Knowledge representation of the Quran through frame semantics

A corpus-based approach

Abdul-Baquee Sharaf Eric S. Atwell

School of Computing

University of Leeds

Leeds, LS2 9JT

United Kingdom

{scsams,eric}@comp.leeds.ac.uk

Table of Contents

Abstract

In this paper, we present our in-progress research tasks for building lexical database of the verb valences in the Arabic Quran using FrameNet frames. We study the verbs in their context in the Quran, and compare that with matching frames and frame evoking verbs in the English FrameNet. We analyze the gaps and make appropriate amendments to the FrameNet by adding new frame elements and relations.

1. Introduction

The Quran is the central religious text of Islam – the world's second largest religion with a growing population of over 1.5 billion Muslims (1). Muslims believe that the Quran contains the words of God revealed on Prophet Muhammad by the Angel Gabriel (2); and that it is free from contradictions or discrepancies (3).

While there has been research in Arabic corpus linguistics (Atwell et al 2008) (Al-Sulaiti & Atwell 2006), or keyword search tools for the Quran (4), to our knowledge no extensive work has been done towards Quranic Corpus Linguistics. The goal of this work-in-progress research is to design a Knowledge Representation (KR) model for the Quran leveraging on the concept of 'frame semantics' as introduced by Fillmore (Fillmore 1978). Based on the concept of frame semantics, researchers in International Computer Science Institute (ICSI), Berkeley, started the FrameNet project (Ruppenhofer et al 2005) (Baker et al 1998) (Fillmore et al 2003) in 1997 to build an online lexicon for English frames which are to capture the semantic and syntactic properties of English predicates based on their usage in the British National Corpus (BNC) (Aston & Burnard 1998). Based on the experience of the English FrameNet, various projects started to build similar lexicon for other languages.

In our research project, we aim to build a FrameNet like lexicon for the verbs in the Arabic Quran. This initial attempt will enable future extension to include predicates other than verbs and to consider other classical Arabic texts as well as Modern Standard Arabic.

This paper is laid out as follows: Section 2 gives background information on Arabic verbs and some linguistic style of the Quran. Section 3 gives a sketch of related works on Quranic and Arabic verbs. Section 4 gives background information on the FrameNet lexicon. Section 4 details our intended research task and the challenges towards its implementation. Section 5 describes Framenet integration projects for other languages. Section 6 reports on the main tasks and challenges of this project. Finally we conclude highlighting the novelty of our research and its expected benefits.

5

2. Backgrounds

2.1 Arabic Verbs

In general, classical Arabic follows Verb-Subject-Object (VSO) order. The majority of Arabic verbs are trilateral, which can be derived to 15 different forms. Each derivation signifies some semantic variations over the original form. Table 1 gives a brief account on the most frequent nine such forms with their semantic significance. (Wright 1996) provides more elaborate discussion. The semantic significance of each derivation form is a subtle aspect of Arabic grammar which has no direct equivalent in the grammar/morphology of English or European languages.

NO	pattern	Semantic significance	Examples
I	فَعَلَ Fa3aLa	**When the 2nd radical is vowelized with (a) it mostly indicates transitive.** **When the 2nd radical is vowelized with (i) it mostly indicates intransitive.**	كَتَبَ to write فَرِحَ to be glad
II	فَعَّلَ Fa33a La	Intensive or extensive meaning of the first form **Convert the intr. In 1st form to transitive** Estimative or declarative	كسَر (to break) and كسَّر (break into pieces) فرّح (to be glad) فرِح (to gladden) كذّب (to lie), كذَب (to call one a liar)
III	فاعَل Faa3aL a	Place effort to perform act upon the obj. Convert prepositional object to direct obj. Use Quality or state to affect another person	قاتله (he tried to kill him) كتب إلى (write to) = كاتب (write to) خاشنه (he treated him harshly)
IV	أَفْعَلَ aF3aLa	Factitive or causative Denominative (derive from noun a tr. Verb) Movement towards a place/time	جلس (to sit down) and أجلس (to dib one sit down) أثمر (to bear fruit ثمر) أشأم (to go to Syria الشام) أصبح (to enter upon the time of morning

(الصباح

V	تفَعّل taFa33aLa	***Express the state into which the obj. of the 2nd form was brought into action*** Reflexive or effective	تكسّر (to be broken in pieces) علّم (to teach) and تعلّم (to become learned)
VI	تفاعَل taFaa3aLa	***Express the state into which the obj. of the 3rd form was brought into action*** ***Convert the tr. Sense of 3rd form to reflexive*** Reciprocity	باعدته (I kept him aloof) فتباعد (so he kept aloof) تماوت (to pretend to be dead) قاتله (he fought with him) and تقاتلا (the two fought with one another)
VII	انْفَعَل inFa3aLa	***Non-reciprocal but reflexive significance of the 1st form*** A person allows an act to be done in reference with him	انكسـر (to break [intr.], to be broken) انهزم (to let oneself be put to flight, to flee
VIII	افْتَعَل iFta3aLa	***Reflexive or middle voice of the 1st form.*** Reciprocal	عرض (to place smth before one) and اعترض (to put oneself in the way, to oppose) اقتتل الناس (the people fought with one another
X	استَفْعَل istaF3aLa	***Convert the factitive significance of the 4th form into the reflexive or middle*** ***A person thinks that the quality expressed in 1st form is applicable to himself*** ***A person seeking what is expressed by 1st form***	أسلم (to give up) and استسلم (to give oneself up, to surrender) حلّ (to be lawful) and استحل (he thought that it was lawful for himself to do) استغفر(to pardon) غفر (to seek pardon)

8

Table 1. Most common forms of Arabic trilateral verbs.

2.2 The Quranic Linguistic Style

According to Muslims, the Quran is divine and contains words of God. It was revealed over a period of 23 years to the Prophet Mohammad in Arabic language. It contains around 78,000 words within the 114 chapters. The central topic of the Quran is to establish the monotheistic creed of God being the only possessor of divine power and only being who deserves to be worshiped. Prophet Muhammad challenged the Arabs to find another text – or a chapter of a text- like the Quran (5). The Quran claims to contain the fairest of statements and a scripture able to raise emotions and sentiments (6).

Following are some of the characteristics of the linguistic styles in the Quran. These features should pose special interests and challenges for computational linguistics solutions.

2.2.1 Scattered information on a same topic

The Quran often talks about a topic scattered within many different verses in different chapters. Consider the following verses (7):

[1]	*Show us the straight path, The path of those whom Thou hast favoured [1:6,7]*
[2]	*Whoso obeyeth Allah and the messenger, they are with those unto whom Allah has shown favour, of the prophets and the saints and the martyrs and the righteous* [4:69]
[3]	*He who holdeth fast to Allah, he indeed is guided unto a right path [2:101]*

In [1] there is a reference to a 'straight/right path' and a reference to a category of people whom God has favoured without highlighting who might be in this category. Verse [2] which is in a different chapter gives four types of people whom God shown favour. In [3], which is again in a different chapter, expands this list of favoured category to include one more.

The Quran also repeats a certain story, for example, of a previous prophet in many chapters but each occurrence adds certain information not present in other occurrences. For example, the Quran tells various aspects of the story of Moses in 132 places distributed among 20 chapters. This feature of the Quran makes a good case for computational solutions towards bringing these scattered occurrences automatically in one thread.

2.2.2 Literal vs. technical sense of a word

The Quran borrows an Arabic word and specializes it to indicate a technical term. Consider for example the word جَنّة/jannah meaning literally 'a garden', but -as a technical term- in the Quran whenever this word is used it refers to 'the paradise' where the believers will abode as reward after the Day of Judgment. However, there are few instances where this word is used in the literal meaning to refer to certain gardens in this world. In the following examples [4] uses the more frequent technical sense and [5] uses the less frequent literal meaning.

[4]	*And vie one with another for forgiveness from your Lord, and for a <u>paradise</u> as wide as are the heavens and the earth, prepared for those who ward off (evil); [3:133]*
[5]	*There was indeed a sign for Sheba in their dwelling-place: Two <u>gardens</u> on the right hand and the left..[34:15]*

2.2.3 Grammatical shift

The Quran often draws the attention of the reader by shifting grammatical agreement in a statement. For example, in [6] the mode changed from 'you' to 'they' and 'them' moving from 2nd person to third person. In [7] the verse shifted from addressing the Prophet alone to addressing the group.

[6]	*when ye are in the ships and they sail with them with a fair breeze* [3:133]
[7]	*O Prophet! When ye (men) put away women..[65:1]*

2.2.4 Verbs associating with different preposition

The Quran exhibits many examples where a certain verb is associated with a preposition which is unusual with this verb, but common with a different verb. Consider [8a] and [8b] below, the Arabic verbs خلا/*khala* means be alone, which is usually followed by the preposition 'with' like 'John was alone with Mary'. However, in this verse the Quran choose to use the preposition 'to' with 'be alone' which sounds unusual to say, 'John was alone to Mary'. However, this is a valid classical Arabic style when a verb borrows a preposition that binds with another verb and uses it to indicate at the same time meaning of both verbs. The Arabic verb ذهب/dhahaba (go) fits well with the preposition 'to' as in: 'John went to Mary'. So, in this verse, the Quran by using a verb (be alone) with a preposition (to) from another verb 'go' conveyed the meaning of 'being alone and going to' at the same time. This unique characteristic made both translations in [8a] and [8b] partially true, highlighting either the sense of the original verb 'be alone with' as in [8a] or the implicit verb with explicit preposition 'go to' as in [8b]. See Ibn-Katheer (2006) on his commentary of this verse.

[8a]	*When they meet those who believe, they say: "We believe;" but when they are <u>alone with</u> their evil ones, they say: "We are really with you: We (were) only jesting."* [2:14 Yusuf Ali Translation]
[8b]	*And when they fall in with those who believe, they say: We believe; but when they <u>go apart to</u> their devils they declare: Lo! we are with you; verily we did but mock.* [2:14 Pickthal Translation]

2.2.5 Metaphors and Figurative

The Quran uses a lot of metaphors and figurative language. In [9] Pickthal used the verb 'shine' but the Arabic verb /ishtala means 'to flare' and shows the analogy of 'old age symptom by many gray hair' with a 'fire burning a bush'. In [10] the Muslim army was so frightened that it felt as if their hearts reached to the throats.

[9]	*My Lord! Lo! the bones of me wax feeble and my head is* <u>shining</u> *with grey hair..*[19:4]
[10]	*When they came upon you from above you and from below you, and when eyes grew wild and hearts reached to the throats* [33:10]

2.2.6 Metonymy

In many verses the Quran uses metonymy. In [11] the Arabic verse literally means 'ask the town' which means (and was translated so) 'ask the people who live in the town'. In [12] 'a thing of planks and nails' is the 'Noah's ark', and in [13] 'eating food' metonymically means the 'need to answer call of nature', see Ibn-Katheer (2006) commenting on this verse.

[11]	*Ask the township where we were, and the caravan with which we travelled hither.* [12:82]
[12]	*And We carried him upon a thing of planks and nails* [54:13]
[13]	*The Messiah, son of Mary, was no other than a messenger, messengers (the like of whom) had passed away before him. And his mother was a saintly woman. And they both used to eat (earthly) food* [5:75]

2.2.7 Imperative vs. non-Imperatives

Arabic verbs are classified into past, present and imperative. Thus, in Arabic the imperative structure can be understood from the type of the verb used. However, in the Quran, although this general rule applies, yet there are many instances where imperative is understood although no imperative verb is used, for example in [14]. The opposite is also true: there are instances where an imperative verb is used, but the verse indicates non-imperative sense, for example [15] where the translator explicitly indicated the non-imperative meaning within brackets.

[14]	*and whoever is minded to perform the pilgrimage therein there is no lewdness nor abuse nor angry conversation on the pilgrimage.* [2:197]
[15]	*O ye who believe! Profane not Allah's monuments nor the Sacred Month nor the offerings nor the garlands, nor those repairing to the Sacred House, seeking the grace and pleasure of their Lord. But when ye have left the sacred territory, then go hunting (if ye will). [5:2]*

3. Related work

(Bielicky and Smarz 2008) describes building a valency lexicon for modern standard Arabic from the Prague Arabic Dependency Treeback (PADT). Their work is built on 'Functional Generative Description (FGD)' theory where verbs have valency frame with many complements known as functors which can further be divided into actants (Actor, Addressee, Patient, Effect and Origin) and adjuncts (like Manner , Means and Location). This FGD concept was adapted for Arabic verbs and various corpus examples were drawn to prove the applicability of FGD for capturing Arabic verb valency. Some cases needed special attention like: diathesis, passive verbs, reflexivity and verb nominals.

(Al-Qahtani 2005) gives an extensive categorization of modern standard Arabic verb valence based on Case Grammar (CG) as described by (Fillmore 1968). Based on the assumption that CG is adequate to classify all verbs of a language and is universal across languages, Al-Qahtani went on to specify valence according to Cook's Matrix Model (Cook 1979) and its extension that includes 24 cells. According to this matrix five cases (Agent, Experiencer, Benefactive, Object, Locative) are plotted horizontally and type of verb (State, Process, Action) vertically. The date was taken from 8327 verbs from a lexicon (Al-Qahtani 2003) and most frequent 200 verbs were exhaustively sorted to a cell in the matrix, and thus proved the suitability of Cook's model for Arabic valence.

(Fiteih 1983) studied the prepositional verbs considering the Quran as his corpus. He could classify four classes of Quranic verbs based on the number and type of nominals and prepositions these verbs allow. There are cases when a verb allows one prepositional object (e.g., *reach to something* as in [16]), or a nominal and a prepositional object (e.g., *send against someone something* as in [17]), or two prepositional objects (e.g., *come forth unto someone from some place* as in [18]), or one nominal object and two prepositional objects [19a] or one prepositional object and two nominal objects [19b].

[16]	*And when he saw their hands* **reached** *not* <u>to</u> *it, he mistrusted them.. [11:70]*
[17]	*For We* **sent** <u>against</u> *them* <u>a furious wind</u>*, [54:19 Yusuf Ali Translation]*
[18]	*Then he* **came forth** <u>unto</u> *his people* <u>from</u> *the sanctuary [19:11]*
[19]	*a. And Allah hath favoured some of you above others in provision [16:71]* *b. He hath* **bestowed** *on* <u>those who strive</u> *a* <u>great reward</u> <u>above</u> *the sedentary[4:95]*

Shamsan (Shamsan 1986) studies the transitivity and intransitivity of Quranic verbs. He analyzed the valences of these verbs and tried to link between the form of these verbs and the semantic significance. He also observed the shift of a verb from intransitive to transitive sense based on semantic characteristics.

(Mir 1989) observed that quite a lot verbs in the Quran are used in idiomatic sense rather than literal meaning of the verb. He went on to list such expressions in the Quran. Some examples are given in the following quote.

When a man's "eyes become cool", it means that he is pleased. A person who "brings down his wing" for you is being kind to you, but if he "bites his fingers" at you, he holds you a severe grudge. If you think you lack the gift of fluent speech, you can pray to God to "untie the knot in your tongue" (Mir 1989: 2-3)

4. FrameNet Lexicon

FrameNet is a lexicon that describes 'Frames' as a schematic representation describing a situation involving various conceptual roles called 'Frame Elements (FE)'. A frame can be 'evoked' by a group of related predicates (mainly verbs, but also nouns or adjectives) called 'Lexical Units (LU)'.

For example, the verb 'buy' along with 'purchase' form the LUs that can evoke the **commerce_buy** frame. This frame has 'core' – frame elements that are essential to the meaning of the frame- FEs (BUYER, GOODS) and has many other non-core FEs (like: DURATION, MANNER, MEANS, MONEY, PLACE, PURPOSE, RATE, REASON, RECIPIENT, SELLER, TIME, UNIT).

Following are some illustrative examples from **commerce_buy** frame description. (The *lexical unit* is in **boldface** and *Frame Elements* are in CAPITALS).

[20]	[BUYER Lee] **BOUGHT** [GOODS a textbook] [SELLER from Abby]
[21]	Will they allow [BUYER you] to **PURCHASE** [MEANS by check?]
[22]	[BUYER Sam] **BOUGHT** [GOODS the car] [MONEY for $12,000].
[23]	[BUYER You] **BOUGHT** [RECIPIENT me] [GOODS three pairs] already!

Currently, the FrameNet project contains more than 10,000 lexical units in nearly 800 hierarchically related semantic frames, exemplified in more than 135,000 annotated sentences. (Ruppenhofer et al 2005).

In addition to frame description, FrameNet also specifies frame-to-frame relations. These relations include: *inheritance, subframe, causative_of, inchoative_of* and *using*. For example, in figure 1, the frame **commerce_buy** inherits from more general **getting** frame, and is inherited by more specific **renting** frame, and is used by two related frames, namely, **importing** and **shopping**.
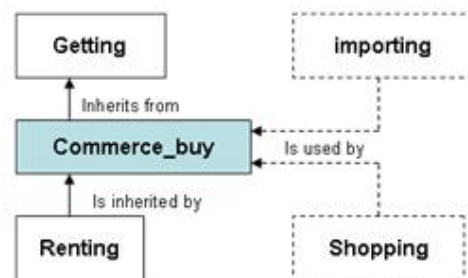

Figure 1. Frame-to-frame relations of the 'commerce_buy' frame

FrameNet also provides annotated sentences. This can be of two types: lexicographically motivated annotation and full-text annotation. In the former, the focus is to record the range of semantic and syntactic combinatory possibilities of a target lexical unit. Annotation of running text, on the other hand intends to exhaustively annotate each word in the text, which is possible thanks to layering techniques. The main layers are: a) Frame Element (FE) specifying frame elements as depicted in example [16] to [19], b) grammatical function (GF) like subject, object, etc., c) phrase type (PT) like noun phrase, verb phrase, etc and d) part-of-speech layer (POS).

Natural texts in many cases do not show up many conceptual frame elements. For this reason FrameNet annotation kept provision for 'Null Instantiation' (NI). This omission can be understood from the context and is called 'Definite Null Instantiation (DNI) like the missing RECIPIENT in [20] or cannot be retrieved but whose type is known like the missing QUARREL sense in [21], or the omission is allowed by the grammar of the sentence like the missing subject in any imperative structure like in [22].

[24]	*John contributed $20.*
[25]	*Bob and Sue would argue all day.*
[26]	*Get out immediately!*

Since the launch of the English FrameNet, many researchers started to use FrameNet for various applications for example, *Machine Translation* (Boas 2002), *Question Answering* (Narayanan & Harabagiu 2004*), information retrieval* (Narayanan & Mohit 2003), *textual entailment* (Burchardt & Frank 2006), and also by incorporating it into domain specific ontology like BioFrameNet project (Dolbey et al. 2006).

5. Multi-lingual FrameNet projects

Since the release of the English FrameNet, researchers started similar projects in other languages. Successful examples are German, Spanish and Japanese.

5.1 German SALSA project

The German FrameNet project known as SALSA (Burchardt et al 2009) builds on the assumption that the English FrameNet is based on coarse-grained semantic classes which describes prototypical situations and thus, can be applied to other languages. During the course of the project, the team have found high correlation between English and German frames. However, they encountered some problems related to non-existence of certain language constructions in English (like some use of datives) and lexicalization differences in certain semantic domains (such as movement). The team went on to exhaustively annotate a large scale German corpus – the TIGER treebank (Bransts et al. 2002) – and in the process they had to encounter issues which were not faced by the FrameNet team, like dealing with idioms, support verb constructions, and metaphors. Idioms are multiword fixed expressions, and hence, the team decided to consider the whole expression as a frame-evoking word. In support verb constructions, the verb only supports a head noun (like 'give lecture') where the 'lecturing' frame should evoked instead of a 'giving' frame. The SALSA team in this case annotated the verbal part with a pseudo frame 'Support' with the noun as SUPPORTED frame element. In case of metaphors, in order to understand the literal source meaning should be transferred to the target intended meaning. The SALSA teams decided to annotate such cases with two frames: one for the target and one for the source.

As FrameNet is still under development, the team had to encounter non-existence of certain lemma senses in the English FrameNet. In these cases, they created proto-frames which define a new Frame following the style of the English FrameNet, and are also included in the frame-to-frame relationships.

The annotation is done using home-made SALTO tool that extends the TIGER syntactic tree to include Frame description. Unlike FrameNet, SALSA annotates frames with only 'core' frame elements.

5.2 Spanish FrameNet

Spanish FrameNet (Subirats & Petruck 2003) uses the English FrameNet lexicon to build a Spanish lexical resource. The project built a subcorpus of sentences from a 300 million word Spanish corpus that contains texts from various genres. (Subirats & Petruck 2003) report some difference in the lexicalization patterns of emotion predicates between English and Spanish as follows:

"While both languages lexicalize the causative meaning with a verb (*sorprender* and *surprise*) and the stative meaning with an adjective (*estar sorprendido* and *to be surprised*), Spanish lexicalizes the inchoative meaning in the reflexive verb *sorprenderse* - 'to get surprised', while English uses a construction with *get* and the adjectival past participle *surprised*. In addition, while English has just one lexical unit *surprised* in the Experiencer_subject frame, Spanish has two: *sorprendido* used in conjunction with *estar* as a stative; and *sorprenderse* which is inchoative."

5.3 Japanese FrameNet

Japanese FrameNet (Ohara et al 2004) is a project started in 2002 based on English FrameNet. It started with a pilot study of motion and communication verbs. Corpus evidence is taken from the Mainichi newspaper corpus. The project team realized that unlike English, Japanese specifies a path along with motion, and thus has verb for 'go across' and another for 'go beyond, go over'. Therefore, they suggest amending Frame elements with BOUNDARY or ROUTE elements.

21

6. The Quranic FrameNet Project

6.1 Main Tasks

The first task is to collect all verbs in the Quran and their context in the verses. The reason we chose to consider only verbs is: first, to start with a feasible scope, and second, in Arabic –as well as other languages- verbs play the most vital predicate role. Malise Ruthven explains further:

Substances and adjective are almost always verbal derivatives, usually participles or verbal nouns. A clerk is a writer [katib], a book is a writ [kitab]. Aeroplanes and birds are thing that fly [tiara and tayr]…it is precisely because Arabic refrains from classifying words into discrete particles, but keeps them instead in a logical and balanced relationship with a central concept. –the verbal root – that it becomes an eminently suitable language for religious expression." (Ruthven 1984:111)

This work of Quranic verbal analysis is being carried out through a machine readable index of the Quran (Abdulbaqi 1955). Each verb will be classified into their form (see verb forms in Table 1), which will help in semantic labeling later. Then, each Quranic verb needs to be studied to find a matching FrameNet lexical unit. For ambiguous cases, several parallel English translations will be consulted. Also, Books of Tafsir (scholarly interpretation of the Quran) for example (Ibn-Katheer 2006) or specialized lexicons and dictionaries (for example (Ibn-Mandhour 1997) or (Penrice 1873)) can be studied for clarification. Through this chosen lexical unit, the corresponding frame in FrameNet will then be studied for appropriateness. To check this 'appropriateness', all target Quranic verb valences must exhibit the core frame elements of the chosen frame.

As an example, consider the *Ingestion* frame as depicted in Table 2 below. This frame has two core elements: ingestibles and an ingestor.

Frame Name	**Ingestion**	
Definition	*An Ingestor consumes food or drink (Ingestible), which entails putting the Ingestible in the mouth for delivery to the digestive system. This may include the use of an Instrument. Sentences that describe the provision of food to others are NOT included in this frame.*	
Core Frame Elements	**Ingestibles**	*The Ingestibles are the entities that are being consumed by the Ingestor.*
	Ingestor **(Sentient)**	*The Ingestor is the person eating or drinking.*
Lexical Units	*breakfast.v, consume.v, devour.v, dine.v, down.v, drink.v, eat.v, feast.v, feed.v, gobble.v, gulp.n, gulp.v, guzzle.v, have.v, imbibe.v, ingest.v, lap.v, lunch.v, munch.v, nibble.v, nosh.v, nurse.v, put away.v, put back.v, quaff.v, sip.n, sip.v, slurp.n, slurp.v, snack.v, sup.v, swig.n, swig.v, swill.v, tuck.v*	

Table 2: FrameNet description of the frame: Ingestion

Next, consider the verb 'eat' in the Quran. It appeared –with derived forms- 100 times. Table 3 below lists a few representative concordance lines. In the majority of the cases, its use was in alignment with FrameNet

descriptions, like the example of line [A]. However, there are examples where 'eat' is used differently, for example lines [B] uses 'eat' to mean 'eating money' which is not a usual ingestible item, and hence it means to 'earn money unlawfully'. Consider also the line [E] where 'seven years' are the 'ingestor' which violates the 'sentient' restriction of FrameNet.

	the sea to be of service that ye	**eat**	fresh meat from thence	16 :14
	And	**eat**	not up your property among	2: 188
	Would one of you love to	**eat**	the flesh of his dead brother?	49 :12
	seven fat kine which seven lean were	**eating**		12 :43
	seven hard years which will	**devou r**	all that ye have prepared for them	12 :48
	they	**eat**	into their bellies nothing else than fire	2: 177
		Devou rer	of unlawful	5: 42

Table 3. Few KWIC lines for <eat> from the Quran

These Quranic usages mandate us to extend the FrameNet to capture these non-ingestible and non-sentient uses. Thus, we suggest following the German SALSA strategy of creating a proto-frame for this special sense of 'eating money'.

As indicated in the previous section, the Quran contains many instances of verbal idioms. In such cases, again we follow the SALSA solution of considering the whole multi-word idiom as a frame-evoking predicate. Similarly, in case of metaphors, we intend to produce two annotations of such verses: one for the literal meaning and another to represent the metaphorical intended meaning.

In addition to exhaustively annotating the subcorpus of verses containing verbs in the Quran, we intend to choose as a case study, full-annotation of chapter 2 'Surah al-Baqarah' as a sample chapter from the Quran. This chapter portrays vibrant use of verbs since 97.5% of its 286 verses contain verbs (Suleiman 1997). We will carry on annotation in three layers, as is the FrameNet practice: Frame Elements layer, Grammatical Function Layer and Phrase Type Layer. In order to annotate Grammatical function, we will resort to reference books which exhaustively analyzed the grammatical function of each verse of each chapter, for example (Salih 1998), and populate the grammatical function layer. It should be noted that because of the vocalized form of the Quranic text, many ambiguities that appear otherwise in modern standard Arabic will not be faced. However, it is evident that many Quranic expressions result in more than one valid syntactic –and semantic- tree. For example, consider [27] which can refer

23

simultaneously to two valid meanings [27a] and [27b] depending on where to pause.

[27]	*This is the book no doubt in it a guidance for those conscious of Allah [2:2]*
[27a]	*This is the book no doubt in it. It is guidance for those conscious of Allah.*
[27b]	*This is the book no doubt. In it a guidance for those conscious of Allah.*

6.2 Representation

To represent the frames and lexical units, we will adhere to the structure of the FrameNet Database as detailed in (Baker et. al 2003). The result will be presented online in the FrameNet style, where color highlighting will help distinguish various frame elements.

6.3 Evaluation and Applications

Our annotated Quran chapters should represent the following three verses which are related semantically but scattered in various locations. Using FrameNet's *Giving* frame and extending it to capture the non-profit charitable sense of spend, the following labeling can be made.

[28]	*and [they DONOR] spend out of [what We have provided for them THEME] [2:3]*
[29]	*and they ask you what [they DONOR] should spend. Say, "[the excess DONATED_AMOUNT]". [2:219]*
[30]	*and they ask you what they should spend. Say, "Whatever [you DONOR] spend of [good DONATED_AMOUNT] is [for parents and relatives and orphans and the needy and the traveler RECIPIENT]. [2:215]*

While [28] talks about the theme of the donated money, [29] qualifies the type of this theme to be from the excess money that is left after spending on the necessary needs. However, [30] gives an answer to the same question as in [29], but specifies the recipient of this spend rather than the type or amount of the money.

Similarly, proceeding with annotation of the Quran governed by frame semantics might reveal interesting findings which might not be captured in books of Tafsir (scholarly interpretation of the Quran).

Annotating the Quran with frame semantics will facilitate efficient search beyond the existing keyword search. A Quranic researcher will be able to search semantic frames and semantic roles in addition to keywords.

Another interesting application of our semantically annotated corpus would be a Question Answering system. A question can be normalized into FrameNet style representation and matched with similar frames in the Quran for potential answers. (Shen and Lapata 2007) showed that FrameNet annotation produces significant improvement in QA systems.

6.3 Challenges

FrameNet is still under development. So for a certain lemma not all senses maybe covered. Also, because FrameNet only uses lexicographical prototype examples, some context usage might be hard to relate. Also, idioms and metaphors pose difficulty in representation. The lack of Arabic NLP tools –as compared to English NLP tools- might cause problems in automation and computational analysis.

7. Conclusions

We have embarked on a novel project towards frame semantics which starts by developing FrameNet frames for Quranic verbs, but can be extended to include non-verbal predicated in the Quran and can further be extended to include predicates in Modern Standard Arabic. To our knowledge no previous attempts has been made towards integrating Arabic verbs to FrameNet frames.

Once completed, this research will benefit a wide audience. It will benefit Arabic NLP researchers considering a full-fledged Arabic FrameNet. It will benefit also the FrameNet community towards achieving a multi-lingual FrameNet project. This research will serve the wide Muslim population for better searching and extracting information from the Quran. In particular, the frame reports of Quranic Verb will interest Arabic linguists is analyzing the valence of the Quranic verbs.

Notes

(1) http://www.adherents.com/Religion_By_Adherents.html

(2) The Quran 26: 192-195

And lo! it is a revelation of the Lord of the Worlds, Which the True Spirit hath brought down. Upon thy heart, that thou mayst be (one) of the warners, In plain Arabic speech. [Pickthal Translation]

(3) The Quran 4: 82

Will they not then ponder on the Qur'an? If it had been from other than Allah they would have found therein much incongruity. [Pickthal Translation]

(4) See for example http://www.searchquran.org

(5) The Quran 10:38

Or say they: He hath invented it ? Say: Then bring a surah like unto it, and call (for help) on all ye can besides Allah, if ye are truthful. [Pickthal Translation]

(6) The Quran 39:23

Allah hath (now) revealed the fairest of statements, a Scripture who's parts resembling each other, paired whereat doth creep the flesh of those who fear their Lord, so that their flesh and their hearts soften to Allah's reminder

(7) Citing verse reference with notion [x:y], x indicates chapter number and y indicates verse number. Unless otherwise mentioned, all translations are taken from Pickthal's translation available at University of Southern California's Centre for Muslim-Jewish engagement website: http://www.usc.edu/schools/college/crcc/engagement/resources/texts/muslim/quran/

References

1- Abdulbaqi, M. (1955) *Aphabatical Index of the Quranic Words* (In Arabic). Dar al-adhami, Beirut. Available online http://www.qurancomplex.com/IdIndex/default.asp?TabID=1&SubItemID=7&l=arb&SecOrder=1&SubSecOrder=7#

2- Al-Qahtani, D (2005) *Semantic Valence of Arabic Verbs*. Librairie du Liban Publishers

3- Al-Sulaiti, L. and E. Atwell (2006). "The design of a corpus of contemporary Arabic." *International Journal of Corpus Linguistics*. 11, 135-171.

4- Aston, G. and L. Burnard, (1998). *The BNC Handbook: Exploring the British National Corpus with SARA*, Edinburgh University Press.

5- Atwell, E., N. Abbas, B. Abu-Shawar, A. Alsaif, L. Al-Sulaiti, A. Roberts and M. Sawalha (2008). "Mapping Middle Eastern and North African diasporas: Arabic corpus linguistics research at the University of Leeds" *In: Proceedings of BRISMES Conference 2008*.

6- Baker, C., C. Fillmore and B. Cronin (2003). "The Structure of FrameNet Database". *Int. Journal of Lexicography*, 16(3), 281-296.

7- Baker, C., C. Fillmore and J. Lowe (1998). "The Berkeley Framenet project." In Proceedings of the 17th International conference on Computational Linguistics. ACL. NJ, USA.

8- Bielicky, V and O. Smarz (2008) "Building the Valency Lexicon of Arabic Verbs", LREC 2008.

9- Boas, H. (2002). Bilingual framenet dictionaries for machine translation. In Proceedings of the 3rd International Conference on Language Resources and Evaluation. Spain.

10- Brants,S., S. Dipper, S. Hansen, W. Lezius and G. Smith (2002). The TIGER Treebank.

11- Burchardt, A. and A. Frank (2006). Approximating texual entailment with LFG and FrameNet frames. In Proceedings of the 2nd Recognising Texual Entailment Workshop. Venice, Italy.

12- Burchardt, A. K. Erk, A. Frank, A. Kowalski, S. Pado, and M. Pinkal(2009): *FrameNet for the semantic analysis of German: Annotation, representation and automation (preprint)*. In Hans Boas (ed.): Multilingual FrameNet. Mouton de Guyter. (In Press)

13- Cook, W. (1979) Case Grammar: Development of the Matrix Model. Washington, D.C. : Georgetown University.

14- Dolbey, A. M. Ellsworth, and J Scheffczyk (2006). BioFrameNet: A domain-specific fragment extension with links to biomedical ontologies. In Proceedings of the Biomedical Ontology in Action.

15- Fiteih, M. (1983) Prepositions and Prepositional Verbs in Classical Arabic. PhD thesis, Univesity of Leeds.

16- Fillmore, C.(1968) The case for case. In Bach, E. W. and Harms, R.T. (Eds), *Universals in Linguistic Theory*.

17- Fillmore, C. (1976). "Frame Semantics and the nature of language." *Annals of the New York Academy of Science*.

18- Fillmore, C., C. Johnson and M. Petruck (2003). "Background to Framenet". *Int. Journal of Lexicography*, 16(3), 235-250.

19- Ibn-Katheer (2006). *Tafseer Al-Quran* (In Arabic). Dar Al-Kutub al-Elmiyyah.

20- Ibn-Mandhour (1997) *Lisan Al-Arab* (In Arabic). Dar Sadir.

21- Mir, M (1989) Verbal Idioms of the Quran. Michigan Series on the Middle East, No. 1.Center for Near Eastern and North African Studies, University of Michigan, Ann Arbor.

22- Narayanan, S. and B. Mohit (2003). Semantic extraction with wide-coverage lexical resources. Companion Volume of the Proceedings of HLT-NAACL. Canada, 2003.

23- Narayanan, S. and S. Harabagiu (2004). Question answering based on semantic structures. In Proceeding of the 20th international conference on Computational Linguistics. Switzerland, 2004.

24- Ohara, K., S. Fujii, T. Ohori, R. Suzuki, H. Saito, and S. Ishizaki (2004). "The Japanese FrameNet Project: An introduction." LREC 2004. The Fourth international conference on Language Resources and Evaluation. Lisbon, Portugal. May, 2004.

25- Penrice, J. (1873). *Dictionary and Glossary of the Koran*. Adam Publishers and Distributors, India.

26- Ruppenhofer, J., M. Ellsworth, M. Petruck, and C. Johnson (2005). "FrameNet: Theory and Practice.

27- Salih, B. (1998). Al-E'rab al-Mufassal Li-Kitabillah al-Munajjal (Arabic). Dar Al-Kutub Al-Elmiyyah. Beirut.

28- Shamsan, A. (1986). Al-Fi'l fil Quran al-Kareem. (In Arabic). Kind Saud University.

29- Shen, D. and M. Lapata (2007). Using semantic roles to improve question answering. In Proceedings of the Conference on Empirical Methods in Natural Language Processing and on Computational Natural Language Learning, Prague, pp. 12–21.

30- Subirats, C.and M. Petruck, (2003). Surprise: Spanish FrameNet. *International Congress of Linguists. Workshop on Frame Semantics, Prague (Czech Republic), July 2003*.

31- Suleiman, F (1997). Al-Fial Fi Surat Al-Baqarah (In Arabic). Maktabat al-Aadab. Cairo

32- Wright, W. (1996) A Grammar of the Arabic Language. Librairie Du Liban, Beirut.