.0013901198	0.013606970	0	+	-	0
26	SUBJ_CAT.ACTIVITY	184	1.682281e-38	0.0410357975	0.185216970	+	-	+	-
27	SUBJ_CAT.ANIMAL	3	2.991085e-01	0.0008122687	0.100072198	0	0	0	0
28	SUBJ_CAT.ATTRIBUTE	11	1.015640e-05	0.0043171207	0.175500193	+	0	0	0

29	SUBJ_CAT.BODY	2	1.113184e-01	0.0010010794	0.175508475	+	0	0	0
30	SUBJ_CAT.COGNITION	6	1.107351e-01	0.0015032527	0.102039772	0	0	0	0
31	SUBJ_CAT.COMMUNICATION	111	1.438955e-23	0.0257490024	0.166489783	+	-	+	-
32	SUBJ_CAT.content	97	2.308254e-57	0.0438024127	0.312884915	-	-	+	-
33	SUBJ_CAT.demonstrative	48	2.700722e-14	0.0138382783	0.169435639	0	-	+	-
34	SUBJ_CAT.EVENT	58	8.769588e-15	0.0144527165	0.152654613	+	-	0	-
35	SUBJ_CAT.GROUP	118	4.388421e-02	0.0015421329	0.009535159	0	+	-	0
36	SUBJ_CAT.HUMAN	1103	4.572411e-214	0.1986206649	0.400310254	+	-	+	-
37	SUBJ_CAT.LOCATION	5	1.439027e-01	0.0012887497	0.102233578	0	0	0	0
38	SUBJ_CAT.NOTION	139	4.132812e-25	0.0262563481	0.144243865	+	-	+	-
39	SUBJ_CAT.OBJECT	34	1.914740e-04	0.0033973846	0.054687320	+	-	0	0
40	SUBJ_CAT.SENSE	10	4.094658e-03	0.0028087462	0.123693280	+	0	0	0
41	SUBJ_CAT.STATE	37	7.648727e-11	0.0105467592	0.158677146	+	-	0	-
42	SUBJ_CAT.SUBSTANCE	7	2.654461e-03	0.0024735434	0.147256246	+	0	0	0
43	SUBJ_CAT.TIME	27	1.748707e-05	0.0056290553	0.109099279	+	-	0	-
44	INTEROG.NO	1982	1.796164e-03	0.0026756160	0.072227678	+	0	0	0
45	INTEROG.YES	18	1.796164e-03	0.0026756160	0.072227678	+	0	0	0
46	NEGATION.NO	1936	1.419533e-13	0.0148269135	0.145130917	+	+	-	-
47	NEGATION.YES	64	1.419533e-13	0.0148269135	0.145130917	+	+	-	-
53	PP.NO	789	1.442884e-133	0.1257489669	0.259907974	0	+	-	-
54	PP.YES	1211	1.442884e-133	0.1257489669	0.259907974	0	+	-	-
55	LOC_ADV.NO	1802	1.652948e-07	0.0070261980	0.030167162	+	-	+	0
56	LOC_ADV.YES	198	1.652948e-07	0.0070261980	0.030167162	+	-	+	0
57	ADVERBIAL.NO	1780	2.207618e-10	0.0086768853	0.034713376	0	-	+	-
58	ADVERBIAL.YES	220	2.207618e-10	0.0086768853	0.034713376	0	-	+	-
59	GOAL.NO	1182	1.764591e-135	0.1213926427	0.248762451	-	+	-	+
60	GOAL.YES	818	1.764591e-135	0.1213926427	0.248762451	-	+	-	+
61	SOURCE.NO	1637	1.140876e-91	0.0712933637	0.208659173	-	-	-	+
62	SOURCE.YES	363	1.140876e-91	0.0712933637	0.208659173	-	-	-	+

Appendix G
Sample of the entire hierarchical configural frequency analysis table

The following table is obtained through conducting an HCFA analysis on the variables SUBJECT NUMBER, SUBJECT PERSON, SUBJECT GENDER and SUBJECT CATEGORY for the three GO verbs *dahaba* = VHB, *maḍā* = MDE and *rāḥa* = RAH.

VERB	SUBJ_ NUM	SUBJ_ PER	SUBJ_ GEN	SUBJ_ CAT	Freq	Exp	Cont.chisq	Obs-exp	P.adj.Holm	Dec	Q
MDE	SING	3RD	MASC	TIME	178	41.5293	448.4607	>	3.86E-55	***	0.094
VHB	SING	1ST	NILL	HUMAN	47	2.5015	791.5854	>	1.10E-39	***	0.03
VHB	PL	1ST	NILL	HUMAN	26	0.2678	2472.4039	>	2.50E-39	***	0.017
MDE	PL	1ST	NILL	HUMAN	18	0.2678	1174.0609	>	6.68E-24	***	0.012
RAH	SING	1ST	NILL	HUMAN	34	2.5015	396.6314	>	9.05E-24	***	0.021
MDE	SING	3RD	FEM	TIME	67	18.497	127.1852	>	1.50E-15	***	0.033
RAH	SING	3RD	FEM	GROUP	43	8.3346	144.1806	>	1.58E-14	***	0.023
MDE	SING	3RD	FEM	HUMAN	12	62.4365	40.7428	<	3.47E-12	***	0.035
VHB	SING	3RD	FEM	OBJECT	33	5.7026	130.6672	>	3.89E-12	***	0.018
VHB	SING	3RD	MASC	TIME	3	41.5293	35.746	<	8.60E-12	***	0.026
RAH	SING	3RD	MASC	TIME	3	41.5293	35.746	<	8.60E-12	***	0.026
RAH	SING	3RD	MASC	HUMAN	233	140.1819	61.4573	>	2.36E-11	***	0.068
RAH	PL	1ST	NILL	HUMAN	9	0.2678	284.7161	>	1.81E-08	***	0.006
VHB	SING	3RD	NILL	HUMAN	0	22.5131	22.5131	<	1.69E-07	***	0.015
RAH	SING	3RD	NILL	HUMAN	0	22.5131	22.5131	<	1.69E-07	***	0.015
MDE	SING	3RD	NILL	HUMAN	0	22.5131	22.5131	<	1.69E-07	***	0.015
RAH	SING	3RD	FEM	TIME	0	18.497	18.497	<	9.90E-06	***	0.012
VHB	SING	1ST	MASC	HUMAN	0	15.5758	15.5758	<	0.000189784	***	0.01
RAH	SING	1ST	MASC	HUMAN	0	15.5758	15.5758	<	0.000189784	***	0.01
MDE	SING	1ST	MASC	HUMAN	0	15.5758	15.5758	<	0.000189784	***	0.01
VHB	SING	3RD	FEM	TIME	1	18.497	16.551	<	0.000194702	***	0.012
VHB	SING	3RD	FEM	HUMAN	27	62.4365	20.1123	<	0.000279774	***	0.025
RAH	PL	3RD	MASC	HUMAN	38	15.0083	35.2216	>	0.000469411	***	0.015
MDE	SING	3RD	FEM	GROUP	26	8.3346	37.4422	>	0.000807962	***	0.012
VHB	SING	3RD	FEM	COMMUNICATION	13	2.2664	50.8331	>	0.000948432	***	0.007
VHB	SING	3RD	FEM	NOTION	18	5.4833	28.5718	>	0.020585197	*	0.008
RAH	DUAL	3RD	FEM	BODY	2	0.006	664.9799	>	0.021156092	*	0.001
VHB	SING	3RD	MASC	HUMAN	189	140.1819	17.0008	>	0.024091839	*	0.036
MDE	SING	3RD	MASC	HUMAN	97	140.1819	13.3018	<	0.042124462	*	0.032
MDE	SING	1ST	NILL	HUMAN	11	2.5015	28.8732	>	0.071869504	ms	0.006
VHB	SING	3RD	FEM	ACTIVITY	13	3.5824	24.7572	>	0.109911935	ns	0.006
MDE	SING	3RD	FEM	ACTIVITY	13	3.5824	24.7572	>	0.109911935	ns	0.006
VHB	SING	2ND	MASC	HUMAN	12	3.1787	24.4799	>	0.140483063	ns	0.006
VHB	PL	3RD	MASC	HUMAN	31	15.0083	17.0395	>	0.216206396	ns	0.011
RAH	SING	3RD	FEM	NOTION	16	5.4833	20.1705	>	0.22079271	ns	0.007
RAH	SING	3RD	FEM	HUMAN	37	62.4365	10.3628	<	0.328032327	ns	0.018
VHB	DUAL	2ND	MASC	HUMAN	2	0.026	149.8898	>	0.391320601	ns	0.001
RAH	SING	3RD	FEM	ACTIVITY	12	3.5824	19.7787	>	0.408910179	ns	0.006
VHB	SING	3RD	FEM	GROUP	20	8.3346	16.3273	>	0.473957245	ns	0.008
VHB	PL	2ND	MASC	HUMAN	4	0.3403	39.3543	>	0.499775569	ns	0.002
MDE	SING	3RD	FEM	NOTION	15	5.4833	16.517	>	0.663068693	ns	0.006
VHB	SING	3RD	MASC	GROUP	6	18.7128	8.6366	<	0.705745473	ns	0.009
MDE	PL	1ST	NILL	GROUP	2	0.0358	107.9226	>	0.731507041	ns	0.001
VHB	SING	1ST	FEM	HUMAN	0	6.9374	6.9374	<	1.118665466	ns	0.005
RAH	SING	1ST	FEM	HUMAN	0	6.9374	6.9374	<	1.118665466	ns	0.005
MDE	SING	1ST	FEM	HUMAN	0	6.9374	6.9374	<	1.118665466	ns	0.005
RAH	SING	3RD	MASC	OBJECT	3	12.8035	7.5064	<	1.385836439	ns	0.007

MDE	SING	3RD	MASC	OBJECT	3	12.8035	7.5064	<	1.385836439	ns	0.007
RAH	PL	3RD	FEM	HUMAN	0	6.6846	6.6846	<	1.436452353	ns	0.004
MDE	PL	3RD	FEM	HUMAN	0	6.6846	6.6846	<	1.436452353	ns	0.004
VHB	SING	3RD	NILL	TIME	0	6.6696	6.6696	<	1.455223847	ns	0.004
RAH	SING	3RD	NILL	TIME	0	6.6696	6.6696	<	1.455223847	ns	0.004
MDE	SING	3RD	NILL	TIME	0	6.6696	6.6696	<	1.455223847	ns	0.004
RAH	SING	3RD	MASC	GROUP	7	18.7128	7.3313	<	2.02237919	ns	0.008
RAH	SING	3RD	FEM	OBJECT	14	5.7026	12.0727	>	2.645640457	ns	0.006
VHB	SING	3RD	MASC	ACTIVITY	1	8.0432	6.1676	<	3.313849284	ns	0.005
VHB	SING	3RD	MASC	COGNITION	5	0.9849	16.3686	>	3.962476365	ns	0.003
MDE	SING	3RD	MASC	GROUP	8	18.7128	6.1329	<	5.125771015	ns	0.007
RAH	SING	3RD	FEM	BODY	4	0.7311	14.6158	>	7.711351464	ns	0.002
VHB	PL	3RD	FEM	HUMAN	1	6.6846	4.8342	<	10.98186411	ns	0.004
VHB	SING	1ST	MASC	TIME	0	4.6144	4.6144	<	11.35334003	ns	0.003
RAH	SING	1ST	MASC	TIME	0	4.6144	4.6144	<	11.35334003	ns	0.003
MDE	SING	1ST	MASC	TIME	0	4.6144	4.6144	<	11.35334003	ns	0.003
VHB	PL	3RD	MASC	TIME	0	4.4463	4.4463	<	13.40362493	ns	0.003
RAH	PL	3RD	MASC	TIME	0	4.4463	4.4463	<	13.40362493	ns	0.003
MDE	PL	3RD	MASC	TIME	0	4.4463	4.4463	<	13.40362493	ns	0.003
RAH	SING	3RD	MASC	ACTIVITY	2	8.0432	4.5405	<	15.08575876	ns	0.004
RAH	SING	3RD	MASC	COMMUNICATION	1	5.0886	3.2851	<	42.85450347	ns	0.003
VHB	DUAL	3RD	FEM	OBJECT	1	0.0466	19.4881	>	52.26697465	ns	0.001
VHB	SING	3RD	NILL	GROUP	0	3.0053	3.0053	<	56.53598531	ns	0.002
RAH	SING	3RD	NILL	GROUP	0	3.0053	3.0053	<	56.53598531	ns	0.002
MDE	SING	3RD	NILL	GROUP	0	3.0053	3.0053	<	56.53598531	ns	0.002
MDE	PL	2ND	NILL	HUMAN	1	0.0547	16.3509	>	60.79633912	ns	0.001
VHB	DUAL	3RD	MASC	ACTIVITY	1	0.0658	13.2677	>	72.70594241	ns	0.001
RAH	SING	3RD	FEM	LOCATION	1	0.0731	11.751	>	80.44466088	ns	0.001
RAH	SING	3RD	FEM	ANIMAL	2	0.4387	5.5573	>	82.25590843	ns	0.001
RAH	SING	3RD	FEM	STATE	2	0.4387	5.5573	>	82.25590843	ns	0.001
VHB	SING	3RD	MASC	STATE	3	0.9849	4.123	>	88.13547985	ns	0.001
RAH	SING	3RD	MASC	ANIMAL	3	0.9849	4.123	>	88.13547985	ns	0.001
RAH	SING	3RD	FEM	COMMUNICATION	5	2.2664	3.297	>	90.56828499	ns	0.002
VHB	PL	3RD	NILL	HUMAN	0	2.4103	2.4103	<	101.6645903	ns	0.002
MDE	PL	3RD	NILL	HUMAN	0	2.4103	2.4103	<	101.6645903	ns	0.002
VHB	DUAL	2ND	NILL	ACTIVITY	0	2.00E-04	2.00E-04	<	108.4740113	ns	0
RAH	DUAL	2ND	NILL	ACTIVITY	0	2.00E-04	2.00E-04	<	108.4740113	ns	0
MDE	DUAL	2ND	NILL	ACTIVITY	0	2.00E-04	2.00E-04	<	108.4740113	ns	0
VHB	SING	3RD	MASC	OBJECT	18	12.8035	2.1091	>	111.1822442	ns	0.003
RAH	DUAL	3RD	MASC	HUMAN	3	1.1465	2.9967	>	123.4427141	ns	0.001
MDE	PL	3RD	MASC	HUMAN	10	15.0083	1.6713	<	132.1489294	ns	0.003
VHB	SING	1ST	MASC	GROUP	0	2.0792	2.0792	<	140.955604	ns	0.001
RAH	SING	1ST	MASC	GROUP	0	2.0792	2.0792	<	140.955604	ns	0.001
MDE	SING	1ST	MASC	GROUP	0	2.0792	2.0792	<	140.955604	ns	0.001
VHB	SING	1ST	FEM	TIME	0	2.0552	2.0552	<	143.6139958	ns	0.001
RAH	SING	1ST	FEM	TIME	0	2.0552	2.0552	<	143.6139958	ns	0.001
MDE	SING	1ST	FEM	TIME	0	2.0552	2.0552	<	143.6139958	ns	0.001
VHB	SING	3RD	NILL	OBJECT	0	2.0562	2.0562	<	143.8513034	ns	0.001
RAH	SING	3RD	NILL	OBJECT	0	2.0562	2.0562	<	143.8513034	ns	0.001
MDE	SING	3RD	NILL	OBJECT	0	2.0562	2.0562	<	143.8513034	ns	0.001
RAH	PL	3RD	MASC	GROUP	0	2.0035	2.0035	<	150.7836287	ns	0.001
MDE	PL	3RD	MASC	GROUP	0	2.0035	2.0035	<	150.7836287	ns	0.001
VHB	SING	3RD	MASC	NOTION	8	12.3111	1.5096	<	150.9305486	ns	0.003
VHB	PL	3RD	FEM	TIME	0	1.9803	1.9803	<	153.9683439	ns	0.001
RAH	PL	3RD	FEM	TIME	0	1.9803	1.9803	<	153.9683439	ns	0.001
MDE	PL	3RD	FEM	TIME	0	1.9803	1.9803	<	153.9683439	ns	0.001
VHB	SING	3RD	NILL	NOTION	0	1.9772	1.9772	<	154.0470744	ns	0.001

RAH	SING	3RD	NILL	NOTION	0	1.9772	1.9772	<	154.0470744	ns	0.001
MDE	SING	3RD	NILL	NOTION	0	1.9772	1.9772	<	154.0470744	ns	0.001
MDE	SING	3RD	MASC	EVENT	2	0.6566	2.7487	>	156.612769	ns	0.001
RAH	DUAL	3RD	MASC	GROUP	1	0.153	4.6872	>	157.6648909	ns	0.001
VHB	DUAL	2ND	FEM	ACTIVITY	0	7.00E-04	7.00E-04	<	189.3741442	ns	0
RAH	DUAL	2ND	FEM	ACTIVITY	0	7.00E-04	7.00E-04	<	189.3741442	ns	0
MDE	DUAL	2ND	FEM	ACTIVITY	0	7.00E-04	7.00E-04	<	189.3741442	ns	0
VHB	SING	3RD	MASC	SENSE	3	1.4773	1.5694	>	205.7227536	ns	0.001
RAH	SING	3RD	MASC	SENSE	3	1.4773	1.5694	>	205.7227536	ns	0.001
VHB	PL	1ST	MASC	HUMAN	0	1.6676	1.6676	<	208.6984307	ns	0.001
RAH	PL	1ST	MASC	HUMAN	0	1.6676	1.6676	<	208.6984307	ns	0.001
MDE	PL	1ST	MASC	HUMAN	0	1.6676	1.6676	<	208.6984307	ns	0.001
VHB	SING	3RD	MASC	BODY	0	1.6415	1.6415	<	213.7432129	ns	0.001
MDE	SING	3RD	MASC	BODY	0	1.6415	1.6415	<	213.7432129	ns	0.001
RAH	SING	3RD	MASC	NOTION	9	12.3111	0.8905	<	237.2283869	ns	0.002
MDE	SING	3RD	MASC	NOTION	9	12.3111	0.8905	<	237.2283869	ns	0.002
VHB	DUAL	1ST	NILL	ACTIVITY	0	0.0012	0.0012	<	238.719621	ns	0
RAH	DUAL	1ST	NILL	ACTIVITY	0	0.0012	0.0012	<	238.719621	ns	0
MDE	DUAL	1ST	NILL	ACTIVITY	0	0.0012	0.0012	<	238.719621	ns	0
RAH	SING	3RD	MASC	BODY	3	1.6415	1.1244	>	250.3251268	ns	0.001
VHB	SING	1ST	MASC	OBJECT	0	1.4226	1.4226	<	264.7723065	ns	0.001
RAH	SING	1ST	MASC	OBJECT	0	1.4226	1.4226	<	264.7723065	ns	0.001
MDE	SING	1ST	MASC	OBJECT	0	1.4226	1.4226	<	264.7723065	ns	0.001
MDE	SING	2ND	FEM	HUMAN	0	1.4158	1.4158	<	266.100233	ns	0.001
VHB	DUAL	2ND	MASC	ACTIVITY	0	0.0015	0.0015	<	267.1012855	ns	0
RAH	DUAL	2ND	MASC	ACTIVITY	0	0.0015	0.0015	<	267.1012855	ns	0
MDE	DUAL	2ND	MASC	ACTIVITY	0	0.0015	0.0015	<	267.1012855	ns	0
VHB	PL	3RD	MASC	OBJECT	0	1.3708	1.3708	<	276.8400834	ns	0.001
RAH	PL	3RD	MASC	OBJECT	0	1.3708	1.3708	<	276.8400834	ns	0.001
MDE	PL	3RD	MASC	OBJECT	0	1.3708	1.3708	<	276.8400834	ns	0.001
VHB	SING	1ST	MASC	NOTION	0	1.3679	1.3679	<	276.8781304	ns	0.001
RAH	SING	1ST	MASC	NOTION	0	1.3679	1.3679	<	276.8781304	ns	0.001
MDE	SING	1ST	MASC	NOTION	0	1.3679	1.3679	<	276.8781304	ns	0.001
VHB	PL	3RD	MASC	NOTION	0	1.3181	1.3181	<	290.2361877	ns	0.001
RAH	PL	3RD	MASC	NOTION	0	1.3181	1.3181	<	290.2361877	ns	0.001
MDE	PL	3RD	MASC	NOTION	0	1.3181	1.3181	<	290.2361877	ns	0.001
VHB	SING	3RD	NILL	ACTIVITY	0	1.2917	1.2917	<	297.1603655	ns	0.001
RAH	SING	3RD	NILL	ACTIVITY	0	1.2917	1.2917	<	297.1603655	ns	0.001
MDE	SING	3RD	NILL	ACTIVITY	0	1.2917	1.2917	<	297.1603655	ns	0.001
RAH	PL	3RD	NILL	HUMAN	1	2.4103	0.8252	<	329.8177868	ns	0.001
MDE	SING	3RD	MASC	ACTIVITY	6	8.0432	0.519	<	331.1344765	ns	0.001
VHB	DUAL	3RD	MASC	HUMAN	0	1.1465	1.1465	<	341.597719	ns	0.001
MDE	DUAL	3RD	MASC	HUMAN	0	1.1465	1.1465	<	341.597719	ns	0.001
RAH	PL	2ND	NILL	ACTIVITY	0	0.0031	0.0031	<	359.3712416	ns	0
MDE	PL	2ND	NILL	ACTIVITY	0	0.0031	0.0031	<	359.3712416	ns	0
VHB	DUAL	1ST	FEM	ACTIVITY	0	0.0033	0.0033	<	363.8137096	ns	0
RAH	DUAL	1ST	FEM	ACTIVITY	0	0.0033	0.0033	<	363.8137096	ns	0
MDE	DUAL	1ST	FEM	ACTIVITY	0	0.0033	0.0033	<	363.8137096	ns	0
RAH	SING	3RD	MASC	COGNITION	0	0.9849	0.9849	<	399.4969511	ns	0.001
RAH	SING	3RD	MASC	STATE	0	0.9849	0.9849	<	399.4969511	ns	0.001
MDE	SING	3RD	MASC	ANIMAL	0	0.9849	0.9849	<	399.4969511	ns	0.001
MDE	SING	3RD	MASC	COGNITION	0	0.9849	0.9849	<	399.4969511	ns	0.001
MDE	SING	3RD	MASC	STATE	0	0.9849	0.9849	<	399.4969511	ns	0.001
VHB	SING	2ND	MASC	TIME	0	0.9417	0.9417	<	415.5760571	ns	0.001
RAH	SING	2ND	MASC	TIME	0	0.9417	0.9417	<	415.5760571	ns	0.001
MDE	SING	2ND	MASC	TIME	0	0.9417	0.9417	<	415.5760571	ns	0.001
MDE	SING	2ND	MASC	HUMAN	4	3.1787	0.2122	>	417.9219186	ns	0.001

VHB	SING	1ST	FEM	GROUP	0	0.9261	0.9261	<	420.5466649	ns	0.001
RAH	SING	1ST	FEM	GROUP	0	0.9261	0.9261	<	420.5466649	ns	0.001
MDE	SING	1ST	FEM	GROUP	0	0.9261	0.9261	<	420.5466649	ns	0.001
VHB	PL	3RD	MASC	GROUP	1	2.0035	0.5026	<	428.7795673	ns	0.001
VHB	PL	3RD	FEM	GROUP	0	0.8923	0.8923	<	431.7092038	ns	0.001
RAH	PL	3RD	FEM	GROUP	0	0.8923	0.8923	<	431.7092038	ns	0.001
MDE	PL	3RD	FEM	GROUP	0	0.8923	0.8923	<	431.7092038	ns	0.001
VHB	SING	1ST	MASC	ACTIVITY	0	0.8937	0.8937	<	432.3486507	ns	0.001
RAH	SING	1ST	MASC	ACTIVITY	0	0.8937	0.8937	<	432.3486507	ns	0.001
MDE	SING	1ST	MASC	ACTIVITY	0	0.8937	0.8937	<	432.3486507	ns	0.001
VHB	PL	3RD	MASC	ACTIVITY	0	0.8611	0.8611	<	444.3411668	ns	0.001
MDE	PL	3RD	MASC	ACTIVITY	0	0.8611	0.8611	<	444.3411668	ns	0.001
VHB	SING	3RD	NILL	COMMUNICATION	0	0.8172	0.8172	<	462.7533389	ns	0.001
RAH	SING	3RD	NILL	COMMUNICATION	0	0.8172	0.8172	<	462.7533389	ns	0.001
MDE	SING	3RD	NILL	COMMUNICATION	0	0.8172	0.8172	<	462.7533389	ns	0.001
VHB	DUAL	1ST	MASC	ACTIVITY	0	0.0073	0.0073	<	465.5845677	ns	0
RAH	DUAL	1ST	MASC	ACTIVITY	0	0.0073	0.0073	<	465.5845677	ns	0
MDE	DUAL	1ST	MASC	ACTIVITY	0	0.0073	0.0073	<	465.5845677	ns	0
VHB	PL	2ND	FEM	ACTIVITY	0	0.0087	0.0087	<	499.1399481	ns	0
RAH	PL	2ND	FEM	ACTIVITY	0	0.0087	0.0087	<	499.1399481	ns	0
MDE	PL	2ND	FEM	ACTIVITY	0	0.0087	0.0087	<	499.1399481	ns	0
VHB	PL	1ST	NILL	ACTIVITY	0	0.0154	0.0154	<	568.2013173	ns	0
RAH	PL	1ST	NILL	ACTIVITY	0	0.0154	0.0154	<	568.2013173	ns	0
MDE	PL	1ST	NILL	ACTIVITY	0	0.0154	0.0154	<	568.2013173	ns	0
RAH	PL	3RD	MASC	ACTIVITY	1	0.8611	0.0224	>	583.7732576	ns	0
VHB	PL	2ND	MASC	ACTIVITY	0	0.0195	0.0195	<	620.2690074	ns	0
RAH	PL	2ND	MASC	ACTIVITY	0	0.0195	0.0195	<	620.2690074	ns	0
MDE	PL	2ND	MASC	ACTIVITY	0	0.0195	0.0195	<	620.2690074	ns	0
VHB	SING	1ST	FEM	ACTIVITY	0	0.398	0.398	<	658.1632995	ns	0
RAH	SING	1ST	FEM	ACTIVITY	0	0.398	0.398	<	658.1632995	ns	0
MDE	SING	1ST	FEM	ACTIVITY	0	0.398	0.398	<	658.1632995	ns	0
VHB	DUAL	3RD	FEM	ACTIVITY	0	0.0293	0.0293	<	664.250295	ns	0
RAH	DUAL	3RD	FEM	ACTIVITY	0	0.0293	0.0293	<	664.250295	ns	0
MDE	DUAL	3RD	FEM	ACTIVITY	0	0.0293	0.0293	<	664.250295	ns	0
VHB	PL	3RD	FEM	ACTIVITY	0	0.3835	0.3835	<	665.7353798	ns	0
RAH	PL	3RD	FEM	ACTIVITY	0	0.3835	0.3835	<	665.7353798	ns	0
MDE	PL	3RD	FEM	ACTIVITY	0	0.3835	0.3835	<	665.7353798	ns	0
VHB	PL	1ST	FEM	ACTIVITY	0	0.0426	0.0426	<	689.0022957	ns	0
RAH	PL	1ST	FEM	ACTIVITY	0	0.0426	0.0426	<	689.0022957	ns	0
MDE	PL	1ST	FEM	ACTIVITY	0	0.0426	0.0426	<	689.0022957	ns	0
RAH	DUAL	3RD	MASC	ACTIVITY	0	0.0658	0.0658	<	724.2547754	ns	0
MDE	DUAL	3RD	MASC	ACTIVITY	0	0.0658	0.0658	<	724.2547754	ns	0
VHB	PL	1ST	MASC	ACTIVITY	0	0.0957	0.0957	<	747.9018092	ns	0
RAH	PL	1ST	MASC	ACTIVITY	0	0.0957	0.0957	<	747.9018092	ns	0
MDE	PL	1ST	MASC	ACTIVITY	0	0.0957	0.0957	<	747.9018092	ns	0
VHB	SING	2ND	FEM	ACTIVITY	0	0.0812	0.0812	<	749.1054038	ns	0
RAH	SING	2ND	FEM	ACTIVITY	0	0.0812	0.0812	<	749.1054038	ns	0
MDE	SING	2ND	FEM	ACTIVITY	0	0.0812	0.0812	<	749.1054038	ns	0
VHB	SING	1ST	NILL	ACTIVITY	0	0.1435	0.1435	<	759.7380333	ns	0
RAH	SING	1ST	NILL	ACTIVITY	0	0.1435	0.1435	<	759.7380333	ns	0
MDE	SING	1ST	NILL	ACTIVITY	0	0.1435	0.1435	<	759.7380333	ns	0
VHB	SING	2ND	MASC	ACTIVITY	0	0.1824	0.1824	<	766.1921369	ns	0
RAH	SING	2ND	MASC	ACTIVITY	0	0.1824	0.1824	<	766.1921369	ns	0
MDE	SING	2ND	MASC	ACTIVITY	0	0.1824	0.1824	<	766.1921369	ns	0