



EUROPEAN COMMISSION
RESEARCH & INNOVATION DG
HUMAN RESOURCES AND
MOBILITY

Mid-term Report

Project No: 238405

Project Acronym: CLARA

Project Full Name: Common Language Resources and their
Applications

Mid-term Report

Period covered: from to 01/12/2011

Date of preparation:

Period number: 1st

Start date of project: 01/12/2009

Date of submission (SESAM):

Project coordinator name:
Prof. Koenraad De Smedt

Project coordinator organisation name:
UNIVERSITETET I BERGEN

Version: 1

Mid-term Report

PROJECT MID-TERM REPORT

| | |
|--|---|
| Grant Agreement number: | 238405 |
| Project acronym: | CLARA |
| Project title: | Common Language Resources and their Applications |
| Funding Scheme: | FP7-MC-ITN |
| Mid-term period covered - start date: | 01/12/2009 |
| Mid-term period covered - end date: | 01/12/2011 |
| Project co-ordinator: | |
| Organisation PIC: | 999974456 |
| Organisation legal name: | UNIVERSITETET I BERGEN |
| Person in charge of scientific aspects: | |
| Title: | Prof. |
| First name: | Koenraad |
| Name: | De Smedt |
| Tel: | +47 95836558 |
| Fax: | +47 55589660 |
| E-mail: | clarin@uib.no |
| Project website address: | http://clara.uib.no |

OVERALL PROGRESS OF THE PROJECT:

Comments:

In all Work Packages, significant progress is visible, although a few delays have occurred due to recruitment and administration problems, sick leave and maternity leave.

In WP2, MPI has extended and developed the Language Archive and its software component. New digital media from researchers worldwide has been added to the MPI archive. A sophisticated search engine is being developed, crossing borders and giving access to language data and metadata hosted at renowned scientific institutes.

In WP3, a bilingual German/Spanish corpus (about 10 M words) is being compiled at UIB, comprising translated documents from the DG Enterprise and Industry Project. At UCPH, corpora are being annotated with respect to regular polysemy in the languages English, Spanish and Danish. At UPF, an ESR has already produced a dissertation.

In WP4, training includes industrial terminology management (IBM), the representation of interrelated domains in termbases, concept diagrams and ontology building, and the integration of wordnets and IE based on text mining (associate partner TEMIS). At NHH, collocations are explored in a corpus of Free Trade Agreements. Tilde works on automated methods for context extraction and enrichment of a multilingual terminology database with knowledge-rich contexts.

In WP5, MPI has extended the ELAN annotation tool and is implementing pattern recognition and search components to be integrated in a multimodal annotation framework, while UCPH is collecting, annotating and exploiting multimodal corpora. UCPH data have been coded with IMDI metadata as required by the MPI repository. Two VS (Jens Edlund and Susan Brown) have been recruited to UCPH; they are teaching at a CLARA summer school related to training for this WP.

In WP6, the scientific work on Machine Translation at CUNI has been focused on developing a dependency treebank for the highly inflected, under-resourced Tamil language, using Prague Dependency Treebank style annotation. At Tilde, work in WP6 has just started due to delayed recruitment. The first task is to investigate existing MT techniques and their application scenarios to under-resourced languages. The work on Proofing Tools has been postponed due to the lack of qualified applicants at Tilde. A change in this position is being considered.

In WP7 (Parsing Technologies and Grammar Models; Lead Participant: CUNI), four institutions (CUNI, UHEL, UTU and UIB) have been cooperating. UHEL works on Finite-State Parsing. At UIB, a morphological component for Wolof has been implemented while a deep grammar for Wolof is being developed using the LFG approach. Experiments at CUNI have showed that structured noun phrases have little effect on a dependency parser but had a statistically significant improvement on MT. Work at UTU is on hold due to maternity leave.

A total of 23 researchers have been employed, thereof 17 Early Stage Researchers, two Experienced Researchers and three Visiting Researchers. All ESR and ER receive training and personal supervision. All recruited ESRs are enrolled in a PhD programme. All ESR and ER positions were advertised on Euraxess and on the CLARA website. Mailing lists were also used, and posters and leaflets were distributed.

Seven out of ten planned events in the Joint Training Programme (WP8) have been held (see overview below). These events have addressed training needs in all Work Packages. In addition, CLARA contributed to the CHAT 2011 Workshop. The training events have played an important role in WP integration by providing a meeting point for the different project partners, the fellows and researchers from outside the project.

Management was intensive especially in the first year. A project website was established for internal and external communication. A management meeting was held to ensure a common understanding of project procedures. Two consortium meetings were held to plan training activities and resolve any

other matters.

CONTRACT DELIVERABLES UPDATE

M - Months

RESR - Researcher

FAC B - Fixed amount contract B (%)

| RECRUITMENT | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------------|-------------------------------|------|-------|-------------|------|-------|------------|------|-------|------------------------------|------|-------|-------------|------|-------|------------|------|-------|------------------------------|------|-------|-------------|------|-------|------------|------|-------|------------------------------|------|-------|-------------|------|-------|------------|------|-------|
| Participants | Early Stage Researchers (ESR) | | | | | | | | | Experienced Researchers (ER) | | | | | | | | | Visiting scientists (VS <10) | | | | | | | | | Visiting scientists (VS >10) | | | | | | | | |
| | Foreseen | | | Implemented | | | Difference | | | Foreseen | | | Implemented | | | Difference | | | Foreseen | | | Implemented | | | Difference | | | Foreseen | | | Implemented | | | Difference | | |
| | M | RESR | FAC B | M | RESR | FAC B | M | RESR | FAC B | M | RESR | FAC B | M | RESR | FAC B | M | RESR | FAC B | M | RESR | FAC B | M | RESR | FAC B | M | RESR | FAC B | M | RESR | FAC B | M | RESR | FAC B | M | RESR | FAC B |
| UiB | 72 | 2 | 0 | 12 | 2 | | 60 | 0 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 |
| UTU | 72 | 3 | 0 | 25 | 3 | | 47 | 0 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 | 3 | 1 | 100 | | | | 3 | 1 | 100 | 0 | 0 | 0 | | | | 0 | 0 | 0 |
| Tilde | 72 | 3 | 0 | 13.5 | 2 | | 58.5 | 1 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 |
| UCPH | 72 | 2 | 0 | 22 | 2 | | 50 | 0 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | | -2 | -2 | 0 | 2 | 1 | 100 | 0 | 0 | | 2 | 1 | 100 |
| UPF | 36 | 1 | 0 | 10.6 | 1 | | 25.5 | 0 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 |
| UHEL | 24 | 1 | 0 | 5 | 1 | | 19 | 0 | 0 | 24 | 1 | 0 | 21 | 1 | | 3 | 0 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 |
| CUNI | 72 | 2 | 0 | 23.8 | 2 | | 48.25 | 0 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 |
| NHH | 30 | 1 | 0 | 16 | 1 | | 14 | 0 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 | 3 | 1 | 100 | | | | 3 | 1 | 100 | 0 | 0 | 0 | 1 | 1 | | -1 | -1 | 0 |
| MPI | 72 | 2 | 0 | 25 | 3 | | 47 | -1 | 0 | 24 | 1 | 0 | 13.6 | 1 | | 10.5 | 0 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 | 0 | 0 | 0 | | | | 0 | 0 | 0 |
| Total | 522 | 17 | 0 | 152.9 | 17 | 0 | 369 | 0 | 0 | 48 | 2 | 0 | 34.6 | 2 | 0 | 13.5 | 0 | 0 | 6 | 2 | 200 | 2 | 2 | 0 | 4 | 0 | 200 | 2 | 1 | 100 | 1 | 1 | 0 | 1 | 0 | 100 |

Comments:

Note: Because the mid-term report had to be handed in 1 month before the mid-term review meeting, it covers the first 21 months of the CLARA project, from the start date 1 Dec 2009 until 1 Sep 2011.

One ESR at MPI has quit prematurely on June 30, 2011 and a new ESR at MPI has been recruited from Aug. 1, 2011. Hence the number of ESRs is higher than planned for ESRs.

One ESR position at Tilde has not yet been filled, due to lack of qualified candidates.

Some recruitments have taken place later than planned, partly due to the administrative and immigration obstacles, partly due to the difficulty of finding qualified candidates.

UCPH has recruited two VS for one month each instead of one VS for two months, due to the need to cover competence needs in different fields.

UCPH was obliged by national rules to employ their VS on a full employment contract rather than a fixed amount contract. This has led to higher costs than the budget allows.

Some VS positions were filled with candidates with a different number of years of experience than anticipated.

INTERNATIONAL CONFERENCES / EVENTS OPEN TO EXTERNAL RESEARCHERS

| Event Number | Participant hosting the event | Type of Event | Month when the event took place | Start date of the event | End date of the event | Total number of researchers outside the network attending the event | Total number of researcher days for researchers from outside the network attending the event | Website of the event |
|--------------|-------------------------------|---------------|---------------------------------|-------------------------|-----------------------|---|--|---|
| 1 | UiB | WORDNETCOURSE | 19 | 20/06/2011 | 23/06/2011 | 3 | | http://clara.uib.no/training-program/course-lexical-resources/ |
| 10 | MPI | INFRASCHOOL | 20 | 05/07/2011 | 12/07/2011 | 5 | | https://clara.uib.no/training-program/clara-summer-school-on-infrastructure-tool-development/ |
| 2 | UiB | CAREERSCHOOL | 22 | 19/09/2011 | 23/09/2011 | | | http://clara.uib.no/training-program/industrial-career-course/ |
| 3 | UTU | TREECOURSE | 13 | 13/12/2010 | 17/12/2010 | | | http://ufal.mff.cuni.cz/clara/treecourse/ |
| 4 | UCPH | EVALCOURSE | 36 | | | | | |
| 5 | UCPH | ANNOTSCHOOL | 21 | 15/08/2011 | 26/08/2011 | | | http://clara.uib.no/training-program/semantic-and-nonverbal-annotation/ |
| 6 | UHEL | MORPHCOURSE | 17 | 11/04/2011 | 15/04/2011 | | | http://www.ling.helsinki.fi/events/CLARA-FSM/ |
| 7 | CUNI | NEWDEVSCHOOL | 27 | | | | | |
| 8 | NHH | TERMCOURSE | 10 | 13/09/2010 | 17/09/2010 | | | https://clara.uib.no/training-program/terminology-course/ |
| 9 | MPI | LRSCHOOL | 8 | 05/07/2010 | 16/07/2010 | 10 | | https://clara.uib.no/training-program/clara-summer-school-in-advanced-resource-creation-architecture/ |

| Total number of researchers outside the network attending the event | Total number of researcher days for researchers from outside the network attending the event |
|---|--|
| 27 | 225 |

Planned number of researcher days for researchers from outside the network attending the event: 208

Remaining number of researcher days for researchers from outside the network attending the event: -17

Comments:

At the time of writing this report, the number of participant days for researchers from outside the network at ANNOTSCHOOL was not yet available.

| | |
|--------------------|--|
| Attachments | |
| Name | |
| Date | |
| Signature | |