

1. Research Area: Human Language Technology

2. Acronym: HLT

3. The Research Area

Human Language Technology (HLT) is emerging as a crucial research area addressing the issue of how to manage large amounts of content that is nowadays available in digital form. The exponential growth of such content, on the one hand makes it necessary both methods and tools for its automatic processing on a large scale; on the other hand it calls for the realization of advanced and innovative services which allow an effective interaction, both person-person and person-machine, through which drastically improving the access to both multimedia and multilingual content.

Under this demand, in the last years the research community on HLT has expanded significantly, both for quantity and quality of its achievements. From the point of view of the methods, research is rapidly converging toward the use of empirical methods (i.e. based on data) whose progresses are constantly measured by means of objective and shared evaluation methodologies. The potential of HLT is recognized by research funding agencies, both in the USA and in Europe. Specifically, at the European level, the FP7 recognizes a prominent role to HLT for its function of increasing communication among the member states and of preserving the richness of their linguistic diversity, and thus their cultural identity.

The scientific challenge is made even more complex in the current web scenario, where we are fronted by the huge information overload, whose analysis requires to consider new forms of content publishing and authoring, and new form of diffusion, access and consume to information streams (e.g. through social networks).

On the applicative side, HLT covers a large range of services and applications to support tasks where the understanding of linguistic interaction is crucial (both based on written texts and on speech), event in contexts where language is not the only communication mean.

4. The research activities in Trento RISE

Within the Trento RISE context, thanks to the converging interests with other Research Areas (particularly Data and Knowledge Engineering, Multidimensional-Multimodal Signal Processing and Understanding, and Social Informatics), our goal is to make Trentino as an area of excellence as for multimedia and multilingual content processing technologies.

The *Piano Trentino della Ricerca* identifies some ICT research areas as strategic. Among them, the HLT area plays a relevant role, both for dimension, research achievements, international visibility, fund rising, social impact and capacity to produce innovation. We have identified four main areas of research and technology development within HLT: (i) large-scale information retrieval and content extraction from both text and speech; (ii) the investigation of semantic and/or cognitively plausible models for linguistic behaviors; (iii) the automatic transcription and translation of documents; (iv) the design and implementation of interactive social agents.

We envision a new generation of applications and services where HLT plays a crucial role. Such systems will be able to: interact with background knowledge and reason on it; aggregate contents from different sources; process continuous data streams, particularly in different languages; automatically translate from one language to another; monitor contents and opinions that are expressed in conversations, as in call centers; be aware of the context, like the time and the location in which the system operates, particularly in case of mobile devices; learn dynamically from experience, avoiding to repeat errors already done; exhibit human-like communication skills, like detecting the emotional impact of language and the use of common-sense knowledge; adapt themselves to new communication forms (e.g. blog, sms, tweeter, etc.), and to different application domains.

5. Relevant keywords

Human Language Technologies, Natural Language Processing, Speech Recognition, Interaction, Information Retrieval, Information Extraction, Machine Translation, Lexical Resources, Machine Learning, Dialogue, Semantics.

6. Industrial relations and innovation activities

The Research Area counts several companies on the territory which are involved in HLT:

- *Pervoice*: FBK spin-off, services for speech transcription. Collaboration on core speech technology as well as on the new Excitement and Eu-Bridge European projects.
- *Cogito*: Expert System spin-off, semantic-based information access. Collaborations for entity-based search within the European project Okkam (now the Okkam company).
- *Celct*: FBK spin-off, evaluation of language technologies. Collaboration on realization of benchmarks for several HLT tasks, as well as organization of evaluation campaigns.
- *Cross Library Services*: FBK and CELI spin-off, with an active involvement of UniTN personnel; multilingual digital libraries. Collaboration on a new industrial project on the use of HLT for high school educational material.
- *IM Service Lab*: LSI spin-off, design of speech and multimodal interaction systems.
- *GST*: FBK spin-off, speech technology, particularly in the health sector.
- *Open-Content*: A company addressing the market of editorial platforms. Collaboration for the integration of HLT into publishing platforms within the LiveMemories project.

Other companies on the territory, which compose the Semantic Valley consortium, have shown interest in language technologies for semantics. Overall, about 50 persons work in HLT trentino-based companies. In addition, the Skill (Semantics Knowledge Innovation Lab) research lab has been recently established as a collaboration between Telecom Italia and Trento RISE, with core interests in HLT.

7. Education in Trentino

There are several courses currently active in Trentino aiming at providing HLT competences:

- Master on Human Language Technologies and Interfaces (HLTI): since 2007-2008, co-organized by UniTN and FBK (responsible Giuseppe Riccardi), about 8 students.
- PhD DISI: about 8 students every year in HLT.
- PhD CIMEC: about 3 students every year in HLT.
- European Masters Program in Language and Communication Technologies (Erasmus Mundus, Bolzano-Trento); Responsible Raffaella Bernardi, Massimo Poesio and Nicu Sebe, about 6 students per year.
- Stream on Filosofia Logica e Informatica at the Facoltà di Lettere e Filosofia (triennale / magistrale), Trento, responsible Marco Baroni.
- Course on Speech and Audio Signal processing at DISI, responsible Maurizio Omologo.
- Stream on language, interfaces and communication at the Cognitive Science Faculty, including the Triennale on Interfaces and Communication Technologies and a Master on Cognitive Sciences.
- Course of Master Degree in Computer Science (Faculty of Mathematical, Physical, and Natural Science): Natural Language Processing and Information Retrieval, responsible Alessandro Moschitti.

All the above initiatives (except the Lauree Triennali) are in English.

Members of the Research Area are also involved in teaching activities at the Computer Science Faculty at Free University of Bozen/Bolzano on HLT courses since the start of the EMLCT program in 2004.

8. Staff

Affiliated “Faculty and Researchers”¹

#	Family Name	First Name	Institution	Position	email
1	Anderson	Andrew	UniTN	Post-Doc	aja.ucl@googlemail.com
2	Barbu	Eduard	UniTN	Post-Doc	eduard.barbu@unitn.it
3	Baroni	Marco	UniTN	Researcher	marco.baroni@unitn.it
4	Bentivogli	Luisa	FBK - CELCT	Researcher	bentivo@fbk.eu
5 *	Bernardi	Raffaella	UniTN	Researcher	bernardi@disi.unitn.it
6	Bertoldi	Nicola	FBK	Researcher	bertoldi@fbk.eu
7	Brugnara	Fabio	FBK	Senior Researcher	brugnara@fbk.eu
8	Cattoni	Roldano	FBK	Researcher	cattoni@fbk.eu
9	Cettolo	Mauro	FBK	Senior Researcher	cettolo@fbk.eu
10	Falavigna	Daniele	FBK	Senior Researcher	falavi@fbk.eu
11	Federico	Marcello	FBK	Research Director	federico@fbk.eu
12	Forner	Pamela	CELCT	Researcher	forner@celct.it
13	Giampiccolo	Danilo	CELCT	Researcher	giampiccolo@celct.it
14	Girardi	Christian	FBK	Researcher	cgirardi@fbk.eu
15	Giuliani	Diego	FBK	Researcher	giuliani@fbk.eu
16	Giuliano	Claudio	FBK	Researcher	giuliano@fbk.eu
17	Gretter	Roberto	FBK	Senior Researcher	gretter@fbk.eu
18	Ivanov	Alexei	UniTN	Post-Doc	ivanov@disi.unitn.it
19	Johansson	Richard	UniTN	Post-Doc	richardjdf.lth.se
20	Lavelli	Alberto	FBK	Senior Researcher	lavelli@fbk.eu
21	Maganti	Hari Krishna	FBK	Post-Doc	maganti@fbk.eu
22	Magnini	Bernardo	FBK	Research Director	magnini@fbk.eu
23	Matassoni	Marco	FBK	Researcher	matassoni@fbk.eu
24	Moschitti	Alessandro	UniTN	Researcher	moschitti@disi.unitn.it
25	Negri	Matteo	FBK	Researcher	negri@fbk.eu
26 **	Omologo	Maurizio	FBK	Research Director	omologo@fbk.eu
27	Pianta	Emanuele	FBK - CELCT	Senior Researcher	pianta@fbk.eu
28	Plank	Barbara	UniTN	Post-Doc	bplank@gmail.com
29	Poesio	Massimo	UniTN	Full Professor	massimo.poesio@unitn.it
30	Popescu	Octavian	FBK	Post-Doc	popescu@fbk.eu
31	Riccardi	Giuseppe	UniTN	Associate Professor	riccardi@unitn.it
32	Roberti	Pierluigi	UniTN	Post-Doc	roberti@disi.unitn.it
33	Ronchetti	Marco	UniTN	Associate Professor	marco.ronchetti@unitn.it
34	Speranza	Manuela	FBK	Researcher	manspera@fbk.eu
35	Stemle	Egon	UniTN	Post-Doc	egon.stemle@unitn.it
36	Strapparava	Carlo	FBK	Senior Researcher	strappa@fbk.eu
37	Tonelli	Sara	FBK	Post-Doc	satonelli@fbk.eu
38	Uryupina	Olga	UniTN	Post-Doc	olga.uryupina@unitn.it
39 *	Vieu	Laure	ISTC-CNR	Senior Researcher	vieu@irit.fr
40	Zamparelli	Roberto	UniTN	Researcher	roberto.zamparelli@unitn.it
41	Zanoli	Roberto	FBK	Researcher	zanoli@fbk.eu

Affiliated “Junior researchers”

	Family Name	First Name	Institution	Position	email

¹ For people who have double affiliation, the primary affiliation at the Research Area is indicated with one asterisk (*) while the secondary affiliation is indicated with two asterisks (**).

1	Abad	Azad	UniTN	PhD student	azad.abad@disi.unitn.it
2	Bartalesi-Lenzi	Valentina	CELCT	res. assistant	bartalesi@celct.it
3	Bayer	Ali Orkan	UniTN	PhD student	aliorkan.bayer@unitn.it
4	Bizazza	Arianna	FBK	PhD student	bisazza@fbk.eu
5	Brugnolli	Novella	UniTN	PhD student	novella.brugnolli@disi.unitn.it
6	Bruni	Elia	UniTN	PhD student	Elia.Bruni@unitn.it
7	Celli	Fabio	UniTN	PhD student	fabio.celli@unitn.it
8	Chowdhury	Faisal	FBK	PhD student	chowdhury@fbk.eu
9	De Souza	José	FBK	PhD student	desouza@fbk.eu
10	Firoj	Alam	UniTN	PhD student	firojalam@gmail.com
11	Fornaciari	Tommaso	UniTN	PhD student	tommaso.fornaciari@unitn.it
12	Ghosh	Arindam	UniTN	PhD student	arindam.ghosh@disi.unitn.it
13	Ghosh	Sucheta	UniTN	PhD student	ghosh@disi.unitn.it
14	Giordani	Alessandra	UniTN	PhD student	agiordani@disi.unitn.it
15	Lebani	Gianluca	FBK	PhD student	gianluca.lebani@unitn.it
16	Jaber	Suhel	FBK	PhD student	jabber@fbk.eu
17	Ju	Qi	UniTN	PhD student	qi.ju@disi.unitn.it
18	Le Dieu	Thu	UniTN	PhD student	dieu.thu@gmail.com
19	Marchetti	Alessandro	CELCT	res. assistant	amarchetti@celct.it
20	Prashant	Mathur	FBK	PhD student	mathur@fbk.eu
21	Mehdad	Yashar	FBK	PhD student	mehdad@fbk.eu
22	Mohammed	Abdul Waheed	FBK	PhD student	amwaheed@fbk.eu
23	Moretti	Giovanni	CELCT	res. assistant	moretti@celct.it
24	Ozbal	Gozde	FBK	PhD student	gozbalde@gmail.com
25	Palmero Aprosio	Alessio	FBK	PhD student	aprosio@fbk.eu
26	Rigo	Stephan	FBK	PhD student	rigo@fbk.eu
27	Ruiz	Nick	FBK	PhD student	nicruiz@fbk.eu
28	Sayadchi	Marie	UniTN	PhD student	marie.sayadchi@disi.unitn.it
29	Severyn	Aliaksei	UniTN	PhD student	severyn@disi.unitn.it
30	Sprugnoli	Rachele	CELCT	res. assistant	sprugnoli@celct.it
31	Stepanov	Eugeny	UniTN	PhD student	stepanov@disi.unitn.it
32	Vecchi	Eva Maria	UniTN	PhD student	evamaria.vecchi-1@unitn.it

9. Laboratories

Research in HLT in Trentino is currently carried on in four distinct laboratories: CLIC, affiliated at CIMEC (UNITN), in Rovereto; LSI, affiliated at DISI (UNITN), in Povo; HLT, affiliated at the Information Technology Centre at FBK (Povo); SHINE, affiliated at the Information Technology Centre at FBK (Povo). Related research activities are carried on at LOA (Laboratory of Applied Ontology, affiliated at ISTC-CNR, in Povo) and at CELCT (Centre for the Evaluation of the Language and Communication Technologies, in Povo).

- UNITN, CLIC, Language, Interaction and Computation Laboratory, coordinators Massimo Poesio and Marco Baroni.
Main research topics: Formal and empirical approaches to natural language semantics; linguistics, computational and cognitive aspects of concepts.
- UNITN, LSI, Language Speech and Interaction, coordinator Giuseppe Riccardi.
Main research topics: interactive and social agents, machine learning, language processing and understanding, both spoken and written, information retrieval, software architectures, multimodality.
- FBK-HLT, Research Unit on Human Language Technologies, coordinators Marcello Federico and Bernardo Magnini.
Main research topics: machine translation, speech recognition, content extraction, development of linguistic resources.
- FBK-SHINE, Research Unit on Speech-Acoustic Scene Analysis and Interpretation, coordinator Maurizio Omologo.

Main research topics: Speech enhancement, robust speech recognition, multi-microphone distant-speech interaction and speaker diarization.

The groups CLIC and LSI (UNITN) currently count 31 people, including 16 PhD students, the groups HLT and SHINE (FBK) count 35 people, including 12 PhD students, CELCT counts 6 people with 4 research assistants, and ISTC-CNR counts one person. In total there are 73 people affiliated at the Research Area, out of which 71 are full time on HLT.

10. National and international relations

The HLT research groups in Trentino are among the more active at the international level, as they are achieving outstanding results and contributing to the advance of the sector. Collaborations with EIT partners include:

- DFKI (Berlin): collaborations on question answering, textual entailment and machine translation.
- LIMSI (Paris) affiliated at University of Paris XI: collaboration on evaluation of HLT tools.

Among the others, collaborations include:

- IBM: collaboration on question answering on the Watson Challenge.
- Bar Ilan University and NICE Corp. (Israel): Textual inference and their applications.
- University of Sheffield: Information Extraction.
- University of Karlsruhe (Germany): Interactive Systems.
- University of Edinburgh (UK): Machine Translation platform Moses.
- HLT group at the Columbia University of New York.
- HLT group at the University of Colorado at Boulder.
- Etc.

11. Research Area organization

- chair: Bernardo Magnini
- vice-chair: Giuseppe Riccardi
- steering committee: Luisa Bentivogli, Raffaella Bernardi, Mauro Cettolo, Marcello Federico, Bernardo Magnini, Alessandro Moschitti, Massimo Poesio, Giuseppe Riccardi.

12. Motivations for a joint initiative

Although the HLT area in Trentino has grown both as dimension and impact, we believe that a better coordination within the Trento RISE context could further exploit its potential. We think it is time for both a rationalization of the logistics and for the implementation of concrete actions to improve cooperation among the HLT research groups in Trentino, for the following reasons.

- Better opportunities for the HLT research in Trentino. It will be beneficial if the different groups on HLT present themselves as a single HLT Research Area, both at the local level (PAT, Trento RISE), national level (MIUR) and international level (EU, EIT), rather than different institutions and departments.
- Better exploitation of the existent HLT resources. Coordinating the activities would allow scale economy, avoid overlaps in research, and the optimization of the overall HLT offer.
- Better education initiatives on HLT. Currently there are a number of independent courses offered in different faculties and departments, while we are missing a shared framework. It would be preferable to revise the current HLT education program considering more integration within Trento RISE.
- Better opportunities for exchange and collaboration. Currently, the distance between the locations of the groups causes inconveniences and a waste of time, which might be overcome with a single location in Povo.

13. Activity planning

September 2011

- Set up a coordination of the activities, with representatives of the HLT research laboratories. The initial goal is to favor the exchange of information about relevant initiatives (e.g. funding proposals, technological transfer, education activities, agreement with PAT).

October 2011

- Organize a public presentation of HLT at Trento RISE.
- Dissemination of the new area at national and international level.

November 2011

- Double affiliations for HLT personnel.
- Set up a web portal for HLT at Trento RISE, with pointers to research activities, seminars, innovation, education opportunities.

December 2011

- Solutions for logistic issues.
- Getting started: seminars on research activities.
- Research plan and plan for new personnel of the HLT Research Area.