This is a repository copy of *Knowledge representation of the Quran through frame semantics: a corpus-based approach*.

White Rose Research Online URL for this paper: http://eprints.whiterose.ac.uk/81364/

**Proceedings Paper:**

**Reuse**
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher’s website.

**Takedown**
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.
Knowledge representation of the Quran through frame semantics
A corpus-based approach

Abdul-Baquee Sharaf Eric Atwell
School of Computing
University of Leeds
Leeds, LS2 9JT
United Kingdom
{scsams,eric}@comp.leeds.ac.uk

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 exist lots of 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 FrameNet like lexicon for the verbs in the 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 it’s expected benefits.

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.

<table>
<thead>
<tr>
<th>NO</th>
<th>pattern</th>
<th>Semantic significance</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Fa3aLa</td>
<td>When the 2nd radical is vowelized with (a) it mostly indicates transitive.</td>
<td>(to write) كتب</td>
</tr>
<tr>
<td></td>
<td></td>
<td>When the 2nd radical is vowelized with (i) it mostly indicates intransitive.</td>
<td>(to be glad) فرح</td>
</tr>
<tr>
<td>II</td>
<td>Fa33aLa</td>
<td>Intensive or extensive meaning of the first form</td>
<td>(to break) كسر</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Convert the intr. In 1st form to transitive</td>
<td>(break into pieces) كسر</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Estimative or declarative</td>
<td>(to be glad) فرح</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(to gladden) كتب</td>
<td>(to call one a liar) كتب</td>
</tr>
<tr>
<td>III</td>
<td>Faa3aLa</td>
<td>Place effort to perform act upon the obj.</td>
<td>(he tried to kill him) قاتله</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Convert prepositional object to direct obj.</td>
<td>(write to) كتب</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Use Quality or state to affect another person</td>
<td>(write to) كتب</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(he treated him harshly) خانته</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>aF3aLa</td>
<td>Factitive or causative</td>
<td>(to sit down) جلس</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Denominative (derive from noun a tr. Verb)</td>
<td>(to bear fruit) الامر</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Movement towards a place/time</td>
<td>(to go to Syria) أينم</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(to enter upon the time of morning) الصباح</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>taFa33aLa</td>
<td>Express the state into which the obj. of the 2nd form was brought into action</td>
<td>(to be broken in pieces) تكسر</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reflexive or effective</td>
<td>(to teach) علم</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(to become learned)</td>
<td></td>
</tr>
<tr>
<td>VI</td>
<td>taFaa3aLa</td>
<td>Express the state into which the obj. of the 3rd form was brought into action</td>
<td>(so he kept aloof) قباععنا</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Convert the tr. Sense of 3rd form to reflexive</td>
<td>(to pretend to be dead) تقلتلا</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reciprocity</td>
<td>(the two fought with one another) تقلتلا</td>
</tr>
<tr>
<td>VII</td>
<td>inFa3aLa</td>
<td>Non-reciprocal but reflexive significance of the 1st form</td>
<td>(to be broken) انكسر</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A person allows an act to be done in reference with him</td>
<td>(to let oneself be put to flight, to flee) انسرم</td>
</tr>
<tr>
<td>VIII</td>
<td>iFta3aLa</td>
<td>Reflexive or middle voice of the 1st form.</td>
<td>(to place smth before one) عرض</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reciprocal</td>
<td>(to put oneself in the way, to oppose) أعرض</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(the people fought with one another) اقتتل الناس</td>
</tr>
<tr>
<td>IX</td>
<td>istaF3aLa</td>
<td>Convert the factitive significance of the 4th form into the reflexive or middle</td>
<td>(to give up) استسلم</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A person thinks that the quality expressed in 1st form is applicable to himself</td>
<td>(to give oneself up, to surrender) استسلم</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A person seeking what is expressed by 1st form</td>
<td>(to be lawful) حسن</td>
</tr>
<tr>
<td></td>
<td></td>
<td>استغل</td>
<td>(he thought it was lawful for himself to do) استغل</td>
</tr>
<tr>
<td></td>
<td></td>
<td>استغفر (to seek pardon) استغفر</td>
<td></td>
</tr>
</tbody>
</table>

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 bring a chapter like the Quran (5). The Quran claims to contain the fairest of statements and a scripture who’s parts resembling each other and it’s topics are paired and 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 few examples of the question left unanswered in [1] and mentions 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 some 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 under one platform.

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 جنّة 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 abide 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 word. 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 paradise 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 gardens 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 is 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 خالa/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 ذهّب/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].

[8a] When they meet those who believe, they say: "We believe," but when they are alone with 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 go apart to 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 heavily metaphors and figurative. 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 as if their hearts reached to the throats.

[9] My Lord! Lo! the bones of me wax feeble and my head is shining 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’.

[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
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 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 proof 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 proofed 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 to it, he mistrusted them. [11:70]
[17] For We sent against them a furious wind. [54:19 Yusuf Ali Translation]
[18] Then he came forth unto his people from the sanctuary [19:11]
[19] a. And Allah hath favoured some of you above others in provision [16:71]
   b. He hath bestowed on those who strive a great reward above 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 under their characteristics.

(Mir 1989) observed that quite a lot verbs in the Quran are used in idiomatic sense rather than literal meaning of the verb. Thus, he went on to list such expressions in the Quran. Some such 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 can also be 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- FE (BUYER, GOODS) and has many other non-core FEs (like: DURATION, MANNER, MEANS, MONEY, PLACE, PURPOSE, RATE, REASON, RECIPIENT, SELLER, TIME, UNIT).

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

- [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](image)

FrameNet also provides annotated sentences. This can be of two types: lexicographically motivated annotation and full-text annotation. In the formal, 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].

[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
German FrameNet project known as SALSA project (Burchardt et al 2009) sits 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 treeback (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 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 builds subcorpus of sentences from a 300 million word
Spanish corpus that contains texts from various genres. (Subirats & Petruck 2003) reports 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 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.

6. The Quranic FrameNet Project: Tasks and Challenges

6.1 Main Tasks
The first task is to collect all verbs in the Quran and their context in the verses. As for 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 birs 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 verses 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: an ingestible and an ingestor.

<table>
<thead>
<tr>
<th>Frame Name</th>
<th>Ingestion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definition</strong></td>
<td>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.</td>
</tr>
<tr>
<td><strong>Core Frame Elements</strong></td>
<td><strong>Ingestibles</strong> The Ingestibles are the entities that are being consumed by the Ingestor. <strong>Ingestor</strong> (Sentient) The Ingestor is the person eating or drinking.</td>
</tr>
<tr>
<td><strong>Lexical Units</strong></td>
<td>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,</td>
</tr>
</tbody>
</table>
Next, consider the verb ‘eat’ in the Quran. It appeared—with derived forms—100 times. Table 2 below lists few representative concordance lines. In 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 is meant ‘earn money unlawfully’. Consider also the line [E] where seven years are the ‘ingestor’ which violates the ‘sentient’ restriction.

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>the sea to be of service that ye eat fresh meat from thence</td>
<td>16:14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>And eat not up your property among</td>
<td>2:188</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Would one of you love to eat the flesh of his dead brother?</td>
<td>49:12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>seven fat kine which seven lean were eating</td>
<td>12:43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>seven hard years which will devour all that ye have prepared for them</td>
<td>12:48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>they eat into their bellies nothing else than fire</td>
<td>2:177</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>Devourer of unlawful</td>
<td>5:42</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

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

As indicated in 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 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 subcorpus of verses containing verbs in the Quran, we intend to choose as case study, full-annotation of chapter 2 ‘Surah al-Baqarah’ as a sample chapter from the Quran. This chapter portrays vibrant use of verbs where 97.5% of it’s 286 verses (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—trees. For example, consider [27] which can refer simultaneously to two valid meaning [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 distinguishing various frame elements.
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 very 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 range of audience. It will benefit Arabic NLP researchers for considering a full-fledged Arabic FrameNet. It will benefit also the FrameNet community towards achieving a multi-lingual FrameNet project. As audience, 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. [Pickthall 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. [Pickthall 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. [Pickthall 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 Pickthall’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


