

PANACEA WP7

Evaluation of components,
integration and resources
produced

WP leader: CNR-ILC

WP participants: UPF, ILSP, UC, LG,
DCU, ELDA

Objectives (in DoW)

WP7: Internal evaluation:

- Evaluation of the integration components
 - Evaluation of components
 - Evaluation/Validation(?) of the integration
- Validation and evaluation of resources produced/information acquired
 - Validation of formats (?) and documentation according to agreed standards/interfaces formats
 - Validation/Evaluation of acquired data

Workplan T1-T6

- Define relevant evaluation protocol/tasks for resources to be produced by PANACEA and collect evaluation data for all the languages involved
- Identify gaps and start constructing missing datasets
- Define baselines for all the different evaluation tasks within PANACEA
- Fix criteria, for all evaluation cycles (?), in D7.1

T7.3: Evaluation of the resources

- What are the key tasks/resources for the success of the Platform?
- What components/resources do we want to evaluate/assess intrinsically? → only those that produce key resources
 - WP4: “parallelness” of corpora;
 - WP5: bilingual dictionaries;
 - WP6: monolingual dictionaries (subcat, lex pref, lex class?)
- Which languages for which component?
- What domain(s)?

T7.3: Evaluation of the resources

- WP4:
 - Crawler; Cleaner/normaliser; Tokeniser (?); Lemmatiser (?); Pos tagger (?); Chunker (?); Shallow parsing (?); (Term extractors (??))
- WP6:
 - subcat frames: (eval against dictionaries, manual)
 - Selectional preferences
 - Verb classes (unsupervised classifier)
 - MW:
- What kind of evaluation/quality assessment?

T 7.2: Evaluation of the platform. Integration of components

- Define our view of evaluation of the integration
 - Fulfill platform requirements: robustness, scalability, usability
 - manual evaluation: Yes/no on requirements
 - who?
- Evaluation as components ws?
 - → ECESS platform (TC-STAR)



Task 7.4: Evaluation for MT

Extrinsic Eval.

- WP5:
 - Sentence aligner (?)
 - Phrase aligner
 - Evaluation= check for comparability
 - Baseline: phrase alignment in Moses
 - Word aligner (?)
 - No intrinsic evaluation? Only extrinsic, MT, eval?
 - Dictionary induction
 - Bilingual dictionaries
 - MW translation
 - Baseline: word alignment in MOSES
 - Evaluation?? Manual or automatic?
 - » ?Compare against existing dictionaries (for “precision”), and possibly storing non matching items separately or with a tag for manual evaluation (this could be done in Panacea as proof-of-concept or deferred to the end users)

→Link to TTC?

→See Accurat for measures/baselines

Task 7.4: Evaluation for MT Extrinsic Eval.

- WP5:
 - Transfer grammars
- Eval:
 - Use EuroParl as general domain for baseline?
 - Need reference translations (500-2000 sentence pairs). What languages?

Survey on the state-of-the-art

- scientific/overview paper on evaluation methodologies
or
- survey of evaluation of components in PANACEA

Workplan: next deadlines

- 15 may: first criteria (baselines, test set, languages)
- Final D7.1: end June
- First Draft: 1 June (inputs from all)
- 7 June (circulate 1° draft harmonised: ILC)

Deliveries

- D7.2. Report of eval. of first integration cycle: Factory and Corpus (D3.2, D4.3)
- D7.3 Report of second integration cycle: Factory, Aligners, Aligned Corpora
- D7.4 Report on third integr. Cycle: everything else (Annotated corpora, Bilingual Dictionary, Lexica, final platform)