The Integrative Processing of Natural Language: The case of Arabic Lexicon

Lazhar Zanned University of Manouba, Tunisia

Abstract

This paper deals with Lexicon in Arabic in the Probabilistic framework elaborated in Zanned (2005). The main results are the following: Lexicon is a network of phonological and conceptual connexions. It shows a multi-layered structuring at the phonological pole and the semantic pole. The paper deals also with NLP. The main claim is that syntax is not central to processing, instead it is based on an integrative simultaneous device. The final claim is that language evolution –in Arabic for instance- is based on diffusive interactive articulatory-cognitive staged processes driven by the necessity to achieve efficiency and distinctiveness in the communicative system at both articulatory and cognitive aspects. Modern languages kept traces of that co-evolution in their inherent structure.[1]

Keywords:

Archiroot
Arabic Lexicon
Lexical access
Probabilistic Model

[1] This article is originally a conference paper presented at *New Directions in Cognitive Linguistics*, 24-26 October 2005, University of Sussex at Brighton, Brighton, UK.

Introduction

The cognitive enterprise has been transforming the foundations of humanities for decades and rethinking many classic issues has been made possible through theoretical insights and scientific findings. Cognitive literature in general, and cognitive linguistics in particular, are becoming much important. New directions are emerging, and new theoretical models are being built to help understand how humans are what they are: intelligent creatures interacting by language use and many other means in complex social interrelationships.

The present paper deals with Natural Language Processing in the Probabilistic framework based on the Archiroot notion regarding the organization of the lexicon in Standard Arabic. It is divided into four parts: the first summarizes the basic principles of the Probabilistic Model. The focus in the following parts will be on the Archiroot implications: The phonetic-semantic bipolarity is dealt with in the second part where the multi-layered structuring of the phonological and the semantic aspects is presented. In the third part, an integrative simultaneous processing device is suggested. It is based on the lexical data in Arabic and deals with lexical access in perception and in production. Part four deals with some evolutionary issues: language is a system evolving to achieve efficiency and distinctiveness by diffusive cumulative processes in both articulatory and semantic aspects.

1- The Probabilistic Model: basic principles

The basic principles of the Probabilistic Model elaborated in Zanned (2005) are summarized as following: Roots in the Arabic lexicon are structures of ordered consonantal positions (slots). The number of positions determines the number of consonantal segments. The order of the consonants determines the identity of the root from both aspects: phonological and semantic. Roots are formed by chaotic combination of consonants from different articulatory places within the limits of the articulatory space. The combinations are established in language by continuous collective use, making up the conventional lexical roots repertoire.

Root formation is governed by the law of the fixed (F) and the variable (V). The value of F and V is one of 28 consonants. So, in theory, every listeme is made of 28 roots. The F element is the consonant which is common to the members of the same listeme, and the V element is the

consonant in which all the members of the listeme differ.

The formation of the triconsonantal root (C1C2C3), the most frequent in the Arabic lexicon, is based on the combination of Fs and Vs following three different types of schemes regarding the number and the order (root positional structure) of Fs and Vs in each one:

(1) Root formation schemes:

a. V1V2V3

The scheme (a) works with three Vs and makes all the triconsonantal roots. It represents the most general consonantal shape.

b. F1F2F3

The scheme (b) works with three Fs and makes a single root. That is the lexical entry as established in language use related to a particular meaning.

c.

c.1 V1V2F V1FV2 FV1V2

The three schemes in (c.1) work with two Vs and one F. They establish the phonetic network between all the roots. This network is based on the resemblance at the F element without any necessary semantic relations. It is the source of rhymes, crosswords, entries in dictionaries and language games, etc.

c.2 F1F2V F1VF2 VF1F2

The three schemes in (c.2) work with two Fs and one V. They are the productive schemes making root listemes. They are based on phonetic and semantic correspondences. Every scheme in c.2 works separately and it makes one root listeme following the values of the Fs and the Vs. However, the values of the Fs and Vs may coincide in the three listemes, giving a hybrid root which is made of three copies of one root. It is called an Archiroot. The Archiroot is the crossing point between three separate listemes as shown in (2) where the Archiroot (**ħzm)** is represented in its three listemes including all the possible 28 roots following the number of the consonantal segments in Arabic:

(2) Root listemes

schemes	F1F2V	F1VF2	VF1F2
values	ħz -	ħ - m	- z m
roots	ħzb	ħbm	bzm
	ħzm1	ħmm	mzm
	ħzw	ħmw	wzm
	ħzf	ħfm	fzm
	ħzθ	ħθm	θzm
	ħzð	ħðm	ðzm
	ħzđ	ħđm	đzm
	ħzt	ħtm	tzm
	ħzd	ħdm	dzm
	ħzţ	ħţm	ţzm
	ħzn	ħnm	nzm
	ħzs	ħsm	szm
	ħzz	**ħzm2**	zzm
	ħzş	ħşm	şzm
	ħzď	ħď'm	ď'zm
	ħzl	ħlm	lzm
	ħzr	ħrm	rzm
	ħzš	ħšm	šzm
	ħzj	ħjm	jzm
	ħzy	ħym	yzm
	ħzk	ħkm	kzm
	ħzq	ħqm	qzm
	ħzx	ħxm	xzm
	ħzγ	ħγm	γzm
	ħz9	ħ9m	9zm
	ħzħ	ħħm	**ħzm3**
	ħzh	ħhm	hzm
	ħz?	ħ?m	?zm

The Archiroot is represented in dictionaries as a single lexical root entry bearing many related or different meanings. It was considered in classical lexicography as polysemic. The root ħzm for example is represented as in (3):

(3) lexical entry: ħzm

-*ħazama*: to bind, to tie in the middle, using any kind of rope.

-*ħazima*: be distressed and affected in the heart.

-*ħazuma*: to be firm and sound, make good judgment.

-*ħaazim*: the quality of being intelligent, possessing firmness, soundness in judgment or knowledge and skill in affairs.

In the Probabilistic framework, a new different analysis is elaborated. It has many theoretical implications at the syntactic, morphological, lexical and semantic levels (see Zanned 2005): What was considered as one polysemic root, is in reality a multiple formation: three consonantal shapes collapsed in one and this made three different meanings relate to that shape. Collapsing is caused by the crossing between listemes which makes the Archiroot. The formation of the Archiroot ħzm is summarized in (4):

(4) Archiroot ħzm

Scheme	F1F2V	F1VF2	VF1F2
Values	ħ z -	ħ - m	- z m
Roots	ħzm1	ħzm2	ħzm3
Meanings	be distressed	soundness	to bind
Archiroot	ħzm		

The Archiroot is the crossing point between three listemes. It is made of three copies (1, 2 and 3). Each copy is a member of one listeme of roots sharing one general notion as represented in (5) where the data is collected from regular dictionaries:

(5) Root listemes

5-a

The first listeme is produced by the scheme F1F2V, where F1= ħ, F2= z and V any other consonant:

ħzb - *ħazaba*: to befall, to distress, press severely or heavily upon.

- *ħaazib*: a severe or distressing event.

ħzm1 - *ħazima* : be distressed, and affected in the heart.

ħzw - *ħazaa*: to compute something, determine by conjecture (quantity, measure) of the thing.

- *ħazaa-ssaraabu-al-šaxsa* : the mirage raised to the eye of the figure of a man or other thing seen from a distance.

ħzn - *ħazina*: to be affected with *ħuzn*: grief, mourning, sorrow/ to be sad, unhappy.

ħzz - *ħazza*, *ħaazzatun*: pain in the heart arising from wrath, fear, physical suffering.

ħzr - *ħazara*: compute by conjecture (quantity, measure, size, proportion, extent, amount,etc), by the eye.

hzl - *?iħza?alla –al-qalbu*: to fear, be distressed.

ħzq - *ħuzuqqun* : narrow in judgement.

5-b

The second listeme is produced by the scheme F1VF2, where F1= ħ, F2= m and V any other consonant:

ħwm - *ħaama*: to seek or plan to get something.

- *ħawwama –fii-al-?amri*: to pause and act with deliberation or in a patient or leisurely manner.

ħmm - *ħumma-al-?amru*: the affair is decided and established.

- *muħaammun*: the quality of someone who is keeping constantly, firmly, steadily, steadfastly or fixedly on an affair.

ħtm - *ħatama*: to decide (an affair or a thing) judicially, to establish or settle it.
to pronounce it; to necessitate it , make it necessary, requisite or unavoidable.

ħdm - ?*iħtadama*: to burn with anger , wrath or rage.

ħtm - *ħaţima*: to become broken in pieces/ *taħaţţama ɣayḍan* : to be broken with anger or rage.

ħsm - *ħasama-?al-?amra*: to decide an affair firmly and shortly.

ħzm2 - *ħazuma*: to make (a judgment, counsel) firm and sound, make good judgment.

- *ħaazim*: the quality of being intelligent, possessing firmness, soundness in judgment or knowledge and skill in affairs.

ħlm - *ħaluma*: treat with *ħilm* (forbearance or clemency, moderation, tranquility on the occasion of emotion or anger.

ħrm - *ħarima* : to persist obstinately, to persist in contention, litigation or wrangling.

ħšm -*?iħtašama*: to be or become confounded or stupefied by shame, and remain speechless and motionless.

ħjm - *?aħjama* : to refrain, forbear, abstain, and desist from.

- *ħijaam*: a muzzle for a camel, when he is excited to prevent his biting.

ħkm - *ħakuma*: to become wise / *ħakama*: to decide judicially/ to prevent, restrain, withhold from acting in an evil or corrupt manner.

ħikmat: wisdom./ *?aħkama*: to render firm, stable, strong, solid, sound or free from defect.

5-c

The third listeme is produced by the scheme VF1F2, where F1= z, F2= m and V any other consonant:

bzm -*bazïm*: leaves of the palmtrees used to tie vegetables in bundles.

wzm -*wazmun*: a bundle of grass, flowers/ bundle of meat that the eagle gathered in the nest.

lzm - lazima: to keep close to…/ hold fast to anything.

- *milzamatun* : vice (two pieces made of wood or iron to hold things tightly together).

rzm - *razama*: to collect together the thing in a garment or piece of cloth./ *razzama* : to collect and tie in bundles./ *rizmatun*: bundle put in one piece of cloth; things put together in a bundle.

jzm - *jazama*: make the oath to be unconditional, absolutely or irreversibly binding.

xzm - *xazama* : to pierce, to perforate namely a thing of any kind, for instance, the nose of a camel, for the ring called *xizaamat* to which the rein is attached.

xazama-širaka-na9lihi: to pierce and tie the thong of his sandal.

- *xazamun*: a certain kind of tree of the bark of which ropes are made.

ħzm 3 - *ħazama* : to bind, to tie.

9zm - *9azam*: to settle or determine his heart or mind firmly upon doing something.

hzm - *?al-hayzamu*: quality of being solid, strong and robust.

?zm - *?azama bi-hi* : he cleaved to him, namely his companion, or to it , namely a place.

-*?azama -al- ħabla*: to make a solid strong rope free from defect.

2- Phonetic-semantic bipolarity

The basic idea in Cognitive Linguistics is that Language is a form of cognition. It was created by convention to allow symbolic communication between humans. Grammar is a vast inventory of conventional linguistic units, structured in hierarchies that overlap and intersect on a massive scale (Langacker 1987, 1991). Grammar is symbolic (Langacker 1987:81). A symbolic unit is bipolar: it is the association between a phonological pole and a semantic pole. Structures in CG are of three types: phonological, semantic, and symbolic. Symbolic structures combine phonological and semantic structures. Each one of these three structures makes a space:

Semantic space is defined as a multifaceted field of conceptual potential within which

thought, and conceptualization unfold.

A phonological structure is part of the phonological space conceived as being the phonic potential (articulatory potential). For example, a vowel chart (high-low, front-back, parameters or features) is a domain in phonological space; a vowel articulation is a location within this range.

The symbolic space is the coordination of the two poles (structures, spaces). Since the symbolic structure is semantic at one pole, phonological at another pole, it is the correspondence between them. Establishing correspondences is essential to cognitive processing. A grammatical construction is based on a global correspondence which may be resolved into local correspondences connecting many individual points or subparts of the global structure. (Langacker 1987:90). Minimal symbolic units combine in a successive way to form progressively larger symbolic expressions (Langacker 1987:97).

The following section deals with how symbolic units are construed in Arabic in the light of a probabilistic framework.

2-1 Phonological pole: a multi-layered structuring

The phonetic-semantic correspondence (symbolization) in Arabic is established at two levels, following the rules governing word formation. The first is the consonantal root level where the correspondence is established between the root and the lexical meaning (concept). The second is the morphological level where the correspondence is established between the syllabic schema (vowels and special derivational or inflectional morphemes, etc.) and their morphological and/or syntactic values and categories.

The focus in this section will be on the first level. The correspondence at this level is a multi-layered one. The layering is conceived as interconnected, hierarchical and parallel working subcomponents of the lexical root formation module in Grammar:

The first layer is represented by the Archiroot, where the correspondence is established between the triconsonantal formation and all the different (un)related meanings.

The second layer is represented by the root listemes, which are three, in general, following the triconsonantal schemata. Each listeme represents all the phonological (articulatory) domain in the articulatory space framed between labials and gutturals. This phonological articulatory domain corresponds to the semantic domain that is the general notion shared between the different roots in the listeme (see section 2.2).

The third layer is represented by the root(s) containing, at the F-position, the consonant(s)

belonging to the different phonological (articulatory) regions: labial, dental, palatal, velar, laryngeal and guttural. Each region at the phonological pole corresponds to a semantic region in the semantic domain (see table (10)).

The fourth layer is represented by the single root which is one of three copies making the Archiroot as the listemes cross. At this level the morphological derivation applies through the correspondence between the lexical semantic features (process, property, state, number of participants, etc.) and the morpho-syntactic values. At this stage the copies of the Archiroot are distinguished and they make separate entries. The distinction is clear and, starting from the verbal form, it spreads all over the derived nominal forms related to that verb, in a way that every copy (root) has its own derivational network. That is the perfect state of the system illustrated by the Archiroot ħzm, but it is not always the case when it comes to deal with other samples in the data following the semantic/conceptual peculiarities (Zanned 2005).

The meaning of the root ħzm3 is related to the dynamic process of *binding* something operated by an agent using an instrument following a determined way in doing so (in the middle). This concept is inherited from its listeme domain. The verbal perfective template corresponding to dynamic processes in Arabic Morphology is C1aC2aC3:

Root copy	ħzm3	: binding, agent, thing, instrument, in the middle.
Verb template	C1aC2aC3	: dynamic process, agent, perfective, past.
Word	*ħazama*	: to bind, to tie in the middle, using any rope.

The meaning of the root ħzm1, inherited from its listeme domain, is related to the static experience of distress which involves an experiencer. The verbal perfective form corresponding to static experiences in Arabic is C1aC2iC3:

Root copy	ħzm1	: experience of distress, experiencer.
Verb template	C1aC2iC3	: static experience, experiencer, perfective, past.
Word	*ħazima*	: be distressed, affected in the heart

The property of soundness related to ħzm2 involves a (zero) agent or experiencer, has its

correspondent verbal perfective form in C1aC2uC3 which expresses a static state or quality, with (zero) agent:

Root copy ħzm2 : quality of soundness, intelligence and firmness.
Verb template C1aC2uC3 : static state or quality, (zero) agent, perfective, past.
Word *ħazuma*: to be intelligent, possessing firmness, soundness in judgment or knowledge and skill in affairs.

In summary, the phonological pole is structured as following:

(6)

- Root consonantal level:
  Archiroot
  Root listeme phonological domain: articulatory space (labial-guttural)
  Root listeme phonological regions : articulatory regions
  Root phonological copies: single copies
- Morphological template level:
  Morphological schemes
  Prosodic schemes: metrical structure, intonation contour, pitch, stress…

2-2 Semantic pole: a multi-layered structuring

Human cognition is based on domains (Langacker 1987:149). They are coherent areas of conceptualization relative to which semantic units may be characterized. Domains are basic and abstract. Basic domains are primitive since they are the fundamental dimension of human experience. Abstract domains are conceptual complexes that may serve as a domain for a higher order concept. A region is a set of mutually interconnected entities that may constitute a higher order concept.

Domains and regions may be bounded or unbounded. Bounded domains or regions are made of limited entities in a finite set. Conceptual structure is a hierarchical arrangement built on discrete features and functions. It encodes such aspects of understanding as category membership (taxonomy) and predicate-argument structure (Jackendoff 2002:346).

Following the basic results elaborated in the previous sections, a root listeme is a finite set at both aspects: phonological and semantic. The phonological continuum in the V-element

corresponds to the semantic continuum:

The semantic aspect of the whole listeme represents a semantic/conceptual domain defined as the most general notion shared between the different roots belonging to the listeme:

(7)

scheme	F-values	attested roots	listeme domain
F1F2V	ħ z –	see 5-a	Emotion, pain, distress
F1VF2	ħ – m	5-b	Wisdom and soundness
VF1F2	– z m	5-c	Binding

The domain corresponding to the listeme is divided into regions which correspond to the phonological articulatory regions. At this level the boundaries between the regions are not necessarily sharp and well defined and the phonological articulatory boundaries may not coincide with the conceptual boundaries (see table 10).

The set of phonological-semantic regions may overlap through time and language use. This may be triggered by the nature of linguistic meaning and by the laws of linguistic change (phonological and semantic processes).

The meaning related to a particular root in a listeme represents a particular facet of the region. It is the basic level where the minimal pairing occurs. Every root represents a coherent cognitive unit. It grasps a particular slice of the experience included in the immediate semantic region which is part of the maximal semantic domain (listeme domain). The slicing occurs without any loss in the region or domain coherence and continuity.

Lexical roots are preset symbolic resources, conventionalized over time to afford a variety of windowing (Talmy 1996), grounding, or viewing (Langacker) frames. This allows the speaker/hearer to find her perspective(s) while communicating and interacting with other people. “Some categories embodied in language…reflect the whole system of options open in particular communicative situations” (Tomasello 1999:166).

The conceptual domain of the listeme (5-b) is “wisdom and soundness”. It is divided into many regions following the phonological-semantic correspondence. The first is the labial region where the values of the Fs are (w) and (m). To this phonological region corresponds the semantic region of “act with deliberation”, which is divided into two facets: the first is related to the root ħwm indicating *planning to achieve an act* (an affair), the second is related to the root ħmm

indicating the *attitude of the agent while working on a affair*. Both facets are complementary: the target is supposed to be one general affair and the root *ħwm* represents the mental state of *planning* before acting. The root ħmm represents the attitude of constancy and steadiness during the process:

ħwm - *ħaama*: to *seek* or *plan* to get something.

- *ħawwama–fii-al-ʔamri*: to *pause* and *act* with *deliberation* or in a *patient* or *leisurely* manner.

ħmm: - *muħaammun*: (adj. m. s.) the quality of someone who is *keeping constantly, firmly, steadily, steadfastly or fixedly on an affair.*

The dental region is divided into two subregions: the [+stop] region and the [–stop, +sibilant] region:

The dental stops (t, d, ţ) region corresponds to the semantic region of "anger". Following the distinctive features, the notion of "anger" is two-graded: a grade of "burning anger" related to the [+voiced, –emphatic] (d) and a grade of "be broken with anger" related to the [–voiced, –emphatic] (ţ):

ħdm - *ʔiħtadama* : (verb, past), to *burn* with *anger , wrath or rage*.

ħţm - *taħaţţama γayḍan* : He is *broken* with *anger or rage.*

The [dental, –stop, +sibilant] region ((s), (z)) corresponds to the "firmness in mental attitude" conceptual region. The firmness is applied to decision making in a short manner in the root *ħsm* with the [–stop, –voiced,+sibilant, –emphatic] (s). And it is applied to judgement making in an intelligent or sound manner in the root *ħzm* with the [–stop, +voiced , +sibilant, –emphatic] (z):

ħsm - *ħasama-ʔal-ʔamra* : to *decide* an affair *firmly* and *shortly*.

ħzm2 - *ħazuma* : to make (*a judgment, counsel*) *firm* and *sound*.

The boundaries between (articulatory-conceptual) regions are not sharp. The root ħtm from

the dental region with the [+stop, –sonorant, –emphatic] (t), for example, should be in theory part of the semantic region of "anger". Instead, it is part of the "firmness in mental attitude" conceptual region which is related to the [dental, –stop, +sibilant] region ((s), (z)). So it corresponds to a facet of the decision making judicially or in an unavoidable way:

ħtm - *ħatama* : to *decide (*an affair or a thing*) judicially*, make it necessary, requisite or *unavoidable*.

The maximal domain related to the listeme is a set of interrelated notions. It is conceived as being the most general semantic units common to all the roots in the listeme. At the level of the phonological-semantic boundaries, a primary division occurs to distinguish the regions: subsets of the maximal domain. These subsets represent the immediate scope for the single root where many cognitive-semantic processes such as selection, profiling and specification operate. Following Langacker (2004) a lexical item represents a conventional way of accessing a set of domains: some domains are central to its value and others peripheral. Centrality may be conceived in the Arabic lexicon as gradual: the central semantic features are those proper to one single root and the peripheral ones are those shared between a set of roots included in a phonological-semantic region.

One of the listemes including the Archiroot (njr), produced by the scheme F1VF2 with F1= n, F2= r and V any other consonant, is related to the maximal domain <*piercing-spreading-carving*>. The phonological region [+back] (velars, laryngeals)) corresponds to a particular semantic region, that is < *piercing-spreading-carving*> applied to a *liquid (water, blood) issuing from a (tubular) container*. So, we have the following structuring, where the facets of the maximal domain are selected and profiled differently in different roots in a gradual way (see Zanned 2005 for more details):

F1VF2 = n-r <piercing-spreading-carving>:

[+velar] corresponds to <*issuing blood* by *piercing* or the *decay* of the container>:

nkr - *nakirat:* (n., f., s.), *blood, pus* etc. *issuing* from the *body*.
nqr - *naqara* : (verb, past), *pierce through*, pierce with the *beak*.
nxr - *naxira* : (verb, past), be *carious* (bone, etc).

[+laryngeal] corresponds to < running liquid (blood)>:

nγr - *naγara* : (verb, past), make *the blood* run out *abundantly*.
n9r - *na9ara* : (verb, past), make the *blood* to spurt out with *a noise*.
nħr - *naħara* : (verb, past), *slaughter* a camel, *wound at the collar bone*.
nhr - *nahara* : (verb, past), the *blood* running with *force, abundantly*.

Selection and delimitation are preset in a schematic format at the level of the root listeme, domain, or region. A single root meaning in a region is related to a particular scope, view or selection (Langacker 2004). The terms in italics summarize these processes: the blood is just issuing from the container with no noise, almost spontaneously in (nkr, nqr, nxr). However, it is running abundantly in (nγr), with force in (nhr), spurting out noisily in (n9r), and in all of these ways in (nħr) by slaughtering.

The lexical semantic pole has the same multilayered structuring as the phonological pole:

(8)
Archiroot meanings
Root listeme semantic domain
Root semantic regions
Root copies meanings

Morphological categories
Prosodic values

The multilayered structuring at both phonological and semantic poles is summarized in (9) where the correspondence is established in a one-to-one fashion:

(9) multilayered structuring in root listemes in the Arabic lexicon:

Symbolic consonantal unit	
Phonological pole	Semantic pole
Archiroot consonants	Archiroot meanings
root listemes domains	semantic domains
root listemes regions	semantic regions
root copies	root copies meanings
morphological schemes	morphological categories
prosodic schemes	prosodic values

The following table represents the correspondence between both poles as applied to the *ħzm3* copy formed by the scheme VF1F2 where F1=z and F2=m. The corresponding consonant at the phonological pole is based on the V-element in the listeme:

Table (10): Phonological- semantic correspondence in the Archiroot framework:

PHONOLOGICAL ARTICULATORY DOMAIN							
							CONCEPTUAL DOMAIN : *BINDING*
	LABIAL	b	*bazïm*: leaves (palmtrees) to tie vegetables in bundles.	Grass	Vegetal rope	IN BUNDLES	
		w	*wazmun*: a bundle of grass, flowers.				
			wazmun: bundle of meat gathered by the eagle in the nest.	Meat			
	DENTAL	r	razama: to collect things together in a garment or piece of cloth.		Cloth		
		l	*milzamatun* : vice (two pieces made of wood or iron to hold things tightly together)		Two pieces		
			lazima: to keep close to…/ hold fast to anything		thing	INTENTIONALLY	
	PALATAL	j	*jazama*: make the oath to be unconditional, absolutely or irreversibly binding.		OATH		
	VELAR	x	- to pierce, to perforate anything, for instance, the nose of a camel, for the ring called *xizaamat* to which the rain is attached. - *xazama-širaka-na9lihi* : to pierce and tie the thong of his sandal.			BY PIERCING	
	GUTTURAL	ħ	*ħazama* : to bind, to tie in the middle, elongated object, instrument : rope			In middle	
		9	-*9uzmatun* : a man's family (?usrat), or near kinsmen on the father's side and his tribe.			By blood	
			9azam: determine his heart or mind firmly upon doing s.thing.			FIRMLY	
	laryngeal	h	*?al-hayzamu*: quality of being solid, strong and robust				
		?	- *?azama bi-hi* : he cleaved to him, namely his companion, or to it , namely a place.				

3- Integrative processing

The focus in Cognitive science is on the computational aspect of the mind (Simon and Kaplan 1989). The general assumption is that the structure of the brain is modular and the cognitive abilities (vision, language, etc) work independently of one another in separate modules (Fodor 1983**).**

However recent studies argue that the brain works in a holistic way: all its parts participate in cognitive processing simultaneously through different mental processors following neural combinations. Any cognitive processing is the terminal stage of a distributed computation (Deacon 1998, Jackendoff 2002, Besson and Schon 2001). Language functions extend into all major lobes of the neocortex (temporal area (auditory), parietal area (tactile), frontal cortex (attention, working memory, planning) of the left hemisphere. This wide distribution suggests a far less localized language system (Deacon 1998: 290-291). That system, like any other system, is continuously active. Its activity is shared between activating signals and inhibitory signals where any cognitive element is made of a large number of neurons distributed over large regions of the brain.

In the literature, a distinction is made between two kinds of processes: encoding (getting information in) and retrieval (getting information out). The basic results of the modern brain imaging techniques is that the left hemisphere (left prefrontal cortex) does more for encoding information into episodic memory and in its retrieval from semantic memory, whereas the right hemisphere (right prefrontal cortex) is more involved in the retrieval from episodic memory. This fact holds for verbal and nonverbal materials (Tulving et al. 1994, Tulving and Lepage 2000, Cabeza et al.1997).

Speaking involves automatic and controlled processes. Controlled processes are related to the communicative intentions which may vary in infinite ways. The conceptualizer (Levelt 1989:9) employs highly controlled processing since there are many alternative means of expression. However, the other processes involved in formulating and articulating the message are highly automatic and take place at a very high speed: the rate of lexical choice and the articulation is two to three words per second. Automatic processes run in a modular fashion: parallel without mutual interference (Levelt 1989:21-22).

Levelt (1989) suggests that a theory of lexical access (in perception and in production) should be judged on two criteria: the speed of access and convergence. Theories based on sequential

processing fail on the first criterion since the speed of access is very high in fluent speech. Various speech errors are a strong argument for a parallel processing. Accessing the correct item among many possible candidates should be dealt with overtly by presenting all the processes leading to the right choice of a particular item to express a particular concept.

Entries in the mental lexicon are interrelated according to both meaning and form. Lexical access in naming objects has three processing stages or levels: visual processing, conceptual categorization, lexical access proper. The speed of lexical access depends on the frequency of its usage (Levelt 1989:231).

3-1 Verbal production: the integrative lexical access

Studies on verbal production mostly focus on non-combinatorial aspects like picture naming (Caramazza and Miozzo 1997; Engelkamp and Rummer 1998). In this kind of task, verbal production is based on activation starting from a visual stimulus and ending at a phonological selection. Two stages are at work: selection of lemmas (Levelt) and selection of word forms or lexemes. Considering that verbal production is not limited to single tasks (naming things or uttering single words), Jackendoff (2002: 211-214) deals with production in a more general frame based on combinatoriality.

He argues that verbal production is a four staged process:

(a)- Thought elaboration: A thought is elaborated in the conceptual system of working memory.

(b)- Call to the lexicon: In order to express this thought, a call to the lexicon is needed to select the word(s) that potentially express(es) it.

(c)- Lexical activation and selection: Many candidates are activated and the item which best matches the intended thought is selected and its structure is bound to working memory.

(d)- Integration: after the activation and selection occur, the item’s syntactic and phonological structures follow shortly and bind to their respective "blackboards". Phonological integration is achieved when the syntactic and morphological integration is completed: syntax determines the order of words and their morphological forms, this determines the order of phonological integration and morphological forms of words to be selected in the light of the syntactic context.

These four stages might be admitted as a broad general frame for verbal production processes in Natural Languages. However, the data from the Arabic lexicon may present some peculiarities related to the nature of its phonological and morphological system, word formation rules and syntactic component which are governed by nonconcatenative processes (McCarthy 1979, 1981, Zanned 2005).

In the probabilistic framework, the Archiroot is the intersection where three different copies of roots coincide, each copy is produced by a particular scheme of c-2 as presented in (1).

The conceptual system elaborates a particular concept (event, action, state, participant(s), time, place, etc). Let it be the action of *binding something in the middle by any kind of rope* or the like, which took place in *the past*. This conceptual set includes two kinds of information packs: the act package (Action: binding (including lexical aspect), Manner: middle, Instrument: rope, Object: elongated thing), the tense-aspect package (past, perfective).

The act package would activate the consonantal roots lexicon in the long term memory, where it gets to the Archiroot (**ħzm).** The tense-aspect package activates the verbal template (trilateral one) in the morphological system where the prosodic templates are stored. Both activations occur simultaneously.

The Archiroot instantiated in three copies (**ħzm1, ħzm2, ħzm3)** presents the selector with three candidates at the phonological consonantal pole, and three verbal stems (C1aC2aC3, C1aC2iC3, C1aC2uC3) at the morphophonological pole through the canonical verbal perfective stem (C1VC2VC3). Then, three semantic sets, three verbal stems are activated and two candidates of each set should be inhibited through the process of selection: inhibition at the semantic level occurs simultaneously with the verbal stem inhibition in a kind of echoing and selection reinforcement:

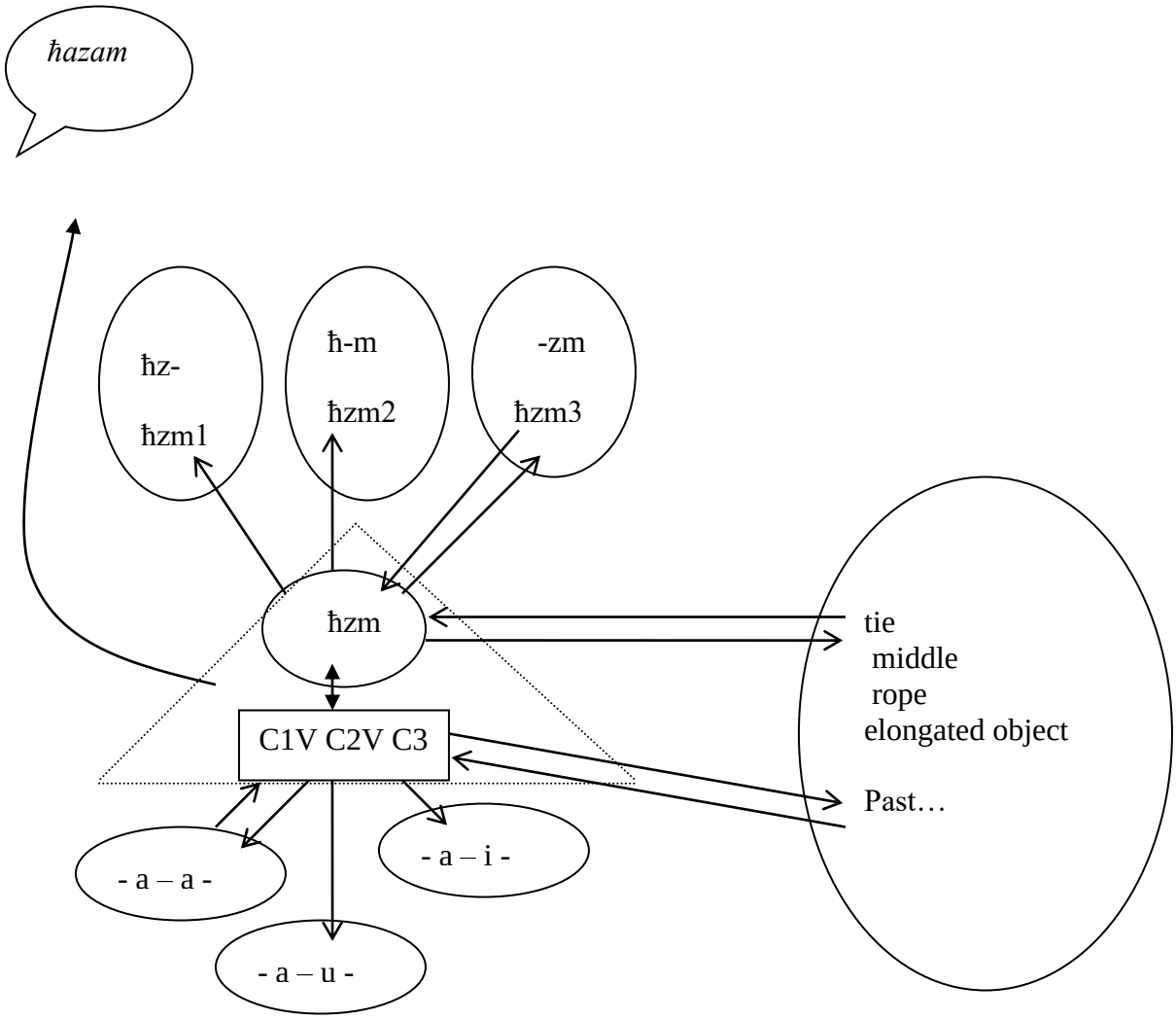
The candidate **ħzm3,** in the case of our sample, is selected by the response going back all the way, simultaneously to the conceptual system matching the candidate with the conceptual content.

The verbal perfective stem (C1aC2aC3) is selected by a response going back to the conceptual system matching the candidate with the tense-aspect features.

Meanwhile, the integrative processor matches between both components of the phonological representation (**ħzm,** C1aC2aC3) giving the preverbal form (**ħazam**) which represents the phonetic articulatory plan for the articulator.

The preverbal form (***ħazam***) is an online construction and its insertion into larger messages (sentences) follows the same processes where the integrative system in the working memory represents the final "blackboard" – following Jackendoff. At this stage, the phonetic plan is established and articulation occurs through the interface with the articulatory system. The articulator converts sequenced phonological segments (bundles of features) to sounds following the sensory- motor instructions.

(11) The Archiroot in verbal production:



3-2 Verbal perception :the integrative lexical access

Following Jackendoff (2002), the auditory-phonological interface presents a sequenced phonetic structure to the phonology. Lexical access cannot be a serial search. It seems that there is no "central executive," processing center, the brain itself is thoroughly decentralized. (Jackendoff 2002: 207). The processing architecture is based on four components: Activation, Binding (or copying), Integration and Resolution:

Activation of a lexical item in long-term memory is accomplished by a call from phonological working memory to the item's phonological structure. It may also be accomplished by spreading activation from associated items within the lexicon (priming). The auditory-phonological interface presents a sequenced phonetic structure to the phonology which establishes the syllabic structure and word boundaries. Lexical entries can be matched with the phonology "blackboard" only through their phonological structure. Binding (or copying) selects and sends an item from the long-term memory to working memory. When a lexical item establishes a phonological connection to working memory, its syntactic and conceptual structures are activated and connected to the appropriate departments of working memory. Integration is the process of combining an item on the "blackboard" with a larger structure being built. Resolution is the integration of the collected multiple processes, during which the multiple possibilities for integration are checked. Full Integration is achieved after the results are communicated from syntactic integration through the syntactic-conceptual interface.

Following the parallel processing, Jackendoff (2002:208) argues that contextual information is part of the conceptual system, thus it is not linked to the phonology blackboard, and has no role in selecting a phonological match.

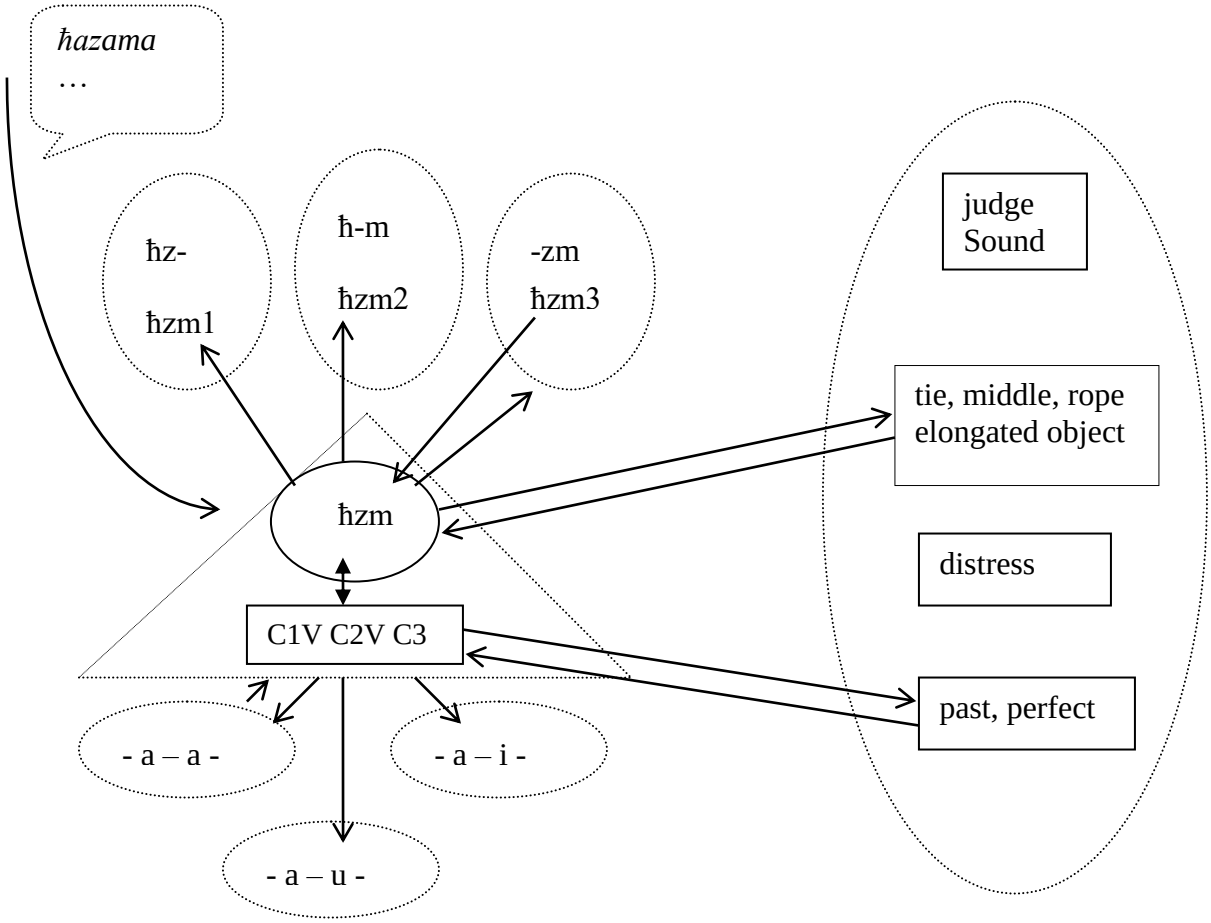
In the probabilistic framework, language perception is conceived as following:

The auditory-phonological interface presents a sequenced phonetic structure (***ħazam***) in our example, to the phonology. The Archiroot consonantal shape (**ħzm)** activates the root copies and presents the selector with three candidates at the phonological consonantal pole: (**ħzm1, ħzm2, ħzm3)**. The vocalic melody shape (— a — a —), activates three verbal stems (C1aC2aC3), (C1aC2uC3), (C1aC2iC3), in the morphology through the canonical tri-literal perfective verbal form (C1VC2VC3).

The conceptual packs related to the different root copies of the Archiroot and the tense-aspect pack, in the conceptual system, are activated simultaneously. At this stage the activation

is spread all over the different subsystems, presenting the integrative processor with a variety of possible candidates established by neural connections through language use. Selection occurs through the inhibitory responses or the lack of reinforcement signals: the different candidates are checked. The matching candidates are selected and those mismatching are eliminated. The integrative processor matches between different related corresponding candidates representing the different phonological, morphological, syntactic and semantic facets: The integrative processor receives reinforcement signals simultaneously from the root consonantal store related to the copy (**ħzm3**), from the verbal template store (C1aC2aC3), from the conceptual system (Action: tie, Manner: middle, Instrument: rope, Object: elongated thing) and (Tense: past, Aspect: perfective). This analysis is limited to one sample (a verbal form), but it may be extended to the sentence level, where other issues should be dealt with.

(12) The Archiroot in verbal perception:



3-3 The integrative simultaneous processing:

In this section, the focus is on the existence of a particular integrative system in cognition, in general, and in language processing, in particular. The integrative system is conceived as the central device in the architecture of NLP. The basic arguments rely on some established principles in cognitive linguistics and on some results established in the probabilistic cognitive model in dealing with the Arabic lexicon and Grammar based on the notion of the Archiroot.

Levelt (1989:232) states that not all the words a speaker uses are stored in the lexicon. Words can be newly constructed on the spot. But languages differ enormously in the amount of lexical encoding they require. Speakers of English rely to a great extent on their store of frequently used words and idioms. The same may be true for construing "utterances" where some sentences are newly constructed on line and others are stored in the memory and become routines in both aspects: cognitive and neural, such as proverbs, collocations, idiomatic expressions, etc.

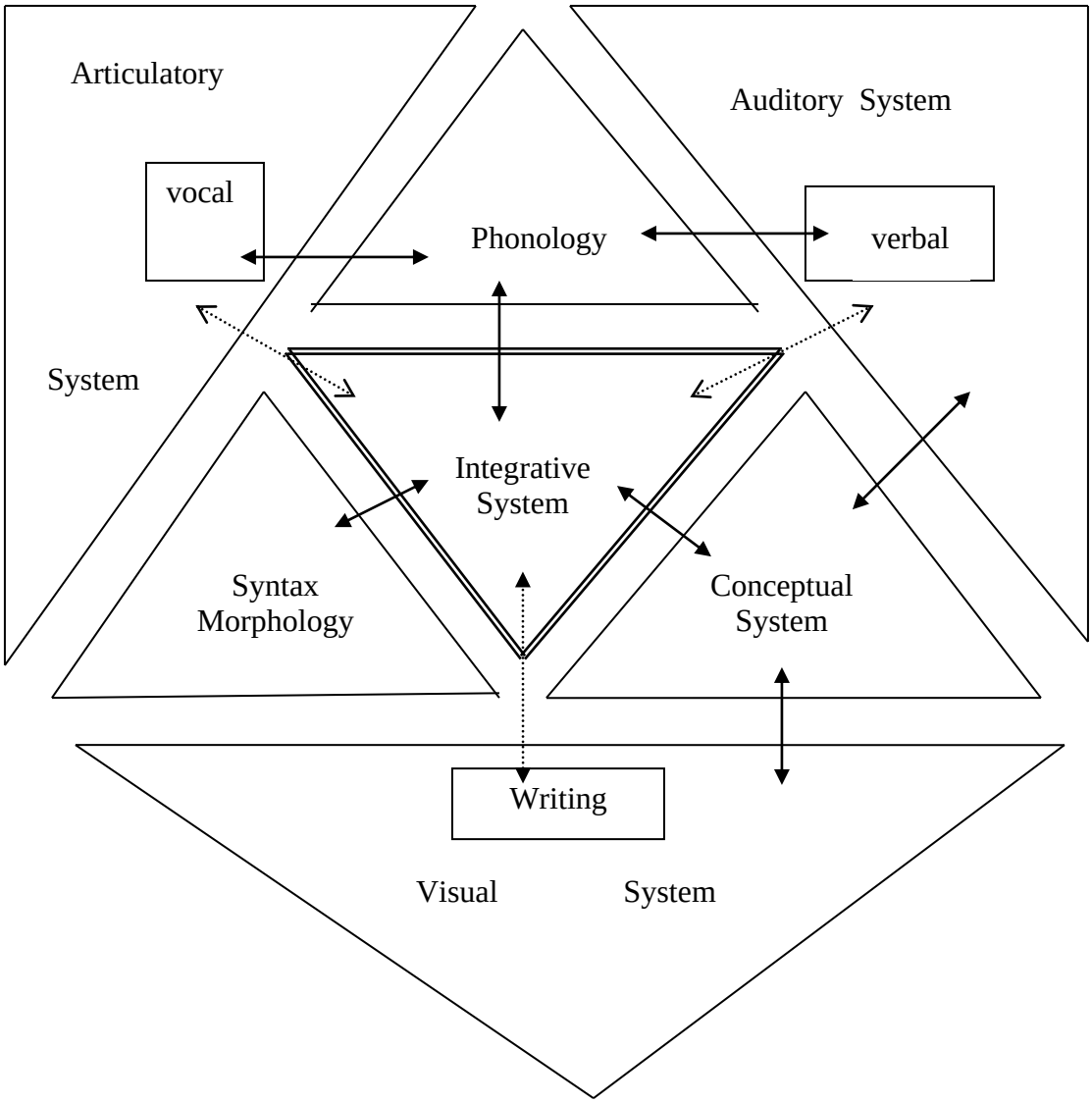
Neural activity may be localized in the brain space or map but cognition is not necessarily situated in a particular region. The functioning of the brain has an organic neural activity and a cognitive activity, the latter being the result of the former. However, linking both of them can not be achieved in an overt fashion. The neural activity has chemical, electrophysiological aspects (traces, manifestations) that may be observed, measured, analyzed and recorded, but the cognitive activity has a symbolic manifestation. Describing the cognitive processing requires relying on the results established in the brain sciences which are making gigantic progress in the organic functional aspect but they are still of "limited" help for linguistics and other disciplines working on symbolic manifestations of cognition.

Grammar is an inherent part of cognitive abilities, and so are its components (phonology, syntax, morphology). Systems are interrelated independent modules based on simultaneity and convergence. Simultaneity requires parallelism and convergence requires a special integrative module where all the particular tasks achieved by a particular module are monitored and constructed in a coherent mass making a cognitive set or whole. Every module is linked to other modules by interfaces that represent the correspondences and the pathways of activating and inhibitory signals through established neural networks.

The integrative system is responsible for checking, matching, and elaborating processes in perception and in production. In the integrative view, syntax is not central to the linguistic processor. Instead, the integrative system is the central part of the working mind and the terminal

and last step in both perception and production processing. The whole architecture is as follows:

(13) Processing architecture in the Arabic lexicon:



4- Evolutionary issues

In this section, the emergence of the phonological combinatorial system is dealt with at the level of the basic consonantal component.

Following Pinker and Bloom (1990), three distinctive features of the humans made the species evolution: know-how, sociality (culture, expertise) and language. Language is a symbolically embodied social institution that developed from previous existing social-

communicative activities (Tomasello 1999:94). Linguistic reference is a social act of interaction: a speaker and a hearer are directing jointly and intentionally their attention to a third thing (the referent) in the world (Tomasello 1999) . The basic function of language is related to social interaction that takes place in a variety of situations involving cooperation between at least two persons.

"Knowing how something originated often is the best clue to how it works" (Deacon1997:23). To establish the steps that language evolution took in its articulatory cognitive and grammatical aspects one can "look within modern language for traces of its past" (Jackendoff 2002:233). It is generally assumed that the evolution of language went through two major stages: the protolanguage and the modern language, where the main difference between the two is the lack of syntax at the first stage (Bickerton 1990). Modern language is built on the protolanguage. Jackendoff (2002:238) summarizes the emergence of language and its evolutionary course in three major phases (pre-protolanguage, protolanguage, modern language) divided into nine steps, some are sequentially ordered, others are parallel: the initial step is the pre-existing primate conceptual structure, followed by the use of symbols in a non-situation-specific fashion which unfold into two parallel steps: (a)- the use of an open, unlimited class of symbols and (b)-the concatenation of symbols. The step (a) is followed by the development of phonological combinatorial system which took place to enlarge the open, unlimited class of symbols, starting, possibly, from first syllables and making distinctions between phonemes. The step (b) is followed by the use of symbol position to convey basic semantic relations. At this level protolanguage is set. The following steps are related to the emergence of hierarchical structuring to convey semantic and grammatical relations by using a variety of symbols, inflections and functions which convey semantic relations.

Combinatoriality is a way of analyzing meaningful symbols and producing new ones in terms of a repertoire of smaller meaningless units (Jackendoff 2002:246). The innovation of phonological structure is a major cognitive advance: concatenation and combination of inherently meaningless phonological units led to an unlimited class of words.

Consonantal combinatoriality is parallel to other phonological, morphological and syntactic aspects based on combination and linear ordering to convey meanings. This evolution opens large possibilities to enhance expressive power and precision. Words (lexical items) are rarely used or heard in isolation. They occur in the flow of speech, inside some complex structures (utterances).

Lexical roots are considered as an extraction of simple consonantal components. This process occurs in language acquisition, since children rarely hear a word in isolation, and they have to extract it from the speech flow (Tomasello 1999:134-135).

Lexical Roots are preset symbolic resources, conventionalized over time to afford a variety of windowing (grounding, viewing) frames when expressing events allowing every speaker/hearer to find his/her perspective(s) while communicating and interacting with other people.

In the probabilistic framework, the formation processes of the listemes based on Fs and Vs, offer a basic argument for evolutionary expanding paths in the language faculty: the human cognitive system evolves in parallel and in an interactive way with the evolution of articulatory abilities. In other terms, the semantic conceptual space evolves in the same way with the phonological articulatory space. Mammalian species categorize objects on the basis of perceptual similarities, in the sense that they treat as similar all phenomena that are identical for some perceptual or motor purposes. Thus they construct categories and image schemas of dynamic events, for example, and internalize their representations in order to be used in a collaborative communicative situation (Tomasello and Call 1997, Tomasello 1999).

Similar events have similar representations corresponding to similar phonetic forms. Similarity may be general, partial or total. These grades are confirmed through the listemes of roots in the Arabic lexicon.

The general similarity between events or concepts corresponds to a general similarity in their mental representation(s), to which corresponds a general similarity in the phonetic form(s) conveying them. The binding domain, for instance, (see table 10), includes many similar events at the general level. To this general domain corresponds the phonetic representation made of the two Fs in their position in the triconsonantal schema in the root listeme.

Partially similar events have partially similar representations which correspond to partially similar phonetic forms. In the binding domain, the partially similar events are grouped in regions corresponding to partially similar phonetic regions represented by the Vs in their position in the triconsonantal schema in the root listeme.

Totally similar events have totally similar representations which correspond to totally similar phonetic forms. Total similarity may be conceived as identity, and this corresponds to a particular phonetic form which is the basic principle in the symbolic bipolarity (sign-meaning) in

the Saussurian conception. This fact is represented by the particular root consonants as established in language use: a particular event corresponds to a particular root. (see scheme (b) in (1)).

Another basic principle is confirmed in the probabilistic view: communicative systems evolve to achieve efficiency. Efficiency is shown in both aspects: semantic precision and phonological distinctiveness.

It is difficult to determine the starting point of the diffusive wave in the phonological and/ or the semantic aspects: if we take any listeme, there is no certainty of indicating which place of articulation at the V-position, which root or which concept was the beginning of the whole. The starting point is chaotic then it is conceived as being determined by coincidence. It is also difficult to establish any directional ordered stages in this process: different remote places in separate articulatory domains or regions may be the starting point of a local diffusive movement which can be pushed to its limits until it meets the boundaries of the neighbouring articulatory domains or regions. The parallel may be established with the semantic level. The main point is that at the end of the evolutionary stages, the correspondence between phonological articulatory aspect and the conceptual semantic aspect is established in a coherent way:

Articulatory space > domain > region > place (consonant)

Semantic space > domain > region > particular meaning

Diffusiveness affects mainly the V(ariable) elements since the fixed (F1 and F2) are established and given. It is the result of the cumulative cultural evolution conceived as invention and social transmission with improvements over time (Tomasello 1999: 5).

Conclusion

Making new claims may disturb the established conceptions regardless their nature. A new claim in humanities is always a hypothesis, and any hypothesis needs to be confirmed and infirmed through interacting argumentations. The broadest claim in the probabilistic model is to suggest a new analysis of Arabic Language, but the only aspect dealt with in this paper is the lexical one. The expected presentation should focus on Grammar and follow its hierarchy, but instead the choice was on dealing with issues related to the lexical component: phonological and semantic organization, lexical processing, and the formation of lexical roots in language

evolution. The issues are broad and the space of the paper is limited, so the presentation is limited and the original contributions may be shaded. The main results are three: the mental lexicon in Arabic is a consonantal and semantic multilayered network, the lexical processing is based on a central integrative device based on simultaneity and convergence, the root formation in Arabic language evolution was diffusive at the articulatory and the semantic/conceptual aspects. The expected part, which is missing in this paper, should follow as the model gets its way.

References:

Besson, M. and Schon, D. 2001. Comparison between language and music: The Biological Foundations of Music. *Annals of New York Academy of Sciences*, New York, 232-258.

Bickerton, D. 1990. *Language and Species*. Chicago: University of Chicago Press.

Cabeza, R., Kapur, S., Craik, F. I. M., McIntosh, A. R., Houle, S., and Tulving, E. 1997. Functional neuroanatomy of recall and recognition: A PET study of episodic memory. *Journal of Cognitive Neuroscience* 9: 254-265.

Caramazza, A. and Miozzo, M.. 1997. The Relation between Syntactic and Phonological Knowledge in Lexical Access: Evidence from the "Tip-of-the-Tongue" Phenomenon. *Cognition* 64: 309-343.

Deacon, T. W. 1998. *The Symbolic Species, The co-evolution of Language and the Brain*. New York, London : Norton & Company.

Engelkamp, J. and Rummer, R. 1998. The Architecture of the Mental Lexicon. In *Language Comprehension: A Biological Perspective*, A. Friederici (ed.). second edition, 133-174, Berlin: Springer.

Fodor, J. 1983. *The Modularity of Mind*. Cambridge, Mass.: MIT Press.

Jackendoff, R. 2002. *Foundations of Language, Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Langacker, R. W. 1987. *Foundations of Cognitive Grammar*, vol.1. Stanford: Stanford University Press.

_____ 1991. *Foundations of Cognitive Grammar*, vol. 2. Stanford: Stanford University Press.

_____ 2004. Grammar as Image: The Case of Voice. In *Imagery in Language: Festschrift in Honour of Professor Ronald W. Langacker*, Barbara Lewandowska-Tomaszczyk and Alina Kwiatkowska (eds), 63-114. Frankfurt am Main: Peter Lang.

Levelt Willem J.M. 1989. *Speaking: from Intention to Articulation*. Cambridge, Mass.: MIT Press.

McCarthy, J.1979. Formal Problems in Semitic Phonology and Morphology. Ph.D. Dissertation. MIT

_____ 1981. A Prosodic Theory of Nonconcatenative Morphology. *Linguistic Inquiry* 12: 373-418.

Pinker, S. and Bloom P. 1990. Natural Language and Natural Selection. *Behavioral and Brain Sciences* 13: 707-726.

Simon, H. A. and Kaplan, C.A. 1989. Foundations of Cognitive Science. In *Foundations of Cognitive Science*. Posner, M. (ed.), 1-47. Cambridge, Mass.: MIT Press.

Talmy, L. 1996. The windowing of attention in language. In *Grammatical Constructions: Their Form and Meaning*. M. Shibatani and S. Thompson (eds), 235-287. Oxford: Oxford University Press.

Tomasello, M. 1999. *The Cultural Origins of Human Cognition*. Cambridge Mass./ London. Harvard University Press.

Tomasello, M., and Call, J. 1997. *Primate Cognition*. New York: Oxford University Press.

Tulving, E., Kapur, S., Markowitsch, H. J., Craik, F. I. M., Habib, R., and Houle, S. 1994: Neuroanatomical correlates of retrieval in episodic memory: Auditory sentence recognition. *Proceedings of the National Academy of Sciences* (USA) 91: 2012-2015.

Tulving E, and Lepage, M. 2000. Where in the Brain Is the Awareness of one's Past? In *Memory, Brain and Belief*, Daniel L. Schacter and Scarry, E. (eds), 208-228. Cambridge, Mass.: Harvard University Press.

Zanned, L. 2005. Root Formation and Polysemic Organization in Arabic Lexicon: a Probabilistic Model. In *Perspectives on Arabic Linguistics* XVII-XVIII, M. T. Alhawary and E. Benmamoun (eds), 85-116. Amsterdam: John Benjamins.