

Other

Farsi (Behrang QasemiZadeh et al.)

Greek (Voula Giouli et al.)

Hebrew (Yaakov Ha-Cohen Kerner, Chaya
Liebeskind)

Hungarian (Veronika Vincze et al.)

Maltese (Lonneke van der Plaas, Mike Rosner)

Turkish (Gülşen Eryiğit et al.)

- **2 annotators/language**

Progress so far

Language Group	Pilot Annotation Phase 1	Pilot Annotation Phase 2
Farsi	finalized	in progress
Greek	finalized	almost there
Hebrew	NA	in progress
Hungarian	finalized	finalized
Maltese	finalized	in progress
Turkish	finalized	finalized

Inter-annotator agreement

Language	Token#	A1	A2	F-Score	Cohen's κ
Farsi	4,913	1.63	1.335	0.708	0.639
Greek	6,313	1.213	1.153	0.638	0.428
Hebrew	NA	NA	NA	NA	NA
Hungarian	4,607	0.456	0.118	0.325	0.914
Maltese	NA	NA	NA	NA	NA
Turkish	NA	NA	NA	NA	NA

some remarks...

- 2nd phase corpora not comparable...
 - newspapers articles
 - Wikipedia
 - readers' comments posted under news articles
 - subcorpora extracted from large pre-processed corpora
- 2 teams have reported to prepare language-specific guidelines
- qualitative reporting is under preparation
- trying to consolidate close co-operation

(language-specific??) issues

- disambiguation between LVCs and IDs
- markable extent (auxiliaries, negative and future particles)
- elliptical LVCs (Farsi, Greek)
- coordinated LVCs that share the same V head