Project No: 238405

Project Acronym: CLARA

Project Full Name: Common Language Resources and their Applications

Periodic Report

Period covered: from 01/12/2009 **to** 30/11/2011   **Date of preparation:** 30/11/2011

Period number: 1st   **Date of submission (SESAM):** 27/01/2012

Project coordinator name:
Prof. Koenraad De Smedt

Project coordinator organisation name: **Version:** 1
UNIVERSITETET I BERGEN

PROJECT PERIODIC REPORT

Grant Agreement number:	238405
Project acronym:	CLARA
Project title:	Common Language Resources and their Applications
Funding Scheme:	FP7-MC-ITN
Periodic report:	1
Period covered - start date:	01/12/2009
Period covered - end date:	30/11/2011

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/

DECLARATION BY THE PROJECT COORDINATOR

I, Prof. Koenraad De Smedt, as co-ordinator of the project (238405, CLARA), hereby confirm that:

- The attached periodic report represents an accurate description of the work carried out in this project for this reporting period;

- The project has achieved most of its objectives and technical goals for the period with relatively minor deviations;

- The project Website is up to date.

- To my best knowledge, the financial statements which are being submitted as part of this report are in line with the actual work carried out and are consistent with the report on the resources used for the project and if applicable with the certificate on financial statement.

- All beneficiaries, in particular non-profit public bodies, secondary and higher education establishments, research organisations and SMEs, have declared to have verified their legal status. Any changes have been reported under section 5 (Project Management) in accordance with Article II.3.f of the Grant Agreement.

PUBLISHABLE SUMMARY

Comments:

CLARA is training a new generation of researchers who will be able to cooperate across national boundaries on the establishment of a common language resources infrastructure and its exploitation for the construction of the next generation of language models with wide theoretical and applied significance.

The scientific objectives of the CLARA research context are twofold: (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.

CLARA supplements basic research competencies in the language and text sciences with specialized knowledge and skills in linguistics, computer science, knowledge engineering, databases, statistical processing and language and speech applications. Participation in advanced research at leading universities is complemented with additional training at industrial partners contributing to careers in industry.

The project has recruited 17 ESR, 2 ER and 3 VS. They work in the following main areas: Designing and Testing Common Infrastructures; Lexical Semantic Modeling; Next generation Domain Modeling; Multimedia and Multimodal Communication Modeling; Applications; and Parsing Technologies and Grammar Models.

A Joint Training Programme of thematic courses covers the various scientific areas of the project. The following events have been held:

- CLARA Summer School in Advanced Resource Creation, Archiving and Usage (Nijmegen, 5-16 July 2010, 10 days, 10 credits)

- CLARA Thematic Training Course on Methods and Technologies for Consolidating and Harmonising Terminological Resources (Bergen, 13-17 Sep. 2010, 5 days, 5 credits)

- CLARA Thematic Training Course on Methods and Technologies for Consolidating and Harmonising Treebank Annotation (Prague, 13-17 Dec. 2010, 5 days, 5 credits)

- CLARA Thematic Training Course on Processing Morphologically Rich Languages (Budapest, 11-15 April 2011, 5 days, 5 credits)

- CLARA Thematic Training Course on Multilingual Lexical Resources and Tools (Bergen, 20-23 June 2011, 4 days, 4 credits)

- CLARA Summer School on Infrastructure Tool Development (Nijmegen, 5-12 July 2011, 10 days, 10 credits)

- CLARA Summer School in Semantic and Nonverbal Corpus Annotation and Evaluation (Copenhagen, 15-26 Aug. 2011, 10 days, 10 credits)

- CLARA Industrial Career Training Course on Product Planning for Next Generation Information Access Technology Solutions (Dubrovnik, 20-23 Sep. 2011, 4 days, 2 credits)

CLARA has also participated at the CHAT Workshop (Riga, 11 May 2011). The following events are planned for 2012:

- CLARA Winter School on New Developments in Computational Linguistics
  (Prague, 13-17 Feb. 2012, 5 days, 5 credits)

- CLARA Thematic Training Course on Evaluation of Human Language
  Technologies (Paris, Nov/Dec 2012, 4 days, 4 credits)

In all Work Packages, significant progress is visible. 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 and Spanish and to some extent Danish. The study involves cooperation with UPF.

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. Rita Temmerman was recruited as a visiting scientist.

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. UCHP data have been coded with IMDI metadata as required by the MPI repository. Two visiting scientists (Jens Edlund and Susan Brown) were recruited to UCPH.

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 reoriented due to the lack of qualified applicants at Tilde.

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 machine translation. Work at UTU is on hold due to maternity leave.

The expected outcome of the project be a pool of highly qualified

researchers who can participate in the development of the next generation of data-intensive language models and applications, and in the establishment of a pan-European infrastructure for language resources. The latter is on the ESFRI roadmap whereas the former will be crucial in tomorrow's solutions for smart information access.

For additional details, news, course reports, research results, project logo, etc. see the project website, http://clara.uib.no

Contact email: clara238405@uib.no

PROJECT OBJECTIVES FOR THE PERIOD

Comments:

The main objectives which were planned in the first period of the
CLARA project consisted of the recruitment of all researchers and the
organization of nine training events. No subprojects are planned to
be complete before the end of the project.

All fellows were planned to be recruited in the first period. One
recruitment could not be filled due to lack of qualified applicants.

The following eight Joint Training Programme events were planned and
successfully held in the first period:

- CLARA Summer School in Advanced Resource Creation, Archiving and
  Usage (Nijmegen, 5-16 July 2010, 10 days, 10 credits)

- CLARA Thematic Training Course on Methods and Technologies for
  Consolidating and Harmonising Terminological Resources (Bergen,
  13-17 Sep. 2010, 5 days, 5 credits)

- CLARA Thematic Training Course on Methods and Technologies for
  Consolidating and Harmonising Treebank Annotation (Prague, 13-17
  Dec. 2010, 5 days, 5 credits)

- CLARA Thematic Training Course on Processing Morphologically Rich
  Languages (Budapest, 11-15 April 2011, 5 days, 5 credits)

- CLARA Thematic Training Course on Multilingual Lexical Resources and
  Tools (Bergen, 20-23 June 2011, 4 days, 4 credits)

- CLARA Summer School on Infrastructure Tool Development (Nijmegen,
  5-12 July 2011, 10 days, 10 credits)

- CLARA Summer School in Semantic and Nonverbal Corpus Annotation and
  Evaluation (Copenhagen, 15-26 Aug. 2011, 10 days, 10 credits)

- CLARA Industrial Career Training Course on Product Planning for Next
  Generation Information Access Technology Solutions (Dubrovnik, 20-23
  Sep. 2011, 4 days, 2 credits)

The following event was also planned for the first period, but have
been postponed in order to maximize training impact:

- CLARA Thematic Training Course on Evaluation of Human Language
  Technologies (Paris, Nov/Dec 2012, 4 days, 4 credits)

A website was established at http://clara.uib.no and continually
maintained with announcements, reports, results, practical
information, etc.

A midterm review meeting was planned and held on September 19,
2011. All beneficiaries and some associated partners were represented,
and all active fellows were present.

The feedback from the REA representive was very positive; the main
findings and adjustments are summarized as follows.

- The project is good and promotes the state of the art in the field through new tools and applications.

- The project holds a good gender balance and there is participation by the commercial sector.

- The training programme is well organized. The consortium must work in order to let the fellows interact among themselves to promote common interests. Reaction: The consortium has organized the Sep. 2011 course so as to promote much interaction and is providing time for a Fellows meeting during the Feb. 2012 course. Furthermore, the fellows have ample opportunity to use social networks.

- Secondments should be implemented if possible for all fellows and potential secondment hosts can be brought into the consortium. Reaction: The consortium has added one associated partner with the view of allowing more secondments. Fellows are being stimulated to carry out secondments.

- Delays may cause unused researcher months. Reaction: Possible redistribution of researcher months will be discussed at the Consortium meeting in Feb. 2012. Partner institutions will bear the costs for extensions allowing fellows to finish their work.

- The project website is good but should be kept updated with a view of wider dissemination, in particular for fellows' activities and for course reports. Reaction: Course reports have been updated. Fellows are requested to keep their webpages up to date.

- There is a potential for new outreach activities. Reaction: Such outreach will probably take the form of a final conference, depending on available funds. The mid-term review report will be further followed up at the consortium meeting.

WORK PROGRESS AND ACHIEVEMENTS DURING THE PERIOD

Comments:

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

In WP2 (Designing and Testing Common Infrastructures; Lead Participant: MPI), MPI has extended and developed the Language Archive and its software component. In the framework of the DOBES program new digital media from researchers worldwide has been added to the MPI archive. A sophisticated Federated Search engine is being developed, crossing borders and giving access to language data and metadata hosted at several highly ranked scientific institutes. It is also connected with the Virtual Language Observatory which combines information about languages (including texts and audio/video recordings) from all over the world, in cooperation with CLARIN.

In WP3 (Lexical Semantic Modeling; Lead Participant: UIB), a bilingual German/Spanish corpus (about 10 M words) is being compiled at UIB. The corpus comprises all documents written by Germany, Austria or Spain and their corresponding translations into German/Spanish from the DG Enterprise and Industry Project of the European Commission, across 13 different domains. The files is being preprocessed in preparation of sentential alignment with SDL Trados and tagging, before starting subsentential alignment of compound words and phraseological units. At UCPH, corpora are being annotated with respect to regular polysemy in the languages English and Spanish and to some extent Danish. The study includes the cross-lingual aspects of semantic corpus annotation mentioned in the work description and involves cooperation with UPF. The latter has defined possible methods to overcome the problem of sparse data for lexical information acquisition and has started developments of strategic techniques to automatically merge lexical information coming from different sources based on automatic mapping of information and merging with graph unification.

In WP4 (Next Generation Domain Modeling; Lead Participant: NHH), the ESRs have been given training in relevant methods and technologies associated with the harmonisation and consolidation of multilingual terminological resources. Attention is paid to industrial terminology management (IBM), the representation of interrelated domains in termbases, concept diagrams and ontology building in terminology management systems (i-Term), systems for single access to multiple terminologies (Rikstermbanken, TermoStat), and the integraton of wordnets and information extraction based on text mining (associate partner TEMIS). At NHH, a corpus of Free Trade Agreement texts is being explored with respect to collocations. At Tilde, work is progressing on automated methods for context extraction and enrichment of a multilingual terminology database (MTDB) with knowledge-rich contexts (KRC). Secondments with TEMIS as well as targeted shorter stays at NHH and TILDE are planned for the next reporting period.

In WP5 (Multimedia and Multimodal Communication Modeling; Lead Participant: UCPH), MPI has extended the ELAN annotation tool with an interface to external components, working with their results and using them to create annotations. This is a result of using an open standard for the said extension interface and promoting it at

scientific meetings, conferences and workshops. Those recognizers are meant to be specialized in delivering specific basic functionality, which can be used as building blocks to solve complex annotation and analysis tasks. Furthermore, MPI is implementing pattern recognition and search components to be integrated in a multimodal annotation framework, and UCPH is collecting, annotating and exploiting multimodal corpora. Some of the UCHP data have already been coded with IMDI metadata as required by the MPI repository as a first step on using UCHP data as testbed for some of the MPI search components. Two visiting researchers (Jens Edlund and Susan Brown) were recruited to UCPH for contributing at a CLARA summer school related to training for this WP.

In WP6 (Applications; Lead Applicant: UTU), the focus has been on Computer Assisted Language Teaching (UTU) and Machine Translation (Tilde and CUNI). The scientific work on Machine Translation at CUNI has been focused on developing a dependency treebank for Tamil (TamilTB) – as a sample under-resourced highly inflected language – using Prague Dependency Treebank style annotation. The base language technologies and resources (lemmatizer, tagger, parser) have been developed and used within this corpus. The current version of TamilTB was released and can be accessed for free. TamiLB will be used for treebased MT experiments in next months. 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 not been filled due to the impossibility of recruiting qualified applicants at Tilde. The profile of this position has been changed to Machine Translation so as to strengthen the work in that area and to increase the chance of recruiting a researcher.

In WP7 (Parsing Technologies and Grammar Models; Lead Participant: CUNI), four institutions (CUNI, UHEL, UTU and UIB) have been cooperating. Work on Finite-State Parsing at UHEL has booked theoretical as well as practical results in the form of computational tools. At UIB, a morphological component for Wolof has been implemented while a deep grammar for Wolof is being developed using the LFG approach, thus taking this technology to a new language family. Experiments at CUNI have compared a dependency parser trained on structured noun phrases and a dependency parser trained on flat noun phrases were conducted; the results showed that the structured noun phrases had little effect on the dependency parser but had a statistically significant improvement on machine translation. Work at UTU is on hold due to maternity leave.

A total of 22 researchers have been employed, among which 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. The final ESR position has however not been filled due to lack of qualified candidates, and later a withdrawing candidate, and is reannounced. As as result of recruitment problems, as well as some sick leave and maternity leave, the use of resources has on the whole incurred a small delay. Researcher months that a partner cannot fill will if necessary be reallocated to other partners.

Eight out of ten planned events in the Joint Training Programme (WP8)

have already been held, with two more planned for 2012. These events have addressed training needs in all Work Packages. In addition, CLARA participated actively at 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. Some events have attracted more external researchers than anticipated.

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. A mid-term review meeting was held in September 2011. A website (http://clara.uib.no) was established and continually maintained with results, reports, announcements, practical information and personal webpages for the recruited researchers.

The project has fully realized its planned joint training programme and has booked good scientific results through publications and the production of new research data and tools by its researchers. Results are disseminated at the project website, through academic publication channels, through participation at scientific meetings and through contact with other projects (e.g. CLARIN and TTC). Further dissemination by means of a final conference is presently being discussed.

DELIVERABLES AND MILESTONES TABLE

RECRUITMENT

Recruiting participant	Fellow name	Type of Contract	category	Location of origin	Gender	Family situation	Start date of secondment/ recruitment	End date of secondment/ recruitment	Working time commitment	Full-time equivalent person-months covered during the reporting period	Declaration of Conformity submitted
MAX PLANCK GESELLSCHAFT ZUR FOERDERUNG DER WISSENSCHAFTEN E.V.	Hugo Garcia Blanco	A	ESR	Spain	Male	No	01/07/2010	30/06/2011	Full Time	12	Yes
MAX PLANCK GESELLSCHAFT ZUR FOERDERUNG DER WISSENSCHAFTEN E.V.	Przemyslaw Lenkiewicz	A	ER	Portugal	Male	Yes	15/07/2010	14/07/2012	Full Time	16.5	Yes
UNIVERZITA KARLOVA V PRAZE	Loganathan Ramasamy	A	ESR	India	Male	No	01/08/2010	31/07/2013	Full Time	16	Yes
MAX PLANCK GESELLSCHAFT ZUR FOERDERUNG DER WISSENSCHAFTEN E.V.	Binyam Gebrekidan	A	ESR	Ethiopia	Male	No	01/09/2010	31/08/2013	Full Time	15	Yes
UNIVERSITETET I BERGEN	Cheikh Bamba M. Dione	A	ESR	Germany	Male	Yes	01/01/2011	31/12/2013	Full Time	11	Yes
HELSINGIN	Senka Drobac	A	ESR	Croatia	Female	No	01/04/2011	31/03/2013	Full Time	8	Yes

YLIOPISTO											
UNIVERSITETET I BERGEN	Carla Parra Escartín	A	ESR	Spain	Female	No	02/05/2011	01/05/2014	Full Time	7	Yes
EBERHARD KARLS UNIVERSITAET TUEBINGEN	Sowmya Vajjala	A	ESR	India	Female	Yes	01/04/2011	30/03/2013	Full Time	8	Yes
EBERHARD KARLS UNIVERSITAET TUEBINGEN	Jianqiang Ma	A	ESR	China (People's Republic of)	Male	Yes	01/04/2011	31/03/2013	Full Time	8	Yes
TILDE SIA	Septina Dian Larasati	A	ESR	Czech Republic	Female	No	01/08/2011	30/04/2013	Full Time	4	Yes
Københavns Universitet	Susan Brown	A	VS <10	United States	Female	Yes	01/08/2011	31/08/2011	Full Time	1	Yes
Københavns Universitet	Jens Edlund	A	VS <10	Sweden	Male	No	01/08/2011	31/08/2011	Full Time	1	Yes
MAX PLANCK GESELLSCHAFT ZUR FOERDERUNG DER WISSENSCHAFTEN E.V.	Anna Lenkiewicz	A	ESR	Portugal	Female	Yes	01/08/2011	31/07/2013	Full Time	4	Yes
NORGES HANDELSHOYSKOLE	Pedro Patiño García	A	ESR	Colombia	Male	No	01/05/2010	31/10/2012	Full Time	19	Yes
HELSINGIN YLIOPISTO	Måns Huldén	A	ER	Finland	Male	Yes	01/12/2009	30/11/2011	Full Time	24	Yes
TILDE SIA	Anne Schumann	A	ESR	Germany	Female	No	16/08/2010	15/02/2013	Full Time	15.5	Yes
EBERHARD KARLS UNIVERSITAET TUEBINGEN	Corina Dima	A	ESR	Romania	Female	Yes	01/06/2010	31/05/2012	Other	4.5	Yes
NORGES HANDELSHOYSKOLE	Rita Temmerman	B	VS >10	Belgium	Female	No	01/09/2010	30/11/2013	Other	1	Yes
UNIVERZITA KARLOVA V PRAZE	Nathan David Green	A	ESR	United States	Male	No	08/10/2010	07/10/2013	Full Time	13.75	Yes
Københavns	Hector Martinez	A	ESR	Spain	Male	No	01/10/2010	30/09/2013	Full Time	14	Yes

Universitet	Alonso										
Københavns Universitet	Magdalena Lis	A	ESR	Poland	Female	No	01/10/2010	30/09/2013	Full Time	14	Yes
UNIVERSITAT POMPEU FABRA	SILVIA NECSULESCU	A	ESR	Romania	Female	No	15/10/2010	14/10/2013	Full Time	13.5	Yes

No. of full-time equivalent months covered during this reporting period: 230.75

M - Months

RESR - Researcher

FAC B - Fixed amount contract B (%)

	RECRUITMENT																																			
	Early Stage Researchers (ESR)									Experienced Researchers (ER)									Visiting scientists (VS <10)									Visiting scientists (VS >10)									
Participants	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		18	2		54	0			0						0			0						0			0						0	
UTU	72	3		20.5	3		51.5	0			0						0		3	1	100				3	1	100		0						0	
Tilde	72	3		19.5	2		52.5	1			0						0			0						0			0						0	
UCPH	72	2		28	2		44	0			0						0			0		2	2		-2	-2		2	1	100				2	1	100
UPF	36	1		13.5	1		22.5	0			0						0			0						0			0						0	
UHEL	24	1		8	1		16	0		24	1		24	1			0			0						0			0						0	
CUNI	72	2		29.75	2		42.25	0			0						0			0						0			0						0	
NHH	30	1		19	1		11	0			0						0		3	1	100				3	1	100		0		1	1		-1	-1	
MPI	72	2		31	3		41	-1		24	1		16.5	1		7.5	0			0						0			0						0	
Total	522	17		187.25	17		334.75	0		48	2		40.5	2		7.5	0		6	2.00	200	2	2		4	0	200	2	1.00	100	1	1		1	0	100

Comments:

One ESR at MPI quit prematurely on June 30, 2011 and a new ESR was recruited as from August 1, 2011. One ESR position at Tilde has not yet been filled, first due to lack of qualified candidates, and then because the selected candidate turned down the offer. 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. One candidate at UTU is having parental leave, this explains why the employment period is longer than the researcher months.

One ER has finalised his employment period by the time of this report.

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. UTU has had trouble recruiting VSes under the correct type of contract. Some VS positions were filled with candidates with a different number of years of experience than anticipated in Annex I.

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	UNIVERSITETET I BERGEN	WORDNETCOURSE	19	20/06/2011	23/06/2011	3	10	clara.uib.no/training-program/course-lexical-resources/
10	MAX PLANCK GESELLSCHAFT ZUR FOERDERUNG DER WISSENSCHAFTEN E.V.	INFRASCHOOL	20	05/07/2011	12/07/2011	5	25	clara.uib.no/training-program/clara-summer-school-on-infrastructure-tool-development/
2	UNIVERSITETET I BERGEN	CAREERSCHOOL	22	19/09/2011	23/09/2011	1	2	clara.uib.no/training-program/industrial-career-course/
3	EBERHARD KARLS UNIVERSITAET TUEBINGEN	TREECOURSE	13	13/12/2010	17/12/2010	43	66	clara.uib.no/training-program/treebank-course/ AND ufal.mff.cuni.cz/clara/treeco urse
4	Københavns Universitet	EVALCOURSE	36					
5	Københavns Universitet	ANNOTSCHOOL	21	15/08/2011	26/08/2011	20	86	clara.uib.no/training-program/semantic-and-nonverbal-annotat ion/
6	HELSINGIN YLIOPISTO	MORPHCOURSE	17	11/04/2011	15/04/2011	14	59	clara.uib.no/training-program/processing-morphologically-rich-languages/ AND www.ling.helsinki.fi/events/CLARA-FSM/
7	UNIVERZITA KARLOVA V PRAZE	NEWDEVSCHOOL	27	13/02/2012	17/02/2012			clara.uib.no/training-program/new-developments-in-cl/
8	NORGES HANDELSHOYSKOLE	TERMCOURSE	10	13/09/2010	17/09/2010	3	13	clara.uib.no/training-program/terminology-course/
9	MAX PLANCK	LRSCHOOL	8	05/07/2010	16/07/2010	19	190	clara.uib.no/trainin

	GESELLSCHAFT ZUR FOERDERUNG DER WISSENSCHAFTEN E.V.							g-program/clara-summer-school-in-advanced-resource-creation-archiving-and-usage/
11	TILDE SIA	CHAT WORKSHOP	18	11/05/2011	11/05/2011	23	23	clara.uib.no/training-program/chat-workshop/

Total number of researchers outside the network attending the event	Total number of researcher days for researchers from outside the network attending the event
131	474

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: -266

Comments:

The participation of external researchers at events exceeded our expectations.

4.3. MILESTONES

Milestone no.	Milestone name	Due achievement date from Annex I	Achieved	Actual / Forecast achievement date
1	TERMCOURSE	2010-06-30	Yes	2010-09-17
2	WORDNETCOURSE	2010-06-30	Yes	2011-06-23
3	LRSCHOOL	2010-07-31	Yes	2010-07-16
4	MORPHCOURSE	2010-09-30	Yes	2011-04-15
5	EVALCOURSE	2010-06-30	No	2012-12-31
7	TREECOURSE	2010-12-31	Yes	2010-12-17
8	INFRASCHOOL	2011-07-31	Yes	2011-07-12
9	ANNOTSCHOOL	2011-08-31	Yes	2011-08-26
10	CAREERSCHOOL	2011-09-30	Yes	2011-09-23
11	NEWDEVSCHOOL	2011-12-31	No	2012-02-13

Comments:

Two courses are postponed to improve their targeted audiences.

ADDITIONAL INFORMATION

Comments:

After the first period, an associate partner was added to the project.

DISSEMINATION ACTIVITIES

Comments:

A website http://clara.uib.no was established with comprehensive
information about the project including the following:

- news with announcements of vacancies, courses, etc.;

- facts about the project, including funding, project period,
  coordination, background, research themes, consortium etc.;

- vacancies with full announcements if still open;

- training program with an overview of courses; program and practical
  information of courses to be held and report of courses already
  held;

- overview of fellows and individual webpages maintained by each
  fellow with personal participation and dissemination activities;

- calendar of events;

- results in terms of scientific publications;

- management information including consortium meetings, references to
  official documents, practical guidelines, contact information, etc.

Scientific results were presented at scientific meetings and published
in scientific publication channels. See the website
http://clara.uib.no/results/ for an overview of scientific
publications, and see also the fellows' individual webpages at the
website for additional activities.

PROJECT MANAGEMENT

Comments:

In the first year. A project website (http://clara.uib.no) was established which not only served for external dissemination (see point 5) but also served internal communication by providing contact information, overviews of procedures, agendas and minutes of meetings, pointers to official documents, etc.

A management meeting was held (15 Feb. 2010) to ensure a common understanding of project procedures and financial guidelines, which was not trivial. Two consortium meetings (22 Mar. and 16 Sep. 2010) were held in order to plan training activities and resolve any other matters. A mid-term review meeting was held in on 19 Sep. 2011. A record of meetings is kept on the website.

The project has not encountered serious deviations. Nevertheless, recruitment of well qualified candidates has proved more difficult than expected. This challenge, in addition to some prolonged sick leave and maternity leave, have caused delays in the project which are not crucial but demanded some corrective changes in the time plan and our current consideration of a reallocation of a few researcher months which would otherwise remain unused.

The coordinator faced challenges in interpreting rules concerning allowances in many special circumstances, answered many questions from the project participants and had to consult the REA representative in several cases.

Conflicts between Marie Curie rules and local or national employment rules has been causing substantial challenges. A still unresolved issue is that of a beneficiary unable to recruit a visiting scientist on a fixed allowance scheme, due to the incompatibility between local social security limitations on non-full employment contracts and the minimal social security requirements imposed by the EU.

More detailed financial planning than expected has been necessary in order to achieve a balance in each partner's contribution to common training activities. This proved to be difficult and will be discussed at the next Consortium meeting.

At the time of the mid-term review meeting it was considered to add an associate partner; this was carried out after the first period.

Gender issues were not encountered since the project naturally obtained a very good gender balance.

Reporting has been more time-consuming than necessary because the reporting procedures are not really user-friendly and mostly involve tedious reformatting of information which is already available elsewhere.

In general, management has been more resource consuming than expected.

FINANCIAL STATEMENTS – FORM C AND SUMMARY FINANCIAL REPORT

Comments:

All Forms C and summary report submitted to EU on 20 Jan 2012.

CERTIFICATES

List of Certificates which are due for this period, in accordance with Article II.4.4 of the Grant Agreement.

Beneficiary	Organisation short name	Certificate on the financial statements provided?	Any useful comment, in particular if a certificate is not provided
1	UiB	No	
2	UTU	No	
3	Tilde	No	
4	UCPH	No	
5	UPF	No	
6	UHEL	No	
7	CUNI	No	
9	NHH	No	
10	MPI	No	

Attachments	FP7-Form Cs Summary-CLARA-submitted-20jan2012.pdf
Name	
Date	
Signature	