AWARD CRITERIA FOR ERASMUS MUNDUS JOINT DOCTORATE PROGRAMMES

DETAILED DESCRIPTION OF THE IMPLEMENTATION OF THE ERASMUS MUNDUS JOINT DOCTORATE FOR “RESEARCH AND ENGINEERING IN LANGUAGE TECHNOLOGY HARMONIZING THE DIVERSITY, DYNAMICS AND DEPTH OF HUMAN LANGUAGES (EMJD RELATE)”

B.1 Academic and research quality

B.1.1 Objectives and Needs Analysis

The Challenge  Language technology is a central enabling technology for the next generation of IT. By further improving algorithms for dealing with human language, IT will be able to overcome existing barriers to understanding. People will be able to effectively use the accumulated digitized knowledge of humankind in any situation, in which knowledge is needed, by just asking their questions. Through powerful services fed by personalized analytics of news and social media they will better understand developments around them and become aware of emerging new processes. Decision makers will understand their customers, patients, clients, citizens or employees much better by using intelligent information services that aggregate and condense their opinions and sentiments. Through reliable automatic translation, we will be able to communicate with people who do not master our mother tongue and we will comprehend texts and films produced in any language.

The ubiquitous information technology itself will understand our questions, needs, desires and emotions, at least to a degree required for delivering greatly enhanced services and for providing proactive assistance.

The spectacular breakthroughs of language technology during the last two years give us just a first modest glimpse of the emerging new type of IT, which understands its users and helps them to understand others and the world around them. Among the applications that already started to change our life are web-based content translation services such as Google Translate or Microsofts Bing Translator, the speech-controlled mobile assistant of Apple Siri, the question answering system behind the electronic TV quiz champion IBM Watson, Nuances speech technology embedded in many call center dialogue systems and Autonmys text analytics acquired by HP for 8bEuro. Many other applications are less visible but also part of the beginning success story. Googles search is gradually becoming more semantic, leading the worlds largest language technology corporation to transform its search technology sector into a new division called Knowledge. Major high technology corporations, European Commission and Parliament, patent offices, intelligence services and numerous translation companies have employed specialized machine translation systems in their daily operations.

Although all of these applications are still far from being perfect, they are becoming useful, they demonstrate the potential of language technology and they prepare the markets for the much bigger things to come. They also create a strong need for experts who can develop, adapt, improve, evaluate, deploy and maintain these technologies and the applications in which they are embedded.

As progress in language technology is speeding up, these experts need to be prepared for coming generations of language technology that will go far beyond the cited examples in their functionality, internal sophistication and utilized methodology. The leading experts will undoubtedly come from academic environments in which they had themselves been able to contribute to the next technology generation during their training and early research. The most exciting and rewarding research in language technology will be dedicated to truly linking language with its meaning, in comprehension, production and translation.

Whereas current applications are already helping us to understand each other and to understand the ever-increasing complexity in our work and our world, they themselves still lack true understanding. Therefore the most ambitious
language technology research aims to make software get at the meaning of our questions, commands and texts. The goal is not so much to simulate all of human intelligence complete with intuition, emotion and creativity, since we could not reproduce anyway what we do not yet understand, but rather to investigate and model those aspects of language and language use that are needed for the targeted applications. This research is strongly inspired but certainly not constrained by our insights into the mechanisms behind human language use. Our ambition is not so much to reach human capabilities in the foreseeable future but rather to exceed human capabilities in selected areas, usually those areas where limitations of human cognition can be overcome by technology. Humans can for instance reason about an important sentence for months and then express their thoughts in an essay, but machines can already do a superficial analysis of millions of pages in a few seconds. Human translators can find a clever and highly intuitive non-literal translation of a metaphor, but automatic translation can use thousands of existing expert translations for selecting appropriate and consistent phrases in the instantaneous translation of a long manual. Machine translation can also translate from and into more languages than any human can ever master.

Now the advantages of networked computing technology need to be extended by even better algorithms for analyzing, producing and translating human language in spoken and written form. The ubiquitous availability of computer programs that better understand human language and human interests, supported by the speed and memory of zillions of untiring microprocessors and supplied with permanent access to the digitized information and knowledge of the world, will considerably extend individual human cognition but also enable a higher level of collective intelligence.

**Application Areas** The range of language technology applications is so wide that no research centre and not even a joint PhD programme could cover them all. The proposed EMJD will focus on a class of applications that improve the understanding among people and their understanding of the world and that overcome communication barriers between people and technology. Many other applications of language technology, e.g., in medical therapy, in forensic contexts, in language learning and in general entertainment will benefit from the planned research but not be directly targeted.

A central application area of the proposed EMJD is novel technology for social intelligence. In a narrower sense, social intelligence can be interpreted parallel to business intelligence and military intelligence. It provides individual and collective deciders with all the information needed for improving the quality of their decisions. For optimal reactions to new challenges and for more effective planning, social commercial, political, educational and service organisations need to understand the needs, priorities, ideas and preferences of their constituency. i.e., citizens, customers, clients, consumers, students, etc. To this end social media, news and other forms of public communication have to be monitored, a task that is not manageable without technologies that can detect and analyse relevant information expressed in human language. In the spirit of an egalitarian and transparent information society, these instruments should also be available to any member of the monitored communities and other target groups so that given sufficient interest and time these people could reach the same degree of understanding.

In wider sense, we mean by social intelligence higher forms of intelligent social computing. The Internet, extended by the densely woven network of information and knowledge known as the web, and by means for accessing this network from any place in the world enables powerful new ways of social interaction. Web users can closely follow processes even if these happen at distant places. They can access virtually all of human knowledge from any location and they can use the web for interacting with large groups of other people through various forms and combinations of synchronous and asynchronous communication.

Potentially the web also permits novel forms of collective work including participation in political, economic, scientific, or cultural processes such as knowledge creation, opinion formation and actual decision making. In the prophecies of technology visionaries, the predicted massive involvement of the affected and the afflicted is sometimes subsumed under the term e-participation. Since such an active involvement of the constituencies will eventually extend the traditional system of democratic decision processes beyond the proven mechanisms of elections, parties and parliaments, it is also termed e-democracy. It will facilitate dynamic issue-centred alliances in which stakeholders and non-appointed experts join forces for the development and promotion of improved solutions to social problems. This vision is also known as liquid democracy. The constructive massive participation of stakeholders requires means for keeping them informed and involved in the public dialogue. The Internet already offers fora for such debates but there are no effective instruments for preserving transparency when the participation reaches a certain scale. Language technology is needed for monitoring, summarising and analysing such large scale deliberations.

Language technology has already started work on large volumes of so-called user-generated in social media. These text genres are challenging for automatic processing because they contain variants of language that differ considerably
from the mainstream language of professional writing as we find it in regular news reports and most books. Twitter is
just an extreme case, but also many other social media such as fora of patients and consumers often exhibit language that
existing analysis software cannot handle. European integration, migration and globalization in business, culture, science
and technology give rise to an even harder problem. Should processes of collective opinion formation be segregated
by the languages of the users? Most decisions on legislation, products, medicines affect people independent of their
native language. Language technology can and will provide the means for dialogue across language barriers. It will also
supply all participants at any time and place with sufficiently comprehensive and up-to-date information and background
knowledge for an informed participation in their preferred language.

Scientific Challenges The attractive complex and scalable application scenario does not only provide the appropriate
desiderata and the types and volumes of data needed for breakthroughs but it also addresses the most exciting open
research questions in language processing. These can be grouped into three major challenges.

1. Overcoming language barriers: Requirements for translation (quality, time and price) vary depending on the
intended purpose. The one-size-fits-all approach of Google Translate is a wonderful tool for accessing information
that otherwise would not be translated at all and for getting a rough overview of found contents. It is not suited
for texts to be published or for information where accuracy is mission-critical. Nearly the entire translation market
consists of quality translation. Overcoming existing quality barriers is a major challenge for research. Most experts
in the field agree that more sophisticated approaches to machine learning and a deeper analysis of language are
needed in order to reach higher levels of quality. The breakthrough will not be achieved by a sudden discovery of
a more powerful learning method but rather by a systematic investigation of quality barriers and new combinations
of learning and deeper syntactic and semantic analysis. Eventually automatic translation needs to be available for
different task-dependent requirements of quality, time and completeness, this presupposes means for optimizing
for each of the three desiderata.

2. Wide-coverage self-adapting understanding: The main paradigm of statistical language technology is the adap-
tation to new tasks by learning from representative data. This will not be sufficient in the long for several reasons:
The combinatorics of language variants, subject domains, text sorts or forms of communications would eventually
result in the need of millions of language models. But even if had such a wealth of models, they could not ade-
quately deal with the mixes of language variants, multilingual documents, expert-lay contributions that we find
in a single forum, sometimes even in one text. in a forum. A huge number of highly specialized models might
also not be the best approach for dealing with the dynamics of language. We opt for an alternative approach to
accommodating variation: wide coverage systems for language analysis and understanding that adapt themselves
to language variants, tasks and situation types. One prerequisite for such an approach are large volumes of hetero-
genous training data that are labeled for variation features.

3. Understanding in the wider context: The contextual nature of utterances is one of the best known but least
mastered properties of human language. However, very often the most relevant context information has to be found
outside the linguistic input. Both in training and in applying language analysis tools extra-linguistic information
can be crucial for success. Todays systems are not able to utilize the information about situation, the speaker/author,
the task and the co-existence with information in other media and modes. Models of discourse, dialogues, tasks,
communication situations and domain knowledge will be investigated for their potential role in language analysis
and production and so will also be connections between linguistically encoded information and visual information.
Two recent developments will become relevant in this endeavor. The first one is the semantification of process
specifications and information repositories ranging from semantically motivated meta-data schemes via linked
open data all the way to ontologies in description logics. The second one is the automatic induction of probabilistic
models from data, for instance, the planned learning of so-called scripts, fuzzy situation models from data.

Methods and Approaches As the investigation and modelling of human language is a truly interdisciplinary endeavour,
the methods of language technology come from several disciplines: computer science, computational and theoretical
linguistics, mathematics, electrical engineering and psychology.

For a while it appeared as if the variety of approaches and techniques required for successful research in language
technology would finally converge on a smaller range of methods. Powerful statistical techniques for automatically

3
learning language models seemed to replace older linguistic and statistical approaches including highly sophisticated intellectually created lexicons and grammars. In the meantime, the convergence has reversed, not only do we witness a return of linguistic and logical methods but we also register an explosion of new machine-learning paradigms and techniques with increased sophistication both in learning algorithms and in the feature sets that determine the learning process. The combination of statistical and linguistic approaches has just begun. In addition we see a growing importance of methods from knowledge processing, action and perception research, and experimental psychology.

Scientists and developers are required nowadays to draw on a wide range of methods, theories, and insights involving linguistic formalisms, machine learning approaches, discrete and non-discrete mathematical methods, programming languages, software engineering methods as well as specialised algorithms for parsing, generation and translation. In addition, successful researchers need some fundamental knowledge about the essential properties of human language including variation across languages and variants of the same language and some hands-on experience with real language data.

It is nearly impossible today to have all the relevant approaches, methods and skills represented at a single university or research centre. Young researchers who get exposure to a wider range of methods and activities usually benefit immensely from this experience. Whenever we could send PhD students to other leading centres of research for extended stays, the results were rather impressive. In the EMJD RELATE we want to go much further than the usual occasional exchange of students and teachers. We have identified themes and systematic combinations of methods and skills as well as modes of cooperation that let larger groups of PhD candidate benefit from the complementarity in methods and experience.

Apart from the strong scientific and technical background, young professionals in the field of LTs are nowadays also required to have:

- the ability to create new products in large established Internet companies, as well as in small SMEs which explore innovative business models;
- the knowledge to push the frontiers in research, keeping in mind the big picture;
- solid intercultural and complementary skills, which should be there when a successful career development leading to management positions in the academia or industry is pursued.

All this cannot be achieved, but through intense cooperation among leading research institutions, like the partners of this consortium. This guarantees the maximisation of the use of knowledge, research facilities.

**Goals of the research, training and exchange program**

The overall aim is to design a joint doctoral program, in which the various research specialisations mentioned above come together in order to form an excellence cluster with a high-level research structure. This is obtained through enhancing already existing cooperation in education and research among the five universities of the consortium and the several associated partners. The development of RELATE is the next consequent step in the evolution of the already existing structures for international education at the partner universities. Specifically, RELATE also reflects the consequence of the success of the EMMC LCT (European Masters Program in Language and Communication Technologies), funded for the editions 2007 to 2012.

The specific goals of RELATE and the concrete actions with which it is intended to realise them are:

1. to improve the career perspectives of PhD candidates in technical competences, intercultural skills, as well as complementary skills, such as social competences, entrepreneurship, research and project management, proposal writing, etc., and to prepare them for life-long learning through trans-national and intersectoral/interdisciplinary skills;
2. to develop a cadre of broadly knowledgeable scientists;
3. to develop strong cooperations and close links between academia across Europe and world-wide;
4. to strengthen Europe's leadership and competitiveness through comprehensive development of human potential.
B.1.2 Added Value and Distinctiveness

There is definitely no EMJD funded in the research area of Language Technology (LT). Due to the interdisciplinary nature of the proposed program, it is indeed generally difficult to find other comparable international doctoral programs, especially, in the European research environment. The presence of only one Marie-Curie Initial Training Network in Speech Communication with Adaptive LEarning (SCALE), which mainly focuses on speech (recognition and synthesis, as well as signal processing) manifests the necessity of a program like the one proposed here for the whole interdisciplinary field of Language Technology. The lack of structured PhD programs in Language Technology in the European Doctorate model is even more noticeable when one compares to the USA NSF-funded PIRE PhD Program on the “Investigation of Meaning Representations in Language Understanding for Speech Reconstruction and Machine Translation Systems” (NSF Partnership for Research and Education), which, though a very distinguished effort, is also restricted as far as the topics it touches upon are concerned, mainly focusing on speech processing for Machine Translation (MT) purposes.

Which are the added values and distinctiveness of RELATE at the European scale?

RELATE

1. combines and utilises different approaches in education and research in the interdisciplinary area of Language Technology;

2. trains on complementary skills, including intercultural competences. This way, it enhances career and employability perspectives;

3. works towards the creation of a truly joint program, trying to smooth out differences on formal issues arising from the various national curricula of the partner universities. This way it enhances attractiveness and visibility of European education;

4. leads to educational, social and economic benefits: (1) multilingual European education in Language Technology is promoted worldwide; (2) the European universities participating in the RELATE further develop as centres of excellence in Europe; and (3) those partners in countries with less developed Language Technology sectors are able to progress language technology for their own languages. Taken together, this increases the competitiveness of European education in Language Technology, which raises the competitiveness of European IT industries, creates a widely accessible multilingual Information Society, and turns the information overload into a wealth of accessible and useful knowledge.

5. introduces a European Doctorate model that is unique in its consistency and coverage in an area that is vital for the multilingual European Community. The model also extends to the heart of the multilingual European society and its needs through its Associated Member (AMs), which have been selected carefully and strategically in order to guarantee the achievement of the objectives described above. The benefits of AM participation in the RELATE and its European Doctorate model lie in the direct mutual relation they build between academia and industry, which is vital for the economic growth of the European society. This not only contributes to the sustainability of the program itself, but also ensures the future of the graduates of the program at the challenging financial era we live in.

B.1.3 Originality, Innovative Aspects and Scientific Quality

In a nutshell, RELATE consists of 2 tracks, the academic and the industrial track. The academic track focuses on basic research and development of scientific methods for Language Technology (LT). The industrial track focuses on the transfer of basic research to technological innovation in the field of LT in close collaboration with the industry.

Training of PhD candidates in RELATE is based on an individual per student research plan and courses. The courses amount to 30 ECTS credits:

1. 15 ECTS credits for courses on scientific subjects of LT;
2. 10 ECTS credits for courses on complementary skills;
3. 5 ECTS credits for courses on intercultural skills.
Research in RELATE is carried out in the five LT research areas elaborated in section B.1.1: technologies for the syntactic analysis of language, technologies for the semantic analysis of language, speech technologies, multilingual and crosslingual information-access technologies, and translation technologies. The research goals of the overall program, as well as the personal career development plan (PCDP) of each PhD candidate, are achieved through:

1. the Doctoral Thesis: the PhD candidates have to write a dissertation (Doctoral Thesis) in one of the five research areas mentioned above. It has to be a personal contribution to the advance of knowledge in the field in collaboration between at least two groups in different partner universities. For the thesis to be brought to a successful end, co-supervision of the PhD candidate by professors of two universities of the consortium is compulsory;

2. presentations, in form of participation and presentation of the research findings included in the Doctoral Thesis in seminars, international conferences, fora, etc., as well as the annual 2-day workshop of RELATE are compulsory;

3. publications in peer-reviewed journals of the scientific results reported in the Doctoral Thesis are required;

4. Thesis Defence: the scientific work of each PhD candidate has to be presented and defended in a colloquium which serves for proof of the scientific qualification of the PhD candidate. The thesis defence follows a joint procedure of both universities at which the PhD candidate carries out the scientific research and complies with local regulations.

Training in RELATE is also related to the five research LT research areas elaborated in section B.1.1, comprising:

1. scientific courses: these are specialised post-graduate courses related to the five LT research areas elaborated in section B.1.1, which reflect the expertise of each partner university and contribute to the global training of the PhD candidates of RELATE in linguistics and mathematics which are necessary for working in the interdisciplinary research field of Language Technology. These courses are selected for each PhD candidate on the basis of their personal career development plan (PCDP), taking into consideration the candidate’s background, as well as the objective conditions for developing their specific research work in the course of their study in RELATE. These courses can be chosen from a range of post-graduate courses offered at all the partners of the consortium in the five LT research areas impacting RELATE.

2. courses for complementary skills: these are courses for the development of
   (a) personal and social competences;
   (b) competence for papers and thesis writing;
   (c) presentation skills;
   (d) research ethics;
   (e) competence for career planning, interview skills, negotiation skills, academic and industrial management;
   (f) competence for grant proposal writing

3. courses for intercultural skills: These courses support acquisition of international and intercultural competences. For example, how to manage international research in academic and industrial environments, how to deal with international customers, etc.

4. summer school: This is a compulsory week-long event for the PhD candidates to be organised by the consortium annually. Courses are offered by local and invited experts in the various current trends of LT. The PhD candidates are awarded 3 ECTS credits per summer school, which may partly or in their entirety be counted to the credits earned through the whole course of the PhD studies as part of the scientific courses described in the previous.

5. winter labs: Each winter, RELATE will organise and host international teams for an intensive 2-week research lab-based workshop on language and speech technology and engineering. Each lab-based workshop team, consisting of internal and external LT experts, PhD candidates and post-docs of RELATE spends 2 weeks together at one of the partner universities (after some advance preparation and in a different partner institution every year), working in close proximity on some challenge problem or promising solution technique that has not yet been well studied. The hope is that some of the teams will have a lasting influence on subsequent research and practice through the
publications, software, and data that they produce. For the PhD candidates and post-docs of RELATE, the biggest benefit is the interaction with other researchers, seeding new and lasting collaborations. Finally, the winter labs are also open to external highly talented masters students and post-docs on a fee-based participation, which gives a unique opportunity to the PhD candidates of RELATE to exchange on their research with a wider and more diverse group of experts and researchers and contributes to the sustainability of the program through wider dissemination. The PhD candidates are also awarded 3 ECTS credits per winter lab, which may partly or in their entirety be counted to the credits earned through the whole course of the PhD studies as part of the scientific courses described in the previous.

6. mobility: This comprises

(a) mobility to one of the consortium partner universities. All PhD candidates must go to at least a second university for at least 6 months, which are defined as not necessarily consecutive. Their activities during their mobility are recognised through the ECTS credit transfer system;

(b) mobility to at least one associated partner. For PhD candidates on the industrial track, mobility to an industrial AM is compulsory and should not exceed one third of the total mobility period. For PhD candidates on the academic track, mobility to an academic AM is meant to provide an additional academic networking opportunity.

The scientific quality of the program is guaranteed through

1. the excellent reputation of the involved partners;

2. the high quality of the supervisors;

3. the collaboration of at least two partners in each doctoral thesis;

4. the courses mentioned above, the summer school, the winter labs, the international and intersectoral mobility which are integral part of the program;

5. the participation of top-level academic and industrial Associated Members (AMs) which, additionally to contributing to the research of the PhD candidates, also provide councelling about the program’s overall training program and research activities, advice related to Quality Assessment, placement possibilities for the program’s PhD candidates, additional research facilities and expertise, as well as further opportunities for career development worldwide.

B.1.4 Inter-sectoral and Inter-organisational collaboration and mobility

RELATE will train PhD candidates in a way that allows them to benefit from progress in multiple fields and to acquire a broad foundation on which to build their careers and adjust to inevitable changes in their specialisations. RELATE therefore will help secure the leading position of Europe in fundamental research in multilingual language technology and put European companies in a position to reap the fruits of the advances in multilingual language and communication technology.

Specifically, RELATE will

1. offer PhD candidates the unique opportunity to carry out study and research in at least two universities of the consortium. This manifests the solid collaboration among the consortium’s partners universities, which is formally established on the legal basis of a Multilateral Agreement;

2. guarantee collaboration with the industry. A research stint with an industrial AM is compulsory for PhD candidates following the industrial track, and optional for those following the academic track;

3. provide additional academic networking opportunities through mobility periods with its world-leading academic AMs;
4. cooperate with doctoral programs at the partner universities and the academic AMs in order to maximise its added value and inter-connectivity and to increase the number of PhD candidates participating in its activities. Such programs comprise: Computational Linguistics at UdS, the study program Informatics, branch Mathematical Linguistics at CUNI, the PhD program in Informatics at the UoE, the PhD program of the Graduate School of the Humanities at RuG, and the Doctor of Philosophy (Dochtúir Fealsúnachta) at DCU.

B.1.5 Expected Innovative Outcomes

The educational outcomes of RELATE conform to those of the third cycle (doctoral level) described by the EQF\(^1\), which are fine-tuned to accommodate the personal competencies of each individual PhD candidate through the definition of a personal career development plan (PCDP). PhD graduates of RELATE will be endowed with:

1. deep scientific competences, substantial authority, autonomy, scholarly and professional integrity, and sustained commitment to innovative, beneficial development;

2. multidisciplinarity, enabling them to transfer knowledge between sub-fields of Language Technology, and across to other fields;

3. complementary skills, enabling them to adhere to the highest standards of ethical professional practice;

4. intercultural skills, enabling them to communicate with and motivate diverse audiences.

In terms of scientific outcomes and achievements in the research of the PhD candidates, RELATE aims at the following scientific and technological outcomes:

1. The primary expected innovations created by the PhD candidates are especially new Natural Language Processing (NLP) models (for all five categories) which will achieve higher quality in several dimensions: higher performance (accuracy in a broader sense), faster learning curves, easier portability to new domains, lower bias with respect to linguistic theory, but also easier incorporation of linguistic prior knowledge at the same time.

2. The second group of expected innovations are of technological nature, related especially to scalability of the implementations of the models. The PhD candidates will be exposed in their research to extremely large data, with the expectation that they will be able to design new (more efficient) techniques for their processing. At the same time, it will be absolutely necessary to process the data in parallel fashion (on computer clusters), so it may be expected that the PhD candidates will be able to design new distributed NLP algorithms.

B.2 Partnership experience and composition

B.2.1 Consortium composition and research staff

The partners in RELATE consortium are all internationally-established players in the area of Language Technology (LT). The composition of the consortium is designed to ensure crucial expertise and critical mass in all the research and training aspects identified in section B.1. Crucially for the success of the project, the partnership emerges from years of intense collaboration both in the research and the educational area, also including participation in successful Erasmus Mundus partnerships. The academic staff in all the partners of the consortium involved in the program are specialists in fields related to Language Technology, including computer science, computational linguistics, syntax, semantics, phonetics, and psycholinguistics. They are all first-rate researchers with an international reputation and experienced teachers.

Dublin City University (DCU): DCU is a young university that has developed a strong research reputation, creating national centres of excellence that collaborate with other universities and industries worldwide. Language technology research is carried out in three closely integrated research centres: the National Centre for Language Technology (NCLT), the Centre for Next Generation Localisation (CNGL) and the Centre for Translation and Textual Studies (CTTS), hosted in the School of Computing and the School of Applied Languages and Intercultural Studies. At DCU the centres host a

\(^1\)European Qualification Framework for Lifelong Learning, European Commission, Educational and Culture, 2008
total of 50+ faculty, post-doctoral researchers and PhD students. Since 2000 more than 30 PhD students have graduated in Language Technologies in DCU. CNGL, NCLT and CTTS have strong research expertise in Machine Translation, Multilingual Treebank-Based Grammar Acquisition and Parsing and Generation with these grammars, Parsing Web 2.0 and User-Generated Content, Error-Sensitive Parsing, Language-Oriented and Multimedia Information Retrieval, Cross-Language Information Retrieval, Professional Translation and Localisation. The centres currently participate in eight EU FP7 projects: EuroMatrixPlus, META-NET, Panacea, Pluto, Cosyne, PetaMedia, Kreshmoi and AXES. The DCU PhD program includes professional development courses (transferable skills such as Entrepreneurship, Intellectual Property Management) and research-focused tutorials and workshops. PhD students come from across Europe, Asia, as well as South and North America.

**The University of Edinburgh (UoE):** UoE offers taught post-graduate degrees in over 160 subjects, with research post-graduate degrees offered by over 130 different academic units. Research in the field of Language Technology is conducted at the School of Informatics, the highest-ranking department in the area of computer science and informatics in the UK in the last two Research Assessment Exercises (2001 and 2008). Within Informatics, the Institute for Language, Cognition and Computation (ILCC) conducts research on all aspects of Language Technology, including the development of linguistic data resources and toolkits, information extraction and retrieval, speech processing, multi-modal dialogue systems, statistical machine translation, and large-scale robust natural language processing. The institute currently comprises 25 academic faculty, 42 research staff, and 56 doctoral students. ILCC routinely engages in international collaborative research, including projects funded by the European Commission (such as EuroMatrixPlus, XPERIENCE, GRAMPLUS, SynProc, Casmacat, James and SpaceBook). It is a partner of the European Master in Informatics (EuMI) and has recently established a joint PhD program with Macquarie University. Recognising the importance of knowledge and skills beyond that of the research area, professional development and careers-related courses are offered to postgraduate students through the University’s Transkills programme, and students are encouraged to enroll. The School also offers postgraduate courses in entrepreneurship and an Entrepreneurship Club with a full schedule of events throughout the year.

**Rijksuniversiteit Groningen (RuG):** The Rijksuniversiteit Groningen provides high quality teaching and research, is internationally oriented, respects differences in ambition and talent, works actively with business, the government and the public, and ranks among the best universities in Europe. Research in Language Technology at the Rijksuniversiteit Groningen is embedded in the Center for Language and Cognition Groningen (CLCG). The research is carried out in a broad range of projects in the Computational Linguistics group and in collaborations with other research groups of the CLCG, and with researchers of computer science and artificial intelligence. Most of the externally funded projects in language technology involve cooperation on a national or international level. In particular, the CL group has participated prominently in seven large projects in the STEVIN initiative (Dutch-Flemish Research Program for Dutch Language and Speech Technology), and in the recent NWO programme on Language and Speech Technology IMIX. Current projects at CLCG include the NWO program “Modelling Textual Organisation (MTO)”, the international projects “Determinants of Dialect Variation” and “Measuring Linguistic Unity and Diversity”, and “Building ICT Research Capacity in Uganda”. Recently, Johan Bos initiated a five year project on computational semantics involving several post-docs and Ph.D. students, entitled “Deep Meaning Annotation Project” (D-Map).

**Univerzita Karlova v Praze (Charles University in Prague (CUNI)):** The Institute of Formal and Applied Linguistics (UFAL) is one of the departments of the School of Computer Science which is one of three schools of the Faculty of Mathematics and Physics of the Charles University in Prague (CUNI). It builds upon the tradition of the Prague Linguistic School which has been an influential factor in linguistics since 1930’s. The Institute is mainly research-oriented, but as an integral part of a School of Computer Science it also provides master and PhD level education in the field of computational linguistics (around 25 PhD students have graduated in the Institute in the last 10 years). The Institute has grown in the past two decades into the current size of approximately 60 professors, teachers, researchers and PhD students employed either directly by the Institute or working on various national or international research projects. The research in the Institute covers a broad range of topics, with formal description of natural languages (the Functional Generative Description of Prof. Sgall constitutes an internationally recognised theory), corpus linguistics (The Prague Dependency Treebank seems to be the most visible achievement internationally), machine translation (both stochastic
and rule based), parsing and speech recognition being just a few examples. Since 2006 the Institute has been a partner of the Erasmus Mundus Program “LCT – European Masters Program in Language and Communication Technologies”.

**Universität des Saarlandes (Saarland University (UdS); coordinator):** The Department of Computational Linguistics and Phonetics (COLI) of Saarland University (UdS) is one of the world’s leading centres for LT, featuring a unique aggregation of basic and applied research projects and information services. The department has four tenured professors positions and a research and teaching staff of about 40. It offers a B.Sc. in Computational Linguistics, an M.A. in Phonetics, an M.Sc. in Language Science and Technology, a PhD in Linguistics as well as in Phonetics, and, last but not least, is the coordinator of the Erasmus Mundus Program “LCT – European Masters Program in Language and Communication Technologies”, in which it provides students with profound knowledge and insight into the various disciplines that contribute to the methods of language and communication technologies and it strengthens their ability to work according to scientific methods. The department has had a joint European PhD program with the University of Edinburgh, which was funded by the German Research Association (DFG), and an international PhD program with the Johns Hopkins University, the Brown University and the Charles University in Prague, which is co-funded by the US National Science Foundation (NSF Partnership for Research and Education (PIRE)) and the German Research Association (DFG). Through Prof. Hans Uszkoreit and Dr. Valia Kordoni, the department has a close relationship to the German Research Centre for Artificial Intelligence (DFKI): Prof. Uszkoreit is the Scientific Director of DFKI and Dr. Kordoni is also affiliated to DFKI as a Senior Researcher in the DFKI Language Technology (LT) Lab, which is located in two sites, Saarbrücken and Berlin. DFKI is Core Partner of ICT Labs, the consortium chosen as the Knowledge and Innovation Community for the ICT area by the European Institute of Technology (EIT). The Department of Computational Linguistics and Phonetics (COLI) at UdS, the coordinator of the RELATE consortium, already supports a full position of a secretary and a full position of a researcher for administration, and will apply for extra funds for one more full position of a researcher to support the RELATE program. In addition, all partner universities of the RELATE consortium provide administrative support for teaching activities both at university level and at faculty level.

**B.2.2 Complementarity and Diversity of the consortium members**

The Consortium consists of the Core Consortium with five European full-partner institutions, selected on the basis of their quality work in Language Technology and long-standing record of previous cooperation. The core institutions have the main responsibility for training of the core skills in summer and winter schools and at home, for research supervision and joint project development. This core consortium is extended with associated partners from several sectors.

This diverse set of core institutions and partners is not gratuitous: Rather, it is aimed at enabling RELATE to achieve its stated objectives of (1) delivering high-quality trans-disciplinary inter-sectoral PhD training in Language Technology (diversifying the sectors and disciplines that we cover); (2) producing a new generation of theorists and practitioners able to carry out world-class research and development in core areas of Language Technology (diversifying the core areas that are covered); and (3) contributing to the technological basis of the multilingual European society in the digital age (diversifying the social and industrial institutions that we contribute to, in their national languages).

Within the consortium, the coordinating institution is responsible for administering, monitoring and coordinating matters pertaining to the consortium, its staff, programme, mobility schemes and candidates, and for communicating with Erasmus Mundus EACEA. While the full-partner institutions share well-recognised expertise in Language Technology (as detailed in Section B.2.1), they have complementary strengths born of different backgrounds, motivations and foci, as detailed below.

**Semantic analysis and inference.** Developments in this area provide an abstraction layer above existing IT technologies that enables bridging and interconnection of data, content, and processes. Partners’ approaches to lexical semantics, sentence-level semantics (including that conveyed through intonation), and textual entailment and inference reflect

- a desire to integrate visual and textual information for image annotation and multi-modal information retrieval (DCU, UoE);
- benefits to be gained from crowd-sourcing semantic annotation (RUG);
- a desire for informative, realistic models of child language acquisition (UoE);
• benefits to be gained from semantics-enhancements to statistical machine translation (DCU, CUNI);
• a desire to model and exploit discourse context in drawing entailments from text (UdS) and in improving the coherence of statistical machine translation (UoE).

Future advances in research and technology development in this area will benefit from awareness of and attention to these different motivations.

**Machine Translation and Cross-Lingual Technologies.** This area demonstrates complementarities across all partners, in terms of methods used (phrase-based methods, factored methods, hybrid methods, as well as syntax-enhanced, semantics-enhanced and discourse-enhanced), languages and language-families being addressed, and the richness and complexity of morphological and grammatical structure being tackled. These are meant to help solve problems in statistical machine translation, such as handling grammatically non-local phenomena including long-distance dependencies and complex coordination, as well as discourse-local phenomena such as ellipsis, anaphoric reference, and coherence relations (both resolution of ambiguous explicit cues to coherence relations and identification of implicit relations), and dealing with lexical and syntactic gaps in the training data. They are also meant to solve other problems experienced in rule-based MT, such as word-sense disambiguation, semantic selection, stylistic preferences, as well as gaps in the lexicon and grammar.

Related to this is complementary work on cross-lingual technologies, which aim to work with multiple languages without the cost of having to translate between them. Cross-lingual technologies can be applied to question answering, allowing users to search for information in several languages, cross-lingual summarization, allowing information to be aggregated and distilled from sources in several languages; and cross-lingual analysis of both the objective and subjective contents of social media.

Particular complementarities in these areas arise from concerns associated with
• morphologically-rich languages (CUNI);
• resource-poor languages (UdS, UoE);
• the needs of cross-lingual information retrieval (DCU)
• adapting systems to a different genre (or mix of genres) (DCU, UoE)
• gathering manually-annotated gold-standard data for training syntactic-, semantic-, and discourse-enhanced methods (CUNI, UoE).

**Robust Language Processing.** Here complementarity arises from partners’ use of different formalisms (e.g., Head-Driven Phrase Structure Grammar, Combinatory Categorial Grammar, Tree-Adjoining Grammar, Lexical-Functional Grammar and Dependency Grammars) that differ in the ease and simplicity in which they capture different types of language regularities, as well as their use of different forms of input, different applications and different drivers. Particular complementarities arise from
• attempting to adapt language tools and models to gradually changing characteristics of web-scale data streams (UoE) and lexical and syntactic characteristics of different domains and genres (RUG);
• multilingual parsers that are not limited by the morphological and typological features of one particular language (UoE, UdS and RuG);
• methods of inducing parsers for low-resource languages from a combination of data and typological features (UoE, CUNI);
• methods for adapting parsers to emerging informal communicative forms, such as blog posts, tweets and clinical notes, as well as to characteristics of spoken language in meetings and on mobile devices (DCU, RuG and UdS);

Advances in research and technology development will again benefit from the complementary motivations of the range of this work.
Speech analysis and synthesis in support of natural dialogue. Here, partners UdS and UoE have demonstrated long-term strong complementarity in the framework of the Marie-Curie ITN SCALE (Speech Communication with Adaptive LEarning). Their work in this area has been motivated by the demand for systems to interact through speech with a wider population, in wider range of environments, and for a wider range of purposes. Originally, speech understanding systems were limited to male voices in quiet environments for transcription or simple tasks. Now systems are needed that can engage in spoken dialogues with the elderly, with children, and with speakers of regional variants, in a range of environments, both indoors and out-of-doors in urban and rural environments.) Particular complementarities arise from:

- application-related concerns to do with structured multi-party events such as meetings (UoE), reference to a changing visual environment (UoE), or spoken question-answer (RuG) and spoken machine translation (DCU, CUNI), tutoring (UoE), and animated conversational agents (UdS, UoE);
- the challenges of personalized speech synthesis for people suffering from degenerative diseases (UoE);
- a desire to generate contextually and socially appropriate intonation (UoE).

Research and training in this area will target the development of better acoustic models that are discriminatively trained, better adapted or normalised to new domains or speakers, more flexible approaches to dialogue modelling, and greater integration of speech with semantics.

B.2.3 Networking and Cooperation Activities

As can be expected from the outward-looking research leaders that make up the RELATE consortium, all partners are involved in national and international collaborations and joint education and research projects. Moreover, the capabilities of the individual partner organisations will be strongly amplified by the planned integration of their respective expertise, experience, skills, resources and networks.

In pursuing its objectives, the EMJD RELATE will cooperate closely with the main professional associations and existing research infrastructures in language technology.

EACL, the European Chapter of the Association for Computation Linguistics has endorsed the goals of RELATE and expressed readiness to actively support them.

The same is true for EAMT, the European Association for Machine Translation.

The European Network of Excellence META-NET is dedicated to supporting research and innovation toward the technological foundations of the multilingual European society in the digital age. The partners of the RELATE consortium are active members of META-NET. The coordinator of RELATE also serves as coordinator of META-NET. With its 47 member organizations in 31 countries the Network is strengthening research and innovation in language technology by opening new avenues in research, building infrastructures for sharing resources and forging the alliance META (Multilingual Europe Technology Alliance) of many stakeholders including besides researchers, provider and user industries as well as the language communities. This alliance now has about 200 member organisations.

Several partners of RELATE are also members of EuroMatrixPlus, the largest ongoing EU research project in machine translation with annual evaluation campaigns, marathon workshops and training events. Partners of the consortium will also use their active role in other networks and professional associations in order to mobilize additional support for the objectives of RELATE.

The International Research Training Group in Language Technology and Cognitive Systems was established in cooperation between Saarland University (Departments of Computational Linguistics, Phonetics, and Computer Science) and Edinburgh University (School of Informatics) – two leading institutions in the fields of computational linguistics, artificial intelligence, and cognitive science, who collaborate in offering a joint post-graduate education programme. There is also an NSF-funded partnership for research and education (PIRE), established in 2005, that runs a collaborative PhD programme between Saarland University, Charles University, Brown University (its Laboratory for Linguistic Information Processing headed by Prof. Eugene Charniak), and Johns Hopkins University (its Center for Language and Speech Processing). PIRE is also affiliated with UdS-COLI’s existing IRTG co-operation with Edinburgh.

In addition to the aforementioned joint post-graduate programs and networks, two of the partners (DCU and UoE) and one academic AM (UniMelb) are members of Universitas 21 (http://www.universitas21.com), an international network of 21 leading research-intensive universities. In 2009, thirteen of the Universitas 21 members signed a Memorandum of Understanding to initiate a U21 jointly-awarded PhD program.
With respect to collaboration on MSc degree programs, three of the partners (UdS, CUNI, RuG) and one academic AM (UniMelb) have been offering the Erasmus Mundus European Masters Programs in Language and Communication Technologies (LCT), strengthening students’ ability to work in industry and research according to scientific methods in the fields of Language Technology, Computational and Theoretical Linguistics, and Computer Science.

The other academic and industrial AMs are also in research cooperation relations with the consortium partners which will be further reinforced and extended in the framework of RELATE.

B.2.4 Professional Sector Participation

One of the key aspects of this proposal is its emphasis on promoting the exchange between academia, industry, and the public sector. The consortium Associate Members (AMs) include the large industrial partners and key players in the field of LT, like Google, Microsoft, IBM, XEROX, Symantec, Thomson Reuters and SMEs, like RightNow, research centres and public research organisations, like the German Research Centre for Artificial Intelligence (DFKI GmbH), the Dutch Language Union (Taalunie), the Institute for Dutch Lexicology (INL), the Institute of the Czech Language of the Academy of Sciences of the Czech Republic, as well as three international universities elsewhere around the globe, the University of Rio Grande do Sul (Brazil), the University of Melbourne (Australia) and the Johns Hopkins University (USA).

All of the AMs will help evaluate the results by taking part to quality assessment mechanisms, as well as help the consortium in fund raising at the local and international level.

Especially, as far as the large/medium/small industrial enterprises who participate as AMs in RELATE are concerned, these offer the PhD candidates the opportunity to carry out their research in close connection to them. Research topics inspired by the applicability and the needs of the industry will be available. Training by placement in the industry is compulsory for the industrial track and also available for the academic track. This prepares PhD candidates who want to pursue a career in the industry in the best possible way. Industrial AMs also participate in the Project Advisory Board (PAB; see Section B.5.1) of RELATE. The cooperation with the industry is based on agreements between the industrial AM and the university at which the PhD candidate is carrying out their PhD research. These agreements cover the objectives and expected outcomes of the cooperation, as well as the intellectual property rights and publishing possibilities. The aim of RELATE is to increase the number of companies which currently cooperate with the various consortium partners.

B.3 European integration and functioning of the programme

B.3.1 EMJD Programme Organisation

RELATE has been jointly developed by the partner institutions to take advantage of the different expertise available in the consortium. The flexible structure of the program leaves the necessary degree of freedom for designing individual paths for each doctoral candidate. 30 ECTS must be achieved in courses and lectures as described in section B.1.3.

The co-tutelle is compulsory: through the joint selection process (see B.3.3), each PhD candidate has two supervisors who define a training and research agenda for him/her (including milestones and mobility periods), and carry out the joint supervision of the candidate. This involves regular meetings (which may also be done over skype) at least every two months. All the agreements between the two supervisors and the PhD candidate are part of the Personal Career Development Plan (PCDP).

An essential cornerstone is that ECTS credits are used throughout the program as a means of mutual recognition of curricular activities. This system applies equally to lectures and seminars taken, as well as to summer schools, winter labs, workshops and structured literature studies. The established Bologna rules will be used to assigning credits. The consortium members have long standing experience in credit transfer among them.

Research activities are also assigned credits in order to allow for the transfer and quantitative measurement of research done at one institution to another partner institution. To this end the PhD candidates submit reports on their research twice a year. Those are assigned ungraded credits.

The previous two measures together allow for a flexible, yet structured program. Within this framework, the pair of supervisors and the PhD candidate are free to choose the optimal career path.
As mentioned in A.1.3, we plan to have one RELATE summer school per year. The organisation will rotate among the partners each year. Each summer school will have a specific topic and internationally renowned experts will be invited. It may also involve a hands-on practical part, as well as a complementary skills training. Where possible the summer schools will be held at internationally well known sites like Royal Academy of Ireland, Dagstuhl or Fribush. Those events will then also be open to external participants. Note that existing external summer schools (e.g., the European Summer School in Logic, Language and Information (ESSLLI)) may also be attended for credit. Moreover, we also plan to have winter labs which, as also mentioned in A.1.3, will host international teams for an intensive 2-week research lab-based workshop on language and speech technology and engineering. Each lab-based workshop team, consisting of internal and external LT experts, PhD candidates and post-docs of RELATE will spend 2 weeks together at one of the partner universities (after some advance preparation and in a different partner institution every year), working in close proximity on some challenge problem or promising solution technique that has not yet been well studied. Finally, there will be a 2-day workshop each year. Its organisation will also rotate among the partners and its focus will be on progress reports presented by the PhD candidates, and more importantly on discussions about their research and its development. This event may also involve external experts, whose function will be that of the external consultant to the research of the PhD candidates. All the aforementioned three activities are important activities where all partners meet together, allowing for exchange of best practice examples and for the identification of new cooperation possibilities.

To conclude, outstanding research environments are provided by the excellent research equipment available in the different partners of the RELATE consortium, and by the expertise of its members. Excellent associated partners recognised worldwide in their specifically concerned fields will be appointed to collaborate with the PhD candidates. Compulsory mobility periods at an associated partner institution in addition to the mobility periods at the consortium institutions will enable such collaboration and thereby cooperation with the research groups of the consortium will be provided.

**B.3.2 Mandatory Mobility Periods**

PhD candidates start at any of the partners involved in the selected mobility. They move to a partner university working on the same topic (at least 6 months; if necessary, the 6 months could be achieved in several shorter periods). An initial list of topics for the first cohort of twenty candidates see 1. Research, as well as courses (scientific or in complementary skills) will be the activity at the partner university.

As indicated in B.1.3 we have two different tracks, an academic track and an industrial track in order to optimally prepare the candidates for a career in either academia and industry. As far as the industrial track is concerned, an additional mobility to an industrial partner must take place for at least two weeks. In addition to scientific exchange and possible use of research equipment, the exchange should provide the PhD candidate with insight into the industrial way of working. Such exchanges could be longer or repeated as necessary. The destinations for doctoral candidates are not restricted to the industrial AMs included in this application. Moreover, other companies may be elected by the students or may join the RELATE network. The Project Steering Committee (PSC; see Section B.5.1) will decide on the recognition of the compulsory mobility period at the industry for the industrial track.

For the academic track, there may be a non-compulsory additional mobility for at least two weeks which will take place to a university in a third country. That period should serve for use of research facilities or scientific exchange with the top-class associated partners. This mobility will normally take place in the second or third year, when PhD candidates have already worked for a while on their research subjects. Such exchanges could be longer or repeated as necessary.

Implementation of mobility will be done using a so-called “mobility account” for each candidate and will be defined from the beginning in the personal career development plan by the supervisors and the students and harmonised by the Project Steering Committee (PSC). Explicit mobility details will be developed on an individual basis in the PCPD and will depend on the chosen track, chosen topic, and the participating partners. The mobility periods are additionally influenced by the timetable of the courses available, the equipment to be used at both universities or at an associated partner institution, etc. By considering all these details, the mobility periods will be realised with maximum efficiency.

Some of my own suggestions (thinking of Gertjan and Johan Bos as well):

and from the old proposal (i dont see it anymore...)
<table>
<thead>
<tr>
<th>Technology</th>
<th>PhD topics</th>
<th>Host Institution</th>
<th>Partner Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>UWC</td>
<td>Text, Video and Speech Retrieval</td>
<td>DCU</td>
<td>UoE</td>
</tr>
<tr>
<td>UWC</td>
<td>Cross-Language Multi-Modal Information Retrieval</td>
<td>DCU</td>
<td>??</td>
</tr>
<tr>
<td>UWC</td>
<td>Multimodal Search in Call Center Archives</td>
<td>CUNI</td>
<td>??</td>
</tr>
<tr>
<td>UWC</td>
<td>Automated indexing of speech data via unsupervised phrase extraction</td>
<td>UoE</td>
<td>??</td>
</tr>
<tr>
<td>UWC</td>
<td>Exploiting gaze tracking in multi-party conversation</td>
<td>UoE</td>
<td>??</td>
</tr>
<tr>
<td>UWC</td>
<td>Enhancing spoken language processing with social dynamics</td>
<td>RuG</td>
<td>??</td>
</tr>
<tr>
<td>UWC</td>
<td>Deep and Shallow linguistic analysis for Semantic Entailment</td>
<td>UoE</td>
<td>??</td>
</tr>
<tr>
<td>UWC</td>
<td>Exploiting Language models for multi-party conversation</td>
<td>UoE</td>
<td>??</td>
</tr>
<tr>
<td>UWC</td>
<td>Sentence retrieval for relation extraction</td>
<td>UdS</td>
<td>RuG</td>
</tr>
<tr>
<td>UWC</td>
<td>Information Extraction in Semantic Web Context</td>
<td>RuG</td>
<td>UdS</td>
</tr>
<tr>
<td>WCSU</td>
<td>Minimally Supervised Tuning to User-Generated Content</td>
<td>DCU</td>
<td>??</td>
</tr>
<tr>
<td>WCSU</td>
<td>Error Sensitive Statistical Parsing</td>
<td>DCU</td>
<td>RuG</td>
</tr>
<tr>
<td>WCSU</td>
<td>Integrating Multiple Knowledge Sources in Statistical Parsing</td>
<td>DCU</td>
<td>??</td>
</tr>
<tr>
<td>WCSU</td>
<td>Statistical Parsing for Opinion Mining</td>
<td>CUNI</td>
<td>??</td>
</tr>
<tr>
<td>WCSU</td>
<td>Unsupervised Semantic Analysis of Parallel Texts</td>
<td>RuG</td>
<td>??</td>
</tr>
<tr>
<td>WCSU</td>
<td>Semantic enrichment of user generated content</td>
<td>RuG</td>
<td>??</td>
</tr>
<tr>
<td>WCSU</td>
<td>Cross-lingual Question Answering for Open Linked Data</td>
<td>RuG</td>
<td>??</td>
</tr>
<tr>
<td>WCSU</td>
<td>Linguistic Analysis of user generated content</td>
<td>RuG</td>
<td>??</td>
</tr>
<tr>
<td>WCSU</td>
<td>Bayesian Topic Modeling for Dialog Systems</td>
<td>CUNI</td>
<td>??</td>
</tr>
<tr>
<td>OLB</td>
<td>Statistical Translation and Information Retrieval and Extraction Technologies</td>
<td>DCU</td>
<td>??</td>
</tr>
<tr>
<td>OLB</td>
<td>Statistical Translation Models, Transfer Learning and Domain Adaptation</td>
<td>DCU</td>
<td>UoE</td>
</tr>
<tr>
<td>OLB</td>
<td>Integrating Multiple Knowledge Sources in Statistical Machine Translation</td>
<td>DCU</td>
<td>??</td>
</tr>
<tr>
<td>OLB</td>
<td>Error Analysis and the Role of Compositionality for High Quality Translation of Phrasal Verbs</td>
<td>UdS</td>
<td>RuG</td>
</tr>
<tr>
<td>OLB</td>
<td>The role of parsing for High Quality Machine Translation of Multiword Expressions</td>
<td>UdS</td>
<td>RuG</td>
</tr>
<tr>
<td>OLB</td>
<td>MT for Languages Pairs (almost) without Parallel Data</td>
<td>CUNI</td>
<td>??</td>
</tr>
<tr>
<td>OLB</td>
<td>Bayesian Techniques for MT Quality Evaluation</td>
<td>CUNI</td>
<td>??</td>
</tr>
<tr>
<td>OLB</td>
<td>Handwriting recognition for non-Latinate orthographic systems</td>
<td>UoE</td>
<td>??</td>
</tr>
<tr>
<td>OLB</td>
<td>Lexical-syntactic cross-linguistic priming in bi-lingual communication</td>
<td>UoE</td>
<td>??</td>
</tr>
<tr>
<td>OLB</td>
<td>Semantics in SMT</td>
<td>UoE</td>
<td>DCU</td>
</tr>
<tr>
<td>OLB</td>
<td>Discourse-informed SMT</td>
<td>UoE</td>
<td>CUNI</td>
</tr>
<tr>
<td>OLB</td>
<td>Cross lingual retrieval of patents</td>
<td>UdS</td>
<td>DCU</td>
</tr>
<tr>
<td>OLB</td>
<td>Machine translation of patents</td>
<td>UdS</td>
<td>??</td>
</tr>
</tbody>
</table>

**B.3.3 Common Standards and Mechanisms: application, selection, admission and review of doctoral candidates (European and third-country)**

**Application procedure**

The application process is organised by the RELATE coordinator at UdS, and reviewed by all members of the consortium. Applicants apply online (by completing and submitting an application form on the RELATE website). In case where an online application is technically not possible the documents may also be sent to the coordinator’s office per snail mail. The following documents are required:

- A brief statement of interest (at most two pages) written in English, focusing on (a) the reasons of the candidate for applying to the RELATE program; (b) the interest of the candidate in the various disciplines that contribute to the methodologies of Language Technologies (Language Technology, Computational and Theoretical Linguistics, and Computer Science); and (c) a list of preferences for the home and partner universities.

- A current Curriculum Vitae, written in English.

- Copies of the BSc and MSc degree certificates, as well as certificates attesting to the fulfillment of the language requirements (see below).

- A transcript of academic records of the candidate’s BSc and MSc curricula studies.

- Two letters of recommendation, uploaded directly by the referees or sent sealed to the coordinator’s office.

- Publications, if available.
• Other additional supporting documents, e.g., a report on competence in languages other than English.

**Admission and Selection criteria**

To guarantee transparent, fair and objective selection criteria, we assess students with respect to their **eligibility** for the programme, as well as **qualitative** and **quantitative** measures.

**Eligibility criteria:** All applicants must meet the application deadline; provide a complete application; demonstrate English language proficiency; and possess an MSc degree or equivalent, by the start of their PhD program. The MSc degree or equivalent should be in the area of (Computational) Linguistics, Language Technology, Cognitive Sciences, Computer Science, Mathematics, Artificial Intelligence, or other relevant disciplines, subject to the consent and approval of the consortium’s Project Steering Committee (PSC).

Admission to the program is not restricted to the scholarship awardees. Depending on the study background of the applicants, some bridging courses up to a certain amount of ECTSs may be required.

**Qualitative measures** will take into account the learning outcomes of the second cycle (level 7) as defined by the EQF, namely:

- **Knowledge** Highly specialised knowledge as the basis for original thinking and research; critical awareness of knowledge issues in Language Technologies (LT) and Computer Science (CS).
- **Skills** specialised problem-solving skills required in research and innovation in order to develop new knowledge and procedures and integrate knowledge from LT and CS.
- **Competence** manage and transform work or study contexts that are complex, unpredictable and require new strategic approaches; take responsibility for contributing to professional knowledge and practice and/or for reviewing the strategic performance of teams.

Furthermore, as foreseen by *The Code of Good Conduct for the Recruitment of Researchers*, the selection process will take into account a candidate’s full range of experience.

**Quantitative measures** include grades awarded during the BSc and MSc studies of the applicant, publication of the MSc thesis, and language assessment.

**Selection of doctoral candidates**

Applicants are selected by the Project Steering Committee (PSC), which includes 2 representatives from each partner university (see Section B.5.1). The PSC evaluate the applications and establish a ranking of all the applications.

Applicants’ portfolios are made available to the RELATE PSC three weeks before the meeting at a secure online site. This practice makes it possible for every partner to check the applications and to solicit further opinions from their own university well in advance of the meeting, and identify those cases which need detailed discussions to be conducted at the meeting. After discussion, a shortlist of accepted applicants is devised. In case the PSC feel the need, applicants are to be interviewed via teleconference at one of the proposed host universities or via SKYPE. This might involve a 20-minute presentation (in English) of the applicant’s latest scientific project and future interests, followed by a 30-minute interview.

Applicants are either granted their choices for host institutions or alternatives are suggested. Applicants should confirm with the coordinator that they accept the proposal of the RELATE consortium about the host institution, and can then proceed to enrollment as PhD candidates with the corresponding partner.

The selection criteria are applied as follows: 1. All criteria are equally weighted except for the English proficiency criterium, which is applied only in case of equal ranking; 2. We use the schema 5, 4, 3, 2, 1, encoding “excellent”, “very good”, “good”, “satisfactory” and “weak”, respectively; 3. Each application is assessed several times; pre-assessment is performed by all partners – the selection committees of all partners perform a two-round evaluation of all candidates; 4. Next round of assessment is done by the PSC; 5. Diverging scores are discussed by the PSC, this process leads to commonly agreed ranking of candidates; and 6. The absence of conflict of interest is guaranteed by making the assessors aware of this concern and by multiple assessment. More specifically to what has been mentioned in the previous, the criteria are: 1. Motivation for following the doctoral program; 2. Study record and examination results from previous...
studies; 3. References (especially, if their recommended by any of our contacts); 4. Knowledge of English; 5. Experience in research and development at universities and/or industry; 6. Overall experience in the field of LT; 7. In case of financial support by any other program, it will be checked whether the candidates fulfill the conditions of the RELATE program. The assessment results are finalised by electronic discussion of the PSC.

RELATE aims for a representative gender balance amongst the applicants, based on an equal opportunity policy during the recruitment and selection phase. All partners of the program are already committed to gender equality, and the Programme adheres to the requirements established in the The Code of Good Conduct for the Recruitment of Researchers.

In sum, the recruitment and admission process respects the rules of the Code of Conduct for Recruitment as laid out in the European Charter for Researchers, according to the principles of: Clarity of information to enquirers, prospective applicants, applicants and their representatives; Fairness in recruitment and admission procedures; Consistency in the application of policies and procedures; Openness and probity; Professionalism, efficiency and timeliness; Diversity and equality. Gender balance is less of an issue in area of language processing because so many women are drawn into the field when they become aware of it. Both broad and targeted advertising of the program is therefore essential.

Closely related is the issue of equal opportunities for male and female researchers which is ensured at all stages. This means in part that the consortium partners will adhere to family-friendly policies, including child-care provision at the annual RELATE meeting and meetings and teleconferences held as close to normal working hours as possible. The RELATE consortium and program attend to the gender balance of committees, and, wherever possible, supervisory teams, as well. Jobs and internships will be advertised to all students, independent of gender or special needs.

Besides the attention paid to gender and special needs within the consortium program, most consortium partners maintain Equality and Diversity Offices at the University level and a Disability Office whose focus is to provide advice and support to individual disabled students, and work with the rest of the University to improve accessibility and awareness in general.

Alternative Ways of Recruitment and Life Long Learning

The procedures outlined above form the standard procedure. However, in special circumstances applications outside of this framework may also be considered. Applicants that apply outside the regular recruitment periods and where special reasons justify a special treatment will be discussed in the Joint Project Committee. If an exception is granted, the other rules and policies outlined above will be applied. If successful, special arrangements will be considered to phase the candidate into the program.

One possible reason for a special arrangement could be candidates who apply from existing industry positions and want to obtain additional qualifications through a PhD in RELATE. In this sense it would already contribute to the Life Long Learning objectives of the European Commission.

Another aspect of Life Long Learning is to enable PhD candidates after the completion of the program to continue to educate themselves, stay flexible and acquire new skills. To this end, we will make sure that suitable complementary skills courses are taken as part of the training.

B.3.4 Quality of Joint Supervision and Monitoring

Co-supervision of the PhD candidate is compulsory. Each candidate will have an appointed supervisor from each university where he/she carries out the research project. A tentative list of topics can be found in Table 1. PhD candidates regularly present their results to the departmental staff and other PhD candidates in seminars. This activity not only trains the PhD candidates in giving presentations, but also permits the other members of the departments and of the other universities to monitor the research activities of the doctoral student. In particular the PhD candidates will have to write a report twice a year. Initially the report will be on

- Courses taken and the ECTS credits being obtained
- A literature review for the chosen topic. This can later became part of the first chapters of the thesis
- A detailed plan for the PhD thesis including a risk analysis and alternatives.

Those reports will be reviewed by the two supervisors of the candidate. The Project Steering Committee as well as the external Project Advisory Board will obtain a copy of all of the reports for all of the candidates. In particular the Project
Advisory Board consisting of internationally well known experts from academia and industry will be able to provide strategic feedback for each of the candidates. Finally those reports will be uploaded on the homepage of the program (without injuring IP issues), in order to inform the other members of the network about the activities of the different research groups.

In later stages the reports will shift in focus to research results. This will in particular involve publications or whenever applicable IPR filed by the candidate. Overall we expect each candidate to write at least one paper in year two and another one in year three. A paper may be replaced by a patent application.

In addition, the doctoral candidates will have to present their research activities during the RELATE annual meetings. The members of the Project Steering Committee, the other professors and additionally the experts from industry will thus have the chance to evaluate the progress achieved by the candidates.

The success of the candidates will be measured by the success in the taught component as well as in the research. Quantity and quality of the publications (or invention disclosures) can be taken as an evaluation metric. The local coordinator at each host institution will monitor the progress (courses taken, presentations, publications) for all of the candidates presently being at this institution. The local coordinator is in contact with the coordinating institution, reporting all credits and achievements into a central data base. This allows to detect deviations from the planing early and enables taking counter measures.

For the thesis defense, both supervisors must be present. The doctoral thesis will be examined by professors of the different universities and by external professors, in order to enhance the quality of the theses. When industrial partners are involved in the research project (especially in the industrial track), they will be invited to participate in the examination procedure.

B.3.5 Degree(s) Awarded

The objectives and implementation of the joint RELATE are in accordance with the current regulations of each partner university participating in the consortium for awarding a doctoral degree. Thus, the degree is recognised by each partner university of the consortium automatically. Activities leading to award of ECTS credits, according to each PhD candidate’s individual study plan, are also fully recognised at the consortium’s partner universities (see Annex “Multilateral Agreement”). In the first phase a double degree will be awarded by the two universities where the PhD candidate is carrying out their research and training work.

Some partner universities (e.g. UoE, UdS and RuG) already now allow for joint degrees based on bilateral individual agreements. This will enable a subset of the consortium to establish joint degrees from the very beginning.

In order to make a joint degree feasible between all of the partners, a working group including project PIs and PhD students will be established:

- Y1: review of best practice, objectives, requirements analysis (academia, industry, society)
- Y2: presenting initial results to public and opening up to external comments and input form academia, industry, society
- Y3: publication of a white paper and implementation according to the results.

In addition to the awarded degrees, a Joint European Excellence Certificate (in the form of a Diploma Supplement) will be issued by the consortium with the seals of all five partner universities.

B.4 Provisions for EMJD candidates and fellowship holders

B.4.1 Information and Promotion Strategy

The main objective of the project information and promotion strategy is to address, attract, prepare and recruit the best possible PhD candidates and post-doc fellows with relevant skills and knowledge, thus ensuring high standards of training and the overall success of the project. Additionally, the consortium aims at making the best practices obtained in the project available to the entire academic community. The prime way of spreading the message will be through the extensive international networks of the RELATE partners, associated partners and regular networking activities. On top of that the following targeted activities will be implemented through:
the central RELATE project website and websites of all partners (e.g., publishing of the Program, all PhD positions will be announced there);

special web portals listing available research fellowships (e.g., all PhD positions will be announced through EURAXESS, national resources such as http://www.jobs.ac.uk and http://www.findaphd.com will be also used, where appropriate);

thematic LT research community-driven mailing lists, including wide-ranging global lists such as Corpora List, MT List, Moses List and Linguist List;

dissemination of information through major national agencies promoting European research and education globally (e.g. DAAD, British Council, etc.);

major social networks like Facebook groups “Speech & Language Technology Enthusiasts” and the “Australasian Language Technology Association”, as well as Twitter, YouTube, Podcampus, etc.;

creation of an RELATE Alumni Association, members of which will be invited to participate in the annual RELATE meeting and will act as multipliers and promoters of the Program worldwide;

other activities foreseen by the Information and Promotion Plan (see below), including participation in major events, production and distribution of promotional materials, like leaflets, posters, as well as promotional entries in special brochures distributed by national agencies, etc.;

Participation in study fairs such as Europosgrados in South America and Study World in Berlin are common practices for the program’s promotional activities.

Planning and monitoring of information and promotion activities will be coordinated by the Central Management Office (CMO, UdS), which, in close consultation with the RELATE partners, will develop (by month 3), orchestrate, execute and annually update the Information and Promotion Plan. The Office will prepare evaluation reports on information and promotion activities for review at the annual RELATE meetings.

Special attention will be paid to ensure adequate representation of women in the program: this will be achieved by spreading the information through targeted mailing lists for women in science and research (e.g., FiT association in Germany, Society of Women Engineers, etc.) and encouraging participation of women through all RELATE promotional materials.

B.4.2 Hosting Services

All consortium partners will appoint the RELATE local coordinators who will be responsible for the hosting services jointly with the International Offices and other respective departments of the partner universities. The goal is to seamlessly incorporate the project’s mobility into the regular university services provided to incoming students at all locations and make use of available infra-structures. At the same time, the role of the local coordinators will be to help the incoming fellows in arranging their stay (especially, during the first couple of months) and getting integrated into the local environment.

Importantly, the CMO will develop a special Guide for local coordinators including a checklist with issues to be taken care of and recommendations on how to tackle problems. This Guide will help to get the local coordinators prepared for handling the first incoming candidates.

The services will cover the following:

Initial preparation

- The CMO will get in contact with all selected candidates to (i) inform them about studying and staying in Europe and (ii) get personal information and requirements (e.g., marital status, children, preferable accommodation, etc.) to be provided to the local coordinators

- The candidates will receive from the local coordinator an information package with the localised details on visa requirements, accommodation and child care options, other useful information.
Visa support
This service will be provided by the local coordinators in cooperation with the International Offices and in accordance with national regulations and each partner university’s regular practices. RELATE will provide all the necessary documents (e.g., invitation letters), advice on procedures, support through Consulates, where needed, etc.

Housing
All partners will make use of available accommodation for incoming PhD candidates. The students will get local advice and support in searching for accommodation and fulfilling the formalities (responsibility of local coordinators).
As for the special housing services available in partner universities, UdS possesses the system of dormitories and residences (on campus and in the city) for ca. 2,500 people. The standard practice is that the international incoming students (incl. PhD candidates) have a priority access to these facilities. Additionally, the student association (ASTA) maintains the database of housing options in the city and provides free consulting to all students looking for accommodation. Also, the university has special residencies for short stays of PhD candidates (e.g., those participating in the mobility).
The UoE offers a great variety of housing and catering options for UK and international students. These range from halls of residence to single rooms and entire flats. Postgraduate students from outside the EU are guaranteed an offer of student accommodation. Posgraduates from within the EU are also eligible for university accommodation. Accommodation places are also reserved for visiting and exchange students.
The University of Groningen is not a campus university, students need to find their own place to live. The Housing Office offers furnished accommodation in Groningen for international PhD students, for a minimum of 1 month up to a maximum of 12 months. Students who stay in Groningen for more than 12 months can still use this service for the first 12 months of their stay.
DCU provides 1,100 rooms for under- and post-graduate students. Extensive information is provided at the DCU webpage at http://www.dcu.ie/accommodation/index.shtml. The DCU International office will advise students on off-campus accommodation possibilities. In addition a large set of public accommodation websites cater for student accommodation.
CUNI offers a number of dormitories, which are located in the city and are easily accessible by public transportation. Foreign students at the Faculty of Mathematics and Physics usually use the Hostivar Hall of Residence, which has a direct tram connection to the Faculty. It is also not difficult to find private accommodation in the city.

Social integration
The International office of each university provides language courses, cultural programs, and integration activities like meetings with local students. The PhD candidates enrolled in RELATE may take part in various scientific and social networking activities specifically organised in cooperation with the graduate centres/schools of the involved universities (e.g., PhD candidates working groups, a web-based social network of PhD candidates, round tables for doctoral and post-doc researchers, etc.).
Moreover, the local coordinators assist the incoming PhD candidates in the following arrangements:

- dealing with employment and respective formalities (e.g., employment contract finalisation, social and pension insurances registration, tax authorities registration, etc.);
- opening local bank accounts, where needed, as well as helping with residence registration;
- registration of children for kindergarten facilities (where needed). Other formalities with the families;
- involvement in various socialising activities in local Universities (e.g., students associations, web-based networks of students, etc.).

Academic services
Continuous tutoring during the program will be provided to the PhD candidates by supervisors and other members of the various research groups. There will also be special advice in defining the individual path for each PhD candidate and regular evaluation of the validity of the individual PCDP. A scientific seminar will be organised annually at a partner university in turns. The seminar is dedicated to the presentation of current research activities at all universities. Finally, the exchange of teaching staff between the universities facilitates the integration of candidates in the new universities.

Local integration
As the study curriculum implies a long stay of PhD candidates and their families in foreign countries, it is critically important for them to get support during their integration into the local environment. All partner universities offer language training programs for foreign students and the participants of RELATE will be encouraged to take part. Moreover,
special integration courses are available from external providers (e.g., VolkshochSchule in Saarbrücken, similar services at other locations), which offer complementary training. All PhD candidates and post-doc fellows will be encouraged to get enrolled into such courses, with support provided to them by the local coordinators.

**Services to PhD candidates with families and special needs**

UdS is certified since 2004 as a family supporting university enabling work-life balance in the scientific context, offering events and advice regarding questions about family and work-life balance or an online babysitter database for university members. Additionally, a kindergarten is available on campus and in university flats students with children have priority.

The UoE has a sizeable number of flats (around 65 accommodation units) available for families. Most of this accommodation is in small self-contained flats with two to four bedrooms, many of them located centrally. There is an online application system for family applicants and a dedicated “Student Homes” web site where interested parties can search for family specific properties.

DCU runs an on-campus crèche (http://www.dcu.ie/info/creche.shtml) with weekly/monthly fees and a subsidy to lone parents who are full-time students.

At RuG, childcare is arranged through SKSG Kinderopvang. International students may be eligible for an allowance paid out by the department of Social Services and Employment (SOZAWE) of the municipality of Groningen under certain conditions.

Moreover, the range of services provided to those PhD candidates with families and children throughout the consortium include advice and support in sorting out respective formalities for the members of the families upon their arrival, obtaining social support to families with children (e.g., subsidies), obtaining places in creche facilities, social and cultural integration.

When it comes to PhD candidates with special needs, the building of all partner universities are endowed for handicapped students and all university partners have Disability Offices that provide advice and support to individual disabled students, and work to improve accessibility and awareness throughout the partner university. The local coordinators will help the individuals needing such kind of services get in touch with the respective University services and solve all issues which may occur.

Additionally, the following services are offered at several partner universities:

1. UdS has a delegate to inform handicapped students and provide service support. Further support is provided by the central Study Counselling Centre, Psychological and Psychotherapeutical Help and Information Centre.

2. The Disability Office at the UoE provides guidance and support for disable students. It supports students with dyslexia, mental health issues and students on the autistic spectrum, as well as those who have physical and sensory impairments. The UoE also offers short-term counselling to help students work through their difficulty, understand themselves better and find ways of managing their situation.

3. The DCU Disability and Learning Support Service (http://www.dcu.ie/students/disability/index.shtml) promotes equality of access and opportunities in DCU for learners with disabilities, long term medical conditions and/or specific learning difficulties.

4. The Office for Students with Special Needs offers the students to discuss their needs and requirements and provides ongoing support through their time at CUNI.

5. The RuG has developed various schemes and facilities for students with a disability. For instance, the Centre for Study Support and Academic Skills has set up a Study Management group for students with a performance disability of some sort. Students take first place in this group. The aim of the group is to gain or to maintain a regular study rhythm.

Typically, a local coordinator (or his/her supporting staff) will undertake the role of the prime facilitator personally responsible for settling arriving PhD candidates and post-docs and their successful integration, which is especially important at the initial stage of the stay (as the PhD candidates will not have sufficient language skills and local knowledge). At the same time, the CMO will maintain the central **Hosting Helpdesk** (telephone and email) which will allow local coordinators and/or the PhD candidates to receive help from the project coordinator in those cases when this is necessary.

**Post-program services** The project will offer the following services to the those who completed their PhD programs:
1. E-Mail Newsletter for the alumni of the program.

2. The Alumni Association of RELATE which will be created during the first two years of the program.

B.4.3 Language policy of the consortium

As the default language of instruction, examinations and learning support in the project is English in all partner institutions, PhD candidates will have to demonstrate a sufficient English language proficiency. At the same time and as has been also highlighted above they will be encouraged and supported to learn local languages as part of their stay in order to ensure their social and cultural integration. Various opportunities, normally free of charge, will be offered at all partner Universities (for instance, language training courses for incoming international students offered by UdS, similar services at all other locations). Where appropriate and possible, the PhD candidates will be able to obtain certain number of ECTS credits for attending such courses.

Although the PhD thesis will be written in English, the extended abstracts in the languages of the home and partner universities at which the student will be enrolled will also be provided. PhD candidates will also be invited to write an extended abstract in their native languages. All of these will be posted on the consortium website, for outreach to additional potential applicants from the same countries.

B.4.4 Candidates’ rights and Fellowship Delivery Measures

The doctoral candidates’ rights are stated in a Doctoral Candidate Agreement with the consortium. Doctoral candidates on an EM grant will be appointed as employees, if necessary also at the consortium coordinator’s institution from where they are sent to their home and host institutions. Health insurance, social security and pension rights are included. The employment salary (2,800 EUR per month) is equivalent to a 0.75 position in Germany. This is the rule also in DFG- (German Science Foundation) graduate schools. It will be paid on a monthly basis by the consortium coordinator.

Travel contribution for Category A fellowships will be paid in four instalments of 1795 EUR each, in months 1, 12, 24 and 36 of the programme. For Category B fellows there is no travel contribution foreseen since the consortium does not include a third country partner.

Immediate post-arrival expenses are sometimes a hardship for candidates. They will be told in advance to bring sufficient money to survive in the first month after arrival, since they will have to open bank accounts and send their bank details to the coordinating institution. The consortium guarantees a pre-payment for the living costs of the first month for candidates from disadvantaged regions, strained economic backgrounds, or suffering unexpected hardship.

Management of employment contracts and fellowships, collection and distribution of participation fee income, payment of travel allowances, and the organisation of summer and winter schools will be centrally done at the coordinating institution, which has recruited the necessary financial sources to appoint special administrative staff for this purpose.

Each partner will be assigned up to 5 fellows as their home candidates and another maximum of 5 within the mobility plan. During the mobility period, monthly payments will also be made by the consortium coordinator. Doctoral candidates will respect further rules and regulations of the institution where they are at home or hosted.

B.4.5 Career Prospects

Building on highly successful previous training initiatives run by RELATE consortium members (including the EM project “LCT – European Masters Program in Language and Communication Technologies”, the International Research Training Group in Language Technology and Cognitive Systems (IRTG), the NSF-funded partnership for research and education (PIRE), the Marie-Curie Initial Training Network in Speech Communication with Adaptive LEarning (SCALE), etc.), the consortium has an outstanding track record in providing excellent career opportunities for graduates: attachment 3, "Other Relevant Annexes” shows that since 2000, the consortium members individually graduated a total of 185 PhD students in Language Technologies (91 at the UoE, 32 at DCU, 31 at UoS, 18 at UG, and 13 at CUNI). 60% (98) of the students (out of the 165 for which first destination data is available) go into high-profile world-wide academic positions (including Beijing University, University of Heidelberg, New York University, Cambridge University, MIT), while 40% (67) of the students go into high-profile industry positions (including Google Research, Toshiba Research, Fujitsu, Harper-Collins, IBM, Accenture, Microsoft) in Europe and world-wide. The RELATE aims at comparable, if not better results:
Academic career prospects will be ensured through high standards in scientific training, provision of skills in scientific teaching, together with transferable skills training (including IP management, leadership, communication). Top-level publications and regular participation in international conferences will build international reputation, as well as networks in the community ensuring that RELATE graduates will be much sought after for global key academic positions in the area of Language Technologies.

Industrial career prospects will be provided through industrial secondments and active participation in the RELATE implementation by industrial leaders (in particular the RELATE industrial AMs). The PhD candidates will gain important complementary skills (e.g. managing research projects in industry, industrial requirements, innovation practices, etc.) improving their attractiveness for industrial employers. The majority of the industrial AMs plan to support post-doc stages for candidates who present an interest in continuing their professional career in industry, or who aim to obtain experiences in the private sector. Job posts in professional network LinkedIn Groups such as Natural Language Processing, Natural Language Processing People, Computational Linguistics, Translation Automation Group etc. attest the massive growth in demand for Language Technology graduates (in particular at the research level) in industry, driven by rapid growth in the areas of machine translation, data analytics (including business intelligence, data mining), multimodal applications and intelligent search. RELATE will strongly contribute to meeting this demand in Europe and at a global level.

The mechanism of an Alumni Network via Facebook will be used to monitor career development of graduates of the program. Furthermore, Alumni (both from industry and academia) will be among the invited speakers at the Student Research Session organised yearly by the RELATE PhD candidates as part of the Annual RELATE week.

B.4.6 The Doctoral Candidate Agreement

The Doctoral Candidate Agreement will be signed between the candidate and the RELATE consortium after the candidate’s selection, defining as precisely as possible the rights and obligations of all parties. It will cover issues such as: (a) the participation costs charged to the candidate, what they do and do not cover; (b) the subject of the candidate’s research activities and the main structure (including period of candidacy), activities (including course requirements), and milestones (up to the awarding of the degree) of their doctoral programme; (c) the nature of the supervisory/monitoring/assessment procedures and the criteria used to assess the candidate’s performance; (d) the candidate’s obligations as regards the delivery of preliminary research results and the state of progress in their thesis preparation, as well as the consequences of not respecting these obligations. The general principles and requirements below foreseen by The European Charter for Researchers are relevant for PhD training and RELATE adheres to them completely.

Also, RELATE adheres to the requirements established in The Code of Good Conduct for the Recruitment of Researchers as described in Section 3. In particular, (i) “funding and salaries” makes sure that the PhD candidates enjoy fair and attractive conditions of salaries with adequate and social security provisions (see above); (ii) “intellectual property rights” are stated in the contract transparently; (iii) “supervision”: PhD projects are assigned already during the first year to a supervisory team, which is a subject to the Doctoral Candidate Agreement; (iv) “complaints/appeals”: the local coordinator at each partner university will act as the “student ombudsman” and be at the candidate’s disposal for mediating in and potentially resolving any possible conflicts between supervisors and PhD candidates, and guarantee confidential and informal assistance.

The employment contract to be concluded between the selected candidate and the host university will be drawn in accordance with the national regulations and regular practices of the host university. The project coordinator will ensure the consistency between the Doctoral Candidate Agreement and the respective employment contract.

B.5 Programme Management and Quality Assurance

B.5.1 Organisational arrangements and cooperation mechanisms

The cooperation mechanisms among the consortium partners of RELATE are regulated by a Consortial Contract that will be signed by the rectors/presidents of the five partner universities before the commencement of the program’s activities. This contract covers the financial and administrative aspects of the program. The academic and scientific aspects of RELATE are governed by the consortial Multilateral Agreement (see Annex “Multilateral Agreement”) which defines the common framework for the relevant doctoral degree regulations in force at each of the partner universities.
Each partner university allocates academic personnel for advising the doctoral students, research facilities and most of the lab equipment expenses for carrying out the research work through own financial resources or projects. Moreover, the consortium allocates personnel for managerial and administrative activities according to the general management structure shown in Figure 1.

As shown in Figure 1, the Central Management Office (CMO) headed and represented by the Project Coordinator (CO) represents the single point of contact between the European Commission and the consortium. The coordinator is responsible for the overall management of the project. In order to reduce the administrative and organisational burden for the CO who will be heavily involved in the academic and scientific coordination of the program activities, the CO will be supported by experienced UdS administrative staff who will aid with all administrative, financial, legal aspects of the project. The tasks of the Central Management Office (CMO) include in particular: (1) managing recruitment, including distributing information to the candidates, receiving and checking their applications, and circulating the applications to the partners; (2) assisting in the organisation of the annual RELATE meetings; (3) maintaining a central project database that contains all contractual documents, financial information, central project plan (tasks, description of work, etc.), annual reports, applicants’ information (application portfolio, study plan, information about internships, feedback from supervisors, annual report charting their progress); (4) recording outreach activities and joint academia-industry exchange events and ensuring that such events are organised regularly; (5) maintaining the website; (6) ensuring central dissemination of information to consortium partners (newsletter, etc.).

The Project Steering Committee (PSC) serves as the forum for discussion and decision making for the project consortium. Each partner will be represented in the PSC by two members - one who will represent the partner from a scientific and content-related perspective, and another one who will represent the partner from a financial, legal, and administrative perspective. As shown in Figure 1 above, while on a day-to-day basis the members of the PSC will be the central contacts and responsible persons for all project-related activities within their organisation (e.g., counseling of PhD candidates, exchange and planning with other partners, internal day-to-day financial monitoring), the CO as a whole is responsible for the co-ordination of the project from the strategic point of view.

The PSC is chaired by the CO and there will be at least one formal PSC meeting per year. The PSC as a whole is in charge of monitoring the program within the consortium, of financial monitoring within the consortium, and of quality assessment. Moreover, any changes in the overall program plan, partners’ resources, tasks and/or personnel are discussed and decided by the PSC by applying the principle of simple majority; in case of parity the CO has the casting vote. Decisions may be met both face-to-face during the formal meetings and remotely, e.g., through web conference. But in both cases, the rules for decision making as described remain in force and apply.

The Project Advisory Board has internationally well known experts from academia, research institutes, public bodies as well as large companies and SMEs. The major role of the Advisory Board is to give advice and guidance to the coordinator and the Project Steering Committee with regard to the programme and to ensure its quality from a broad perspective, taking into account the views and requirements of all different kinds of stakeholders. Where needed, additional expert advice may also be sought for particular problems from the most suitable AMs who confirmed their general support for this program.
Last but not least, at each partner **local coordinators** are responsible for the day-to-day implementation and QA of the doctoral program and its cooperation activities (e.g., counseling of PhD students, guidance with visa, living, insurance matters). Local coordinators in cooperation with student supervisors help students to realize their mobility scheme. According to B.1.3, the minimal scheme comprises stays at two universities and one AM, but students will be encouraged by their supervisors to gradually build research contacts with partner universities much beyond this scheme, in order to strengthen the cooperation network of the consortium. Implementation details on hosting students at partner universities are given in section B.4.2.

Finally, the International Office (IO) of each partner university also plays an important role in the support of the PhD candidates for mobility arrangements, like visa requirements, access to accommodation facilities, local administrative issues, etc.

An extensive **feed-back system** will be established to enable continuous and effective information flow among all partners and Associated Members. Specifically, important news and information of interest for all partners are collected centrally (see above, central database of the CMO) and distributed whenever necessary as newsletters to all partners. Day-to-day communication will be primarily carried out by email and, whenever necessary, web-conference or phone. General information (including also advertisement and/or announcement of events) will be published on the project website.

### B.5.2 Participation Costs

Estimated implementation costs are based on the actual budget for Master students in similar areas of the faculties split up into basic costs and laboratory costs. They are composed of salaries, equipment, rent, and depreciation. They do not take into account the cost of central services such as library and sports facilities, but they do take into account costs of running a laboratory for research and experiments on Language Technology. The laboratory component is essential to the program, since one of the central ideas is to familiarise the candidates thoroughly with the latest Language Technologies. This is essential to make them into experts for future global markets offering research, R&D and highly competitive LT products.

To estimate the implementation cost of RELATE, the additional requirements in these areas have been taken into account, based on 60 candidates (3 editions) in the programme at any time.

<table>
<thead>
<tr>
<th>Basic cost/candidate</th>
<th>Lab cost/candidate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EUR</strong></td>
<td><strong>EUR</strong></td>
</tr>
<tr>
<td>Staff</td>
<td>3,100</td>
</tr>
<tr>
<td>Equipment</td>
<td>300</td>
</tr>
<tr>
<td>Rent</td>
<td>250</td>
</tr>
<tr>
<td>Depreciation</td>
<td>20</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>3,670</strong></td>
</tr>
</tbody>
</table>

**Total estimated implementation cost RELATE**: 3,670 (basic) + 6,350 (lab) = **10,020 EURO**

**Participation fee** of 8,000 EURO yearly/candidate are calculated as follows:

**Basic cost share**: 3,600 EURO. This almost covers the RELATE basic cost. The additional RELATE costs include staff (1 assistant is added) and the summer and winter schools. These are financed by means acquired by the coordinating institution from other sources (DAAD, most notably).

**Lab cost share**: 4,400 EURO. In RELATE the doctoral candidates do all their research in labs and they need to use the lab equipment (machinery, but also the assistance of the technical people working in the labs), which is expensive. The test materials and the various kinds of software are expensive to obtain through licenses, many of which are expensive industrial ones. We charge laboratory-based participation fees to cover at least part of the costs. The laboratory fee will be spent exclusively on obtaining project data, buying equipment time if necessary, providing additional laboratory space for the candidates, and giving the candidates a laboratory buddy at each institution.

For those PhD candidates without an Erasmus Mundus grant, the participation costs in the research and training activities must be covered by themselves or by the research groups where these PhD candidates develop their research activities. Grants from Erasmus Mundus will be received by the coordinator and distributed among the partners which
compensate their own doctoral candidates according to the following scheme: number of candidates in primary supervision (2/3 of the participation costs/candidate) and secondary supervision (1/3 of the participation costs/candidate).

B.5.3 Development and sustainability plan as well as Complementary funding

Development and sustainability plan

The ultimate aim of the consortium is to create a sustainable network which will remain active beyond the actual funding period. As such, the project’s further development and sustainability is firmly based on the following two pillars: (1) the contributions and commitment of the consortium partners and (2) the contributions and commitment of the Associated Members (AMs).

Regarding the former, the sustainability of the program is formally based on the commitment of the consortium partners as expressed in both the Multilateral Agreement, as well as in the Consortial Contract. Through this formally agreed cooperation, the project will also contribute strongly to establishing direct links between researchers working in the partner universities. This kind of intangible results along with the existing joint projects that are already running between the partners (see section D of the application form) and their joint interest in the research areas of RELATE represent the basis and starting point for the cooperation among the partners after the end of the project.

Also, RELATE includes several inter-sectoral associated members (see Section 2.4) which will play a vital role in the program and its sustainability plan. First of all, these AMs will play an important role in the promotion of the program by directly supporting the dissemination and mobility activities (e.g., through secondments, dissemination events, etc.). Many of them have committed to providing internships for the PhD candidates as the RELATE areas of research are of major interest to them and PhD graduates of RELATE will constitute a unique source of experts needed by these AMs in the future.

In addition, the AMs have committed to supporting dissemination of the project results and activities, e.g., by establishing contacts with local players and industry and by using their standard dissemination channels. Last but not least, the AMs will also act as fund raisers for the program not only by directly providing complementary funding (see below, section B.5.3 for details), but also by promoting the program to potential external fund raisers (e.g., through direct contacts with representative offices of European and global international development agencies, relevant RTD initiatives and projects, etc.).

The extended sustainability plan will be developed by the CMO of UdS in cooperation with all partners. It will be updated regularly according to the existing needs of the network so as to prepare and ensure the sustainability of the program even after the funding period. The plan will be presented and discussed in the course of the annual RELATE meetings.

Overall, the RELATE network is expected to grow during the implementation period by establishing collaborations with even more institutions than what is now there in our already confirmed AMs list (especially, from the industry and private sector). Moreover, by involving PhD candidates financed by other funds such as those aforementioned from our industrial cooperations, national research foundations and international exchange services (e.g., DAAD, British Council, etc.) or scholarship schemes from the involved universities (see section B.5.3 below), the consortium expects to increase this number to 18-20 PhD candidates per year within the development of the course.

The main benefits for the academic partners include increased scientific excellence through high-level publications and new / strengthened partnerships with other academic research institutions. Moreover, the project will contribute strongly to the reinforcement of academia-industry relationships and cooperation (through mobility, joint research projects, etc.) and thus, support the technology transfer from academia to industry which is of utmost importance to both sides. Industrial partners (large enterprises, SMEs), on the other hand, benefit from high-quality students who have been trained to serve their requirements. Similarly, the career prospects of PhD candidates are significantly improved as mentioned earlier. Last but not least, through the increased collaboration among all partners and the relevance of the program to the needs of both academia and industry, the overall sustainability of the program will be ensured.

Complementary funding

As indicated in the section above, the use and acquisition of complementary sources of funding is an important feature of RELATE. Such funds may be used (1) to cover the difference between actual tuition fees and what is covered
by Erasmus Mundus, (2) to extend in exceptional and well motivated cases only the duration of the PhD for up to 4 years, and (3) to pay for doctoral students travel and/or publication costs that are not covered by the program or the partners’ own budget.

The basic complementary support will be provided by resources and funding that are already in use by the different partners:

- **UdS**: own resources, current projects (for instance, the DFG-funded “Cluster of Excellence – Multimodal Computing and Interaction”, the BMBF-funded “Softwareinnovationen für das digitale Unternehmen”, the EU-funded Marie-Curie ITN SCALE), German Academic Exchange Service (DAAD) for non-German candidates, European Funding (People Programme) for candidates from outside Germany, German Research Foundation (DFG) and Ministry of Education and Research (BMBF) funding, Google, local GradUS programme for individual support to PhD students, cooperation with industry, fellowships funded by the provincial government. The presidium of UdS provides financial support for structured PhD programs from own budget. That refers to and covers funding of training programs, coordination costs, and invitation of guest lecturers with such structured PhD programs. Finally, the presidium also offers administrative support for preparing and implementing parts of RELATE by “manpower” contributions from the aforementioned GradUS programme.

- **DCU**: the main sources of research funding in DCU are the Science Foundation Ireland (SFI), the Irish Research Council for Science, Engineering and Technology (IRCSET), Enterprise Ireland (IE), EU Project funding and direct Industry-funded research projects. DCU is the most successful Irish university in attracting external research funding per capita of university staff. In particular, in the area of language technologies DCU is currently the lead institution in the SFI and Industry partner funded Centre for Next Generation Localisation (CNGL) as well as being a partner or the lead in eight P7 EU projects (EuroMatrixPlus, META-NET, Panacea, Pluto, Cosyne, PetaMedia, Kreshmoi and AXES). DCU will dovetail these funding sources with the RELATE program to achieve optimal synergies.

- **UoE**: own resources (the Informatics Graduate School), funding from ILCC, studentships from the Scottish Informatics and Computer Science Alliance (SICSA), a research pool supported by the Scottish Funding Council, the Engineering and Physical Sciences Research Council (EPSRC) and the Economic and Social Research Council (ESRC); members of staff routinely obtain funds for students through the Google Faculty Research Awards program and the Microsoft Research Studentship program.

- **RuG**: own resources of the Humanities Graduate School, the Center of Language and Cognition Groningen, and various on-going research projects (for instance those funded by Taalunie (through the STEVIN programme), NWO and other national or international programs).

- **CUNI**: own resources of the Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics; several running projects, funded both by national organisations (such as Grant Agency of Czech Republic and Ministry of Education of the Czech Republic) and by EU (projects such as Faust, Metanet, Khresmoi, EUDAT, MosesCore).

All partners will support and assist the PhD students in obtaining and getting access to these complementary sources of funding. Furthermore, the AMs will directly provide additional complementary funding:

- Additional work contracts and/or scholarships for PhD students will be provided by the research centres / professional bodies cooperating with the consortium partner universities (namely DFKI, Instituut voor Nederlandse Lexicologie, etc.).

- IBM offers the mediation of job opportunities and industry placements in the IBM Thomas J. Watson Research Centre. Similarly, other AMs (Xerox, Microsoft, Symantec) have offered support and transfer of PhD candidates to local industry which provides further placement of students and dissemination activities.

- Additional funding for PhD candidates’ mobility will be provided by DFKI.

- The University of Rio Grande do Sul (UFRGS, Brazil) offers the provision of human resources, equipments and spaces necessary for the implementation of the project in addition to the hosting of PhD candidates in the various
LT research areas in which it specialises. The same holds for the University of Melbourne (Australia) and the Johns Hopkins University (USA).

B.5.4 Internal Evaluation and External Quality Assessment

RELATE implements a two-fold assessment strategy including both internal and external evaluation measures. Internal evaluation will be carried out on program level and on individual project level:

- The overall program evaluation will be based on the annual online Postgraduate Research Experience Survey (PRES) which will be extended by additional questions focusing on unique characteristics of RELATE. Overall, the evaluation questionnaires will be drawn up carried out at three different stages: (1) Upon arrival of the doctoral candidates to find out about their expectations; (2) at the end of each year of their study to identify status quo, strengths and weaknesses of the program, as well as suggestions for improvement; (3) one year after graduation to identify the impact of the program. All results will be managed and stored in the central database of RELATE by UdS.

- The progress and quality of individual research projects is undertaken through the Joint Assessment Procedure which includes standard progress checks (e.g., bi-annual progress reports), as well as selected qualitative and quantitative assessment criteria (e.g., number of publications, contribution to trainings, mastery of skills and methods, etc.) (see section B.3.5 for details). All outputs of the assessment procedure will be accumulated in the central database too, in order to facilitate sharing information about the progress across all partners.

The Joint Assessment Procedure will be managed by PSC. Local coordinators are responsible for collecting all needed materials from students and from their supervisors, as well as for delivering the materials to PSC via the database. Local supervisors are responsible for continuous monitoring of student progress. In case there is a research project with substantially sub-optimal progress, the supervisor and local coordinator consult PSC immediately, without waiting for the end of the regular evaluation period.

In addition to the standard local evaluation measures and to the RELATE Joint Assessment procedure, special attention will be paid to the quality of published scientific results accross all RELATE individual research projects. RELATE Best Paper Award for an outstanding journal article or conference paper written by one or more RELATE students will be announced yearly by PSC.

External evaluation will be carried out both by a central instance and locally:

- The Project External Advisory Board consists of independent members from industry and leading academic experts. Once a year, the board will provide an evaluation report which covers all aspects of the program (scientific, academic, administrative/organisational, financial). Where necessary (e.g., in sub-optimal areas), corresponding actions for improvement will be elaborated jointly and presented in the report.

- In addition, external evaluation will be carried out on local level by the respective university bodies in charge for quality assurance of teaching and research:

  1. At UdS, evaluation of doctoral research and procedures is carried out by GradUS; methodological support (e.g., design of academic evaluation surveys) is centrally provided by QUALIS.
  2. DCU: DCU runs a Quality Promotion Office (http://www.dcu.ie/qpu/) which carries out regular internal quality reviews of DCU units and programs in accordance with the Framework for Quality in Irish Universities regulations. Additional external reviews are carried out every 5 years.
  3. UoE: At UoE, like at all UK universities, research is subject to a national research assessment that aims to provide quality profiles of research activity across the higher education sector. The next assessment (the Research Evaluation Framework, or REF) is scheduled for 2014. The School of Informatics was the highest-ranking department in the UK in Computer Science and Informatics in the previous two research assessments, in 2001 and 2008.
  4. RuG: At the RUG, research is subject to periodical evaluation based on the Standard Evaluation Protocol of the KNAW (Royal Dutch Academy of Sciences).
  5. CUNI: The quality of research at CUNI is regularly evaluated by the Ministry of Education of the Czech Republic and by the Council for Research, Development and Innovations.