Agreement Morphology: A Mathematical Approach to Integrate Arabic Syntax and Semantics

Yousuf Aboamer

A dissertation presented for the degree of

Doctor of Philosophy

Faculty of Linguistics and Literary Studies

Bielefeld University

Germany

14.01.2021

Contents

List of Tables

List of Figures

Abstract

In this dissertation, I apply a theory of semantic composition, as introduced by Marcus Kracht, to represent the meaning of Arabic sentences. Kracht's theory is a modification of Kees Vermeulen's work on referent systems. In Kracht's approach, the renaming process of variables is not decided by directionality, as suggested in the original version of referent systems, but by the morphosyntactic information associated with each variable. Arabic morphological constituents are allowed to be discontinuous, which maintains the pure surface treatment on the one hand and proves the possibility to save the concatenative paradigm for Arabic morphology on the other. An empty string is proposed in cases of null-subject or covert copula clauses, specifically challenging constructions of Arabic syntax. The semantics of Arabic temporal morphosyntactic markers is represented depending on Wolfgang Klein's theory of tense. As a part of my work, I develop a computational lexicon for a fragment of Arabic depending on the proposed analysis.

Keywords:
Agreement Morphology - Argument Structure - Referent Systems - Discontinuity of Constituents -Arabic Sentence Structure - Tense – Aspect

Zusammenfassung

In dieser Dissertation wende ich eine Theorie der formalen Modellierung von Sprachen und Grammatiken auf das Arabische an. Diese Theorie der semantischen Komposition wurde von Marcus Kracht eingeführt, der die Arbeit von Kees Vermeulen an Referentensystemen erweiterte, um Variablen je nach der verfügbaren morphosyntaktischen Information und nicht nach Direktionalität umbenennen zu können. Die morphologischen Konstituenten des Arabischen können dabei als diskontinuierlich repräsentiert werden, so dass einerseits eine reine Oberflächenbehandlung beibehalten und andererseits gezeigt werden konnte, dass es möglich ist, in einem Paradigma der konkatenativen Morphologie zu bleiben. Zwei spezielle Probleme der Syntax des Arabischen, verblose Sätze und Nullsubjekt-Sätze, wurden gelöst, indem eine leere Zeichenkette in beiden Fällen verwendet wird. Die Semantik arabischer zeitlicher morphosyntaktischer Markierungen wird nach Wolfgang Kleins Tempustheorie dargestellt. Bestandteil der Arbeit ist eine computerlinguistische Implementierung der vorgenommenen Analyse. Dazu gehört insbesondere der Aufbau eines entsprechend strukturierten Lexikons eines Fragments des Arabischen.

Schlagwörter:
Kongruenzmorphologie - Argumentstruktur - Referentensysteme - Diskontinuität von Konstituenten - Arabische Satzstruktur - Tempus – Aspekt

Acknowledgements

While working in this dissertation, I received great support from many people, and I am quite sure that their invaluable support was decisive for me to reach this moment. No words could express my gratitude to Prof. Marcus Kracht and Prof. Ralf Vogel for their academic guidance. Since my first day at the University of Bielefeld, or even during the application process of my scholarship, Prof. Kracht has spared no effort to help me on both academical and human levels. I was also very fortunate to receive real support from Prof. Vogel. I gained great insights from his discussions and suggestions to solve the problems I faced in my research, even before he became my supervisor after Prof. Kracht had suffered from health problems. I am also so grateful to the committee members, Prof. Ali Farghaly and Prof. Jutta Hartmann, for their time and their comments.

I would like to thank my colleagues at Bielefeld University: apl. Prof. Jens Michaelis, Dr. Johanna Domokos, apl. Prof. Joana Cholin, Prof. Petra Wagner and Mr. Daniel Milne-Plückebaum. I also acknowledge the efforts of my colleagues at Cairo University: Prof. Emad Abdulatif, Prof. Husam Qasim, Dr. Amr Otafy and Mr. Mohammad Mansy for their support. Finally, I am deeply indebted to my family for everything they did for me to help me concentrate on my research.

Many thanks to the Egyptian Ministry of Higher Education, the German Academic Exchange Service (Deutsche Akademische Austauschdienst “DAAD”) and the Faculty of Linguistics and Literary Studies (LiLi) for the financial support.

Chapter 1

Introduction

1.1 Goals of this dissertation

This dissertation examines Arabic morphology in connection with the demands set by compositional semantics. It explores how to integrate Arabic syntax and semantics using the overt morphological features in the string-to-meaning translation. I apply a theory of semantic composition as introduced by Marcus Kracht (2016) to Arabic. It is a modified version of Kees Vermeulen's theory of referent system (henceforth, RS) (Vermeulen, 1995). Kracht has built on the work of Vermeulen on the semantic composition by the renaming of the variables allowing the variables to be identified depending on the morphosyntactic features associated with them. I will refer to Kracht's approach as "Agreement Morphology and Referent Systems" and will use the initials (AMRS) in the rest of the dissertation. In this dissertation, I attempt to answer the following questions:

1. What are the theoretical bases of AMRS?
2. How can Arabic non-concatenative morphology be described within AMRS?
3. To what extent could we handle structural aspects of Arabic sentence depending on a morphologically motivated approach and how?
4. How well does the theory show universal properties of language?
5. What are the main challenges that the theory meets?

To answer these questions, I did not limit my work to just formulating a description of the Arabic sentence within AMRS, but developed a computational lexicon for (a fragment of) Arabic to test the theoretical bases of the theory. The aim of this test is to prove the applicability of the solutions proposed to solve the problems and to find out if there are limitations or any lack of correspondence between the theoretical assumptions and the practical implementation. A basic motivation of this work exceeds the mere description of the Arabic sentence within AMRS. While it may seem that the work is specifically devoted to a specific language, the implicit goal is to examine Kracht's claim that AMRS shows the general properties of language.

Why Arabic? I argue that Arabic is an excellent case study for several reasons. It is a relatively different case study, compared to Indo-European languages like English and German. Arabic has a very rich and complex morphology. The most important feature of Semitic languages in general, and of Arabic in particular, is that their words are formed from discontinuous morphemes. This is usually referred to in the literature by introflectional morphology, nonconcatenative morphology or transfixing. From a structural viewpoint, Arabic exhibits a number of interesting phenomena like its flexible word order, the possibility to drop the subject pronoun, i.e., pro-drop and copula-less clauses in a tensed environment. The real challenges posed by the morphosyntactic nature of Arabic can be considered an indication of the power of the calculus to handle similar languages or at least the phenomena that Arabic shares with other languages in the world. Furthermore, the formal representation of the meaning of Arabic sentences has received very little attention. To the contrary of English and German, to which a large amount of work has been devoted, the formal representation of sentence meaning in Arabic is a very deserted area of NLP. This work is an attempt to fill part of this gap.

Kracht posits a number of assumptions that constitute the base of his theory on the one hand, and exhibit his view of the general properties of the languages of the world on the other. Like other formal frameworks, natural language is designated in AMRS as a set of strings. Smaller units combine in some way to compose complex units until the full sentence is composed. While it shares this common description of natural language with other formal frameworks, it is strictly surface-oriented. The pure surface

orientation is twofold: (i) Constituents are composed of strings only via concatenation and reduplication, but no kind of deletion or movement is allowed. This is in lines with non-transformational frameworks, like Head-driven Phrase Structure Grammar, Lexical-Functional Grammar, Categorial Grammar and Dependency Grammar. (ii) It maintains the principle of surface compositionality (Hausser, 1980), according to which the semantic representation of a linguistic expression must have (concrete) surface syntactic motivation and the complex surface expressions must be (explicitly) composed of basic surface ones.

The third assumption of AMRS is the disregard of syntactic structure. The key role in the combination of strings is played by the so-called argument structure. The argument structure functions as an interface between syntax and semantics. Each semantic unit has an argument structure that specifies its behavior when it combines with another unit. When two units are to be combined, one takes the other as its argument. The argument structure declares this information, i.e., the role of the unit in the combination, how many arguments a unit may have and what are the morphosyntactic features the argument must have to combine with the functor. It seems to me as an attempt to swim against the current. That is, while the vast majority of generative grammars are much more concerned with syntax and give it the absolute priority, AMRS does not only promote morphology over syntax, but it ignores the syntactic structure almost completely. I can assume that Kracht does not use the word "morphosyntactic" as is usually used by linguists because it presupposes the dualism of morphology that studies word formation and syntax that studies word combination. It is rather used as an umbrella term to gather all the features that govern the combination of strings and nothing beyond that.

Kracht defines language as a set of signs. Each sign consists of two structures: the morphological structure and the semantic structure. The morphological structure contains the string and the relevant phonological and morphological properties that specify other strings it can combine with. This is independent of the semantic structure. The semantic structure of the sign consists of two components: the argument structure and the semantic representation. The argument structure is considered the main component of the sign and it provides an interface between syntax and semantics. The argument

structure contains information about the role of the sign with respect to the merge operation, i.e., whether it is a head or an argument and the variable's relevant identification information. It specifies which name a variable takes to be identified for merge and the name of the variable after merge. The argument structure is simply the engine of the semantic composition. It is a modification of the definition of RS as introduced by Kees Vermeulen. The main idea of the theory of RS is semantic composition by the renaming of variables. When two variables x and y are identified with each other, they are named to x^1. This mechanism has solved the problem of choosing the names of the variables. Kracht keeps this compositional mechanism, but makes the identification of variables depend on the match between the morphosyntactic features. The second component of the semantic structure is the meaning of the unit represented using discourse representation structures as introduced in (Kamp, Van Genabith, & Reyle, 2011; Kamp & Reyle, 1993).

In my work, I follow a bottom-up approach. This is also consistent with the calculus, since complex units are composed of smaller ones. Thus, the starting point would be the formal description of Arabic morphological units. Arabic morphology has given rise to various paradigms because of its nonconcatenative nature, most notably the autosegmental analysis (McCarthy, 1981). This work attempts, however, to investigate the possibility to save the concatenative paradigm, if discontinuity is admitted. This is the route that has been taken by Kracht with respect to the analysis of German[1]. Nevertheless, in order to describe Arabic morphology in a strictly surface-oriented way, a number of problems must be dealt with, especially the discontinuity of morphemes, matching among morphemes, and morphophonemic alternations resulting from the interaction among different types of morphs. A combinatorial function is used to combine Arabic discontinuous morphemes in diverse ways: forward concatenation, backward concatenation, forward wrapping, reduction, forward transfixation and, beyond multicontext free grammars (MCFGs), reduplication. However, the definition of this combinatorial function is not the only challenge, because roots behave differently when they com-

[1]Kracht focused on the separable prefixes and the perfect form, but a more similar case in German is the ablaut like singen, sang, gesungen (see (Sahel & Vogel, 2013, p. 139-140)).

bine with the patterns[2]. I benefited from the work of traditional Arabic grammarians and reformulated their categorization of Arabic roots in an attribute value matrix that specifies which root can combine with which vocalism. Morphophonemic alternations are listed in the lexicon as morphs of the same morpheme. This ensures the purely surface-oriented treatment of Arabic word formation and prevents overgeneration.

Like morphology, the structural aspects of Arabic are another source of complexity. In many languages, the verb is always in the center of the analysis, as is the case in English and German. In the first, the sentence has a fixed order and the verb is always in the core. In German, the position of the verb depends on the type of the sentence, main or subordinate. Arabic has two different types of sentences: verbal sentences and verbless sentences. In both types, one of the elements can be dropped from the surface. In the verbal sentence, the subject can be dropped in some cases, and the verbless sentences are constructed in the present tense without a verb or a copula. AMRS allows for empty strings; they do exist but are only invisible (Kracht, 2007). I show how agreement morphology helps specify the role of the dropped element and therefore construct the sentence smoothly.

The proper semantics of any sentence must also account for the parameters that affect its truth conditions. In principle, AMRS allows adding as many parameters as one wishes (Kracht, 2016, 2007). In this work, I add only three parameters: properties, tense and aspect. There are other parameters like world, person and location, but I include only the first three. The argument structure also contains information about the mechanism of handling these parameters and their meanings are represented in the DRSs. Although the concept of time in natural language has been the focus of a large amount of research in a variety of disciplines, I found a high degree of inconsistency in the literature on temporality in Arabic in the adopted approaches, findings and the usage of terms. Also, the phenomenon of sequence-of-tense in Arabic, for which the use of parameters is significantly important, has received little attention. This also applies to time in noun phrases. In this regard, I agree with Comrie (1985, 1976) and Fassi (2012)

[2]The function of the pattern is done in AMRS by the so-called handler, i.e, a function specifies the position of each letter of the root and the position of each letter in the vocalism. In some cases, the root merges with a combination of vowels and consonants and such a morph is referred to in this dissertation as “affixed vocalism” (af-vocalism)

that Arabic is a tense-aspect language, i.e., its temporal expressions exhibit both tense and aspect but follow Klein's point of view (W. Klein, 1994) that both are relations. Also, I argue that Arabic, like Japanese, should be represented as a non-sequence-of-tense language.

The scope of this work is restricted to the version of Arabic used nowadays as the basic means of communication among Arabic speaking countries. This version is commonly referred to as Modern Standard Arabic (MSA) to distinguish it from two other versions of Arabic, namely, Classical Arabic and Regional Arabic. Modern Standard Arabic is the language used in the media, newspapers, scientific writings and educational institutions.

This dissertation is structured as follows: in this introduction, I introduce the theoretical framework of this dissertation, the questions I try to answer, the goals I aim to achieve and the scope of my work. In addition, I explain my approach on the representation of Arabic text and the glossing rules. I also give a brief introduction to the structure of Arabic and its levels of use.

In Chapter two, I review the history of the development of formal semantics and refer to the most important contributions that paved the way to the emergence of formal semantics and pay more attention to those which shaped its history in the last fifty years, most notably type-theory, Montague Grammar and Discourse Representation Theory. I also illustrate the major criticisms directed to these approaches and how they should be solved in AMRS. Then, I move to a specific review of the literature on the formal representation of the meaning of Arabic sentences. Although there is total agreement on the lack of effort in Arabic logic-based meaning representation, I shed light on three attempts to represent the meaning of Arabic sentences, classify them and show their limitations.

Chapter three describes AMRS. I start with the definition of the semantic units and move to the detailed explanation of the two components of the units. I first introduce the basic notions that constitute the morphological component of the units. These notions include, among others, glued string, fractured glued string and handler. I also show how AMRS differentiates between the morph and the morpheme. Then, a special attention is paid to the semantic component in general and to the argument structure

in particular. I explain the four elements of each argument identification statement, i.e., variable, diacritic, attribute value matrix and parameters and show the role of each element in the merge operations. I also show how the discourse representation structure combines with the argument structure to form the semantic structure. I close this chapter with an explanation of how each of these components is computed in the merge operation.

Chapter four is devoted to the description of Arabic morphology within AMRS. I start by showing the gap between morphology and semantics in the current approaches to Arabic morphology and how to fill this gap in AMRS. I focus on the most related notions to the nature of Arabic morphology and how they are applied to it. To be more particular, the focus is placed on the discontinuity of Arabic morphemes, or the fractured glued strings using Kracht's terminology, and how they are combined by handlers. Two other points are dealt with in this chapter, namely reduplication and morphophonemic alternations in Arabic weak verbs.

In chapter five, I move from the simple morphological units to how they combine to form a full sentence. I focus on the most problematic issues in Arabic sentence structure. Arabic shares most of these issues with other languages of the world, and therefore the same solutions I propose to Arabic can be applied in similar cases. I specifically show how the powerful mechanism of the proposed calculus accounts for complex phenomena such as flexible word order, null-subject, copula-less clauses in a tensed environment and the reduplication of arguments.

Chapter six is devoted to three of the parameters that affect the truth conditions of the sentence, but I pay more attention to tense and aspect. I apply a three-parameter theory of time as introduced by (W. Klein, 1994) to Arabic. The chapter also represents an attempt to resolve the inconsistency in the literature on Arabic temporarily. Therefore, I discuss the most common assumptions related to tense and aspect in Arabic. I also show how to represent time in noun phrases and how to classify Arabic with respect to the phenomenon of sequence-of-tense.

Chapter seven is divided into two parts: part one gives a description of the developed lexicon and how it can be used within the computational implementation of AMRS in OCaml. I show how the lexical entries are inserted and explain the code notations. I

also show the results of the merge of lexical entries and end with a full example of a complete sentence. The second part discusses the challenges that AMRS meets both theoretically and practically. These challenges include, among others, the challenges posed by the absence of diacritics in the modern writing system of Arabic, the compositional treatment of multi-word expressions, specifying the semantic role for the two kinds of objects and the computational complexity.
In the conclusion, I give a general discussion of the results as well as some concluding remarks.

1.2 Arabic text representation

In this work, I use two processes to represent the Arabic text: transcription and transliteration. The international phonetic alphabet (IPA) is used to represent Arabic phonemes along with Leipzig Glossing Rules in the dissertation. Practically, I use a modified version of Buckwalter's transliteration model to represent the lexical entries in the developed lexicon. While the IPA makes the text more easily readable and understandable for those who are not familiar with Arabic letters, I found it more practical to use Buckwalter's transliteration model in developing the lexicon, since it has been originally developed for Arabic NLP applications. Table (1.1) shows the Arabic letters and their representations in the IPA and in Buckwalter's model. Nevertheless, for technical reasons, I made some modifications to Buckwalter's model, because it contains some symbols that make it XML-unfriendly, especially my lexicon is created in an XML format. These modifications include, for example, the differentiation between the two functions of the semi-vowels waːw and jaːʔ . Each of the two letters behaves in some contexts as a consonant and is referred to in Arabic as ħuruːfu lliːni and in other contexts as a vowel and is referred to as ħuruːfu lmaddi. Along with the ʔalif, the three letters can be described as long vowels compared to the three short vowels alfatħah, adˤdˤmah and alkasrah. Therefore, I use two different symbols; one to refer to the first function of each letter and one to indicate the second function. That is, I will use the three lower case letters "a", "i" and "u" for the short vowels and the upper-case letters "A", "I" and "U" for the long vowels. Also, unlike Buckwalter's model, in which the

different forms of the hamzah are represented using six different symbols, I use one symbol, the apostrophe, to represent the different forms of the hamzah. The distinction among these forms depends on what vowels come before or after the hamzah.

In the transcription process, I always follow the pronunciation rather than a letter-to-letter correspondence. For example, the second letter in the definite article is pronounced in some cases and is dropped in other cases. I follow the pronunciation in the two cases. In the glossing process, I follow a word-by-word alignment. The examples are first written in Arabic, followed by the transcription, then the grammatical categories and finally the translation. Short vowels and one of the special characters are represented above or below a letter.

Arabic consonant	IPA equivalent	Buckwalter's transliteration model
Consonants		
ء	ʔ	'
ب	b	b
ت	t	t
ث	θ	t
ج	ʤ	j
ح	ħ	H
خ	x	x
د	d	d
ح	ð	*
ر	r	r
ز	z	z
س	s	s
ش	ʃ	$
ص	$s^ʕ$	S
ض	$d^ʕ$	D
ط	$t^ʕ$	T
ظ	$ð^ʕ$	Z

ع	ʕ	E
ج	ɣ	g
ف	f	f
ق	q	q
ك	k	k
ل	l	l
م	m	m
ن	n	n
ه	h	h
و	w	w
ي	j	y

Table 1.1: Arabic consonants in IPA and in Buckwalter's transliteration model

Arabic vowels	IPA equivalent	Buckwalter's transliteration model
Vowels		
بَ	a	a
بِ	i	i
بُ	u	u
ا	aː	A
ي	iː	y
و	uː	-

Table 1.2: Arabic vowels in IPA and in Buckwalter's transliteration model

Arabic character	IPA equivalent	Buckwalter's transliteration model
Special characters		
ة	t, h	p
ـٰ	-	o
بِّ	-	~

Table 1.3: Arabic special characters in IPA and in Buckwalter's transliteration model

1.3 Introduction to Arabic

Arabic belongs to the Semitic language family, which includes, in addition to Arabic, Hebrew, Aramaic and Amharic. The members of this family are spoken in the Middle east and Ethiopia (McWhorter, Leven, & Blandford, 2004, Part I, P 65). Semitic languages have been divided on geographical criteria into three groups: northwestern, northeastern and southwestern. Arabic was a member of the southwestern group (Holes, 2004, p. 10). While some languages of the Semitic group died long time ago, Arabic is currently the most spoken living Semitic language in the world with a number of speakers approaching 415 million, according to the CIA World Fact Book (CIA, 2018). These are native speakers of Arabic while a billion Muslims use Arabic in their daily prayers. It is the/an official language of 25 countries in a vast area extending from the Arabian Peninsula to the Fertile Crescent, in addition to some other countries in the middle east and Africa, mostly the countries of the Arab world. This list of countries includes, in alphabetical order, Algeria, Bahrain, Chad, Comoros, Djibouti, Egypt, Eritrea, Iraq, Jordan, Kuwait, Lebanon, Libya, Mauritania, Morocco, Oman, Palestine, Qatar, Saudi Arabia, Somalia, Sudan, Syria, Tanzania, Tunisia, United Arab Emirates and Yemen. Since 1973, Arabic has become an official language of the United Nations.

1.3.1 Arabic varieties

The literature on Arabic language differentiates generally between three levels of Arabic: Classical Arabic (CA), Modern Standard Arabic (MSA) and Arabic dialects.

Classical Arabic (CA)

Historically, before the 7^{th} century, Arabic was only spoken by a number of tribes in the Arabian Peninsula. However, after the emergence of Islam in the 7^{th} century, several political, religious and environmental factors played a significant role in changing the status of Arabic to become an international language, most importantly, the expansion of the new empire, the centrality of Arabic as a holy language on the one hand and the language of administration on the other, as well as its being the language of scientific writing and research. Arabic of that time is usually referred to in the literature as Classical Arabic. In the 8^{th} century Arabic grammarians started to codify Arabic for the first time (Farghaly, 2010, p. 45), depending on four resources: the language of the Quran, the sayings (Ahadith) of the Prophet, the language of poetry until a specific time and the language as used by Arabs only based on strictly temporal and spatial criteria. The goal was to keep the language purity and to avoid the mistakes of the non-native speakers. At the present time, Classical Arabic is used in very limited contexts, especially by some scholars who are significantly influenced by Quranic studies or in some limited academic environments. Also, a significant amount of the lexicon of Classical Arabic is no longer used, particularly in everyday life and the media, since it is closely related to a specific time and environment.

Modern Standard Arabic (MSA)

There is no commonly agreed upon definition of Modern Standard Arabic, but there is a general agreement that Modern writings represent the basis of this form of Arabic (Ryding, 2005, p. 8). This includes all forms of written word such as the language of written media, street signs and advertisements. In addition, MSA is considered the language of public speaking and news broadcasts on radio and television ((Ryding, 2005, p. 5) and (Badawi, 1979, p. 90)). MSA is now the common language among the Arab world. It is considered a modern version of Classical Arabic, since both agree mostly in phonology, morphology and structure, but differ significantly in lexicon and style

(Procházka, 2010, p. 43) because it (MSA) is largely influenced by modern civilization (Badawi, 1979, p. 89)[3].

The beginning of Modern Standard Arabic dates back to the end of the 18^{th} century and the beginning of the 19^{th}, mainly when Arab intellectuals started translating European works into Arabic. At that time, there was a civilization gap between Europe and the Arab countries, and soon those intellectuals recognized the shortcomings of Arabic and its insufficiency to capture the new concepts, notions and technical or intellectual terms (Procházka, 2010, p. 43). The modernization of Arabic continued after the first world war, especially the lexicon and the style, by means of coining neologisms or borrowings from other languages, particularly from English and French (Procházka, 2010, p. 43). It is worth noting that MSA, despite being the official language, is not the native level that children acquire from their parents or their community. In the real situation, children acquire and speak their regional dialects as a native spoken level, but learn the standard level in educational institutions.

Regional Dialects

Arabic represents one of the world's most famous and diglossic language situations (Kaye, 2001, 1994). In his study of Arabic diglossia, Ferguson (1959) has differentiated between two levels of Arabic: high and low varieties. He depended in his classification on a number of criteria, notably the function, i.e., the situations in which a specific level of language is used. According to Ferguson, high language is used in sermons in churches or mosques, political speeches in the parliament, university lectures ... etc, while low language is used in situations like conversations with family, friends and colleagues and in folk literature. Badawi (1979) has argued that Ferguson's division does not reflect the real situation of Arabic. Badawi has distinguished between fusʕha attura:θi (فُصحَى التُّرَاثِ 'Classical Standard Arabic') and fusʕha lʔasʕri (فُصحَى العَصرِ 'Modern Standard Arabic'). Also, he made a distinction between three levels of colloquial Arabic

[3]I deal with MSA and CA as having the same morphology and syntax.

in Egypt: ʕaːmmijjatu lmuθaqqafiːn (عَامِّيَّةُ المُثَقَّفِينَ 'colloquial of the educated'), ʕaːmmijjatu lmutanawwiriːn (عَامِّيَّةُ المُتَنَوِّرِينَ 'colloquial of the enlightened') and ʕaːmmijjatu lʔummijjiːn (عَامِّيَّةُ الأُمِّيِّينَ 'colloquial of the illiterate'). Nevertheless, Badawi's claims are not fully agreed upon and other linguists have rejected his classification, because these levels are not identifiable linguistic levels. Furthermore, Arabic dialects can be divided geographically into several dialects, including Egyptian Arabic, Levantine Arabic, Gulf Arabic among others (Habash, 2010, p. 2).

1.3.2 Structure of Arabic

Phonology and script

The Arabic alphabet contains twenty eight consonants, twenty six pure consonants and two with a double function depending on the context. In addition, Arabic has three short vowels which are written above or under the letter and they are referred to as diacritics or ħarakaːt (حَرَكَات 'vowels'). The three short vowels have three corresponding long vowels. In addition, Arabic uses other characters with specific meanings. These characters include the reduplication symbol or aʃʃaddah (اَلشَّدَّة), the set of tanwiːn symbols (تَنوِين) and the non-vowel symbol or sukuːn (سُكُون) (see the transcription and transliteration models, for the names and the forms of Arabic letters in table (1.1). The total number of characters used in the Arabic orthographic system is 60 unique characters for the letters, diacritics and punctuation marks (Attia, 2008a, 17). In addition, Arabic is written from right to left using the Arabic script. Letters are joined in a systematic way and there is no capitalization. In the modern writing system, the diacritics are usually absent. This increases the level of ambiguity. These issues make Arabic typographically different from the Latin character set, and this poses a challenge to computers to view Arabic fonts correctly: first when computers were limited to use ASCII system, and now in some cases even after the introduction of the Unicode system (Attia, 2008a, 17)[4].

[4]I have faced many challenges during writing my dissertation in Latex. Even when I used specific packages developed to correctly represent the Arabic fonts, I encountered some problems like the correct representation of the tanwiːn (تَنوِين) and the clash between these packages and some other packages.

Morphology

One of the most famous features of Semitic languages, and particularly of Arabic, is their introflectional morphology (Watson, 2010, p. 50). This feature is also referred to in the literature as discontinuity of morphemes, nonconcatenative morphology (McCarthy, 1981), or transfixing (Bauer, 2003). Unlike concatenative languages, like English, which depend on the linear concatenation of morphemes as in (un-desir-able-ness), Arabic words are mostly formed by the merge of the so-called roots and patterns. Roots are pure consonants, typically three, that bear the lexical meaning shared by all words derived from the root. Vowels are inserted between the letters of the roots to form the stems which concatenate with prefixes, suffixes and clitics to form the actual words, as shown in figure (1.1). Since roots carry the basic meaning, derived words usually share this common meaning and differ in their morphology. I will discuss Arabic morphology from a theoretical and computational perspective in some detail in chapter four.

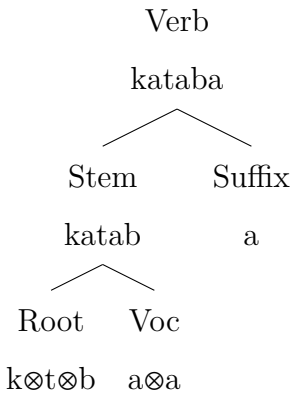


Figure 1.1: Word formation in Arabic

Syntax

Arabic has two types of sentences: verbal sentences and copula-less sentences. A verbal sentence is composed of a verb-predicate and at least one argument, i.e., the subject. In case of transitive verbs, the verb takes two arguments and the order is flexible. That is, Arabic is mainly a Verb-Subject-Object language, but this order can vary significantly. A verbal sentence with an intransitive verb can be ordered as Subject-Verb or Verb-Subject. This applies to transitive and ditransitive verbs. The order is free as long as

the change in word order does not result in ambiguity. Another main feature of Arabic verbal sentences is the possibility to drop the subject, if it can be inferred from the context. This feature is referred to as the null-subject or the pro-drop. The second type of sentences in Arabic is the copula-less sentence or verb-less sentence. In the second type, neither an overt verb nor a copula appears at the surface of the sentence. The two basic predicative constituents are a noun phrase (subject) and another noun phrase, an adjective phrase or an adverbial phrase (predicates). I will discuss the syntactic structure of Arabic in some detail in chapter five.

I acknowledge that some parts of this dissertation have been published in the following papers[5]:

- Kracht, M.& Aboamer, Y., "Argument Structure and Referent Systems", 12th International Conference on Computational Semantics (IWCS). 19-22 September 2017 Montpellier (France), 2017.

- Aboamer, Y.& Kracht, M., "Representing Meaning of Arabic Sentence Dynamically and more Smoothly.", In the Proceedings of the 4th International Conference on Arabic Computational Linguistics (ACLing 2018) November 17-19, 2018
  - Dubai (UAE). The Journal of Procedia Computer Science, Elsevier., 2018.

- Aboamer, Y.& Kracht, M., "A Purely Surface-oriented Approach to Handling Arabic Morphology.",In the proceedings of the 24th Conference on Formal Grammar in conjunction with the 31st European Summer School in Logic, Language and Information (ESSLLI 2019), Riga, Latvia. Lecture Notes in Computer Science (LNCS), Springer., 2019.

[5]The application of the system to Arabic is my part in the enterprise.

Chapter 2

Review of Related Work

This chapter has several goals. It reviews the most important contributions on formal semantics, and therefore shows the relation between this work and the literature in formal semantics. Also, it introduces basic notions and terminology that will be frequently used in this dissertation. I believe that both goals are significant; however, the Arabic research community was at the core of my concerns during writing this chapter. The lack of efforts, the scarcity of specialists and the limited amount of works on formal semantics in the Arab world made it almost without adherents. This chapter attempts to fill this gap by introducing the influential contributions on formal semantics along with reviewing the limited attempts devoted to handling Arabic.

2.1 Formal semantics from static to dynamic

Semantics is generally defined as the study of meaning; however, what is meant by meaning may differ significantly depending on the linguistic discipline and the proposed theory of analysis. Lexical semantics is concerned with the denotative (conceptual or cognitive) meaning of the word, i.e., the relation between a word and what it refers to (Murphy, 2010, p. 29-32). Also, the semantic relations between these denotative meanings such as synonymy, hyponymy, antonymy … etc., lie within the scope of lexical semantics. Diachronic semantics is devoted to the study of how and why meanings of linguistic expressions evolve across time (Deo, 2015). It studies, for example, the semantic change of the meaning of the word "holiday" from "holy day", i.e., a religious feast,

to its current meaning “holiday”, i.e., a day without work, (Leech, 1974, p. 123-124). Formal semantics is defined as “the discipline that employs techniques from symbolic logic, mathematics, and mathematical logic to produce precisely characterized theories of meaning for natural languages” (King, 2006). Formal semantics aims at explaining relations like entailments, i.e., the relations between premises and valid conclusions expressed as natural language sentences, clarifying important issues in linguistic theory, most importantly the phenomenon of structural ambiguity and using abstract mathematical structures (models) to analyze natural language expressions (Winter, 2016).
There is a general consensus that Richard Montague was the first one to explicitly represent the relationship between syntax and semantics of natural language in a formal way. Prior to the ground-breaking works of Richard Montague, namely: “English as a Formal Language” (Montague, 1970a), “Universal Grammar” (Montague, 1970b) and “The Proper Treatment of Quantification in English” (Montague, 1973), many linguists have, in fact, doubted that it is possible to represent natural languages formally due to the very high level of ambiguity and the lack of structure compared with formal languages. However, three extremely important contributions had inspired Montague’s work, and without these contributions, it was extremely difficult to see the achievement of Montague; namely, Alfred Tarski’s semantic concept of truth, Frege’s principle of compositionality and the invention of lambda calculus by Alonzo Church.

2.1.1 The concept of truth

The notion of truth values that has been introduced into logic for the first time by Gottlob Frege, is a fundamental notion in formal semantics. According to Frege, the meaning of a (declarative) sentence is a truth value. This means that there is a relation between the meaning of a sentence and its truth; and to know its meaning means to know how the world can be made for it to be true, i.e., which conditions must be achieved in the world for it to be true (Lohnstein, 2011, p. 2-3). A sentence is true only if it denotes a fact; otherwise, it is false and the sentence cannot be true and false at the same time. So, (1) describes a fact that either actually exists or does not exist. That is, this sentence is true only if there is a person called “Merkel”, there is a place referred to as “Germany”, there is a person referred to as “chancellor” and “Merkel”

is the “chancellor of Germany”. Therefore, based on the idea that the meaning of a sentence denotes a fact, we can determine the meaning of (1) as true. If “Merkel” is not the “chancellor”, then we say that the meaning of (1) is false.

(1) Merkel is the Chancellor of Germany.

However, the way in which natural language is organized allows, through recursive rules of syntax, for the potential existence of infinite nonsynonymous well-formed sentences, the fact that led to the emergence of Frege’s principle of compositionality. But before I move to the Fregean principle of compositionality, I refer to Tarski’s semantic theory of truth, in which he has given the concept of truth its theoretical basis within a formal language that reconstructs the relationship between language and the world in a systematic formal way (Lohnstein, 2011, p. 5). For Tarski, an adequate theory of truth for a language L must satisfy *Convention T*, which he defines, for each sentence x of L, as (Glanzberg, 2018):

Definition 1. *Convention T*: x is true if and only if P.

x is a variable for the name of the sentence of L (object language) and P is the name of x in the metalanguage. This is shown in the following sentence[1]

(2) “Grass ist grün” is true if and only if grass is green.

In (2) the object language is German and the metalanguage is English. This is the case of base clauses but truth can also be defined recursively using connectives like $\wedge$ (logical and) and $\vee$ (logical or) in the following way (Glanzberg, 2018) for any two sentences x and y of L:

1. $x \wedge y$ is true if and only if x is true and y is true.

2. $x \vee y$ is true if and only if x is true or y is true

As an example of these recursion clauses, consider (3).

(3) “Grass ist grün und Schnee ist weiß” is true if and only if grass is green and snow is white.

[1] I modified the examples of Glanzberg (2018) to be clearer. Instead of using English as an object language and a metalanguage, I used German as an object language and English as a metalanguage.

2.1.2 Principle of compositionality

Groenendijk and Stokhof (2005) count two reasons for the centrality of compositionality in the heart of formal semantics:

1. The use of logical languages as representational devices, be it that of first-order logic, that of intensional type theory or some other language.

2. The notion of "creativity" of language, or of language users.

The second point is related to the ability of the competent speakers of a natural language to compose an infinite number of complex meanings from a finite number of word meanings. The syntax of any logic language contains the rules that define how to compose complex expressions from simpler ones. The principle of compositionality, also known as Fregean principle of compositionality, or simply Frege's principle, specifies that the meaning of a complex expression is determined by the meaning of its parts and the syntactic rules by which they are organized. Two conditions must be achieved to satisfy the principle of compositionality (Partee, ter Meulen, & Wall, 1990, p. 318-319):

1. The syntax is given by a recursive specification, starting with a stipulation of basic expressions of given categories.

2. The semantics is given by a parallel recursive specification, including a stipulation of the semantic values for the basic expressions and for each syntactic rule a single semantic rule.

The following two statements (taken from: (Lohnstein, 2011, p. 60)) explain to what extent this principle can affect the meaning of a sentence. Both statements are composed of the same atomic statements p and q, the connector $\vee$ and the negation symbol $\neg$. They explain how their meanings differ depending on their syntactic structures.

(4) $\neg(p \vee q)$

(5) $(\neg p) \vee q$

The two statements have different meanings because (4) is true if and only if both p and q are false, but (5) is true if and only if p is false or q is true. This simply shows that

the meaning of complex expressions does not depend only on the meanings of its parts but also on their structures. Frege's principle of compositionality didn't inspire only Montague's work, but it was essential for all the subsequent work on formal semantics because of its central role in articulating the relation of semantics to syntax in formal languages or the correspondence between the syntactic structure of a formula and its semantics representation (Partee et al., 1990, p. 317-319).

In model-theoretic semantics, the meaning of any expression, simple or complex, is not interpreted in the absolute sense, but with respect to a model. A model is an abstract representation of the world that contains an arbitrary set of entities (discourse domain) and their relations and properties (formulae), which are evaluated as either true or false with respect to that model (Cann, Kempson, & Gregoromichelaki, 2009, p. 68).

2.1.3 First-order logic

First-order logic, also called predicate logic or predicate calculus, is a system for representing propositions (sentences or statements) in a formal notation. It is more powerful than propositional logic, since it addresses certain aspects of logic which require an analysis of the internal structure of atomic propositions and includes quantification (Crystal, 2008, p. 381). This means that first-order logic was introduced as an alternative for propositional logic to deal with two limitations: the inadequacy of propositional logic to analyze the internal structure of sentences and the problem of quantification. In first-order logic, sentences are analyzed in terms of the relation between predicates and arguments. Also, to analyze sentences that include quantifiers like "some" and "all", the two quantifiers $\forall$ and $\exists$ are introduced.

Before Montague, first-order logic was the standard method of the semantic analysis of natural language because it was much more powerful than the propositional logic. However, it was not expressive enough to deal with the richness of natural language. Therefore, it was not considered an inadequate means of representing the natural language. One of the limitations of first-order logic is the lack of a uniform semantics for noun phrases, which behave in many respects syntactically alike (Janssen, 2020) as shown in (6), (7) and (8).

(6) Peter sleeps.
sleep'(Peter')

(7) Every student sleeps.
$\forall$x [student'(x) $\rightarrow$ sleep'(x)]

(8) Some students sleep.
$\exists$x [student'(x) $\wedge$ sleep'(x)]

Another limitation is that quantifiers are only admitted over individual variables but not over predicates or functions. The language of first-order logic has been enriched in different ways. One of these ways was second order-logic (PL2). In the language of second-order logic quantifiers are not restricted to the properties of individuals, but it is also possible to quantify over the properties of properties. For example, if a t-shirt is blue, and blue is a color, it cannot be inferred that the t-shirt is a color, but rather of a color. The predicate "is a color" can only be applied to the property of the object, not to it directly (Partee et al., 1990, p. 231-232). Nevertheless, the second-order logic was not also expressive enough to capture the semantic structures of natural language.

2.1.4 Type logic and λ-operator

Type theory is a "system of semantically motivated categories". It was "designed so that restrictions on well-formedness stated in terms of types can guarantee that any well-formed expressions will be semantically well-defined" (Partee et al., 1990, p. 339). It was introduced as an attempt to classify semantic expressions into categories (types) in a similar way to what is done in the syntactic theory. That is, just as syntactic expressions are categorized into nouns, verbs, adjectives etc., type logic was designed to give semantic expressions a comparable categorization (Lohnstein, 2011, p. 126). There are two basic types, e for entity and t for truth value. The denotations of type e are individuals, whereas the denotations of type t are truth values. The predicate types show how many expressions, and of which types they must connect to form a type t . An expression of the type $\langle e, t\rangle$ can be seen as a function from type e to type t. The set of types is defined recursively as follows:

- e is a type
- t is a type
- if a and b are types, then $\langle a, b \rangle$ is a type.
- Nothing else is a type.

The following example shows how syntactic and semantic rules work in parallel.

(9) Peter visited London.

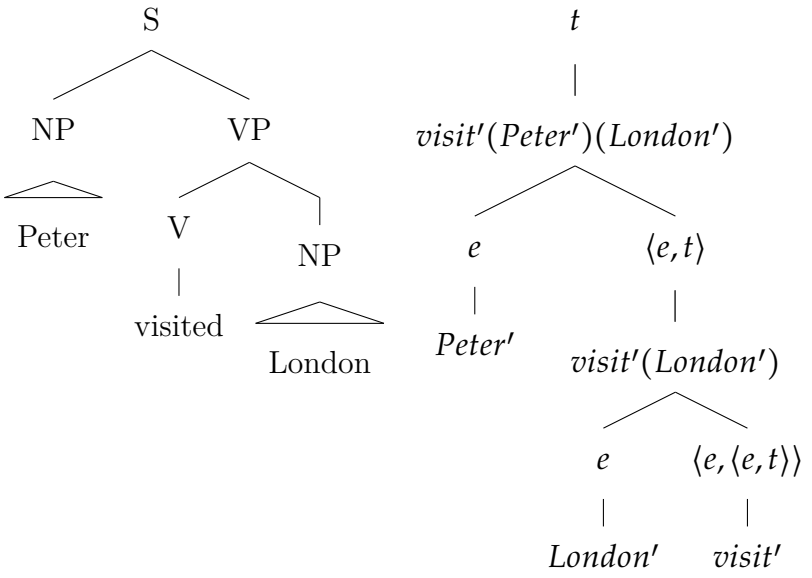


Figure 2.1: Parallel work of syntactic and semantic rules

In the pure type logic, the type $\langle e, t \rangle$ is a functional type. It indicates that it requires an argument of type e and yields a value of type t. However, a non-logical expression like $sleep'$, which is of type $\langle e, t \rangle$, does not indicate what argument it takes and what value it yields. This limitation could be overcome through the extension of type logic by the λ-operator (Lohnstein, 2011, p. 155-159). This operator was introduced by Alonzo Church (1940) "to permit the construction of expressions that unambiguously and compositionally denote functions" (Partee et al., 1990, p. 338). The λ-abstraction has been considered a very useful tool in semantics because many of the basic syntactic structures can be compositionally dealt with as involving function-argument application and

many of the less basic constructions can be given a compositional semantics involving λ-abstraction (Partee et al., 1990, p. 339). With the help of λ-operator, "sleep" is represented as $\lambda x[sleep'(x)]$, which stands for a function that assigns for each variable x the function value $sleep'(x)$, instead of $sleep'$. That is, $\lambda x[sleep'(x)]$ has the type $\langle e,t\rangle$ exactly like $sleep'$, x has the type e, $sleep'(x)$ has the type t and the term $\lambda x[sleep'(x)]$ has the same denotation as $sleep'$, i.e., the characteristic function.

2.1.5 Noun phrases and determiners

Noun phrases include several natural language expressions; some of which can be simple and contain only a noun, like proper nouns, mass nouns and plural nouns. Others are more complex like definite nouns and quantified nouns (Carnie, 2006, p. 64-66). Proper nouns, which denote specific objects like the names of persons (John, Mary, Peter) and places (London, Leipzig, Bielefeld), are considered of type e (Lohnstein, 2011, p. 182). Quantified nouns are represented with the help of the two quantifiers $\forall$ and $\exists$ as shown in (10) and (11).

(10) Every man sleeps. $\forall x\,[man'(x) \rightarrow sleep'(x)]$

(11) A man smokes. $\exists x\,[man'(x) \wedge smoke'(x)]$

This treatment of quantified nouns is problematic from two sides: first, there are many other possibilities of quantification. Lohnstein (2011, p. 187-188) gives examples for these possibilities in German as shown in (12).

(12) (i) Mindestens ein Student raucht. 'At least one student smokes.'
(ii) Zwei Studenten rauchen. 'Two students smoke.'
(iii) Mehr als die Hälfte der Studenten rauchen. 'More than the half of students smoke.'
(iv) Die meinsten Studenten rauchen. 'Most of students smoke.'
(v) Viele Studenten rauchen. 'Many students smoke.'
(v) Wenige Studenten rauchen. 'Few students smoke.'

The second problem is related to the difference between the syntactic structures of the sentences with quantified nouns and their semantic structures, i.e., they are not

derived compositionally (Lohnstein, 2011, p. 187–190). According to Montague, each syntactic category should have its corresponding semantic type, and therefore if several linguistic expressions have the same syntactic category, then they should have the same semantic type. Using λ-abstraction, Montague has proposed a uniform representation of noun phrases as "generalized quantifiers". That is, quantified nouns are considered expressions of type $\langle\langle e,t\rangle,t\rangle$ and determiners take nouns as their arguments, i.e., they are expressions of type $\langle\langle e,t\rangle,\langle\langle e,t\rangle,t\rangle\rangle$. In this way, the different expressions which are syntactically classified as noun phrases can be interpreted in a uniform way (Partee et al., 1990, p. 358-360). We can now reconsider the representations of sentences (10) and (11) in figure (2.2).

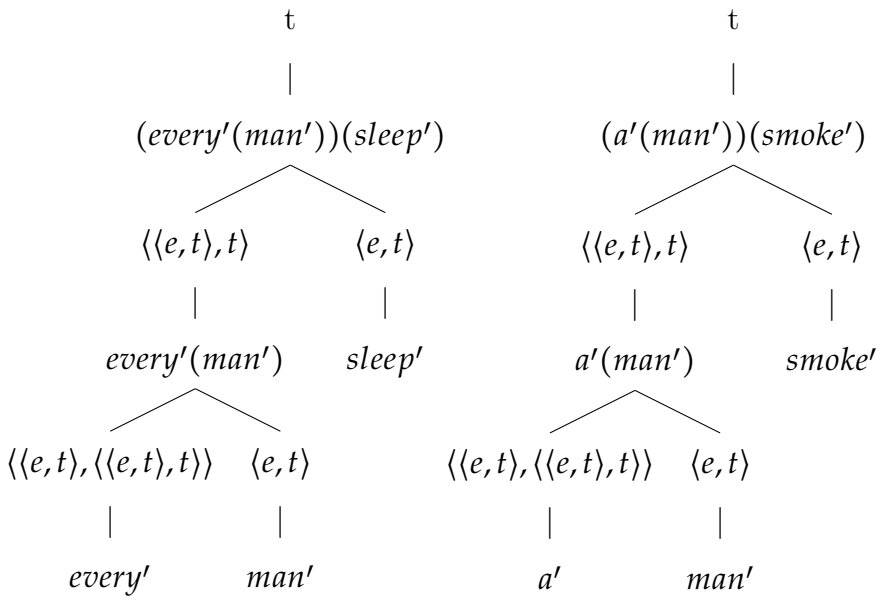


Figure 2.2: Uniform representation of quantified nouns

The semantic representations of the determiners can be determined using the λ-term as follows:

- every $\lambda Q\lambda P\forall x[Q(x) \rightarrow P(x)]$
- a $\lambda Q\lambda P\exists x[Q(x) \wedge P(x)]$
- no $\lambda Q\lambda P\neg\exists x[Q(x) \wedge P(x)]$

The translations of (10) and (11) can be given in (13) and (14) respectively.

(13) (i) $\lambda Q \lambda P \forall x[Q(x) \rightarrow P(x)](man')(sleep')$
(ii) $\lambda P \forall x[man'(x) \rightarrow P(x)](sleep')$
(iii) $\forall x[man'(x) \rightarrow sleep'(x)]$

(14) (i) $\lambda Q \lambda P \exists x[Q(x) \wedge P(x)](man')(smoke')$
(ii) $\lambda P \exists x[man'(x) \wedge P(x)](smoke')$
(iii) $\exists x[man'(x) \wedge sleep'(x)]$

As for proper nouns, it is convenient to be treated as having type e, since they denote unique entities, but this means that there is no uniform treatment of all noun phrases, although they have the same syntactic category NP (Cann, 1993, p. 172). Montague proposed to treat quantified nouns like "every man" and proper nouns like "John" as denoting sets of properties of individual concepts. "The individual concept of John is the function that picks out John at each possible world and time" (Partee, 1976, p. 59).

2.1.6 Montague grammar

It is commonly agreed upon that Richard Montague was the first to explicitly state the possibility to comprehend both natural and formal languages in the same way. Before Montague, both linguists and philosophers doubted that natural languages can be given a compositional semantics not only due to its high level of ambiguity but also because it is too unstructured to be represented by any formal language (Partee et al., 1990, p. 333-334). However, Montague clearly said that it is possible to connect the syntax and semantics of natural language in the same way as formal languages, and consequently to comprehend the syntax and semantics of both kinds of languages with a single natural and mathematically precise theory (Montague, 1970a). To fully understand Montague grammar, the following principles should be kept in mind (Janssen, 2020):

- He constructs a systemic relationship between syntax and semantics. That is, for each syntactic rule that composes an expression, there must be a semantic rule that determines the meaning of that expression.

- Syntactic rules are translated into intensional logic (IL). It was developed mainly to allow to specify certain models and specific moments of time. The symbol ^

is prefixed to expressions to indicate the intension, i.e., the extension of a given expression at a specific world and time.

- He applies the model-theoretic semantics. That is, expressions are interpreted with respect to a specific model.

In the following, I summarize Partee's explanation of Montague Grammar in (Partee, 1976). Montague grammar has two components: a syntactic component and a semantic component. The syntactic component contains a set of recursive syntactic rules that define the set of well-formed formulae starting with the smallest elements and specifying how they combine to form larger ones. That is, there are two basic categories: t for the category of sentences, i.e., the expressions that have a truth value and e for entity-expressions. From these two basic categories, complex categories can be formed recursively. No words or phrases of English are assigned the category e, but it composes with the category t the remaining categories. For example, the category of intransitive verb phrases is defined as t/e, the category of common noun phrases is defined as $t//e$ and the category of term-phrases is defined as t/IV. The following table, (depending on (Partee, 1976) but with minor modifications and examples), contains the categories that Montague described for a fragment of English, their abbreviations, names in PTQ and their nearest linguistic equivalents.

Category	Abbrev	Name in PTQ	Equiv	Example
t	—	Truth-value expression	Sentence	—
e	—	Entity expression	(Noun phrase)	—
t/e	IV	Intransitive verb phrase	Verb phrase	run
t/IV	T	Term	Noun phrase	John
IV/T	TV	Transitive verb phrase	Transitive verb	love
$t//e$	CN	Common noun phrase	Noun	man
IV/IV	IAV	IV-modifying adverb	VP-adverb	rapidly

Table 2.1: Some categories in Montague grammar

All the syntactic rules build up larger phrases from smaller ones, whereas the semantic

rules operate in two stages: first, every syntactic rule is translated into an expression of intensional logic. For example, for each category A/B or $A//B$, there is a syntactic rule of the form:

$$\text{If } \alpha \in P_{A/B} \text{ and } \beta \in P_B, \text{ then } F_i(\alpha, \beta) \in P_A$$

where F_i is a specification of the syntactic mode combination of the constituent phrases, $P_{A/B}$ is the set of phrases for each category A/B, P_B is the set of phrases for each category B and P_A is the set of phrases for each category A. The function F_i yields a phrase of the category A from a phrase of the category A/B and a phrase of the category B. For example, let $love \in IV/T$ (Transitive Verb, TV) and $John \in t/IV$ (Term, T), then $F_i(love,\ John) \in t/e$ (Verb Phrase, IV). Applying this function, we get $(love\ John)$. Again, let $Mary \in t/IV$ (Term, T), and apply the same function $F_i(Mary,\ love\ John)$ to get a sentence or t.

Then, each syntactic rule is translated into intensional logic, as shown in the following translation of the previous rule:

> If $\alpha \in P_{A/B}$ and $\beta \in P_B$, and α, β translate into α', β' then $F_i(\alpha, \beta)$ translates into $\alpha'(\hat{}\beta')$

where $\hat{}\beta'$ means the intension of β'. After translating syntactic rules into expressions of intensional logic, a possible-worlds semantics is defined for the given expressions of intensional logic.

All over the last four decades Montague's framework has been thoroughly discussed and scrutinized. Kracht (2008) sheds light on the criticisms that have been raised over Montague's framework and concentrates on the problem of the identification of objects across sentences. This limitation is usually recognized as the donkey sentences which cannot be properly analyzed using Montague's generalized quantifier as shown in the translation (15) into predicate logic.

(15) Every farmer who owns a donkey beats it.

It has been claimed that the truth conditions of (15) should be expressed with the existential quantifier translated as a wide-scoping universal quantifier as shown in (16). However, this is not the translation that can be derived compositionally (Cann et al., 2009, p. 148).

(16) $\forall x \, \forall y \, [[(farmer'(x) \wedge donkey'(y) \wedge own'(x, y)] \rightarrow beat'(x, y)]$

Montague tried to solve this problem using the so-called quantifying-in rule which is beyond the mere application of function. The problem of anaphoric pronouns attracted attention and led to the development of discourse representation theory.

2.1.7 Discourse representation theory

The emergence of dynamic semantics goes back to the work of Kamp and Heim in Discourse Representation Structure and File Change Semantics respectively as an attempt to deal with the problem of anaphora by making a connection between the anaphoric pronoun and its antecedent (Vermeulen, 1995). The introduction of the two theories changed the perspective of dealing with the semantics of natural language from a static perspective to a dynamic one.

In general, DRT is a theory on the semantics of natural language that aims at finding a way to represent the meaning of sentences and texts in a logical language (form). This type of dynamic semantics theories focuses on the context dependence of meaning, as clearly shown in anaphoric pronouns. They have become more concerned with "information change potential" of the meaning of a sentence in addition to its truth conditions. This makes it different from earlier static semantics theories which were only concerned with reference, truth and satisfaction (Kamp et al., 2011, p. 4).

Discourse representation structures (DRSs)

Discourse representation structures (DRSs) are interpretations of sentences and texts in the form of abstract structures (Kamp & Reyle, 1993). It is simply a way to represent the meaning of sentences of natural languages in a formal way. According to Kamp and Reyle (1993), DRSs are obtained through the application of certain rules to the input sentence. Such rules are referred to as "construction rules". A DRS is simply composed of two components: set of discourse referents and set of DRS-conditions to combine the predicates with the chosen discourse referents (Kamp & Reyle, 1993).
The two components are represented in a two-part box; the upper part, referred to as

the head, contains the set of discourse referents and the lower part, referred to as the body, contains the set of conditions.
The general representation of DRSs is show in figure (2.3).

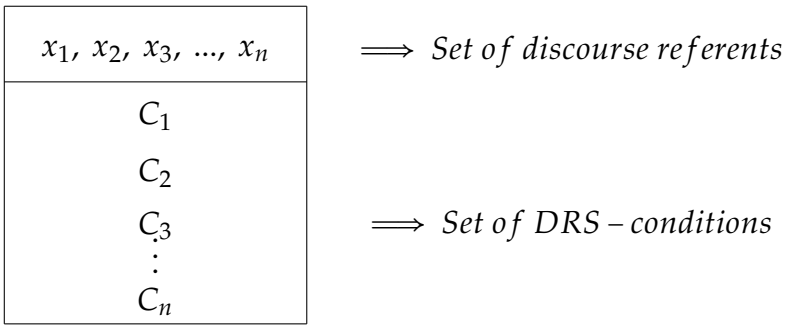


Figure 2.3: General representation of DRSs

Take (17) as an example.

(17) Mary read Hamlet.

First, formal representatives, or discourse referents, must be assigned for the indicated individuals, say x for Mary, and y for Hamlet. Then, we use the so-called DRS-conditions to combine the predicates with the chosen discourse referents (Kamp & Reyle, 1993). This is shown in figure (2.4)

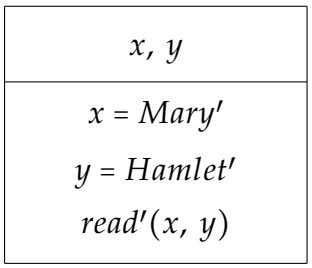


Figure 2.4: DRS of the sentence: Mary read Hamlet.

This DRS consists of two components:

- set of discourse referents, called the universe of the DRS.
- set of DRS-conditions.

The universe of this simple DRS has two discourse referents; x and y, and its set of conditions contains $\{x = Mary', y = Hamlet', read'(x,y)\}$.

Syntax of DRS

The set of DRSs can be defined as follows (Kracht, 2016, p. 75-76)):

Definition 2. (DRS): A DRS is a pair $[\ V\ :\ \Delta]$, where V is a finite set of variables and Δ is a finite set of formulae. The set of DRSs is constructed as:

1. if x is a variable, then $[\{x\}\ :\ \varnothing]$, also written as $[x\ :\ \varnothing]$, is a DRS.
2. if ϕ is a formula, then $[\varnothing :\ \{\phi\}]$, also written as $[\varnothing\ :\ \phi]$, is a DRS.
3. if $[V_1\ :\Delta_1]$ and $[V_2\ :\Delta_2]$ are DRSs, then so are:
   (a) $[V_1\ \cup\ V_1\ :\Delta_1\ \cup\ \Delta_2]$
   (b) $\neg\ [V_1\ \cup\ \Delta_1]$.
   (c) $[V_1\ :\Delta_1]\ \vee\ [V_2\ :\Delta_2]$.
   (d) $[V_1\ :\Delta_1]\ \Longrightarrow\ [V_2\ :\Delta_2]$.

Pronominal anaphora, accessibility and binding

The significance of the DRS lies in its treatment of the problem of pronominal anaphora (the relation between an anaphoric pronoun and its antecedent). In DRS, the relation is considered between an anaphoric pronoun and a discourse referent already presented in the representation, not between a pronoun and a noun phrase. Let's consider, for example, the two anaphoric pronouns "she" and "it" in (18).

(18) Mary read Hamlet. She liked it.

The idea is to find for each pronoun a discourse referent to which the pronoun refers. The DRS of (18) is shown in figure (2.5).

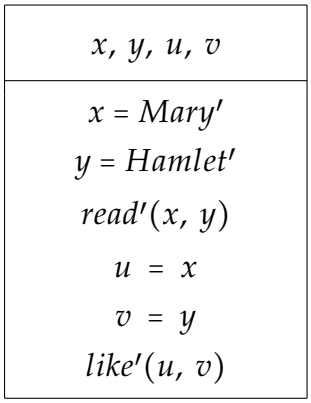


Figure 2.5: DRS of: Mary read Hamlet. She liked it.

To properly handle the relation between discourse referents in the different DRSs, the central notion of accessibility was introduced. Discourse referents are used as a double function tool. That is, they serve as antecedents for anaphoric expressions such as pronouns and at the same time as the bound variables of quantification theory. The second function entails some kind of mutual scope relation among discourse referents which leads to the concept of the sub-DRS or a DRS which occurs in a larger DRS (Kamp et al., 2011). Such a relation is captured by the concept of accessibility.

Definition 3. Accessibility is a relation which must hold between the linked discourse referents and which obtains if, informally speaking, the pronoun occurs within the logical scope of its antecedent.

The idea is that the anaphoric pronouns can refer only to the discourses that are accessible. The accessibility relation is defined in the following way (Kracht, 2016, p. 76)

- Let $\delta = [V : \Delta]$ be a DRS, then δ is accessible to every $\gamma \in \Delta$.
- Let $\delta' \Longrightarrow \delta''$ be a DRS, then δ' is accessible to δ'', but not vice versa.
- Let $\delta' \vee \delta''$ be a DRS, then neither of them is accessible to the other.

The merge of two DRSs is defined taking the union of the two DRSs.

Definition 4. Merge of DRSs if $\delta = [V_1 : \Delta_1]$ is a DRS and $\delta' = [V_2 : \Delta_2]$ is a DRS, then $\delta \cup \delta' \doteq [V_1 \cup V_2 : \Delta_1 \cup \Delta_2]$

Recall the famous donkey sentence “Every farmer who owns a donkey beats it.”. This sentence can be represented using the following DRS:

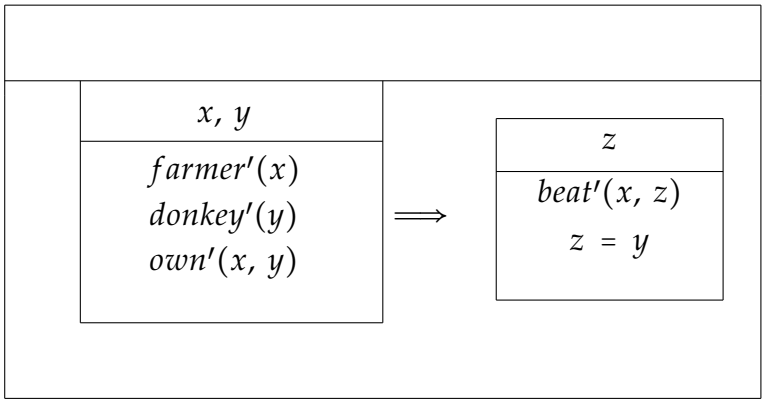


Figure 2.6: DRS for the donkey sentence

Following the point of view of model theoretic semantics that the place of possible worlds is taken by models, DRSs are evaluated, in terms of truth and falsity, relative to a model. A model is defined as “a certain information structure, relative to which it is possible to evaluate the expressions of a some given language (DRSs in this case) and in particular to evaluate the sentences of that language in respect of truth and falsity” (Kamp & Reyle, 1993, p. 93). Take (19) as an example.

(19) A brave policeman chased a dangerous criminal.

This sentence can be translated using DRS in figure (2.7).

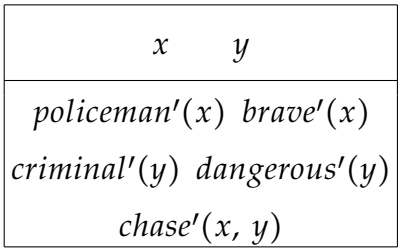


Figure 2.7: DRS of: A brave policeman chased a dangerous criminal.

This DRS is evaluated as true in a model M if and only if there is an x such that x is a brave policeman, a y such that y is a dangerous criminal and x *chased* y.
The previous sentence consists of four word types: an (indefinite) article, an adjective,

a noun and a (transitive) verb with the following DRSs, taking into consideration that adjectives and nouns have a similar DRS. This is given in figure (2.8).

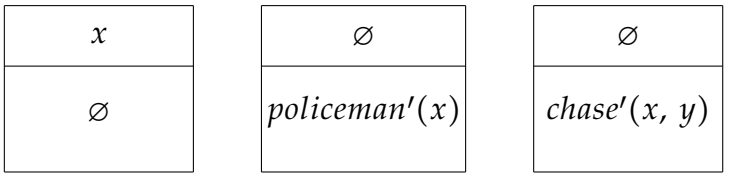


Figure 2.8: DRSs for three different word types

The union of each two DRSs is taken to represent the semantics of their corresponding constituent. That is, we take the union of the DRS of the indefinite article "a" and the DRS of the adjective phrase "brave man" to get the semantics of the determiner phrase "A brave policeman". The result of the union operation is shown in figure (2.9).

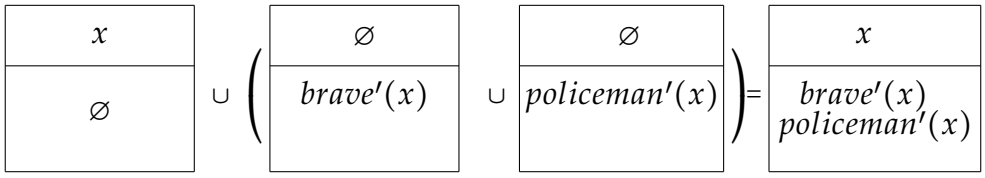


Figure 2.9: Union of the DRSs of the phrase "A brave policeman"

The same operation is applied to the other determiner phrase "a dangerous criminal" as represented in figure (2.10)

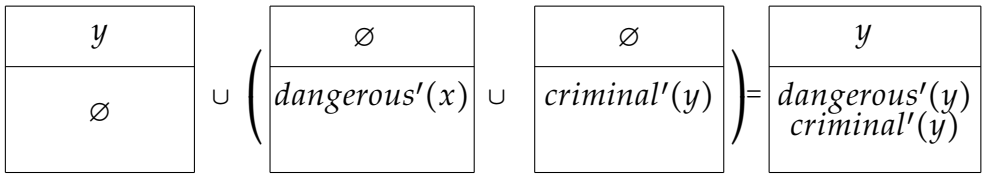


Figure 2.10: Union of the DRSs of the phrase "a dangerous criminal"

Similarly, the union of the DRS of the transitive verb "chased" and the determiner phrase "a dangerous criminal" is shown in figure (2.11)

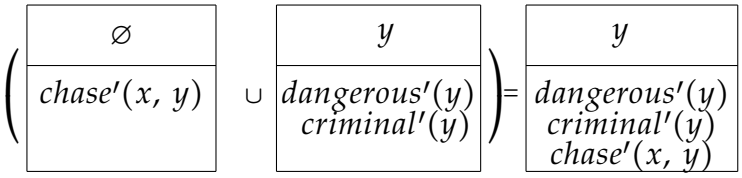


Figure 2.11: Union of the DRSs of the verb “chased” and the phrase “a dangerous criminal”

Finally, the union of the DRS of the determiner phrase “A brave policeman” and the verb phrase “chased a dangerous criminal” is taken to get the semantics of the complete sentence as shown in figure (2.12)

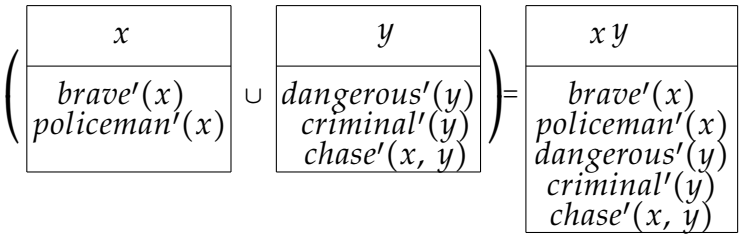


Figure 2.12: Union of the DRSs of the phrase “chased a dangerous criminal” and the phrase “a brave policeman”

The idea that the compositional operation is simply accomplished by taking the set union was introduced by Henk Zeevat (1989). He called this operation the merge and gave it this general form:

$$\delta \bullet \delta' = (V \cup V',\ \Delta \cup \Delta')$$

The merge of two DRSs is defined with respect to a specific model if and only if both are true in this model. However, the solution of the problem of the object identification across sentences, introduced by the DRT, through the systematic use of free variables, has advantages and disadvantages. On the one hand, the variables become more dynamic and are allowed to be bound retroactively. However, the problem of choosing proper names for the variables remains without an effective solution. The solution, suggested by Zeevat, that the meaning is a union of the condition of the two semantics $\Delta \cup \Theta$ is far from perfectness; because it requires an extremely careful choice of variable

whenever we come to draw a new meaning (U. Klein & Kracht, 2011; Kracht, 2016, p. 72). This is clear in the following example.

(20) Peter loves Mary.

Let the DRS for "Mary" introduce the referent x, "Peter" introduce y and DRSs be merged by the operation $\langle U_1, C_1\rangle \bullet \langle U_2, C_2\rangle = \langle U_1 \cup U_2, C_1 \cup C_2\rangle$. The result can be represented as follow:

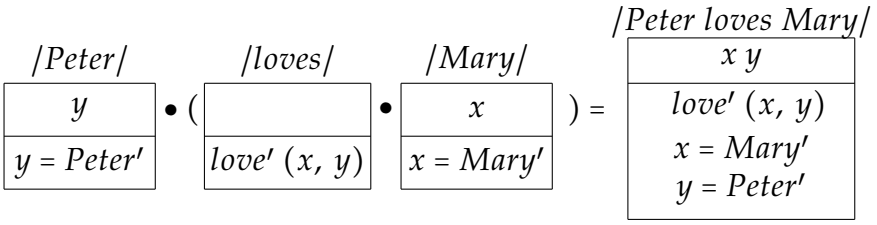


Figure 2.13: Merge operation of the sentence: Mary love Peter.

This is the right meaning of "Mary love Peter" but not for "Peter love Mary". The problem is precisely related to the haphazard identification of the variables. A solution for the accidental identification of variables has been suggested by Vermeulen (1995) using the so-called referent systems (RS) to manage the choice of variable names. Vermeulen has introduced an alternative approach to the merge operation, providing a flexible mechanism for working with variables, which distinguishes the variables themselves from their names, and from the information stored in them (values). In AMRS, Kracht has modified the work of Vermeulen allowing the variables to be identified depending on the overt morphosyntactic features rather than the position of the argument. We introduce both RS and AMRS in some detail in chapter three.

2.2 Arabic from a formal perspective

There is a consensus that the amount of work and research on Arabic computational models of semantics is very limited (Habash, 2010). The research community has neglected this area of research due to its higher complexity and subtlety (Habash, 2010; Haddad & Yaseen, 2003). The term "computational models" is used in the literature

in its broad meaning, i.e., all the approaches and the attempts related to the computational processing of meaning, including developing semantically annotated corpora or machine-readable lexical resources. However, the formal representation of meaning of the Arabic sentence represents a special case of neglect.

Scholars usually count the same reasons for this neglect, be it the high degree of complexity or the lack of collaboration between specialists in related areas like logic, computer science and linguistics. Although I agree with what they say, I believe these are not the only reasons. Arabic linguists have been much more concerned with the description of Arabic within famous frameworks, particularly the Chomskyan versions of generative grammar. They focused on syntactic structure and disregarded semantics. The majority of work and research followed this tradition. Some other attempts have been devoted to non-transformational approaches that relied on the lexicon rather than the transformation rules, including LFG and HPSG. Nevertheless, syntax has always been what really mattered. As for NLP, Arabic has received increasing attention in the last three decades, and has, in fact, achieved a considerable progress in some areas, especially morphology and syntax, but the amount of research in semantics is far smaller than other areas of NLP. This was the case before the dramatic movement in this area towards the statistical methods, like Support Vector Machine, Hidden Markov Models, Maximum Entropy Models ... etc. and the subsequent advances of deep learning algorithms. Accordingly, the focus has been moved to issues related to these approaches like the improvement of the accuracy of the developed tools, moving to dialects, changing the amount or the type of data ... etc. In such a situation, the logic-based representation of meaning in Arabic has been left almost without apostles.

According to El-Sayed (2019), only three attempts have been introduced to formally represent the meaning of Arabic sentence in the last three decades. Each decade has witnessed one attempt. These attempts have adopted either a Montagovian formalization, or a unification-based[2] syntax semantics interface. Despite this modest number of works, Elsayed has devoted his paper to provide a survey and a classification of them, explain their structure and point out their limitations. El-Sayed classifies the three

[2]Unification-based formalism is a term used to refer to a variety of formal devices, including HPSG applied by the researchers, in which the central operation is unification (Crystal, 2008; Shieber, 2003) syntax semantics interface

attempts into:

- Syntactic meaning representation: (Al-Johar, 1999; Al-Johar & McGregor, 1997).
- Surface semantic representation: (Haddad & Yaseen, 2001, 2003, 2005; Haddad, 2007).
- Analytical syntax-semantics representation (El-Sayed, 2015, 2011).

Al-Johar and McGregor (1997) attempted to represent Arabic statements formally using Generalized Phrase Structure Grammar (GPSG) in the syntactic treatment whereas the meaning is expressed depending on Montague semantics. They argued that unlike the Arabic traditional theories of the ʔamal (اَلعَمَل 'governance')[3] and the ʔisnaːd (اَلإِسنَاد 'subject/predicate relationship')[4], which provide neither a sufficient and full formal description for Arabic statements nor a mechanism for the meaning representation, GPSG is a powerful grammar theory for the syntactic analysis of Arabic. They ascribed the problems of formal representation of meaning of Arabic sentences to the difference between Arabic and Indo-European languages, to which model-theoretic semantics was applied, word order and sentence consistency (Al-Johar & McGregor, 1997). However, they did not manage to completely disregard both theories, but to use them as a guide to assure the validity of the treatment for Arabic (Al-Johar, 1999, p. 9). That is, at the phrase level, the theory of governance was used to examine the construction of each phrase, and at the sentence level the predicate was treated as a function from the subject to the sentence (Al-Johar, 1999, p. 90-91).

They proposed a classification of Arabic sentences into three groups: verbless sentences (nominal sentences) (21), verbal sentences (22) and sentences of individuals (23). Depending on the type of the sentence, they proposed a formal representation that corresponds to its structure, as shown in (21).

[3]Governance is a syntactic principle wherein certain words cause others to inflect in particular ways- not in agreement with the ʔaːmil (اَلعَامِل 'governing word'), but as a result of its effect (Ryding, 2005).

[4]It is the assignment of the predicate to the subject. Arabic sentences are classified depending on the first word in the sentence. In the nominal sentence the faːʔil (اَلفَاعِل 'subject') or the mubtadaʔ (اَلمُبتَدَأ 'topic') is called the musnad ʔilajhi (اَلمُسنَد إِلَيه) and the xabar (اَلخَبَر 'predicate') is called the musnad (اَلمُسنَد). In the verbal sentence, the subject is called the musnad ʔilajhi and the predicate is called the musnad.

Word/Phrase	Logical Form	Formal representation
PN “ʔaħmad”	logical constants	Ahmed
Indef NP “tˤaːib”	one-place predicate	$\lambda x.$tˤaːib(x)
Indef AP “tˤaːib muʤdd”	one-place predicate joined by $\wedge$	$\lambda x.$ (tˤaːib$(x) \wedge$ muʤdd(x)
Def NP “atˤtˤaːib”	generalized quantifier	$\lambda Q[\exists y[\forall x[$tˤaːib$\{x\} \Leftrightarrow x = y] \wedge Q\{y\}]]$
Intrans V “naʤaħa”	one-place predicate	$\lambda x.$naʤaħa(x)
Trans V “kataba”	two-place predicate	$\lambda x \lambda y.$kataba(y, x)
Ditrans V “darrasa”	three-place predicate	$\lambda x \lambda y \lambda z.$darrasa$(z, y, x)$
Preposition “fiː”	two-place predicate	$\lambda x \lambda y.$fiː(y, x)

Table 2.2: Syntactic categories and their meanings by Al-Johar and McGregor

(21) أَحمَدُ طَالِبٌ مُجِدٌّ

ʔaħmadu tˤaːlibun muʤddun
Ahmed.nom.sg.m student.nom.sg.m hard-working.nom.sg.m
‘Ahmed is a hard-working student’ (Al-Johar, 1999; 9)

(22) دَرَّسَ مَازِنٌ عَلِيًّا بَاسكَالَ

darrasa mːzinun ʕalijjan baskaːa
teach.3.sg.m.pst Mazin.nom.sg.m Ali.acc.sg.m Pascal.acc.sg.f
‘Mazin taught Ali Pascal’ (Al-Johar, 1999; 99)

(23) مَن أَحمَدُ ؟

man ʔaħmadu
who.nom Ahmed.nom.sg.m
‘Who is Ahmed?’ (Al-Johar, 1999, p. 100)

They adopted a simple approach to provide the syntax-semantic interface. Since their basic motivation was to develop a natural language interface from Arabic to SQL, each Arabic query was provided with syntactic and semantic analysis. To do this, Arabic expressions were classified into syntactic categories and each category was provided with its formal meaning as shown in table (2.2):
Once sentences are syntactically parsed according to the rules of GPSG, the constituent phrases are provided with their semantic representations as shown in table (2.2). As an example, we consider the analysis of (24).

(24) اَلطُّلَابُ نَاجِحُونَ

atˤtˤulaːbu naːʤiħuːna
student.nom.pl.m.def successful.nom.pl.m
'The students are successful.'
$\exists y[\forall x[\text{tˤaːib}\{x\} \Leftrightarrow x = y] \wedge \text{naːʤħ}\{y\}]$ (Al-Johar, 1999, p. 95)

Another attempt to represent the meaning of the Arabic sentence formally within a unification-based syntax-semantics interface was led by Haddad and Yaseen (Haddad & Yaseen, 2001, 2003, 2005; Haddad, 2007). They have introduced their project in several publications either in a co-authorship format or individually by Haddad. Despite their argument of the absence of a formal theory that is capable of handling all phenomena involved in Arabic semantic processing, they tried to provide an Arabic syntax-semantics interface using the theory of HPSG (Pollard & Sag, 1994). Specifically, they managed to apply a lambda-conversion process to represent the meaning of Arabic sentences formally depending on the compositional rules that express the meaning of syntactic categories of Arabic (Haddad, 2007). In their model, Arabic determiners, as al (ال 'the'), kull (كُلّ 'all') and baʕdˤ (بَعض 'some') are dealt with as generalized quantifiers represented as follows:

$$\| \mathit{Quant} \| \Rightarrow \lambda R \lambda S \, (\mathit{Quantifier}(R, S))$$

These Arabic generalized quantifiers combine a restriction R and scope S and therefore the definite Arabic determiner is represented as :

$$\| \text{ال} \| \Rightarrow \lambda R \lambda S (\ \text{ال}(x, R \wedge S))$$
$$\| \mathit{the} \| \Rightarrow \lambda R \lambda S \, (\mathit{the}(x, R \wedge S))$$

To consider the function of the definite Arabic determiner, the example in (25), taken from (Haddad & Yaseen, 2003), shows their approach.

(25) يَتَعَلَّمُ اَلوَلَدُ اَلعَرَبيَّة

jataʕallamu lwaladu lʕarabijata
study.3.sg.m.prs boy.nom.sg.m.def Arabic.acc.sg.f.def
'The boy studies Arabic.' (Haddad and Yaseen, 2003)

The sentence in (25) is classified as a verbal sentence and is analysed as:

$$\|VS\| \xrightarrow{sem} \|Subj\|(\|Obj\|(\|Verb\|)) \quad (1)$$

$$\lambda R\lambda S(\text{ال}\ (x,\ R \wedge S))\ (\|\text{وَلَدُ}\|)\ (\|\text{يَتَعَلَّمُ العَرَبيَّة}\|) \quad (2)$$

$$\lambda R\lambda S(\ \text{The}\ (x,\ R \wedge S))\ (\|\ boy\ \|)\ (\|\ studies\ the\ Arabic\ \|)$$

$$\lambda S(\text{ال}\ (x, \text{وَلَدُ}(x) \wedge S))\ (\|\text{يَتَعَلَّمُ العَرَبيَّة}\|) \quad (3)$$

$$\lambda S(\ The(x, boy(x) \wedge S))\ (\|\ studies\ the\ Arabic\ \|)$$

$$\text{ال}\ (x, \text{وَلَدُ}(x) \wedge \text{ال}\ (y, \text{اَلعَرَبيَّة}(y) \wedge \text{يَتَعَلَّمُ}(x,y))) \quad (4)$$

$$The(x, boy(x) \wedge\ The(y, Arabic(y) \wedge study(x,y)))$$

Following (Bos, Mastenbroek, MacGlashan, Millies, & Pinkal, 1994) in their λ-DRT, Haddad and Yaseen managed to combine the λ-conversion process in the DRT to handle the problems of representing anaphoric aspects in Arabic. DRT is very well-known for dealing with cross-sentential anaphora, but in a less elegant way with phenomena such as quantification, which pushed some logicians to attempt to take the advantages of both approaches and combine DRT and the compositionality of Montague (Muskens, 1996). λ-DRT combines features of DRT and Montague-style extended type theory. DRSs are taken as basic meaning representation structures, and the λ-abstraction is allowed over the DRSs to derive the representations of complex types. Another ingredient in the

language of λ-DRT is the merge operation $\otimes$ by taking the union of the sets of the discourse markers and the conditions (Bos et al., 1994).

El-Sayed (2011, 2015) provides another attempt to represent Arabic formally through the application of Montague grammar to Arabic. Following Montague's approach in developing a syntax to a fragment of English, El-Sayed divides syntactic categories into: primitive categories and complex categories. He uses almost the same category indexes, definitions and names proposed by Montague. He uses, for example, the indices (e, t, T, IV, TV) to refer to the sentence, entity expression, term, intransitive verb phrase and transitive verb phrase respectively. The syntactic derivation of expressions in Montague grammar is represented by trees; and using Montague's rule-to-rule translation method, Arabic expressions are translated into the language of intensional logic. The whole approach is illustrated by the example in (26)

(26) حَضَرَ جَميعُ اَلطُّلَابِ

ħadˤara ʤamiːʕu atˤtˤullaːbi
come.3.sg.m.pst all.nom student.gen.pl.m.def
'All students came.' (El-Sayed, 2015)

El-Sayed argues that Montague's syntactic and semantic rules and derivations can be applied to Arabic expressions in a straightforward way. He first specifies the syntactic categories in (26) as follows: atˤtˤullaːbi (اَلطُّلَّابِ 'students') is a common noun (CN), ʤamiːʕu (جَميعُ 'all') is a determiner (DET) and ħadˤara (حَضَرَ 'came') is an intransitive verb (IV). The syntactic analysis is represented bottom-up and right-left as shown in figure (2.14)

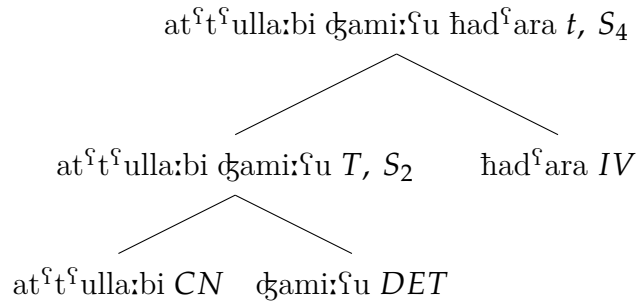


Figure 2.14: Analysis tree of the sentence: ħadʔara ʤamiːʕu atʔtʔullaːbi

Using the rule-to-rule translation, El-Sayed translates (26) into the language of intensional logic as shown in figure (2.15).

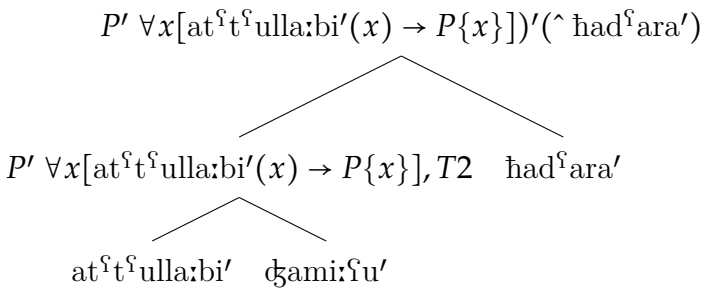


Figure 2.15: Rule-to-rule translation of sentence: ħadʔara ʤamiːʕu atʔtʔullaːbi

In spite of his argument that logical computation of Arabic using Montague grammar is considerably achievable, he refers to two limitations (El-Sayed, 2015):

- Montague's syntactic and semantic rules do not fit well with certain Arabic expressions of quantification.
- Some Arabic quantifiers cannot be represented using these rules.

Each of the three attempts have tried to apply one of the famous approaches in formal representation of natural languages. The three applied approaches have problems, regardless of the language used as a case study, be it Arabic or some other language. The meaning representation in (Al-Johar, 1999) seems a simple first-order logic, rather than a montagovian framework, as they claim, and it lacks the expressiveness to capture the semantic structures of Arabic. El-Sayed's attempt is more mature in this direction; however, Montague grammar cannot deal with the anaphoric pronouns. Montague tried to

solve the problem of object identification using the so-called Quantifying-In rule which involves means that are beyond the mere application of function. Also, the systematic use of free variables in DRT forces using the same names of variables to be able to take the set theoretic union (Kracht, 2016).

2.3 Summary

In this introductory chapter, I tried to give a brief history of formal semantics with focus on the concept of compositionality and how it has shaped the history of semantics in the last forty years. In his influential work, Richard Montague has presented a first attempt to represent the relationship between syntax and semantics in a compositional way. His framework has attracted great attention and has also been thoroughly scrutinized, most notably for the identification of objects across sentences (Kracht, 2008). The problem of anaphoric pronouns has led to the development of discourse representation theory (Kamp et al., 2011; Kamp & Reyle, 1993). The DRT has solved this problem by the systematic use of free variables. However, taking the set union on the set of markers, the idea of Henk Zeevat (1989), has created the problem of choosing the names of the variables. The theory of RS, due to Kees Vermeulen (1995) has provided an elegant mechanism for working with variables in which the sets of variables are made disjoint, and the names of the variables are assigned locally, instead of assigning them globally in Zeevat's approach. Kracht (2016) has modified this version of RS to solve the problem of directionality in the renaming process. Arabic has received little attention in this area and the limited number of works have applied either Montague framework or a unification-based formalism. Therefore, they have the same problems.

Chapter 3

Agreement Morphology and Referent Systems: A Theory of Semantic Composition

This chapter explains in detail the theory of Agreement morphology and referent systems (AMRS) as introduced mainly and thoroughly in (Kracht, 2016) and in sparse in an array of publications (Aboamer & Kracht, 2019; Kracht & Aboamer, 2017; U. Klein & Kracht, 2011; Kracht, 2008, 2007). It involves a key role for overt morphological features in the string-to-meaning translation. Due to its strict surface orientations, the approach is basically lexicalist; all the linguistic information is encoded in the lexicon, because the rules of sign formation are supposed to be universal and the difference between languages should be only in the lexicon. The lexicon is composed of lexical entries, referred to as signs. Each sign is a combination of its morphological description and semantics. A sign δ is a triple $\delta = \langle \mu, \alpha, \Delta \rangle$, such that μ is a morpheme, α is the argument structure and Δ is the meaning of the sign. The morpheme is the morphological component of the sign and both the argument structure and the meaning compose its semantic structure. I first introduce the theory of RS and then each part of the three respectively following Kracht in the usage of notions and symbols.

3.1 Referent systems

The introduction of Referent systems RS was motivated by the attempt to find a practical way to avoid the accidental identification of variables and provide a flexible mechanism for working with them (Vermeulen, 1995). Following the notions used in computer science, a distinction is made in RS between the variable itself, as a location in the memory, the syntactic variable or its name and the information stored in it. Vermeulen defines a referent system as follows:

Definition 5. (Referent System): A referent system is a triple $\langle I, R, E \rangle$, where:

1. R is a finite set of referents;
2. I is a partial injective function from NOM (a stock of names) to R, Import function;
3. E is a partial injective function from R to NOM, Export function.

Both the import and export functions are used to manipulate referents (variable names). The idea is similar to the idea of supply and demand in economy where each company says what it offers or what it actually needs. Similarly, a referent system with an import function tells us what referents are needed from the context and under which name. On the other hand, the export function says which referents are offered and under which name. So, if we have, for example, $I(A) = x$, this means x is imported under the name A and if we have $E(x) = A$, then x is exported under the name A. As we can notice, in the first case, the referent (variable name) is picked from the NOM set because the referent does not exist but it is introduced. In contrast, the referent in the second case does exist in the system.
The merge operation of two referent systems is defined by taking the disjoint union rather than the set union as suggested by Zeevat. The definition of the merger of two referent systems is as follows: (Kracht, 2016; Vermeulen, 1995)

Definition 6. Let $\mathfrak{R} = (\alpha : x : \beta)$ and $\mathfrak{R}' = (\gamma : y : \epsilon)$ be two referent systems. The merger of the two systems $\mathfrak{R} \bullet \mathfrak{R}'$ is defined as:

1. if $\beta \neq \gamma$ (the export name of the first referent does not equal the import name of the second referent), then, $\mathfrak{R} \bullet \mathfrak{R}$' is one of the following four possible referent systems:

   (a) $\{\alpha : x^1 : \beta, \gamma : y^2 : \epsilon\}$ given neither $\alpha = \gamma$ nor $\beta = \epsilon$;

   (b) $\{\alpha : x^1 : \beta, - : y^2 : \epsilon\}$ given $\alpha = \gamma$ but $\beta \neq \epsilon$;

   (c) $\{\alpha : x^1 : -, \gamma : y^2 : \epsilon\}$ given $\alpha \neq \gamma$ but $\beta = \epsilon$;

   (d) $\{\alpha : x^1 : -, - : y^2 : \epsilon\}$ given $\alpha = \gamma$ and $\beta = \epsilon$.

2. if $\beta = \gamma$ (which means that x and y refer to the same variable), then the resulting referent system is $\{\alpha : x^1 : \epsilon\}$.

This model can be explained as follows[1]: each unit is provided with its semantics, if it has, and a referent system. The referent system is composed of a variable and its combinatoric information. The combinatoric information specifies the name under which the variable is exported and the name under which the variable is imported. It is important to note that the name can be empty. We can apply this description to the sentence in (27).

(27) Mary bought a car.

The sentence in (27) is composed of four lexical items, whose referent systems are shown in figure (3.1) according to their occurrence in the sentence.

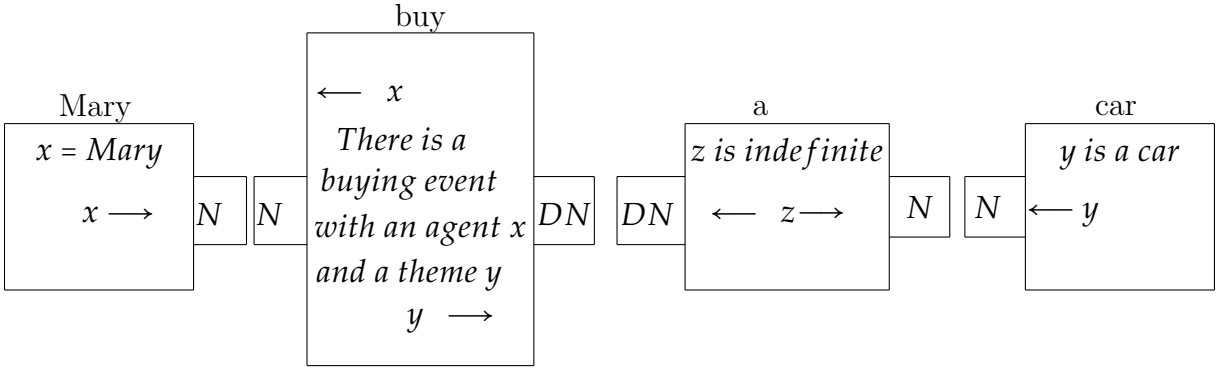


Figure 3.1: Referent systems of the lexical items of: Mary bought a car.

[1]I benefited a lot in this part from (Keil, 2005), but I used a different example and tried to simplify the explanation of the model.

We can simply observe that both “Mary” and “car” have similar representations. They differ only in the semantics. The proper name “Mary” has the semantics $x = Mary$ and its semantic variable is associated with the name N in the right side, which means that this referent is exported under the name N and it occurs, in this example, to the left of the verb. It can also be seen that “Mary” has no import name, i.e., the import name is empty. The name is randomly chosen. It is the first letter of the word “noun”. The lexical item “car” is almost represented in the same way; it has the semantics $y\ is\ a\ car$, is associated with the name N in the left side, because it occurs to the right of the indefinite article, and has no import name. The indefinite article “a” is associated with the semantics $z\ is\ indefinite$ and its semantic variable has two names: an import name N in the right side, which means that it imports the referent z under the name N and export name DN, for “determiner and noun” in the left side, which means that the referent z is exported under the name DN. The lexical unit “buy” has the semantics $there\ is\ a\ buying\ event\ with\ an\ agent\ x\ and\ a\ theme\ y$. It has two referents: one has the import name N, and the second has the import name DN. The referent system of the verb “buy” imports two variables, but one from its left side, i.e., the subject or the agent and the other from the left side, i.e, the object or the theme.

The combination process in RS, also called the merge, is fully associative. This means that any two lexical items can combine, and the result is also a referent system. But the result of the merge varies depending on the match or the mismatch between the export and the import names. If we reconsider the representations of the lexical items of (27), we observe that the variable in the referent system of the determiner “a” is imported from the lexicon under the name N and the variable in the referent system of the lexical item “car” is offered under the same name N, which means that the import name and the export name match and variables are unified. Therefore, they can merge as shown in figure (3.3).

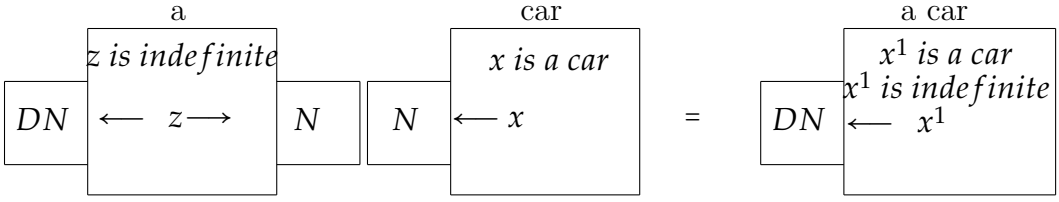


Figure 3.2: Merge of the two items “a” and “car”

The merge result of the two items “a” and “car” gives “a car”. The merge result is computed as follows:

- the two lexical items combine;
- the union of the two semantics is taken;
- the import name disappeared;
- the identified variable is renamed to x^1.

The last point in the computing process shows how variables are renamed if they are identified. Nevertheless, the calculus allows any two adjacent referent systems to combine, even if the variable is not identified. In this case, the disjoint union of the variables is taken, as shown with the merge of the two items “car” and “buy” in figure (3.3).

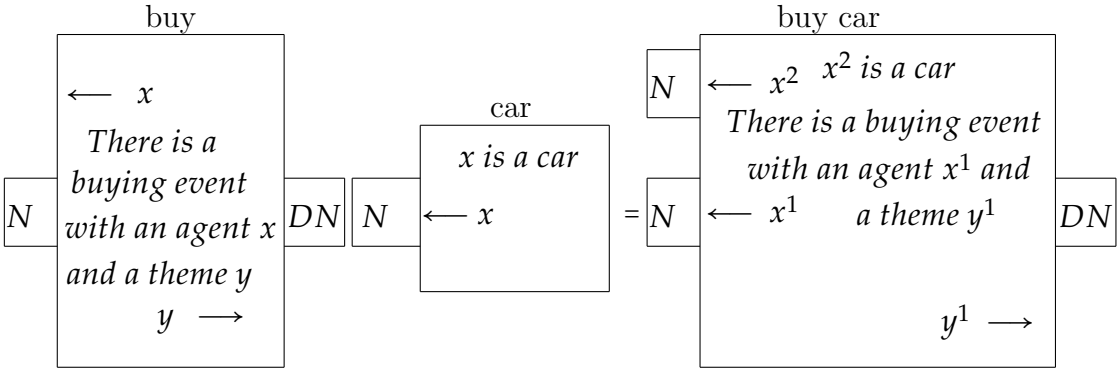


Figure 3.3: Merge of the two items “car” and “buy”

In (3.3), the referent system of the lexical item “buy” imports two referents: one under the name DN from its right side and the second under the name N from its left side. It merges with the referent system of the lexical item “car”, in which the variable is exported under the name N. Although one of the two imported referents in the referent system of “buy” is imported under the name N, it does not unify with the variable exported under the same name because of the direction. The name N is opposite to the DN, rather than the similar N. In this case, the two referent systems are allowed to merge, but the variables are renamed differently. That is, the merge of the two referent systems results in a third referent system in which the variable in the left referent system (of the functor) is renamed as x^1 instead of x and y to y^1 respectively, and the

variable of the right referent system (of the argument) is renamed as x^2. The rest is similar to the case of unified variables, i.e., the two lexical items concatenate and the union of the semantics is taken. Recalling the definition of the referent system, we can define the corresponding sets and functions of each lexical item. I first define the set of names (NOM) and the set of referents (R).

- NOM = $\{N, DN\}$
- R = $\{x, y, z\}$

Then, we specify the sets and functions.

- Mary: R = $\{x\}$, I = $\varnothing$, E = $\{\langle$ N, x$\rangle\}$
- buy: R = $\{x, y\}$, I = $\{\langle$ x, N$\rangle, \langle$ y, DN$\rangle\}$, E = $\varnothing$
- a: R = $\{z\}$, I = $\{\langle$ z, N$\rangle\}$, E = $\{\langle$ DN, z$\rangle\}$
- car: R = $\{y\}$, I = $\varnothing$, E = $\{\langle$ N, y$\rangle\}$

Before we move to the problems of this model, we can summarize its basic features in the following three points:

1. The merge operation is fully associative and the correct translation does not require a constituent structure.
2. The merge operation takes the disjoint union of the variables rather than the set union and therefore the name of the variable is not significant.
3. The semantic composition depends on the renaming of variables whether they are unified or not.

The model works well in cases similar to (27), because if the hearer can interpret sentences as s/he perceives them, s/he can easily identify the agent and theme of the verb depending on their position. This is the case in the languages with fixed word order either SVO or OVS. However, the model starts suffering when complexities appear, especially with the languages with free word order and VSO languages as by definition a VSO language allows two arguments (subject and object) to occur next to each other

unlike SVO languages. To be more specific, the original version of referent systems does not allow two arguments with the same name to occur at the same side. In the subordinate clause of a sentence like (28), the two verb arguments appear in the left side of the verb, which occurs at the end of the sentence.

(28) ..., dass das Mädchen das Buch liest.
..., that the.nom.sg.n girl.nom.sg.n the.acc.sg.n book.acc.sg.n read.3.sg.prs
'The mother said that the girl is reading the book'

In this case, the referent system of the verb is supposed to import its two arguments with the same name and from the same direction, because the definite article "das" 'the (neutral)' takes the same form in the nominative and accusative cases. This is shown in figure (3.4).

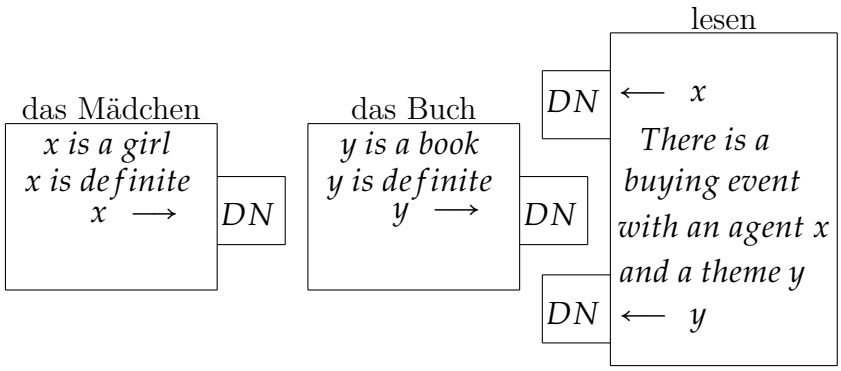


Figure 3.4: Referent systems of the verb "lesen" and its two arguments

This case is problematic because the verb cannot import two arguments with the same names from the same side. It can take only one argument per name (Keil, 2005). Therefore, the theory fails to deal with a big number of similar cases due to its full reliance on directionality. Similarly, Arabic is primarily described as a VSO language but it has relative free word order, including the possibility to have two arguments at the same side, and directionality cannot play a significant role.

Kracht (2016) has built on the work of Vermeulen and introduced a modified version of RS, in which he disregarded the distinction between left and right context and presented it as a relation between a functor and its arguments. He also dealt with names as attribute value structures (AVSs) which contain morphosyntactic features. This makes it look like the argument structure, whose main function is to declare how the functor's

semantic arguments are realized on the surface. I devote the rest of this chapter to present Kracht's approach but I can generally list the main changes in the following four points:

i Variables are identified based on the match of morphosyntacic information rather than the import and export names.

ii This means that every variable is provided with its relevant identification information.

iii The export and import functions have different roles with respect to the merge operation, rather than specifying the order.

iv The new role of the import function is to specify what name a variable takes to be identified for merge, whereas the export function states the name of the variable after merge.

3.2 The morphological component

At the simplest level of description, a natural language is simply described in AMRS as a set of strings over a given alphabet A. A^* denotes the set of all strings over A. The concatenation of two strings x and y is denoted by $x\hat{}y$ or simply xy. Concatenation is associative, that is, $(x\hat{}y)\hat{}z = x\hat{}(y\hat{}z)$, $\langle A^*, \hat{}, \varepsilon\rangle$ constitutes a monoid (Partee et al., 1990). In the sequel, variables for strings are formed using a vector arrow, e.g. $\vec{x}$.

3.2.1 Glued strings

Kracht suggests the notion "glued string" to describe how strings concatenate with each other. The lexical units are described as glued strings rather than just strings. A glued string is a string with two context conditions: one for the left context and one for the right context. These conditions specify the properties of the string $\vec{x}$ such that $\vec{x}$ can appear in $\vec{u}\,\vec{x}\,\vec{v}$. The left and right conditions are simply requirements. So, before the definition of the glued string, the notion of a requirement is introduced.

Definition 7. (Requirement): A requirement is a pair $(s, \vec{x})$, where s is a sign[2] and $\vec{x}$ a string. Roughly, the context must contain one of the strings with sign + as suffix (if on the left) and as prefix (if on the right), while avoiding all the strings with sign −.

The glued string can be defined as follows:

Definition 8. (Glued String): A glued string is a triple $j = \langle L, \vec{x}, R \rangle$, where L is a set of left requirements, $\vec{x}$ is a string, and R is the set of right requirements.

The following two explanatory examples, taken from (Kracht, 2016) illustrate the notion. (/␣/ denotes the blank):

1. $j = \langle \{(+, \text{ch}), (+, \text{s}), (+, \text{sh}), (+, \text{x}), (+, \text{z})\}, \text{es}, \{(+, ␣)\} \rangle$
   This example codes the fact that the plural morph /es/ in English is only suffixed to words ending in /ch/, /s/, /sh/, /x/ or /z/, while it must strictly be at the end of the word.

2. $j_1 = \langle \{(+, \text{b})\}, \text{bal}, \varnothing \rangle$
   This example from Hungarian specifies that the instrumental form /bal/ appears only after /b/.

3.2.2 Occurrence

A string occurs in another string if the first is a substring of the second. Let $\vec{x}$ and $\vec{y}$ be two strings. An occurrence of $\vec{x}$ in $\vec{y}$ is a pair $o = \langle \vec{u}, \vec{v} \rangle$ such that $\vec{y} = \vec{u}\,\vec{x}\,\vec{v}$. For example, let $\vec{y}$ = /disagreement/ , $\vec{u}$ =/dis/ and $\vec{v}$ = /ment/, then $\vec{x}$ =/agree/. If $o_1 = \langle \vec{u_1}, \vec{v_1} \rangle$ is an occurrence of $\vec{x_1}$ in $\vec{y}$ and $o_2 = \langle \vec{u_2}, \vec{v_2} \rangle$ is an occurrence of $\vec{x_2}$ in $\vec{y}$, then it is said that o_1 is to the left of o_2 and (o_2 is to the the right of o_1) if $\vec{x_1}$ is a prefix of $\vec{u_2}$. If $\vec{u_2} = \vec{u_1}\vec{x_1}$, it is said that o_1 is immediately to the left of (left adjacent to) o_2 and o_2 is immediately to the right of (right adjacent to) o_1. The two occurrences o_1 and o_2 are said to be contiguous if one of them is left or right adjacent to the other. Otherwise, they overlap.

[2]There should not be a confusion with the term sign as the units of the lexicon. A sign here is just plus + or minus −

3.2.3 Morphological class

The notion of "glued string" is developed to handle the conditions of combination. However, in many cases this notion does not suffice to handle all possible conditions. This is the actual case of string combination in many languages. Kracht (2016, p. 46) gives two examples to show that the combination of strings goes beyond the mere phonology. The first example is the German word /Bank/ which has two meanings: if it means the "monetary organization", it has the plural form /Banken/ and if it means the "bench", it takes the plural form /Bänke/. This cannot be captured with the phonology alone, but the meaning should also be considered. The second example is vowel harmony in Hungarian, which does not depend entirely on phonology. Arabic also has similar cases that will be discussed in chapter four. Therefore, an additional mechanism is introduced to capture such cases. This is the notion of "morphological class". These are properties of individual morphs, not morphemes, that control the behavior of a morph under combination. When two morphs are to be combined, one of them takes the role of the argument, while the second takes the role of the functor. When a morph m_1 takes a morph m_2 as its argument, the two give a third morph m_3. This can be written in a form of a function:

$$m_1(m_2) = m_3$$

The German perfect form /aufgemacht/ 'opened' is composed, for example, form two discontinuous morphs, i.e., /auf⊗mach/ and /ge⊗t/. In the combination process, the morph /ge⊗t/ takes the morph /auf⊗mach/ as its argument and the combination process results in the perfect form:

$$m_1(m_2) = m_3$$
$$/\text{ge}\otimes\text{t}/(/\text{auf}\otimes\text{mach}/) = /\text{aufgemacht}/$$

The morph that takes the role of the functor (m_1, in this case, i.e, /ge⊗t/) has two classes, an "ingoing class" and an "outgoing class". The ingoing class specifies what class the argument (m_2, in this case /auf⊗mach/) must have to combine with m_1. The outgoing class states what class the combination of m_1 and m_2 has. Since the

number of features that guide the combination can be large and the morphological classes themselves can be rather complex, an attribute value matrix (AVM) in the following form is suggested:

$$\begin{bmatrix} \text{ATTRIBUTE}_1: & \text{valueset}_1 \\ \text{ATTRIBUTE}_2: & \text{valueset}_2 \\ \ldots & \ldots \\ \text{ATTRIBUTE}_n: & \text{valueset}_n \end{bmatrix}$$

If $n = 0$, the AVM is empty, and if $n \geq 1$, the ATTRIBUTE_i is a name of an attribute such as case, gender, number ... etc. and the valueset_i is a set of admissible values for the ATTRIBUTE_i. Here, given an attribute a, $\text{rg}(a)$ denotes the set of admissible values for a. So, we require that a $\text{valueset}_i \subseteq \text{rg}(\text{ATTRIBUTE}_i)$ for every $i \leq n$. For instance, the gender attribute in German has the range $\text{rg}(\text{GEND}) = \{\text{m}, \text{f}, \text{n}\}$, and the number attribute has the range $\text{rg}(\text{NUM}) = \{\text{sg}, \text{pl}\}$. Hence, the following is a legal AVM.

$$\left[\ \text{NUM}: \quad \{\text{sg}, \text{pl}\}\ \right]$$

Sets of values encode underspecification. Using logical notation, we may write instead:

$$\left[\ \text{NUM}: \quad \text{sg} \vee \text{pl}\ \right]$$

$\top$ denotes the set of all values. So we have

$$\left[\ \text{GEND}: \quad \top\ \right] \equiv \left[\ \text{GEND}: \quad \{\text{m}, \text{f}, \text{n}\}\ \right]$$

where $\equiv$ denotes logical equivalence. Conjunction and disjunction may be used to combine AVMs. The following equivalence is evident from the definition of AVMs.

$$\begin{bmatrix} \text{NUM}: & \{\text{pl}\} \\ \text{CASE}: & \{\text{nom}\} \end{bmatrix} \equiv \left[\ \text{NUM}: \quad \{\text{pl}\}\ \right] \wedge \left[\ \text{CASE}: \quad \{\text{nom}\}\ \right]$$

When an attribute receives the empty set as a value, this means that we have an empty

disjunction, which is defined to be false ($\bot$):

$$\left[\ \text{CASE}: \ \varnothing\ \right] \equiv \left[\ \text{CASE}: \ \bot\ \right]$$

The usual laws of logic can be applied. Consider, for example, two attributes (say, CASE and NUM) and use the laws of distribution:

$$\left[\begin{array}{ll}\text{NUM}: & sg \vee pl \\ \text{CASE}: & nom\end{array}\right] \equiv \left[\begin{array}{ll}\text{NUM}: & sg \\ \text{CASE}: & nom\end{array}\right] \vee \left[\begin{array}{ll}\text{NUM}: & pl \\ \text{CASE}: & nom\end{array}\right] \equiv$$

$$\left[\ \text{CASE}: \ nom\ \right] \wedge \left(\left[\ \text{NUM}: \ sg\ \right] \vee \left[\ \text{NUM}: \ pl\ \right]\right) \equiv$$

$$\left(\left[\ \text{CASE}: \ nom\ \right] \wedge \left[\ \text{NUM}: \ sg\ \right]\right) \vee \left(\left[\ \text{CASE}: \ nom\ \right] \wedge \left[\ \text{NUM}: \ pl\ \right]\right)$$

Abstractly, an AVM is a partial function f from attributes to sets of admissible values. If f is undefined on a, it may be extended by putting $f^{+}(a) \coloneqq \mathrm{rg}(a)$ (the completion of f). Kracht introduces the definition feature space to codify the description of the AVM as a function from a set A of attributes to a set V of values.

Definition 9. (Feature space) A feature space is a triple $\sigma = \langle A, V, rg\rangle$ such that A is a finite set of attributes, V is a finite set of values and $\mathrm{rg} : A \to \wp(V)$ a function, such that for all $a \in A$, $\mathrm{rg}(a) \neq \varnothing$. σ- matrix is a partial function $f : A \hookrightarrow \wp(V)$, such that such that for all $a \in A$ $f(a) \subseteq rg(a)$.

For example, articles in German inflect for case, number and gender. So, we may put $A \coloneqq \{\text{CASE}, \text{NUM}, \text{GEND}\}$, $V \coloneqq \{\text{nom}, \text{acc}, \text{gen}, \text{dat}, \text{sg}, \text{pl}, \text{m}, \text{f}, \text{n}\}$. German have a set of prepositions that must be followed by the dative case like “von” und “zu”. A dative preposition can take an article as its argument regardless of its number, gender or definiteness and the result is a combination of the preposition and the article in the dative. In such a case, the f(NUM), f(GEND) and f(STATE) are equal to the rg(NUM),

rg(GEN) and rg(STATE) as shown in the following AVM:

$$\begin{bmatrix} \text{POS}: & art \\ \text{NUM}: & \top \\ \text{GEND}: & \top \\ \text{STATE}: & \top \end{bmatrix}$$

So, by convention we may write f(STATE) = {def, indef}. This is simply because a dative preposition accepts any combination of values of the three attributes. That is, in "zu einer", the article is indefinite feminine singular, in "zu dem" it is definite masculine singular and in "zu den" it is definite plural.

3.2.4 Discontinuity, Reduplication and Handlers

Discontinuity is used in grammatical analysis to refer to the splitting of a construction by the insertion of another grammatical unit (Crystal, 2008, p. 147). The concept of discontinuity is extremely important for Arabic morphology. I will discuss it in more detail in chapter four, but I introduce it here as a part of AMRS, and because Kracht has admitted the discontinuity of constituents for the analysis of German. It is important to find out if the two cases are similar or there is any difference. In the optimal case, glued strings concatenate with each other in a definite and non-exceptional linear way. Consider, for example, the plural forms /pictures/ and /Bilder/ in English and German respectively. Both are composed of a singular form and a plural suffix via the linear concatenation, i.e, /picture/ and /s/ in English and /Bild/ and /er/ in German. However, this is not always the case. If we reconsider the German perfect form /aufgemacht/ and compare it with plural forms, we notice that they are actually two different cases of word formation. In /pictures/ and /Bilder/ neither of the two morphs is split, but in /aufgemacht/ the two morphs are split. Thus, the two morphs /auf⊗mach/ and /ge⊗t/ are not described as glued string, but they are examples of the so-called fractured glued strings.

Definition 10. (Fractured glued string) A fractured glued string is a sequence of glued strings. If $\gamma_0, \gamma_1, \cdots, \gamma_{m-1}$ are glued strings, then $g := \gamma_0 \otimes \gamma_1 \otimes \cdots \otimes \gamma_{m-1}$ denotes the

fractured glued string, formed from the γ_i in this order. γ_i is called the ith section of g. m is called the dimension of g, referred to as $dim(g)$. The unique fractured string with dimension 0 is denoted by ζ.

The two morphs in the German perfect form can be written as /auf⊗mach/ and /ge⊗t/. The content of a string is defined as:

Definition 11. (String content) If $\gamma = \langle L, \vec{x}, R \rangle$ is a glued string, then $c(\gamma) = \vec{x}$. Furthermore, $c(\otimes_{i<n}\gamma_i) = c(\gamma_0)^\frown c(\gamma_1)^\frown ...^\frown c(\gamma_{n-1})$.

Context free grammars are not equipped to describe discontinuity. To deal with discontinuous constituents or, more particularly, to combine two fractured glued strings, (Kracht, 2016), following (Seki, Matsumura, Fujii, & Kasami, 1991) suggests using a combinatorial function called handler. A handler can be defined as follows:

Definition 12. (Handler) A handler is a sequence H of sequences of pairs (i, b), where i is a natural number and b a boolean. The members of H are called its sections. A pair (i, b) is said to occur in H, in symbols $(i, b) \in$ H, if there is a section of which (i, b) is some member. The pairs occurring in H are called its parts. Parts may have several occurrences. The result of applying H to two fractured strings g and h such that $g = \gamma_0 \otimes \gamma_1 \otimes \cdots \otimes \gamma_{m-1}$ and $h = \eta_0 \otimes \eta_1 \otimes \cdots \otimes \eta_{n-1}$ is defined as follows:
Put:

$$(i, b)\,(g, h) = \begin{cases} \gamma_i & \text{if } b = true \\ \eta_i & \text{else} \end{cases}$$

Now, for the sequence $\mathrm{h}_i = (i_0, b_0), (i_1, b_1), \cdots, (i_{p-1}, b_{p-1})$, we put

$$\mathrm{h}_i(g, h) := (i_0, b_0)(g, h)^\frown (i_1, b_1)(g, h)^\frown \cdots ^\frown (i_{p-1}, b_{p-1})(g, h)$$

Finally, let $\mathrm{H} = (\mathrm{h}_0, \mathrm{h}_1, \cdots \mathrm{h}_{q-1})$ have q sections, then:

$$\mathrm{H}(g, h) := \mathrm{h}_0(g, h) \otimes \mathrm{h}_1(g, h) \otimes \cdots \otimes \mathrm{h}_{q-1}(g, h)$$

A handler is used if and only if it is proper. A proper handler is defined as:

Definition 13. (Proper Handler) A handler H is proper if for all numbers i, j and booleans b, if H contains (i, b) and $j < i$ then H also contains (j, b). The dimension of a handler H is defined by:

$$dim\ \mathrm{H} = (\{i : (i, true) \in \mathrm{H}\}, \{i : (i, false) \in \mathrm{H}\})$$

If H is proper, dim H is a pair of numbers, such that:

$$0 \text{ is the empty set } \phi \text{ and } n+1 = \{0, 1, ..., n\}.$$

A handler $\mathrm{H}(g,h)$ is defined if and only if H is proper and dim H = $(dim(g), dim(h))$ i.e. if all sections of the two fractured strings are used in H. This combinatorial function allows a sequence of glued strings to be combined in diverse ways: forward concatenation, backward concatenation, forward wrapping, reduction, forward transfixation and, beyond MCFGs (Seki et al., 1991), reduplication (Kracht & Aboamer, 2017) as shown in the following examples:

- Forward Concatenation:
  Put $F := \langle\langle (0, \text{true}), (0, \text{false})\rangle\rangle$.
  Then $F\left(\vec{x}, \vec{y}\right) = \vec{x}\,\vec{y}$

- Backward Concatenation:
  Put $B := \langle\langle (0, \text{false}), (0, \text{true})\rangle\rangle$.
  Then $B\left(\vec{x}, \vec{y}\right) = \vec{y}\,\vec{x}$

- Forward Wrapping:
  Put $W := \langle\langle (0, \text{true}), (0, \text{false}), (1, \text{true})\rangle\rangle$.
  Then $W\left(\vec{x} \otimes \vec{v}, \vec{y}\right) = \vec{x}\,\vec{y}\,\vec{v}$

- Reduction:
  Put $R := \langle\langle (0, \text{true}), (1, \text{true})\rangle\rangle$.
  Then $R\left(\vec{x_0} \otimes \vec{x_1}\right) = \vec{x_0}\vec{x_1}$

- Transfixation:
  Put $T := \langle\langle (0, \text{true}), (0, \text{false}), (1, \text{true}), (1, \text{false})\rangle\rangle$.
  Then $T\left(\vec{x_0} \otimes \vec{x_1}, \vec{y_0} \otimes \vec{y_1}\right) = \vec{x_0}\vec{y_0}\vec{x_1}\vec{y_1}$

Thus, the handler allows us to deal with another important feature, namely reduplication. Reduplication is an important feature for Arabic word formation, there are languages in which processes like reduplication is the primary morphological operation (McCarthy, 1981). The general rule of changing the singular form to the plural one in

Malay is the reduplication of the singular form and the insertion of a hyphen between the two parts. Therefore, the plural form of the word /ayam/ “chicken” is /ayam-ayam/. In multiple context free grammars (Kasami, Seki, & Fujii, 1987), no component is allowed to appear in the value of the function more than once. This is not the case in the above-mentioned example. To handle such cases, the following handler for the plural form in Malay must be used.

$$D := \langle\langle (0, \text{false}), (0, \text{true}), (0, \text{false})\rangle\rangle.$$

Then

$$D(\text{-}, \text{ayam}) = \text{ayam\^{}-\^{}ayam}$$

3.2.5 Morphs and morphemes

The actual units of expressions of language are the morphs. They comprise three components. The first is the exponent, which is typically a glued string. The second is a sequence of selectors, which determine what arguments the morph takes. And the third is a rank function. This function is only needed for empty morphs, to prohibit infinite derivations, and will concern us no further. Morphemes, which are the only meaning bearing units, are sets of morphs that share a common semantics. Thus, the lexical units pair meaning representations with morphemes, not morphs. This accounts for the fact that morphemes can have many different surface forms.

Definition 14. (Selector) A selector is defined formally as triple $\sigma = (M, N, H)$, where M and N are morphological classes and H is a handler. M is called the in-class of σ and N is its out-class.

The role of the selector is to specify what happens when a morph is applied to another morph. The application of one morph (the functor) to another (the argument) is only defined if the in-class of the functor unifies with the out-class of the argument. Given, for example, a functor $\sigma = (M, N, H)$ and an argument $\sigma' = (M', N', H')$, the application of σ to σ' is defined in first instance as follows:

$$\sigma \cdot \sigma' := (M', N, H \circ H')$$

However, underspecification must be handled properly. The way this is standardly done is that the out-class is not actually underspecified, but is a function of its in-class, which is genuinely underspecified. Underspecified values in the out-class are copies of the actual in-class values. Indeed, the proper way to view selectors is as pairs (f, H) where f is a function from fully specified morphological classes to fully specified morphological classes, and H is a handler. This function is given by the pair of AVSs as follows: for each attribute ATT, the associated function f is the following:

- ATT is given a value a in M and a value b in N. Then f is defined on all nonempty values $a' \subseteq a$ and returns b.
- ATT is given no value in M but value b in N. Then f is defined on all nonempty values $a' \subseteq \mathrm{rg}(\mathrm{ATT})$ and returns b.
- ATT is given a value a in M but no value in N. Then f is defined on all nonempty values $a' \subseteq a$ and returns a'.
- ATT is neither given a value in M nor in N. Then f is defined on all $a \subseteq \mathrm{rg}(\mathrm{ATT})$ and returns a.

The product of (M, N) and (M', N') is specified by computing the values of each occurring attribute.

Definition 15. (Morpheme) A morpheme is a set of morphs that share the same semantics but differ in the form.

Finally, given two morphemes M and N, the combination $M \star N$ is the set of all $m(n)$ such that $m \in M$ and $n \in N$.

3.3 The semantic component

I have already introduced the morphological part of the signs. Now, I move to its semantic structure. The semantic structure is composed of two distinct components; one for the semantic representation and the other is an interface that shows how semantic structures are merged.

3.3.1 Argument structure

The merge of morphemes depends on sharing variables and the renaming process as supposed in RS. When a functor F and an argument A are to be combined, both have an argument structure and the merge process is denoted by $F \bullet A$. The argument structure is a sequence of argument identification statements (AISs). Each argument identification statement is defined as follows:

Definition 16. An argument identification statement (AIS) is a quadruple $\langle x : \delta : \Xi :: P \rangle$ such that:

- x is a variable;
- δ is a diacritic;
- Ξ is a pair of AVSs;
- P is a pair of parameter.

There are two types of diacritics; the first is vertical diacritics, whose function is to specify the role of the referent with respect to the merge. A set of vertical diacritics = $\{\triangle, \triangledown, \diamond, -\}$, where:

- the diacritic $\triangle$ means that the referent is an argument,
- the diacritic $\triangledown$ means that the referent is a functor,
- the diacritic $\diamond$ means that the referent is an adjunct, and
- the diacritic - means that the AIS is a referent carrier.

So far, the merge operation proceeds as normal only with one extra piece of information that specifies the role played by the referent with respect to the functor argument relation and the variables are renamed as in the original form of RS. The merge process results in the following possibilities; each case is followed by an illustrative example as shown in figure (3.8).

- $\langle x : \triangledown : A ::\rangle \bullet \langle x : \triangle : A ::\rangle = \langle x : - : A ::\rangle$

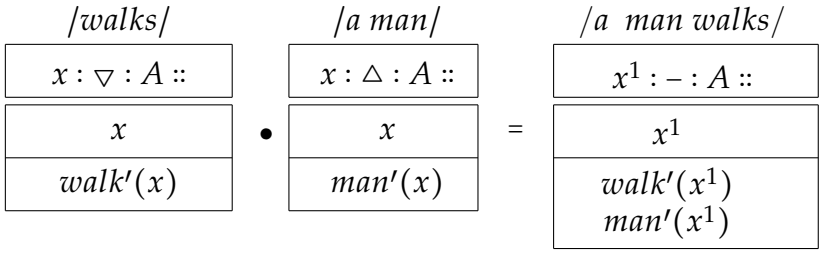


Figure 3.5: Merge operation of /walks/ and /a man/ using vertical diacritics

- $\langle x : \diamond : A \rightarrow B ::\rangle \bullet \langle x : \triangle : A ::\rangle = \langle x : \triangle : B ::\rangle$

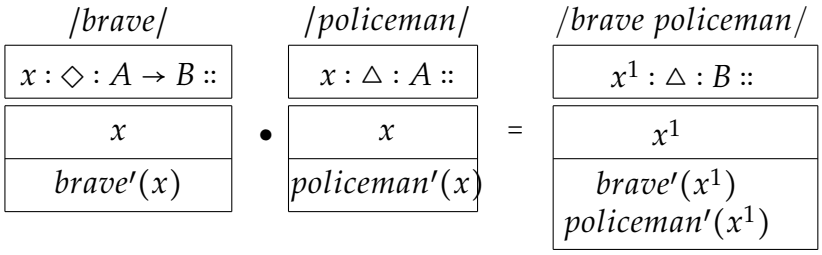


Figure 3.6: Merge operation of /brave/ and /policeman/ using vertical diacritics

- $\langle x : \triangledown : A :\rangle \bullet \langle x : \diamond : C \rightarrow A :\rangle = \langle x : \triangledown : C ::\rangle$

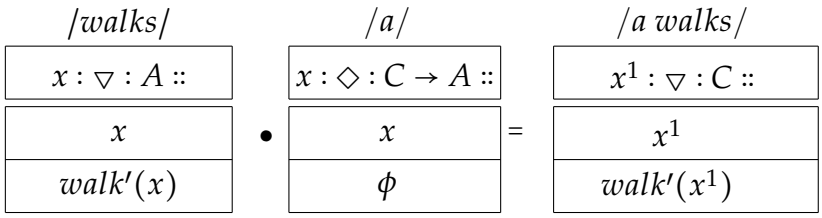


Figure 3.7: Merge operation of /walks/ and /a/ using vertical diacritics

- $\langle x : \diamond : A \rightarrow B ::\rangle \bullet \langle x : \diamond : C \rightarrow A ::\rangle = \langle x : \diamond : C \rightarrow B ::\rangle$

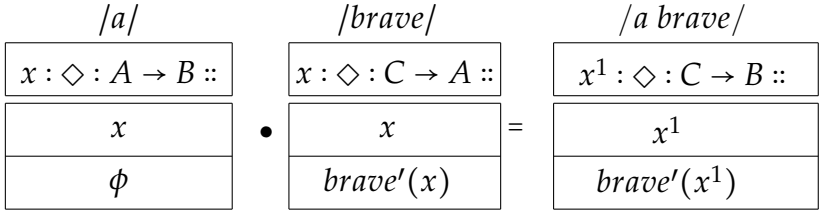


Figure 3.8: Merge operation of /a/ and /brave/ using vertical diacritics

This type of information (vertical diacritics) is described as hierarchical information (U. Klein & Kracht, 2011). Depending only on hierarchical information is not sufficient since it does not help specify the position of the argument whether to the right or the left. It determines only the role of each constituent in the merge operation. Consequently, a different type of diacritic referred to as horizontal diacritics (U. Klein & Kracht, 2011; Kracht, 2016). A set of horizontal diacritics = $\{⊜, ⊚, ◯\}$, where:

- the diacritic ⊜ means that the referent expects an argument on the right side,
- the diacritic ⊚ means that the referent expects an argument on the left side, and
- the diacritic ◯ means that the referent is only exported.

This type of information is referred to as "linear information" (U. Klein & Kracht, 2011) and is represented in the way shown in figure (3.9):

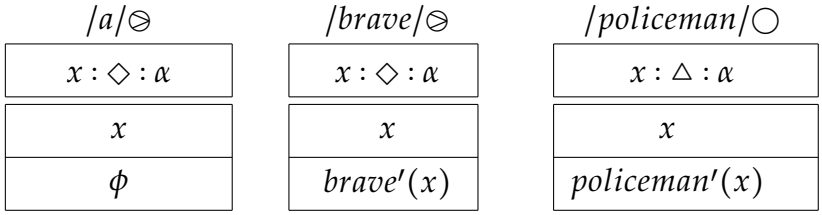


Figure 3.9: Merge operation of /a/ and /brave/ using vertical diacritics

The idea is that both the indefinite article "a" and the adjective "brave", when merged with a noun, play the role of the functor, as identified by the ◇, and both expect the argument to be on the right side, which is specified by the ⊜. Consequently, when "policeman" occupies the right position either directly to the indefinite article resulting in "a policeman" or to the adjective giving "brave policeman" which comes, in turn, to

the right side of “a” to get “a brave policeman”. In such cases, the functor expects only one argument on either side; however, the question that may arise is when the functor expects two argument as in the case of transitive verbs. Kracht has solved this problem by allowing a secure link between the morphosyntax and the argument structure by “the fact that they are sequences of identical length” (Kracht, 2016). Let’s consider the representation of the English transitive verb “chase” in figure (3.10).

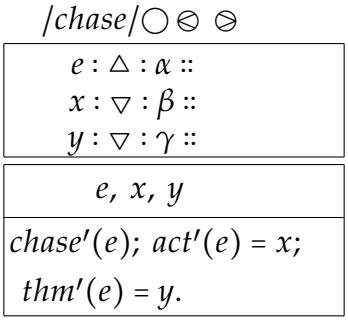


Figure 3.10: Representation of the English transitive verb

The linear information above the argument structure states that the verb /chase/ expects two arguments; the first is on its left side (subject) and the second is on its right side (object). Correspondingly, the referent x (the subject) in the argument structure comes in order before the referent y (the object). This order may vary from one language to another. The significance of properly handling word order is not restricted, in case of English, to the position occupied by the subject and the object with respect to the verb. The case of ditransitive verbs requires a proper handling of the positions of the two objects. That is, in order to avoid ungrammatical compositions of sentences, the verb must specify which of the two object comes first and which follows. Compare, for example, the two sentences in (29) and (30).

(29) I gave him my book.

(30) *I gave my book him.

Ξ is a pair of AVSs. The first specifies the name under which the variable is imported and the second determines the name under which the variable is exported. If the diacritic associated with the variable is △, this means that it has only an export name and

therefore, the second AVS is left empty. If the diacritic is $\triangledown$, this means it has only an import name, and therefore, the first AVS is left empty. If it has the diacritic $\diamond$, this means it has both import and export names. If the variable has both names, the operation is called transformation. Abstractly, when the variable has both import and export names, the functor specifies a function from an import name to export names. If an attribute takes the same value in both of them, the symbol $\surd$ is used in the export AVS. Consider, for example, the agreement between the adjective and the noun in German. In the accusative case, the adjective imports an accusative noun and exports a combination of both of them in accusative case as well. This is shown in the following simple pair of AVSs.

$$[(\text{CASE} : \{\text{acc}\})], \quad [(\text{CASE} : \surd)]$$

The merge of two AVSs is a function composition and the result is a new pair of AVSs. This requires that the export name in the argument and the import name in the head are unifiable.

The fourth component of the AIS is parameter. Parameters are pairs (ρ, ν) such that ρ is the parameter name and ν is a variable. The calculus allows adding as much parameters as one wishes such as properties, tense, aspect, world ... etc. The variable can be used in both Ξ and P. Two parameters like properties and predication time can be introduced in the following way:

$$\begin{bmatrix} \text{PROP} : & p \\ \text{PRED} : & t \end{bmatrix}$$

Similar to names, the AIS has a pair of parameter AVSs. So, a parameter P can be defined as $P = (U, V)$ such that U and V are parameter AVSs, each is a pair (ρ, ν). Given two parameters $P = (U, V)$ and $Q = (U', V')$, P and Q are unified if and only if U and V' have the same parameter name. This is shown in the following example.

$$P = [\text{PROP} : \text{p}] \quad Q = [\text{PROP} : \text{q}]$$

Both P and Q share the same name PROP, so they are unified during the merge, even if they have different variables. By the introduction of the parameters, a simple complete

AIS looks like this

$$\langle x : \triangle : [CASE : nom] :: [PROP : p] \rangle$$

This is a single AIS. The argument structure is a sequence of AISs. Formally it is defined as follows:

Definition 17. (Argument Structure) An "argument identification statement" (AIS) is quadruple $\langle x : \delta : \Xi :: P \rangle$ such that: x is a variable, δ, is a diacritic, Ξ is a pair of AVSs and P is a pair of parameter AVSs. An argument structure is $\alpha = \langle \mu_i : 1 \leq i \leq n \rangle$ of AISs, such that if $n > 1$ then μ_1 is not empty.

The second component of the semantic structure of signs is the DRS. Since I have already explained the concept of DRS in detail in chapter one, I do not repeat it again. Now, a full definition of the sign can be introduced.

Definition 18. (Sign) A sign is a triple $\delta = \langle \mu, \alpha, \Delta \rangle$, such that μ is a morpheme (= set of morphs), α is the argument structure and Δ is a DRS, such that every unbound referent of Δ occurs in α. Each morph of μ should have the same dimension, which must be equal to the length of α.

3.4 Merge of signs

The merge of any two signs depends on the match of all components and the result of the merge is a new sign. This means that a pair of substitutions σ_1 and σ_2 are executed: the first (σ_1) on the head sign and the second (σ_2) on the argument. The merge operation of two signs includes the following three steps:

1. computing the two argument structures,
2. computing the two DRSs, and
3. computing the two morphs.

Variables are unified depending on the match between the morphosyntactic features in the AVSs. When two variables x and y are identified, then $\sigma_1(x) = \sigma_2(y)$, but if they are not identified, then $\sigma_1(x) \neq \sigma_2(y)$. The following example, (from (Kracht & Aboamer,

2017)), shows how this works. The verb /rennen/ 'run' specifies the morphosyntactic information of its subject.

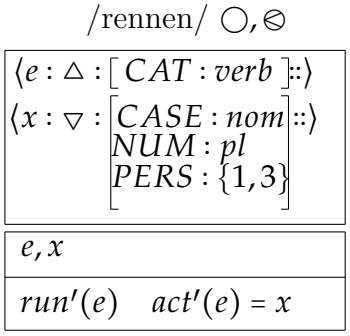


Figure 3.11: Representation of the German verb "rennen"

This can be illustrated as follows: the variable x is imported under the name nominative first person plural or nominative third person plural. If the two pronouns /wir/ 'we' and /uns/ 'us' have the following two representations,

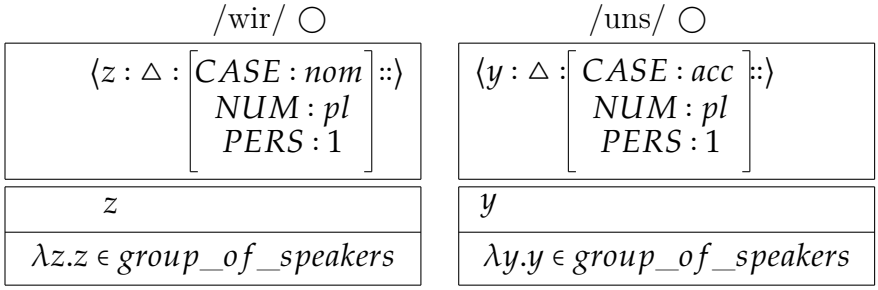


Figure 3.12: Representation of the German pronouns "wir" and "uns"

then, the merge of AISs in both /rennen/ 'run' and /wir/ 'we' succeeds and a new constituent is composed, i.e., /wir rennen/ 'we run'. This simply means that $\sigma_1(x) = \sigma_2(z)$. In contrast, the merge of AISs in both /rennen/ 'run' and /uns/ 'us' does not succeed because the identification is blocked by the mismatch between the values of the case attribute in the two signs. The head imports an argument in the nominative case but the argument offers an accusative case. Consequently, the ill-formed constituent /uns rennen/ 'us run' is not allowed.

In the case of /rennen/ 'run' and /wir/ 'we', there is only one AIS in the argument structure of each sign. The real situation may include, however, more than one AIS.

The verb /schreiben/ 'write' requires, for example, two arguments: one argument for the subject and one argument for the object, and therefore its argument structure contains more than one AIS. If none of the AISs imports a variable, the argument structure is referred to as saturated.

Definition 19. (Saturated argument structure) A saturated argument structure is an argument structure whose AISs do not import any variable.

This definition is important, because there are two types of merge, depending on whether a complement is saturated or not, namely, proper merge and fusion. This distinction is stated by the diacritics. That is, in addition to the previously defined set of diacritics, a special diacritic is used to state if an AIS allows for fusion, namely ▼ and ⧫.

Definition 20. (Proper merge) A proper merge is executed if the argument structure of the complement is saturated.

Definition 21. (Fusion) Fusion is executed if the argument structure of the complement is not saturated.

The distinction between proper merge and fusion is important for cases like the infinitives when used as complements. Consider, for example, the following two control constructions (from (Kracht & Aboamer, 2017)).

(31) Bert promised Mary to leave.

(32) Bert persuaded Mary to leave.

Both (31) and (32) include a control predicate, but (31) is an instance of the subject control and (32) is an instance of the object control. In both cases, the control verb, /promise/ and /persuade/ respectivley, specifies the argument of the subordinate verb. The control verb /promise/ specifies its subject /Bert/ as the actor of the subordinate, i.e., /Bert promised Mary to leave(he)/. On the other hand, the control verb /persuade/ specifies its object /Mary/ as the actor of the subordinate, i.e., /Bert persuade Mary to leave(she)/. If the merge only is allowed the subordinate verb must also be saturated, and therefore, the omitted /he/ and /she/ cannot be accounted for. Fusion allows the

complements to bring arguments to the new structure.

In the normal case, one variable is identified under merge; however, it is allowed to identify several variables. Therefore, the merge or the fusion can be either monadic or polyadic depending on the number of AISs that merge at a time. A merge or a fusion is described as $n-ary$ based on the n AISs that merge at a time. This means that there are four possible cases:

1. monadic proper merge: if $1-ary$ AISs merge and the complement does not import any variable,

2. polyadic proper merge: if $n-ary$ AISs merge such that $n>1$ and none of AISs of the complement imports a variable,

3. monadic fusion: if $1-ary$ AISs merge and the complement imports a variable, and

4. polyadic fusion: if $n-ary$ AISs merge such that $n>1$ and at least one of AISs of the complement imports a variable.

If the argument structure has several AISs, i.e., in the polyadic merge, access conditions apply. Compare, for example, the representation of the English verb /give/ and its German equivalent /geben/.

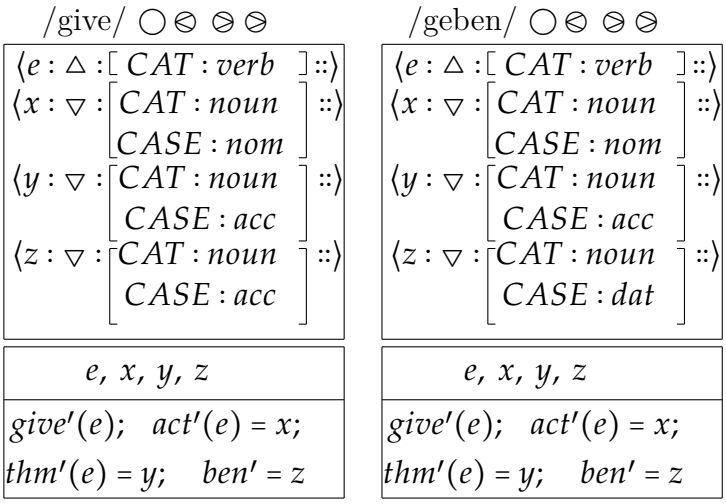


Figure 3.13: Representation of the English verb “give” and the German verb “geben”

The argument structure of each verb imports three arguments: a subject, a direct object and an indirect object. However, a significant difference should be noted here. The semantic representation of the English verb /give/ accounts for one possible word order, i.e., its arguments have the following fixed word order:

$$x : \triangledown : sub \quad e : \triangle : give \quad y : \triangledown : indobj \quad z : \triangledown : dobj$$

This word order gives well-formed sentences like (33).

(33) He gave me the book.

Changing the positions of the arguments results in odd sentences. German, like Arabic, has flexible word order. A sentence like (33) can be expressed by any of the following sentence:

(34) Er gab mir das Buch.
he.nom.sg.m give.3.sg.m.pst me.dat.1.sg the.acc.n.sg book.acc.n.sg
'He gave me the book.'

(35) Mir gab er das Buch.
me.dat.1.sg give.3.sg.m.pst he.nom.sg.m the.acc.n.sg book.acc.n.sg
'He gave me the book.'

(36) Das Buch gab er mir.
the.acc.n.sg book.acc.n.sg give.3.sg.m.pst He.nom.sg.m me.dat.1.sg
'He gave me the book.'

This means that the merge in English must follow a specific order, i.e., the AISs are accessed one by one in a restricted order. German, on the other hand, has a different kind of access, which allows to skip an AIS, if its features do not match. The first type is referred to as E-access and the second is called G-access. Both are included in the following definition of access (Kracht & Aboamer, 2017; Kracht, 2016).

Definition 22. (Access) Let $\alpha = \langle \mu_i : 1 \geq i \geq m \rangle$ be an argument structure and v is an AIS.

- v E-accesses μ_k iff $k = m$ and $\mu_m \bullet v$ succeeds.
- v G-accesses μ_k iff k is the largest index and $\mu_k \bullet v$ succeeds.

3.5 Cancellation under agreement

Two main approaches can be identified with respect to constituents composition, namely cancellation and agreement. In the first version of categorial grammar as introduced by, Ajdukiewicz (1935), expressions were classified as basic types and sentences. A type B is composed via a cancellation schema from a type $\frac{B}{A}$ and a type A as follows (Morrill, 2011, p. 4):

(37) $\frac{B}{A}\ A \longrightarrow B$

In the subsequent works on categorial grammar, the definition of types has been modified from this non-directional fractions to a form that involves slashes that express the order, but the cancellation approaches has remained. That is to say, instead of (37), the slash notation has been introducedas shown in (38) (Steedman, 1993).

(38) $A\backslash B \longrightarrow B$

(38) means that the expression of type $A\backslash B$ takes an argument from type A on its left side to produce an expression from type B (Morrill, 2011, p. 5-6).
In HPSG, constituents are composed via agreement, which is included as a category, i.e., agr-cat in the feature structure. The contrast between (39) and (40) is accounted for depending on the information provided by the agr-cat in as shown in (3.14) and (3.15) (Sag, Wasow, & Bender, 1999, p. 109-110).

(39) Kim walks.

(40) *We walks.

Kim

$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{noun} \\ \text{AGR} & \begin{bmatrix} \text{PER 3rd} \\ \text{NUM sg} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SPR} & \langle\ \rangle \end{bmatrix} \end{bmatrix}$$

walks

$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} \\ \text{AGR}\ \boxed{1} & \begin{bmatrix} \text{PER 3rd} \\ \text{NUM sg} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SPR} & \left\langle \begin{bmatrix} \text{NP} \\ \text{AGR}\ \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

Figure 3.14: Lexical entries for “Kim” and “walks” in HPSG

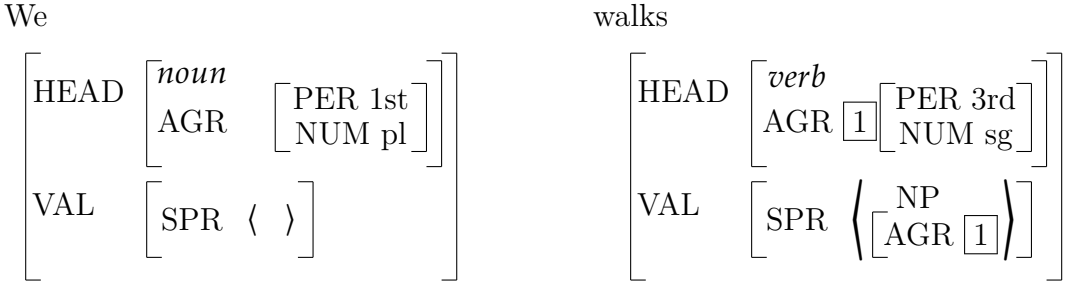


Figure 3.15: Lexical entries for “we” and “walks” in HPSG

AMRS connects both approaches, i.e., cancellation is performed depending on agreement. Constituents are composed through the merge operation between a head and an argument. The argument structure of the head specifies the argument selection requirements, and if the argument structure of the argument meets these selection requirements, they compose a complex constituent and the selection requirements are canceled. The example in (41) shows how the agreement helps not only compose complex constituents but also resolve the disambiguation.

(41) Die Fragen haben wir verstanden.
the.acc.pl questions.acc.pl have.aux we.nom.1.pl understand.prf
‘We understood the questions’

The argument structure of the verb /verstehen/ ‘understand’ imports two arguments for the subject and the accusative object. The AISs contain the selection requirements of each argument. The definite noun phrase /die Fragen/ ‘the questions’ have the same form in the accusative and nominative cases. Although it occurs in the first position and the pronoun /wir/ ‘we’ occurs after the auxiliary verb /haben/ ‘have’, the latter is in the nominative case, which means it cannot function as an object, and therefore the only possible choice to form a sentence is to take /die Fragen/ ‘the question’ as an object, i.e., in the accusative case, taking into consideration that German allows for the G-access. So, assume that the head has the following representation.

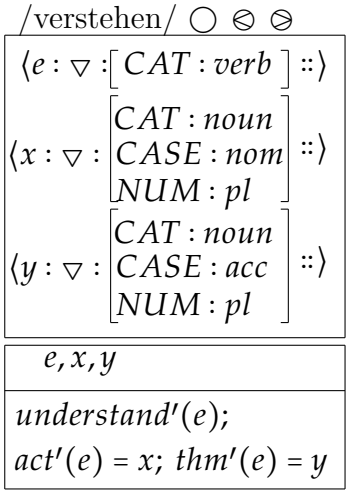


Figure 3.16: Representation of the head “verstehen”

The two arguments have three representations: one for the pronoun /wir/ ‘we’ and two for the noun phrase /die Fragen/ ‘the questions’.

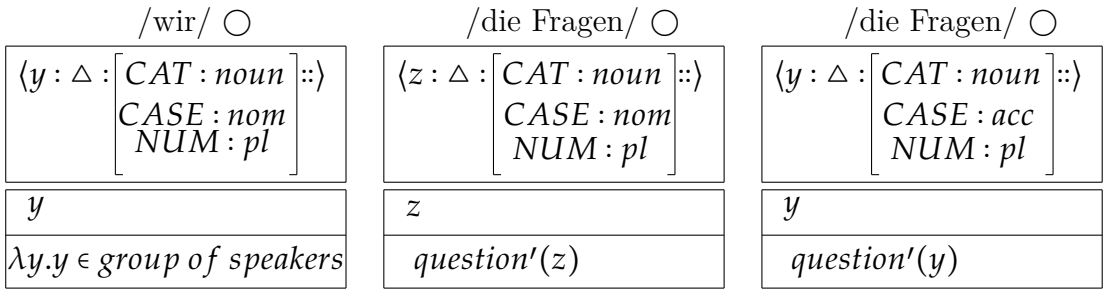


Figure 3.17: Representation of the arguments “wir” and “die Fragen”

The head merges with /wir/ ‘we’ and with /die Fragen/ ‘the questions’ in the accusative case, and in each merge process the selection restrictions disappears (is canceled) in the argument structure of the complex constituent.

3.6 Summary

In this chapter, I introduced AMRS in some detail. I used the notions and terminologies as used by Kracht. Since AMRS is an extended version of the theory of RS, I preferred to first explain the idea of RS and its drawbacks and then describe how AMRS represents an extension of it. I showed how RS has presented a smart solution for dealing

with variables that eliminates the problem of choosing the names of the variables. Nevertheless, RS has faced real challenges because it has depended on word order in the renaming process, which does not work with the languages that have flexible word order. Kracht has extended RS and introduced his own theory, AMRS, in which he has disregarded the directionality and allowed the variables to be identified depending on the agreement of the morphosyntactic features. In AMRS, the lexicon is a collection of signs and each sigh has two components: the morphological component and the semantic structure. In the morphology, I distinguished between the glued string and the fractured glued string. The latter is used for the discontinuous morphemes whose merge process is accomplished using a combinatorial function called handlers. In the semantics, I introduced the four components of the AISs: variables, diacritics, the pair of AVSs and parameters. A sequence of AISs composes the argument structure. The meaning of the sign is represented as a DRS. I also showed the merge mechanism of the signs and how the ARMS combines both agreement and cancellation when two signs merge to compose a complex constituent.

Chapter 4

A Purely Surface Oriented Approach to Handling Arabic Morphology

This chapter is devoted to the description of Arabic morphology within AMRS. The nature of Arabic morphology represents an interesting case for the pure surface-oriented treatment. Arabic is very well-known for its non-concatenative morphology, and therefore, has given rise to various paradigms, particularly the autosegmental analysis (McCarthy, 1981). In this chapter, I show that the concatenative paradigm can be saved by admitting the discontinuity of morphemes. Since one of the aims of this dissertation is to end up with a morphosemantic lexicon for (a fragment of) Arabic, I describe the components of the lexicon in this chapter. To remain surface oriented, constituents are allowed to be composed only via concatenation and reduplication, but no rule can delete, add, or modify any string. In addition to allowing the discontinuity of morphemes, two other problems are dealt with, namely the matching among morphemes and the morphophonemic alternations resulting from the interaction among different types of morphs. I discuss the morphological system of Arabic and the various paradigms proposed for the study of Arabic word formation and show how the proposed approach differs from these paradigms. Then, I show how to handle the morphology of Arabic within AMRS.

4.1 Arabic morphology: Theoretical and computational approaches

There is no uniform morphological system for word formation in all languages of the world, but languages differ significantly in the way in which they compose their words. Isolating languages, like Classical Chinese, can be described as languages without morphology because sentences are composed of monosyllabic units that cannot usually be divided to smaller ones (Watson, 2002, p. 124). Other languages, like Turkish, Hungarian and to some extent English[1], have an agglutinative morphology, in which words are (mostly) composed by a linear non-exceptional concatenation of morphemes as in (use-less-ness), which is composed of three distinct and isolatable morphemes: "use", "less" and "ness".

Arabic morphology is referred to as templatic root and pattern morphology. It differs from that of English or other Indo-European languages because it is, to a large extent, based on discontinuous morphemes. Words in Arabic are derived from roots, which bear the core meaning of their derivatives, by inserting vowels and maybe other consonants. Roots are relatively invariable, discontinuous bound morphemes, typically containing three consonants in a certain order, and interlocked with vowels and other consonants to form stems (Ryding, 2005, p. 45). Let us consider the following example: the triliteral (3 consonantal) root /d⊗r⊗s/ /د ⊗ ر ⊗ س/ is supposed to bear the meaning of studying and from which words like /darasa/ /دَرَسَ/ 'he studied', /durisa/ /دُرِسَ/ 'be studied' and /daːris/ /دَارِس/ 'student/studying' are derived. This process has extensively and very productively evolved in order to cover a vast array of meanings associated with each semantic field (Ryding, 2005, p. 46). The vast majority of Arabic words are derived in this way and therefore can be analyzed as consisting of two different bound morphemes: a root morph and a vocalism morph that interlock to form stems, and neither of them can occur alone in the sentence. Some examples of how roots and vocalisms can be

[1]English has both agglutination and flexion, and therefore its classification as an agglutinative or a flexional is a matter of degree (Payne, 2017)

merged to get stems are given in table (4.1).

Root M	vocalism M	Merging Way	Stem	Form
/k⊗t⊗b/	/aː⊗i/	CVVCVC	/kaːtib/	Active Participle
/k⊗t⊗b/	/ma⊗uː/	maCCVVC	/maktuːb/	Passive Participle
/k⊗t⊗b/	/ma⊗a/	maCCVC	/maktab/	Common-locative

Table 4.1: Interlocking of roots and vocalisms to form stems

McCarthy (1981) has proposed a prosodic theory for the analysis of the non-concatenative morphology, in which the devices of autosegmental phonology were applied. So, similar to autosegmental phonology, each morphological form is linked to some phonological material by lines of association (Watson, 126). In his model, he differentiated three tiers: the root, a sequence of consonants, the template, a sequence of CVs (C for consonants and V for vowels) and the vocalism, a sequence of vowels. This is shown in the figure (4.1).

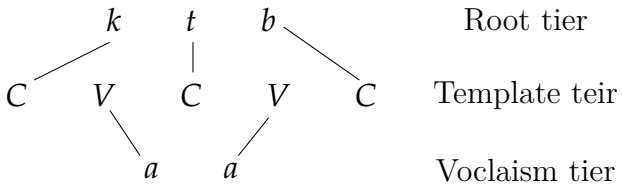


Figure 4.1: Three tiers for Arabic words

The pronounceable string is then obtained by folding the root tier and the vocalism tier in one single tier as shown in figure (4.2) (Watson, 127).

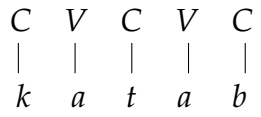


Figure 4.2: Folding the root tier and the vocalism tier in one single tier

Eventually, McCarthy and Prince (1996) developed a theory of prosodic morphology in which they claimed that templates are defined in terms of the authentic units of

prosody: mora (μ), syllable (δ), foot (F), prosodic word (PrWd). The stem /katab/ can be represented as shown in the following figure.

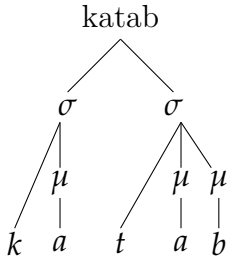


Figure 4.3: Stem /katab/ in the prosodic morphology

After composing stems, they concatenate with suffixes and prefixes to form words as illustrated, for example, in the Arabic masculine sound plural form /muʕalimuːna/ /مُعَلِمُونَ/ 'teachers' obtained from the concatenation of the strings /muʕalim/ /مُعَلِمِ/ 'teacher' and /uːna/ /ونَ/, a nominal suffix for the masculine plural (and a verbal suffix for masculine plural in the indicative mood). In the previous example, the stem concatenates with a suffix. In /jaktub/ /يَكتُب/ 'he writes', the verb stem /ktub/ /كتُب/ concatenates with the prefix /ja/ /يَ/,[2], the present form for the masculine third person singular. The stem can be surrounded by both a prefix and a suffix as in /jaktubuːna/ /يَكتُبُونَ/ 'they write', composed of the stem /ktub/ /كتُب/, the prefix /ja/ /يَ/ and the suffix /uːna/ /ونَ/.

As a result of such degree of richness, the complexity of word formation, and the exceptional degree of ambiguity in the writing system, which present special challenges to text-processing applications, Arabic morphology has been the focus of research in natural language processing for a long time (Al-Sughaiyer & Al-Kharashi, 2004; Soudi, Neumann, & Van den Bosch, 2007; Dichy & Farghaly, 2007; Sawalha & Atwell, 2008; Habash, 2010). Researchers have adopted different approaches in the treatment of Ara-

[2]In this part, I follow the standard assumption that the both the consonant /j/ and the short vowel /a/ compose together the verb prefix, but in chapter six, I propose a different point of view.

bic morphology, both theoretically and computationally (Dichy & Farghaly, 2007). Most of the efforts have been particularly devoted to addressing morphological analysis, generation and disambiguation. Arabic morphotactics have been sufficiently described and handled using finite state operations (Kay, 1987; Beesley, 1998, 2001; Attia, Pecina, Toral, & Van Genabith, 2013; Aboamer & Farghaly, 2015). Reinhard and Gibbon (Reinhard & Gibbon, 1991) suggested prosodic inheritance morphology to deal with both concatenative and non-concatenative morphology. Al-Sughaiyer and Al-Kharashi (Al-Sughaiyer & Al-Kharashi, 2004) provide a comprehensive survey for Arabic morphological analysis techniques. Furthermore, Dichy and Farghaly (2007) list the following main six approaches:

- The root and pattern approach
- The lexeme-based approach
- The machine learning and statistical approach
- The stem-based approach
- Stems, based on root and pattern
- Stem-based lexical resources, including root-pattern and grammar-lexis information.

Depending on one of these approaches, many Arabic morphological analysers and generators have been developed in the last two decades, particularly by the works of Tim Buckwalter in his Arabic Morphological Analyser (BAMA) (Buckwalter, 2002, 2004), Habash, Rambow and Roth in their system for Morphological Analysis and Disambiguation for Arabic (MADA) (Habash, Rambow, & Roth, 2009), the Linguistics Data Consortium in its Standard Arabic Morphological Analyser (SAMA) (Maamouri, Graff, Bouziri, Krouna, & Kulick, 2010), Sawalha's Fine-Grained Morphological Analyser and Part of Speech Tagger for Arabic Text (Sawalha, Atwell, & Abushariah, 2013) and Attia's AraComLex (Attia et al., 2013) among others. Some of these systems are open-source and available for research and evaluation, while others are available in encrypted

formats or used in proprietary commercial applications (Attia et al., 2013). Recent works have attempted either to improve the accuracy of the analyser (Abdelali, Darwish, Durrani, & Mubarak, 2016), to add some other features (Taji, Khalifa, Obeid, Eryani, & Habash, 2018) or to focus on a specific dialect (Habash, Eskander, & Hawwari, 2012). Generally, there is a gap between morphology and semantics in the current approaches to Arabic morphology and the focus has been placed on the issues of word formation and decomposition. We cannot go higher than morphology. It is true that some morphological analyzers have been enriched with some semantic information, but the formal meanings of the units were completely absent. The treatment of morphology within AMRS has both a context free grammar on the one hand and a mechanism to deal with discontinuity on the other. This allows the computer to understand /k⊗t⊗b/, and /a⊗a/ as units, not states, because they have meanings. Recall that German "trennbare Präfixe" (separable prefixes), like /auf/ in /aufmachen/, appear in some cases as two parts. This can occur, for instance, in the imperative, as in /Mach bitte das Fenster auf/ "Open the window, please". However, /mach/ and /auf/ together form one unit (as they can only be interpreted together) but this unit is discontinuous. This also applies to Arabic roots and vocalisms. Furthermore, there is clear psycholinguistic evidence that Arabic consonantal roots are distinct components of the Arabic mental lexicon (Frisch & Zawaydeh, 2001). Shortly, the proposed approach is different from previous and current approaches in (i) it is compositional; morphological units have meanings and the meaning of complex expressions is determined by the meaning of these smaller units and their order; (ii) morphology is not distinguished from syntax; rather, it is a lexicalist approach on a par with categorial grammar; (iii) it is more restrictive; since it is purely surface-oriented, it allows only grammatical constituents. This eliminates the overgeneration.

4.2 A concatenative paradigm for Arabic morphology

AMRS differentiates, based on the combination behavior, between two types of strings: glued strings and fractured glued strings. Roots and vocalisms are fractured glued string, whereas prefixes, suffixes and clitics are glued strings. A very simple example of glued

strings in Arabic is the feminine morpheme /t/, which is only followed by space (forget now about case markers) and consequently no other strings can follow it directly. Moreover, /t/ is not allowed to be preceded by a space. So, the Arabic glued string /t/ and its right and left requirements can be coded as a glued string j such that:

$$j = \langle \{(-, \sqcup)\}, t, \{(+, \sqcup)\} \rangle$$

The concatenation of glued strings can be described in the light of the definition of occurrence[3]. For example, if $\overrightarrow{y}$ = /muʕalimuːna/ 'teachers', $\overrightarrow{u}$ = /mu/ and $\overrightarrow{v}$ = /uːna/, then, $\overrightarrow{x}$ = /ʕalim/. The possible cases of occurrence can be illustrated as follows:

If $o_1 = \langle \overrightarrow{u_1}, \overrightarrow{v_1} \rangle$ is an occurrence of $\overrightarrow{x_1}$ in $\overrightarrow{y}$ and $o_2 = \langle \overrightarrow{u_2}, \overrightarrow{v_2} \rangle$ is an occurrence of $\overrightarrow{x_2}$ in $\overrightarrow{y}$, then it is said that o_1 is to the left of o_2 and (o_2 is to the right of o_1) if $\overrightarrow{u_1}\overrightarrow{x_1}$ is a prefix of $\overrightarrow{u_2}$ as shown in figure (4.4).

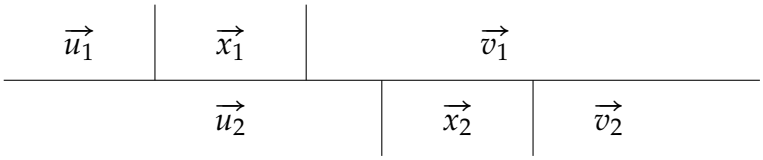


Figure 4.4: Two contiguous occurrences o_1 and o_2

If $\overrightarrow{u_2} = \overrightarrow{u_1}\overrightarrow{x_1}$ as in figure (4.5), it is said that o_1 is immediately to the left of (left adjacent to) o_2 and o_2 is immediately to the right of (right adjacent to) o_1.

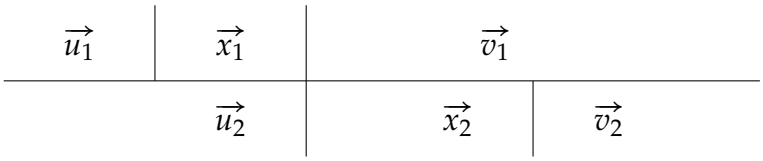


Figure 4.5: Two adjacent occurrences o_1 and o_2

The two occurrences o_1 and o_1 are said to be contiguous if one of them is left or right adjacent to the other. Otherwise, they overlap. Now, reconsider the word /muʕalimuːna/[4].

[3]Recall the definition of "occurrence" in section (3.2.2)

[4]Figures (4.6) and (4.7) involve a random word segmentation, not a morpheme-based segmentation. The aim is to explain the notion of occurrence not to segment the word into its morphemes.

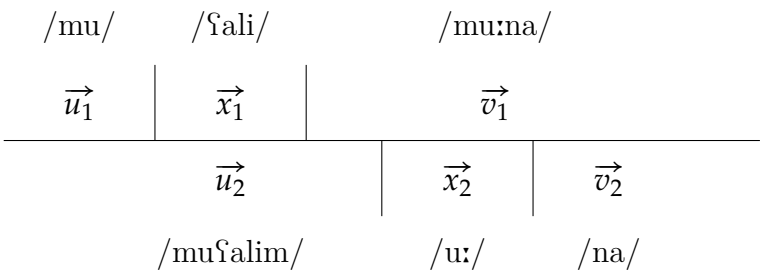


Figure 4.6: Example of two contiguous occurrences o_1 and o_2 in Arabic

The same example is shown in figure (4.6), but o_1 is immediately to the left of (left adjacent to) o_2, i.e., $\overrightarrow{u_2} = \overrightarrow{u_1}\,\overrightarrow{x_1}$.

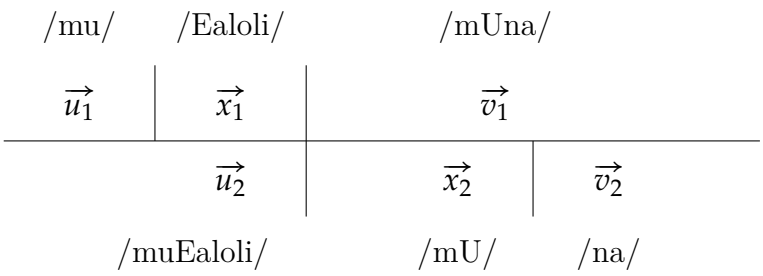


Figure 4.7: Two adjacent occurrences o_1 and o_2

As illustrated in chapter three, the right and left conditions do not suffice alone to handle all possible conditions of combination, and therefore each string is provided with a pair AVSs that compose its morphological class. The feminine morpheme in Arabic /t/ is used in most cases to change the gender from masculine to feminine as /kaːtib/ /كَاتِب/ 'writer' and /kaːtibah/ /كَاتِبَة/ 'writer (female)'; /muʕalim/ /مُعَلِم/ 'teacher' and /muʕalimah/ /مُعَلِمَة/ 'teacher (female)'; /ʤamiːl/ /جَمِيل/ 'beautiful' and /ʤamiːlah/ /جَمِيلَة/ 'beautiful (female)' and /tˤawiːl/ /طَوِيل/ 'tall' and /tˤawiːlah/ /طَوِيلَة/ 'tall (female)'. However, in some cases this suffix is ignored because the noun or the adjective refers by nature to a female as in /tˤaːliq/ /طَالِق/ 'divorced' and /ħaːʔidˤ/ /حَائِض/ 'menstruating'. Also, there are masculine words that end with /ة/

/t/ like /xaliːfah/ /خَلِيفَة/ ‘successor’. Consequently, the combination between words like /tˤaːliq/ or /ħaːidˤ/ and the feminine morpheme should be prevented. These properties are included in the morphological class of each morph. Morphological classes are very significant for the combination process. When a root morph combines with a vocalism morph, the vocalism morph takes the role of the functor (say m_1), and the root morph, m_2, takes the role of the argument.

$$m_1(m_2) = m_3$$
$$\text{vocalism(root)} = \text{stem}$$
$$\text{/a⊗ a/(/k⊗ t⊗ b/)} = \text{/katab/}$$

The example shows that to properly handle the combination between roots and vocalisms, the properties of m_1 and m_2 before the combination should first be handled, and the result is the properties of m_1 and m_2 after combination, i.e., m_3. The combination in the previous example results in a past stem in the active form. Compare that with the following combination:

$$m_1(m_2) = m_3$$
$$\text{vocalism(root)} = \text{stem}$$
$$\text{/u⊗ i/(/k⊗ t⊗ b/)} = \text{/kutib/}$$

The result of the combination, in this case, is the past stem in the passive form.

The root /k⊗t⊗b/ and the vocalism /a⊗a/ are instances of the so-called fractured glued strings. As is the case in the German perfect form, both the roots and vocalisms are allowed to be discontinuous and the combination process is led by the handler. The derivation of the stems /katab/ “he wrote”, /kutib/ “was written” and /kaːtib/ are handled using a similar function because in the three cases the parts of the root have the same position and the parts of the vocalisms as well. In the three cases, the first letter of the root /k/ comes at the beginning followed by whatever the first letter in the vocalism, /a/, /u/ or /aː/, then the second letter of the root /t/, followed by the second letter of the vocalism, /a/ or /i/ and finally the stem ends with the last letter of the root. Thus, the three combinations can be handled using the following handler:

$$\mathrm{H} := \langle\langle(0, false), (0, true), (1, false), (1, true), (2, false)\rangle\rangle$$

In the case of /katab/, if we apply a function that maps from each part of the handler to its corresponding string in the two fractured strings, we get the following (taking into consideration that the vocalism takes the role of the functor):

- $(0, false) \rightarrow$ /k/
- $(0, true) \rightarrow$ /a/
- $(1, false) \rightarrow$ /t/
- $(1, true) \rightarrow$ /a/
- $(2, false) \rightarrow$ /b/

The result of applying this handler to the two fractured glued strings /a⊗a/ and /k⊗t⊗b/ is /katab/ as shown in the following.

H (a⊗a,k⊗t⊗b) := kˆaˆtˆaˆb = katab

Similarly, in the case of /kutib/, we get:

- $(0, false) \rightarrow$ /k/
- $(0, true) \rightarrow$ /u/
- $(1, false) \rightarrow$ /t/
- $(1, true) \rightarrow$ /i/
- $(2, false) \rightarrow$ /b/

The result of applying the same handler to the two fractured glued strings /u⊗i/ and /k⊗t⊗b/ is /kutib/ as shown:

H (u⊗i,k⊗t⊗b) := kˆuˆtˆiˆb = kutib

Also, in the case of /kaːtib/, we get:

- $(0, false) \rightarrow$ /k/
- $(0, true) \rightarrow$ /aː/
- $(1, false) \rightarrow$ /t/
- $(1, true) \rightarrow$ /i/
- $(2, false) \rightarrow$ /b/

The result of applying the same handler to the two fractured glued strings /aː⊗i/ and /k⊗t⊗b/ is /kaːtib/ as follows:

H (aː⊗i,k⊗t⊗b) := kˆaːˆtˆiˆb = kaːtib

Reduplication is also an important feature in Arabic word formation that can be accounted for using handlers. In some cases, Arabic tends to duplicate a specific letter (string) to get a new word as in /kattaba/ 'made someone write' from the root /k⊗t⊗b/ and the vocalism /a⊗a/ where the second letter in the root, /k/, occurs twice. This also applies to /darrasa/ 'teach' from the root /d⊗r⊗s/ and the vocalism /a⊗a/ where the second letter of the root, /r/, is reduplicated. In these examples, reduplication results not only in a different form of the verb, but also a different meaning. Both /kataba/ and /kattaba/ have the same root /k⊗t⊗b/ and the same inserted vowels /a⊗a/. This also applies to /darasa/ and /darrasa/, derived from the same root /d⊗r⊗s/ and the same inserted vowels /a⊗a/. The only difference between the two forms lies in the reduplication of the second consonant, which changes the meanings from 'he wrote' to 'made someone write' and from 'learn' to 'teach' respectively.

In some other cases, reduplication occurs in the root itself even before it combines with any vocalism, when its second and third consonants are identical as in /marra/ 'he passed', /farra/ 'he escaped', /hadda/ 'he pulled down' and /madda/ 'he extended'. In these cases, reduplication is represented orthographically with "shaddah" or "tashdid" above the reduplicated string as in /رَّ/ and /دَّ/. Verbs which involve this kind of reduplication are referred to as "doubled" or "geminate" verbs. Nevertheless, the reduplicated letter may orthographically appear, in some cases, twice and without "shaddah". Compare, for example, (42) and (43).

(42) مَرَرتُ عَلَى أَحمَدَ

marartu ʕala: ʔaħmada
pass.1.sg.pst over Ahmed.gen.sg.m
'I passed over Ahmed'

(43) مَرَّ عَلَى أَحمَدَ

marra ʕala: ʔaħmada
pass.3.sg.pst over Ahmed.gen.sg.m
'He passed over Ahmed'

In (42) and (43), the verb is in the past form, but the subject in (42) is a first person and in (43) a third person, which results in a change in the form of the geminate verb. Quadriliteral roots may also involve a kind of reduplication. This happens when a four-consonant root is composed by the reduplication of the first two consonants as in /z⊗l⊗z⊗l/ /ز ⊗ ل ⊗ ز ⊗ ل/ and /z⊗ʕ⊗z⊗ʕ/ /ز ⊗ ع ⊗ ز ⊗ ع/.
As mentioned in section (3.2.4), AMRS goes beyond MCFG and allows components to appear in the value of the function more than once. In this way, the plural in Malay could be handled. Arabic is not an exception; handlers are allowed to combine the multi occurrence of any string. Although reduplicated letters appear orthographically as one letter above which a "shaddah", they are actually two letters: a consonant followed by itself and a vowel. So, in the computational lexicon, I separate the two occurrences using the "sukoon" symbol. In other words, instead of /kattab/, I will generate /katotab/. This also applies to /darras/ and /daroras/, /marra/ and /marora/ and /hadda/ and /hadoda/, but not to /zalzal/ or /zaʕzaʕ/, since it a different form of reduplication. Thus, if I put,

$$D := \langle\langle (0,\text{false}), (0,\text{true}), (1,\text{false}), (1,\text{true}), (1,\text{false}), (2,\text{true})\,(2,\text{false})\rangle\rangle$$

then

$$D\ (a \otimes o \otimes a, k \otimes t \otimes b) = k\hat{}a\hat{}t\hat{}o\hat{}t\hat{}a\hat{}b = katotab$$

$$D\ (a \otimes o \otimes a, d \otimes r \otimes s) = d\hat{}a\hat{}r\hat{}o\hat{}r\hat{}a\hat{}s = daroras$$

Arabic roots have different types and the combinatorial functions deal only with the order of the parts of the merged strings, but which root merges with which vocalism is a different question. In the real implementation, I utilize the morphological class technique to sub-categorize the roots into several types and each category is allowed to merge with specific vocalisms. The matching between root and vocalism is discussed in section (4.4).

4.3 Morphs and morphemes in Arabic

In section (3.2.5), I differentiated between the notion "morph" and "morpheme" and showed that strings, which share a common semantics but differ in form, are considered morphs of the same morpheme. The feminine morpheme /t/ /ة/ has two forms: one if it is only followed by a space and another if it is followed by suffixes. The two cases are represented in (44).

(44) اِلتَقَت أَمِيرَةُ وِيلزَ بِأَمِيرَتِنَا

iltaqat ʔamiːratu wiylza biʔamiːratinaː
meet.3.sg.pst.f princess.nom.sg.f Wales.gen.sg with-our-princess.gen.sg.f
'The princess of Wales met our princess'

In (44), the feminine morpheme has two different forms in /ʔamiːratu/ /أَمِيرَةُ/ 'princess' and /biʔamiːratinaː/ /بِأَمِيرَتِنَا/ 'with our princess', respectively. The masculine form of both is /ʔamiːr/ /أَمِير/ 'prince'. Since AMRS is a purely surface-oriented approach, no rule is used to change /ة/ to /ت/ when it is followed by a string other than space[5]. This means that both share the same semantics and only differ in form, and each of them has its own behavior when it combines with other strings. Thus, both /ة/ and /ت/ are considered two morphs of the same morpheme.

[5]Punctuation marks are also allowed to follow the /ة/. This is in contrast with /ت/, which, when used as a feminine morph with nouns and adjective, are followed neither by a space nor by a punctuation mark

Also, the active participle in Arabic has two different forms depending on the number of the root consonants. If the root consists of only the basic three or four consonants, it is in the base form, which is also referred to in traditional Arabic grammar as muʤarrad (مُجَرَد, literally the 'stripped'); otherwise it is morphologically complex, referred to as maziːd (مَزيد, literally 'increased')(Ryding, 2005, p. 434). Active participles of the base form of the triliteral roots are derived by inserting /aː/ after the first root consonant and /i/ after the second root consonant. Therefore, the active participle of /k⊗t⊗b/ /ك ⊗ ت ⊗ ب/ is /kaːtib/ /كَاتِب/ 'writer', /dˁ⊗r⊗b/ /ض ⊗ ر ⊗ ب/ is /dˁaːrib/ /ضَارِب/ 'hitter' and /l⊗ʔ⊗b/ /ل ⊗ ع ⊗ ب/ is /laːʔib/ /لَاعِب/ 'player'. On the other hand, if the verb past stem has more than three consonants- either by adding from one to three letters to the basic three-root consonants, reduplicating a specific letter or even if the root has four basic letters, the active participle is derived by replacing the prefix of the present stem with /m/ followed by /u/ and inserting /i/ after the penultimate root consonant. This means that two steps should be done: the first is to get the present stem form and the second is to merge the active participle morph with the present stem morph. Examples of various root types are given to show the derivation process of the active participle form.

- The past stem form /ʔablaɣ/ / أَبــلَــغ / 'informed' has the root /b⊗l⊗ɣ/ /ب ⊗ ل ⊗ غ/ plus the letter /ʔ/ / أ/. Its present stem form is /jubliɣ/ /يُبلِغ/ 'inform' and therefore applying the aforementioned rule of active participle, the result is /mubliɣ/ /مُبلِغ/ 'informer'.

- The past stem form /ʔintaqam/ / اِنــتَــقَم / 'revenged' has the root /n⊗q⊗m/ /ن ⊗ ق ⊗ م/ plus the letter /ʔ/ / اِ/ and the letter /t/ /ت/. Its present stem form is /jantaqim/ / يَنتَقِم/ 'revenge' and from which the active

participle /muntaqim//مُنتَقِم/ ‘avenger’ is obtained.

- The past stem form /ʔistaxraʤ/ /اِستَخرَج/ ‘extracted’ has the root /x⊗r⊗ʤ/ /خ ⊗ ر ⊗ ج/ plus the letter /ʔ/ /ا/, the letter /s/ /س/ and the letter /t/ /ت/. Its present stem form is /jastaxriʤ/ /يَستَخرِج/ ‘extract’ and from which the active participle /mustaxriʤ/ /مُستَخرِج/ ‘extractor’ is obtained.

- The past stem form /zalzal/ /زَلزَل/ ‘shook’ has quadriliteral root /z⊗l⊗z⊗l/ /ز ⊗ ل ⊗ ز ⊗ ل/. Its present stem form is /juzalzil/ /يُزَلزِل/ ‘shaker’ and from which the active participle /muzalzil/ /مُزَلزِل/ ‘shaking’ is obtained.

The first three examples have triliteral roots but unlike verbs like /kataba/ and /darasa/, whose past forms contain only radical consonants, the past forms of these examples contain more consonants than the three radical ones. The last example is derived from a quadriliteral root, whose four consonants are radical. The four active participles have the same af-vocalism morph[6], i.e., /mu⊗i/. This means that Arabic has two different morphs for the active participle morpheme: the first is /aː⊗i/ and the second is /mu⊗i/. The first is used to derive the active participle from the triliteral base form and the second to derive the active participle from both the base quadriliteral root and the complex morphological forms of both types.

From this perspective, Arabic broken plural is also highly allomorphic; for a given singular pattern, two different plural forms may equally be frequent, and sometimes, for some singulars as many as three further statistically minor patterns are also possible

[6]I differentiated between pure vocalism morphemes and affixed vocalism morphemes. If the vocalism has only vowels like the /a⊗a/ in /katab/, it is vocalism, but if it is affixed with any consonant as in /ma⊗uː/ in /maktuːb/, it is referred to as af-vocalism.

(Soudi, Cavalli-Sforza, & Jamari, 2002). Some examples are presented in table (4.2) below.

Root	Singular Form	Affixes	Plural Form	Gloss
/b⊗ħ⊗r/	/baħr/	/i⊗aː/	/biħaːr/	seas
		/u⊗u/	/buħuːr/	
		/ʔa⊗u/	/ʔabħur/	
/d⊗r⊗b/	/darb/	/u⊗uː/	/duruːb/	ways
		/ʔa⊗aː/	/ʔadraːb/	
		/ʔa⊗u/	/ʔadrub/	
/ʃ⊗x⊗sˤ/	/ʃaxsˤ/	/u⊗uː/	/ʃuxuːsˤ /	persons
		/ʔa⊗aː/	/ʔaʃxaːsˤ/	
		/ʔa⊗u/	/ʔaʃxusˤ/	

Table 4.2: Different plural forms of the same singular form

I do not attempt to count all Arabic morphemes and their corresponding morphs; however, two other important cases will be discussed in detail in the next two sections. The first is the different forms of the same vocalism morpheme that interlock with the root morph to compose the verb stems, and this case will be discussed in section (4.4). The second is the different forms of the same root, as clearly seen in the different surface forms of the weak roots, which will be discussed as part of the problem of morphophonemic alternations in section (4.5).

4.4 Root and vocalism matching

Roots behave differently when they combine with vocalisms to form stems. That is, each root chooses specific vocalisms to combine with to produce its past and present stems. In table (4.3), the six possibilities of combination between triliteral roots and their corresponding vocalisms are shown with an example for each possibility.

Root	Past stem	Pres stem	Imperative stem
/ك ⊗ ت ⊗ ب/ /k⊗t⊗b/	/كَتَب/ /katab/	/كتُب/ /ktub/	/كتُب/ /ktub/
/ض ⊗ ر ⊗ ب/ /dˤ⊗r⊗b/	/ضَرَب/ /dˤarab/	/ضرِب/ /dˤrib/	/ضرِب/ /dˤrib/
/ف ⊗ ت ⊗ ح/ /f⊗t⊗ħ/	/فَتَح/ /fataħ/	/فتَح/ /ftaħ/	/فتَح/ /ftaħ/
/ش ⊗ ر ⊗ ب/ /ʃ⊗r⊗b/	/شَرِب/ /ʃarib/	/شرَب/ /ʃrab/	/شرَب/ /ʃrab/
/ح ⊗ س ⊗ ن/ /ħ⊗s⊗n/	/حَسُن/ /ħasun/	/حسُن/ /ħsun/	/حسُن/ /ħsun/
/ح ⊗ س ⊗ ب/ /ħ⊗s⊗b/	/حَسِب/ /ħasib/	/حسِب/ /ħsib/	/حسِب/ /ħsib/

Table 4.3: Six possibilities of combination between triliteral roots and their corresponding vocalisms

The six possibilities of the combination process in the table can be illustrated as follows:

1. The root /k⊗t⊗b/ combines with the morph /a⊗a/ to form the past stem and with the morph /u/ for the present and imperative.

2. The root /dˤ⊗r⊗b/ combines with the morph /a⊗a/ to form the past stem and with the morph /i/ for the present and imperative.

3. The root /f⊗t⊗ħ/ combines with the morph /a⊗a/ to form the past stem and with the morph /a/ for the present and imperative.

4. The root /ʃ⊗r⊗b/ combines with the morph /a⊗i/ to form the past stem and with the morph /a/ for the present and imperative.

5. The root /ħ⊗s⊗n/ combines with the morph /a⊗u/ to form the past stem and with the morph /u/ for the present and imperative.

6. The root /ħ⊗s⊗b/ combines with the morph /a⊗i/ to form the past stem and with the morph /i/ for the present and imperative.

Arabic classical grammarians attempted to handle this problem by classifying roots into basic and sub-categories. Roots are generally classified in terms of the number of consonants into triliteral and quadriliteral. The two basic root types are then sub-categorized into different types. I relied upon this classification and formulated it in AVMs that represent the input and output classes of the morphs. Particularly, each root morph has a morphological class with four attributes: type, sub-type, form and stem second (ʕajnu lfiʕli عَينُ الفِعْلِ) or stem-ʕayn.

- Type: This attribute differentiates between the two basic types of roots: triliteral and quadriliteral. This means that the type attribute in the morphological class of any root has a set of values that has two elements and consequently has the following range:

  $$\mathrm{rg}(\mathrm{TYPE}) = \{tri, quad\}$$

  Examples of verbs that are derived from triliteral roots include /ʤalasa/ /جَلَسَ/ 'he sat down', /qaraʔa/ /قَرَأَ/ 'he read' and /naːma/ نَامَ/ 'he slept'. On the other hand, verbs like /barhana/ /بَرهَن/ 'he proved' and /baʕθara/ /بَعثَرَ/ 'he scattered' are examples of verbs derived from quadriliteral roots.

- Subtype: This attribute classifies Arabic roots depending on the nature of the letters of the root, not on their number, and position of specific letters within the root. Roots are divided into:

  – Sound (assaːlim اَلسَّالِم): sound roots are free of the two semi-vowels, the hamza and the reduplication, like /k⊗t⊗b/.

  – Geminate (almudʕaʕʕf اَلمُضَعَّف): geminate roots contain a reduplicated

letter, like /m⊗r⊗r/.

- First hamzated[7] (mahmuːzu lfaːʔi مَهمُوزُ الفَاء): first hamzated roots begin with hamza like /ʔ⊗k⊗l/. This value is referred to in the lexicon as "fhamzated".
- Second hamzated (mahmuːzu lʕajn مَهمُوزُ العَين): the second consonant in this type of roots is hamza like /s⊗ʔ⊗l/. This value is referred to in the lexicon as "samzated".
- Third hamzated (mahmuːzu llaːm مَهمُوزُ اللَّام): the third consonant in this type of roots is hamza like /q⊗r⊗ʔ/. This value is referred to in the lexicon as "thamzated".
- Assimilated (almiθaːl اَلمِثَال): roots which begin with either the semi-vowel waːw or jaːʔ are called assimilated, like /w⊗q⊗f/ and /j⊗b⊗s/.
- Hollow (alʔaʤwaf اَلأَجوَف): hollow roots have either the semi-vowel waːw or jaːʔ in the middle, like /q⊗w⊗l/ and /b⊗j⊗ʕ/.
- Defective (annaqisʕ اَلنَاقِص): defective roots have either the semi-vowel waːw or jaːʔ in the end, like /d⊗ʕ⊗w/ and /m⊗ħ⊗j/.
- First and third weak (allafiːf almafruːq اَللَّفِيف المَفرُوق): in this type both the first and the third letters are semi-vowels, like /w⊗ʃ⊗j/. This value is referred to in the lexicon as "ftweak".
- Second and third weak (allafiːf almaqruːn اَللَّفِيف المَقرُون): in this type both the second and the third letters are semi-vowels, like /ʃ⊗w⊗j/. This value is referred to in the lexicon as "stweak".

Thus, the subtype attribute has a set of values that contains ten elements and consequently has the following range:

[7]Hamza means glottal stop

$$\text{rg(Subtype)} = \{sound, geminate, fhamzated, shamzated, thamzated, assimilated, hollow, defective, ftweak, stweak\}$$

- Form: As mentioned before, Arabic roots are either triliteral or quadriliteral, but both can host other consonants. That is, three other non-radical consonants can be added to the triliteral roots, while the quadriliteral roots can only have two more consonants. In practice, not every lexical root occurs in all different forms, but they vary from one another, and dictionaries normally list all the forms in which a lexical root regularly appears (Ryding, 2005, p. 434). If the root only consists of the three or four consonants, it is in the base form or form I; otherwise it is in one of the forms II to X. Therefore, in terms of the form, the root takes a Roman number extending form I to X in case of triliteral roots and from I to IV in case of quadriliteral roots. For instance, from the root /k⊗t⊗b/, /katab/ 'he wrote' is in form I, /kattab/ 'made someone write' is in from II, /ka:tab/ 'correspond' is in form III, /ʔktab/ 'made someone write' is in form IV, /taka:tab/ 'wrote to each other' is in form VI and so on until we reach Form X. Thus, the form attribute in the morphological class has the following range:

$$\text{rg(FORM)} = \{I, II, III, \cdots, X\}$$

- Stem second: The short vowel that follows the second root consonant of the verb is traditionally referred to as ħarakatu ʕajni lfiʕli (حَرَكَةُ عَينُ الفِعْلِ 'stem-second vowel')(Ryding, 2005) and I showed in the beginning of this section that the stem-second vowel may vary from one stem to another. I capture these possibilities using an attribute with a value set that has six values. These values specify the stem-second vowel in past and present forms, and therefore, this attribute has the following range:

$$\text{rg(ʕayn)} = \{\text{au}, \text{ai}, \text{aa}, \text{ia}, \text{uu}, \text{ii}\} \ .$$

Thus, the root /k⊗t⊗b/ has the following morphological class:

$$\begin{bmatrix} \text{TYPE}: & \textit{tri} \\ \text{SUBTYPE}: & \textit{sound} \\ \text{FORM}: & \textit{I} \\ \text{STEM-ʕAYN}: & \textit{au} \end{bmatrix}$$

This morphological class says that /k⊗t⊗b/ is triliteral , sound, has the form I and its second consonant is followed by /a/ in the past stem, i.e., /katab/ and by /u/ in the present stem, i.e., /ktub/. Compare this with the morphological class of the root /dˤ⊗r⊗/:

$$\begin{bmatrix} \text{TYPE}: & \textit{tri} \\ \text{SUBTYPE}: & \textit{sound} \\ \text{FORM}: & \textit{I} \\ \text{STEM-EAYON}: & \textit{ai} \end{bmatrix}$$

Both roots share the same value for the attributes TYPE, SUBTYPE and FORM, but differ in the attribute STEM-ʕAYN. /k⊗t⊗b/ takes *au* but /dˤ⊗r⊗/ takes *ai*. The morphological classes of the roots contain only output-classes. On the other hand, vocalism morphs are provided with an in-class (for the hosted root morph) and an out-class (for the result of the merge). The merge is defined only if the out-class of the root matches the in-class of the vocalism. Consider, for example, the in-class of the vocalism morph /u/.

$$\begin{bmatrix} \text{TYPE}: & \textit{tri} \\ \text{SUBTYPE}: & \{\textit{sound}, \textit{fhamzated}, \textit{shamzated}, \textit{thamzated}\} \\ \text{FORM}: & \textit{I} \\ \text{STEM-EAYON}: & \textit{au} \end{bmatrix}$$

This in-class of the vocalism morph /u/ says that it hosts a triliteral sound or hamzated root in the first form, whose second consonant is followed by /a/ in the past stem and /u/ in the present stem. The out-class of the root /k⊗t⊗b/ and the in-class of the vocalism morph /u/ match, and can, therefore, be merged

using the following handler:

$$H := \langle\langle (0, \text{false}), (1, \text{false}), (0, \text{true}), (2, \text{false})\rangle\rangle$$

$$H(u, k\otimes t\otimes b) := k\hat{}t\hat{}u\hat{}b = ktub$$

This applies to the vocalism morph /i/ which merges with the root morph /dˤ⊗r⊗b/ to give the present stem /drib/. It is now clear that both */drub/ and */ktib/ are not allowed because neither the vocalism morph /u/ can merge with the root morph /dˤ⊗r⊗b/, nor the vocalism morph /i/ is allowed to merge with the root morph /k⊗t⊗b/. This is because the associated classes, in either cases, do not match.

4.5 Morphophonemic alternations

The interaction among the different types of morphs may result in phonological alternations. I showed that Arabic roots are broadly classified into two types in terms of the presence or the absence of the two semi-vowels: waːw and jaːʔ . They are generally referred to as weak roots, whose stems may undergo changes when they inflect. Consider, for example, the different forms of the same hollow roots /b⊗j⊗ʕ/ /ب⊗ي⊗ع/ in (47), (46) and (47).

(45) بَاعَ مُحَمَّدٌ السَّيَّارَاتِ

baːʕa moħammadun assajjaːraːti
sell.3.sg.m.pst Mohammad.nom.sg.m car.acc.pl.f.def
'Mohammad sold the cars.')

(46) بِعتُ السَّيَّارَاتِ

biʕtu assajjaːraːti
sell.1.sg.n.pst car.acc.pl.f.def
'I sold cars')

(47) يَبِيعُ مُحَمَّدٌ السَّيَّارَاتِ

jabiːʕu moħammadun assajjaːraːti
sell.3.sg.m.prs Mohammad.nom.sg.m car.acc.pl.f.def
'Mohammad sells the cars'

As shown in the examples, the same root morpheme /b⊗j⊗ʕ/ /ب⊗ي⊗ع/ has three different forms:

- /b⊗aː⊗ʕ/ /ب ⊗ ا ⊗ ع/ in /baːʕa/ /بَاعَ/ 'he sold', masculine third person singular in the past form;
- /b⊗ʕ/ /ب ⊗ ع/ in /biʕtu/ /بِعتُ/ 'I sold', masculine or feminine first person singular in the past form, and
- /b⊗iː⊗ʕ/ /ب ⊗ ي ⊗ ع/ in /jabiːʕu/ /يَبِيعُ/ 'he sells', masculine third person singular in the present form.

Arabic grammarians attribute such morphophonemic changes to the rule of "origin of ʔalif" in verbs like /baːʕa/ /بَاعَ/ and /qaːla/ /قَالَ/ derived from /b⊗j⊗ʕ/ and /q⊗w⊗l/ respectively.

They argue that the "ʔalif" is returned to its original jaːʔ or waːw and provide other justifications for the dropping of the second letter of the root in /biʕtu/ for instance. In all cases, I am not concerned with discussing the rules that justify such changes because in order to ensure the purely surface-oriented treatment of Arabic word formation, no rule is used to delete or modify any string. That is, no rule is used, for example, to change the "ʔalif" of the hollow verb to its origin its original jaːʔ or waːw in specific forms or to drop it in /qultu/ /قُلتُ/ 'I said'. This applies to other cases in which the interaction among different morphs may result in some changes. Instead, each root is a morpheme that has a set of morphs that correspond to its all possible forms under merge. Thus, for the root morpheme /b⊗j⊗ʕ/, I list three morphs:

- /b⊗j⊗ʕ/, from which /jabiːʕu/ /يَبِيعُ/ 'he sells' and the imperative form /biːʕuː/

/بِيعُوا/ ‘sell’ are derived.

- /b⊗ʕ/, from which /biʕtu/ /بِعتُ/ ‘I sold’ and the jussive form /jabiʕ/ /يَبِع/ ‘sell’ are derived.
- /b⊗aː⊗ʕ/, from which /baːʕa/ /بَاعَ/ ‘he sold’ and /baːʕuː/ /بَاعُوا/ ‘they sold’ are derived.

Similarly, for the root morpheme /q⊗w⊗l/, I list three morphs:

- /q⊗w⊗l/, from which /jaquːlu/ /يَقُولُ/ ‘he says’ and the imperative form /quːluː/ /قُولُوا/ ‘say’ are derived.
- /q⊗l/ from which /qultu/ /قُلتُ/ ‘he says’ and the jussive form /qul/ /قُل/ ‘say’ are derived.
- /q⊗aː⊗l/, from which /qaːla/ /قَالَ/ ‘he said’ and /qaːluː/ /قَالُوا/ ‘they said’ are derived.

This not only generates the grammatical forms but also disallows the ungrammatical ones like */jaqaːluː/ */يَقَالُوا/ or */jaqultuː/ */يَقُلتُ/ and */jabaːʕu/ */يَبَاعُوا/ or */jabiʕtu/ */يَبِعتُ/. The different forms of the hollow roots are important from two perspectives: they represent another instance for the distinction between the notions morpheme and morph, and their proper handling ensures the theoretical surface orientation of AMRS.

4.6 Summary

In this chapter, I described the morphological component of the morphosemantic lexicon of (a fragment of) Arabic. I introduced two important aspects of Arabic morphology,

i.e., discontinuity of morphemes and reduplication, and I solved two problems, namely, the root and vocalism matching and the morphophonemic alternations. At the very beginning, I reviewed the most famous theoretical and computational approaches to Arabic morphology. I showed that there is a gap in the current approaches between morphology and semantics and argued that AMRS differs from these paradigms in its compositionality, its disregarding the distinction between morphology and syntax and its being more restrictive. Arabic roots and vocalisms are dealt with as fractured glued strings, i.e., they are allowed to be discontinuous. The combination process is executed through handlers. This mechanism saves the concatenative paradigm but requires finding out the matching between the roots and the vocalisms. This is done using the morphological classes. Each root is provided with an out-class that specifies with which vocalism it combines, and each vocalism is provided with an in-class that says which morph it hosts. I also presented some examples of morphemes that have different morphs and showed that all morphs of the same morpheme should be inserted in the lexicon, as is the case of the hollow roots. In the next chapter, I move from the mere word formation to more complex constituents and show how Arabic sentence can be handled within AMRS.

Chapter 5

Arabic Sentence: A Syntax-Semantics Interface

In this chapter, I use AMRS as an interface between the syntax and semantics of Arabic. I move from the morphological component to the semantic structure of the units. Similar to morphology, the sentence structure of Arabic involves complex phenomena such as flexible word order, null-subject and copula-less clauses in a tensed environment. Although word order in Arabic is relatively free, morphological information plays a crucial role in the assignment of semantic roles. Also, agreement has a basic role in the structure of Arabic phrases and clauses. I start the chapter with a distinction between word order and sentence type and then show how AMRS provides a powerful compositional mechanism for handling the two types of sentences and their related phenomena.

5.1 Word order versus sentence type

It is crucial for the proper semantic analysis to identify the constituents of the sentence, which in turn requires a correct classification of sentence type (Farghaly, 2010). In languages like English or German, the verb lies in the core, and the position of other elements is determined relative to it. But the case is different in Arabic because a distinction should be first made between two terms: sentence type and word order. A good deal of discussion has been devoted to word order, and the focus has been placed

on whether Arabic is a SVO or a VSO language. This results in a confusion between the two clearly distinct issues and in some cases, word order alterations have been used to refer to sentence types. This way of dealing with Arabic sentence structure neglects a significant part of the Arabic sentences, namely those which do not contain a verb, neither in the first position nor in the second, i.e., copula-less (verbless) sentences.

5.1.1 Verbal sentences

Verbal sentences contain a verb (phrase) as one of their basic two predicative constituents. The verb (phrase) composes with its subject (noun phrase) and maybe some other complements a sentence. To the contrary of the traditional approach, which differentiates verbal sentences with a noun (phrase) in the first position from those that begin with a verb (phrase), I deal with both as different forms of word order of the same sentence type. In this regard, I am not concerned with the position of the arguments relative to the verb as a criterion of classification, but with the licit versus illicit word orders, on the one hand, and with the meaning of the sentence on the other. Each verb has a sub-categorization frame that specifies the number and the type of arguments it takes. That is, intransitive verbs take one argument. The verb and the argument can exchange their positions; however, the sentence maintains grammaticality and its semantics.

(48) مَاتَ الرَّجلُ

maːta arraʤulu
die.3.sg.pst.m man.nom.sg.m.def
'The man died'

(49) اَلرَّجلُ مَاتَ

arraʤulu maːta
man.nom.sg.m.def die.3.sg.pst.m

‘The man died’

Adverbial phrases are allowed to appear at any position resulting in the following possibilities (given X stands for the adverbial phrase, S is the subject, V is the verb):

XVS - VXS - VSX - XSV - SXV - SVX

In the case of transitive verbs, the following word orders are allowed:

(50) كَتَبَ مُحَمَّدٌ الدَّرسَ

kataba muħammadun addarsa
write.3.sg.m.pst Mohammad.nom.sg.m lesson.acc.sg.m.def
‘Mohammad wrote the lesson’

(51) مُحَمَّدٌ كَتَبَ الدَّرسَ

muħammadun kataba addarsa
Mohammad.nom.sg.m write.3.sg.m.pst lesson.acc.sg.m.def
‘Mohammad wrote the lesson’

(52) اَلدَّرسَ كَتَبَ مُحَمَّدٌ

addarsa kataba muħammadun
lesson.acc.sg.m.def write.3.sg.m.pst Mohammad.nom.sg.m
‘Mohammad wrote the lesson’

(53) كَتَبَ الدَّرسَ مُحَمَّدٌ

kataba addarsa muħammadun
write.3.sg.m.pst lesson.acc.sg.m.def Mohammad.nom.sg.m
‘Mohammad wrote the lesson’

As can be observed from the examples, verbal sentences can have four different word orders:

VSO - SVO - OVS - VOS

The verb only occurs in the first or the second position and does not appear in the last position. This fact means that there are two illicit forms, and they should not be allowed, namely, *SOV and *OSV. So, the following two sentences are ungrammatical:

(54) *مُحَمَّدٌ الدَّرسَ كَتَبَ

*muħammadun addarsa kataba
Mohammad.nom.sg.m lesson.acc.sg.m.def write.3.sg.m.pst
'Mohammad wrote the lesson'

(55) *اَلدَّرسَ مُحَمَّدٌ كَتَبَ

*addarsa muħammadun kataba
lesson.acc.sg.m.def Mohammad.nom.sg.m write.3.sg.m.pst
'Mohammad wrote the lesson'

Ditransitive verbs take two arguments in the accusative case in addition to the subject with a relatively flexible order. That is, in the case of ditransitive verbs, 12 word orders are possible, 6 if the sentence begins with a verb and 6 if it begins with a noun. Again, the verb cannot be preceded by two of its arguments, i.e., it occurs either in the first or the second position. Also, adverbs can appear anywhere. Moreover, Arabic has a special controversial category of verbs that may take three objects. This category includes /ʔaxbara/ /أَخبَرَ/, /xabbara/ /خَبَّرَ/, /ʔaʕlama/ /أَعلَمَ/, /ʔaraː/ /أَرَى/, /ʔanbaʔa/ /أَنبَأَ/, /nabbaʔa/ /نَبَّأَ/, /ħaddaða/ /حَدَّثَ/ and all have the same meaning 'to tell or to inform'. The behavior of this group of verbs is shown in (56).

(56) أَخبَرَ الحَارِثُ الزَائِرَ المَكَانَ مُغلَقًا

ʔaxbara alħaːriθu azzaːʔira almakaːna muɣlaqan.
tell.3.sg.pst guard.nom.def visitor.acc.def place.acc.def closed.acc
'The guard told the visitor that the place was closed.'

However, this construction is not very frequent in the modern usage of Arabic, and it is much more preferred to use a /ʔanna/ /أَنَّ/ 'that' clause instead of the second and

the third objects as in (57).

(57) أَخبَرَ الحَارِثُ الزَائِرَ أَنَّ المَكَانَ مُغلَقٌ

?axbara alħa:riθu azza:?ira anna almaka:na muɣlaqun.
tell.3.sg.pst guard.nom.def visitor.acc.def that place.acc.def closed.nom.def
'The guard told the visitor that the place is closed.'

Actually, there is no consensus on the acceptability of this construction. That is, while Danks (2011, p. 107) reports that he examined the entries of Wehr (1994) and found a small number of verbs that take three objects as shown in (58).

(58) أَبدَلَ المَلِكُ خَدَّامَهُ ذَهَبًا بَيتَهُ

?abdala almaliku xadda:mahu ðahaban bajtahu
exchange.3.sg.pst king.nom.def his-servant.acc.def gold.acc his-house.acc
'The king gave his servant gold in exchange for his house.'

Dickins and Watson (2009) consider these indirect objects adverbial complements (see the footnote in Danks (2011, p. 106)).

5.1.2 Verbless sentences

Verbless clauses are not a unique phenomenon of Arabic. Many other languages share it with Arabic such as Hebrew and Russian. First, it is important to differentiate verbless clauses from the so-called small clauses that occur in many languages like English, Spanish, German and also in Arabic and cover a wide range of constructions including, among others, consider-type and find-type (González-Rivera, 2010, p. 75).

(59) I consider [Bielefeld beautiful].

(60) Ich finde [Bielefeld schön].
I.nom.1.sg find.1.sg.prs Bielefeld.acc beautiful
'I find Bielefeld beautiful'

(61) أَرَى بِيلِيفِلدَ جَمِيلَةً

ʔarː Bielefeld ʤamiːlatan.
find.1.sg.prs Bielefeld.acc beautiful.acc.f
'I find Bielefeld beautiful'

The first two constructions, in English and German, can be separately dealt with as a subject and predicate containing an empty verb (be) (González-Rivera, 2010, p. 76), and without a verb in the case of Arabic, as indicated in the following:

(62) Bielefeld is beautiful.

(63) Bielefeld ist schön.
Bielefeld.nom be.3.sg.prs beautiful.
'Bielefeld is beautiful'

(64) بِيلِيفِلدُ جَمِيلَةٌ

Bielefeld ʤamiːlatun.
Bielefeld.nom beautiful.nom.f
'Bielefeld is beautiful'

Following Stowell (1981) in his argument that predication is possible without verbs, den Dikken (2006) defines a small clause as "a subject-predicate structure lacking tense". As is clear, small clauses in English and German, for example, are not allowed to form a main clause without the verbs "be" and "sein" respectively. They do not accept main be-less clauses in a tensed environment (González-Rivera, 2010, p. 92). On the other hand, Arabic allows for the construction of main clauses without a verb or an overt copula.

Arabic classical grammarians referred to the possibility to construct a main clause from a noun phrase as a mubtadaʔ (مُبتَدَأً 'subject/topic of verbless sentences') and another noun phrase (65), an adjective phrase (66), a prepositional phrase (67), a verbal clause (68), or a verbless clause (69) as a xabar (خَبَر 'predicate').

(65) عَلِيٌّ أُستَاذٌ فِي الجَامِعَةِ. NP (NP PP)

ʕalijjun ʔustaːðun fiː lʤaːmiʕati
Ali.nom.sg.m professor.nom.sg.m in university.gen.sg.f.def
'Ali is a professor at the university'.

(66) عَلِيٌّ ذَكِيٌّ. NP AP

ʕalijjun ðakijjun
Ali.nom.sg.m intelligent.nom.sg.m
'Ali is intelligent'.

(67) عَلِيٌّ فِي الجَامِعَةِ. NP PP

ʕalijjun fiː lʤaːmiʕati
Ali.nom.sg.m in university.gen.sg.f.def
'Ali is in the university'.

(68) عَلِيٌّ ذَهَبَ إِلَى الجَامِعَةِ. NP (VP PP)

ʕalijjun ðahaba ʔilaː alʤaːmiʕati
Ali.nom.sg.m go.3.sg.m.pst to university.gen.sg.f.def
'Ali went to the university'.

(69) عَلِيٌّ أَبُوهُ أُستَاذٌ فِي الجَامِعَةِ. NP (NP (NP PP))

ʕalijjun ʔabuːhu ʔustaːðun fiː lʤaːmiʕati
Ali.nom his-father.nom professor.nom in university.gen.def
'Ali, his father is a professor at the university'.

All sentences that contain a noun phrase in the initial position, including SVO and OVS, are traditionally classified as "nominal sentences". This is not my approach in this dissertation. I argue that the difference is related to contextual, rather than structural or semantic, matters. That is, in the case of the NP-initial sentences, the focus of the sentence is the NP, whereas the action itself is the focus of the VP-initial ones. Consider, for example, (70) and (71) as two different word orders of the same sentence.

(70) مُحَمَّدٌ ضَرَبَ هِندًا

muħammadun dˤaraba hindan
Mohammad.nom.sg.m hit.3.sg.m.pst Hend.acc.sg.f
'Mohammad hit Hend'.

(71) ضَرَبَ مُحَمَّدٌ هِندًا

dˤaraba muħammadun hindan
hit.3.sg.m.pst Mohammad.nom.sg.m Hend.acc.sg.f
'Mohammad hit Hend'.

(70) can be seen as an answer of the question: Who hit Hend? But (71) can be considered an answer of the question: What did Mohammad do? From a syntactic viewpoint, the subject is the same in the two cases, namely, /muħammadun/ /مُحَمَّدٌ/ and also the object, /hindan/ /هِندًا/. In both sentences, /muħammadun / has the semantic role of the agent and /hindan/ has the semantic role of the patient. The transitive verb /dˤaraba / /ضَرَبَ/ 'he hit' takes them as its two arguments and the both (70) and (71) have the same semantic representation.

x, y
$x = Mohammd'$ $y = Hend'$ $hit'(x, y)$

Figure 5.1: DRS for both /muħammadun dˤaraba hindan/ and /dˤaraba muħammadun hindan/

The shift between the two orders VSO and SVO in the language of Arabic newspapers provides a piece of evidence. Watson (1999) observes that the headlines of Arabic newspaper usually take the SVO order, while the same message is repeated in the first line but in VSO order. As an example consider (72) and (73).

(72) أَلمَانيَا تُنَظِّمُ بُطُولَةَ أُورُوبَا [Headline SVO]

ʔalmaːnjaː tunaðˤðˤimu butˤuːlatu ʔuruːbː
Germany.nom organise.3.sg.prs championship.acc.sg.f Europe.gen
'Germany organizes the European Championship.'

(73) تُنَظِّمُ أَلمَانيَا بُطُولَةَ أُورُوبَا [First Sentence VSO]

tunaðˤðˤimu ʔalmaːnjaː butˤuːlatu ʔuruːbː
organise.3.sg.prs Germany.nom championship.acc.sg.f Europe.gen
'Germany organizes the European Championship.'

The headline sentence begins with the subject to attract the attention of the reader to the winner of the right of hosting the championship. In this case, the name of the winner is the focus. It is Germany, not Turkey. This is not the case in the first sentence because the reader already knows who the winner is. Traditionally, (73) is dealt with as a simple verbal sentence that consists of a verb and its subject and object. On the other hand, (72) is dealt with as a sentence, whose predicate is a complete verbal sentence. To account for this, Arabic grammarians assume a hidden pronoun that plays the role of the subject of the verb and refers to the initial NP as indicated in (74).

(74) مُحَمَّدٌ ضَرَبَ (هُوَ) هِندًا

muħammadun dˤarab-a (huwa) hind-n
Mohammad.nom.sg.m hit.3.sg.m.pst (he).nom.3.sg.m Hend.acc.sg.f
'Mohammad hit Hend'.

The most important syntactic principle in the traditional Arabic grammatical theory is alʕamal (اَلعَمَل 'governance'). According to this principle, every case marker is definitely an effect of a specific governor- typically overt, and can be covert (abstract). The notion of "abstract governor" has generally been developed to deal with the cases of the absence of the overt governor, and particularly to deal with the absence of the assigner

of the nominative case for the initial NP (the subject or the topic) in both verbless and verbal sentences. In NP-initial sentences, the NP, which occupies the first position, is assigned the nominative case by an abstract governor, no matter what follows. However, a debate has arisen on what assigns the nominative case of the NP-predicate. Some grammarians have proposed that it is the subject that assigns the nominative case of the predicate, others have suggested that both the subject and the abstract governor assign it together, and a third group has assumed both the subject and the predicate assign the case of each other mutually. However, none assumed that verbless clauses have a hidden copula.

In modern Arabic linguistics, one debate has arisen on whether verbless sentences contain a silent copula (González-Rivera, 2010, p. 116). While Bakir (1980) and Fassi (1993) suggest that there is a silent copula in Arabic verbless sentences, Benmamoun (2008, 2000) rejects this assumption and provides the following two arguments against the existence of the silent copula:

- The overt copula assigns an accusative case to the predicate as in (75), but in copula-less clauses, both the subject and the predicate are assigned a nominative case as in (76).

(75) كَانَ الجَوُّ بَارِدًا

kaːna	alʤawwu	baːridan
be.3.sg.pst.m	weather.nom.sg.m.def	cold.acc.sg.m

'The weather was cold'.

(76) اَلجَوُّ بَارِدٌ

alʤawwu	baːridun
weather.nom.sg.m.def	cold.nom.sg.m

'The weather is cold'.

- In Arabic, the copula appears in past and future and only disappears in the present, which requires, given a silent copula, a deletion rule.

In generative transformational grammar, small and verbless clauses are considered a projection of a functional head, i.e., (pred^0) (Bowers, 2001), (5.2: a), tense (Benmamoun, 2008, 2000),(5.2: b), or an abstract functional head that can be a verb, a tense, or a copula, overt or covert (Den Dikken, 2006) (5.2: c). The latter was adopted by González-Rivera (2010) to deal with Spanish verbless small clauses.

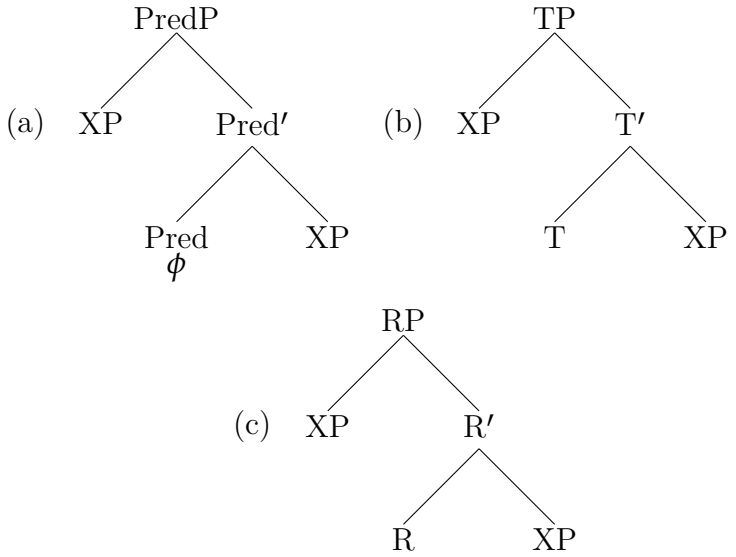


Figure 5.2: Syntactic structure of verbless sentences in generative grammar

While it is allowed in the transformational paradigm to have a deleted copula using a deletion rule, in non-transformatinoal frameworks, no deletion rules are allowed. In Lexical functional grammar, for instance, neither deletion rules, nor empty categories are allowed (Attia, 2008a, p. 95). The question has been whether to deal with the post-copula complement, and as part of the discussion the predicate in the copula-less clauses, as an open complement (XCOMP) or as a closed complement (PREDLINK). Dalrymple et al. (2004) have claimed that a unified approach cannot be realized either cross-linguistically, or within the same language. To provide an answer to this question, they have differentiated between two situations:

- The optional existence of the copula

  If the occurrence of the copula is optional, then the predicate can function as

a sentential head and consequently subcategorize for the subject, as is the case in the Japanese adjectives. In this case, a single-tier analysis, in which the adjective functions as the head of the predicate clause, is suggested (Attia, 2008b; Nordlinger & Sadler, 2007). The following example is taken form (Dalrymple et al., 2004).

(77) hon wa akai
book Top red
'This book is read'

$$\begin{bmatrix} \text{PRED} & \text{'red <(↑SUBJ)>'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'book'} \end{bmatrix} \end{bmatrix}$$

Figure 5.3: Dalrymple's Analysis for the Japanese sentence: hon wa akai.

- The obligatory existence of the copula

  On the other hand, if the occurrence of the copula is obligatory, then the predicate cannot function as a sentential head and consequently cannot subcategorize for the subject, as is the case in the Japanese nominals. Instead, a double-tier analysis, in which the copula functions as the head of the predicate clause, is suggested (Attia, 2008b; Nordlinger & Sadler, 2007). In this case, the post-copula complement can be represented either as an open function, i.e., XCOMP, or as a closed function, i.e., PREDLINK, depending on the individual properties of each language.

  Attia (2008b) chooses the double-tier closed function as a default for a wide range of coupla constructions. Following Nordlinger and Sadler (2007), Attia deals with Arabic copula-less clauses as having a main predictor, null-be, to satisfy the coherence condition in LFG and to provide some information like tense and negation as shown in the example (78), its phrase structure and its f-structure.

(78) هُوَ طَالِبُ

huwa t$^{\text{ʕ}}$a:libun
he.nom.3.sg.m student.nom.sg.m
'He is a student'

S ⟶ NP ϵ NP ∨ AP
(↑ SUBJ)=↓ (↑ PRED)='null-be <SUBJ, PREDLINK>' (↑ PREDLINK)=↓
(↑ TENSE)=pres (↓ GEND)=(↑ SUBJ GEND)
(↓ NUM)=(↑ SUBJ NUM)

$$\begin{bmatrix} \text{PRED} & \text{'null-be <(↑SUBJ) (↑PREDLINK)>'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'he'} \end{bmatrix} \\ \text{PREDLINK} & \begin{bmatrix} \text{PRED} & \text{'student'} \end{bmatrix} \\ \text{TENSE} & \text{present} \end{bmatrix}$$

Figure 5.4: Attia's Analysis for the Arabic sentence: huwa t$^{\text{ʕ}}$a:libun.

In head driven phrase structure, Sag, Wasow and Bender (1999) have introduced a constructional analysis of the absence of the copula in African American Vernacular English (AAVE). In AAVE, copula-less sentences are possible as in (79).

(79) You in trouble. (Bender, 2001)

To remain surface-oriented, they rejected both empty categories and deletion rules. Instead, they suggested a structural analysis of the copula-less clauses as finite ones composed of a noun phrase and a predicative phrase. These clauses are referred to as zero-copula-phrases.

Shortly, three different analyses have generally been proposed for Arabic verbless clauses: small clause analysis, silent copula analysis and tense projection analysis. Each one has been formalized according to the adopted approach. The small clause analysis is not suitable for Arabic verbless sentences because they are tensed. Benmamoun (2008, 2000) uses the case assignment as a basic argument against the silent-copula analysis and claims that Arabic verbless sentences are projections of a tense phrase. Aoun et al. (2010) follow the same approach. This analysis entails that both the copula and the T

select for the same set of complements which is not true. Normally, the T has either VP or NegP as its complements. In this analysis, it exactly behaves like a copula selecting for the same type of phrases. Moreover, the functional T analysis does not explain why both copula sentences and copula-less sentences have the same range of complements. The same applies to the existential clauses, which are completely the same regardless of the existence or the absence of the copula (Alotaibi, 2018, p. 222-224).

Since the accusative case is assigned in the past and future by the copula, i.e., the case and its marker follow the overt occurrence of the assigner in the clause, the covertness of the assigner entails the absence of its effect. This does not mean there is neither a covert copula nor an invisible state verb. This can be described as the default case, in which the two nominative cases are assigned by an invisible state verb (state of being). The assumption that there is a state verb is motivated by the fact that the copula-less clauses are sentences that happen through a time. This means that some verb should be assumed in the (80) that has the meaning /juːʤadu/ ‘exists’ or /jakuːnu/ ‘is’.

(80) عَلِيٌّ فِي العَمَلِ

ʔalijj-un fiːl-ʕamal-i
Ali.nom.sg.m in the-wor.gen.sg.m.def
‘Ali is at work.’

This was even suggested by classical Arabic grammarians, who allowed a predicative relation between a noun phrase, as a subject, and an adverbial phrase, as a predicate, assuming the adverbial phrase is related to either a hidden noun phrase /kaːʔin/ ‘being’ or a verb /ʔistaqara/ ‘stay (in a certain place)’. On the one hand, I agree with the principle that an invisible element exists, but disagree, on the other, with the possibility of having a hidden noun phrase as the real predicate modified by the overt adverbial phrase. This can be described a problem transmission from a problematic overt adverbial phrase that occurs in the position of the predicate to another problematic covert noun phrase.

This general rule is overridden by some other special rules related to the verbless clauses, i.e., special rule > general rule. Arabic verbless sentences can be preceded, for example,

by a special group of articles referred to as ʔinna waʔaxawaːtuhaː (إِنَّ وَأَخَوَاتُهَا) resulting in a syntactic change by assigning an accusative case to the subject as shown in (81). Similarly, if the copula appears at the surface, it yields a syntactic change by assigning an accusative case to the predicate.

(81) إِنَّ الرَّجُلَ مَرِيضٌ

ʔinna arraʤula mariːdˤun
Indeed man.acc.sg.m.def sick.nom.sg.m
'Indeed the man is sick.'

Benmamoun's second argument is related to the need of a deletion rule in the present tense. This argument can be valid if all frameworks allow for deletion. As is well known, nontransformational approaches do not allow for such deletion rules. A string can simply be empty. The proposed mechanism deals with Arabic in the same way. In the applied approach, nouns are typical arguments, whereas both adverbs and adjectives function as adjuncts, and each verb takes a specific number of arguments to constitute a sentence. When the argument structure of the noun exports a referent that matches one of those imported by the verb, the number of arguments required by the verb is reduced by one. Each verb can be thought of as having two natural numbers (i, j), where i is the number of arguments required by the verb, and j is the number of residual arguments to complete the sentence, such that prior to any merge $i = j$, and every single merge operation reduces j by 1. The sentence is complete when $j = 0$. This is similar to Tesnière's idea of valency and Chomsky's idea of sub-categorization frames.
To make matters specific, the following points are considered in my treatment of Arabic sentence structure:

- Subjects and predicates should sharply be distinguished from lexical categories such as nouns and verbs.

- Sentences are composed of two basic components, namely, a subject and a predicate.
- The subject, as a component, can be a noun (phrase), a complimentizer phrase, or a non-finite clause.
- the predicate[1], as a component, can be composed of one of the following:
    1. a verb without any complement (intransitive verb);
    2. a verb with a complement/complements;
    3. a copula-construction, or
    4. a zero (covert) copula-construction.

Only (transitive) verbs, as a lexical category, can function alone as a predicate, i.e., can occupy the position of the predicate component. Other lexical categories cannot and need to combine with some other categories. The abstract functional head, suggested by den Dikken (2006), is a relatively practical analysis. However, I disagree with this analysis in the assumption that the relator is distinct from the XP predicate. I do not assume a third component as den Dikken's relator is part and parcel of the predicate, both structurally and semantically.

5.2 Arabic sentence: A syntax-semantics interface

In this section, I show how to apply AMRS to both types of Arabic sentence and their related phenomena.

5.2.1 Simple verbal sentence

The most important features of Arabic simple verbal sentences are (i) the predicate is a verb phrase; (ii) the verb agrees with the subject in gender; and (iii) the order of the predicative constituents can vary from the basic one (Ryding, 2005, p. 65). Let's

[1]I mean "in the syntactic notion

consider the example (82) and see how this modified version of RS can be used as a syntax-semantics interface for this type of sentences.

(82) نَجَحَ عَلِيٌّ

naʤaħa ʔalijjun
succeed.3.sg.m.pst Ali.nom.sg.m
'Ali succeeded.'

This VS word order of (82) is represented in (5.5). In the merge operation, the verb takes the role of the functor and the subject takes the role of the argument. If the output name of the argument matches the input name of the functor, the merge succeeds, and the variables are renamed in the same way of the original form of RS. In this example, the variables match and the merge succeeds.

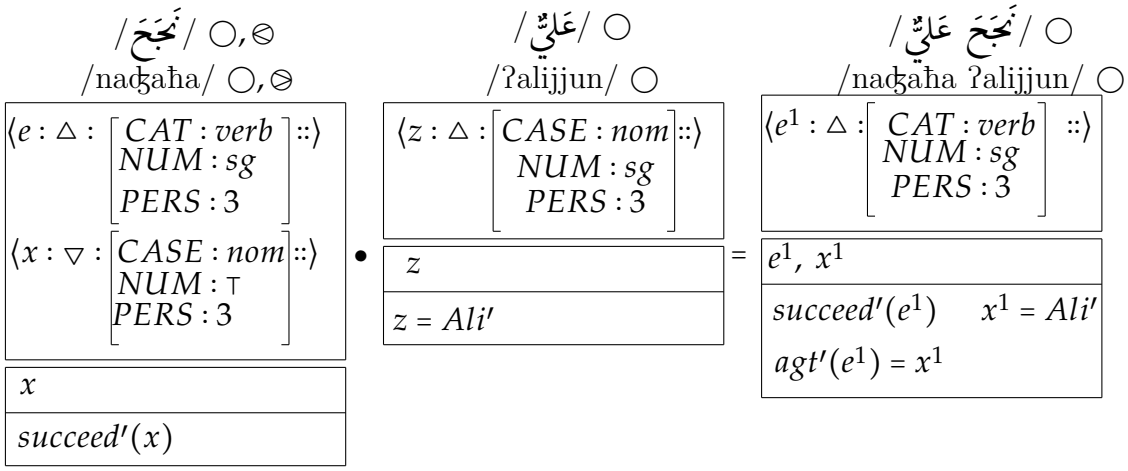


Figure 5.5: Representation and the merge of the sentence: naʤaħa ʔalijjun.

To account for the SV word order, another possibility is inserted but the horizontal diacritic ⊜ is substituted with ⊜. This says that the verb expects an argument from either sides. So, instead of having only one structural possibility in the case of intransitive verbs, I add two: one to account for sentences that begin with the verb and the

other for sentences that begin with the noun. The whole process proceeds in the same way as shown in the semantic representation of (83) in (5.6).

(83) عَلِيٌّ نَجَحَ

ʔalijjun naʤaħa
Ali.nom.sg.m succeed.3.sg.m.pst
'Ali succeeded.'

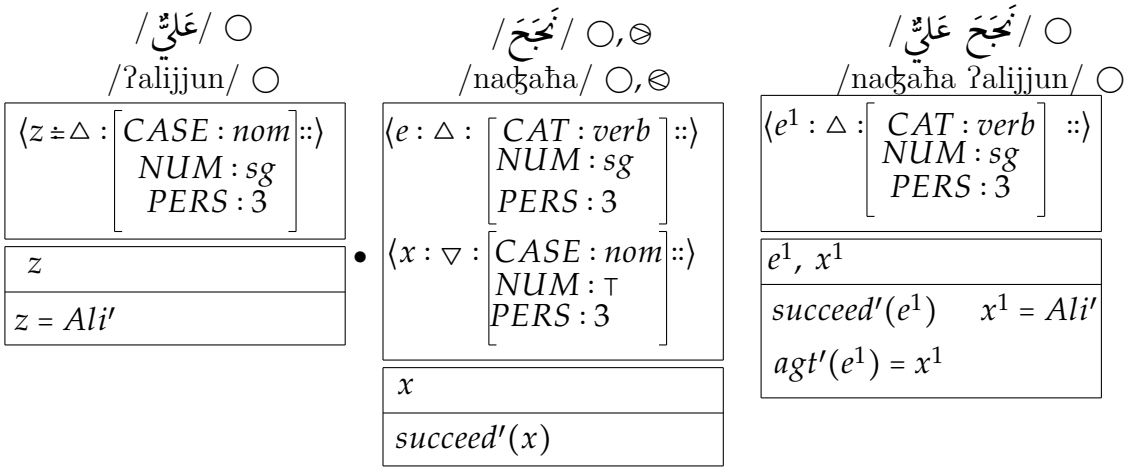


Figure 5.6: Representation and the merge of the sentence: ʔalijjun naʤaħa

5.2.2 Verbal sentences with transitive verbs

As is the case with intransitive verbs, the argument identification structures of transitive verbs are not accessed one by one in a restricted order, but they can be skipped if their information does not match. In other words, the argument identification structures are G-accessed not E-accessed[2]. As an example of handling verbal sentences with transitive verbs, I show how (84) is handled.

(84) مُحَمَّدٌ يُشَاهِدُ فِيلمًا

[2]See the two types of access in the access definition in section (3.4).

muħammadun juʃaːhidu fiːlman
Mohammad.nom.sg.m watch.3.sg.m.prs movie.acc.sg.m
'Mohammad is watching a movie.'

In this order, the verb takes two arguments: one from its right side and one from its left side. The merge in AMRS is fully associative. So, the verb can first merge with the object and then the obtained constituent combines with the subject. The verb may also combine with the subject and then the two with the object. I follow the first choice. The verb /juʃaːhidu/ /يُشَاهِدُ/ 'is watching' combines with the direct object /fiːlman/ /فِيلمًا/ 'a movie' as shown in (5.7).

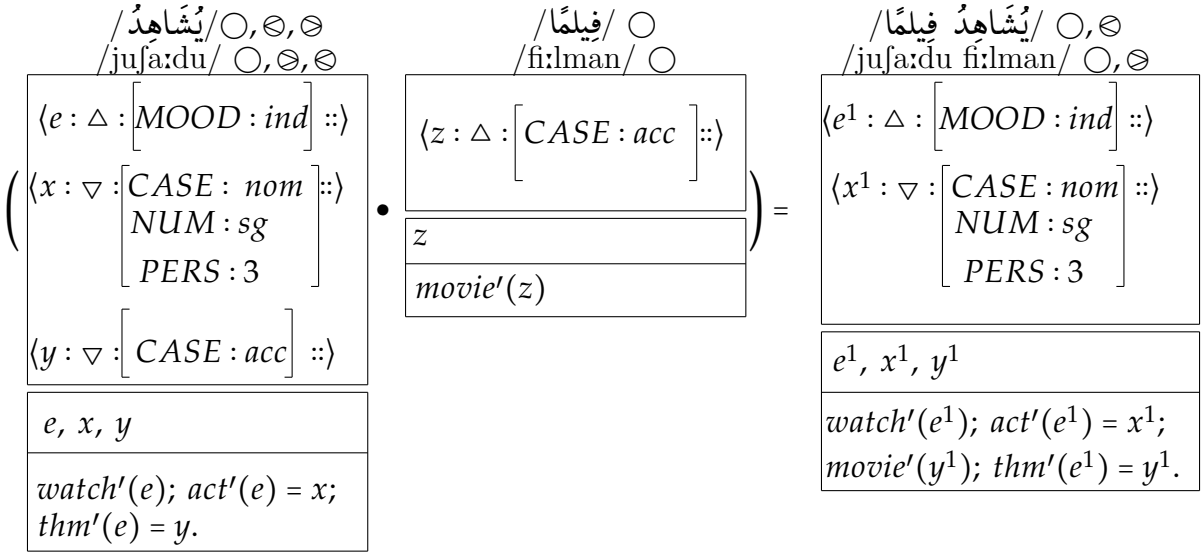


Figure 5.7: Merge of the verb /juʃaːdu/ and its object /fiːlman/

The renaming process depends on whether the variable is identified or not; if it is identified, it is appended 1, denoted here as x^1 instead of $x1$, and if it is not identified, the variables of the right DRS are appended 1, and those of the left DRS are appended 2. Here, the verb imports two arguments: one under the name nominative singular third person and the other under the name accusative. The y variable in the DRS of the verb

/juʃaːhidu/ is identified with the z variable in DRS of the noun phrase /fiːlman/, so they are renamed to x^1. The other two variables in the DRS of the verb are not identified; therefore, they are appended [1] as well because it is the left DRS. Then, both can merge with the subject /muħammadun as represented in (5.8).

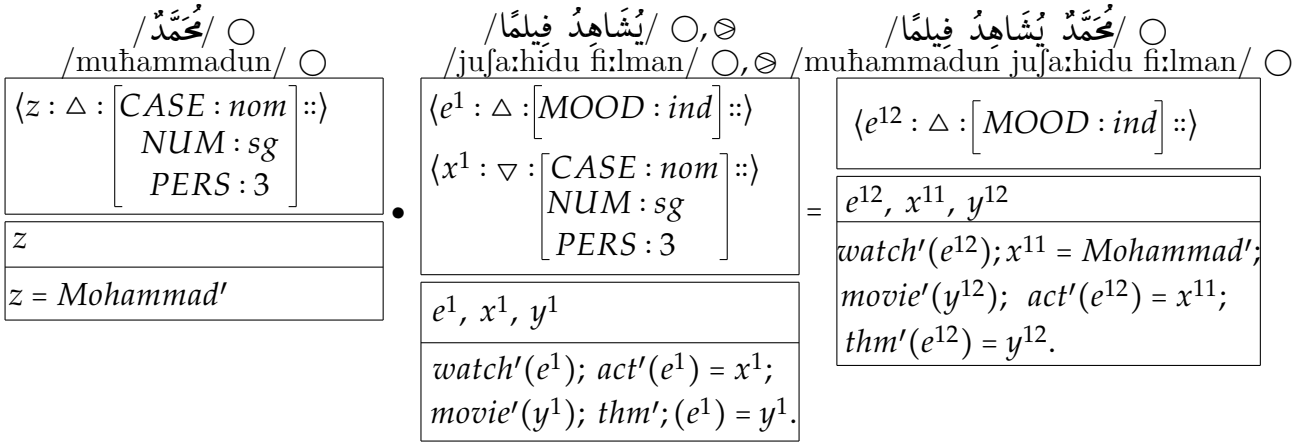


Figure 5.8: Merge of the subject /muħammadun/ and the verb phrase /juʃaːdu fiːlman/

Using G-access, the merge between the verb and its two arguments succeeds whether the order is /muħammadun juʃaːhidu fiːlman/, i.e., SVO, or /fiːlman juʃaːhidu muħammadun/, i.e., OVS. The merge succeeds, if the intersection between the argument structure of the verb /juʃaːhidu/ and that of any of its two arguments is not empty. However, even using G-access, this semantic structure does not account for the VP-initial clauses, i.e., VSO and VOS. The problem lies in the direction of the vertical diacritics. The above structure states that the verb expects two arguments: one from its right side and one from its left side. The flexibility, offered by the G-access mechanism, allows the two arguments to exchange their positions, as long as they surround the verb, but both cannot appear in the same side. Therefore, we make a disjunction relation between two structures, such that the verb can have one argument at each side or two arguments (only) at its left side. This is represented in (5.9) and (5.10) in which I follow the VSO order.

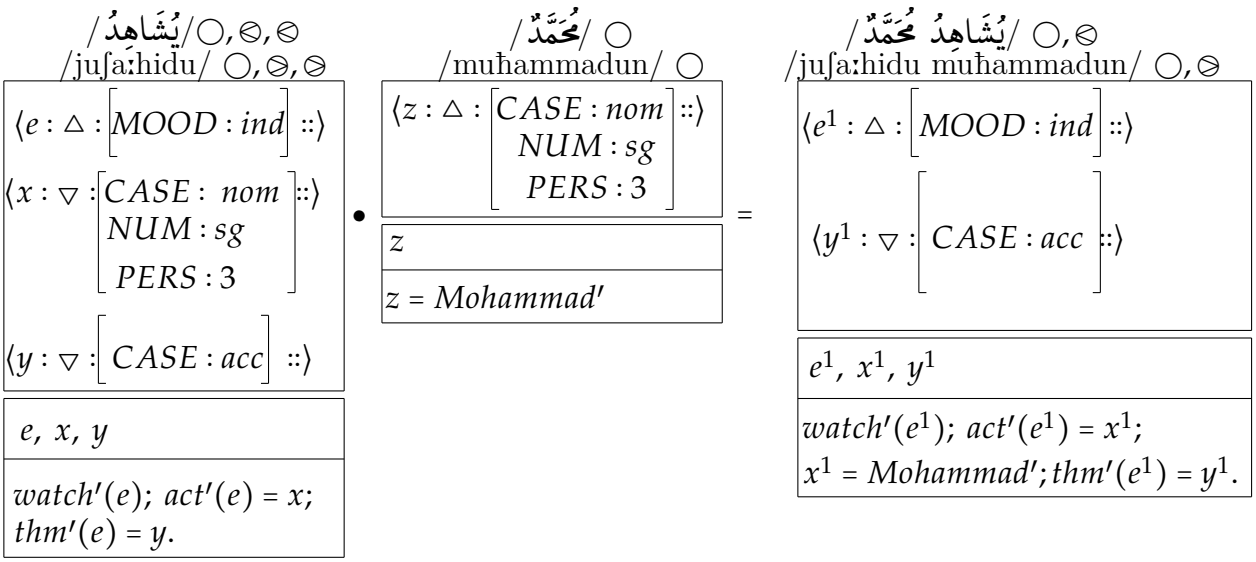


Figure 5.9: Merge of the verb /juʃaːdu/ and the subject /muħammadun/

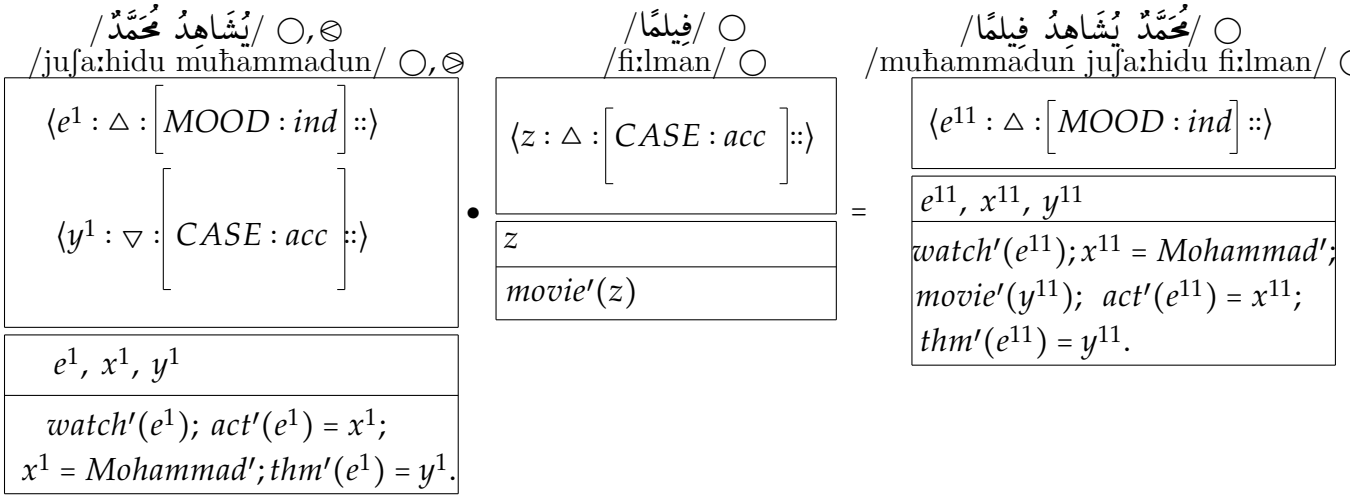


Figure 5.10: Merge of the subject and the verb /juʃaːdu muħammadun/ with the object the subject /fiːlman/

5.2.3 Verbal sentences with ditransitive verbs

As mentioned above in (5.1.1), Arabic verbless sentences can begin either with the verb or with one of its arguments. If a sentence begins with one of the arguments, its syntactic structure is similar to its German counterpart. In German main sentences, the verb occurs at the second position and the arguments can flexibly change their

positions. This is the same in Arabic. The verb can only be preceded by one argument, regardless of which one it is, i.e., it occupies the second position. It is true that the calculus is powerful enough to account for the flexibility of Arabic word order, but it suffers from a limitation in the case of ditransitive verbs. Unlike German, which distinguishes morphologically between accusative and dative objects, the two types of objects are assigned in Arabic accusative case markers. In this part, I only focus on the power of the calculus to account for the syntactic structure, and this limitation will be discussed in the evaluation of the theory. (85) is an example of a sentence with a ditransitive verb. The sentence begins with the verb followed by the subject and the indirect object, and the direct object comes at the very end.

(85) مَنَحَ أَحمَدُ عَليًّا جَائِزَةً

manaħa	ʔaħmadu	ʕalijjan	ʤaːʔizatan
give.3.sg.m.pst	Ahmed.nom.sg.m	Ali.acc.sg.m	prize.acc.sg.f

'Ahmed gave Ali a prize.'

There are no constraints on the order of the merge. The verb can consume its arguments in any order, regardless of their positions. It can follow the order of the sentence starting by consuming the subject followed by the indirect object and ending by the direct object. It can jump to the direct object skipping both the subject and the direct object and then consume either of them. Here, I show one possibility, say the order of the sentence as it appears. The three merge steps are represented in (5.11), (5.12) and (5.13).

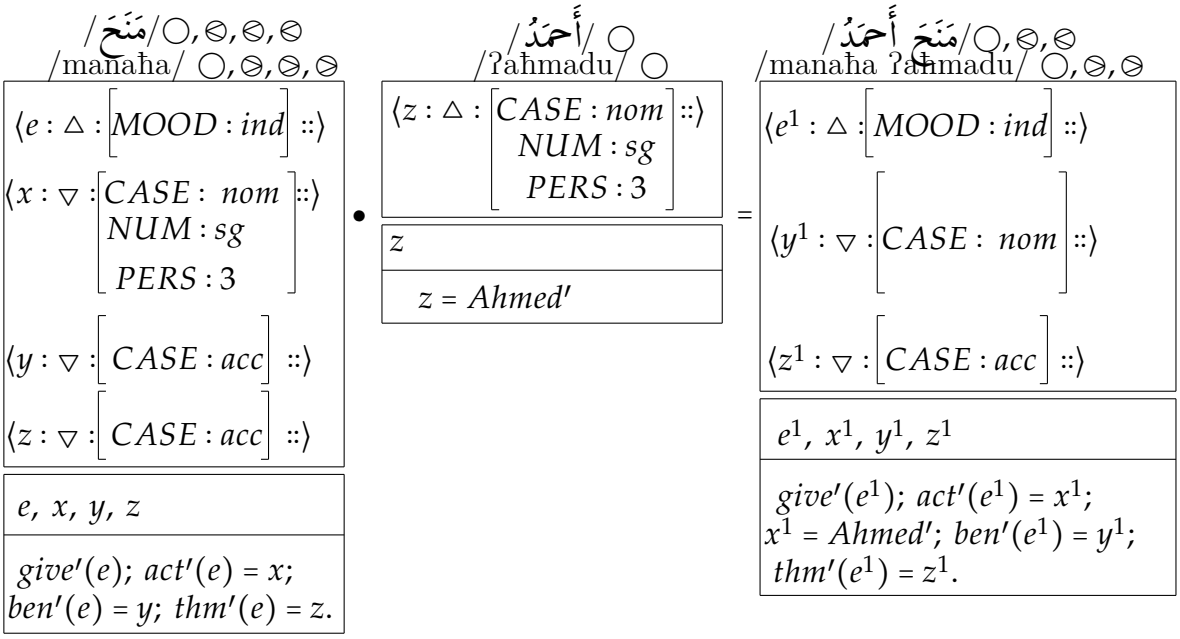


Figure 5.11: Merge of the verb /manaħa/ and the subject /ʔaħmadu/

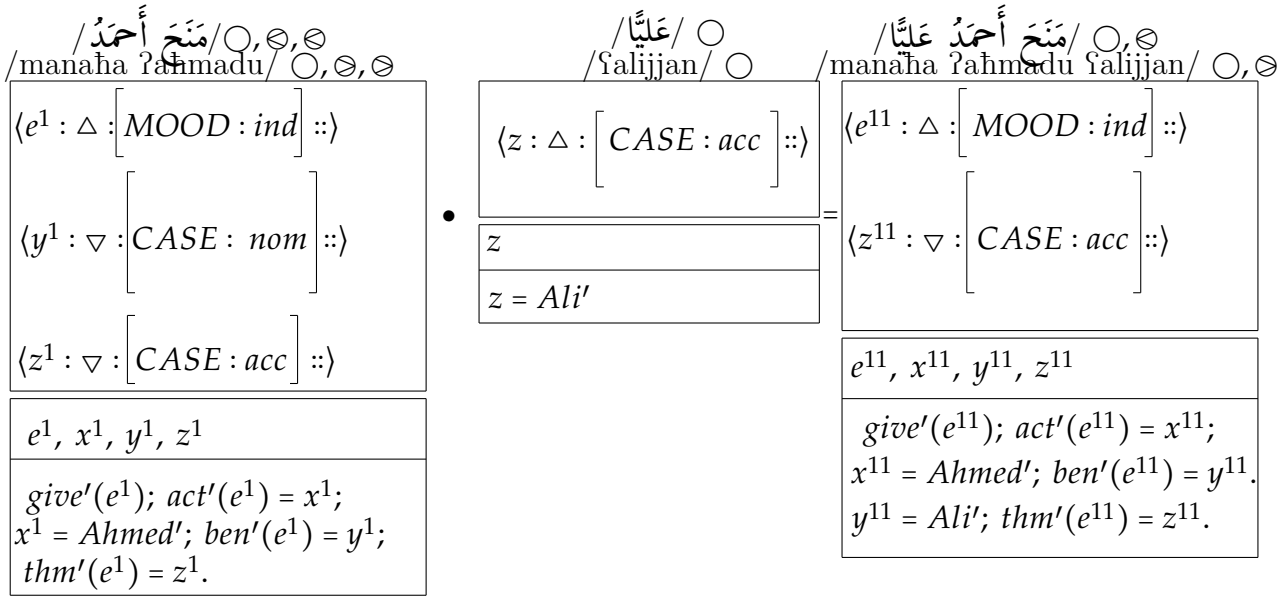


Figure 5.12: Merge of the verb and the subject /manaħa ʔaħmadu/ with the indirect object /ʕalijjan/

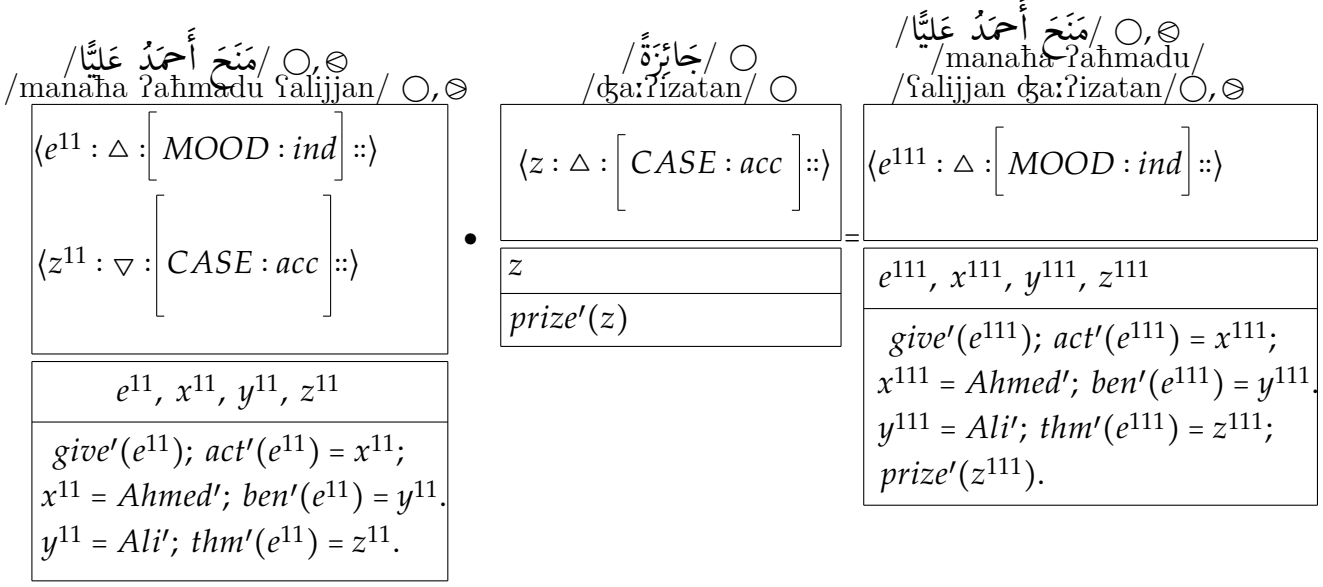


Figure 5.13: Merge of the verb, the subject /manaħa ʔaħmadu ʕalijjan/ with the direct object /ʤa:ʔizatan/

It is also very common to use the preposition /l/ /ل/ 'to' before the indirect object which is consequently assigned the genitive case as shown in (86). This does not make a big difference in the merge process because in such a case, the verb imports a prepositional phrase as its indirect object instead of importing a noun phrase.

(86) مَنَحَ أَحمَدُ جَائِزَةً لِعَلِيٍّ

manaħa	ʔaħmadu	ʤa:ʔizatan	liʕalijjin
give.3.sg.m.pst	Ahmed.nom.sg.m	prize.acc.sg.f	to-Ali.gen.sg.m

'Ahmed gave Ali a prize.'

5.2.4 Pro-drop vs reduplicated arguments

In addition to flexible word order, verbal sentences exhibit two other contradictory phenomena: pro-drop and reduplication of arguments. Arabic, like some other languages

such as Italian and Spanish, allows for the possibility to drop the subject pronoun, when the information provided by the subject pronoun can be recovered (Farghaly, 2010). To the contrary, verbal sentences may contain one argument that is expressed twice. Consider, for example, the following two sentences.

(87) أَدَى المُهِمَّةَ

ʔadaː almuhimmata
do.3.sg.m.pst job.acc.sg.f.def
'He did the job.'

(88) اَلكِتَابَ قَرَأَهُ مُحَمَّدٌ

alkitaːba qaraʔahu moħammadun
book.acc.sg.m.def read.3.sg.m.pst+it.acc.sg.m Mohammad.nom.sg.m.
'The book, Mohammad read it.'

In (87), the subject is dropped; however, its information can be recovered, thanks to the verb subject agreement. That is, the subject of the verb /ʔadaː/ /أَدَى/ 'did' is a third person masculine singular. The null-subject is inserted in the lexicon as an empty string, and the merge process proceeds exactly in the same way as the overt subject because each verb has a specific number of required arguments to constitute a sentence. This number is equal to the imported variables in the argument structure of the verb. When the argument structure of the noun (overt or dropped) exports a referent that matches one of those imported by the verb, the variable is consumed and the number of the required arguments is reduced by one. The representation in (5.14) shows how the empty string of the null-subject combines with the verb and its object.

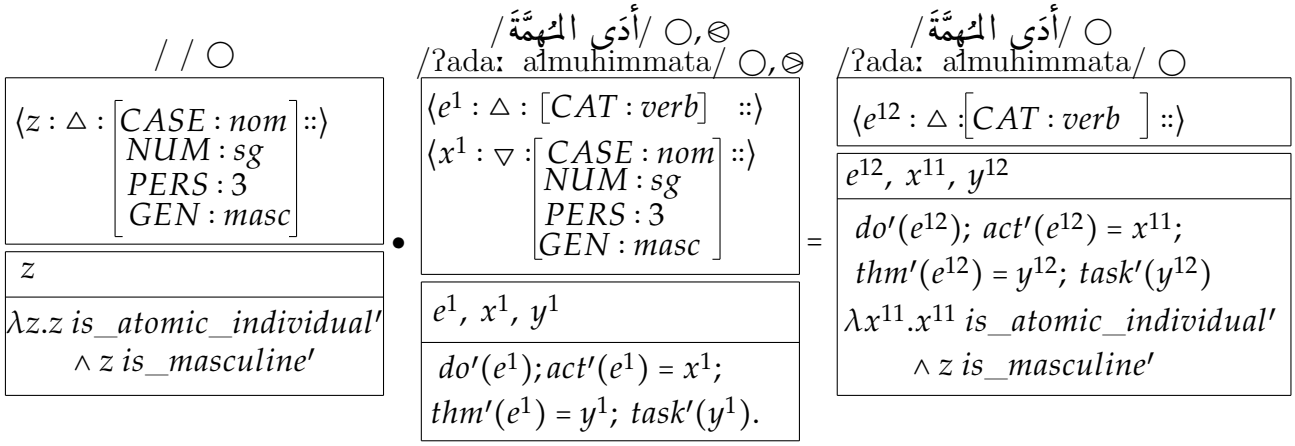


Figure 5.14: Merge of the empty string of the null-subject with the verb and its object

On the other hand, the theme argument in (88) is expressed twice: one time as a noun phrase in the accusative case and the second one as an anaphoric pronoun that agrees with its antecedent. I suggest two solutions: the first is to deal with these anaphoric pronouns as a special group of agreement suffixes that attach to the verb. In this case, the number of the arguments imported by the verb remains unchanged. The second is to allow the verb to have another argument, such that the third argument must be an object pronoun that is equal to its antecedent and agrees with it. I preferred the second solution in the practical implementation. Thus, to represent (88), I allow the verb /qaraʔa/ /قَرَأَ/ 'read' to import three arguments: the subject /muħammadun/, the object /alkitaːba/ and the object pronoun /hu/ /هُ/ 'it' which refers to the object /alkitaːba/ and agrees with it in gender, case and number. It is also worth noting that the pronoun and its antecedent can differ in case. This is usually referred to in the literature as "left dislocation" as shown in (89) taken from Ross (1967, p. 424).

(89) My father, he's Armenian, and my mother, she's Greek.

To account for this syntactic structure, the verb /qaraʔa/ has the two representations

in (5.15).

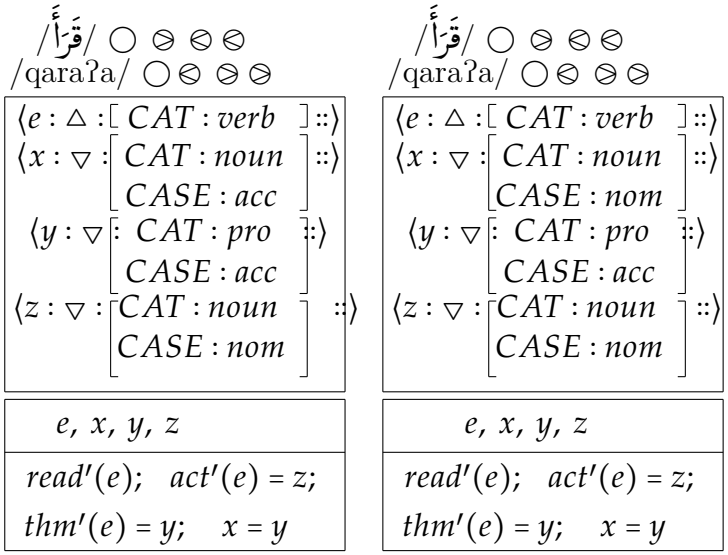


Figure 5.15: Two representations for the verb /qaraʔa/

The two representations account for one possible word order but for two possible cases of the first argument. The antecedent can have either a nominative case or an accusative case, but the word order remains as follows:

$$Arg_1(noun)\ Verb\ Arg_2(pro)\ Arg_3(noun)$$

The three arguments can be represented as shown in (5.16).

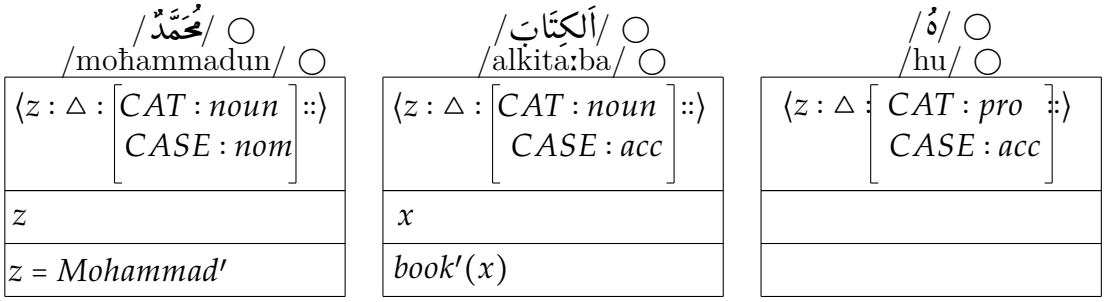


Figure 5.16: Representations for three arguments

Again, the merge operation is associative. The verb can merge with its three imported arguments regardless of the order of the merge. If it is the case that the verb first merges with the pronoun, the result will be as described in (5.17).

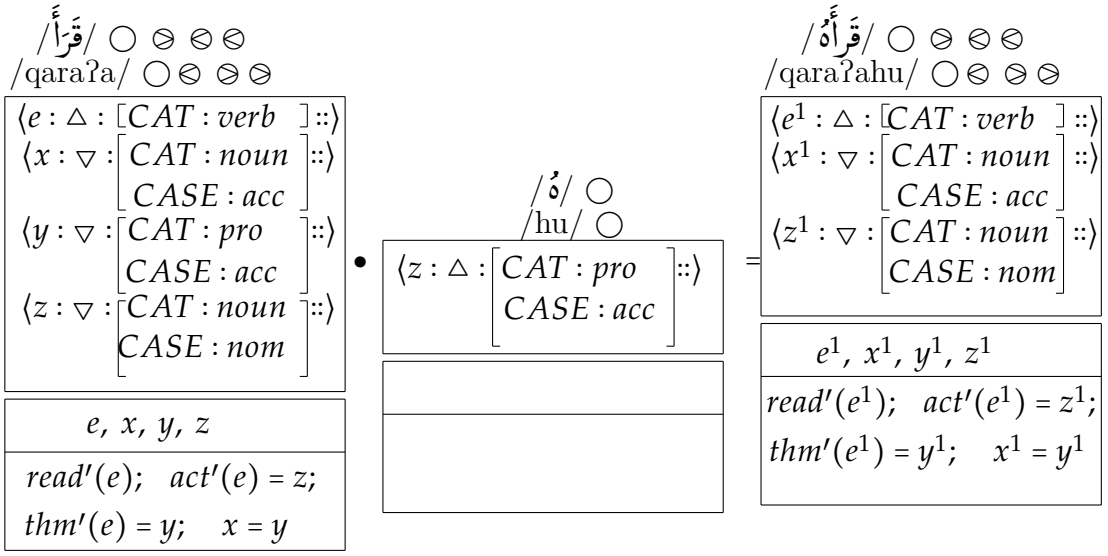


Figure 5.17: Merge of the verb /qaraʔa/ and the object pronoun

The verb phrase can then merge either with the subject or with the antecedent of the pronoun. If it merges with the subject, the result in (5.18) is obtained.

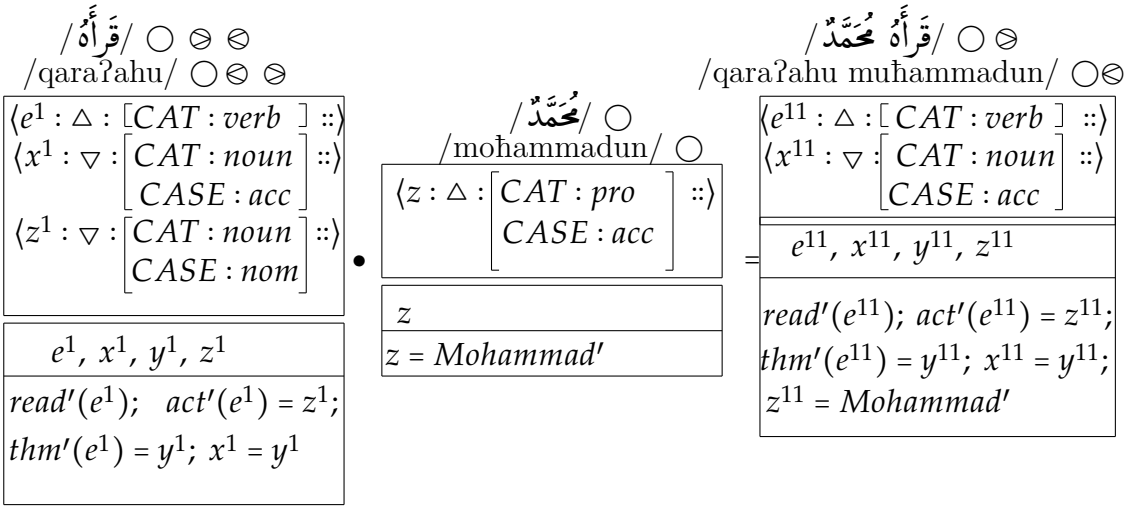


Figure 5.18: Merge of the verb /qaraʔa/ and the object pronoun with the subject

Finally, the whole clause can merge with the remaining argument to get a whole sentence with a duplicated argument as shown in (5.19).

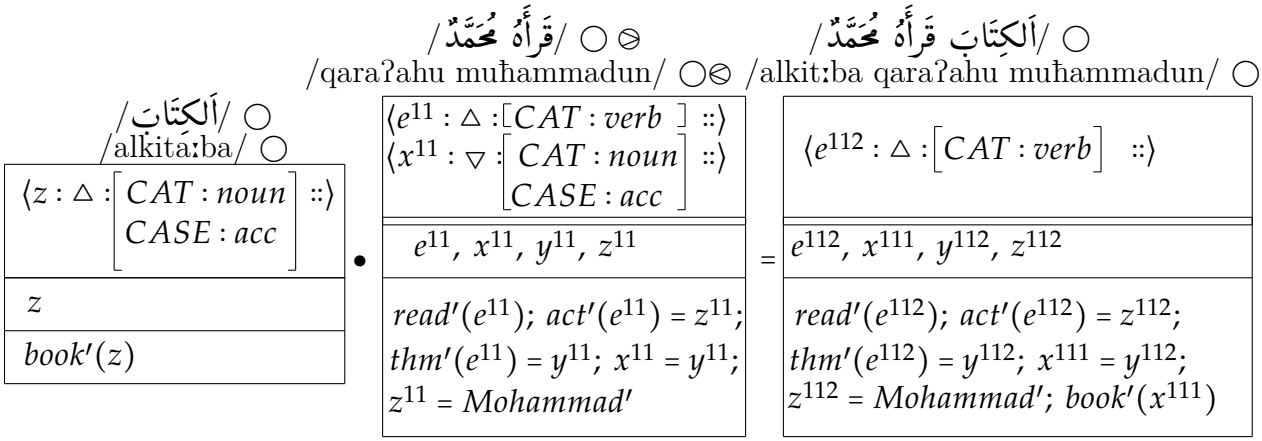


Figure 5.19: Whole sentence with a duplicated argument

In the previous example, I showed how to capture the three arguments if the first argument has the accusative case. Similarly, the verb can import them if the first argument is in the nominative case. The difference lies in the value of the case attribute.

5.2.5 Verbless sentence

In section (5.1.2), I discussed in some detail the proposed analyses for the copula-less structures both in transformational and nontransformational approaches. I also argued for the covert copula, and accordingly the copula in Arabic has two semantic structures:

1. The argument structure of the overt copula imports a noun phrase in the nominative case and another noun/adjective phrase in the accusative case, or an adverbial phrase.

2. The empty string of the covert copula (or state verb) is distinct from the overt one in two points:

   - The two imported phrases are in nominative case.
   - It is always in the present tense.

The following two representations in (5.20) show the difference between the two types of copula [3].

[3]Tense and properties are dealt with as parameters and they will be discussed in detail in the chapter devoted to parameters. In this representation, I ignored tense and introduced the property variable p.

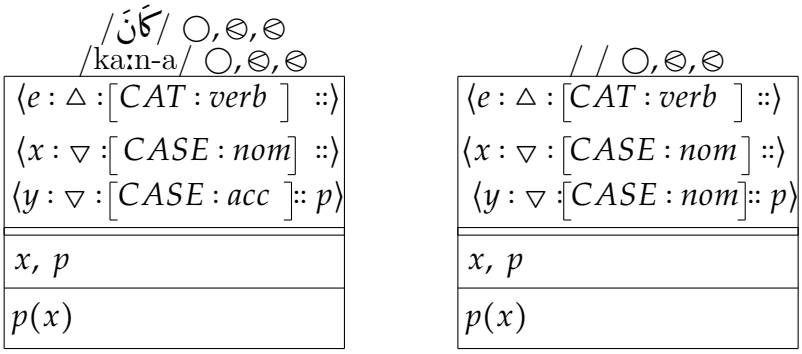


Figure 5.20: Two representations of the overt copula and the empty string

As an example, I explain how the merge process proceeds in the case of the absence of the copula as in (90).

(90) عَلِيٌّ طَبِيبٌ

ʔalijjun tˤabi:bun
Ali.nom.sg.m doctor.nom.sg.m
'Ali is a doctor'

Assuming that /ʔalijjun/ /عَلِيٌّ/ 'Ali' and tˤabi:bun /طَبِيبٌ/ 'doctor' have the two semantic structures in (5.21), the covert copula, i.e., the empty string, imports either of them first and then the other. If it first merges with the second argument to compose the predicate component, then the result will be as shown in (5.22).

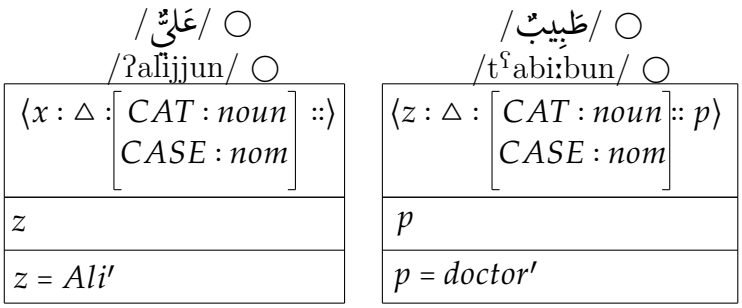


Figure 5.21: Semantic structures of /ʔalijjun/ and /tˤabi:bun

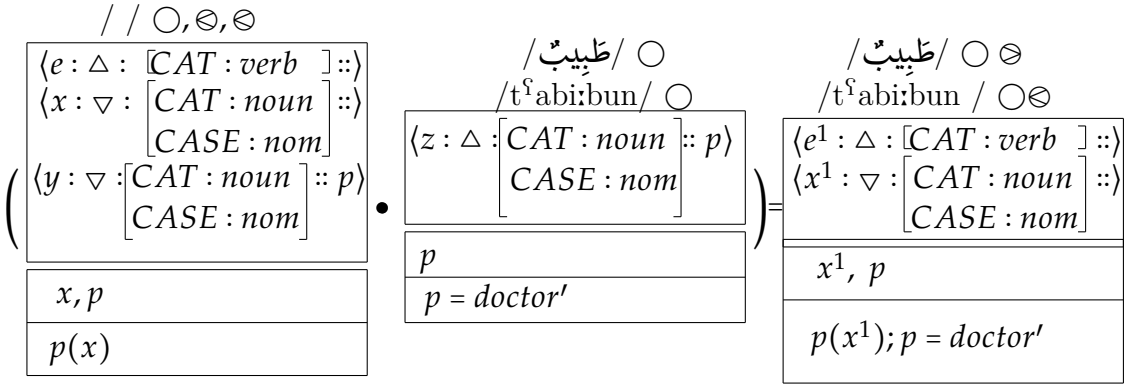


Figure 5.22: Merge of the empty string of the covert copula with the second argument

Then, the predicate component can merge with the subject as shown in (5.23).

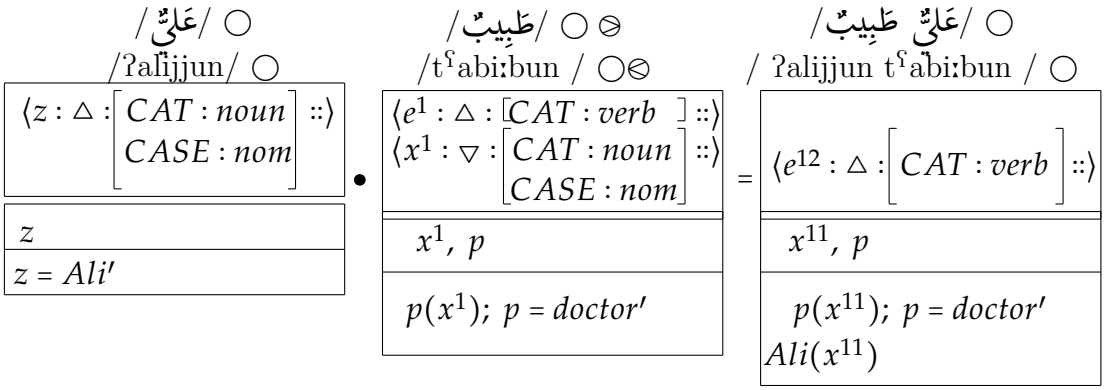


Figure 5.23: Merge of the predicate component with the subject

5.3 Summary

AMRS is a morphologically motivated theory of semantic composition. This raises questions on its adequacy to handling the syntactic aspects of language. From a structural perspective, Arabic represents an excellent case study to show the powerful mechanism of AMRS to account for complex phenomena such as flexible word order, null-subject and copula-less clauses in a tensed environment. In this chapter, I introduced in some detail the syntactic structure of Arabic. I started with a distinction between two notions: word order and sentence type and showed that Arabic has two different types of sentences: verbal sentences and verbless (copulaless) sentences, and it also has flexible

word order. Verbless sentences are main sentences without a verb or an overt copula. Three famous analyses have been introduced to deal with the absence of the copula, namely, the small clause analysis, the silent copula analysis and the functional head analysis. I dealt with the null-subject and covert copula as empty strings. The null-subject information can be recovered thanks to the verb subject agreement. The empty string of the covert copula differs from the overt copula in two aspects: it is always in the present, and the two imported phrases are assigned the nominative case. The latter is considered a default case, whereas the existence of the copula has a special rule which overrides the rule of the general case. All the solutions provided in this chapter are supported with examples that show the powerful mechanism of AMRS to handle the syntactic structure of Arabic and its related phenomena.

Chapter 6

Parameters: Properties, Tense and Aspect

The proper semantics of any sentence must account for the parameters that affect its meaning. These parameters include, among others, properties, tense, aspect, and world. AMRS includes parameters as the fourth component of the argument identification statement. By adding it, an AIS is a quadruple $\langle x : \delta : \Xi :: P \rangle$, where x is a variable, δ is a diacritic; Ξ is a pair of AVSs; and P is a pair of parameters. Theoretically, the calculus allows to add as many parameters as one needs, but I add only three: properties, tense and aspect. In this chapter, I show the significance of parameters and the mechanism of handling them.

6.1 Properties

Kracht (2016) provides three reasons for introducing properties. The first is because they are different from all kinds of entities. The second is because they help solve the technical problems of assigning meanings to inflectional morphemes and finally due to the lack of a direct mechanism for abstraction. That is, in addition to objects and events, properties should be assumed. To argue for his assumption, Kracht compares the two sentences in (91) and (92).

(91) John is the fool.

(92) People call John a fool.

While (91) refers to John as that individual, (92) denotes John who has the prop-

erty of being fool. Another piece of evidence comes from syntax. Objects appear after demonstrative pronouns, but properties do not, as shown in (93) (Kracht, 2016, p. 171).

(93) *People call John this fool.

This means that noun phrases can denote either an object or a property. He gives an additional piece of syntactic evidence from what is referred to as "lack of discourse markers" in Hungarian. When a noun phrase denotes a property, no back anaphoric reference is allowed.

Moreover, Kracht argues in favor of properties from a semantic point of view. The semantic evidence comes basically from the so-called non-intersective adjectives. It has largely been agreed upon that adjectival modifiers are not semantically equal, but there are different kinds of adjectival modifications (Lassiter, 2015). This means that nouns and adjectives do not combine in the same way. Therefore, a typology of the relationship between nouns and adjectives has long suggested in formal semantics. For an arbitrary adjective A and an arbitrary noun N, the relationship between A and N can be one of the following, using entailment and substitution failure test, due to Siegel (1976) (examples are taken from (Lassiter, 2015; Partee, 2003)):

- A is an intersective adjective if and only if for all N, $[\![AN]\!] = [\![A]\!] \cap [\![N]\!]$. An example is vegetarian.

  (94) Al is a vegetarian farmer; Al is a cellist.
  $\models$ Al is a farmer.
  $\models$ Al is a vegetarian cellist.

- A is a subsective adjective if and only if for all N, $[\![AN]\!] \subseteq [\![N]\!]$. An example is skillful.

  (95) Al is a skillful farmer; Al is a cellist.
  $\models$ Al is a farmer.
  $\not\models$ Al is a skillful cellist.

- A is a non-subsective adjective if and only if for all N, there is at least one N such that $[\![AN]\!] \not\subseteq [\![N]\!]$. An example is alleged.

(96) Al is an alleged forger; Al is a pickpocket.
⊭ Al is a forger.
⊭ Al is an alleged pickpocket.

- A is a privative if and only if for all N, $[\![AN]\!] \cap [\![N]\!] = \phi$. An example is fake, since no fake gun is a gun.[1].

This means that in addition to intersective adjectives, subsective adjectives include non-intersective adjectives. As shown in (95), having the property of being a farmer, Al is skillful, but this does not mean Al is a skillful cellist. A skillful farmer might be a very bad cellist or might also be a skillful one, but it is not a must. If an adjective is intersective, it can be supposed that the adjective attributes a property to an object, but this cannot be the case if the adjective is non-intersective, because such a type of adjectives modifies a property rather than attributing a property to an object (Kracht, 2016, p. 174). This applies to the gradable adjectives like /tall/, /small/ and /big/. In order to get an appropriate interpretation of a modified noun like a /small elephant/ or a /big turtle/, the standard of comparison, which can only be calculated with respect to a relevant comparison class, should be specified (Kennedy 2007). A very expensive T-shirt is clearly very cheap if compared to a modern car, and the biggest housefly is obviously very small if compared even to the smallest elephant and so on. (97) is an illustrative example form Partee (1995).

(97) Win is a tall 14-old-boy.
Win is a basketball player.
⊭ Win is a tall basketball player.

According to Kracht, the mystique of the semantics of the non-intersective adjectives can be avoided using properties- either as sorts of their own or as individual concepts. The problem of associating meanings to morphemes is another piece of evidence provided by Kracht in the favor of properties. AMRS presumes a meaning for each and every morpheme, e.g., the plural morpheme, such that every time the plural morpheme, for instance, is attached to a given stem, it gives a group of things satisfying this property. But consider the example in (98) from Kracht (2016).

[1]No full consensus on the existence of this type of adjectives. Partee (2003) argues that it does not truly exist

(98) quattuor magni mures.
four big mice.

If a meaning is associated each time to the plural morpheme, the result would be a group of big mice in addition to its being a group of big things and a group of four elements. This begs the question: is a group of big mice a group of big things? Again, the problem is related to the relevant comparison class. That is, on the one hand, it can be considered a group of big things if we talk about mice, but, on the other, they are definitely a group of small things if we talk about mammals. Kracht's suggested solution is to take adjectives and nouns as denoting properties and the numeral as forming a group of properties (Kracht, 2016, p. 175). The final reason for the necessity of properties is the lack of an abstraction mechanism to get a property from individuals and groups. So, instead of dealing with nouns as objects and adjectives as modifiers of these objects, AMRS proposes that the meaning of adjectives is a function that takes a property as its argument and returns another property. The given property is then attributed to the noun. This simply means that adjectives modify properties rather than objects, and the result is a modified property rather than a modified object. Parameters are associated with particular variables, and whenever an object variable is shared, the parameter variable is shared in parallel, as shown in figure (6.1) (Kracht, 2016).

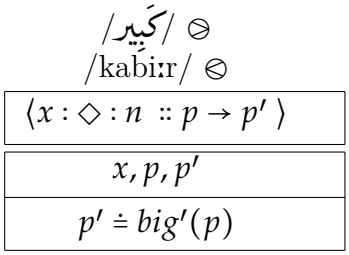


Figure 6.1: Representation of the adjective /kabiːr/

The representation of the adjective /kabiːr/ /كَبِير/ 'big' in (6.1) shows how adjectives

modify properties; it takes p as its argument and gives a modified property p'. Nouns, on the other hand, do not modify parameters as represented in figure (6.2).

/فَأْر/ ○
/faʔr/ ○

$\langle x : \triangle : \begin{bmatrix} CAT : noun \\ PERS : 3 \\ GEND : masc \\ NUM : sg \end{bmatrix} :: p \rangle$
x, p
$p \doteq mouse'$

Figure 6.2: Representation of the noun/faʔr/

The result of the merge between the adjective and the noun is a property of a modified property as reflected in figure (6.3).

/فَأْرٌ كَبِير/ ○
/faʔrun kabiːrun/ ○

$\langle x : \triangle : \begin{bmatrix} PERS : 3 \\ GEND : masc \\ NUM : sg \end{bmatrix} :: p' \rangle$
x, p, p'
$p \doteq mouse'$; $p' \doteq big'(p)$.

Figure 6.3: Merge of the noun /faʔrun/ and the adjective /kabiːrun/

Unlike non-intersective adjectives, the merge of an intersective adjective and a property is a conjunction of a modified object and a property, since intersective adjectives modify objects. This means that intersective adjectives do not take properties as arguments but objects as shown in the representations of the adjective /nabaːtiː/ /نَبَاتِي/ 'vegetarian'

and the noun /ʃaxsˤ/ /شَخص/ ‘person’ in figure (6.4).

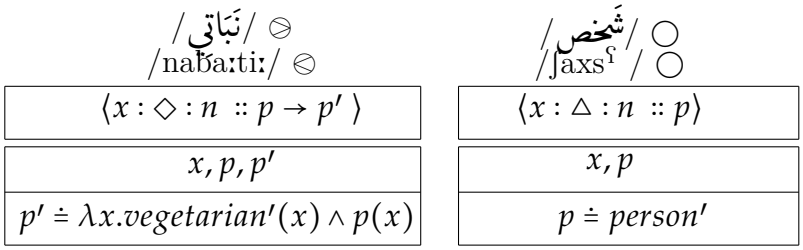


Figure 6.4: Representation of the noun /ʃaxsˤ/ and the adjective /naba:ti:/

Given the variable is identified, the obtained result is represented in figure (6.5).

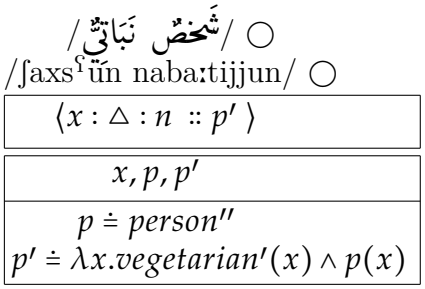


Figure 6.5: Merge of the noun /ʃaxsˤun / and the adjective /naba:tijjun/

Properties are divided into two types: individual level like “man” and stage level like “president” (Kracht, 2016, p. 180-181). In the definition of the AIS, a parameter is supposed to consist of a pair $\langle \rho, y \rangle$, where ρ is the name of the parameter and y is a variable, but so far only variables are used because only one parameter, i.e., properties, is dealt with. However, from now, both parameter names and variables will be used, since more than one parameter will be included. That is, noun phrases are considered properties in a specific time. The attribute PROP is used for the former and the attribute PRED is used for the latter as shown in figure (6.6).

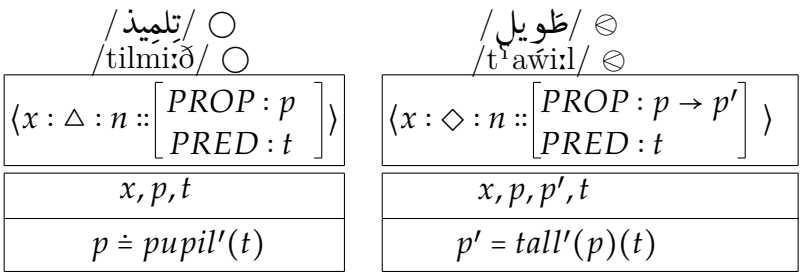


Figure 6.6: Representation of the noun /tilmiːð/ and the adjective /tˤawiːl/

Figure (6.6) shows that the attribute value concept used for names is also used for parameters. That is, in the case of names, attributes like NUM for number and PERS for person are used, and likewise, in the case of parameters, PROP is used for properties and PRED for prediction time. Parameters are handled using the so-called parameter handling statement, or PHS, which is defined, given P is a set of parameter names, as:

Definition 23. A parameter handling statement over P is either a partial function from P and values from the set of referents; this type is called simplex. Or it is a partial function from P with values being pairs of referents. This type is called duplex (Kracht, 2016, p. 181-182).

6.2 A three-parameter theory of time

The concept of time in natural language has been the focus of a large amount of research in a variety of disciplines, and there have been several attempts to explicitly and precisely formalize the relation between time and the use of tense in language (W. Klein, 1994; Comrie, 1985, 1976; Declerck, 1995, 1991, 1986; Reichenbach, 1947). These attempts can be dated back to Aristotle; however, the work of Reichenbach (1947) is considered a turning point in the way of describing this relation. In his influential work, he has suggested three points of time that should be considered:

- Point of speech, defined as the "time point of the token" and referred to with the initial S.
- Point of event, that specifies the time of the predicate in the sentence and referred to with the initial E.

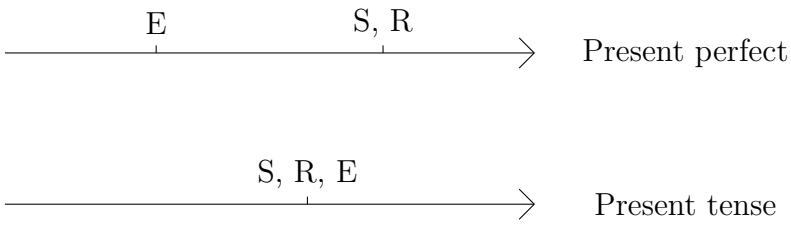
- Point of reference, referred to with the initial R.

The point of reference varies from one tense to another, but it is relevant to all tenses. The example in (99), taken from (Reichenbach, 1947, p. 288), shows how the three points are specified on the time axis (see also (Declerck, 1986)).

(99) Peter had left.

The point of the speech is when the token is produced, the point of the event is when Peter went, and the point of reference is in between (however is determined by the context). In this example, it is clear that the time of event is before the time of reference, and both are before the time of speech, as shown at the time axis:

This applies to other tenses as well. However, some time points can be simultaneous. That is, in the present perfect, the point of speech and the point of reference are simultaneous, and both are after the point of event. In the present tense, the three time points are simultaneous. This can also be described using the time axis:



Reichenbach's system has been criticized for several reasons, most importantly, for the over-generation of possible tenses, for providing only one reference point for all tenses, even complicated ones which require more than one reference point and finally for its inadequacy to deal with the so-called "during the preceding week" time (Declerck, 1986). Comrie (1985) has tried to avoid such limitations and to develop a general theory of tense. He defines tense as "the grammaticalization of location in time", i.e., the grammaticalized expressions of when a situation is located in time. This location in

time is relative to a deictic center that can be the present moment or a reference point that can be understood from the context. If the present moment, i.e., now, is taken to be the deictic center, the grammatical expression of time reference is referred to as "absolute tense", and this category includes past, present and future. If "now" is the deictic center, situations are located before (anterior) in the past tense, simultaneous in the present tense or after (posterior) in the future:

- Present tense: E simul S, e.g. I do my homework.
- Past tense: E before S, e.g. I did my homework.
- Future tense: E after S, e.g. I will do my homework.

Comrie has completely disregarded the point of reference in describing the absolute tenses and considered only two points: the point of speech and the point of event. On the other hand, relative tenses are described relative to a point of reference given by the context. This reference point can be the present moment, but the difference between absolute and relative lies in the obligatory specification of the present time as a reference point in the former and the optionality in the latter. Non-finite verb forms in English are examples of relative tenses that have relative time reference.
A further distinction has been made between "pure absolute tenses", in which the point of situation is directly described relative to the point of speech as in the past tense, and "absolute-relative tenses", in which the point of situation is related to the point of speech via one or more intermediate reference times as in the past perfect (Declerck, 1995). Absolute-relative tenses include past perfect, future perfect, future in future and future in past. They are described as follows:

- Past perfect: E before R before S, e.g. Before he joined the university, he had become a millionaire.
- Future perfect: E before R after S, e.g. By this time next year, I will have completed my PhD.
- Future in the future: E after R after S, e.g. At midnight, I will be traveling to USA.

- Future in the past: E after R before S, e.g. I did not imagine, he would get a position at Google.

Two observations should be noted:

(i) Comrie's analysis does not deal with continuous forms as separate tenses.

(ii) It does not distinguish between past tense and present perfect.

Both observations are justified in the light of Comrie's basic assumption of tense as only "location in time". That is, the difference between "John plays" and "John is playing" is not one of tense because in both sentences the situation holds at the present time, i.e., both have absolute present tense, but rather is one of aspect which is a different way of "viewing the internal temporal constituency of a situation" (Comrie, 1976, p. 3). The present tense form represents the situation as "a single unanalyzable whole", whereas the present continuous form makes an explicit reference to "the internal temporal structure". On the other hand, Comrie has not distinguished between past tense and present perfect except for "current relevance" and not in terms of location in time. That is, the term "perfect" refers to a past situation with a present relevance as in "His arm has been broken." (Comrie, 1976, p. 12).

In terms of the way of viewing the internal temporal constituency of the situation, Comrie distinguishes between perfective as a way of looking at the situation from outside without distinguishing the internal structure, and imperfective as a way of looking at the situation from inside. Similar to tense, aspect, as a functional category, can be expressed by different means, including inflectional morphology or periphrasis. A basic difference between tense and aspect in Comrie's theory is the existence versus the non-existence of a relation to a deictic center. He states that aspect is non-deictic, i.e., is completely independent of any relation to any other point time (Comrie, 1985, p. 14). His theory has been reconsidered by Klein (1994), particularly for two points: the temporal relation expressed by tense and whether aspect also expresses a temporal relation or not. Klein, whose terminology and implementation is adopted in this work, has proposed three parameters (W. Klein, 1994, p. 3-9):

- Time of topic (TT), the time for which an assertion or a claim is made.

- Time of situation (TSit), the time of the predication.
- Time of utterance (TU), the time at which the utterance is made.

The example in (100), taken from (W. Klein, 1994), illustrates what is meant by each of these tree parameters.

(100) The light was on.

The situation here is the being-on of the light, and its time (TSit) is the time when the light was on. A distinction is made here between this time of situation and the time for which the sentence "the light was on" is made (TT). The time when the sentence is uttered is the time of utterance (TU). It is clear that the TT is before the TU; however this does not imply that the light was still on at the TU. Klein argues that the time of situation may endure beyond the specific time span described by the sentence. This means that the being on of light can continue after the particular time span for which the claim is made (say, for example, when I entered the room). The three times are thought of as intervals rather than just points. This entails relations like inclusion and exclusion rather than just anteriority, posteriority or simultaneity.
Two distinct relations are expressed by these three parameters; one is expressed by tense and the second is expressed by aspect:

- Tense expresses the relation between TT and TU.
- Aspect expresses the relation between TT and TSit.

This means that while aspect, in Comrie's analysis, is non-deictic that does not provide any temporal relation, it is a way of relating the time of situation to the time of topic in Klein's analysis. There are three possible cases (W. Klein, 1994, p. 99-100):

- Tsit is fully including TT (TT INCL TSit).
- Tsit is partially including TT (TT AT TSit). This can be PARTIALLY BEFORE or PARTIALLY AFTER.
- Tsit is excluding TT (TT EX TSit). This can be BEFORE or AFTER.

If intervals are reduced to points, there will be a slight change in the description of the relations, as done, for example, by Kracht in the description of Latin (Kracht, 2016). That is, the three basic relations are identical (instead of INCL and AT), before and after (instead of EX). This also applies to tense, but the relation is between TT and TU. Thus, given t and t' are two time points, then the three basic relations between t and t' are:

- t is identical to t' ($t = t'$)
- t is before t' ($t < t'$)
- t is after t' ($t > t'$)

Three relations are expressed by tense, which correspond to the three well-known tenses: past, present and future. The relation are described as:

- Past tense: Time of topic is before time of utterance ($TT < TU$).
- Present tense: Time of topic is identical to time of utterance ($TT = TU$).
- Future tense: Time of topic is after time of utterance ($TT > TU$).

Also, there are three relations expressed by aspect that correspond to the three aspects: perfective, imperfective and prospective.

- Perfective: Time of situation is before time of topic ($TSit < TT$).
- Imperfective/Ongoing: Time of situation is identical to time of topic ($TSit = TT$).
- Prospective/Futurate: Time of situation is after time of topic ($TSit > TT$).

However, this reduction results in losing some important details such as the distinction between perfect and perfective aspects. I will keep this reduction in the representation of tenses and consider the inclusion relation in the representation of aspects. If the perfect aspect is considered, there will be twelve possible cases instead of nine (see (Bohnemeyer, 2003)).

Tense	Aspect			
	Perfective	imperfective	Perfect	Prospective
Past	$TT < TU; TT \supseteq TSit$	$TT < TU; TT \subset TSit$	$TT < TU; TT > TSit$	$TT < TU; TT < TSit$
Present	$TT = TU; TT \supseteq TSit$	$TT = TU; TT \subset TSit$	$TT = TU; TT > TSit$	$TT = TU; TT < TSit$
Future	$TT > TU; TT \supseteq TSit$	$TT > TU; TT \subset TSit$	$TT > TU; TT > TSit$	$TT > TU; TT < TSit$

Table 6.1: Combination of tense and aspectual relations

6.3 Temporal relations in Arabic

There is no identical way in which all languages in the world express the temporal relations, but they considerably vary in the means of specifying such relations, including inflectional endings, stem change, periphrastic constructions and suppletive forms (W. Klein, 1994, p. 123). Unfortunately, the literature on temporality in Arabic shows a high degree of inconsistency not only in the adopted approaches and findings but also in the usage of terms. Also, some phenomena have received little attention like time in noun phrases and the so-called sequence-of-tense for which the use of parameters is significantly important. Questions like "How does Arabic express temporality?" and "How can the meaning of Arabic morphosyntactic markers and expressions of temporality be represented?" should clearly be answered. I investigate how Arabic expresses temporal relations in terms of the relation between TT and TU on the one hand, and between TSit and TT on the other. Arabic is very well-known for the nature of its derivational morphology. Derived words also inflect for different grammatical categories depending on their type. This derivational nature has caused a debate on how Arabic expresses temporality and on whether Arabic expresses tense or aspect. In the following, I try to follow the origins of these problems, discuss the different assumptions and suggest a consistent treatment of Arabic temporal expressions both morphsyntactically and semantically.

6.3.1 First insights of classical Arabic grammarians

The first insights of the classical Arabic grammarians show that they were not mainly concerned with temporality but accidentally dealt with it in their classification of word types[2]. They tended to use descriptive language and some examples rather than specific terms. A first remark dates back to Si:bawajh (765 - 796) in his classification of word types. He said that the verb, as a word type, takes three different forms, a form for elapsed events, like /kataba/, a form for ongoing events, like /jaktubu/ and a form for prospective events, like /uktub/. Taking the general context of the remark into consideration, I argue that temporality was not the main concern of Si:bawajh's sub-categorization of verb forms. He was basically concerned with what makes the verb distinct from the noun, particularly in terms of the relation of each of them to time. That is, the noun has a meaning in itself and time is not part of it, whereas the verb has a meaning in itself and time is part of it.

In his definition of the first verb form, he used the word "elapsed" only, which denotes an event that happened in the past. This applies to the third form, which refers to an event that has not occurred yet (or will happen in the future). Only in defining the second form, he used two parameters: is being and has not elapsed. It seems that Si:bawajh was aware that the past form solely indicates an elapsed event both relative to the time of speech or any other time point (span) because the completion and the non-completion of past events are not determined in Arabic by the verb form in isolation. However, as mentioned before, Si:bawajh was only concerned with the morphological form of the verb, not with its semantics when combined with other words, say /ka:na/ /كَانَ/ 'was' or /ka:na qad/ /كَانَ قَد/ 'was already'. This is also the case of the imperative form. This assumption can be supported by the examples, given by Si:bawajh, that show the ungrammaticality of using past forms with future or present adverbs like /ɣadan/

[2]I cited the opinions of Arabic classical grammarians from (Gād, 2017).

/غَدًا/ ‘tomorrow’ or /aljawma/ /اَليَوم/ ‘today’ (M. Bahloul, 2007, p. 41) as in (101). On the contrary, /jaktubu/ denotes both the time of the event (the present) and gives the meaning of progressiveness.

(101) *أَتَاكَ غَدًا

ʔatːka ɣadan
come.3.sg.m.pst+you.acc.2.sg.m tomorrow
‘*He came to you tomorrow.’

The following grammarians have mostly followed the steps of Siːbawajh in his tripartite classification of verb forms. ʔibnu ssarraːʤ (? - 929), for example, has differentiated between three tenses denoted by the three word forms, namely, past, present and future. The same classification has been adopted by almubarrid (825 - 899) but he referred to the possibility of using the present form to denote both the present and the future. ʔibnu ʤinni (? - 1002) has exceeded the mere focus on the single verb form and mentioned how the verb combines withe prefix /sa/ /سَ/ ‘will’ to denote the absolute future, since the present form can denote both present and future.

The analysis of azzaʤʤːʤiː (855 - 923) was more informative. Although he has adopted a bipartite classification into past and future, he has used for the first time the concept of time points to describe the verb forms. He defines the past as an event that elapsed and at least two times exceeded it: the time in which it took place and the time in which it has been informed about. The future has not occurred yet and no time has exceeded it. Even though azzaʤʤːʤiː has not considered the present a distinct tense, he deals with it from the same perspective. That is the “verb of current status” is what is being during the speaker’s utterance. It is neither elapsed nor expected to occur in a later time. This analysis is very similar to Comrie’s analysis of the so-called absolute tenses, since each time of event is described relative to the time of the speaker’s utterance. Or even more accurately, except for the present, the two analyses are identical regarding the past and the future.

I conclude this part with a piece of evidence for my argument that temporality was not the main motivation behind the classification of verb forms into present, past and

future. The term mudˤaːriʕ(مُضَارِع), which is translated to 'present' in the modern literature, was used for forms like /jaktubu/ because Arabic grammarians have considered these forms similar to "nouns", since they inflect for "mood" and nouns inflect for "case". The first attempts to analyze temporality in the modern Arabic linguistics have largely been influenced by this traditional classification. This is why, many attempts have tried to classify Arabic as either a tense-oriented or an aspect-oriented language, focusing only on the verb forms as the grammatical means of expressing temporality.

6.3.2 The dilemma of tense aspect classification

A main reason for this disagreement has arisen from the assumption that each verb form expresses one and only one side of temporality- either the time of the situation relative to the current moment, and in this case it expresses tense, or it specifies the completion versus the noncompletion of the situation or the inside versus the outside way of viewing the internal temporal constituency, and consequently it is aspect-oriented. Comrie (1976, p. 78-82) and Fassi (2012) have argued that Arabic temporal expressions denote a tense/aspect combination, or what they referred to as "combined tense/aspect oppositions" and "polyfunctionality of tense/aspect-forms" respectively. To the contrary of the tense/aspect combination, Kuryłowicz (1973) claims that Arabic is neither aspect-oriented like "Slavic" nor tense-oriented like "Romance". In addition to the two-side and the neither-nor assumptions, there are two other one-side points of view: the tense-oriented and the aspect-oriented. The four assumptions are discussed in this part (see also, (M. Bahloul, 2007, p. 37-42).

- Arabic is an aspect-oriented language

  Much of the modern literature on temporality in Arabic starts from Wright's analysis of time in Arabic (Wright & Caspari, 1896), in which he considered two temporal forms of verbs. He has claimed that the two forms denote the (state) of the action rather than its tense. That is, each form determines whether an act is finished or unfinished relative to other acts. Therefore, he rejected using the terms "preterite" and "future" and suggested using the "perfect" vs "imperfect" instead of the tripartite classification (Wright & Caspari, 1896, VI p 51). Wright

has supposed that Arabic grammarians have not succeeded in connecting the two verb forms to time. Instead, he has proposed a bipartite classification in terms of the action being finished (perfect) or unfinished (imperfect). It is worth noting that Wright has exceeded the mere verb form and paid attention to the structures that express temporality. This means that the perfect and imperfect are considered two umbrella terms under which there is a number of structures, and each structure has its own semantics. Several scholars have adopted this analysis and described actions as either completed (finished) or non-completed (unfinished), but the term "aspect" started to appear instead of Wright's term "state".

- Arabic is a tense-oriented language

  Some modern Arabists and grammarians have followed the traditional approach in describing verb forms as expressing tense. The proponents of this approach have not introduced a significant contribution to what has already been introduced by the traditional Arabic grammarians neither terminologically nor analytically.

- Combined tense/aspect oppositions

  After differentiating between the meanings of tense and aspect, Camire (1976, p. 78-82) claimed that the difference between the two Arabic forms can be neither one of aspect nor one of tense. He argued that the debate of tense/apsect classification arises if the two forms are considered in isolation (without temporal adverbs), as in (102) and (103); but if a specific time reference is considered, as in (104) and (105), it becomes clear that Arabic expresses "combined tense/aspect oppositions".

(102) جَلَسُوا عَلَى البَابِ

ʤalasuː ʕala albaːbi
sit.3.pl.m.pst at door.gen.sg.m.def
'They sat at the door.'

(103) اَللَّهُ يَعلَمُ مَا تَفعَلُونَ

Allahu jaʕlamu ma: tafʕalu:na
Allah.nom.sg.def know.3.sg.ind.prs what.acc do.2.pl.ind.prs
'Allah knows what you do.'

(104) آجِيئُكَ إِذَا احمَرَّ البُسرُ

a:ʤi:ʔuka ʔiða ʔiħmarra albusru.
Come.1.ind.prs+you.acc if become-red.3.sg.ind.prs dates.nom.def
'I came to you if the dates become red.'

(105) اَللَّهُ يَحكُمُ بَينَكُم يَومَ القِيَامَةِ

Allahu jaʕħkumu bajnakum jawma alqija:mati
Allah.nom judge3.ind.prs between+you.gen day.acc judgement.gen
'Allah will judge between you on the Day of Judgement.'

Similarly, Fassi (2012, p. 93-118) argues that Arabic verb forms are polyfunctional. Again, Fassi compares the semantics of the verb forms in isolation and when accompanied by auxiliaries or other temporal adverbs. In case of isolation, the two verb forms express the semantics of "past" and "present" as in (106) and (107); but when embedded under auxiliaries, they function as perfect and imperfect 'participles' as in (108) and (109) Fassi (2012, p. 95).

(106) كَتَبَ رِسَالَةً

kataba risa:latan
write.3.sg.pst message.acc.sg.f
'He wrote a message.'

(107) يَكتُبُ رِسَالَةً

jaktubu risa:latan
write.3.sg.ind.prs message.acc.sg.f
'He is writing a message.'

(108) كَانَ كَتَبَ رِسَالَةً

ka:na kataba risa:latan
be.3.sg.m.pst write.3.sg.pst message.acc.sg.f
'He had written a message.'

(109) كَانَ يَكتُبُ رِسَالَةً

ka:na jaktubu risa:latan
be.3.sg.m.pst write.3.sg.ind.prs message.acc.sg.f
'He was writing a message.'

- Arabic is neither tense-oriented nor aspect-oriented

  Comparing Arabic to Slavic languages on the one hand, and to Romance languages from another hand, Kurylowicz (1973) (depending on (M. Bahloul, 2007, p. 42)) claims that Arabic verb forms do not denote a particular time reference nor are they aspectual. Instead, the two forms express an opposition of "anteriority" and "simultaneity".

Both the tense-oriented assumption and the neither-nor one are affected by an extreme view adopted by some authors who restrict the notion of aspect to the morphological opposition between "perfective" and "imperfective" in Slavic languages. However, the more liberal perspective claims that other languages can express the same category but in different ways (W. Klein, 1994, p. 16). I disregard these two extreme assumptions.
Now, given Arabic is an aspect-oriented language, i.e., verb forms only express a completed or a non-completed situation, the question that arises is: "Relative to when is it completed or finished?". The difference between (110) and (111) in English is not of tense, since both are in the past (if we take the time of utterance as the reference point).

(110) John hit his sister.

(111) John was hitting his sister.

But if (110) and (111) are used as subordinates, as in (112) and (113) respectively, the time of utterance cannot be taken as the reference time.

(112) When I was in my room, John hit his sister.

(113) When I was in my room, John was hitting his sister.

This simply means that the temporal expressions establish two distinct relations: one with the time of utterance and one with the time to which the expression is produced (time of topic in Klein's notions). This applies to (114) and (115) when used as main sentences or subordinate ones, and consequently the one-side relation is not accepted.

(114) عَلِيٌّ كَتَبَ رِسَالَةً

ʕalijjun kataba risaːlatan
Ali.nom.sg.m write.3.sg.pst message.acc.sg.f
'Ali wrote a message.'

(115) كَانَ عَلِيٌّ يَكتُبُ رِسَالَةً

kaːna ʕalijjun jaktubu risaːlatan
be.3.sg.m.pst Ali.nom.sg.m write.3.sg.ind.prs message.acc.sg.f
'Ali was writing a message.'

The last (and the only valid) assumption is the one adopted by Comrie and Fassi, i.e., Arabic expresses both tense and aspect. However, Comrie's analysis itself has been criticized and argued against (W. Klein, 1994; Declerck, 1991; Salkie, 1989). According to Comrie's analysis, "absolute tenses" are only described relative to the time of speech and the time of event, but if we consider (116) (taken from (Declerck, 1991)), we find that the time of event is located relative to a reference time (yesterday), which is in turn before the time of speech. This shows that Comrie's disregarding of the reference

time in the description of the so-called asolute tense is not justified (Declerck, 1991, p. 235).

(116) John was here some time yesterday.

I argue that the following three points should be considered together. (i) Temporality is expressed in Arabic by morphosyntactic expressions that exceed the mere verb form. (ii) Arabic temporal expressions denote both aspect and tense since the two grammatical categories denote two distinct relations. (ii) Taking into consideration the combination of tense and aspcetual relations, Arabic expresses the common combinations in European languages (French, German, Greek and Latin) (Kracht, 2016, p. 187)- either by the mere morphology or by morphosyntactic combinations.

6.3.3 Means of expressing tense and aspect

Another controversial issue is related to the means of grammatically expressing temporality in Arabic. Bahloul (2007, p. 29-32) counts different points of view including his own opinion.

- Tense/Aspect is an abstract morpheme that has no specific phonological realization. Benmamoun (2000, 1993) argues that the suffix /a/ in kataba is a pure agreement morpheme and the prefix /ja/ in /jaktubu/ is a default form of the verb.

- The first /a/ in the stem /katab/ denotes the completion of the situation, whereas the /u/ in /jaktub/ designates recurrence in the actual world (Er-rayyan, 1986, p. 75-122).

- The first /a/ in /katab/ expresses the tense/aspect category, whereas this category is expressed in /jaktub/ by the first /a/ and the absence of the first vowel in the stem (R. Bahloul, 1991).

- The first /a/ is the only marker of the tense and aspect category in both forms.

- Tense/Aspect is expressed by the suffix in /kataba/ and by the prefix in /jaktubu/.

Although Benmanoun's view seems convincing with respect to the suffix of /kataba/, it does not explain why there are two different verb stems. In line with this hypothesis, Hallman (2015) claims that the imperfective form in Arabic does not denote imperfectivity, but rather it marks the default form of the verb. However, the idea of a default form is also controversial because if we suppose that /jaktub/ is the default, it entails that /katab/ is derived from it, say by means of stem change, which needs to be proven. This also neglects the most dominant feature of Arabic morphology, in which the word form plays the central role in specifying its classification, function and semantics. If the two grammatical categories are expressed by abstract morphemes, then having two distinct stem forms is meaningless.

Moreover, Aoun et al. (2010, p. 22-23) argue for this assumption depending on their observation of the verbs /lajsa/ 'not' and /maː zaːla/ 'be still' claiming that both denote negation in the present whereas they host the suffixes of the past forms. This contradiction demonstrates that the vocalism does not express temporality. It is true that /lajsa/[3] appears on the surface like a past form, since it hosts past suffixes, and it is also true that it is used (mostly) to negate the present, but to the contrary of regular verbs, temporality is not part of /lajsa/. It is not devoted to only negate the present, but it negates the past (117), present (118), future (119) or is even used for the negation of generic sentences (120).

(117) أَليسَ قَد نَهَى ... عَن زِيَارَةِ القُبُورِ

ʔalajsa qad naha: ... ʕan zijaːrati lqubuːri
not already prevent.pst ... from visiting.gen graves.gen.def
'Has not ... forbidden visiting the graves'

(118) لَيسَ المُدِيرُ غَاضِبًا

lajsa l-mudiːru ɣaːdˤiban
not manager.nom.sg.m.def angry.acc.sg.m
'The manager is not angry.'

[3] /lajsa/ is a special word in Arabic because it can be used as a verb that only conjugates in the past form or as a negation particle. In the first case, it behaves like /kaːna/ and its sisters, i.e., it takes two arguments and assigns for the first a nominative case and for the second an accusative case. But when used as a negative particle, it only gives the meaning of negation. Thus, /lajsa/ is (mostly) a verb because it affects the case. That is, when it negates a copula-less sentence, it changes the nominative case of the predicate to accusative. It is also the locus of gender, number and person.

(119) يَومَ يَأتِيهِم لَيسَ مَصرُوفًا عَنهُم

jawma jaʔjiːhim lajsa masˁruːfan ʕanhhum
day.acc come.3.ind.prs+them.acc not averted.acc.sg from+them.gen
'On the Day it comes to them, it will not be averted from them'

(120) لَيسَ الفَقِيرُ فَقِيرَ المَالِ.

lajsa lfaqiːru faqiːra lmaːli
not poor.nom.sg.m.def poor.acc.sg.m money.gen.def
'The poor is not the poor in wealth.'

As for /maː zaːla/, it is composed of two separate words: the past form /zaːla/ and the negative particle /maː/. This verb hosts the suffixes of the past simply because it is the past form of the root /z⊗y⊗l/ [4], but its meaning is idiomatic. The case of /maː zaːl-a/ is very similar to the compositional meaning of the present form of the verb in Arabic with the negation particle /lam/ / لَم/ 'not'. No doubt that the verb form /janʤaħ/ /يَنجَح/ 'succeed/pass' denotes the present, but in a sentence like (121), it negates the past, regardless of the verb form or the suffixes it hosts. This meaning is a result of the combination, as is the case in /maː zaːla/.

(121) لَم يَنجَح فِي الاَمتَحَانِ

lam janʤaħ fiː limtaħaːni
not succeed.3.sg in exam.gen.sg.def
'He did not pass the exam.'

The other three assumptions can be discussed together because they are closely related to each other. That is, all of them propose that the first /a/ is the grammatical marker of the tense/aspect category in /kataba/, whereas they disagree on the grammatical marker in /jaktubu/. In fact, the three assumptions arise several questions. First, let's concentrate on /kataba/. If we suppose that the first /a/ expresses the tense/aspect

[4] Aoun et al. (2010, p. 23) claim that /zaːla/ is derived form the root /z⊗w⊗l/ which is incorrect, because there is a differenece between /zaːla/ /jazuːlu/.

category, then, what is the function of the second /a/? Is it a distinct morpheme? It is known that the second vowel can be /a/ as in /katab/ 'write', /i/ as in /laʕib/ 'play' or /u/ as in /ħasun/ 'became comely'; on what basis the second vowel is chosen?

Bahloul (2007, p. 33) claims that the choice is mostly determined by the verb's valence. He states that the second /a/ considerably specifies transitivity. He gives some examples to support his view of the function of the second /a/. It is true that many verbs of this form are transitive, but there are also a large number of intransitive verbs in the same form as /wahana/ 'became weak', /fasada/ 'became spoiled (corrupted)', /baraza/ 'jut/emerge', /saqaʕa/ [5].

Bahloul (2007, p. 32) uses passivization as another piece of evidence for treating the first vowel in /kataba/ as the morpheme which represents the aspect–tense category. He states that passivization is achieved by changing the two vowels of the stem into /u i/, i.e., the /a/ indicates the active voice and the /u/ the passive voice. He asserts that the /u/ is the only alternate to the /a/, and this alternation emphasizes the aspect–tense features of the verbal event. This argument can be argued against using the passive form of the hollow verbs. A verb like /qaːla/ 'he said' has no /u/ in its passive form. It is changed into /qiːla/ 'was said', and the same applies to /xaːfa/ 'he feared' and /xiːfa/ 'was feared'. This means that /u/ is not the only vowel used in the passive form. Therefore, the assumption of the first /a/ in /kataba/ as the only marker of the tense/aspect category is far from perfection.

Bahloul (2007, p. 35) claims that the same vowel /a/ in /jaktubu/ denotes the tense/aspect category, and the only difference is in its position as part of the prefix, rather than of the stem, i.e., outside the root. Again, he uses passivization as a piece of evidence for his argument because the vowel /a/ of the prefix is changed into /u/ in the passive. However, he himself fails to propose a function for the second vowel stating that "its function is not as straightforward and as easy to discern as the one for the perfect". This simply means, we have left with two distinct bound morphemes, one whose function is to express tense/aspect, and the second with no function.

Finally, the standard assumption has several problems, above all both the suffix of

[5]Bahloul's argument might statistically be true, but since he does not support his argument with concrete statistics, the counter examples cannot be regarded as exceptions or irregular forms.

/kataba/ and the prefix of /jaktubu/ have no fixed form. They have different forms to satisfy the agreement principle. In addition to the /j/ in /jaktubu/, there are three other forms: /ʔ/, /t/ and /n/, and all function as agreement prefixes and not as temporality morphemes. This is more problematic in the past form, because the suffix is completely dropped in the conjugation of the verb with the first person as in /katabtu/ and /katabnaa/.

I suggest that the tense/aspect category is expressed by the two vowels in the stem form. That is, a verb like /kataba/ is composed of three morphemes: two of them are discontinuous, namely, the root /k⊗t⊗b/ and the vocalism /a⊗a/, and the suffix /a/ as shown in figure (6.7). /jaktubu/ is composed of four morphemes: two discontinuous morphemes, i.e., the root /k⊗t⊗b/ and the vocalism /a⊗u/, the prefix /j/ and the suffix /u/, as shown in figure (6.7).

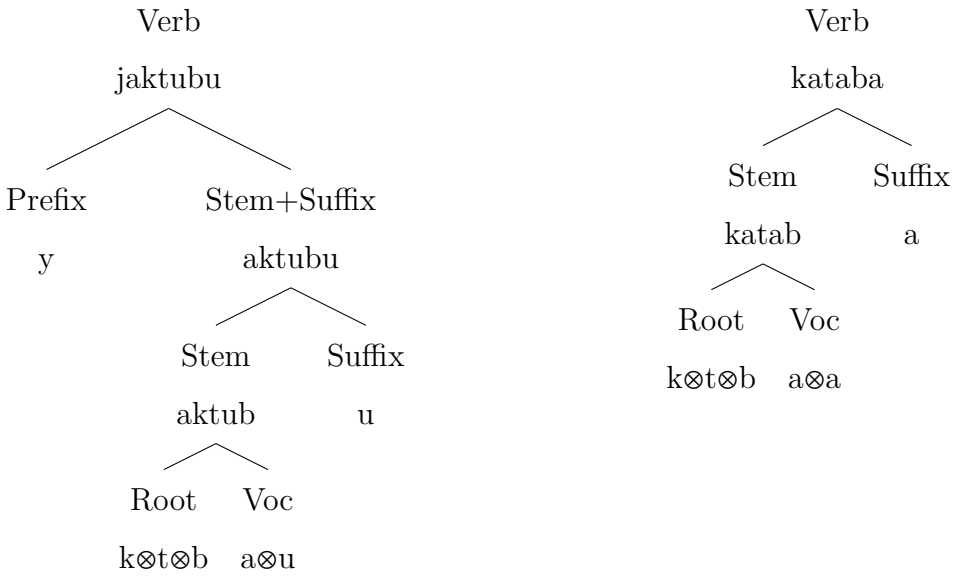


Figure 6.7: Composition of the two verb forms /kataba/ and /jaktubu/

The discontinuous morpheme /a⊗a/ gives the past meaning before receiving any prefixes or suffixes. This also applies to the present form. The discontinuous morpheme /a⊗u/ gives the meaning of time before the stem receives the suffix and the prefix. One question that might arise is why the /a/ in /jaktubu/ precedes the /k/, while it follows it in /kataba/. I assume this /a/ is not part of the prefix, it is part of the vocalism morpheme. It precedes the root morpheme for phonological reasons because Arabic does not allow

the sequence of two consonants without any vowel after any of them and the initial consonant of any word must be followed by a vowel.

There are various pieces of evidence for my proposal. First, this /a/ is fixed with all agreement prefixes. That is, as mentioned above, there are four allomorphs for the agreement prefix of the present stem: /ʔ/, /j/, /t/ and /n/, and there is an array of suffixes, but the /a/ does not change in all forms. Secondly, every time a change happens to the past tense vocalism morpheme, this happens in parallel to the present tense vocalism. Consider, for example, the conversion from active to passive in table (6.3.3) below.

Past		Present	
Active	Passive	Active	Passive
katab	kutib	aktub	uktab
laʕib	luʕib	alʕab	ulʕab
ʃarib	ʃurib	aʃrab	uʃrab

Table 6.2: Conversion from active to passive

A third piece of evidence comes from geminate verbs like /dalla/ /دَلَّ/ 'he guided', /jadullu/ /يَدُلُّ/ 'he guides'. In this group of verbs, not only /a/ changes its position, but also /u/, and again for the same phonological reason. That is, since each geminate verb contains a reduplicated consonant without any vowel in between, the /u/ is forced to change its position in order to prohibit the sequence of two consonants. This proposal is largely supported by the fact that the /u/ in the geminate verbs gets back its position once the reduplication is decomposed, say when the verb is followed by the feminine plural suffix /na/ as in /jadlulna/ /يَدلُلنَ/ 'they [female] guide'.

Table (6.3) shows a set of temporal relations in Arabic. Each row shows the form, the means of expressing the temporal relation, the meaning of the form and finally a description of the form.

Form	Means	Temporal relations	Description
katab	vocalim	$TT < TU;\ TT \supseteq TSit$	perfective past
akotub	vocalism	$TT = TU;\ TT \subset TSit$	imperfective present
sajakotub	porclictic + vocalism	$TT > TU;\ TT \supseteq TSit$	future
kana katab	aux + vocalism	$TT < TU;\ TT > TSit$	perfect past
kana jakotub	aux + vocalism	$TT < TU;\ TT \subset TSit$	imperfective past
kana sajakotub	aux + porclictic + vocalism	$TT < TU;\ TT < TSit$	prospective past
sayakunu katab	aux + porclictic + vocalism	$TT > TU;\ TT > TSit$	perfect future
sayakunu jakotub	aux + porclictic + vocalism	$TT > TU;\ TT \subset TSit$	imperfective future

Table 6.3: Temporal relations in Arabic

I show how markers install both tense and aspect. Consider, for example, the representation of the root /k⊗t⊗b/ in figure (6.8).

/ك ت ب/○
/k⊗t⊗b/○

$$\langle e : \triangle : \begin{bmatrix} CAT : verb \\ TENSE : * \\ ASP : * \end{bmatrix} :: \begin{bmatrix} \quad \end{bmatrix} \rangle$$

$$e$$

$$write'(e)$$

Figure 6.8: Representation of the root /k⊗t⊗b/

The past and present vocalisms are represented in (6.9) and (6.10) (taking into consideration that t_1 is the time of topic, t_2 is the time of situation and t_3 is the time of utterance).

/a⊗a/ ○

$$\langle e : \blacklozenge : \begin{bmatrix} CAT : verb \rightarrow \surd \\ TENSE : * \rightarrow past \\ ASP : * \rightarrow perf \end{bmatrix} :: \begin{bmatrix} TT : & t_1 \\ TSit : & t_2 \\ TU : & t_3 \end{bmatrix} \rangle$$

$$e, t_1, t_2, t_3$$

$$time'(e) = t_2; t_1 < t_3; t_1 \supseteq t_2$$

Figure 6.9: Representation of the past vocalisms

/a⊗u/ ○

$$\langle e : \blacklozenge : \begin{bmatrix} CAT : verb \rightarrow \surd \\ TENSE : * \rightarrow pres \\ ASP : * \rightarrow imperf \end{bmatrix} :: \begin{bmatrix} TT : & t_1 \\ TSit : & t_2 \\ TU : & t3 \end{bmatrix} \rangle$$

$$e, t_1, t_2, t_3$$

$$time(e) = t_2; t_1 = t_3; t_1 \subset t_2;$$

Figure 6.10: Representation of the present vocalisms

Compositionally, the meaning of the stem is obtained by a merge of the morphological units, i.e., root and vocalism, and their meanings. This is shown in the representation of the stems /katab/ and /aktub/ in figures (6.11) and (6.12).

/كَتَب/ ○

/katab/ ○

$$\langle e : \triangle : \begin{bmatrix} CAT : verb \\ TENSE : past \\ ASP : perf \end{bmatrix} :: \begin{bmatrix} TT : & t_1 \\ TSit : & t_2 \\ TU : & t_3 \end{bmatrix} \rangle$$

$$e, t_1, t_2, t_3$$

$$time'(e) = t_2; t_1 < t_3; t_1 \supseteq t_2$$

Figure 6.11: Representation of the stem /katab/

/كتُب/ ○

/aktub/ ○

$$\langle e : \triangle : \begin{bmatrix} CAT : verb \\ TENSE : pres \\ ASP : imperf \end{bmatrix} :: \begin{bmatrix} TT : & t_1 \\ TSit : & t_2 \\ TU : & t_3 \end{bmatrix} \rangle$$

$$e, t_1, t_2, t_3$$

$$time'(e) = t_2; t_1 = t_3; t_1 \subset t_2$$

Figure 6.12: Representation of the stem /aktub/

In complex tenses the auxiliary /kaːna/ /كَانَ/ 'was' is made sensitive to whether the stem form has a past or a present stem; whether $TU > TT$ or $TU = TT$ respectively. This means that the auxiliary will be considered a tense marker in complex tenses, whereas the aspect is provided by the stem form.

6.3.4 Time in the Arabic noun phrase

For a long time, temporality has been linked to verbs and research on the expression of temporality has only focused on notions related the verb, mainly tense, aspect and aktionarts (W. Klein, 2018). This means that for a long time, the temporal interpretation of the verb arguments has been considered dependent on the time of the main verb. However, Enç (1986) has proposed that the temporal interpretation of the noun phrase can be independent of the temporal interpretation of the main verb as in (122).

(122) Every fugitive is now in jail. (Enç, 1986, p. 409)

Attempting to capture the semantics of the tense using the formal theory of tense logic (Prior, 1967), (122) is represented as shown in (123).

(123) $\forall x[fugitive(x) \rightarrow in-jail(x)]$

But it is impossible to be fugitive and in jail at the same time. The intended meaning of the sentence is: the individuals, who were fugitive at some time in the past, are now in the jail. This means, tense morphemes cannot semantically be considered sentential operators, since the noun phrases can have their own temporal interpretations.
Kracht (2016, p. 190-195) claims that there are only two differences between verbs and noun phrases in terms of their dependency on time: noun phrases do not express this dependency morphologically, and they do not handle the parameters but only make use of one of them. That is, in a sentence like (124), the noun phrase /Praesident/ is a time dependent property that must hook itself onto a parameter.

(124) Der Praesident war in seiner Schulzeit ein schlechter Schueler. (Kracht, 2016, p. 191)

The.Def.Nom.Sg.M president.Nom.Sg.M. be.3.Sg+Pst in his.Gen.Sg+F school-time.Gen.Sg.F ein.Indef.Nom+Sg+M bad.Nom.Sg.M studen.Nom+Sg.

'In his school days the president was a bad student'.

He says that the /Praesident/ is a time dependent property. In (124), the situation is the being bad student of the president, and its time (TSit) coincides with the time of topic (president's school time). This means that the prediction time of the noun phrase is the time of topic. This idea becomes clearer if the adjective /heutige/ is used to modify the noun phrase:

(125) Der heutige Praesident war in seiner Schulzeit ein schlechter Schueler. (Kracht, 2016, p. 191)

The.Def.Nom.Sg.M present-day.Nom.Sg.M president.Nom.Sg.M. be.3.Sg+Pst in his.Gen.Sg+F school-time.Gen.Sg.F ein.Indef.Nom+Sg+M bad.Nom.Sg.M studen.Nom+Sg.

'In his school days the present-day president was a bad student.'

Using the adjective /heutige/ makes the time of utterance as the predication time of the noun phrase. The use of the adjective /damalige/ to enforce the time of topic gives a pragmatically odd sentence as illustrated in (126).

(126) ?Der damalige Praesident war in seiner Schulzeit ein schlechter Schueler. (Kracht, 2016, p. 191)

The.Def.Nom.Sg.M of-that-time.Nom.Sg.M president.Nom.Sg.M. be.3.Sg+Pst in his.Gen.Sg+F school-time.Gen.Sg.F ein.Indef.Nom+Sg+M bad.Nom.Sg.M studen.Nom+Sg.

'The president of that time was in his school days a bad student.'

Moreover, the predication time of the subject noun phrase can be the time of situation (TSit):

(127) Im Jahr 1953 hielt der Praesident eine große Rede. (Kracht, 2016, p. 192)
In-the.Def.Gen.Sg.N year.Gen.Sg.N 1953 hold.3Sg.Pst The.Def.Nom.Sg.M president.Nom.Sg.M a.Indef.Acc.Sg.F big.Acc.Sg.F speech.Acc.Sg.F

'In 1953 the president held a big speech'.

However, contrary to Kracht's claim that noun phrases do not express time morphologically, Nordlinger and Sadler (2004) have introduced evidence from a number of languages around the world that not only verbs morphologically inflect for tense, aspect and mood, but also NP constituents can have TAM markers. They provided a cross-linguistic comprehensive survey for the TAM marking on NP constituents. Using data from a range of languages, particularly from North and South America, they claim that the nominal TAM inflection provides specific temporal information to the NP independently of the temporal information provided by the inflection of the clausal head-verb.

In Tariana [6] around forty percent of nouns in texts are tense-inflected ((Nordlinger & Sadler, 2004) depending on (Aikhenvald, 2003)). Nouns in Tariana have four tense clitics: -miki-ri, for the past masculine singular, -miki-ru for the past feminine singular, -miki for the past plural and -pena for the future. Examples include: correio-miki-ri (post.office-PST-NF) 'old/former post office' and du-sa-do-miki-ru(3SG.NF-spouse-FEM-PST-FEM) 'his late spouse', wa-tfimari-pena (1PL-son.in.law- FUT) 'our future son-in-law' and pi-ya-dapana-pena (2SG-POSS-house-FUT) 'your future house' (Nordlinger & Sadler, 2004).

From a different perspective, Tonhauser (2005) rejects the assumption of Nordlinger and

[6] Tariana is an Arawak language and the only one spoken in the linguistic area of the Vaupes river basin in the territory of the Upper Rio Negro in northwestern Brazil (Aikhenvald, 2003).

Sadler that Paraguayan Guaran´[7] is inflected for tense but for aspect. I am not much concerned with this disagreement but with the provided evidence that noun phrases can express their time dependency morphologically and independently of the main verb. In the following three examples (taken from (Tonhauser, 2005)), the noun phrase has one of three cases: to be free of temporal markers (128), to be marked with the nominal suffix -kue (129) or with -r˜a (130).

(128) Kuehe a-hecha pa'i-pe.
Yesterday I-see priest-PE
'Yesterday I saw the priest.'

(129) Kuehe a-hecha pa'i-kue-pe.
Yesterday I-see priest-KUE-PE
Yesterday I saw the former priest.

(130) Kuehe a-hecha pa'i-r˜a-me.
Yesterday I-see priest-RA-PE
'Yesterday I saw the seminarist/future priest.'

In (128), the noun phrase pa'i 'priest' is free of temporal markers, which means that the referent of the noun has the property assigned by the noun at the time of the event (the speaker's seeing of the priest). This is not the case in (129) and (130), where the noun is marked with -kue and -r˜a, which means that the referent of the noun has the property assigned by the noun before and after the time of the event respectively.
Arabic can also provide a supporting evidence for the argument that noun phrases may express temporal relations morphologically. Arabic participles may function as nouns, adjectives, adverbs or even verb substitutes (Ryding, 2005, p. 102). Both (131) and (132) are of the same sentence type, i.e., copula-less sentences. This type of sentences is only allowed in the present since the copula appears in the past and future. Therefore, the two sentences are supposed to express the same temporal relations.

[7]Paraguayan Guaran´ is a Tupian language spoken in Paraguay, the Argentinean province of Corrientes and in several municipalities of the Brazilian state of Matto Grosso do Sul (Rendón, n.d.).

(131) عَلِيٌّ كَاتِبُ الرِّسَالَةِ

ʕalijjun ka:tibu rrisa:lati
Ali.nom.sg.m writer.nom.sg.m message.gen.sg.f
'Ali is the writer of the message.'

(132) عَلِيٌّ كَاتِبٌ الرِّسَالَةِ

ʕalijjun ka:tibun rrisa:lata
Ali.nom.sg.m writer.nom.sg.m message.acc.sg.f
'Ali is going to write the message.'

However, the time of writing is not the same in the two sentences. In (131), the writing time of the message is in the past, but it is in the future in (132). The active participle in (131) functions as a noun, and it composes with the following noun what is referred to as "idafa construction". When the active participle is used in the idafa construction, it typically denotes the past. But if it functions like a verb as in (132), it denotes the future. In both cases, it expresses time morphosyntactically. In (131), the active participle is suffixed with the tanwīn of the nominative case, and the noun /rrisa:lat/ 'message' is assigned a genitive case. But in (132) it is suffixed with dammah, and the noun /rrisa:lat/ 'message' is assigned an accusative case.

Kracht's idea is to account for these facts using parameters. That is, all elements are basically supposed to share the three time parameters- whether they make use of them, or they just pass them to other elements. In the noun phrase, the predication time of the noun is passed either to the time of topic or the time situation. Thus, a stage dependent property can be represented as shown in figure (6.13).

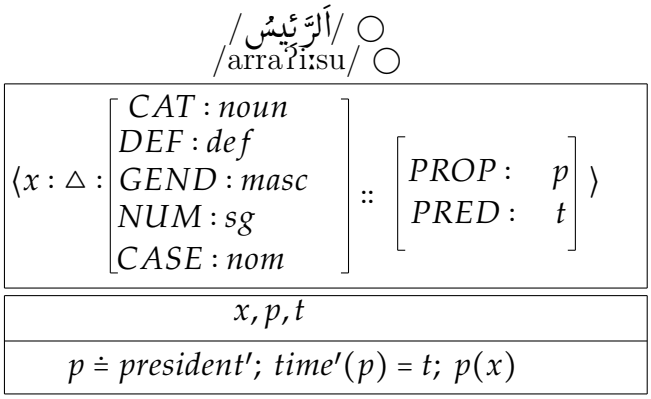


Figure 6.13: Representation of the noun phrase /arraʔiːsu/

Only one parameter, i.e., the time of situation, is used, and the other two parameters are neglected in this representation because they are not needed. However, in the semantic representation of an adjective like /alħaːliː/ /اَلحَالِي/ 'current' or /assaːbiqu/ /اَلسَّابِقُ/ 'former', more parameters are used as shown in their representations in figures (6.14) and (6.15).

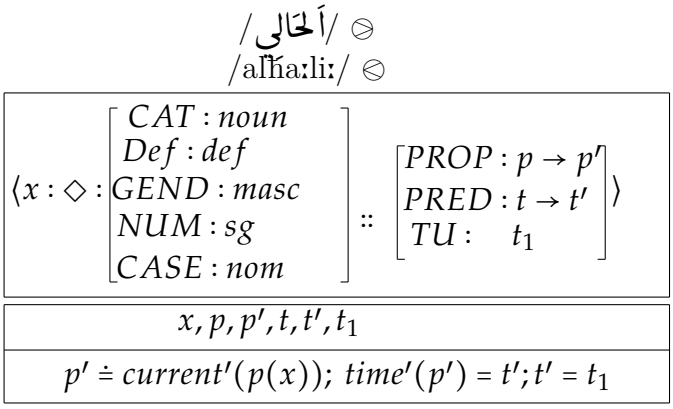


Figure 6.14: Representation of the adjective /alħaːliː/

/اَلسَّابِقُ/ ○
/assaːbiqu/ ○

$$\langle x : \Diamond : \begin{bmatrix} CAT : noun \\ Def : def \\ GEND : masc \\ NUM : sg \\ CASE : nom \end{bmatrix} :: \begin{bmatrix} PROP : p \rightarrow p' \\ PRED : t \rightarrow t' \\ TU : \quad t_1 \end{bmatrix} \rangle$$

$$x, p, p', t, t', t_1$$

$$p' \doteq former'(p(x));\ time'(p') = t';\ t' < t_1$$

Figure 6.15: Representation of the adjective /assaːbiqu/

The representation of the adjective /alḥaːliː/ shows that it takes the time of situation of the noun and maps it to its own time of situation ($t \rightarrow t'$), which equals the time of utterance ($t = t'$). The same happens in the DRS of the adjective /asaːbiqu/, but the time of situation of the adjective is before the time of utterance.

6.3.5 Tense in Arabic embedded clauses

Parameters are very significant for the phenomenon of sequence of tense (SOT), used as a basic argument by Enç (1987) against the treatment of tense as a sentential operator. That is, in a sentence like (133), there are two possible interpretations: either Mary's pregnancy was before John's hearing or both were simultaneous with each other. The first reading is referred to as the simultaneous interpretation, and the letter is called the back-shifted interpretation.

(133) John heard that Mary was pregnant. (Enç, 1987, p. 635)

This topic has attracted much attention in the last two decades, since languages behave differently in terms of the embedded clauses tenses. That is, while two readings are possible for (133), there are languages that do not allow the simultaneous reading in the past under past constructions; Japanese is an example. In the so-called non-sequence-of-tense languages like Japanese, past-under-past constructions yield only one interpretation, rather than two, namely, the back-shifted interpretation, as shown in (134) and (135) (taken from (Ogihara, 1996).

(134) Taroo-wa Hanako-ga ninsinsi-te i-ru to it-ta
Taro-TOP Hanako-NOM be-pregnant-PRES that say-PAST
'Taro said that Hanako was [simultaneous] pregnant.'

(135) Taroo-wa Hanako-ga ninsinsi-te i-ta to it-ta
Taro-TOP Hanako-NOM be-pregnant-PAST that say-PAST
'Taro said that Hanako was [back-shifted] pregnant.'

Russian is similar to Japanese, in which the simultaneous reading requires the embedded clause to be in the present as is clear in (136).

(136) Pjetja skazal, cto Misha placet. (Kracht, 2016, p. 214)
Pjetja said that Misha is crying
'Pjetja said that Misha was crying.'

Now, the question is whether Arabic is a SOT language or a non-SOT one. Let's consider (137).

(137) قَالَ عَلِيٌّ إِنَّ هِندًا كَانَت مَرِيضَةً

qaːla ʕalijjun ʔinna hindan kaːnat mariːdˤatan
say.3.sg.pst.m Ali.nom.sg.m that Hend.acc.sg.f be.3.sg.pst.f sick.acc.sg.f
'Ali said that Hend was sick.'

This Arabic sentence has only a back-shifted interpretation, and the simultaneous interpretation is not possible. The simultaneous reading requires the dropping of the copula in the embedded clause, which means that the embedded clause should be in the present as in (138).

(138) قَالَ عَلِيٌّ إِنَّ هِندًا مَرِيضَةٌ

qaːla ʕalijjun ʔinna hindan mariːdˤatun
say.3.sg.pst.m Ali.nom.sg.m that Hend.acc.sg.f sick.nom.sg.f
'Ali said that Hend is sick.'

This suggests that Arabic is classified as a non-SOT language like Japanese and Russian. Formally, a verb like /qaːla/ /قَالَ/ 'said' specifies the temporal relations of its complement. In (137), each verb has three time parameters, time of utterance, time of topic and time of situation respectively. The main verb is in the past, which means that the time of topic (t_1) is before the time of utterance (u_1); and in the perfective form, which means that the time of topic (t_1) includes the time of situation (s_1), i.e., ($t_1 < u_1$; $t_1 \supseteq s_1$). Also, the main verb decides the values of the values of the parameters of its complement /kaːa/ /كَانَ/ 'was' such that only the back-shifted reading is allowed.

6.4 Summary

In this chapter, I dealt with the fourth component of the AIS, i.e., parameters. They are a very significant feature and affect the meanings of sentences. There are several parameters, and AMRS allows to include any number of parameters. I included only three: properties, tense and aspect. Properties provide solutions for three problems: assigning meanings to inflectional morphemes, the semantics of the non-intersective adjectives and the lack of a direct mechanism for abstraction. Noun phrases are considered properties, and the adjectives are modifiers of these properties. Handling the temporal parameters in Arabic was very challenging due to the high degree of inconsistency in the literature on temporality in Arabic in the adopted approaches, findings and the usage of terms. I discussed the different assumptions of how Arabic expresses temporality and showed how to formally represent the meaning of Arabic morphosyntactic markers and expressions of temporality. Taking into consideration the combination of tense and aspectual relations, I argued that Arabic grammatically expresses the common com-

binations in European languages by vocalisms, auxiliaries and clitics. Both perfective past and imperfective present are only expressed by vocalisms. The other relations are expressed by combinations. The meanings of Arabic temporal forms are represented by using the notions of Klein's theory of time. Furthermore, I discussed two important, but largely neglected, issues, i.e., time in noun phrases and time in Arabic embedded clauses. I classified Arabic as a non-sequence-of-tense language Japanese, and therefore only the back-shifted reading should be allowed. I also showed that Arabic provides a supporting piece of evidence for the argument that noun phrases may express temporal relations morphologically.

Chapter 7

Implementation and Evaluation

In the theoretical part of this work, I argued that the proposed calculus is powerful enough to handle both the morphological and structural aspects of Arabic. In the morphology, constituents are allowed to be discontinuous. This is very important in order to deal with the nonconcatenative nature of Arabic. Also, although the semantic structure is independent of morphology, the morphosyntactic features determine whether two morphs can be combined or not. This allows to capture different word orders of sentences depending on the match between these morphosyntactic features rather than the order of the constituents as is the case in the original version of RS. As part of this dissertation, I developed a computational morphosemantic lexicon for (a fragment of) Arabic. This lexicon can be used within the computational implementation of the calculus developed by Marcus Kracht (in OCaml). The system can be downloaded at http://wwwhomes.uni-bielefeld.de/mkracht/referent/, where a detailed description of the system can be found (Kracht, 2016). The developed lexicon is attached in the appendix. While it is true that the applied calculus is designed to deal with the semantics in combination both with morphology and syntax, i.e., there are no borders between purely syntactic or purely morphological formation, I managed, for the sake of simplicity, to start with word formation and moved after that to the sentence level. After describing the developed morphosemantic lexicon, I discuss some limitations of AMRS both theoretically and practically.

7.1 Lexical units

Lexical units are morphemes not words. A list of roots is first created. Each lexical entry has a unique identifier followed by the argument structure as shown in snip of XML in figure (7.1)[1].

```
- <entry id="schreiben">
  - <argx>
    - <arg>
        <var name="e0"/>
        <dia name="u"/>
        <avi/>
        <avo voice="ast" mood="ast" tense="ast" asp="ast" trs="trans" cat="root"/>
      - <pavs>
          <vp name="gf1" out="x0"/>
          <vp name="gf2" out="x1"/>
        </pavs>
      </arg>
    </argx>
```

Figure 7.1: The unique identifier and the argument structure of the root /k⊗t⊗b/

The unique identifier of the root /k⊗t⊗b/ is “schreiben”, defined by <entry id = “schreiben”>. The argument structure is a quadruple: a variable defined by the attribute <var name = “e0”>, a diacritic defined by <dia name = “u”>. This means that the referent functions as an argument with respect to the merge operation, and this is denoted by the symbol △ in the output[2]. An attribute value matrix follows the specification of the diacritic. Since roots function only as arguments, their input attribute value matrix <avi> is empty, but the output attribute value matrix <avo> contains six attributes and their values. The first four attributes, i.e., voice, mood, tense and aspect, have an empty value “ast” denoted by ⋆. All of them will be decided later, not by the root itself. The last two attributes “trs”(for transitivity) and “cat” (for category) are provided with the two values “trans” and “root” respectively. Finally, there are two parameters: gf1 and gf2 for grammatical functions 1 and 2 respectively. The two functions assign the arguments of the verb or the subject and the object. The semantic roles of the two arguments are specified in the discourse representation structure.

[1]This only applies to verbs. Other categories, like nouns and adjectives, are inserted with their suffixes as one single entry for the sake of simplicity and to save time.

[2]Recall the set of vertical diacritics in section (3.3.1).

```
- <mor>
   - <exp>
      - <gst>
            <pre/>
            <text>k</text>
            <post/>
        </gst>
      - <gst>
            <pre/>
            <text>t</text>
            <post/>
        </gst>
      - <gst>
            <pre/>
            <text>b</text>
            <post/>
        </gst>
    </exp>
   - <marg>
      - <ma>
            <hdl idref="null3"/>
            <mi/>
            <mo stem-Eayon="ai" form="I" sub-type="sound" type="triliteral" morph="root"/>
        </ma>
    </marg>
```

Figure 7.2: Morphological component of the root /k⊗t⊗b/

Figure (7.2) shows the morphological unit or the exponent <exp> of the root /k⊗t⊗b/. Arabic roots are fractured glued string. The fractured glued string /k⊗t⊗b/ is composed of three glued strings <gst>: /k/, /t/ and /b/. Neither of the three glued strings specifies left or right conditions, since there are no phonological specifications for the combination process. This is why, both the left conditions <pre> and the right conditions <post> of the three glued strings are empty. Nevertheless, the morphological component contains a handler and a morphological class. The handler has the reference ID “null3”, defined by <hdl idref = “null3”>. The set of handlers is defined as shown in figure (7.3).

```
</hdl>
<hdl id="null3">
  - <unit>
        <prt fct="true" pos="0"/>
    </unit>
  - <unit>
        <prt fct="true" pos="1"/>
    </unit>
  - <unit>
        <prt fct="true" pos="2"/>
    </unit>
</hdl>
```

Figure 7.3: Specific handler for the sound roots

This is a specific handler for the sound roots. Similar to other specific handlers, each part has two attributes: “fct” which takes a boolean value and “pos” for the position of the string. The three boolean values are “true” because roots are the starting point of the merge process. When a root morph combines with some other vocalism morph, the handler of the latter controls the sequence of the strings of both of them.

In addition to the handler, the exponent has a morphological class, which in turn has two attribute value matrices: input class, denoted by “mi”, and output class, denoted by “mo”. The input class specifies the required properties of morph that can combine with the root. Again, since the root morph is always an argument in the merge process, its input class is empty. The output class specifies the properties of the root /k⊗t⊗b/. It contains four attributes and their values. According to these values, the root /k⊗t⊗b/ is triliteral , sound, in the first form and the second stem vowels are /a/ in the past and /u/ in the present. These are the individual properties of this root which specify its behavior under merge and therefore the vocalism with which it can merge.

The third component of the lexical unit is the DRS. The DRS has two components: the head and the body. The head, denoted by “head”, contains three variables: “e0”, “x0” and “x1”, and the body contains the set of conditions. The thematic roles of the subject and object are specified in the body of the DRS. That is, the subject (“x0”) functions as the agent of the verb as denoted by the first equation, and the object (“x1”) functions as the theme as denoted by the second equation. The semantic structure of the root /k⊗t⊗b/ is shown in figure (7.4).

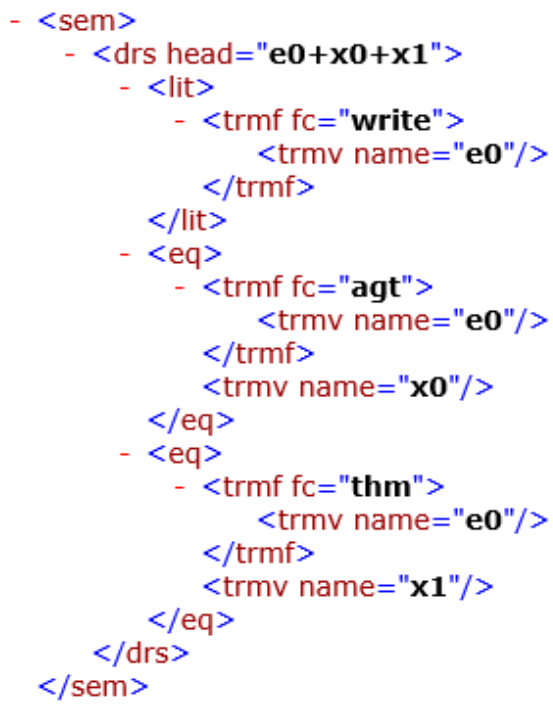


Figure 7.4: Semantic structure of the root /k⊗t⊗b/

Each lexical entry is constructed in the same way, but the attributes and their values differ from one entry to another. When two morphs are combined, the combination results in a new constituent. At the word level, the match between the output class of the root morph and the input class of the vocalism morph results in a stem. In case of the root /k⊗t⊗b/, it merges, for example, with the vocalism morph /a⊗a/ and the output is the stem /katab/ (as shown in the output of the stem /katab/ in figure (7.5)) which then concatenates with its corresponding prefixes and suffixes. The first line in the output is the identifier, followed by the stem and some morphological information. The first box in the output is the argument structure, the second is the semantics of the parsed string, i.e., a typical DRS with a head part and a body part, and the third box is the parse terms.

The argument structure of the stem exports an event variable, e. It is categorized as a stem in the perfective aspect and the past tense. It is also classified as transitive and in the active form. The mood attribute has no value. The AVS of the parameter has five parameters: two for the grammatical arguments, i.e., the agent and the theme, and the

three temporal parameters: time of topic, time of utterance and time of situation.

Welcome to Referent Systems Version 6.1.

Input: /katab/

(1)

Id:en21

✱.(○, 0)./k·a·t·a·b/⟨ [FORM : *I*; MORPH : *root*; STEM-EAYON : *ai* ; SUB-TYPE : *sound*; TYPE : *triliteral*] : 0 $^{0}_{\circ}$⟩

⟨e :△: [ASP : *perf* ; CAT : *stem* ; MOOD : * ; TENSE : *past* ; TRS : *trans* ; VOICE : *act*] :: [gf2 : x0 ; gf1 : x ; TU : t1 ; TT : t0 ; TSit : t] ⟩

e; t; t0; t1; x; x0

time′(e) = t; before′(t0, t1);
incIde′(t0, t); write′(e);
agt′(e) = x; thm′(e) = x0.

Parse terms:
✱. ◰$_f$ (p1trans : mr2, schreiben : mr0)

1

Figure 7.5: System's output of the stem /katab/

When the stem merges with the suffix /a/, for example, the complete form of the verb

gives one result for each directionality of the arguments as shown in the output of the verb /kataba/ in figure (7.6).

Welcome to Referent Systems Version 6.1.

Input: /kataba/

(1)

Id:en22

✱.(∘, 0)./k·a·t·a·b·a/⟨ FORM : *I*; MORPH : *root*; STEM-EAYON : *ai*; SUB-TYPE : *sound*; TYPE : *triliteral* : 0 $^{0}_{\circ}$⟩:⟨: 1 $^{0\,0}_{\bullet\circ}$⟩:⟨: 1 $^{0\,0}_{\circ\bullet}$⟩

⟨e :△: ASP : *perf*; CAT : *verb*; GEND : *masc*; MOOD : *; NUM : *sg*; PER : *3rd*; TENSE : *past*; TRS : *trans*; VOICE : *act* :: TU : t1; TT : t0; TSit : t⟩

⟨x : ▽ : CASE : *nom*; CAT : *noun*; GEND : *masc*; NUM : *sg*; PERS : *3rd* :: [ ]⟩

⟨x0 : ▽ : CASE : *acc*; CAT : *noun* :: [ ]⟩

e; t; t0; t1; x; x0

time′(e) = t; before′(t0, t1);
incIde′(t0, t); write′(e);
agt′(e) = x; thm′(e) = x0.

Parse terms:

✱. ◰$_f$ (atrasn : mr3, ◰$_f$(p1trans : mr2, schreiben : mr0))

1

Figure 7.6: System's output of the verb /kataba/

The argument structure of the verb includes three AISs: one for the event variable, one for the agent and one for the theme. The verb exports the event variable and imports both the agent and the theme. The agent is imported under the name nominative, noun, masculine, singular and third person, whereas the theme is imported under the name accusative noun. The parameter AVSs are ignored for the agent and the theme, whereas the parameter AVS of the verb includes three temporal parameters.

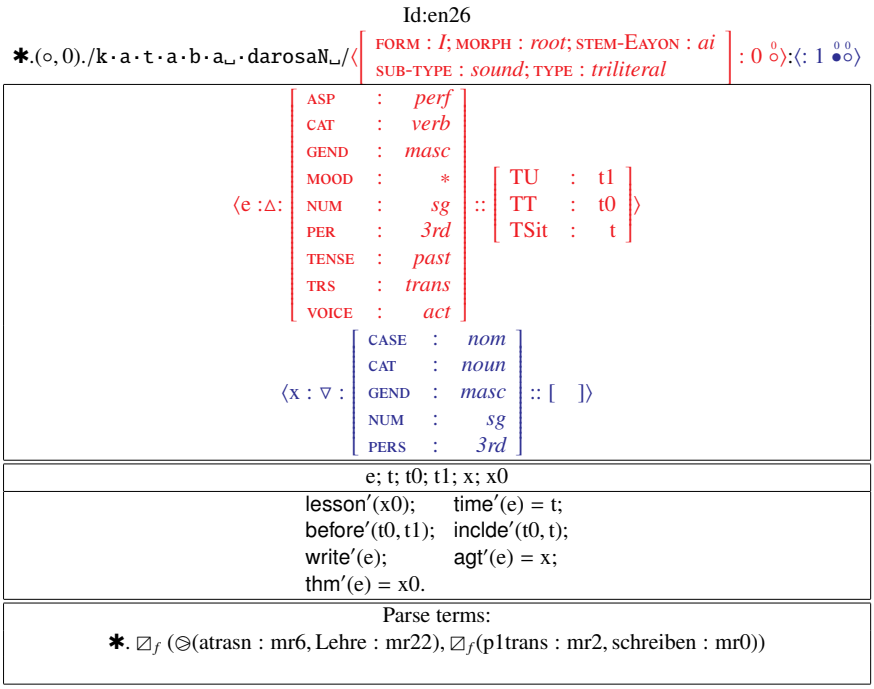
7.2 Phrases and simple clauses

Although the proposed calculus assumes that there is no syntactic structure and depends completely on morphology in the merge process until the whole sentence is constructed, I'd prefer to follow the common way of constructing sentences in explaining the computational implementation of the system. Also, I mention some comments on this point in the evaluation. Thus, we can imagine that words are already composed of their constituent morphemes, and they are ready to combine to form phrases. Theoretically, several variables can be identified under merges. That is to say, the verb /kataba/ can merge with the subject and the object together. But I prefer to first show the merge between the verb and one of its arguments, say with the object. This is shown in the output of the phrase /kataba darosaN/ in figure (7.7). Once the verb /kataba/ merges with the theme /darosaN/, the selection restrictions of the theme disappear in the resulting constituent.

Welcome to Referent Systems Version 6.1.

Input: /kataba darosaN /

(1)



1

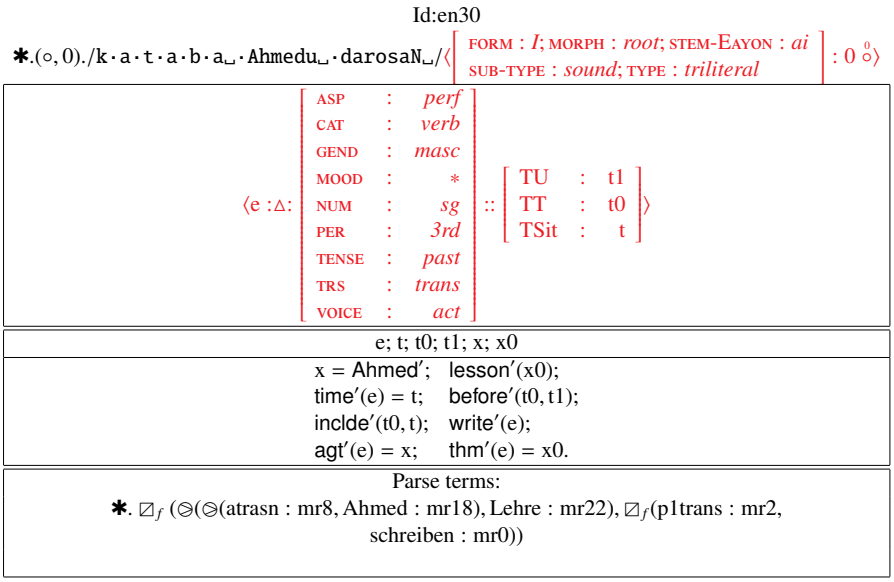
Figure 7.7: System's output of the phrase /kataba darsoaN/

A full simple sentence can then be obtained by the merge of the verb phrase /kataba darosaN/ with the subject /Ahmedu/ as shown in the output of the sentence /kataba Ahmedu darosaN/ in figure (7.8).

Welcome to Referent Systems Version 6.1.

Input: /kataba Ahmedu darosaN /

(1)



1

Figure 7.8: System's output of the sentence /kataba Ahmedu darsoaN/

Now, the argument structure contains only the AIS of the exported event variable. In the head part of the DRS, there are six variables for the event, the agent, the theme and the three temporal variables. In the body box, the computed semantics is represented in the form of functions and equations. The agent of the verb is x, represented as

$agt'(e) = x$ and $x = Ahmed$. The theme of the verb is x_0 and $x_0 = lesson$. The semantics of time is represented as follows: the time of topic, t_0, is before the time of utterance, t_1, which denotes the past tense; and the time of situation, t, is before the time of topic, t_0, which denotes the perfective aspect.

7.3 Evaluation

It is true that the proposed calculus provides a powerful mechanism to account for complex phenomena such as discontinuity of morphemes, flexible word order, null-subject and copula-less clauses in a tensed environment, but it meets also some challenges and there are questions that need to be answered.

7.3.1 Multi-word expressions

One of the most basic challenges of compositionality is multi-word expressions and to be more specific "non-decomposable multi-word expressions". The notion was suggested by Sag et al. (2002) as a response to the attempts to provide a semantic compositional analysis to some idioms (Nunberg (1994) as an example). The meanings of some MWEs can be predicted from the meanings of their parts and the way they are organized, even with specific interpretations of the meanings of these parts. "Spill the beans" is an example (Sag et al., 2002), which means to "remove the secrets". The meaning is obtained from "spill" in the "removal" sense and "beans" in the "secret" sense. However, there are many counter examples that cannot be described in the same way as in "kick the bucket". Kracht (2007) defines language as a set of signs, and each sign is composed of an expression and a meaning. The grammar of language is the finite set of constructions to form signs. A language L is described as a compositional language only if it has a compositional grammar, i.e., a finite set of operations to generate L. Kracht assumes that there is always some way in which the signs are composed and combined with each other. This way can be specified depending on the phonological constrains, as reflected in the definition of glued strings and fractured glued strings, or based on the morphosyntactic agreement. First, it seems obvious that phonology cannot play any role in this regard. Secondly, taking advantage of the morphosyntactic

agreement requires the possibility to predict the behavior of the signs under merge. However, there are MWEs that are syntactically-idiomatic whose construction cannot be predicted from its smaller components like “ad hoc”, “by and large” and “wine and dine”. In fact, MWEs represent a challenge to the compositionality principle in general not to Kracht’s theory in specific. Another syntactic problem is the variation. I argue that this framework can, from a theoretical viewpoint, deal with the decomposable MWEs, even in cases of syntactic variation by allowing the parts of the expression to be discontinuous. The non-decomposable MWEs can only be handled as a single unit with a specific meaning, which is simply against the compositionality principle. MWEs can be described as s cross-linguistic phenomenon, and therefore Arabic is not an exception. It has different kinds of MWEs, including compound nouns such as /ʤaːdu lħaqi/ / ‘Gadu alhaq’, idioms like /zawbaʕatun fiː finʤaːnin/ ‘a storm in a tea cup’ collocations such as /assijuːlatu lmuruːrijatu/ ‘Traffic fluidity’ … etc. Arabic MWEs are not dealt with in this dissertation, but the same challenges are expected for Arabic.

7.3.2 Discontinuity between Arabic morphemes and German separable prefixes

Kracht proposes that German verbs with separable prefixes are inserted in the lexicon as discontinuous roots as /auf⊗les/ ‘pick up’ while the perfect suffix morphs are as /ge⊗t/, /ge⊗en/, /t/ and /en/. This is the same solution proposed to Arabic roots and vocalisms as in /k⊗t⊗b/ and /a⊗a/. This is problematic from two sides: it ignores the fact that Arabic roots, even if they contain different parts, are single morphemes, whereas these German “roots” are not composed of different parts but from two distinct morphs. So, the Arabic root can be associated with one meaning, but one question is whether the meaning of /auf⊗les/, for example, is obtained compositionally from the meanings of /auf/ and /les/, or it is associated with its meaning noncompositionally. Also, the sequence of parts in the Arabic roots is always fixed, but the sequence of parts in the case of German verbs with separable prefixes is not fixed. German verbs that are composed of a stem and a separable prefix are different cases: they may maintain the main sequence of their parts as in the subordinate clauses (139) or exchange the order in

the main clauses (140) and also in the imperative (141). This in addition to the perfect form. The discontinuity in Arabic morphemes is purely morphological. The roots and vocalisms combine to compose the stem and nothing beyond that. In this point, the perfect form in German is similar, but in the main clauses and in the imperative, it goes beyond the morphology, since other constituents are inserted between the parts of the fractured string.

(139) ..., wenn der Zug abfährt.
..., when the.def train.nom.sg.m leave.3.sg.prs
'..., when the train leaves.'

(140) Ich mache die Tür auf.
I.nom.1.sg make.1.sg.prs the.def door.acc.f.sg on(open)
'I will open the door.'

(141) mach bitte die Tür auf.
make.2.sg.imp please the.def door.acc.f.sg on(open)
'Please open the door.'

A final note in this point is related to the strict prohibition of movement. Kracht argues that his framework does not allow any kind of movement, and he substitutes the movement rules with the handler. If /auf⊗les/ is dealt with as one morpheme, then it may appear at the surface either as /auf⊗les/ or as/les⊗auf/. Even using a combinatorial function instead of the movement rule, the real situation says that there is a kind of movement, and the difference is only in the way in which the movement is done.

7.3.3 Absence of diacritical markers

The overt morphological information plays a significant role in the identification of variables. In fact, the AISs are allowed to be accessed either in a strict order, i.e., E-accessed, which is not the case of Arabic, or to be skipped depending on the match of the morphosyntactic information, i.e., G-accessed. However, the writing system of modern standard Arabic poses real challenges to AMRS, both at the morphological and syntactic levels.

Diacritics and word formation: Morphologically, Arabic words are formed from roots and vocalisms. The latest are supposed to be written as diacritics above or below the consonants, but this is not the actual case. Diacritics are mostly absent in the written Arabic texts. Diacritical marks do not appear in 98% of Arabic texts (Habash, Diab, & Rambow, 2012). The absence of diacritics results in an exceptional degree of ambiguity in the writing system. Consider the string /ktb/ as actually written. This string may mean an array of words. It can mean /kataba/ 'he wrote', /kutaba/ 'was written', /kattaba/ 'made someone write' and /kutub/ 'books'.

Diacritics and sentence formation: Diacritics also play a significant role at the syntactic level. They are used among others as case markers. Arabic nouns inflect for three cases: nominative, accusative and genitive. In the normal cases, nouns are suffixed with a dammah in the nominative case as in /alkitabu/, 'the book', a fatha in the accusative case as in /alkitaba/, 'the book' and a kasrah in the genitive case as in /alkitabi/, 'the book'. It is noteworthy to mention that there are some other markers used to indicate these cases, and the same diacritic may function in some cases differently. The absence of these diacritics represents a real challenge for the proposed approach since it depends on them in specifying the syntactic functions and consequently the semantic roles of the arguments. Kracht's main argument against Vermeulen's original version of RS and at the same time his main motivation to extend it is its compositional inadequacy in the case of languages with free word order like Arabic. However, even Kracht's solution requires a fully diacritised sentence to be practically applicable. In a sentence like (142), neither word order nor the overt morphological marker can help identify the semantic roles of the arguments. That is, if the basic word order of the verbal sentence is followed, namely VSO, a misleading meaning, in which the employee functions as the agent whereas the manager functions as the patient, is obtained. But this is not the intended meaning. Unfortunately, depending on the overt morphological features, as Kracht proposes, is useless because the case markers are absent and the agreement between the verb and its subject is not sufficient since both the subject and the object have the same morphological features except for the case. Both are singular, masculine and definite. This means that both can syntactically function as the subject or the ob-

ject, but logically it is expected that the manager is the one who punished the employee since the opposite is difficult to be imagined.

(142) عَاقب الموظف المدير

ʕaːqb almwðˤðˤf almdiːr

punish.3.sg.pst.m employee.??.sg.m.def manager.??.sg.m.def

'The manager punished the employee.' or 'The employee punished the manager'

The two possible ways, in which the AISs are accessed, presuppose that arguments are connected to heads either by the order or by the marking, i.e., languages are either marking dependent, like many languages from the Indo-European family, including German, Greek, Armenian and the Slavonic languages, or zero-marking languages which typically have a very little morphology like Chinese and Vietnamese (Tallerman, 2014, p. 140). In the second group of languages, the order is crucial, but the writing system of modern standard Arabic represents a challenging case in which the relationship between the head and its arguments cannot be captured neither by the order nor by the markers.

Clearly, I do not say that Arabic is a zero-marking language or even the writing system of modern standard Arabic is fully neutral, despite the fact that Arabic dialects have lost case (Holes, 2004) and the fact that case is not usually marked in the orthography (Habash, Gabbard, Rambow, Kulick, & Marcus, 2007). In many cases, nouns are inflected for case using letters rather than diacritics, and therefore the absence of diacritics is insignificant for the identification of the syntactic functions. (143) is a modification of (142) but the verb arguments are in the plural form.

(143) عَاقب الموظفين المديرون

ʕaːqb almwðˤðˤfiːn almdiːruːn

punish.3.sg.pst.m employee.acc.pl.m.def manager.nom.pl.m.def

'The managers punished the employees.'

The syntactic functions of the verb arguments in (143) can be determined even in the absence of the diacritics because Arabic sound plural expresses case (also gender and number) orthographically. In (143), the verb can specify that /almdiːruːn/ /ٱلمديرون/ 'the managers' is the subject, and /almwðˤðˤfiːn/ /الموظفين/ 'employees' is the object depending on the distinction between the nominative marker and the accusative marker.

Riau Indonesian is a more complicated case because it cannot be classified in terms of the relationship between the verb and its arguments. The three common ways in which languages indicate this relationship are: constituent order, case marking and agreement (Tallerman, 2014, p. 189). However, Gil (2004) describes the language, spoken in informal situations by the inhabitants of the Riau province in east-central Sumatra, as a strongly isolating language with no inflectional morphology, relatively little derivational morphology and very flexible free word order. According to Gil, the most significant feature of Riau Indonesian is the pervasiveness of underspecification: grammatical categories like number, tense, aspect ... etc lack overt features. The expression in (144), taken from (Gil, 2004), shows the underspecification in Riau Indonesian.

(144) Makan ayam

eat chicken

'an association of eating and chicken'

In (144), /Makan/ 'eat' is underspecified for tense and aspect, and there is no overt marker for number or definiteness in /ayam/ 'chicken'. More important is the underspecification of the semantic roles of the two components of the expression, and consequently

the expression can be associated with an array of interpretations that can only be reduced depending on the actual context. Finally, the two components can exchange their positions, and the new expression maintains the same interpretations (Gil, 2004). This means that the relationship between the head and its arguments in Riau Indonesian does not follow any of the three above-mentioned ways.

Gil argues that Riau Indonesian is significantly different from European language and should not be described by the same theory of language. That is, there is no distinction between syntactic categories of noun, adjective, verb and sentence, nor between lexical categories and their phrasal projections. This simply means that it lacks any hierarchic syntactic structure. Also, the semantics of Riau Indonesian centers around the notion "association" rather than the predicate-argument relation. Moreover, from a semantic viewpoint, any expression can be interpreted as an event, state, time, place or thing, and the meaning of a construction XY, from the two expressions X and Y, can be associated to any interpretation from the interpretations of X and Y. The two orders of (144) can have the following 13 meanings if one, i.e., Makan, is interpreted as belonging to the category of things and the other, i.e., ayam, to the category of events (Yoder, 2010).

(145) Makan ayam/Makan ayam

(1) The chicken is eating.

(2) The chickens are eating.

(3) A chicken is eating.

(4) The chicken was eating.

(5) The chicken will be eating.

(6) The chicken eats.

(7) The chicken has eaten.

(8) Someone is eating the chicken.

(9) Someone is eating for the chicken.

(10) Someone is eating with the chicken.

(11) The chicken that is eating.

(12) Where the chicken is eating.

(13) When the chicken is eating.

This means that the free association of X and Y to any of the five categories results in 25 possible combinations, and for each combination there are 13 possible meanings. This generates hundreds of meanings for a two-word sentence, which is definitely unconvincing (Yoder, 2010).

7.3.4 Border removal and overlap of functions

AMRS depends on a basic principle, to which I refer as border removal. Simply, AMRS does not draw borders between the purely morphological and the purely syntactic formation. The language is simply a sequence of strings (glued or fractured), and these strings combine in some way. Let's focus on the verb and ask the following question: which morph imports the verb arguments? Is it the whole verb as a string or one of its components? Is it the root, the vocalism or the suffix? The lexical entry of the root /k⊗t⊗b/ (see figure 7.1) has two parameters for the subject and the object of the verb, and their semantic roles are specified in the DRS of the root. However, the proper treatment of the phenomenon of reduplicated arguments, or the so-called "left dislocation", requires an additional argument, i.e., a pronoun that agrees with its antecedent. I solved this problem by allowing the suffix to import three arguments rather than two as in the argument structure of the root. Although the proposed solution is practically effective, since the arguments are correctly imported with thematic roles, the theoretical justifications remain unclear. Questions like: Is it the function of the suffix to export the arguments? If so, what is the significance of the two parameters in the root entry? In case of simple structures, i.e., a verb and its two arguments, the root can import the subject and the object; however, in case of three arguments, one of the three arguments will not be imported. Two solutions can be suggested: the first is to insert two entries for each root, and the second is to insert two suffixes and to allow the suffix to import the arguments. Both solutions have their limitations. Having two entries for each root is theoretically unjustified in addition to the expected redundancy. On the other hand, if one of the two suffixes is allowed to import the arguments, this leads to an overlap in the functions of roots and suffixes.

```
- <argx>
  - <arg>
      <var name="e0"/>
      <dia name="u"/>
      <avi/>
      <avo voice="ast" mood="ast" tense="ast" asp="ast" trs="trans" cat="root"/>
    - <pavs>
        <vp name="gf1" out="x0"/>
        <vp name="gf2" out="x1"/>
      </pavs>
    </arg>
  </argx>
```

7.3.5 Multiplicity of Morphs

This point is closely related to the border removal and the overlap of functions. As explained, suffixes are allowed to import three arguments to account for the reduplication of one of the arguments, but this suffix morph cannot be used to account either for the normal case of transitive verbs, i.e., the case in which the verb has only two arguments for the subject and object, or for the ditransitive verbs. In the case of transitive verbs the number of arguments are different, and although in the case ditransitive verbs the number of the arguments is the same, there are two other problems. The first is that only two arguments have similar morphosyntactic information, but the third must be a pronoun in the case of the reduplicated argument, whereas it can be either a noun or a pronoun in the normal case of ditransitive verbs. Secondly, while in the case of reduplicated arguments a specific word order is forced starting by the antecedent either in the nominative or in the accusative case, followed by the verb connected to the anaphoric pronoun and finally the last argument, ditransitive verbs allows in the normal case for a more flexible word order. Unfortunately, this problem cannot be solved by adding multiple morphs to the same morpheme, i.e., the suffix, because the number of the imported variables and the morphosyntactic features are first specified in the argument structure, and the order is then specified in the morphology. This means that the only solution is to add multiple entries for the same suffix to account for the different possible cases. Practically, I chose this approach, but the question is whether these entries represent the same morpheme or they are different. Also, if this is the only way, it will be extremely difficult to consider all possible cases at the morphological and syntactic levels because the final result will be a huge number of entries for what is supposed to be the same morpheme.

7.3.6 Specifying the semantic role for the two kinds of objects

Ditransitive verbs like /give/ require three arguments: the agent, the theme and the recipient. The verb should be able to determine in some way how to specify these roles. Consider, for example, (146) and (147), taken from (Kracht, 2016, p. 94).

(146) They called him an idiot.

(147) He gave Albert the car.

The verb can distinguish the direct object from the indirect object depending on the order, since changing the order of the two object produces ungrammatical sentences (Kracht, 2016, p. 94).

(148) *They called an idiot him.

(149) *He gave the car Albert.

In German, the verb can also distinguish the accusative object from the dative object. In a sentence like (150), the verb can specify that /dem Kind/ 'the child' is the dative object and /das Spiel/ 'the toy' is the accusative object depending on the overt morphological markers.

(150) Der Vater gab dem Kind das Spielzeug.
The-DEF father-NOM+SG+M give-3SG+PST the-DEF child-DAT+SG+N the-DEF toy-ACC+SG+N

'The father gave the toy to the child.'

However, the issue is not simple as it may appear. Consider, for example, the Arabic equivalent of (150) in (151).

(151) أَعطَى الوَالِدُ الطِّفلَ اللُّعبَةَ
ʔaʕtˤaː alwaːlidu atˤtˤifla alluʕbata
give.3.sg.pst father.nom.sg.m.def child.acc.sg.m.def toy.acc.sg.f.def

'The father gave the toy to the child'.

The two objects have almost the same morphosyntactic information, and therefore they have two similar argument structures. Both are definite, singular and in the accusative case. They differ only in the gender. The child is masculine and the toy is feminine. Nevertheless, the gender does not play any role in this regard. Also, the two arguments can exchange their position. Thus, neither the order nor the overt morphological features can help to distinguish the direct object form the indirect one. This case is different from ambiguous cases, in which two possible interpretations are semantically acceptable. Kracht (2016, p. 130) gives the example in (152) where the determiner fails to exhibit the syntactic functions of the nouns.

(152) ..., dass die Katze die Mutter sieht.
..., that the cat sees the mother.
..., that the mother sees the cat.

Kracht suggests two solutions for such cases: the first is to produce all the possible analyses. In case of (152), there are two possibilities: one in which /die Katze/ 'the cat' is the subject and /die Mutter/ 'the mother' is the object, and the other is the opposite. The applicability of the proposed solution is because both can function as agents and themes. The second solution is to make the verb first merge with the rightmost entry. That is, the verb first merges with /die Mutter/ and then both merge the /die Katze/, and thus this sentence is interpreted as '..., that the cat sees the mother'. (153) and (154), taken from (Kracht, 2016, p. 95), show that the direct and indirect objects in English can be syntactically indistinguishable, and therefore the verb is not allowed to first combine with the indirect object.

(153) John gave the farmer the slave.

(154) John gave the slave the farmer.

But in the Arabic sentence the only accepted meaning is when the "toy" is the theme of the verb and the "child" is the recipient. The opposite is semantically odd because "the child" cannot be given to "the toy", but the two objects are allowed to exchange their positions. Even providing the entries with semantic features cannot completely solve the problem. They can be useful in some cases, since each semantic role has

(mostly) some typical features. The problem lies in the impossibility of generalization. For example, it can be said that agents are mostly human, but it cannot be generalized that agents are human. The impossibility of generalization renders adding semantic features to determine the semantic roles as not optimal.

7.3.7 Deviation from the basic meanings

An important motivation of AMRS is to be compositionally surface-oriented, and therefore it depends on the morphology and assumes neither syntactic structure nor movement. Two assumptions can be made in this regard: morphs are the basic expressions and if two morphs merge, they compose a complex expression. Using Frege's words, morphs are the parts whose meanings and the way they are organized determine the meaning of complex expressions. This entails that every morph should have a meaning, and this meaning is unique. I discuss only one manifestation of the deviation from the basic meaning of temporal forms in the language of newspapers, namely, the deviation from the basic meaning of the present. The most prominent deviation, that has been observed, is the use of the imperfective present in the headlines to denote the perfective past (See Gad (2017, p. 114-119)[3]. Consider, for example, the headline in (155) (taken from (Gād, 2017, p. 115).

(155) اَلسَّادَاتُ يُعلِنُ: نَحنُ مَعَ شَعبِ المَغرِبِ وَالحَسَنِ رَغمَ إِسَاءَةِ السِّيَاسِيِّنَ المَغَارِبَةِ لَنَا

assa:da:tu juʕlinu naħnu maʕa ʃaʕbi

Assadat.nom.sg.m.def state.3.sg.ind.prs we.nom.1.pl with people.gen.sg.m

almaɣribi walħasani raɣma ʔisa:ʔati

Morocco.gen.sg.m and-alħasan.gen.sg.m despite offense.gen.sg.f.def

issija:sijji:na almaɣ:ribati lana:

Politicians.gen.pl.m.def Moroccan.gen.pl.m to-us.gen.pl

'Al-sadaat states: we are with Morocco and Al-hasan despite the offense of the

[3]This observation has been made depending on a corpus of two newspapers: اَلأَهرَام al-ʔahra:m 'The Pyramids' and اَلشَرق الأَوسَط a-ʃʃarq al-ʔawsaʕ 'The Middle East'. The corpus has been collected by Gad, and it covers the period from 1960 to 2015 including headlines, articles, reports and news (Gād, 2017, p. 111).

Moroccan Politicians to us.'

In this headline, the verb /juʕlinu/ /يُعلِنُ/ 'state' is used in the present form, and the main meaning is supposed to be imperfective present, i.e., the time of topic equals the time of utterance and is included in the time of situation ($TT = TU$ & $TT \subset TSit$). However, the present form in (155) is used to denote the perfective past. In the body of the news, the past form /ʔaʕlana/ /أَعلَنَ/ 'stated' is used with adverbs of the past like /ʔams/ /أَمس/ 'yesterday'. While in (155) the intended meaning of the present form (as denoting the past) is emphasized in the body of the news by using the past form along with a past adverb, in (156) (taken from (Gād, 2017, p. 119)) the present form is used with a past adverb in the headline. In the normal use of Arabic, the headline in (156) is supposed to be ungrammatical in addition to the semantic deviation form the main meaning of the present form.

(156) هِلِيكُوبتَرُ تَابِعَةٌ للّقُواتِ المُسَلَّحَةِ المِصرِيَّةِ تَحُومُ فَوقَ رُؤُوسِ المُتَظَاهِرِين قُربَ القَصرِ الجُمهُورِيّ وَسَط القَاهِرَةِ أَمس

hilikubtaru taːbiʕatun lilquwwati almusallaħati almisrijjati
aircraft.nom.sg related.nom.sg to-forces.gen.pl.def armed.gen.def Egyptian.gen.def
taħuːmu fawqa ruʔuːsi lmutaðˤaːhiriːna qurba lqasˤri
hover.3.sg.ind.prs over head.gen.pl protester.gen.pl.def near palace.gen.sg.def
lʤumhuːrijji wasaðˤa lqaːhirati ʔams
republican.gen.sg.def downtown Cairo.gen.sg.def yesterday.

'A helicopter of the Egyptian armed forces hovers over the heads of the protesters near the Republican Palace downtown Cairo yesterday.'

I meant to show by (155) and (156) that the actual use of language may involve deviations from the basic meaning of the expression and violations of the standard rules.

7.3.8 Computational complexity

Computational lexicon-based approaches to natural languages are generally described as very time-consuming and tedious. AMRS is not an exception. It is not only a completely lexicalist approach but also a purely surface-oriented one. All information that governs the behavior of the strings needs to be stored in the lexicon. The problem of Arabic weak verbs is an obvious manifestation of this complexity. Three distinct entries for each hollow verb are inserted in the lexicon along their morphosyntactic information. That is to say, instead of applying general rules to deal with cases of dropping, adding or changing certain strings, I had to insert multiple entries (morphs) for the same morpheme, and each morph has its own morphosyntactic behavior.

7.4 Summary

I devoted this chapter to two significant points: the description of the developed morphsemantic lexicon and the evaluation of AMRS both theoretically and practically. The lexicon can be used within the computational implementation of the calculus in OCaml (Kracht, 2016). At the word level, roots first merge with vocalisms to compose stems. Both are considered discontinuous, i.e., fractured glued strings, and both have meanings. Then, stems concatenate with prefixes and suffixes to form words. Then, words combine to compose phrases and sentences. As an example, I showed the analysis of the sentence /kataba Ahmedu darsoaN/ [write.3.sg.pst.m Ahmed.nom.sg.m lesson.acc.sg.m.indef] 'Ahmed wrote a lesson'. In this example, I showed how the root /k⊗t⊗b/ first merges with /a⊗a/ to form the stem /katab/. I also explained how the argument structures and DRSs of the root and vocalism merge. Then, I showed that the stem merges with the suffix /a/ to form the 3rd person singular in the past. After I proved that AMRS is basically adequate for handling Arabic, I counted a number of theoretical and practical challenges the theory meets. These challenges include the absence of some overt morphological markers due to the Arabic modern writing system, the inadequacy to handle some types of multi-word expressions, the overlap of functions resulting from the boarder removal between morphology and syntax, the multiplicity of morphs that may lead to numerous entries for the same morpheme, the problem of specifying the

semantic role for the two kinds of objects (in some cases) and finally the semantic deviation from the basic meaning of expressions.

Chapter 8

Conclusion

There have been several attempts to compositionally handle natural language. In this dissertation, I investigated a theory of semantic composition and applied it to Arabic. The theory, introduced by Marcus Kracht, is an extension of the work of Kees Vermeulen on referent systems. I began the dissertation with questions on the bases of the theory, how to handle Arabic morphology and structure within the proposed calculus, to what extent the theory shows universal properties of language and the challenges it meets both theoretically and practically.

In the introductory chapter, the main assumptions of AMRS are indicated, whereas a detailed description of the semantic signs is given in chapter 3. The pure surface orientation, the maintenance of the principle of surface compositionality and the promotion of the role of the overt morphological information in building up complex constituents are generally the main assumptions of AMRS. Semantic signs are inserted in the lexicon, and each sign has two components: a morphological component and a semantic component. The morphological component is composed of a morpheme and its morphological information that provides clues to how it combines with other morphemes. The semantic component is composed of an argument structure and a DRS. The argument structure is a modification of the definition of RS. The mechanism of the renaming of the variables is kept as it is in the original version of RS but the clues, that indicate how variables are identified, are provided by morphology.

To reduce the complexity, I first dealt with word formation. The most important result in this part is the proof that the concatenative paradigm can be saved for Arabic word

formation if the discontinuity of morphemes is admitted. Arabic morphs are described as fractured glued strings that can be combined through a combinatorial function. Two main problems are solved: root and vocalism matching and morphophonemic alternations. The proper combination of roots and vocalisms is secured as follows:

- Each morph has a morphological class with four attributes:
  - Type: {triliteral , quadriliteral}
  - Sub-type: {sound, geminate, hollow ....}
  - Form: {I, II, II,....}
  - Stem-Eayon: {au, aa, ai, ia, ii, uu}
- Roots are only provided with out-classes because they only play the role of argument and never behave as a functor.
- Vocalism morphs are provided with in-class (for the hosted root morph) and out-class (for the result of the merge).
- The merge is defined only if the out-class of the root matches the in-class of the vocalism.

The second problem is the morphophonemic changes as particularly seen in the case of hollow roots. I considered all these changes as morphs of the same root morpheme, and consequently all of them are listed in the lexicon. This is consistent with the applied calculus and also prohibits overgeneration.

Then I moved to sentence structure. Arabic involves two types of sentences: verbal sentences and verbless sentences. I handled several challenges related to both types. The empty string, that bears the agreement morphological information, is suggested for both the absence of copula and the null-subject. The case of reduplication of arguments in verbal sentences is dealt with by allowing the verb to import one more argument provided that the extra argument is a pronoun, and it agrees with its antecedent. Free word order is accounted for using the G-access mechanism, and the adverbial phrase is allowed to appear at any position in the sentence.

AMRS allows any number of parameters in the AISs; however, I added only three:

properties, tense, and aspect. Following Kracht (2016) in his treatment of properties, nouns are dealt with as properties, rather than objects, and adjectives are considered functions that take properties as arguments and return properties that can then be attributed to nouns. As for tense and aspect, I argued for Comrie and Fassi's point of view that Arabic expresses both but following Klein, I dealt with both as relations. So, using Klein's notions, I represented the meaning of Arabic temporal morphosyntactic markers in terms of the relation between time of topic (TT) and time of utterance (TU) on the one hand, and between time of situation (TSit) and time of topic (TT) on the other. I also showed that Arabic expresses the common temporal relations in European languages. These relations are expressed either by morphology only or by morphosyntactic combinations.

I devoted chapter 7 to the description of the developed lexicon and the evaluation of the theory. In particular, I showed how Arabic morphs are listed in the lexicon with a focus on roots and vocalisms as fractured glued strings. Then, I described how these morphs are combined to compose phrases and clauses. I also referred to some challenges that AMRS meets. These challenges include the compositional treatment of multi-word expressions, the absence of diacritical marker in the modern writing system of Arabic, the overlap of functions due to the border removal between morphology and syntax, multiplicity of morphs, specifying the semantic role for the two kinds of objects when both the order and the morphological information do not suffice, the cases of deviation from the basic meanings, and the computational complexity. In the following points, I summarize the results of this dissertation:

- I presented a new attempt to both theoretically and practically represent the meaning of Arabic sentences dynamically and more smoothly. To the best of my knowledge, this is the first attempt to combine Arabic morphological units with their semantic representations.

- I proved that the concatenative paradigm can be saved in the treatment of Arabic morphology. This support Kracht's claim that admitting discontinuity can save the concatenative paradigm.

- In Arabic copula-less clauses, I proposed an empty string that functions as a

covert copula or a state verb and considered this construction as a default case. This default case can be overridden by rules of special cases. This empty string provides this type of sentences with tense, and it differs from the overt copula in two aspects: it imports two arguments in the nominative case, and it is in the present tense.

- I suggested two solutions to deal with the phenomenon of "reduplicated argument": either to deal with the anaphoric pronouns as a group of agreement suffixes or to allow the verb to import an extra argument. In the implementation, I followed the second solution.

- Like Comrie and Fassi, I argued that Arabic expresses both tense and aspect. However, I rejected all the current assumptions regarding the means of expressing tense and aspect. Particularly, I suggested that temporality is expressed in Arabic by vocalisms, auxiliaries, and clitics. If the stem forms are isolated, temporality is only expressed by vocalisms.

- The last point entails rejecting that the first vowel in /jaktub/ is part of the prefix, but it is part of the vocalism discontinuous morpheme.

- I classified Arabic as a non-sequence of tense language like Japanese and represented the temporal relations in Arabic embedded clauses depending on this classification.

- I showed that Arabic can provide a supporting piece of evidence for the argument that noun phrases may express temporal relations morphologically. This is in line with Nordlinger and Sadler (Nordlinger & Sadler, 2004) and against Kracht's claim that noun phrases do not express their dependency on time morphologically (Kracht, 2016).

- I evaluated the theory of AMRS and showed that it provides a powerful mechanism to handle complex phenomena both morphologically and structurally; however, it meets some challenges both theoretically and practically.

- As part of my work, I built a computational lexicon for (a fragment of) Arabic depending on the proposed theoretical analysis.

This work can be extended in the future both theoretically and practically. From a theoretical perspective, it can be developed to include topics like quantification and conjunctions. These two topics are additional significant indicators for the power of the calculus. Also, the phenomenon of discontinuity of constituents needs to be cross-linguistically investigated for a better understanding of this significant phenomenon. Practically, more work can be devoted to increase the entries of the lexicon. Also, more parameters can be studied and added, since the calculus allows for any number of parameters.

References

Abdelali, A., Darwish, K., Durrani, N., & Mubarak, H. (2016). Farasa: A fast and furious segmenter for arabic. In Proceedings of the 2016 conference of the north american chapter of the association for computational linguistics: Demonstrations (pp. 11–16).

Aboamer, Y., & Farghaly, A. (2015). Mariam Comlex: A Bi-Directional Finite State Morphological Transducer for MSA. In The 29th Annual Symposium on Arabic Linguistics at the University of Wisconsin-Milwaukee, USA.

Aboamer, Y., & Kracht, M. (2019). A purely surface-oriented approach to handling arabic morphology. In International conference on formal grammar (pp. 1–17).

Aikhenvald, A. (2003). A grammar of tariana, from northwest amazonia. Cambridge University Press.

Ajdukiewicz, K. (1935). Die syntaktische konnexitat. Studia philosophica, 1–27.

Al-Johar, B. (1999). A portable natural language interface from arabic to sql. (Unpublished doctoral dissertation). PhD thesis, University of Sheffield.

Al-Johar, B., & McGregor, J. (1997). A logical meaning representation for arabic (lmra). In Proceedings of the 15th national computer conference, riyadh, saudi arabia (pp. 31–40).

Alotaibi, A. S. (2018). The copula in arabic: Description and analysis (Unpublished doctoral dissertation). University of Essex.

Al-Sughaiyer, I. A., & Al-Kharashi, I. A. (2004). Arabic morphological analysis techniques: A comprehensive survey. Journal of the Association for Information Science and Technology, 55(3), 189–213.

Aoun, J., Benmamoun, E., & Choueiri, L. (2010). Arabic syntax. Cambridge: Cambridge.

Attia, M. (2008a). Handling arabic morphological and syntactic ambiguity within the lfg framework with a view to machine translation. PhD dissertation, University of Manchester.

Attia, M. (2008b). A unified analysis of copula constructions in lfg. In In butt, m. and king, t. (ed.) proceedings of the lfg08 conference, csli publications (Vol. 3, p. 105).

Attia, M., Pecina, P., Toral, A., & Van Genabith, J. (2013). A corpus-based finite-state morphological toolkit for contemporary arabic. Journal of Logic and Computation, 24(2), 455–472.

Badawi, E.-S. (1979). mustawaja tu al-ᶜarabijjati al-muᶜa ṣirati fi miṣr. Dar Al-maQ:rif.

Bahloul, M. (2007). Structure and function of the arabic verb. Routledge.

Bahloul, R. (1991). Morphology of the verbal system in arabic. ms, Cornell University.

Bakir, M. J. (1980). Aspects of clause structure in arabic. Indiana U. Linguistics Club.

Bauer, L. (2003). Introducing linguistic morphology. Edinburgh University Press Edinburgh.

Beesley, K. R. (1998). Arabic morphology using only finite-state operations. In Proceedings of the Workshop on Computational Approaches to Semitic languages (pp. 50–57).

Beesley, K. R. (2001). Finite-state morphological analysis and generation of Arabic at Xerox Research: Status and plans in 2001. In ACL Workshop on Arabic Language Processing: Status and Perspective (Vol. 1, pp. 1–8).

Bender, E. (2001). Syntactic variation and linguistic competence (Unpublished doctoral dissertation). Ph. D. dissertation, Stanford.

Benmamoun, E. (1993). Functional and inflectional morphology problems of projection, representation and derivation.

Benmamoun, E. (2000). The feature structure of functional categories: A comparative study of arabic dialects. Oxford University Press.

Benmamoun, E. (2008). Clause structure and the syntax of verbless sentences. Current Studies in Linguistics Series, 45, 105.

Bohnemeyer, J. (2003). Relative tense vs. aspect: The case reopened. Handout of a

talk given at SULA, 2, 14–16.

Bos, J., Mastenbroek, E., MacGlashan, S., Millies, S., & Pinkal, M. (1994). A compositional drs-based formalism for nlp applications.

Bowers, J. (2001). Predication. Oxford: Blackwell.

Buckwalter, T. (2002). Buckwalter Arabic Morphological Analyzer Version 1.0 Linguistic Data Consortium, University of Pennsylvania, LDC Catalog No: LDC2002L49. (Tech. Rep.). (ISBN 1-58563-257-0)

Buckwalter, T. (2004). Buckwalter Arabic Morphological Analyzer Version 2.0. Linguistic Data Consortium, University of Pennsylvania, 2002. LDC Catalog No.: LDC2004L02 (Tech. Rep.). ISBN 1-58563-324-0.

Cann, R. (1993). Formal semantics: an introduction. Cambridge University Press.

Cann, R., Kempson, R., & Gregoromichelaki, E. (2009). Semantics: An introduction to meaning in language.

Carnie, A. (2006). Syntax: A generative introduction. BLACKWELL PUBLISHING.

CIA. (2018). Cia world fact book.

Comrie, B. (1976). Aspect: An introduction to the study of verbal aspect and related problems (Vol. 2). Cambridge university press.

Comrie, B. (1985). Tense (Vol. 17). Cambridge university press.

Crystal, D. (2008). A dictionary of linguistics and phonetics (the language library). John Wiley & Sons Incorporated.

Dalrymple, M., Dyvik, H., & King, T. H. (2004). Copular complements: Closed or open. In Proceedings of the lfg04 conference (pp. 188–198).

Danks, W. (2011). The arabic verb: form and meaning in the vowel-lengthening patterns (Vol. 63). John Benjamins Publishing.

Declerck, R. (1986). From reichenbach (1947) to comrie (1985) and beyond: Towards a theory of tense. Lingua, 70(4), 305–364.

Declerck, R. (1991). Tense in english: Its structure and use in discourse. Routledge.

Declerck, R. (1995). Is there a relative past tense in english? Lingua, 97(1), 1–36.

Den Dikken, M. (2006). Relators and linkers: The syntax of predication, predicate inversion, and copulas (Vol. 47). MIT press.

Deo, A. (2015). Diachronic semantics. Annu. Rev. Linguist., 1(1), 179–197.

Dichy, J., & Farghaly, A. (2007). Grammar-lexis relations in the computational morphology of arabic. In Arabic Computational Morphology (pp. 115–140). Springer.

Dickins, J., & Watson, J. C. (2009). Transitivity. Encyclopedia of Arabic Language and Linguistics, 4, 528–535.

El-Sayed, H. (2011). Philosophical investigations into formalization and compositionality of language: a montagovian analysis of arabic quantification as a case study (Unpublished doctoral dissertation). PhD thesis, University of Leeds.

El-Sayed, H. (2015). Arabic between formalization and computation. International Journal of Languages, Literature and Linguistics, 1(1), 25–29.

El-Sayed, H. (2019). A survey and classification of arabic logic-based meaning representations (almr).

Enç, M. (1986). Towards a referential analysis of temporal expressions. Linguistics and Philosophy, 405–426.

Enç, M. (1987). Anchoring conditions for tense. Linguistic inquiry, 633–657.

Er-rayyan, M. R. H. (1986). Toward the construction of a temporal system for natural language in the light of the data of the arabic and english languages.

Farghaly, A. e. (2010). Arabic computational linguistics. CSLI Publications.

Fassi Fehri, A. (1993). Issues in the structure of arabic clauses and words (vol. 29). Studies in natural language and linguistic theory.

Fehri, A. F. (2012). Key features and parameters in arabic grammar. J. Benjamins Publishing Company.

Ferguson, C. A. (1959). Diglossia. word, 15(2), 325–340.

Frisch, S. A., & Zawaydeh, B. A. (2001). The psychological reality of OCP-Place in Arabic. Language, 77(1), 91–106.

Gād, H. (2017). Alttarākību alddaltu alā alzzamani fī lughati alṣṣaḥāfati almu āṣirati: dirāsatun naḥwiyyatun.

Gil, D. (2004). Riau indonesian sama: Explorations in macrofunctionality. In Haspelmath, M. (ed) Coordinating constructions, 58, 371-424.

Glanzberg, M. (2018). Truth. In Zalta, E. N. (ed.), The Stanford Encyclopedia of Philosophy (Fall 2018 Edition), URL = <https://plato.stanford.edu/archives/fall2018/entries/truth/>.

González-Rivera, M. (2010). On the internal structure of spanish verbless clauses (Unpublished doctoral dissertation). The Ohio State University.

Groenendijk, J., & Stokhof, M. (2005). Why compositionality.

Habash, N. (2010). Introduction to arabic natural language processing. Synthesis Lectures on Human Language Technologies, 3(1), 1–187.

Habash, N., Diab, M. T., & Rambow, O. (2012). Conventional orthography for dialectal arabic. In Lrec (pp. 711–718).

Habash, N., Eskander, R., & Hawwari, A. (2012). A morphological analyzer for egyptian arabic. In Proceedings of the twelfth meeting of the special interest group on computational morphology and phonology (pp. 1–9).

Habash, N., Gabbard, R., Rambow, O., Kulick, S., & Marcus, M. (2007). Determining case in arabic: Learning complex linguistic behavior requires complex linguistic features. In Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (emnlp-conll) (pp. 1084–1092).

Habash, N., Rambow, O., & Roth, R. (2009). MADA+ TOKAN: A toolkit for Arabic tokenization, diacritization, morphological disambiguation, POS tagging, stemming and lemmatization. In Proceedings of the 2nd international conference on Arabic language resources and tools (MEDAR), Cairo, Egypt (Vol. 41).

Haddad, B. (2007). Semantic representation of arabic: A logical approach towards compositionality and generalized arabic quantifiers. International Journal of Computer Processing of Oriental Languages, 20(01), 37–52.

Haddad, B., & Yaseen, M. (2001). Towards understanding arabic: A logical approach for semantic representation. In Arabic nlp workshop, acl'01.

Haddad, B., & Yaseen, M. (2003). Towards semantic composition of arabic: A λ-drt based approach. In Mt summit ix, workshop on machine translation for semitic languages: Issues and approaches, amta, new orleans.

Haddad, B., & Yaseen, M. (2005). A compositional approach towards semantic representation and construction of arabic. In International conference on logical aspects of computational linguistics (pp. 147–161).

Hallman, P. (2015). The arabic imperfective. Brill's Journal of Afroasiatic Languages

and Linguistics, 7(1), 103–131.

Hausser, R. R. (1980). Surface compositionality and the semantics of mood. In Speech act theory and pragmatics (pp. 71–95). Springer.

Holes, C. (2004). Modern arabic: Structures, functions, and varieties. Georgetown University Press.

Janssen, T. (2020). Montague semantics. In Zalta, E. N. (ed.), The Stanford Encyclopedia of Philosophy (Spring 2020 Edition), forthcoming URL = https://plato.stanford.edu/archives/spr2020/entries/montague-semantics.

Kamp, H., & Reyle, U. (1993). From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory. Kluwer Academic Publishers, Dordrecht, Holland.

Kamp, H., Van Genabith, J., & Reyle, U. (2011). Discourse representation theory. In Handbook of philosophical logic (pp. 125–394). Springer.

Kasami, T., Seki, H., & Fujii, M. (1987). Generalized context-free grammars, multiple context-free grammars and head grammars. Preprint of WG on Natural Language of IPSJ.

Kay, M. (1987). Nonconcatenative finite-state morphology. In Proceedings of the third conference of the European chapter of the Association for Computational Linguistics (pp. 2–10).

Kaye, A. S. (1994). Formal vs. informal in arabic: Diglossia, triglossia, tetraglossia, etc., polyglossia—multiglossia viewed as a continuum. Zeitschrift für arabische Linguistik(27), 47–66.

Kaye, A. S. (2001). Diglossia: the state of the art. International journal of the sociology of language, 2001(152), 117–129.

Keil, B. (2005). An associative semantics for basic sentences in malagasy. Heinz and Ntelitheos, 131–140.

King, J. (2006). Formal semantics. In Lepore, E., and Smith, B. C. (eds.) "The Oxford handbook to the philosophy of language.

Klein, U., & Kracht, M. (2011). Direction and obviation in plains cree: A referent systems approach. In Proceedings of sinn und bedeutung (Vol. 15, pp. 381–396).

Klein, W. (1994). Time in language. Psychology Press.

Klein, W. (2018). Looking at language (Vol. 317). Walter de Gruyter GmbH & Co KG.

Kracht, M. (2007). Compositionality: The very idea. Research on Language and Computation, 5(3), 287–308.

Kracht, M. (2008). Compositionality in montague grammar. Edouard Machery und Markus Werning Wolfram Hinzen, editor, Handbook of Compositionality, 47–63.

Kracht, M. (2016). Agreement morphology, argument structure and syntax. Unpublished manuscript.

Kracht, M., & Aboamer, Y. (2017). Argument structure and referent systems. In Iwcs 2017—12th international conference on computational semantics.

Kuryłowicz, J. (1973). Verbal aspect in semitic. Orientalia, 42, 114–120.

Lassiter, D. (2015). Adjectival modification and gradation. Handbook of contemporary semantic theory, 143–167.

Leech, G. (1974). Semantics. england. Penguin Books Ltd.

Lohnstein, H. (2011). Formale semantik und natürliche sprache. Springer.

Maamouri, M., Graff, D., Bouziri, B., Krouna, S., & Kulick, S. (2010). Ldc standard arabic morphological analyzer sama v. 3.1 (Tech. Rep.). LDC Catalog No. LDC2010L01. ISBN 1-58563-555-3.

McCarthy, J. J. (1981). A prosodic theory of nonconcatenative morphology. Linguistic inquiry, 12(3), 373–418.

McCarthy, J. J., & Prince, A. (1996). Prosodic morphology 1986.

McWhorter, J. H., Leven, J., & Blandford, J. (2004). The story of human language. Teaching Company.

Montague, R. (1970a). English as a formal language.

Montague, R. (1970b). Universal grammar. Theoria, 36(3), 373–398.

Montague, R. (1973). The proper treatment of quantification in ordinary english. In Approaches to natural language (pp. 221–242). Springer.

Morrill, G. (2011). Categorial grammar: Logical syntax, semantics, and processing. Oxford University Press.

Murphy, M. L. (2010). Lexical meaning. Cambridge University Press.

Muskens, R. (1996). Combining montague semantics and discourse representation.

Linguistics and philosophy, 19(2), 143–186.

Nordlinger, R., & Sadler, L. (2004). Nominal tense in crosslinguistic perspective. Language, 776–806.

Nordlinger, R., & Sadler, L. (2007). Verbless clauses: revealing the structure within.

Nunberg, G. (1994). Ivan a sag & thomas wasow. idioms. Language, 3.

Ogihara, T. (1996). Tense, scope and attitude ascription. Dordrecht, NL: Kluwer.

Partee, B. (1976). Montague grammar.

Partee, B. (1995). Lexical semantics and compositionality. An invitation to cognitive science: Language, 1, 311–360.

Partee, B. (2003). Are there privative adjectives. In Conference on the philosophy of terry parsons, university of massachusetts, amherst.

Partee, B., ter Meulen, A. G., & Wall, R. (1990). Mathematical methods in linguistics (vol. 30). Studies in Linguistics and Philosophy. Dordrecht, NL: Kluwer Academic Press.

Payne, T. E. (2017). Morphological typology. In A. Y. Aikhenvald & R. M. W. Dixon (Eds.), The cambridge handbook of linguistic typology (p. 78–94). Cambridge University Press. doi: 10.1017/9781316135716.003

Pollard, C., & Sag, I. A. (1994). Head-driven phrase structure grammar. University of Chicago Press.

Prior, A. N. (1967). Past, present and future (Vol. 154). Clarendon Press Oxford.

Procházka, S. (2010). Arabic. In Brown, K., and Ogilvie, S. (eds.) Concise encyclopedia of languages of the world..

Reichenbach, H. (1947). Elements of symbolic logic.

Reinhard, S., & Gibbon, D. (1991). Prosodic inheritance and morphological generalisations. In Proceedings of the fifth conference of the European chapter of the Association for Computational Linguistics (pp. 131–136).

Rendón, J. G. (n.d.). Language contact in paraguayan guarani . In The oxford handbook of language contact.

Ross, J. R. (1967). Constraints on variables in syntax.

Ryding, K. C. (2005). A reference grammar of modern standard arabic. Cambridge university press.

Sag, I. A., Baldwin, T., Bond, F., Copestake, A., & Flickinger, D. (2002). Multiword expressions: A pain in the neck for nlp. In International conference on intelligent text processing and computational linguistics (pp. 1–15).

Sag, I. A., Wasow, T., & Bender, E. M. (1999). Syntactic theory: A formal introduction (Vol. 92). Center for the Study of Language and Information Stanford, CA.

Sahel, S., & Vogel, R. (2013). Einführung in die morphologie des deutschen.

Salkie, R. (1989). Perfect and pluperfect: what is the relationship? Journal of linguistics, 25(1), 1–34.

Sawalha, M., & Atwell, E. (2008). Comparative evaluation of Arabic language morphological analysers and stemmers. Coling 2008: Companion volume: Posters, 107–110.

Sawalha, M., Atwell, E., & Abushariah, M. A. (2013). Salma: standard arabic language morphological analysis. In 1st International Conference on Communications, Signal Processing, and their Applications ICCSPA (pp. 1–6).

Seki, H., Matsumura, T., Fujii, M., & Kasami, T. (1991). On multiple context-free grammars. Theoretical Computer Science, 88(2), 191–229.

Shieber, S. M. (2003). An introduction to unification-based approaches to grammar. Microtome Publishing.

Siegel, M. E. (1976). Capturing the adjective.

Soudi, A., Cavalli-Sforza, V., & Jamari, A. (2002). The arabic noun system generation. In Proceedings of the international symposium on the processing of arabic (pp. 69–87).

Soudi, A., Neumann, G., & Van den Bosch, A. (2007). Arabic computational morphology: knowledge-based and empirical methods. In Arabic Computational Morphology (pp. 3–14). Springer.

Steedman, M. (1993). Categorial grammar. Lingua, 90(3), 221–258.

Stowell, T. A. (1981). Origins of phrase structure (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Taji, D., Khalifa, S., Obeid, O., Eryani, F., & Habash, N. (2018). An arabic morphological analyzer and generator with copious features. In Proceedings of the fifteenth workshop on computational research in phonetics, phonology, and morphology

(pp. 140–150).

Tallerman, M. (2014). Understanding syntax. Routledge.

Tonhauser, J. (2005). Towards an understanding of the meaning of nominal tense. In Proceedings of sinn und bedeutung (Vol. 9, pp. 475–488).

Vermeulen, C. F. M. (1995). Merging without mystery or: Variables in dynamics semantics. Journal of Philosophical Logic, 24(4), 405–450.

Watson, J. (1999). The syntax of arabic headlines and news summaries. Arabic grammar and linguistics, 161.

Watson, J. (2002). The phonology and morphology of arabic. Oxford University Press on Demand.

Watson, J. (2010). Arabic as an introflecting language. In Brown, K., and Ogilvie, S. (eds.) Concise encyclopedia of languages of the world..

Wehr, H. (1994). Arabic-english dictionary: The hans wehr dictionary of modern written arabic. Ithaca, NY: Spoken Language Services.

Winter, Y. (2016). Elements of formal semantics: An introduction to the mathematical theory of meaning in natural language. Edinburgh University Press.

Wright, W., & Caspari, C. P. (1896). A grammar of the arabic language: translated from the german of caspari, and edited with numerous additions and corrections.

Yoder, B. (2010). Syntactic underspecification in riau indonesian. Work Papers of the Summer Institute of Linguistics, University of North Dakota Session, 50(1), 1.

Zeevat, H. (1989). A compositional approach to discourse representation theory. Linguistics and Philosophy, 12(1), 95–131.

Arabic Morphosemantic Lexicon

```
<?xml version="1.0"?>

-<dict xsi:noNamespaceSchemaLocation="refsysschema.xsd" xmlns:xsi=
"http://www.w3.org/2001/XMLSchema-instance">

<name string="Arabische Verben" vlg="de"/>

<name string="Arabic Verbs" vlg="en"/>

<!-- General handlers do not work properly, so this is added fortesting
purposes. -->


<hdlg lg="1" gen="lg" id="post"/>


-<hdl id="pre1">


-<unit>

<prt fct="true" pos="0"/>

<prt fct="false" pos="0"/>

</unit>

</hdl>


-<hdl id="pre2">


-<unit>

<prt fct="true" pos="0"/>

</unit>


-<unit>

<prt fct="false" pos="0"/>

</unit>

</hdl>


-<hdl id="post1">


-<unit>

<prt fct="false" pos="0"/>

<prt fct="true" pos="0"/>

</unit>

</hdl>


+<hdl id="post2">
```

```
-<hdl id="post3">
-<unit>
<prt fct="false" pos="1"/>
<prt fct="false" pos="0"/>
<prt fct="true" pos="0"/>
</unit>
</hdl>
-<hdl id="null3">
-<unit>
<prt fct="true" pos="0"/>
</unit>
-<unit>
<prt fct="true" pos="1"/>
</unit>
-<unit>
<prt fct="true" pos="2"/>
</unit>
</hdl>
-<hdl id="perf">
-<unit>
<prt fct="false" pos="0"/>
<prt fct="true" pos="0"/>
<prt fct="false" pos="1"/>
<prt fct="true" pos="1"/>
<prt fct="false" pos="2"/>
```

```
</unit>
</hdl>

-<hdl id="null">

-<unit>
<prt fct="true" pos="0"/>
</unit>
</hdl>
<!-- This is the list of roots -->

<!-- Triliteral roots -->


-<entry id="schreiben">

-<argx>

-<arg>
<var name="e0"/>
<dia name="u"/>
<avi/>
<avo voice="ast" mood="ast" tense="ast" asp="ast" trs="trans" cat="root"/>

-<pavs>
<vp name="gf1" out="x0"/>
<vp name="gf2" out="x1"/>
</pavs>
</arg>
</argx>

-<mor>

-<exp>

-<gst>
<pre/>
<text>k</text>
```

```
<post/>
</gst>

-<gst>
<pre/>
<text>t</text>
<post/>
</gst>

-<gst>
<pre/>
<text>b</text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="null3"/>
<mi/>
<mo stem-Eayon="ai" form="I" sub-type="sound" type="triliteral" morph=
"root"/>
</ma>
</marg>
</mor>

-<sem>

-<drs head="e0+x0+x1">

-<lit>

-<trmf fc="write">
<trmv name="e0"/>
</trmf>
</lit>
```

```
-<eq>

-<trmf fc="agt">
<trmv name="e0"/>
</trmf>
<trmv name="x0"/>
</eq>

-<eq>

-<trmf fc="thm">
<trmv name="e0"/>
</trmf>
<trmv name="x1"/>
</eq>
</drs>
</sem>
</entry>

-<entry id="geben">

-<argx>

-<arg>
<var name="e0"/>
<dia name="u"/>
<avi/>
<avo voice="ast" mood="ast" tense="ast" asp="ast" trs="ditrans" cat="root"
/>

-<pavs>
<vp name="gf1" out="x0"/>
<vp name="gf2" out="x1"/>
<vp name="gf3" out="x2"/>
</pavs>
</arg>
</argx>
```

```
-<mor>

-<exp>

-<gst>
<pre/>
<text>m</text>
<post/>
</gst>

-<gst>
<pre/>
<text>n</text>
<post/>
</gst>

-<gst>
<pre/>
<text>h</text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="null3"/>
<mi/>
<mo stem-Eayon="ai" form="I" sub-type="sound" type="triliteral" morph=
"root"/>
</ma>
</marg>
</mor>

-<sem>

-<drs head="e0+x0+x1+x2">
```

```
-<lit>

-<trmf fc="geben">
<trmv name="e0"/>
</trmf>
</lit>

-<eq>

-<trmf fc="agt">
<trmv name="e0"/>
</trmf>
<trmv name="x0"/>
</eq>

-<eq>

+<trmf fc="thm">
<trmv name="x1"/>
</eq>

-<eq>

-<trmf fc="ben">
<trmv name="e0"/>
</trmf>
<trmv name="x2"/>
</eq>
</drs>
</sem>
</entry>
<!-- List of Derivational Morphs -->

<!-- Perfct stem morphs -->

<!-- Active form -->
```

```
+<entry id="p1trans">
```

-<entry id="p1ditrans">

```
-<argx>

-<arg>
<var name="e0"/>
<dia name="duft"/>
<avi voice="ast" mood="ast" tense="ast" asp="ast" trs="ditrans" cat="root"
/>
<avo voice="act" mood="ast" tense="past" asp="perf" trs="ditrans" cat=
"stem"/>

-<pavs>
<vp name="TT" out="t1"/>
<vp name="TSit" out="t0"/>
<vp name="TU" out="t2"/>
</pavs>
</arg>
</argx>

-<mor>

-<exp>

-<gst>
<pre/>
<text>a</text>
<post/>
</gst>

-<gst>
<pre/>
<text>a</text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
```

```
<mi stem-Eayon="au+aa+ai" form="I" sub-type="sound" type="triliteral"
morph="root"/>

<mo/>

<hdl idref="perf"/>

</ma>

</marg>

</mor>

-<sem>

-<drs head="e0+t0+t1+t2">

-<eq>

-<trmf fc="time">

<trmv name="e0"/>

</trmf>

<trmv name="t0"/>

</eq>

-<lit>

-<trmf fc="before">

<trmv name="t1"/>

<trmv name="t2"/>

</trmf>

</lit>

-<lit>

-<trmf fc="include">

<trmv name="t1"/>

<trmv name="t0"/>

</trmf>

</lit>

</drs>

</sem>

</entry>
```

```
<!-- List of suffixes -->

<!-- List of perfect suffixes (subjects) -->


-<entry id="atrasn">

-<argx>

-<arg>

<var name="e0"/>

<dia name="duft"/>

<avi voice="act" mood="ast" tense="past" asp="perf" trs="trans" cat="stem"
/>

<avo voice="act" mood="ast" tense="past" asp="perf" trs="trans" cat="verb"
 gend="masc" num="sg" per="3rd"/>

-<pavs>

<vp name="gf1" in="x0"/>

<vp name="gf2" in="x1"/>

</pavs>

</arg>

-<arg>

<var name="x0"/>

<dia name="d"/>

<avi cat="noun" gend="masc" num="sg" pers="3rd" case="nom"/>

<avo/>

</arg>

-<arg>

<var name="x1"/>

<dia name="d"/>

<avi cat="noun" case="acc"/>

<avo/>

<pavs/>

</arg>

</argx>
```

```
-<mor>

-<exp>

-<gst>
<pre/>
<text>a</text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>
</marg>
</mor>

-<mor>

-<exp>
```

```
-<gst>
<pre/>
<text>a </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>
</marg>
</mor>

-<mor>

-<exp>

-<gst>
<pre/>
<text>a</text>
```

```
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>
</marg>
</mor>

-<mor>

-<exp>

-<gst>
<pre/>
<text>a </text>
<post/>
</gst>
</exp>
```

```
-<marg>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>
</marg>
</mor>

-<mor>

-<exp>

-<gst>
<pre/>
<text>a</text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="post1"/>
```

```
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>
</marg>
</mor>

-<mor>

-<exp>

-<gst>
<pre/>
<text>a </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>
```

```
-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>
</marg>
</mor>
</entry>

-<entry id="atrasnlef">

-<argx>

-<arg>
<var name="e0"/>
<dia name="duft"/>
<avi voice="act" mood="ast" tense="past" asp="perf" trs="trans" cat="stem"
/>
<avo voice="act" mood="ast" tense="past" asp="perf" trs="trans" cat="verb"
 gend="masc" num="sg" per="3rd"/>

-<pavs>
<vp name="gf1" in="x0"/>
<vp name="gf2" in="x1"/>
<vp name="gf3" in="x2"/>
</pavs>
</arg>

-<arg>
<var name="x0"/>
<dia name="d"/>
```

```
<avi cat="noun" gend="masc" num="sg" pers="3rd" case="nom"/>
<avo/>
</arg>

-<arg>
<var name="x1"/>
<dia name="d"/>
<avi cat="noun" case="acc"/>
<avo/>
<pavs/>
</arg>

-<arg>
<var name="x2"/>
<dia name="d"/>
<avi cat="pro" case="acc"/>
<avo/>
<pavs/>
</arg>
</argx>

-<mor>

-<exp>

-<gst>
<pre/>
<text>a</text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="post1"/>
<mi/>
```

```
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>
</marg>
</mor>

-<sem>

-<drs head="">

-<eq>
<trmv name="x1"/>
<trmf fc="x2"/>
</eq>
</drs>
</sem>
</entry>

-<entry id="aditrans">
```

```
-<argx>

-<arg>
<var name="e0"/>
<dia name="duft"/>
<avi voice="act" mood="ast" tense="past" asp="perf" trs="ditrans" cat=
"stem"/>
<avo voice="act" mood="ast" tense="past" asp="perf" trs="ditrans" cat=
"verb" gend="masc" num="sg" per="3rd"/>

-<pavs>
<vp name="gf1" in="x0"/>
<vp name="gf2" in="x1"/>
<vp name="gf3" in="x2"/>
</pavs>
</arg>

-<arg>
<var name="x0"/>
<dia name="d"/>
<avi cat="noun" gend="masc" num="sg" pers="3rd" case="nom"/>
<avo/>
</arg>

-<arg>
<var name="x1"/>
<dia name="d"/>
<avi cat="noun" case="acc"/>
<avo/>
<pavs/>
</arg>

-<arg>
<var name="x2"/>
<dia name="d"/>
<avi cat="noun" case="acc"/>
```

```
<avo/>
<pavs/>
</arg>
</argx>

-<mor>

-<exp>

-<gst>
<pre/>
<text>a</text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>

-<ma>
```

```
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>
</marg>
</mor>

-<mor>

-<exp>

-<gst>
<pre/>
<text>a </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>
```

```
-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>
</marg>
</mor>

-<mor>

-<exp>

-<gst>
<pre/>
<text>a</text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
```

```
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>
</marg>
</mor>

-<mor>

-<exp>

-<gst>
<pre/>
<text>a </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
```

```
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>
</marg>
</mor>

-<mor>

-<exp>

-<gst>
<pre/>
<text>a</text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
```

```
</ma>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>
</marg>
</mor>

-<mor>

-<exp>

-<gst>
<pre/>
<text>a </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
```

```
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>
</marg>
</mor>

-<mor>

-<exp>

-<gst>
<pre/>
<text>a</text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>
```

```
-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>
</marg>
</mor>

-<mor>

-<exp>

-<gst>
<pre/>
<text>a </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
```

```
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>
</marg>
</mor>
</entry>
<!-- List of object pronouns -->

-<entry id="Hu">

-<argx>

-<arg>
<var name="x1"/>
<dia name="u"/>
<avi/>
<avo cat="pro" case="acc"/>
<pavs/>
</arg>
```

```
</argx>

-<mor>

-<exp>

-<gst>
<pre/>
<text>Hu </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<mi/>
<mo/>
<hdl idref="null"/>
</ma>
</marg>
</mor>

-<sem>
<drs/>
</sem>
</entry>
<!-- List of subject nouns -->

-<entry id="Ahmed">

-<argx>

-<arg>
<var name="x0"/>
<dia name="u"/>
<avi/>
```

```
<avo cat="noun" gend="masc" num="sg" pers="3rd" case="nom"/>
<pavs/>
</arg>
</argx>

-<mor>

-<exp>

-<gst>
<pre/>
<text>Ahmedu </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="null"/>
</ma>
</marg>
</mor>

-<sem>

-<drs head="">

-<eq>
<trmv name="x0"/>
<trmf fc="Ahmed"/>
</eq>
</drs>
</sem>
</entry>

-<entry id="n1">
```

```
-<argx>

-<arg>
<var name="x0"/>
<dia name="u"/>
<avi/>
<avo cat="noun" gend="masc" num="sg" pers="3rd" case="nom"/>
<pavs/>
</arg>
</argx>

-<mor>

-<exp>

-<gst zero="0">
<pre/>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="pre1"/>
<mi/>
<mo/>
</ma>
</marg>
</mor>

-<mor>

-<exp>

-<gst zero="0">
<pre/>
<post/>
```

```
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>
</marg>
</mor>

-<mor>

-<exp>

-<gst zero="0">
<pre/>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="null"/>
<mi/>
<mo/>
</ma>
</marg>
</mor>

-<sem>

-<drs head="">

-<eq>
```

```
<trmv name="x0"/>
<trmf fc="Ahmed"/>
</eq>
</drs>
</sem>
</entry>
<!-- List of object nouns -->

-<entry id="Lehre">

-<argx>

-<arg>
<var name="x1"/>
<dia name="u"/>
<avi/>
<avo gen="masc" cat="noun" num="sg" pers="3rd" case="acc"/>
<pavs/>
</arg>
</argx>

-<mor>

-<exp>

-<gst>
<pre/>
<text>darosaN </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<mi/>
```

```
<mo/>
<hdl idref="null"/>
</ma>
</marg>
</mor>
-<sem>
-<drs head="">
-<lit>
-<trmf fc="lesson">
<trmv name="x1"/>
</trmf>
</lit>
</drs>
</sem>
</entry>
-<entry id="Aly">
-<argx>
-<arg>
<var name="x1"/>
<dia name="u"/>
<avi/>
<avo cat="noun" case="acc"/>
<pavs/>
</arg>
</argx>
-<mor>
-<exp>
-<gst>
```

```
<pre/>
<text>AlyaN </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<mi/>
<mo/>
<hdl idref="null"/>
</ma>
</marg>
</mor>

-<sem>

-<drs head="">

-<eq>
<trmv name="x1"/>
<trmf fc="Ali"/>
</eq>
</drs>
</sem>
</entry>

-<entry id="present">

-<argx>

-<arg>
<var name="x2"/>
<dia name="u"/>
<avi/>
<avo cat="noun" case="acc"/>
```

```
<pavs/>
</arg>
</argx>

-<mor>

-<exp>

-<gst>
<pre/>
<text>jAizataN </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<mi/>
<mo/>
<hdl idref="null"/>
</ma>
</marg>
</mor>

-<sem>

-<drs head="">

-<lit>

-<trmf fc="present">
<trmv name="x2"/>
</trmf>
</lit>
</drs>
</sem>
</entry>
```

```
<!-- List of adj -->

-<entry id="neu">

-<argx>

-<arg>
<var name="x0"/>
<dia name="du"/>
<avi gen="masc" cat="noun" num="sg" pers="3rd" case="acc"/>
<avo/>
<pavs/>
</arg>
</argx>

-<mor>

-<exp>

-<gst>
<pre/>
<text>jadIdaN </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<mi/>
<mo/>
<hdl idref="post1"/>
</ma>
</marg>
</mor>

-<sem>
```

```
-<drs head="x0">

-<lit>

-<trmf fc="neu">
<trmv name="x0"/>
</trmf>
</lit>
</drs>
</sem>
</entry>
<!-- List of adv -->

-<entry id="today">

-<argx>

-<arg>
<var name="x0"/>
<dia name="du"/>
<avi cat="noun"/>
<avo/>

-<pavs>
<vp name="TT" in="t"/>
</pavs>
</arg>
</argx>

-<mor>

-<exp>

-<gst>
<pre/>
<text>Alyauma </text>
```

```
<post/>
</gst>
</exp>

-<marg>

-<ma>
<mi/>
<mo/>
<hdl idref="post1"/>
</ma>
</marg>
</mor>

-<mor>

-<exp>

-<gst>
<pre/>
<text>Alyauma </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<mi/>
<mo/>
<hdl idref="pre1"/>
</ma>
</marg>
</mor>

-<sem>

-<drs head="">
```

```
-<lit>
<trmf fc="t is in today"> </trmf>
</lit>
</drs>
</sem>
</entry>

-<entry id="today2">

-<argx>

-<arg>
<var name="x0"/>
<dia name="du"/>
<avi cat="verb"/>
<avo/>

-<pavs>
<vp name="TT" in="t"/>
</pavs>
</arg>

-<arg>
<var name="x0"/>
<dia name="d"/>
<avi cat="noun" case="nom"/>
<avo/>
<pavs/>
</arg>

-<arg>
<var name="x0"/>
<dia name="d"/>
<avi cat="noun" case="acc"/>
<avo/>
```

```
<pavs/>
</arg>
</argx>

-<mor>

-<exp>

-<gst>
<pre/>
<text>Alyauma </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<mi/>
<mo/>
<hdl idref="post1"/>
</ma>

-<ma>
<mi/>
<mo/>
<hdl idref="post1"/>
</ma>

-<ma>
<mi/>
<mo/>
<hdl idref="post1"/>
</ma>
</marg>
</mor>
```

```
-<mor>

-<exp>

-<gst>
<pre/>
<text>Alyauma </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<mi/>
<mo/>
<hdl idref="pre1"/>
</ma>

-<ma>
<mi/>
<mo/>
<hdl idref="pre1"/>
</ma>

-<ma>
<mi/>
<mo/>
<hdl idref="pre1"/>
</ma>
</marg>
</mor>

-<sem>

-<drs head="">

-<lit>
```

```
<trmf fc="t is in today"> </trmf>
</lit>
</drs>
</sem>
</entry>

-<entry id="today3">

-<argx>

-<arg>
<var name="x0"/>
<dia name="du"/>
<avi cat="verb"/>
<avo/>

-<pavs>
<vp name="TT" in="t"/>
</pavs>
</arg>
</argx>

-<mor>

-<exp>

-<gst>
<pre/>
<text>Alyauma </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<mi/>
```

```
<mo/>
<hdl idref="pre1"/>
</ma>
</marg>
</mor>

-<sem>

-<drs head="">

-<lit>
<trmf fc="t is in today"> </trmf>
</lit>
</drs>
</sem>
</entry>
<!-- Verb to be -->

-<entry id="be">

-<argx>

-<arg>
<var name="e0"/>
<dia name="u"/>
<avi/>
<avo voice="act" mood="subj" tense="pres" asp="imperf" cat="verb"/>
<pavs/>
</arg>

-<arg>
<var name="x0"/>
<dia name="d"/>
<avi cat="noun" case="nom" def="def"/>
<avo/>
</arg>
```

```
-<arg>
<var name="x1"/>
<dia name="d"/>
<avi cat="noun" case="nom" def="indef"/>
<avo/>

-<pavs>
<vp name="PORP" in="p"/>
</pavs>
</arg>
</argx>

-<mor>

-<exp>

-<gst zero="0">
<pre/>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>

-<ma>
```

```
<hdl idref="post1"/>
<mi/>
<mo/>
</ma>
</marg>
<rank out="5" in="any"/>
</mor>

-<sem>

-<drs head="p+x0">

-<lit>

-<trmf fc="p">
<trmv name="x0"/>
</trmf>
</lit>
</drs>
</sem>
</entry>

-<entry id="Ali">

-<argx>

-<arg>
<var name="x0"/>
<dia name="u"/>
<avi/>
<avo cat="noun" case="nom" def="def"/>
<pavs/>
</arg>
</argx>

-<mor>
```

```
-<exp>

-<gst>
<pre/>
<text>Ali </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<mi/>
<mo/>
<hdl idref="post1"/>
</ma>
</marg>
</mor>

-<sem>

-<drs head="">

-<eq>
<trmv name="x0"/>
<trmf fc="Ali"/>
</eq>
</drs>
</sem>
</entry>

-<entry id="doctor">

-<argx>

-<arg>
<var name="x0"/>
<dia name="u"/>
```

```
<avi/>
<avo cat="noun" case="nom" def="indef"/>

-<pavs>
<vp name="PROP" out="p"/>
</pavs>
</arg>
</argx>

-<mor>

-<exp>

-<gst>
<pre/>
<text>tabIbuN </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<mi/>
<mo/>
<hdl idref="post1"/>
</ma>
</marg>
</mor>

-<sem>

-<drs head="">

-<eq>
<trmv name="doctor"/>
<trmf fc="p"/>
```

```
</eq>
</drs>
</sem>
</entry>

-<entry id="clever">

-<argx>

-<arg>
<var name="x0"/>
<dia name="du"/>
<avi cat="noun" case="nom"/>
<avo/>

-<pavs>
<vp name="PROP" out="q" in="p"/>
</pavs>
</arg>
</argx>

-<mor>

-<exp>

-<gst>
<pre/>
<text>mAhiruN </text>
<post/>
</gst>
</exp>

-<marg>

-<ma>
<mi/>
<mo/>
<hdl idref="post1"/>
```

```
</ma>
</marg>
</mor>

-<sem>

-<drs head="x0+p+q">

-<eq>
<trmv name="q"/>
<trmf fc="clever of p"/>
</eq>
</drs>
</sem>
</entry>
</dict>
```