

# Open Research Online

---

The Open University's repository of research publications and other research outputs

## Building the Emirati Arabic FrameNet

### Conference or Workshop Item

How to cite:

Gargett, Andrew and Leung, Tommi (2020). Building the Emirati Arabic FrameNet. In: Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet, European Language Resources Association, Marseille, France pp. 70–76.

For guidance on citations see [FAQs](#).

© 2020 European Language Resources Association



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Accepted Manuscript

Link(s) to article on publisher's website:

<https://www.aclweb.org/anthology/2020.framenet-1.10>

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's [data policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](https://oro.open.ac.uk)

# Building the Emirati Arabic FrameNet

A. Gargett(1), T. Leung(2)

(1) Hartree Centre (Science & Technology Facilities Council),  
United Kingdom (andrew.gargett@stfc.ac.uk)

(2) United Arab Emirates University,  
United Arab Emirates (leung@uaeu.ac.ae)

## Abstract

The Emirati Arabic FrameNet (EAFN) project aims to initiate a FrameNet for Emirati Arabic, utilizing the Emirati Arabic Corpus. The goal is to create a resource comparable to the initial stages of the Berkeley FrameNet. The project is divided into manual and automatic tracks, based on the predominant techniques being used to collect frames in each track. Work on the EAFN is progressing, and we here report on initial results for annotations and evaluation. The EAFN project aims to provide a general semantic resource for the Arabic language, sure to be of interest to researchers from general linguistics to natural language processing. As we report here, the EAFN is well on target for the first release of data in the coming year.

**Keywords:** Emirati Arabic, FrameNet, corpus linguistics

## 1. Introduction

The Emirati Arabic FrameNet (EAFN) project aims to initiate a FrameNet for Emirati Arabic, utilizing the Emirati Arabic Corpus (EAC, Halefom et al. 2013). The goal is to create a resource comparable to the initial stages of the Berkeley FrameNet (Baker et al. 1998). A FrameNet (FN) is a corpus-based resource, documenting the semantics of a natural language by linking the “lexical units” (or form-meaning pairings) of the language, such as words, to “frames”. Frames represent the background knowledge against which lexical units are understood. This background knowledge typically surfaces in how a lexical unit is used in some situation, together with syntactically related units, termed “frame elements”. For example, lexical units such as *accuse*, *blame* and *esteem* all have in common a JUDGEMENT frame, since they typically involve “a Cognizer making a judgment about an Evaluatee” (such frame elements are usually presented capitalized).

This notion of a “Frame Semantics” has been pursued by Charles Fillmore and colleagues for over 4 decades, with a vast body of research to support the approach (e.g. Fillmore 1982, Fillmore et al. 2003), much of which can be accessed from the Berkeley FrameNet website.<sup>1</sup> Fillmore’s key insight is that an individual’s use of specific items in their language is structured by the background knowledge referred to above. Thus, expressing notions of judging draws upon a “domain” of vocabulary whose elements somehow presuppose a schematization of human judgment and behavior involving notions of worth, responsibility, judgment, etc.” (Fillmore 1982). This enables generalizations to be made about natural language patterns in terms of frames, which the FN seeks to capture.

A FN for a natural language thereby provides a rich and highly nuanced model of the syntactic and semantic patterns of the language. A FN project has the potential to add a number of valuable component resources to any existing corpus:

- a) Fine-grained information about grammatical roles and relations.
- b) A searchable database of semantically oriented annotations.

- c) Easily accessible and semantically organized example sentences, especially useful for language learning and teaching.
- d) Detailed annotations in a gloss language, such as English in the case of the EAFN project, also a significant resource for language learning and teaching.

The EAFN will be an invaluable resource for primary theoretical research on Emirati Arabic, as well as for additional forms of research crossing a number of disciplines, including natural language processing, information retrieval, corpus linguistics, second language acquisition teaching and research, machine translation, psycholinguistics, and artificial intelligence. FNs are currently available for such major languages as English, German (Rehbein et al. 2012) and Japanese (Ohara 2012). FNs typically accompany a corpus resource of some description, in the target language, and the EAFN will employ data from the EAC for this purpose.

### 1.1 The Emirati Arabic Corpus

The Corpus of Emirati Arabic (EAC) was established and licensed by the Department of Linguistics at the United Arab Emirates University (Halefom et al. 2013). The EAC is a three-million-word corpus of Emirati Arabic. The data of the EAC was drawn from various naturalistic sources such as radio and TV interviews, and daily conversations. It also consists of some scripted conversations such as TV dramas and documentaries.

While the current size of the EAC is incomparable with other full-fledged corpora (e.g. British National Corpus), the EAC is the first annotated corpus of spoken Arabic (cf. other annotated corpora which are based on Modern Standard Arabic). It also serves as a useful tool for other potential research.

The EAC is fully annotated using the International Phonetic Alphabet (IPA). Narrow transcriptions are used in which detailed phonetic information instead of the citation form is described. In addition to the phonetic details, the EAC also provides further annotation including morphological boundaries (\mb), glossing (\ge), part of speech (\ps), and translation (\ft). For in-stance, Tables 1 and 2 contain two annotated examples from the EAC.

<sup>1</sup> <https://framenet.icsi.berkeley.edu/fndrupal/>

|      |  |        |      |       |         |      |            |
|------|--|--------|------|-------|---------|------|------------|
| \ref | EAC002.3                               |        |      |       |         |      |            |
| \tx  | do:k                                   | hawi:h | fu:  | jales | jso:lɛf | wja  | lɛhma:r    |
| \mb  | do:k                                   | hawi:h | fu:  | jales | jso:lɛf | wja  | lɛ-hma:r   |
| \ge  | look                                   | Hawih  | what | doing | talk    | with | the-donkey |
| \ps  | v                                      | N      | wh   | v     | v       | prep | d-n        |
| \ft  | look, hawih is talking with the donkey |        |      |       |         |      |            |

Table 1: Example from the EAC

|      |                       |
|------|-----------------------|
| \ref | EAC002.59             |
| \tx  | wes <sup>s</sup> alt  |
| \mb  | wes <sup>s</sup> al-t |
| \ge  | arrived-2sp           |
| \ps  | v-pro                 |
| \ft  | Did you arrive?       |

Table 2: Example from the EAC

## 1.2 The Emirati Arabic FrameNet Project

The EAFN project aims to describe the range of semantic and syntactic combinations of each word in our collection in each of its senses. As mentioned earlier, this collection is a sub-corpus extracted from the EAC. To add FN information, annotators perform computer-assisted annotation of example sentences, these annotations then being collected in the EAFN database.

Currently, the Berkeley English FrameNet (BEFN), which began in 1997, consists of in excess of 13,000 entries for senses of lexical units, over 190,000 manually annotated sentences, representing more than 1200 frames. Started in 2015, the EAFN project originally planned for the initial 2-year phase to collect 1000 senses of lexical units, and up to 10,000 annotated sentences, and an expectation of over 100 frames for Emirati Arabic. However, while a substantial amount of these initial objectives was met, the project was delayed until earlier this year, due to a change in circumstances for the first author. We are currently planning to complete the project by the end of this current year. Delivery of this database will represent a huge advance in knowledge about the language, and lay the groundwork for development of a rich array of corpus-based and other resources, including descriptive, computational and teaching and learning resources, for Emirati Arabic.

Our project aims to make a significant contribution to the level of resources for Arabic, and especially Emirati Arabic. The only comparable work to date is from outside the region, for example, the Leeds University Corpus, where within the Computer Science Department, the Corpus of Quranic Arabic has been developed. However, our project differs from such previous work, in that it aims to deliver large-scale information about deep-level syntactic (grammatical roles) as well as semantic (argument roles) information for this dialect of Arabic. This will involve developing novel collection materials, much of which involves using the BEFN.

Regarding research outcomes, the project aims to deliver a store of primary linguistic information about syntactic and semantic patterns of Emirati Arabic, in a detailed and

searchable database of such patterns in this language. The information stored in this database will include:

- 1) Raw sound files (from the current Emirati Arabic Corpus).
- 2) Arabic and English Transcriptions of the data (a variety of texts in Emirati Arabic).
- 3) Annotations in the International Phonetic Alphabet of the files listed in (1) above (from the current Emirati Arabic Corpus).
- 4) FrameNet annotations, including Frame Element (FE) components for each lexical unit:
  - a) Frame Element (FE) name for lexical unit
  - b) Grammatical function (e.g. subject, object, etc)
  - c) Phrase type (e.g. noun phrase)

## 2. Method

The annotation in this project combines manual and automatic annotation techniques, and integrates these at several points, as explained below.

### 2.1 FrameNet Annotation

Formally, FN annotations are sets of triples that represent the FE realizations for each annotated sentence, each consisting of the frame element's name (for example, Food), a grammatical function (say, Object) and a phrase type (say, noun phrase). Working these out for a newly encountered language requires a range of decisions to be made. The first stage of our project involved developing a manual annotation protocol, as well as preparing the sub-corpus of EAC texts for annotation (e.g. extracting citation forms for lexical units).

Developing a FN typically proceeds as follows (Fillmore and Atkins 1998, Fillmore et al. 2003, Boas 2009):

- 1) Select the words to be analyzed.
- 2) Starting from the primary corpus (for the proposed project, this is the Emirati Arabic Corpus), define frame descriptions for these words by:
  - a) first, providing in simplified terms a description of the kind of entity or situation represented by the frame,

- b) next, choosing labels for the frame elements (entities or components of the frame),
  - c) finally, collecting words that apparently belong to the frame.
- 3) Next, focus on finding corpus sentences in the primary corpus that illustrate typical uses of the target words in specific frames.
  - 4) Then, the sentences from (3) are annotated by tagging them for frame elements.
  - 5) Finally, lexical entries are automatically prepared and stored in the database.

Building a FN for a language from scratch involves a range of decisions, both linguistic and non-linguistic, raising questions about having sufficient data, about the kind of information to include (dependent on the size and scale of the project aims), and also about the tools required to carry out the work. Relatedly, there are questions about the overall approach to building the FN, such as, whether to employ largely manual or automatic techniques, there being advantages and disadvantages on both sides. As can be seen from the above outlines of a procedure for annotating frames, the complexity of annotating semantic information means manual annotation would be expected to yield higher quality data, although relatively much more expensively, whereas automatic annotation would potentially yield much more, lower quality data albeit far more cheaply.

In our project, we have combined manual and automatic annotation procedures, to maximize quality and yield, over the longer term of the project itself. Having a foundation of manually annotated frames provides for the EAFN a solid core on which to build our database. On the other hand, we faced a lengthy lead-in time for developing suitable software tools for the automatic annotation, and so having the manual annotation track enabled an immediate start on frame collection. Further, and perhaps more importantly, the manually collected gold-standard can be used to evaluate the output of automatic annotation, and in turn, manual annotators are able to evaluate the results of automatic annotation.

It might at first seem counter-intuitive that such a resource can indeed be constructed automatically, given the semantic complexity of natural language. Ambiguity abounds in daily communication, making the proposal that a computer system could somehow automatically perform accurate and reliable annotation a somewhat dubious one. However, it turns out that a key factor in being able to achieve this is the generality of the notion of frame, in particular its definition in usage-based terms: this definition leads us to expect that there is a significant overlap between the set of frames in one language and a completely unrelated language, since a frame consists of knowledge about the situations in which a specific language is used, and a significant number of such situations are common across languages. For example, while currencies and even protocols for proper financial arrangements may differ from country to country, the Transaction frame, wherein goods are exchanged for tokens or other goods of equal worth, is ubiquitous across language settings, covering a

range of activities, such as buying, selling, bartering, trading, and the like. The automatic side of the project aims to build resources able to leverage this generality of frames, and thereby interface the English FN with an Arabic language resource, in order to capture frames common across each language. Of course, this generality is known to be limited (e.g. Boas 2009), although, we have anticipated this with the manual annotation side of our project, which provides a capacity within our project for discovering frames unique to (Emirati) Arabic. Of course, we acknowledge the difficulty of the challenge involved in being able to build such a resource for generating frames across distinct languages (on this, see e.g. recent work by Czulo et al. 2019).<sup>2</sup> However, we are heartened by a range of results, particularly using more recent, scaled up data-driven approaches to Machine Translation, where Deep Neural Networks are making significant gains in automating the task of relating the semantics of one language to another,<sup>3</sup> and such work is already yielding impressive results (e.g. ElJundi et al. 2019).

### 2.1.1 Manual annotation

One standard approach to building a large-scale resource like a FN is to construct a representative sample of the language, to carry out any required corpus analysis. Manual annotation on the EAFN follows this route, and starts from a sub-corpus specially selected from the EAC for this task.

In spring 2014, a research collaboration was established between the UAE University and the University of Birmingham with the aim of enriching the EAC by providing frame annotations. In particular, the research purpose is to annotate the EAC by adopting the framework laid out by the Berkeley FN (Baker et al. 1998). Researchers at the UAEU manually annotated the EAC with frames. Manual annotation was initiated with native Arabic speaker annotators being trained by the main EAFN researchers in frame annotation, in line with the protocols established by the Berkeley English FN (see section 2.1 above). Annotators then carried out annotation of sentences sample from the EAC.

Below are two examples of the same lexical unit ʔəmʃii which stems from the tri-consonantal root mʃʔ. All conceptual frames are arrived at through corpus-driven techniques, rather than through native speaker introspection. Note that for these initial stages of the EAFN, labels for frames and FEs have been largely drawn from the Berkeley English FN, although we fully anticipate this will need to be revised as the project further develops.

<sup>2</sup> We would like to thank an anonymous reviewer for pointing this out (complete with reference).

<sup>3</sup> For a very recent example of this, see work by the Tsinghua University Natural Language Processing Group (<https://github.com/THUNLP-MT/THUMT/>)

|                             |                               |
|-----------------------------|-------------------------------|
| <code>\entryid</code>       | EAC001.2                      |
| <code>\root</code>          | mʃʔ                           |
| <code>\lexeme</code>        | ʔəmʃii                        |
| <code>\gloss</code>         | Walk                          |
| <code>\pos</code>           | verb                          |
| <code>\frame</code>         | Self-Motion                   |
| <code>\corefe1_label</code> | Goal                          |
| <code>\corefe1_item</code>  | Deliila                       |
| <code>\corefe1_gloss</code> | Deliila                       |
| <code>\corefe2_label</code> | Path                          |
| <code>\corefe2_item</code>  | xəTTi                         |
| <code>\corefe2_gloss</code> | along with                    |
| <code>\example</code>       | ʔəmʃii ʕalaa xəTTi deliila    |
| <code>\free_trans</code>    | I walk on the path to Deliila |

Table 3: Example from the EAFN

|                             |                             |
|-----------------------------|-----------------------------|
| <code>\entryid</code>       | EAC0016                     |
| <code>\root</code>          | mʃʔ                         |
| <code>\lexeme</code>        | ʔamʃii                      |
| <code>\gloss</code>         | Walk                        |
| <code>\pos</code>           | verb                        |
| <code>\frame</code>         | Self_Motion                 |
| <code>\corefe1_label</code> | Path                        |
| <code>\corefe1_item</code>  | fi ha Siihra                |
| <code>\corefe1_gloss</code> | in the desert               |
| <code>\example</code>       | ʔamʃi fi ha Siihra          |
| <code>\free_trans</code>    | I will walk in this desert. |

Table 4: Example from the EAFN

The initial annotation process was carried out iteratively in two phases. During the initial development phase, annotators built the database using backslash entries, as demonstrated in Tables 1 to 4. In the second phase this backslash database was converted into an XML database using custom built parsers; for this phase, the initial annotation protocol can be refined, involving reconsideration of the range of categories required for annotating frames in Emirati Arabic, as well as the procedures for this annotation. For this first round of annotations, these phases gave rise to the foundation of the EAFN database; subsequent rounds of annotations continue to employ both phases, enabling a relatively flexible arrangement. Furthermore, this approach to building a database requires a minimal setup of a laptop on which to run a text editor, making the task highly mobile and relatively technology independent, with annotators employing relatively lightweight tools. Note that the flexibility of such a set-up potentially facilitates collecting such data in a more typical fieldwork type setting. Finally, by extending the custom parsers for the backslash database, we can extract the required information as XML, thereby making our database (re)usable in a range of ways.

### 2.1.2 Automatic annotation

This side of the project brings together a variety of Natural Language Processing tools, aiming to construct a state-of-the-art system for automatically generating frames for Emirati Arabic. There have been a variety of attempts to

use the Berkeley English FN to help build FNs in other languages (De Cao et al. 2008, Tonelli et al. 2009), often by linking existing electronic resources, such as a dictionary, in a target language to the English FN in some way, in order to label items from this language with frames from the English FN.

Along these lines, our approach makes use of the English FN (i.e. the Berkeley English FN), and the English and Arabic Wiktionaries. In order to link these resources, we have customized available NLP tools, and also built such tools from scratch, in order to use these resources to derive candidate frames for the EAFN, based on those from the English FN. A major part of this work has involved using the tools made available by the Ubiquitous Knowledge Processing (UKP) Lab at the University of Darmstadt in Germany.<sup>4</sup> In particular, we employed tools for parsing the English and Arabic Wiktionaries, the Java-based Wiktionary Library (JWKTL),<sup>5</sup> and the UBY database<sup>6</sup> (Gurevych et al. 2012).

Considering the UBY database first, we built tools for extracting information from the UBY database, in order to bridge the English Wiktionary and the English FN. This database stores a wealth of Wiktionary-related information across a range of languages, such as English and Arabic, as well as links to other resources, in particular the English FN. We extracted the following information from this:

- 1) For each English Wiktionary lexeme:
  - a) Its written form
  - b) Its sense
- 2) For each English FrameNet lexical unit matched to an English Wiktionary lexeme:
  - a) Its index in the English FN
  - b) Its UBY definition [essentially a gloss]

As well as supplying a ready-made parser for the English Wiktionary, the JWKTL library provides the means for customizing a parser for the Arabic Wiktionary; while wiktionaries largely overlap in their format, there can be significant differences from one language to another.

Actual entries in individual language wiktionaries contain information about a specific lexeme in that language, but also, importantly for our purposes, links to translations of this lexeme in wiktionaries of other languages; e.g. the English Wiktionary entry for book links to the Arabic Wiktionary entry for كِتَاب (this Arabic word being a direct translation of the English).

Using the newly customized parser for the Arabic Wiktionary, and the one already available for the English Wiktionary, we were able to collect information from both wiktionaries, as follows – for each lexeme in the English Wiktionary, we collected:

- 1) Word form
- 2) Part-of-speech
- 3) All possible definitions for this lexeme

<sup>4</sup> <https://www.informatik.tu-darmstadt.de/ukp/>

<sup>5</sup> <https://dkpro.github.io/dkpro-jwktl/>

<sup>6</sup> <https://dkpro.github.io/dkpro-uby/>

- 4) The lexeme in the Arabic Wiktionary which the English lexeme has been linked to. For each of these Arabic lexemes, we also collected:
  - a) Word form
  - b) Part-of-speech
  - c) Definition [supplied in English]

Now, these links between the English to Arabic Wiktionaries are one-to-many, in that there are many possible Arabic word forms for each English lexeme. This means we need to carry out a disambiguation of some kind, if we are to properly align the FN and Wiktionary resources. Taking this need for disambiguation into account, we proceed with the alignment in two stages:

- 1) First, for each English Wiktionary lexeme from the UBY database, we split the list of English Wiktionary definitions, and calculate a measure of the similarity between this lexeme’s UBY definition and its Wiktionary definition. For this work, we used the Gensim word2vec tools,<sup>7</sup> and trained models for this based on the so-called “1 Billion Word Language Model Benchmark”.<sup>8</sup> We use this similarity measure as part of an automatically derived overall confidence score, which we later use when comparing competing frame entries in the database.
- 2) Second, we align the English Wiktionary definitions with the Arabic Wiktionary definitions, again calculating a similarity measure between these definitions (with the same set-up for Gensim word2vec referred to above), as another automatically derived component of the above-mentioned confidence score.

The automatically collected frame annotations of items from the Arabic Wiktionary, currently consist of lexical units (i.e. pairing of lemma and frame), including confidence measure derived from measuring the strength of the match between the English FN and Wiktionary definitions, on the one hand, and between English and Arabic gloss-es/definitions, on the other. Future work will involve extending this work to include annotations of Frame Elements.

## 2.2 Corpus progress

While the initial release of the EAFN is still un-der development, immediately below we provide a snapshot of the current data collection, for the initial stages of each collection track. In the next section, we present more detailed evaluations of both the automatic and manual collection efforts.

Currently the EAFN covers verbs only. For manually gathered entries, we have collected 29 frames, and 360 LUs. As we show later in this section, in initial evaluation studies, we have found reasonably high inter-annotator agreement for the manual annotation. We have also implemented a fully automatic procedure for collecting entries, for which we have gathered 630 frames and 2100

LUs. Of course, such results need to be treated with a great deal of caution, and indeed initial evaluation of this data suggests only a fraction of this data is expected to be of sufficient quality to justify its being retained for the initial release of the EAFN database.

While we are listing manually and automatically collected entries separately at this stage, these will be collected together for the initial release of the database.

Finally, we should also emphasize that the two sources of language are different in dialectal terms: the manual track works directly from the EAC, and so the yield is dialect-based, whereas the automatic track works from the Wiktionary, which is in fact closer to the Modern Standard Arabic dialect. This combination of dialects within the same resource raises many issues, and we intend to begin addressing these during the latter part of the current project, which constitutes the initial development stage of the EAFN. However, it is likely that more comprehensive solutions to the issues raised will be solved in later stages of the EAFN, once we have completed the initial release of the database.

## 3. Evaluation

Semantic annotation is fraught with issues regarding lack of reliability and accuracy, making quality control of data a key component of any project in this area. While our project is still at an early stage of development, we are working toward an initial release of our data, for which we are developing a comprehensive evaluation regime, incorporating both the manual and automatic annotation tracks. A description of this, as well as some early results, are included in the rest of this section.

### 3.1 Manual track

#### 3.1.1 Procedure

We are currently piloting several evaluation tasks, targeting accuracy of judgements about frames and the core elements of those frame. For these tasks, we first extract a random sample from the EAC, and annotators then carry out annotation of this data according to the annotation protocols we have developed (see Section 2 above). We then proceed to apply various measures of agreement between the annotators

We have several measures of the quality of this data, centering on degrees of overlap in the annotations of two of the annotators currently involved in the collection efforts at the UAEU. The statistic we are using here is Cohen’s kappa coefficient  $k$ :

$$k = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

Where  $\Pr(a)$  models the probability of observed agreement among raters, and  $\Pr(e)$  captures chance agreement; the higher the value for  $k$ , the better the agreement between annotators. There are various interpretations of such scores, for example, 0.60 is often considered a threshold, with

<sup>7</sup> <https://radimrehurek.com/gensim/>

<sup>8</sup> <http://www.statmt.org/lm-benchmark/>

scores above this being taken to indicate “substantial agreement” (Landis & Koch 1977).  $k$  enables quantifying the inter-annotator agreement (IAA), particular for qualitative data, which is closer to our evaluation task, involving as it does detailed semantic knowledge.<sup>9</sup>

### 3.1.2 Results

For comparison of frame annotations on our sample, we achieve the following:  $k = 0.790$  (p-value  $\ll .001$ ,  $N = 31$ ). For annotation of core FEs, we achieve,  $k = 0.899$  (p-value  $\ll .001$ ,  $N = 31$ ). This shows that using the protocol we have devised, annotators are achieving very good levels of agreement for judgements about FEs, and acceptable agreement for judgements about frames.

## 3.2 Automatic track

### 3.2.1 Procedure

Evaluating the automatic annotation provides a key point of convergence between the two tracks. For this, the manual annotators evaluated the output of the automatic system, their responses to the automatically generated frames requiring them to draw on their intuitions, which have their foundations in their direct experience building the manual collection of frames. Feedback from the annotators is crucial to pinpoint where further development on the automatic system will be required. In this way, our aim is that the automatic track more closely approximates the results from the manual track.

The procedure we followed here involved manual annotators going through individual, automatically generated LUs, complete with brief information about the target LU, as well as the frame assigned to this LU. Each annotator was given a total of 198 randomly sampled lexical units to evaluate. Annotators rated this on the following 5-point scale: 1 = Completely correct, 2 = Mostly correct, 3 = Acceptable, 4 = Mostly incorrect, 5 = Completely incorrect. The sample was further split according to two conditions: either (1) the rendering of the lexical unit in Arabic script included vowel information, or (2) it did not. For Arabic script, information about vowels can help disambiguate LUs, and potentially influence the ratings assigned for any specific LU. We are interested in investigating such aspects of the automatic collection process more closely.

The key statistic we are using here is Cohen’s kappa coefficient, the same statistic used for measuring agreement during evaluation of the manual annotation task. The difference for the task of evaluating the automatic annotation, is that this task results in ordered data (a likert scale), and so we need to use weighted kappa coefficients; specifically, we are using squared weights, whereby disagreements are weighted according to their squared distance from perfect agreement.

<sup>9</sup> For all of this, we have used the **irr** package in R, which has been specifically designed for modelling “interrater reliability and agreement.”

### 3.2.2 Results

Table 5 presents the results of this evaluation, with evaluation categories used by both annotators across the top and down the leftmost column, and inside the table showing how scores matched for each item. From this, we can see that by far the largest number of matches is where annotators agree that an item is “completely correct”, and the next highest being where one an-notator thought that an item was “mostly correct” and the other annotator thought the same item was “completely correct”.

When ignoring the vowel vs. no-vowel condition, we achieve the following:  $k = 0.443$  (p-value value  $\ll .001$ ,  $N = 198$ ). However, when taking into consideration the vowel vs. no-vowel condition, this score improved somewhat:  $k = 0.602$  (p-value value  $\ll .001$ ,  $N = 83$ ).

Overall, we can see that general agreement between annotators is quite low, despite the overall largest match being “completely correct”. This suggests possible problems and indeed errors for many of the automatically collected frames. On the other hand, when we partition the data set, and extract those items with vowel information, for this subset, the IAA improves considerably, suggesting that such information is an important component to incorporate in future automatically acquired collections for the EAFN.

|   | 1  | 2  | 3 | 4 | 5 |
|---|----|----|---|---|---|
| 1 | 24 | 15 | 4 | 1 | 1 |
| 2 | 2  | 7  | 0 | 2 | 0 |
| 3 | 1  | 1  | 0 | 2 | 0 |
| 4 | 1  | 3  | 1 | 3 | 1 |
| 5 | 2  | 2  | 0 | 1 | 9 |

Table 5: Evaluation of automatic track (1 = Completely correct, 2 = Mostly correct, 3 = Acceptable, 4 = Mostly incorrect, 5 = Completely incorrect)

## 4. Conclusion

We have presented early results for the first iteration of the Emirati Arabic FrameNet (EAFN). The EAFN is a general semantic resource for the Arabic language, which is sure to be of interest to a range of researchers, from those in linguistics, to others working within natural language processing. The project is divided into manual and automatic tracks, based on the predominant techniques being used to collect frames in each track. Despite a hiatus, work on the EAFN has recommenced; we have here reported on initial results for annotations and evaluation of these annotations which have been carried out in both tracks. The EAFN is well on target for the first release of data in the coming year.

## 5. Acknowledgements

We would like to thank S. Al Helou and W. Ghaban for help with the evaluation. We would also like to thank M.

Elkaref for some early technical help with the JWKTL library, as well as valuable discussions about carrying out computational linguistic modelling of Arabic. Finally, we would also like to thank J. Ruppenhofer for discussions about the English FrameNet.

Processing: Volume 1-Volume 1 (pp. 276-285).  
Association for Computational Linguistics

## 6. Bibliographical References

- Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998). The Berkeley Framenet project. In Proceedings of the 17th international conference on Computational linguistics-Volume 1 (pp. 86-90). Association for Computational Linguistics
- Boas, H. C. (Ed.). (2009). Multilingual FrameNets in computational lexicography: methods and applications (Vol. 200). Walter de Gruyter.
- Czulo, O., Torrent, T. T., da Silva Matos, E. E., da Costa, A. D., & Kar, D. (2019). Designing a Frame-Semantic Machine Translation Evaluation Metric. In 2nd Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT 2019) (pp. 28-35).
- De Cao, D., Croce, D., Pennacchiotti, M., & Basili, R. (2008). Combining word sense and usage for modeling frame semantics. In Proceedings of the 2008 Conference on Semantics in Text Processing (pp. 85-101). Association for Computational Linguistics.
- ElJundi, O., Antoun, W., El Droubi, N., Hajj, H., El-Hajj, W., & Shaban, K. (2019). hULMonA: The Universal Language Model in Arabic. In Proceedings of the Fourth Arabic Natural Language Processing Workshop (pp. 68-77).
- Fillmore, C. (1982). Frame semantics. *Linguistics in the morning calm*, (pp. 111-137). Seoul, South Korea: Hanshin Publishing Co.
- Fillmore, C. J., & Atkins, B. S. (1998). FrameNet and lexicographic relevance. In Proceedings of the First International Conference on Language Resources and Evaluation, Granada, Spain (pp. 28-30).
- Fillmore, C.J., C.R. Johnson, and M.R. L. Petruck. (2003). "Background to FrameNet." *International Journal of Lexicography*. 16.3: 235-250.
- Gurevych, I., Eckle-Kohler, J., Hartmann, S., Matuschek, M., Meyer, C. M., & Wirth, C. (2012). Uby: A large-scale unified lexical-semantic resource based on LMF. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (pp. 580-590). Association for Computational Linguistics.
- Halefom, G, T. Leung and D. Ntelitheos. (2013). A Corpus of Emirati Arabic. NRF Grant 31 H001, United Arab Emirates University.
- Landis, J.R. and G.G. Koch (1977). The measurement of observer agreement for categorical data. *Biometrics*. Vol. 33, pp. 159—174.
- Ohara, K. (2012). Semantic Annotations in Japanese FrameNet: Comparing Frames in Japanese and English. In LREC (pp. 1559-1562).
- Rehbein, I., Ruppenhofer, J., Sporleder, C., & Pinkal, M. (2012). Adding nominal spice to SALSA-frame-semantic annotation of German nouns and verbs. In Proceedings of the 11th Conference on Natural Language Processing (KONVENS'12) (pp. 89-97).
- Tonelli, S., & Giuliano, C. (2009). Wikipedia as frame information repository. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language