

# Building the Emirati Arabic Framenet

Andrew Gargett (STFC, UK) & Tommi Leung (United Arab Emirates University)

## Aims

- Initiate a Framenet for Emirati Arabic, utilizing the Emirati Arabic Corpus (EAC, Halefom et al. 2013)
- Create a resource comparable to the initial stages of the Berkeley Framenet (Baker et al. 1998)

## Challenges

- Very few linguistic or computational resources currently available
- The process of creating a Framenet from scratch is highly resource intensive

## Approaches

- Re-purpose available resources (e.g. corpora, software), to facilitate cross-disciplinary work (i.e. corpus linguistics, digital humanities)
- Attempt to automate as much of the process as possible
- Combined evaluation of manual and automatic “tracks”

## Key features of the EAFN

- Fine-grained information about grammatical roles and relations.
- A searchable database of semantically oriented annotations
- Easily accessible and semantically organized example sentences, especially useful for language learning and teaching
- 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

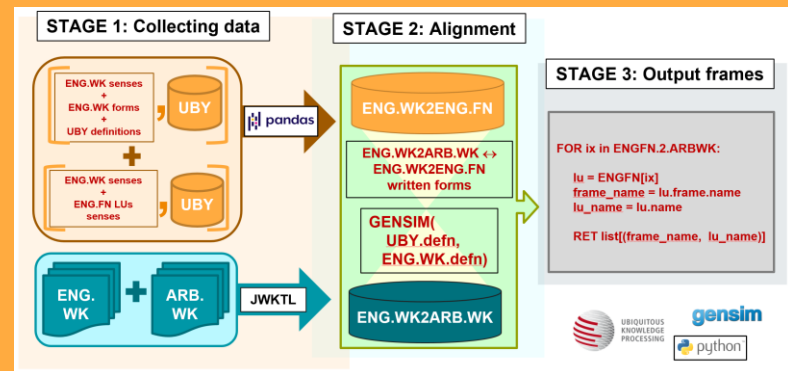
## Corpus progress

- **Manual track:** 29 frames + 360 Lexical Units
- **Automatic track:** 630 frames + 2100 LUs

## Plans

- **Mid-2020:** Re-starting both manual and automatic annotation tracks
- **Later 2020:** Planned release of first version of the corpus

## Automatic track pipeline



## Evaluation

- **Manual track:** measuring inter-annotator agreement (kappa scores)
- **Automatic track:** ratings by annotators (plus vowels vs. no-vowels conditions)

## Results

- **Manual track:**
  - Frame annotation:  $k=0.790$  (p-value  $\ll .001$ ,  $N=31$ )
  - Annotation of core FEs:  $=0.899$  (p-value  $\ll .001$ ,  $N=31$ )
- **Automatic track:**
  - Vowels included:  $k=0.443$  (p-value  $\ll .001$ ,  $N=198$ )
  - Vowels not included:  $k=0.602$  (p-value  $\ll .001$ ,  $N=83$ ).

## Conclusions

- First iteration of the Emirati Arabic Framenet
- Integrating manual and automatic approaches
- Initial results for manual track good, but automatic track was mixed
- Aiming for initial release of the EAFN in later 2020