

CLARA Mid-term Review Meeting

Dubrovnik, Sep. 19, 2011

SP3-People-ITN 238405

Tour de table

- REA: Concepción (Conchita) Perez Camaras
- Independent expert: Αθηνά Μαρκαντώνη

Beneficiaries

1. UIB: Koenraad De Smedt and Annhild Fetveit
2. UTU: Detmar Meurers
3. TILDE: Inguna Skadiņa
4. UCPH: Bolette Pedersen
5. UPF: Núria Bel
6. UHEL: Krister Lindén
7. CUNI: Markéta Lopatková
8. NHH: Gisle Andersen and Marita Kristiansen
9. MPI: Przemysław (Przemek) Lenkiewicz

Associate partners

1. FFZG: Marko Tadić
2. Uni Research (previously AKSIS): Victoria Rosén

Coordinator's report

- a. Scientific
- b. Training
- c. Networking
- d. Management

Scientific: Need for training

Europe needs trained researchers with competences and skills in:

- Research in Language Resources and Technologies
- Building language infrastructures (CLARIN, Meta-Net, etc.)
- Language industries producing ICT applications

Overall research objectives

1. To develop the next generation of data-intensive language models and applications by integrating approaches across language and country boundaries
2. To contribute to the establishment of a pan-European infrastructure for language resources.

Research themes

1. Designing and Testing Common Infrastructures
2. Lexical Semantic Modeling
3. Next generation Domain Modeling
4. Multimedia and Multimodal Communication Modeling
5. Applications (MT, CALL, proofreading)
6. Parsing Technologies and Grammars

Some scientific highlights so far

- UIB
 - Wolof morphology and grammar
 - Spanish-German parallel corpus from DG Enterprise and Industry for SMT
- UTU:
 - Language aware search engine
 - Semi-automatic creation of domain treebank
 - Chinese word segmentation

Scientific highlights (contd.)

- TILDE
 - Pattern-based context extraction in Russian automotive domain
- UCPH
 - Multimedia and multimodal annotation

Scientific highlights (contd.)

- UPF
 - Automatic merging of lexical resources
 - Automatic acquisition of lexical information for low-frequent words
- UHEL
 - New version of HFST toolkit
- CUNI
 - Tamil dependency treebank
 - NP bracketing in dependency parsing and SMT

Scientific Highlights (contd.)

- NHH
 - MWE terms in Spanish-English parallel corpus
- MPI
 - Infrastructure solutions in cooperation with CLARIN and Language Archive
 - Automated annotation

Multidisciplinarity and industry interest

- Recruited researchers come from Linguistics and Language Studies as well as Computer and Information Sciences
- Broadly focused research activities with potential applications in ICT for private and public sector, language teaching, etc.
- TILDE, an SME in the language industries, is a beneficiary (Level 1) in the project

Training

- Joint Training Programme of 10 training courses and schools, with cooperation from beneficiaries, associate partners and third parties
- 7 out of 10 planned training events held, one additional training event held
- Local training at host institutions
- Secondments

Joint training programme

1. [CLARA Summer School in Advanced Resource Creation, Archiving and Usage \(Nijmegen, 5-16 July 2010, 10 days, 10 credits\)](#)
2. [CLARA Thematic Training Course on Methods and Technologies for Consolidating and Harmonising Terminological Resources \(Bergen, 13-17 Sep. 2010, 5 days, 5 credits\)](#)
3. [CLARA Thematic Training Course on Methods and Technologies for Consolidating and Harmonising Treebank Annotation \(Prague, 13-17 Dec. 2010, 5 days, 5 credits\)](#)
4. [CLARA Thematic Training Course on Processing Morphologically Rich Languages \(Budapest, 11-15 April 2011, 5 days, 5 credits\)](#)
5. [CLARA Thematic Training Course on Multilingual Lexical Resources and Tools \(Bergen, 20-23 June 2011, 4 days, 4 credits\)](#)

Joint training programme

6. [CLARA Summer School on Infrastructure Tool Development \(Nijmegen, 5-12 July 2011, 10 days, 10 credits\)](#)
7. [CLARA Summer School in Semantic and Nonverbal Corpus Annotation and Evaluation \(Copenhagen, 15-26 Aug. 2011, 10 days, 10 credits\)](#)
8. [CLARA Industrial Career Training Course on Product Planning for Next Generation Information Access Technology Solutions \(Dubrovnik, 20-23 Sep. 2011, 4 days, 2 credits\)](#)
9. CLARA Winter School on New Developments in Computational Linguistics (Prague, Feb. 2012, 5 days, 5 credits)
10. CLARA Thematic Training Course on Evaluation of Human Language Technologies (Paris, Nov/Dec 2012, 4 days, 4 credits)

All course descriptions are on the CLARA website

Additional collaborative training event

- CHAT (Workshop on Creation, Harmonization and Application of Terminology Resources), May 11, 2011
 - Necessary follow-up of Terminology course
 - Organization and participation by Tilde, NHH; cooperation with TTC, Meta-Nord

Local training programmes

- ESRs get personal supervisors and are integrated in research groups and projects (e.g. INESS, CLARIN, TTC)
- ESRs participate in seminars, workshops and conferences, group meetings, courses
- ERs work more independently and learn to manage research groups and projects

Secondments

- None carried out yet (too early)
- Planned:
 - UCPH to/from MPI
 - Tilde to/from NHH
 - UPF to Temis
 - Tilde/NHH to Temis

Complementary skills

- Career course (Dubrovnik, Sep. 20-23, see website)
- Local courses in language competence, academic writing, presentation, philosophy of science, business ethics, etc.
- Training through experience in writing articles and presentations in research groups, at workshops and conferences

External participation at events

- All training events have been open to external participation
- Most events have had external lecturers
- Attendance of external researchers has varied between events
- Positive feedback from all participants

Networking

- Cooperation in the network
 - Extensive cooperation in JTP courses has led to interesting sharing of knowledge and exchange of perspective
 - Co-tutelle between UIB and UPF for shared PhD supervision
 - ESR at Tilde is PhD student at CUNI
 - UHEL cooperates with UTU on FS methods
 - Collaboration potential on research outputs may be somewhat underexploited

Associate partners (incl. private sector)

- HASRIL: participated at Morphology course
- Comperio: does not participate due to personnel change
- TEMIS: participated at Terminology course and has contact with UTU
- ELDA: will participate in Evaluation course
- FFZG: participates in Career course
- AKSIS: participated at Treebank course and in local ESR supervision
- LC: participated at Lexicon course
- Tilde uses terminology extraction tool by FFZG

Cooperation with external groups

- UTU collaborates with EF and RCEAL at Univ. of Cambridge
- MPI collaborates with Fraunhofer and U. de Catalunya
- UHEL collaborates with Aalto Univ., Lingsoft, Tromsø Univ. and Apertium
- ESR at Tilde is PhD student at Univ. of Vienna
- FP7 and CIP Projects: CLARIN, Meta-Nord, TTC
- VS affiliations: CVC Brussels, U. of Indiana, Boulder U, KTH Stockholm

Management: coordination

- Consortium agreement in place
- Consortium meetings (Copenhagen, Bergen, Luxembourg)
- Manager and controller at coordinator:
Annhild Fetveit

Management: communication

CLARA website

- External communication: vacancies, news, consortium information, events
- Internal communication: management information, people, meetings, events, etc.

So far fellows have not created any web-based communication channels among themselves (but they interact intensively at events)

It has been difficult for administrations to reach a full understanding of procedures and budgets

Recruitment

- Euraxess has been used for all ESR and ER positions
- Despite wide advertisement, finding the right candidates has sometimes been difficult
- Immigration obstacles from non-EU countries have led to delays
- Researchers have been informed of their contractual rights and obligations and their salary and allowances
- Career plans have been made

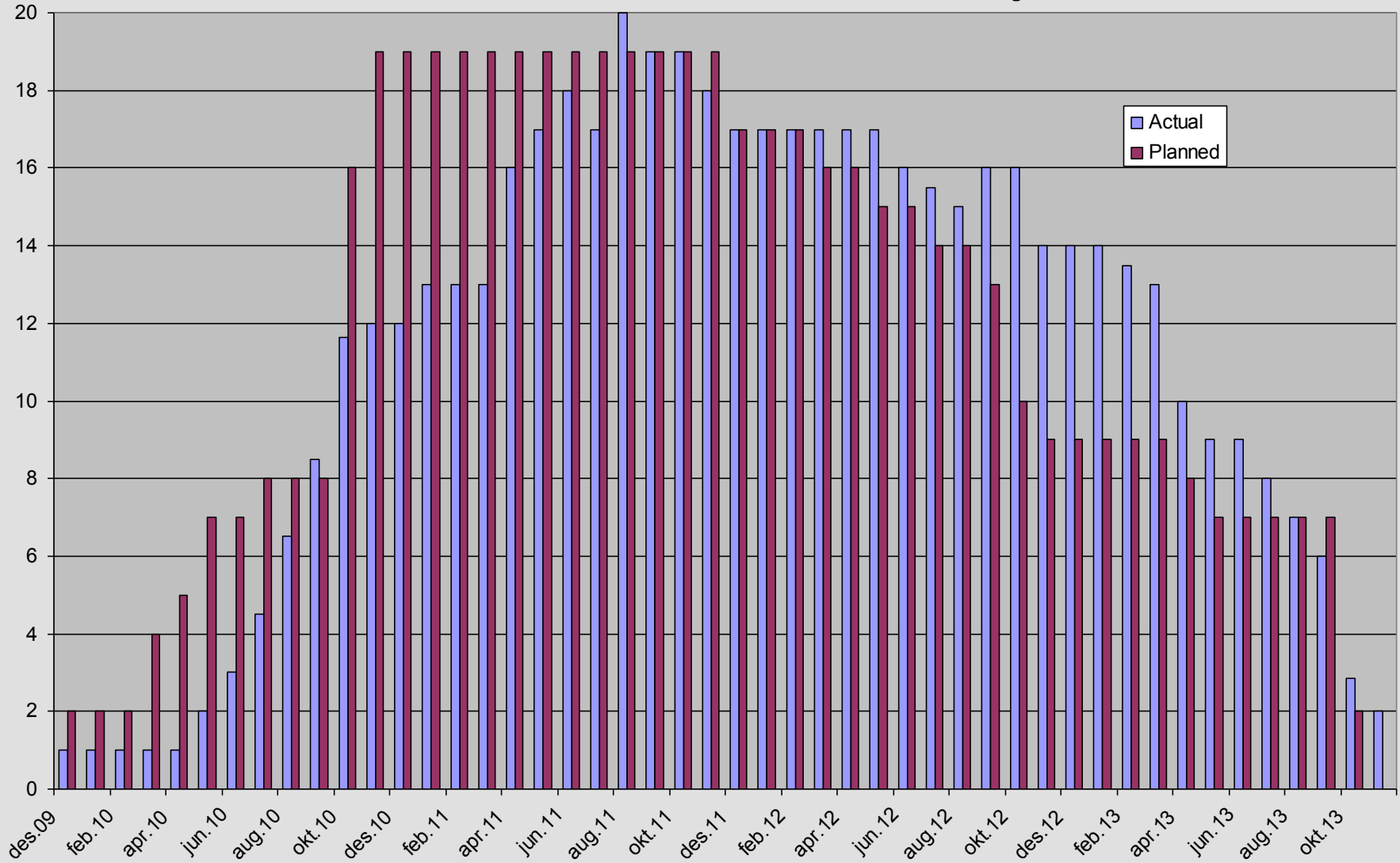
Recruited fellows - statistics

As per 1 September 2011:

- ESR: 17 (16 positions out of 17 filled)
- ER: 2 (2 positions filled)
- VS: 4 (3 positions, months as planned), contributing to training events

- 50% males, 50% females

CLARA fellows – actual/planned



CLARA

Financial aspects

- Budget vs committed costs, cat A-C
- Prepayment status
- Cat. E&F status and challenges

Budget vs committed costs A-C

| All amounts in EUR | Budget | Committed | Deviation |
|-----------------------|------------------|------------------|--------------|
| UiB | 344.254 | 343.505 | 749 |
| UTU | 283.285 | 296.333 | -13.048 |
| Tilde | 193.452 | 187.252 | 6.200 |
| UCPH | 364.958 | 367.289 | -2.331 |
| UPF | 126.189 | 121.033 | 5.157 |
| UHEL | 242.976 | 237.444 | 5.532 |
| CUNI | 205.934 | 203.596 | 2.338 |
| NHH | 159.526 | 161.086 | -1.560 |
| MPI | 398.686 | 398.905 | -219 |
| Total | 2.319.260 | 2.316.442 | 2.818 |

Financial aspects: Cat. A-C

- Low margins in budget
 - High mobility and travel allowances, more experienced VS (UTU, NHH), more than 1 VS needed (UCPH)
 - Full contract rather than fixed amount VS due to national rules (UCPH)
 - Need for top-up due to national rules (UIB, NHH, UCPH)
- Problems with recruitment have led to some contracts that extend beyond the planned project period (5 months at UIB will be financed by UIB itself)

Pre-payment status

- Payments after submission of Declarations on the Conformance (DoCs) – condition in consortium agreement
- Pre-payment amount from EU 2 168 351
- Paid out to partners EUR 2 045 194
- Remaining amount EUR 123 157
 - Relates to 1 ESR not employed, VSs recently employed and management

Category E & F - Joint training

- Cat E
 - Expenses are not balanced; while income is proportional to researcher months, expenses are related to contribution to joint training activities
- Cat F
 - External participant days vary strongly from planned number of days
- Balancing measures within the consortium will be taken at next consortium meeting (February 2012)

Outputs, dissemination

- Training events and vacant positions are well announced on CLARA website and other channels
- Researcher profiles and research outputs (publications) are on CLARA website
- Brochures and posters were printed and distributed at major events (e.g. EU Language Days, Luxembourg, 2010)

Proposed re-orientations

- Problems with recruitment at Tilde requires a shift of theme (from proofreading to translation)
- Project period may need adjustment due to parental leave at UTU
- Some adjustments to the schedule for recruitment and events in order to optimize participation
- Some individual timing adjustments due to leave of absence, etc.

Final remarks

- Challenges:
 - Finding qualified and eligible candidates
 - PhD studies for ESRs at industrial partner
 - Immigration of non-EU fellows (no fast track)
 - Struggle between national rules and MC rules
- Positive aspects:
 - Researchers with knowledge of typologically different languages are a major enrichment of multilingual NLP

Fellows: ESR and ER

- Cheikh M. Bamba Dione (ESR,M) UiB 2011-01-01 / 2013-12-31
- Carla Parra (ESR, F) UiB 2011-05-01 / 2014-04-30
- Corina Dima (ESR,F) UTU 2010-06-01 / 2012-05-31 (on parental leave)
- Jianqiang Ma (ESR,M) UTU 2011-04-01 / 2013-03-31
- Sowmya Vajjala (ESR,F) UTU 2011-04-01 / 2013-03-31
- Anne-Kathrin Schumann (ESR,F) Tilde 2010-08-16 / 2013-02-15
- Septina Dian Larasati (ESR,F) Tilde 2011-08-01 / 2013-04-30
- Hector Martinez Alonso (ESR,M) UCPH 2010-10-01 / 2013-09-30
- Magdalena Lis (ESR,F) UCPH 2010-10-01 / 2013-09-30
- Silvia Neculescu (ESR,F) UPF 2010-10-15 / 2013-10-14

Fellows (contd.)

- Måns Huldén (ER,M) UHEL 2009-12-01 / 2011-11-30 (absent due to illness)
- Senka Drobac (ESR,F) UHEL 2011-04-01 / 2013-03-31
- Loganathan Ramasamy (ESR,M) CUNI 2010-08-01 / 2013-07-31
- Nathan David Green (ESR,M) CUNI 2010-10-08 / 2013-10-07
- Pedro Patiño García (ESR,M) NHH 2010-05-01 / 2012-10-31
- Przemysław Lenkiewicz (ER,M) MPI 2010-07-15 / 2012-07-14
- Binyam Gebrekidan (ESR,M) MPI 2010-09-01 / 2013-08-31
- Anna Lenkiewicz (ESR,F) MPI 2011-08-01 / 2013-07-30
- Hugo García Blanco (ESR,M) MPI 2010-07-01 / 2011-06-30 (contract ended)