Multilingual FrameNet (MLFN)

Online meeting
February 5, 2017

Welcome!

- Thanks for attending
- Zoom software
- Recording meeting
- Mics, background noise

Overview

- Parallel annotation exercise
- Aligning existing FrameNets
- Future plans

Parallel Annotation Exercise

- Choice of text
    - Translations available, freely distributable
    - The problem of translationese
    - TED Talk (2006): Ken Robinson: "Do schools kill creativity?"
    - Wikipedia article "Solar system"
    - UN Universal Declaration of Human Rights?
    - Others?

Data preparation

- Get texts, split into 1 sentence per line
  - Remove time marking, "(Laughter)"
  - Dealing with long quotation
- Upload lexicon -- only WFs that occur in the talk or all?
- Upload labels
  - GFs and PTs
  - "Other" and POS-specific layers
  - Language-specific layers and labels
  - Frames and FE names will come from Berkeley FN
- Send names & e-mails of participants and desired status (annotator or admin)

Hosting sites

- Data for annotation prep on GitHub
https://github.com/ICSI-berkeley/
multilingual_fn
- Open source site-- anyone can view, download
- Need permission to modify, choose person
- We hope to host parallel annotation exercise on NSF-supported "cyberinfrastructure" sites
  - Jetstream and XSEDE

Annotation

- (Tiago)

Alignment of Existing Data

ICSI created a single MySQL database of the data from all the projects as of 2015. Summary on the next slide

- We plan to redo this in a new, graph database, rather than a relational DB.

- We urgently need the latest data from each project, and often we'll need help interpreting it

Project	Frames	Frames ~ ICSI	LUs ~ICSI	Unmatched Frames	Unmatched LUs	~Frames 1 LU)	<> Frames (1 LU)
JFN	977	904	3117	73	288	541	61
SpaFN	1196	1145	1154	51	115	182	19
SALSA	1023	254	992	769	834	254	769
FNBrCopa	384	35	140	349	985	29	318
FNBr	1092	1081	242	11	9	108	4
FBK	38	36	211	2	6	36	2
UNIPI_2	33	31	38	2	2	21	1
UNIPI_3	37	35	32	2	2	19	1
UNIPI_1	37	35	177	2	3	35	2
SweFN	1186	1099	30704	87	7012	921	65
ICSI_FN_1-6	1205	1205	13315	0	0	1059	
FrFN	148	80	1325	68	1265	73	64
CFN	323	238	3429	85	518	215	78

Aligning lexical DBs

- Aligning frames
  - Equivalence by frame name or ICSI FN ID number
  - Translation equivalent LUs using lexical resources
    - Multilingual Open WordNet (Bond & Foster 2003)
    - Wiktionary
    - BabelNet (Navigli. & Ponzetti 2012)
- Cross-lingual frame relations besides equivalence
  - Subset/superset ?≈ inheritance, perspective on
  - Scenarios and subframes, narrow vs. broad
- Can frame relations in Berkeley FN be projected?

Future plans

- More MLFN meetings

- Publicity and training sessions

- Systematic tracking of FN-related research, bibliography

- Support for NLP, e.g. new data formats, LLOD, NLTK

- Annotation tools for instructional purposes/ crowdsourcing

- Expand FrameNets by cross-linguistic projection