



International Directory of Language Technologies

Presented to the
Language Technologies Research Centre

Prepared by Infolog Inc.
March 2006

TABLE OF CONTENTS

FOREWORD	8
ALLEMAGNE	9
LABORATOIRES	9
<i>Applied Informatics / Cooperative Systems, Technical University of Munich</i>	9
<i>Artificial Intelligence Research Koblenz</i>	10
<i>Chair of Computer Science VI: Human Language Technology and Pattern Recognition</i>	11
<i>Competence Center Semantic Web</i>	12
<i>Computational Linguistics Lab, University of Potsdam</i>	13
<i>Computational Linguistics, University of Erlangen-Nuremberg</i>	14
<i>Computational Psycholinguistics Group, University of Saarlandes</i>	16
<i>Experimental Phonetics Group at the Institute of Language Processing, University of Stuttgart</i>	17
<i>German Research Center for Artificial Intelligence</i>	18
<i>Institute of Communication Acoustics</i>	19
<i>Institute of Human-Machine Communication</i>	20
<i>Intelligent Information and Communication Systems</i>	21
<i>Jenna University Language and Information Engineering Lab</i>	22
<i>Language Technology Lab, DFKI</i>	23
<i>Linguistics Department, University of Konstanz</i>	24
<i>Manfred Pinkal Computational Linguistics Group</i>	25
<i>Phonetics Institute, University of Saarlandes</i>	26
ORGANISMES ET ASSOCIATIONS	27
<i>Bavarian Archive for Speech Signals</i>	27
AUSTRALIE	28
LABORATOIRES	28
<i>Centre for Language Technology, Macquarie University</i>	28
<i>Department of Computer Science and Software Engineering, University of Melbourne</i>	29
<i>Knowledge Management Research Group, University of Sydney</i>	31
<i>Speech Hearing and Language Research Centre, Macquarie University</i>	32
<i>Sydney Language Technology Research Group, University of Sydney</i>	33
ORGANISMES ET ASSOCIATIONS	34
<i>Australian Computer Science Society</i>	34
<i>Australian Speech Science and Technology Association</i>	35
<i>KnowledgeNet (ARC Research Network for Next Generation Web Technologies)</i>	36
<i>National Information and Communication Technology Association of Australia</i>	37
<i>Smart Internet Technology CRC</i>	38
BELGIQUE	39
LABORATOIRES	39
<i>Center for Computational Linguistics, Université Catholique de Louvain</i>	39
<i>Center for Processing Speech and Images (PSI)</i>	41
<i>Department of Electronics and Information Processing, Vrije Universiteit Brussels</i>	42
<i>Digital Speech and Signal Processing, Ghent University</i>	43
<i>Théorie des Circuits et Traitement du Signal, Faculté Polytechnique de Mons</i>	44
ORGANISMES ET ASSOCIATIONS	45
<i>Belgian Acoustical Society (ABAV)</i>	45
CHINE ET TAIWAN	46

LABORATOIRES	46
<i>Institute of Information Science, Academia Sinica</i>	<i>46</i>
<i>National Taiwan University , Department of Computer Science and Information Engineering</i>	<i>47</i>
<i>National Tsing Hua University, National Tsing Hua University.....</i>	<i>48</i>
<i>Spoken Language Group, Aacademia Sinica</i>	<i>49</i>
CORÉE	50
LABORATOIRES	50
<i>Postech Intelligent Software Lab.....</i>	<i>50</i>
DANEMARK	51
LABORATOIRES	51
<i>Center for Language Technology, Faculty of Humanities, University of Copenhagen.....</i>	<i>51</i>
<i>Department of Computational Linguistics, Copenhagen Business School.....</i>	<i>52</i>
<i>Department of Nordic Studies and Linguistics, University of Copenhagen</i>	<i>53</i>
ESPAGNE	54
LABORATOIRES	54
<i>Artificial Intelligence section, Software Dept., Technical University of Catalonia.....</i>	<i>54</i>
<i>Compilers and Languages, Escuela Superior de Ingeniería Informática, Orense</i>	<i>55</i>
<i>Database Lab, University of A Coruña.....</i>	<i>56</i>
<i>Department of Artificial Intelligence, Technical University of Madrid</i>	<i>57</i>
<i>IXA Group.....</i>	<i>58</i>
<i>Laboratory for Research and Development in Artificial Intelligence.....</i>	<i>59</i>
<i>Natural Language Processing and Information Retrieval Group at National Distance Learning University (UNED).....</i>	<i>60</i>
<i>Natural Language processing Group, Technical university of Catalonia.....</i>	<i>61</i>
<i>Speech Processing Group (VEU), Technical University of Catalonia.....</i>	<i>62</i>
<i>The Research Group on Signals, Telematic and Communications, University of Granada.....</i>	<i>63</i>
ORGANISMES ET ASSOCIATIONS	64
<i>Catalan Association of Artificial Intelligence</i>	<i>64</i>
ÉTATS-UNIS	65
LABORATOIRES	65
<i>Brain and cognitive sciences, Rochester University.....</i>	<i>65</i>
<i>Cambridge research Lab</i>	<i>66</i>
<i>Center for advanced information processing.....</i>	<i>67</i>
<i>Center for automated learning and discovery.....</i>	<i>68</i>
<i>Human Language Technology Research institute</i>	<i>69</i>
<i>Center for Human-Computer Communication.....</i>	<i>70</i>
<i>Center for intelligent information retrieval.....</i>	<i>71</i>
<i>Center for language and speech processing.....</i>	<i>72</i>
<i>Center for machine translation</i>	<i>74</i>
<i>Center for spoken language research.....</i>	<i>75</i>
<i>Center for Study of Language and Information.....</i>	<i>76</i>
<i>Center of Excellence for Document Analysis and Recognition</i>	<i>77</i>
<i>Computational Linguistics and Information Processing Laboratory.....</i>	<i>78</i>
<i>Computational Linguistics at Stanford University.....</i>	<i>79</i>
<i>Columbia Natural Language Processing Group.....</i>	<i>80</i>
<i>Computing research laboratory</i>	<i>81</i>
<i>Department of Cognitive and Linguistic Sciences, Brown University</i>	<i>82</i>
<i>Haskins Laboratories.....</i>	<i>83</i>
<i>Human-Computer Interaction Institute.....</i>	<i>84</i>

<i>USC Natural Language Group</i>	85
<i>International Computer Science institute, Speech Group</i>	86
<i>Linguistic Grammars Online</i>	87
<i>Maryland Information and Network Dynamics Lab (MIND)</i>	88
<i>MIT Media Laboratory</i>	89
<i>Research Cluster on Language, Center for Cognitive Science, Rutgers State University</i>	90
<i>Robust Speech Processing Group Center for Spoken Language Research</i>	91
<i>The Signal, Speech and Language Interpretation Lab at the University of Washington</i>	92
<i>Speech Technology and Research Laboratory</i>	94
<i>Spoken Language Systems</i>	95
ORGANISMES DE FINANCEMENT ET ASSOCIATIONS	96
<i>American association of artificial intelligence</i>	96
<i>Applied Voice Input/Output Society (AVIOS)</i>	97
<i>Association for machine translation of America</i>	98
<i>IEEE Signal Processing Society</i>	99
<i>Linguistic data Consortium</i>	100
<i>Retrieval Group</i>	101
<i>Open Mind Initiative</i>	102
FINLANDE	103
LABORATOIRES	103
<i>From data to Knowledge research unit</i>	103
<i>Helsinki Institute for Information Technology, Basic Research Unit</i>	104
<i>Laboratory of Acoustics and Audio Signal Processing</i>	105
<i>Multimodal Interfaces Group, Adaptive Informatics Research center</i>	106
<i>Semantic Computing Research Group</i>	107
FRANCE	108
LABORATORIES	108
<i>Analyse et traitement informatique de la langue française (ATILF)</i>	108
<i>(Computerized Analysis and Processing of French)</i>	108
<i>École nationale supérieure de télécommunications, Groupe Information, Interaction, Intelligence</i>	109
<i>Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur, Human-Machine Communication Department</i>	110
<i>Laboratoire des sciences de l'information et des systèmes, INCOD</i>	111
<i>(Information and Systems Science Laboratories)</i>	111
<i>Laboratoire d'Automatique Documentaire et Linguistique</i>	112
<i>Automated Documents and Linguistics Laboratory</i>	112
<i>Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur – Computer Sciences Laboratory for Mechanics and Engineering Sciences</i>	113
<i>Laboratoire de Linguistique Formelle</i>	114
<i>(Formal Linguistics Laboratory)</i>	114
<i>LORIA Laboratoire lorrain de recherche en informatique et ses applications (The Lorraine Laboratory of IT Research and its Applications)</i>	115
<i>Laboratoire informatique d'Avignon</i>	116
<i>Avignon IT Laboratory</i>	116
<i>Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen</i>	117
<i>(Caen IT, Image, Machine and Instrument Research Group)</i>	117
<i>Groupe de traitement du langage parlé, Laboratoire d'Informatique et de Mécanique pour les Sciences de l'Ingénieur</i>	118
<i>(Spoken Language Processing Group, IT and Mechanics Laboratory for Engineering Sciences)</i>	118
FRENCH ORGANIZATIONS AND ASSOCIATIONS	119
<i>Association francophone de la communication parlée</i>	119
<i>(The Francophone Association of Spoken Communications)</i>	119

<i>Association pour la Recherche Cognitive</i>	120
<i>(The Cognitive Research Association)</i>	120
<i>Association pour le traitement automatique des langues</i>	121
<i>(The Association for the Automatic Processing of Languages)</i>	121
<i>European Language Resources Association (ELRA)</i>	122
<i>Association Internationale de la communication parlée</i>	123
<i>(International Speech Communication Association)</i>	123
<i>Réseau Francophone de l'Ingénierie de la Langue (FRANCIL)</i>	124
<i>(Francophone Network of Language Engineering)</i>	124
<i>Technolangue Portal</i>	126
HONG KONG	127
LABORATOIRES	127
<i>Holiday Centre for Intelligent Applications of Language Studies</i>	127
INDE	128
LABORATOIRES	128
<i>AU-KBC Research Center, Madras Institute of Technology</i>	128
<i>Center for Development of Advanced Computing</i>	129
<i>Computer Vision and Pattern Recognition Unit, Indian Statistical Institute of Calcutta</i>	130
IRLANDE	131
LABORATOIRES	131
<i>Digital Signal Processing, University College Dublin</i>	131
<i>Information Management Group, Dublin College University</i>	132
<i>Knowledge and Data Engineering Group, Trinity College Dublin</i>	133
<i>Language and Intelligence Group, Dublin City University</i>	134
<i>National Center for Language Technologies</i>	135
<i>TCD Computational Linguistics Group</i>	136
ISRAËL	137
LABORATOIRES	137
<i>Department of Computer Science, Bar-Ilan University</i>	137
<i>Department of Computer Science, University of Haifa</i>	138
<i>Israël Institute of Technology, Computational Linguistics Group</i>	139
<i>Knowledge Center for Processing Hebrew</i>	140
ITALIE	141
LABORATOIRES	141
<i>Center for Scientific and Technological Research, Cognitive and Communication Technologies Division</i>	141
<i>Cognitive and Communication Technologies Division, ITC-Irst</i>	142
<i>Department of computer and systems science, University of Rome</i>	143
<i>Department of Information Engineering, University of Padova</i>	145
<i>European Academy Bolzano, Institute for Specialised Communication and Multilingualism</i>	145
<i>Institute of Cognitive Sciences and Technologies</i>	146
<i>Istituto di Elaborazione della Informazione</i>	147
<i>Interactive Sensory System Division, SSI</i>	148
<i>ISTI, Knowledge Discovery and Delivery Laboratory</i>	149
<i>Istituto di Linguistica Computazionale del CNR</i>	150
<i>Knowledge Representation Research Center, Free University of Bozen-Bolzano</i>	151
<i>Laboratory fo Applied Ontology</i>	152
ORGANISMES ET ASSOCIATIONS	153
<i>Center for the Evaluation of Language and Communication Technologies</i>	153

<i>FoLLI, Association of Logic, Language and Information</i>	154
<i>Tuscan Word Centre</i>	155
JAPON	156
LABORATOIRES	156
<i>Advanced Telecommunications Research Institute International ATR</i>	156
<i>Computational Linguistics Laboratory</i>	157
<i>Itaku Laboratory</i>	158
<i>Kasuya Lab (Klab)</i>	159
<i>Language Media Laboratory</i>	160
<i>National Institute of Advanced Industrial Science and Technology (AIST)</i>	161
<i>Shikano Lab</i>	162
<i>Tokyo Institute of Technology</i>	163
ORGANISATIONS ET ASSOCIATIONS	164
<i>Asian-Pacific Association for Machine Translation</i>	164
<i>Information Processing Society of Japan</i>	165
PAYS BAS	166
LABORATOIRES	166
<i>Computational Linguistics and AI Group, Tilburg University</i>	166
<i>Mathematics and Computer Science? Faculty, University of Twente</i>	167
<i>Institute for Logic, Language and Computation</i>	168
<i>Dutch Research School for Information and Knowledge Systems</i>	169
<i>Human Computer Studies Laboratory, University of Amsterdam</i>	170
<i>IRIS Information Retrieval and Information, Universtiy of Nijmegen</i>	171
<i>The Leiden University Centre for Linguistics</i>	172
<i>Utrecht Institute of Linguistics OTS</i>	173
ORGANIMES ET ASSOCIATIONS	174
<i>Acoustical Society of the Netherlands</i>	174
<i>European Network of Excellence in Human Language Technologies</i>	175
ROYAUME UNI	176
LABORATOIRES	176
<i>Center for communication Interface Research</i>	176
<i>The Center for Speech Technology Research</i>	177
<i>Center for vision, Speech and Signal Processing</i>	178
<i>Natural Language and Information Processing Group</i>	179
<i>Natural Language Processing and AI</i>	180
<i>Natural Language Processing Research Group</i>	181
<i>Institute for Communication and Collaborative Systems</i>	182
<i>School of Computer Science, University of Manchester, IMG, NLE</i>	183
<i>Scool of computing, Robert Gordon University</i>	184
<i>Computing Science at Aberdeen</i>	185
<i>Language and computation group</i>	186
<i>National Center for Text Mining</i>	187
<i>Natural Language and Computational Linguistics</i>	188
<i>Oxford Computational Linguistics Group</i>	189
<i>UCL Department of Phonetics and Linguistics</i>	190
<i>University Center for Computer Corpus Research on Language</i>	191
SUÈDE	192
LABORATOIRES	192
<i>Department of Linguistics, Lund University</i>	192

<i>Department of Linguistics, Stockholm University</i>	193
<i>Graduate School of Language Technology</i>	194
<i>Human Computer Interaction and Language Engineering Laboratory (HUMBLE), Swedish Institute of Computer Science</i>	195
<i>Interaction and Presentation Laboratory</i>	196
<i>Natural Language Processing Laboratory, Linköping University</i>	197
<i>Speech and Communications Technology Group</i>	198
SUISSE	199
LABORATOIRES	199
<i>Artificial intelligence Laboratory, Federal Institute of Technology in Lausanne</i>	199
<i>Department of Linguistics, University of Geneva, Language Technology Laboratory</i>	201
<i>Institute of Computational Linguistics, University of Zurich</i>	202
ORGANISMES ET ASSOCIATIONS	203
<i>European Association for Machine Translation</i>	203
<i>European Association for Speech, Signal and Image Processing</i>	204

Foreword

This document is the result of a research mandate granted by the Language Technologies Research Centre (LTRC) to Infolog Inc., a research and communications company. It is a directory indexing the major university laboratories abroad that are performing research and development in the language technologies industry. Moreover, it comprises the major funding institutions and organizations pertaining to this field.

This research work has been limited to a certain number of countries, in order to circumscribe the contents of this document to countries with the largest research pools in language technologies. The majority of institutions that were brought together are found in North America and Western Europe. Included in this list are institutions from the United States, the United Kingdom, France, Italy, Germany, Spain, Ireland, Belgium, Holland as well as the Scandinavian Peninsula. Laboratories in Australia, Israel and Asia (China, Japan, Korea, India and Hong Kong) are also described in this document. It is however important to underscore the fact that only institutions with English or French information texts were considered for this directory. This constraint had only a minor influence on the data gathered, as most activities of the language technologies field occur in English, regardless of the official language of the host country.

This directory details language industry research occurring in these diverse regions. The publication of this information aims to foster greater cooperation between various sector stakeholders and to allow Canadian organizations to eventually establish contacts with their colleagues abroad. Ultimately, this tool will simplify actions taken by the LTRC, the Language Industry Association of Canada (AILIA), the National Research Council of Canada (NRC), as well as government participants who support language industry initiatives.

The directory is structured in such a way as to separate the data profiles of every country studied into nineteen (19) separate sections. Each section is also subdivided into two parts. The research laboratories list is featured first, with those of organizations and associations following.

All these institutions are described in this directory with individual profiles. These profiles include all useful contact information (electronic as well), the names of individuals to contact, as well as a brief description of research projects undertaken, resulting products or, in the case of organisms, their mission and scope.

This directory was produced in several phases which occurred between January and March of 2006. Language industry stakeholders herein showcased were compiled from lists provided by the various national language technologies organizations of countries involved. The European Union Portal was especially useful to the research team, exhibiting an excellent degree of cohesion within Europe. The criteria considered to determine whether a laboratory was large enough to be included within the directory were laboratories comprising at least 20 (twenty) researchers. The contents of organizational data profiles included were drafted from information gathered on their Webpages.

Good reading.

Allemagne

Laboratoires

Applied Informatics / Cooperative Systems, Technical University of Munich
TU München, Institut für Informatik, Boltzmannstr. 3, D-85748 Garching bei München Tel: +49--89--289--18656 Fax: +49--89--289--18657 Website: http://www11.in.tum.de/index.html.en Contact: Evelyn Genkow Email: gemkow@informatik.tu-muenchen.de Number of positions: 26
Research projects
The goal of the area Knowledge Management is to acquire, use and preserve knowledge in organizations and communities. We thereby apply modern technologies and concepts such as the Semantic Web and ontologies. Our chair organises the Münchner Knowledge Management Kreis since 2000, which promotes information exchange between industry and university.

Artificial Intelligence Research Koblenz

Universität Koblenz-Landau, Campus Koblenz Institut für Informatik AG Künstliche Intelligenz
Universitätsstr. 1 D-56070 Koblenz Germany

Tel: +49(0)261/287-2772 Fax: +49(0)261/287-2731

Website: <http://www.uni-koblenz.de/FB4/Institutes/IFI/AGKI>

Contact: Christoph Wernhard

Email: wernhard@uni-koblenz.de

Research projects

The working group Artificial Intelligence belongs to the most renowned research groups in Germany in the fields of automated deduction, formal verification, logic programming, and autonomous agents. The main topic of research is the development of intelligent software-systems. Ultimately, machines should be able to work autonomously in complex environments in order to effectively support users in their decisions and actions. Sample applications are finding relevant information in a large amount of data, but also controlling intelligent robots in natural and simulated environments.

Today model computation in propositional logic is used for complex tasks in planning, verification and diagnosis. For wider application fields, extensions of language expressivity are necessary: for example full first order predicate logic, non-monotonic methods, preferred models, supported models, answer set programming and description logics. Within the project, application relevant aspects of such techniques are investigated and implemented efficiently. Practical adequacy of the approach is demonstrated with a case study from the field of document management.

Chair of Computer Science VI: Human Language Technology and Pattern Recognition

Chair of Computer Science VI, (Lehrstuhl für Informatik VI), Professor Dr.-Ing. H. Ney,
Ahornstr. 55 , D-52056 Aachen, Germany

Tel: +49 (241) 80-21600 Fax: +49 (241) 80-22219

Website: <http://www-i6.informatik.rwth-aachen.de/>

Contact: Hermann Ney

Email: ney@informatik.rwth-aachen.de

Number of positions: 20

Research projects

TC-STAR Technology and Corpora for Speech to Speech Translation? TC-STAR is a concentrated six year effort for advanced research in all core technologies for speech to speech translation: speech recognition, translation, and synthesis. The project will target a selection of unconstrained conversational speech domains i.e. broadcast news and speeches and a few languages relevant for Europe's economy and society: Chinese, European English and European Spanish. The technical challenges and objectives of the project will focus on the development of new algorithms and methods, integrating relevant human knowledge which is available at translation time into a data-driven framework. Examples of such new approaches are the integration of linguistic knowledge in the statistical approach of spoken language translation, the statistical modelling of pronunciation of unconstrained conversational speech in automatic speech recognition, and new acoustic and prosodic models for generating expressive speech in speech synthesis.? TC-STAR is supported by the *European Union*.

Competence Center Semantic Web

DFKI GmbH, Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany

Tel: +49-681-302-5325, Fax: +49-681-302-5338

Website: <http://www3.dfki.uni-kl.de/ccsw/>

Contact: Paul Buitelaar

Email: paulb@dfki.de

Number of positions:

Research projects

- KSpace Content-based Multimedia Analysis;
- SmartWeb Mobile Broadband Access to the Semantic Web;
- NEWS News Intelligence Technology for the Semantic Web;
- Direct-Info Media Monitoring and Multimodal Analysis;
- VleWs Connecting Information and Knowledge in saarland.de;
- EPOS Evolving Personal to Organizational Knowledge Spaces;
- INKASS Ontology Engineering and Lifecycle Mechanisms for Electronic Knowledge Trading Marketplaces.
- FRODO Framework for Distributed Organizational Memories;
- Clockwork Web-based knowledge management.

Computational Linguistics Lab, University of Potsdam

Computational Linguistics, Department of Linguistics, University of Potsdam, Campus Golm / Haus 24, Karl-Liebknecht-Str. 24-25, D-14476 Potsdam-Golm, Germany

Tel: ++49 0331 / 977-2016 Fax: ++49 0331 / 977-2761

Website: <http://www.ling.uni-potsdam.de/cl/cl/>

Contact: Peter Staudacher

Email: staudach@ling.uni-potsdam.de

Number of positions: 69

Research projects

Research and Development at the Computational Linguistics Group in Potsdam

Linguistic Database for Information Structure Project D1 provides the technical infrastructure for building, maintaining and retrieving the linguistic data collected by the SFB 603 "Information Structure". Besides providing suitable software and hardware, D1 supports the individual projects in their task of data annotation.

Parsing of Pathology Reports The project develops a prototypical system to process and retrieve pathologic image descriptions and diagnostic findings, especially in pulmonary diseases. http://www.ling.uni-potsdam.de/cl/cl/res/forsch_lunge.html/SiCoPS/lang_en/SiCoPS.html

Automatic Summarisation The project develops a system for the automatic summarisation of documents (both single documents and collections). http://www.ling.uni-potsdam.de/cl/cl/res/forsch_summar.html/SiCoPS/lang_en/SiCoPS.html

Dealing with Uncertainty The project is developing methods for dealing with uncertainty about user input in spoken dialogue systems.

Digital Lexicon In cooperation with the project "Digital Lexicon of Current German" at the Berlin-Brandenburg Academy of Sciences and Humanities, we develop a highly-detailed morpheme lexicon of German.

Rhetorical Structure Analysis Rhetorical structures describe inter-clausal phenomena within a text that are driven by semantic and some syntactic properties. We analyse them automatically.

URML - An Underspecified Markup Syntax for Rhetorical Structure Annotations As a step toward flexible, extendible annotation schemes, we propose an XML format for annotating underspecified rhetorical structure trees. We discuss the various design decisions involved, illustrate the format with an example, and sketch some applications.

Discourse Marker Lexicon DiMLex Discourse markers are conjunctions and other functional words, that indicate the type of correlation between adjacent text spans, e.g. because --> causal, despite --> concessive. Our lexicon of German discourse markers provides information about syntax and semantics via XML/XSLT and is used for automatic text analysis and text

generation.

Computational Linguistics, University of Erlangen-Nuremberg

Friedrich-Alexander-Universität Erlangen-Nürnberg, Abteilung für Computerlinguistik,
Bismarckstr. 6 und Bismarckstr. 12 (Rechnerraum 0.320), 91054 Erlangen

Tel: (+49 09131) 85-29250 Fax: (+49 09131) 85-29251

Website: http://www.linguistik.uni-erlangen.de/en_contents/index.php

Contact: Jörg Kapfer

Email: jmkapfer@linguistik.uni-erlangen.de

Number of positions: 116

Research projects

- JSLIM (Java Surface Linear Internal Matching): Interpreter for left-associative grammars implementing Database Semantics
- Textmill
- GLDV Study Group on Parsing in Morphology, Syntax and Semantic
- An interactive Malaga-Tutorial (Elke Ertel)
- Malaga, a left-associative grammar system, implemented by Björn Beutel
- Cooperation project with Korea
- Grammatical Induction of German Syntax (Markus Schulze)
- Semantic Analysis of German text in the framework of Leftassociative Grammar (Manfred Hupfer)
- Database-supported storage of corpora (Thomas Künneth)
- Left-associative morphology of the Korean language in Malaga (Prof. Dr. Kiyong Lee, Ph.D.)
- LAPT&DA: Dictionary system with automatic partitioning in domain specific dictionary parts and its dynamic activation while tagging unknown texts

Computational Psycholinguistics Group, University of Saarlandes

FR 4.7 Psycholinguistik, Universität des Saarlandes, Gebäude 17.1, Room 1.15, D-66041 Saarbrücken, Germany

Tel: + 49 (681) 302 6555 Fax: +49 (681) 302 6561

Website: <http://www.coli.uni-saarland.de/groups/MC/>

Contact: Matthew W. Crocker

Email: crocker@coli.uni-sb.de

Number of positions: 20

Research projects

- ALPHA: Adaptive Mechanisms for Human Language Processing
- Parallelism in Human Parsing
- Assoziationsnormen für Wörter und Bilder im Fernsehen
- AMLaP-2001 Conference: Architectures and Mechanisms for Language Processing
- European Graduate College in Language Technology and Cognitive Systems

Experimental Phonetics Group at the Institute of Language Processing, University of Stuttgart

IMS -Experimentelle Phonetik- Azenbergstrasse 12 D-70174 Stuttgart, Germany

Tel: +49 (0)711 121-1379 Fax: +49 (0)711 121-1366

Website: <http://www.ims.uni-stuttgart.de/phonetik/>

Contact: Sabine Dieterle

Email: Sabine.Dieterle@ims.uni-stuttgart.de

Number of positions: 20

Research projects

SmartWeb *Speech synthesis in a system enabling mobile access to the Semantic Web (*

Prosody Acquisition *First language acquisition of acoustic correlates of word stress*

Prosody Production *A Computational Model of Target Oriented Production of Prosody*

SmartKom *Speech output component for a multimodal system*

Schwerpunktprogramm Sprachproduktion *The Neuroanatomic Foundation of Language Production: Visual Examination Using Functional Magnet Resonance Imaging (fMRI)*

Speech Synthesis *Modeling of the linguistic context for speech synthesis*

Alignment Project *Segmentation of Speech Data from the Mannheim Institute of German Language (IDS)*

German Research Center for Artificial Intelligence

DFKI GmbH Stuhlsatzenhausweg 3 (Building D 3 2) D-66123 Saarbrücken

Tel: +49 (0)681 302 5151 Fax: +49 (0)681 302 5341

Website: <http://www.dfki.de/web/>

Contact: Brigitte Selzer

Email: Brigitte.selzer@dfki.de

Number of positions: 38

Research projects

Gnowsis

Duration: 2002 - (2005) **Short description:** Gnowsis ist a Software Tool to gather, order and structure the miscellanous Information and Data that can be found on personal computers by the principles of the semantic web.

NEWS

Duration: from April 2004 to March 2006

Sponsor: NEWS is funded by the European Commission's Information Society Programme (IST), FP6 - 001906

Short Description: The goal of the NEWS project is to develop News Intelligence Technology for the Semantic Web.

SmartWeb

Duration: from April 2004 to April 2007

Sponsor: German Federal Ministry of Education, Science, Research and Technology (BMBF)

Short Description: The goal of the SmartWeb project is to lay the foundations for multimodal user interfaces to distributed and composable semantic Web services on mobile devices. The SmartWeb consortium brings together experts from various research communities: mobile services, intelligent user interfaces, language and speech technology, information extraction, and semantic web technologies.

RISE

Duration: from January 2004 to December 2005

Sponsor: German Federal Ministry of Education, Science, Research and Technology (BMBF)

Short Description: The goal of Rise is to develop a wide solution for the reuse of products, methods, models and experience in software engineering. Initial point for the development work is the exposure with defect and feature (internal commissions) in software engineering. The outcome of this is a specifically fitted approach for a practical software engineering for small and medium companies.

Institute of Communication Acoustics

Institute of Communication Acoustics Building IC 1/132, Ruhr-University Bochum, 44780 Bochum, Germany

Tel: +49 (0)234 32-22496 Fax: +49 (0)234 32-14165

Website: http://www.ika.ruhr-uni-bochum.de/index_en.htm

Contact: Rainer Martin

Email: Rainer.Martin@rub.de

Number of positions: 28

Research projects

Spoken-dialogue-systems quality: A large number of information services is now available which make use of speech technology (speech recognition, dialogue management, speech synthesis) in order to create user-friendly access to information. The quality of these so-called spoken-dialogue systems depends on the technology used, but also on the conditions they are used in (environmental acoustics, transmission channels, etc.). Based on available knowledge, a relatively new working area of the project group is the quality assessment of spoken-dialogue systems in real application scenarios. Assessment is performed on a global (i.e. referring to the system as a whole) or on a diagnostic level (e.g. assessing the impact of transmission channels on speech recognition or on speech synthesis, dialogue management assessment).

Speech-transmission quality: The quality of speech transmission systems is another major research topic of the group. The aim is to collect auditory data to verify and extend a computational model for speech-transmission quality. Obviously, speech-transmission quality is closely linked to speech-communication quality in general. Speech-communication devices have to meet the demands of actual users. These depend on the specific purpose of the devices, the users' situations, their experience with similar devices, and several other factors. Therefore, speech-communication quality is not a static entity, but it "happens" in a specific situation – for each user individually! Nevertheless, the planners of speech-communication devices such as speech-transmission systems typically have to satisfy the needs of large user groups. To this end, it is mandatory to be able to measure quality auditorily, or – even better – to predict instrumentally what level of quality can be reached with a specific device, the measurements and predictions being based on physical characteristics of the device and the speech signals.

Quality elements of speech synthesis: Speech synthesis systems have now been available for several years, in a quality that is sufficient for many simple applications (for example automatic announcements or email readers for the blind). For other requirements, natural voice continues to be by far superior. One of the group's objectives is to create a synthetic voice which sounds as natural, intelligible and flexible as a human voice. While much research effort has been put into intelligibility so far, naturalness and flexibility continue to be problematic.

Institute of Human-Machine Communication

Rcisstr. 16, Gebäude S6, 2. Stock, Zimmer S2640

Tel: +49 (0)89 289-28554 Fax: +49 (0)89 289-28535

Website: <http://www.mmk.ei.tum.de/>

Contact: Robert Lieb

Email: lieb@ei.tub.de

Number of positions: 51

Research projects

Research at the Institute for Human-Machine Communication focuses on the fundamentals of a widely intuitive, natural, and therefore multimodal interaction between humans and information processing systems. All forms of interaction, i.e. modalities, that are available to humans, are to be investigated for this purpose. Both the machine's representation of information and the interaction technique is to be considered in this context, like text and speech, sound and music, haptics, graphics and vision, gesture and mimics, and emotions.

Signal Processing means the theory and application of filtering, coding, transmitting, estimating, detecting, analyzing, recognizing, synthesizing, recording, and reproducing signals by digital or analog devices or techniques. The term signal includes audio, video, speech, image, communication, medical, musical, and other signals in continuous or discrete (i.e. sampled) form. Competence in Signal Processing is vital for the development of new techniques in Human-Machine Communication.

Our research in speech processing aims to develop algorithms and systems which are able to automatically recognize continuous speech under real-world conditions. For that purpose, **statistical classifiers** as well as **hybrid systems** are being investigated. Most methods are based on stochastic Hidden Markov models (HMMs), which are utilized as reference models for speech sounds (phonemes). Words and complete sentences can be built up from the phoneme models. The sentences are analysed by a speech understanding module, giving an interpretation of the meaning. Special problems have to be solved due to the great variability in pronunciation as well as to the strong dependence from the speaker. Here, we favourably apply pronunciation variants and adaptive classifiers.

Intelligent Information and Communication Systems

Fern Universität in Hagen Intelligent Information and Communication Systems

Informatikzentrum Universitätsstraße 1 58084 Hagen Germany

Tel: 02331-987 374 Fax: 02331-987 392

Website: <http://pi7.fernuni-hagen.de/homepage/>

Contact: Hermann Helbig

Email: hermann.helbig@fernuni-hagen.de

Number of positions: 20

Research projects

Knowledge Representation with Multilayered Extended Semantic Networks (the MultiNet paradigm)

The WOCADI parser (formerly: NatLink; WOCADI is an acronym for WOrd CIAss based DIambiguating) is a computer program written in the Scheme programming language that transforms a German text into a formal semantic representation using the MultiNet (multilayered extended semantic networks) formalism.

Morpho-lexical analysis of the text.

Syntactico-semantic analysis of the text.

Jenna University Language and Information Engineering Lab

Fürstengraben, 3007743 Jena, Germany

Tel: +49 3641 9-44324 Fax: +49 3641 9-44321

Website: http://supreme.coling.uni-jena.de/component/option,com_frontpage/Itemid,1/

Contact: Joachim Wermter

Email: wermter@coling-uni-jena.de

Number of positions: 20

Research projects

The JULIE Lab is a member of the EU-funded Network of Excellence entitled Semantic Interoperability and Data Mining in Biomedicine (NoE 507505). Its goal is to establish Europe as the international scientific leader in medical and biomedical informatics. The long-term goal of the network will be the development of generic methods and tools supporting the critical tasks of the field; data mining, knowledge discovery, knowledge representation, abstraction and indexing of information, semantic-based information retrieval in a complex and high-dimensional information space, and knowledge-based adaptive systems for provision of decision support for dissemination of evidence-based medicine.

Morphosaurus is a unique and powerful medical language tool that the Department of Medical Informatics in Freiburg, Germany, developed in cooperation with the JULIE Lab and the Health Technology Master Program of Parana Catholic University in Curitiba, Brasil. The project is mainly funded by Deutsche Forschungsgemeinschaft (DFG), grant KL 640/5-1 and 5-2.

The JULIE Lab is one of the founding groups of the BioTem-Initiative. This is a consortium of leading German research groups in the fast-growing field of biomedical text mining. Our goal is to make the importance of biomedical text mining visible to the German research funding authorities and to ultimately establish a German virtual center for text mining on an institutional level.

Language Technology Lab, DFKI

DFKI GmbH Forschungsbereich Sprachtechnologie? Stuhlsatzenhausweg 3 / Building 43.1 D-66123 Saarbrücken Germany

Tel: +49 681 302 5282 Fax: +49 681 302 5338

Website: <http://www.dfki.de/lt/v>

Contact: Hans Uszkoreit

Email: uszkoreit@dfki.de

Number of positions: 62

Research projects

- Exploiting - and automatically extending - ontologies for content processing
- tighter integration of shallow and deep techniques in processing.
- Enriching deep processing with statistical methods.
- Combining language checking with structuring tools in document authoring.
- Document indexing for German and English.
- Automatically associating recognized information with related information and thus building up collective knowledge.
- Automatically structuring and visualizing extracted information.
- Processing information encoded in multiple languages, among them Chinese and Japanese.

Linguistics Department, University of Konstanz

Wolke, Irene Office, G 213

Tel: +49 7531 88-4166 Fax: +49 7531 88-4160

Website: <http://ling.uni-konstanz.de/pages/index-en.html>

Contact: Markus Bader

Email: markus.bader@uni-konstanz.de

Number of positions: 67

Research projects

The International Lexicon: Contours and Alignment in Read Texts

The Universals Archive is available in the form of a searchable archive, enabling its on-line users to retrieve universals in terms of any of the individual words or combinations of words that occur in their formulation or in their documentation. It is also possible just to browse through the Universals Archive.

Manfred Pinkal Computational Linguistics Group

Fachrichtung 4.7 Allgemeine Linguistik? Universität des Saarlandes Im Stadtwald - Gebäude C 7.2 Postfach 15 11 50 66041 Saarbrücken Germany.

Tel: +49-681-302-4344 Fax: +49-681-302-4351

Website: <http://www.coli.uni-saarland.de/groups/MP/page.php?id=index>

Contact: Manfred Pinkal

Email: pinkal@coli.uni-sb.de

Number of positions: 31

Research projects

CHORUS is concerned with semantic processing of ambiguous natural language utterances. It aims at combining underspecification techniques and preferences.

CORTE deals with the automatic recognition, extraction and processing of definitions in German court decisions by use of linguistic technology.

Dialog empirically investigates flexible dialogue management strategies in complex mathematics tutoring dialogues and develops an experimental prototype system gradually embodying the empirical findings.

SALSA creates a semantically annotated corpus resource and investigates methods for its utilisation.

TALK This project aims to generalise the TRINDI and SIRIDUS "Information State Update" (ISU) approach to dialogue management to develop adaptive multimodal dialogue systems.

InDiGen aims at developing an integrated approach to discourse and sentence planning which captures the interaction of discourse marker selection, ellipsis, and discourse structure.

COLLATE is a joint effort of the DFKI, Department for Computational Linguistics & Phonetics and the Department for Computer Science at Saarland University to establish a German competence center for language technology in Saarbruecken.

MiLCA is a joint project on e-learning, in which different computational linguistics locations interlace their curricula by a variety of virtual courses. MiLCA Saarbrücken develops teaching material on Computational Semantics, Algorithms for Natural Language Processing, and Dialogue Processing.

Phonetics Institute, University of Saarlandes

Universität des Saarlandes, FR 4.7 Phonetik, Geb. 17.2, Raum 4.11, Postfach 15 11 50, D-66041 Saarbrücken, Germany

Tel: + 49 (681) 302 2926 Fax: +49 (681) 302 4684

Website: <http://www.coli.uni-saarland.de/groups/WB/phonetik/>

Contact: William Barry

Email: wbarry@coli.uni-sb.de

Number of positions: 88

Research projects

SecurePhone

The SecurePhone project integrates a biometric recogniser in a 3G/B3G enabled PDA. The text-dependent speaker verification is the main responsibility of Saarland University.

Logox

German and English text-to-speech synthesis on the basis of microsegment concatenation, in collaboration with G DATA.

Mary

A modular German text-to-speech synthesis system, in collaboration with DFKI.

German Intonation

Verification and extension of a model of German intonation (see also GToBI)

NaDia

Phrase-based speech synthesis, in collaboration with CLT

PhonASR

Phonetic features, speaking rate and subword units in acoustic decoding for automatic speech recognition

Speaking lift

Speech recognition for the lift in the General Linguistics building, in collaboration with Computational Linguistics and Computer Science

PrAnK

Prosodic analysis of the German KielCorpus

NECA

A net environment for embodied emotional conversational agents, in collaboration with OEFAI (coordinator), DFKI Freeserve, ITRI and Sysis

Phon2

Online phonetics course, initially aimed at native speakers of Bulgarian, English, French and German, in collaboration with University of Reading (coordinator), University of Sofia and University of Pau.

Voice profiles

Classification of healthy and pathological voice profiles on the basis of the microphone and electroglottographic signals, in collaboration with the ENT clinic of Caritasklinik St. Theresia in Saarbrücken and the Department of Neurology and the Department of Neurosurgery of the University of the Saarland in Homburg, funded by the German Research Council (DFG).

Organismes et associations

Bavarian Archive for Speech Signals

Bavarian Archive for Speech Signals c/o Institut für Phonetik, Universität München
Schellingstr. 3 / II 80799 München

Tel: +49-89-2180-2758 Fax: +49-89-2800362

Website: <http://www.bas.uni-muenchen.de>

Contact: Florian Schiel

Email: Schiel@phonetik.uni-muenchen.de

Number of positions:

Research projects

The *Bavarian Archive for Speech Signals (BAS)* is a public institution hosted by the University of Munich. This institution was founded with the aim of making corpora of current spoken German available to both the basic research and the speech technology communities via a maximally comprehensive digital speech-signal database. The speech material will be structured in a manner allowing flexible and precise access, with acoustic-phonetic and linguistic-phonetic evaluation forming an integral part of it.

Tasks: The last few years have seen an abrupt increase in the demand for large speech-signal data collections, both on the part of academic investigators carrying out basic research as well as on the part of engineers from industry working in the new integrated field of speech and information technology. There are many reasons for this. Primarily, however, the sudden increase in demand must be attributed to the breakneck pace of hardware and software development in speech signal processing. The increasing number of techniques for acoustic-phonetic signal processing, and the increasing amount of speech data that can be efficiently handled and processed together generate an accompanying demand not only for linguistically interesting text material (which of course emerges automatically from the modern printing industry) but also for reliably acquired and phonetically evaluated spoken language material. A number of national and international initiatives (such as BDFON, PHONDAT, LDC, SPEX or COCOSDA) have, it is true, already resulted in the collection and distribution of large speech corpora. However, they exhibit a variety of formats, corresponding to the variety in the aims pursued. For German, a central institution was clearly lacking that could carry out such tasks within a long-term perspective. *BAS* will be responsible in Germany for these tasks for distributable corpora of spoken German, collecting, maintaining and making them available in standardized form.

In addition, *BAS* will develop its own procedures for automatic labelling and segmentation, making the results available with the distributed speech corpora.

Australie

Laboratoires

Centre for Language Technology, Macquarie University
Macquarie University, NSW 2109, Australia Tel: +61 2 9850 7111 Website: http://www.clt.mq.edu.au/ Contact: Robert Dale Email: ltinfo@ics.mq.edu.au Number of positions: 28
Research projects

Research within the Centre for Language Technology is largely carried out within the context of projects. Current projects are listed here; past projects can be found here. See the web pages of individual members of the Centre for other details of research interests.

AnswerFinder (coordinator: Diego Molla) This project looks at how existing document bases can be used as a source of answers to questions posed in natural language AnswerFinder combines traditional approaches to natural language processing with current robust approaches. In the indexing stage, AnswerFinder exploits the linguistic information in the target documents. The result is a simplified representation of the logical form of the document sentences. In the retrieval stage, the question is thoroughly analysed. A fall-back retrieval procedure finds the sentences whose logical forms indicate that they contain the answer, and extracts the answers. **DDD** (coordinator: Robert Dale) This project is concerned with dynamic document delivery: the use of natural language generation techniques to provide customized and up-to-date information via the web. **The Meeting Room Project** (coordinator: Steve Cassidy) The meeting room project is aimed at building speech technology applications in the context of a meeting room. One of our core goals is to build technology that can be deployed without invasive or complex instrumentation (eg. headsets, large array microphones). We are also interested in the confluence of speech technology and language technology and how both can be applied to make speech technology in the meeting room be a viable and useful resource. **PENG** is a computer-processable controlled natural language designed for writing unambiguous and precise specifications and use cases. PENG covers a strict subset of standard English and is precisely defined by a controlled grammar and a controlled lexicon. In contrast to other controlled languages, the author does not need to know the grammatical constraints of the language explicitly. ECOLÉ, a look-ahead text editor, points out the grammatical constraints while the text is written.

**Department of Computer Science and Software Engineering,
University of Melbourne**

4th Floor, ICT Building, University of Melbourne, 111 Barry St., Carlton Vic 3053, Australia.

Tel: +61 3 8344 1301 Fax: +61 3 9348 1184

Website: <http://www.cs.mu.oz.au/research/index.html>

Contact: Ramamohanarao Kotagiri

Number of positions: 77

Research projects

Human Language Technology

Human language and communication are extremely complex, naturally occurring phenomena. The analysis of language, whether in speech, text or multimodal forms, is a significant computational challenge. As the quantity of information on the Web grows, the need for automated language analysis systems becomes more pressing. Equally, as mobile and embedded computers proliferate, the need for natural linguistic interaction between humans and machines becomes more acute. Language technologies are beginning to address these needs. In this project we are pursuing open research questions in the following areas: language modelling, multimodal linguistic annotation, high performance computing for natural language processing, electronic documentation of endangered languages, structured models for linguistic data, and digital libraries for language resources.

Information Discovery

Large document collection, such as those maintained by corporations, provide challenges in several ways. There is the challenge of storing the data efficiently; the challenge of indexing it, so that it can be searched by content; and there is the challenge of knowing how to implement searching so as to return the documents that are likely to be relevant to the query posed by the searcher. In this project we consider the efficient and effective implementation of Information Retrieval systems for multi-gigabyte text collections. We are particularly interested in storage and indexing solutions that provide fast heuristic searching; and in techniques that offer the prospect of scaling to cope with very large collections indeed. Experimental work includes involvement with the US-funded TREC investigation.

Machine Learning and Data Mining

Finding patterns in large collections of data and using these patterns for reasoning is a challenging task. In this project we are interested in efficient algorithms for mining a number of fundamental types of patterns and methods for using these patterns in classification.

Knowledge Management Research Group, University of Sydney

School of Information Technologies, Madsen Building, University of Sydney, NSW 2006

Tel: +61 2 9351 4291 Fax: +61 2 9351 3838

Website: <http://kmsg.it.usyd.edu.au/index.html>

Contact: Joseph Davis

Email: jdavis@it.usyd.edu.au

Number of positions: 21

Research projects

Knowledge Management (KM) is an emerging research area that addresses conceptual and practical questions related to mapping, codifying, extracting, and utilizing the various forms of knowledge in organisations and the design and implementation of tools and systems to facilitate the sharing of such knowledge. It draws on and integrates developments in a number of topics such as document and workflow management systems, natural language processing, knowledge discovery and data mining, text and web mining, organisational memory systems, peer-to-peer computing, social network analysis, among others.

Research Themes:

- Knowledge Discovery;
- Knowledge Representation;
- Knowledge Sharing and Transfer;
- Knowledge Economy.

Speech Hearing and Language Research Centre, Macquarie University

Department of Linguistics, Macquarie University, NSW Australia 210

Tel: + 61 2 9850 8788 Fax: + 61 2 9850 9199

Website: <http://www.ling.mq.edu.au/shlrc/index.htm>

Contact: Linda Cupples

Email: Linda.cupples@ling.mq.deu.au

Number of positions: 67

Research projects

Members of the Speech Hearing and Language Research Centre (SHLRC) are concerned with investigating the perceptual, acoustic, articulatory, psycholinguistic, and sociolinguistic processes involved in producing and understanding speech and language. Our focus is on examining speech, hearing and language processes in people with 'typical' speech and language abilities as well as people with impairments in speech, hearing or language. Our aim is to understand better the processes underlying speech, hearing and language use as well as the ways in which those processes can be affected by various forms of disorder including acquired brain damage, hearing impairment, and developmental disability.

The Centre includes speech, hearing and language specialists with diverse backgrounds in areas such as audiology, corpus linguistics, experimental phonetics and phonology, psychology, speech and language pathology, and sociolinguistics. Members of the Centre bring to bear a range of theoretical perspectives from a number of scientific disciplines including cognitive neuropsychology, clinical linguistics, psycholinguistics, sociolinguistics, speech science and technology, and physiology. Many different methodological approaches are employed in the implementation of our research such as experimental design and quantitative methods, case studies, discourse analysis, qualitative methods, epidemiology, longitudinal research and intervention studies. Use of these different methodologies enables members to address interrelated issues of theoretical and practical importance in a complementary and often integrated way, thus enhancing the Centre's overall research strength.

Sydney Language Technology Research Group, University of Sydney

Sydney Language Technology Research Group, School of Information Technologies,
University of Sydney, NSW 2006

Tel: +61 2 9351 4174 Fax: +61 2 9351 3838

Website: <http://www.cs.usyd.edu.au/~rcdmnl/>

Contact: Joseph Davis

Email: jdavis@it.usyd.edu.au

Number of positions: 20

Research projects

- Data Mining of Semi-Structured Data;
- Core Ontology of the Basque Language;
- Core Ontology of the Basque Language (Some early results) ;
- Diachronic Phonology;
- Intelligent Self-Learning Parser-editor for Dictionary Databases (Project Description) ;
- Electronic Reference materials for Language Learning and Translation - Proof of Concept ;
- Workflow Methodology for the Automatic Regeneration of Multi-media Linguistic Publications (Project Description);
- Software Architecture and Engineering for LT Systems;
- Evidence Based Parsers;
- Document Classification Workbench;
- Named Entity Recognition;
- The Computation of an Ontolexicon;
- A Workbench for the markup of Systemic Functional Linguistics Linguistic;
- Methods for Extracting Tacit Knowledge for Knowledge Management.

Organismes et associations

Australian Computer Science Society

Level 3 160 Clarence Street, SYDNEY NSW 2000

Tel: +61 2 9299 3666 Fax: +61 2 9299 3997

Website: <http://www.acs.org.au/>

Contact: Dennis Furini

Email: dennis.furini@acs.org.au

Research projects

The Australian Computer Society (ACS) is the recognised association for Information & Communications Technology (ICT) professionals, attracting a large and active membership from all levels of the ICT industry. A member of the Australian Council of Professions, the ACS is the public voice of the ICT profession and the guardian of professional ethics and standards in the ICT industry, with a commitment to the wider community to ensure the beneficial use of ICT.

The society was founded in 1966. Its objectives are to further the study, science and application of Information Technology; promote, develop and monitor competence in the practice of ICT by people and organisations; maintain and promote a Code of Ethics for members of the Society; define and promote standards of knowledge of ICT for members, promote the formulation of effective policies on ICT and related matters; extend the knowledge and understanding of ICT in the community; promote the benefits of membership of the Society and promote the benefits of employing members of the Society.

ACS members work in all areas of business and industry, government and academia, and are qualified and experienced ICT professionals committed to the Society's Code of Ethics and Code of Professional Conduct and Professional Practice. ACS membership denotes a commitment to professionalism.

Principal Object

To promote the development of Australian information and communications technology resources.

Australian Speech Science and Technology Association

Website: <http://www.assta.org/>

Contact: Dennis Burnham

Email: secretary@assta.org

Mission

The Australasian Speech Science and Technology Association (ASSTA) is a scientific association which aims to advance the understanding of speech science and its application to speech technology in a way that is appropriate for Australia and New Zealand. In order to fulfil these aims the Association seeks to provide significant exchange between those actively involved in speech science and technology research and practice, between this multidisciplinary body of professionals and the community in general as well as other national and international bodies having related aims. The Association regularly updates its constitution to reflect its evolving needs and is incorporated in the Australian Capital Territory as a non-profit scientific association with specific objectives designed to promote its aims.

KnowledgeNet (ARC Research Network for Next Generation Web Technologies)

ARC, Macquarie University, NSW Australia 210

Website: <http://www.comp.mq.edu.au/research/semantic/index.html>

Contact: Abhaya Nayak

Email: abhaya@ics.mq.edu.au

Number of positions: 100

Mission

The World Wide Web (WWW) was designed as an information space, with the expectation that it would be useful not only for human-human communication, but also that machines would be able to participate and help. To realize its full potential, the Semantic Web (widely taken to be the Next-Generation Web) augments the WWW by adding machine understandable content to web resources, providing extra information on content, and enabling agents acting on behalf of humans to reason with the information obtained. Internationally, the Semantic Web is currently in an early, but highly active, state of development and exploitation. To fully exploit the power afforded by the Semantic Web, we need to harness the multifarious research activities conducted in cognate areas including agent technologies, web technologies, web security, knowledge engineering, information visualization, image processing, and collaborative technologies. The activities targeted in KnowledgeNet, this Research Network for Next Generation Web Technologies, will enable Australian researchers to collaborate on internationally competitive research programs that will help to shape the key technologies at the foundation of the next generation of the WWW and lead to the realization of the vision of an eAustralia.

KnowledgeNet has evolved from an initial seed-funding bid titled "Intelligent Applications Through The Semantic Web" (SR0354476). It has drawn some new members and activities from the bid "Transforming Knowledge Spaces: Open Technologies for Research Collaboration and Research Communication" (SR0354839), and has also attracted a number of other researchers with complementary strengths.

KnowledgeNet is hosted by Macquarie University that has an outstanding record in fostering research excellence, in attracting ARC and other competitive research funding, in commercialization of research, and in public education.

National Information and Communication Technology Association of Australia

Australian Technology Park Bay 15 Locomotive Workshop Australian Technology Park
Eveleigh NSW 1430 Australia

Tel: : +61 2 9209 4750 Fax: + 61 2 9209 4748

Website: <http://nicta.com.au/>

Contact: Clare Gill

Email: clare.gill@nicta.com.au

Mission

Information and Communications Technology (ICT) has dramatically altered the world's social and economic landscape.

Australia has enormous, untapped ICT talent of an international standard. Established as part of *Backing Australia's Ability* — an Australian Government initiative to promote science and innovation — National ICT Australia (NICTA) aims to capitalise on this untapped potential.

NICTA was formed by the Federal Government's Department of Communications, Information Technology and the Arts and the Australian Research Council. NICTA's consortium partners are the Australian Capital Territory Government, the New South Wales Government, the University of New South Wales, and the Australian National University. A national laboratory has emerged from this powerful union based on the foundations of research, commercialisation, education, and collaboration. With a focus on use-inspired basic research, NICTA attracts, develops, and networks exceptional talent for Australia's future prosperity.

NICTA's research efforts focus on the technology challenges facing industry, community, and the national interest. The central drivers of NICTA's research are the *Priority Challenges*.

The creation of technology for use in society, the market, and the environment inspire the Priority Challenges. NICTA's two initial Priority Challenges are *Trusted Wireless Networks* and *From Data to Knowledge*.

Smart Internet Technology CRC

Smart Internet Technology CRC Bay 8 Suite 9/G12 Australian Technology Park Eveleigh
NSW 1430

Tel: +612 8374 5080 Fax: +612 8374 5090

Website: <http://www.smartinternet.com.au/SITWEB/index.jsp>

Contact: Darrel Williamson

Email: darrell.williamson@smartinternet.com.au

Mission

To ensure efficient and effective world class research with significant technological and commercial outcomes.

To provide integration between technological innovation and social and behavioural research.

To provide high-value Internet-based technologies, solutions and services that are competitive in the international marketplace.

To provide actual and future benefits 5 times the value of the resources used.

To produce knowledgeable, skilled and enterprising Information and Communication Technology (ICT) professionals required for Australia.

To develop enabling technologies, solutions and services to increase productivity and deliver efficiency gains for Australia through pervasive use of the Internet.

To assist, through technology transfer, the commercial advancement of Australian Industry, including Small to Medium Enterprises (SMEs).

Belgique

Laboratoires

Center for Computational Linguistics, Université Catholique de Louvain

Katholieke Universiteit Leuven, Center for Computational Linguistics
Maria-Theresiastraat 21, B-3000 Leuven

Tel: +32-16-325088 Fax: +32-16-325098

Website: <http://www.ccl.kuleuven.be/about.en.php>

Contact: Geert Adriaens

Email: geert.adriaens@ccl.kleuven.ac.be

Number of positions: 12, unknown number of students

Research projects

Machine Translation : EUROTRA, METAL, METIS, METIS-II.

Computational Syntax and Semantics : Semantics and Machine Translation, Formal Semantics for Discourse, Linguistic Specifications for Future Industrial Standards, MODUL: Monolingual Linguistic Specifications for Dutch, LS-GRAM+ (DUTCH), LINGUADUCT: Contextual interpretation of natural language texts by means of abductive reasoning and inductively acquired knowledge, MultiMeteo (components for Dutch), Formal Semantics, Universals and Machine Translation, Computacionele syntaxis en semantiek, Parallel Natural Language Processing: PARLEX, A probabilistic grammar of Spanish, SEMADUCT: Combining deductive and inductive techniques for lexical semantics.

Computational Lexicography : PROTON, Lexically anchored linguistic knowledge representation for natural language processing applications, The Construction of Bilingual Dictionaries.

Computational Phonology : FONILEX: A phonological database with the pronunciation of Dutch in Flanders.

Corpus Linguistics : ANNO: Annotated linguistic database for Dutch, CGN: Corpus Gesproken Nederlands, D-coi: Dutch Language Corpus Initiative.

Automatic Transcription : ATraNoS: Automatic Transcription and Normalisation of Speech

Document Handling : ALCOGRAM, MENELAS, SALOMON: Towards automatic legal text categorization and legal data extraction, SECC: A Simplified English Checker/Corrector.

Computer Aided Language Learning & Course Development : A.D.A.M. & E.V.E.: Automated Document Analysis and Manipulation and Extensible Variety of Exercises, CTP: Comprehensive Training Package, Word Drum, ZÉRO FAUTE, DECIDE: Development of European Courses on Information and Datacom Engineering, E.E.T.: Extended English Language TEaching Vocabulary-list, MYTHE: Multimedia Young-children Thesaurus for

Educational Purpose.

Center for Processing Speech and Images (PSI)

U.Leuven - ESAT/PSI Kasteelpark Arenberg 10 3001, Heverlee BELGIUM

Tel: +32-16-32.1057 Fax: +32-16-32.1723

Website: <http://www.esat.kuleuven.be/psi/spraak/>

Contact: Patrick Wambacq

Email: patrick.wambacq@esat.kuleuven.be

Number of positions: 70

Research projects

The Center for Processing Speech and Images (PSI) is with 70 people one of the big research units within the Electrical Engineering (ESAT) Department. Research covers a large area in the field of audio-visual information processing.

Speech research projects The combination of speech and image processing within the same group is an important advantage for the future, which is quite unique within the European context. Our researchers come from different backgrounds: engineers, physicists, mathematicians, linguists... This multi disciplinarity gives PSI the possibility to couple fundamental innovation with practical solutions in a broad spectrum of applications. The large number of European projects underlines our international reputation. PSI has been a starting point of successful spin-offs and a reliable partner for industry.

MIDAS: Missing Data Solutions.

ESSCorTeSS: Exponential Sinusoidal Speech compression for Corpus-based Text-to-Speech Synthesis.

JASMIN-CGN: Jongeren, Andertaligen, Senioren en Machine Interactie voor het Nederlands.

SPACE: Speech Algorithms for Clinical and Educational applications.

Akoestische Adaptatie van HMMs voor Computerondersteund Leren.

TEMPLATE: Data Driven Speech Recognition.

FLaVoR -- Flexible Large Vocabulary Recognition: Incorporating Linguistic Knowledge.

**Department of Electronics and Information Processing, Vrije
Universiteit Brussels**

Vrije Universiteit Brussel (VUB) Dept. of Electronics and Information Processing, (ETRO)
Pleinlaan 2 - B-1050 Brussel – Belgium Room 5K321

Tel: +32 2 629 29 30 Fax: +32 2 629 28 83

Website: <http://www.etro.vub.ac.be/>

Contact: jpcornel@etro.vub.ac.be

Email: Jan Cornelis

Number of positions: 107

Research projects

- Speech intelligibility.
- Voice modification and conversion.
- Transplantation of voice characteristics.
- Data to speech systems.
- Time and pitch scaling of speech and audio.
- Modelling and parametric coding of audio.
- Psycho-acoustic re-quantization of audio.

Digital Speech and Signal Processing, Ghent University

Universiteit Gent, Vakgroep Elektronica en Informatiesystemen St. Pietersnieuwstraat 41, B-9000 Gent, Belgium

Tel: +32-9-264 33 66 Fax: +32-9-264 35 94

Website: <http://www.elis.ugent.be/>

Contact: Jan Van Campenhout

Email: Jan.VanCampenhout@elis.ugent.be

Number of positions: 16, unknown number of students

Research projects

The Digital Speech and Signal Processing (DSSP) research group at the Electronics and Information Systems (ELIS) department of Ghent University concentrates on speech analysis, synthesis and recognition, and on automatic music content analysis.

AUTONOMATA (Automata for deriving phoneme transcriptions of Dutch and Flemish names) This project is about the creation of a grapheme-to-phoneme (g2p) convertor that performs better on the transcription of Dutch and Flemish names than the standard g2p's one can buy.

SPACE (Speech algorithms for clinical and educational applications) This project is about the application of automatic speech recognition and synthesis techniques for the assessment of pathological voices (dysarthry, speech of the deaf, etc.) and for the assessment of pronunciation and reading proficiency of students (dyslexia, non-natives).

Théorie des Circuits et Traitement du Signal, Faculté Polytechnique de Mons

MULTITEL-TCTS Lab, Faculté Polytechnique de Mons, Parc Initialis, Copernic Avenue, B-7000 MONS, Belgium

Tel: +32 65 374730 Fax: +32 65 374729

Website: <http://tcts.fpms.ac.be/index.php>

Contact: Joel Hancq

Email: joel.hancq@fpms.ac.be

Number of positions: 28

Research projects

Speech synthesis relates to the automatic production of speech from text. We have developed and patented an in-house speech synthesis algorithm, MBROLA, and currently coordinate the MBROLA Project for Free Multilingual Speech Synthesis (34 languages are now supported, thanks to international collaboration). Current research focuses on new signal processing models of speech, on voice conversion for speech synthesis systems, as well as on real-time interfaces for audio synthesis. Applications : multimedia applications, technical aids for handicapped people, man-machine communication, email readers, consumer products.

Through several European and national projects, but also in collaboration with the MULTITEL ASBL research center, the ASR team addresses many aspects of the speech recognition problem:

Robust speech recognition: through several projects (RESPITE, MODIVOC, AURORA) we have deeply investigated the problem of robust speech recognition: spectral subtraction, wiener filtering, noise estimation, multi-band, mixture of experts, missing data, microphone arrays, ...

Keyword spotting: the problem of keyword spotting is crucial in real life application in order to partly handle spontaneous speech in man-machine dialogues. Next to keyword spotting, the problem of rejection of out of vocabulary words or poorly recognized word, has also been studied through the estimation of relevant confidence levels.

Model adaptation: fast model adaptation allows to update the statistical models with very few data. Noise adaptation and speaker adaptation have been studied, more particularly in the framework of the hybrid HMM/ANN systems.

Organismes et associations

Belgian Acoustical Society (ABAV)

ABAV General Secretariat CSTC – WCTB avenue Pierre Holoffe 21? B-1342 Limelette
Belgium

Tel: +32 16 327128 Fax: +32 16 327984

Website: <http://www.abav.be/>

Contact: Gerrit Vermeir

Email: gerrit.vermeir@bwk.kuleuven.ac.be

Mission

The Belgian Acoustical Society (ABAV) was founded in 1966, as a national society without financial profit goal. The society gathers all interested persons from universities, research laboratories, consulting offices, industries, governmental organisations, ... with activities in the field of acoustics in its broadest sense: physical acoustics, industrial noise control, building acoustics, environmental protection, electro-acoustics, noise protection, ...

The society has at the moment approximately 125 members (effective, associate, student and honorary) and 40 supporting members.

The ABAV Society has the following goals:

to create a permanent working frame for information exchange between members emanating from any of the acoustical disciplines; to collect and to exchange information on all topics of concern for the society; to create a forum for discussion and exchange about noise management, noise protection and noise control in the interest of creating acceptable, our agreeable acoustical environment; to contribute to the development and promotion of acoustics as a scientific discipline; to establish strong links with similar national and international associations.

Chine et Taiwan

Laboratoires

Institute of Information Science, Academia Sinica
<p>INSTITUTE OF INFORMATION SCIENCE, ACADEMIA SINICA, 128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan Tel: +886 2 2788 3799 Fax: +886 2 2782 4814 Website: http://www.iis.sinica.edu.tw/ Contact: Jane Win-Shih Liu Email: janeliu@iis.sinica.edu.tw Number of positions: 54</p>
Research projects
<ul style="list-style-type: none">- Adaptive Intelligent Internet Agents Research Laboratory.- Advanced Networking Technologies and Services Group.- Automation Laboratory.- Computer Systems and Communication Laboratory.- Intelligent Agent Systems Laboratory.- Intelligent Transportation Systems.- Machine Learning and Classification Laboratory.- Verification Automation Laboratory.- Web-based Collaboratory Laboratory.

National Taiwan University , Department of Computer Science and Information Engineering

#1 Roosevelt Rd. Sec. 4, Taipei, Taiwan 106, ROC Dept. of CSIE, NTU

Tel: + 886 2 3366 4888 ext. 203 Fax: + 886 2 2362 8167

Website: <http://www.csie.ntu.edu.tw/>

Contact: Tei-Wei Kuo

Email: chairman@csie.ntu.edu.tw

Number of positions: 32

Research projects

- Acoustic Processing: Features, modeling, training & pronunciation variation;
- Linguistic Processing & Language Modeling;
- Search/Decoding Algorithm and Keyword Spotting;
- Multilingual Information Management;
- Language Engineering.

National Tsing Hua University, National Tsing Hua University

Department of Computer Science, National Tsing Hua University, 101, Kuang Fu Rd, Sec.2, HsingChu, Taiwan 300 R.O.C

Tel: +886-3-5714787 Fax: +886-3-5723694

Website: www.cs.nthu.edu.tw

Contact: Biing-Feng Wang

Email: bfwang@cs.nthu.edu.tw

Number of positions: 41

Research projects

Research on Indexing and searching techniques for multimedia; focus on rhythm, chord recognition, sound synthesis, and voice recognition;

Parallel language design, optimizing compilers for embedded systems.

Spoken Language Group, Academia Sinica

INSTITUTE OF INFORMATION SCIENCE, ACADEMIA SINICA, 128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan

Tel: 886-2-2788-3799 Fax: 886-2-2782-4814

Website: <http://sovideo.iis.sinica.edu.tw/SLG/index.htm>

Contact: Hsin-Min Wang

Email: whm@iis.sinica.edu.tw

Number of positions: 29

Research projects

- Integration of Speech and Speaker Recognition for Biometric Person Authentication.
- Development of Linguistic and Phonological Knowledgebase for Next-Generation.
- Automatic Speech Recognition.
- A Study on Database Technology for Electronic Audio/Video Digital Archive.

Corée

Laboratoires

Postech Intelligent Software Lab.
San 31, Hyoja-Dong, Pohang, 790-784, Korea Tel: +82 54 279 5581 Fax: +82 54 279 2299 Website: isoft.postech.ac.kr Contact: Prof. Gary Geunbae Lee Email: gblee@postech.ac.kr Number of positions: 39
Research projects
<ul style="list-style-type: none">- Speech Recognition and TTS.- Information Retrieval & Language Processing.- Multi-Lingual Information Processing.- Bio-Informatics and Text Mining.- Spoken Dialogue Management.- Robust Spoken Language Understanding.

Danemark

Laboratoires

Center for Language Technology, Faculty of Humanities, University of Copenhagen

Njalsgade 80, 2300 Copenhagen S.

Tel: +45 35329090 Fax: +45 35329089

Website: <http://www.cst.dk/uk/index.html>

Contact: Bente Maegaard

Email: bente@cst.dk

Number of positions: 20

Research projects

- Infost - Information structure in the Danish language. The project ends March 2006.
- LEXADV Danish, Norwegian and Swedish adverbs.
- MULINCO MULTiLINGual Corpus of the University of COpenhagen.
- OntoQuery Ontology-based Querying.
- PaTrans Unification-based MT system for translating patents English => Danish
- Validation Centre Written Language Resources
- DanDokCenter Danish Documentation Centre for Research Results of HLT.
- DefSum Tool for summerization of Danish texts
- DIALOG PC-based system for spoken Danish dialog conc. flight reservations
- EAGLES Evaluation of Natural Language Processing Systems
- Enabler European National Activities for Basic Language Resources. The project ended 31 August 2003.
- IATE - Services for the Development of an Interactive Terminology Database System.
- LINDA Linguistic Specifications for Danish.

Department of Computational Linguistics, Copenhagen Business School

CBS, Solbjerg Plads 3 DK-2000 Frederiksberg, Copenhagen

Tel: +45 3815 3123

Website: http://uk.cbs.dk/forskning_viden/fakulteter_institutter_centre/institutter/sprog/id

Contact: Sabine Kirchmeier-Andersen

Email: ska.id@cbs.dk

Number of positions: 23

Research projects

The research at the Dept. of Computational Linguistics is conducted within three core fields

- **Formal descriptions of natural-language syntax and semantics.**
- **LSP knowledge and data modeling.**
- **Statistical language processing.**

Special attention is given to the following areas:

Databases with linguistic content: dynamical terminology and knowledge-bases with user-defined interfaces facilitate the systematisation of concepts in concept systems.

Information search: ontology-based systems offer a more intelligent and user-friendly search than the text-string based search available today.

Natural-language interfacing: enables flexible communication with the computer through speech and writing – in Danish.

Machine translation: translation memory, or automatic translation where rule-based systems are combined with probabilistic methods – a solution to the global, linguistic challenges that face companies today.

Corpus linguistics: modern linguistic research with the aid of corpora – collections of text annotated with grammatical information ("treebanks") form the empirical basis of our work.

Department of Nordic Studies and Linguistics, University of Copenhagen

Njalsgade 80, DK-2300 Koebenhavn S, Denmark

Tel: +45 35 32 86 40 Fax: +45 35 32 86 35

Website: <http://www.cphling.dk/>

Contact: Nina Grønnum

Email: ng@cphling.dk

Number of positions: 40

Research projects

The Department of General and Applied Linguistics is part of the Faculty of Letters, and is housed in the university's building complex on Amager (KUA). The department was founded in 1988 by an amalgamation of the Institute of Linguistics, the Institute of Phonetics, the Institute of Applied and Mathematical Linguistics, and the Center for Audiologopedics.

The amalgamation of these formerly autonomous units has resulted in a department with a wide selection of more or less intimately connected activities. Common to all is a close association with general linguistic theory and methodology (linguistics taken in its widest sense). It is thus considered important that research on Danish and other languages contribute to general theoretical and typological language descriptions, just as general linguistic theory and methodology are being investigated as a natural tool for analytical and descriptive linguistic research. Apart from the more theoretical disciplines, the department also hosts a number of applied ones.

From September 1999 until December 2001 IAAS collaborated with Center for Person Communication, Aalborg University, and TDC (Danish TeleCom), to improve and further develop Danish concatenative text-to-speech synthesis. The project was funded by the Danish Ministry for Research, Technology and Development. The system is being made commercially available by Speech-Ware.

Espagne

Laboratoires

Artificial Intelligence section, Software Dept., Technical University of Catalonia
Departament de Llenguatges i Sistemes Informàtics. Universitat Politècnica de Catalunya. C/Jordi Girona Salgado, 1-3, Mòdul C6. 08034 Barcelona - Spain Tel: +34 93 401 56 42 Fax: +34 93 401 70 14 Website: http://www.lsi.upc.es/dept/investigacion/sectia/home.html Contact: René Alquézar Email: alquezar@lsi.upc.es Number of positions: 45
Research projects
EUROWORDNET: Building a multilingual wordnet with semantic relations between words Application of knowledge-based systems to the management of wastewater treatment plants and biotechnological processes. Simultaneous elimination of carbon, nitrogen and phosphorous Construction and Integration of Knowledge Bases for Knowledge-Based Systems. ITEM: Textual Information Retrieval in a multilingual environment using NL Techniques.

Compilers and Languages, Escuela Superior de Ingeniería Informática, Orense

Escuela Superior de Ingeniería Informática, Departamento de Informática, Edificio Politécnico, Campus As Lagoas s/n, 32004 Orense (SPAIN)

Tel: +34 988 387280 Fax: +34 988 387001

Website: <http://coleweb.dc.fi.udc.es/>

Contact: Manuel Vilares Ferro

Email: vilares@uvigo.es

Number of positions: 20

Research projects

- Who, What, Where, How Much, How, Why? NLP tools, machine learning and Bayesian statistics for building high-precision query robots in financial domains.
- Extraction of multilingual economic information (ETIMON)
- Computer-based processing of multilingual economic information (TIEMPO) coordinated by
- Information retrieval for question-answering in economic texts (RIBETE)
- Computer-based processing of multilingual economic information (TIEMPO).
- Application of Language Engineering to Collaborative Systems and Desktop Publishing
- Extracting information from stock exchange news to assess market attitude.
- Natural Language Information Retrieval Systems for Cognitive Evaluations of Information.
- Cluster of 30 nodes with architecture x86.
- Querying Structured Textual Databases
- Information Retrieval and Extraction Applying Linguistic Knowledge.
- Automatic Analysis of Verbal Constructions in Spanish (syntactic level)
- Translation units: empirical analysis.
- Design of Tabular Analyzers for Natural Languages.

Database Lab, University of A Coruña

Laboratorio de Bases de Datos Facultad de Informática Campus de Elviña s/n 15071 A Coruña (España)

Tel: +34 981 16 70 00 Ext. 1306 Fax: +34 981 16 71 60

Website: <http://cedeira.dc.fi.udc.es/PlainConsulta/MainPage.do?forward=MainPage.do>

Contact: Nieves Rodríguez Brisaboa

Email: brisaboa@udc.es

Number of positions: 20

Research projects

The Database Lab (LBD) of the University of A Coruña was created in 1994, and since then it has been headed by Nieves R. Brisaboa (brisaboa@udc.es). Nowadays, the LBD is integrated by researchers and teaching staff from the Computer Science Department of the University of A Coruña, who work in the following research areas:

- Software Engineering
- Deductive Databases
- Digital Libraries and Federation of Digital Libraries
- Geographic Information Systems
- Compression, Indexing and Text Retrieval
- Natural Language Processing

Department of Artificial Intelligence, Technical University of Madrid

Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, 28660 Boadilla del Monte, Madrid, España

Tel: +34-91-352-4819 Fax: +34-91-336-7444

Website: <http://www.dia.fi.upm.es/indexeng.htm>

Contact: Ana Garcia Serrano

Email: agarcia@dia.fi.upm.es

Number of positions: 41

Research projects

The department of artificial intelligence comprises many groups working on language technology issues:

Grupo de Investigación de Sistemas Inteligentes (ISYS) (Intelligent Systems Research group)

Leaders: Associate professor García Serrano, Ana Ma. (agarcia@dia.fi.upm.es), Associate professor Molina González, Martín (mmolina@fi.upm.es)

Members: 3 Assistant professor/s, 8 Research assistant/s, 3 Student/s

Research topics: Knowledge Based Systems (Second Generation Architectures, Real Applications), Distributed Artificial Intelligence, Computer Supported Cooperative Work, Natural Language.

Grupo de Investigación en Validación y Aplicaciones Industriales

Leaders: Associate professor Cardeñosa Lera, Jesús (carde@fi.upm.es)

Members: 2 Associate professor/s, 4 Research assistant/s, 5 Student/s

Research topics: Applied Artificial Intelligence, Knowledge Based Systems, Quality Systems Processes, Reengineering (BPR), Reusability. Domain Analysis, Computer Support Cooperative Work (CSCW), Automated Negotiation Systems, Natural language/Machine translation.

Grupo de Ontologías (Ontologies)

Leaders: Associate professor Gómez Pérez, Asunción (asun@dia.fi.upm.es)

Members: 1 Assistant professor/s, 3 Research assistant/s, 7 Student/s

Research topics: Methodologies for building Ontologies, Formal evaluation of Ontologies, Tools for building Ontologies, Ontology-based applications in areas like: e-commerce, knowledge management, intelligent integration information, intelligent search on internet, etc.

Laboratorio de Inteligencia Artificial (LIA) (Artificial Intelligence Laboratory)

Leaders: Full professor Pazos Sierra, Juan (jpazos@fi.upm.es), Associate professor Gómez Pérez, Asunción (asun@dia.fi.upm.es)

Members: 4 Associate professor/s, 3 Assistant professor/s, 4 Research assistant/s, 7 Ph. D. student/s, 3 Master student/s, 30+ Student/s

Research topics: Expert Systems, Natural Language, Knowledge Based Systems Validation, Artificial Intelligence, Knowledge Sharing and Reuse, Molecular Computing.

IXA Group

Ixa Group, 649 Posta kutxa, 20080 Donostia (Basque Country)

Tel: +34 943 015 019 Fax: +34 943 015 590

Website: <http://ixa.si.ehu.es/ixa>

Contact: Eneko Agirre

Email: e.agirre@ehu.es

Number of positions: 48

Research projects

The IXA Group was created with the aim of promoting the modernization of Basque by means of developing basic computational resources for it. As a result of our work four applications are currently available for common users: a spelling checker, a lemmatization based web-crawler, a lemmatization based on-line bilingual dictionary and a generator of weather reports.

The development of language technology for minority languages differs in several aspects from its development for widely used languages. The high capacity and computational power of present computers, combined with the scarcity of human and linguistic resources motivates the design of new and different strategies.

Human Language Technologies will make an essential contribution to the success of the information society, but most of the working applications are only available in English. For those working with minority languages, a great effort is needed to face this challenge.

MEANING: developing multilingual web-scale language technologies.

HERMES: news databases: cross-lingual information retrieval and semantic extraction.

XUXENG: Design and implementation of a prototype of grammar checker for Basque.

Laboratory for Research and Development in Artificial Intelligence

LIDIA. Department of Computer Science Faculty of Informatics University of A Coruña, 15071.

A Coruña, Spain

Tel: 81 16 70 00 Fax: 981 16 71 60

Website: <http://www.dc.fi.udc.es/lidia/lidia.html>

Contact: Amparo Alonzo Betanzos

Email: ciamparo@udc.es

Number of positions: 20

Research projects

- Projects are being conducted in the following areas :
- Knowledge Systems and Artificial Intelligence;
- Symbolic Processing Models, Knowledge Representation and Uncertainty;
- Data mining.

Natural Language Processing and Information Retrieval Group at National Distance Learning University (UNED)

Dpto. De Lenguajes y Sistemas Informaticos, ETSI Informatica, UNED, C/ Juan del Rosal, 16 Ciudad Universitaria, 28040 Madrid

Tel: (+34) 91 398 7983 Fax: (+34) 91 398 6535

Website: <http://nlp.uned.es/index.php>

Contact: Felisa Verdejo Maillo

Email: felisa@lsi.uned.es

Research projects

Intelligent Information Access

Multilingual Information Access.

- Foreign-language search assistants for Document Retrieval, Question Answering and Information Synthesis tasks.
- Multilingual website search engines.
- Multilingual Phrase browsing and Terminology Retrieval.
- Shallow and efficient NLP for Information Access
- Intelligent organization, visualization and browsing of search results
- Evaluation of Information Access systems.
- Co-organizers of the Cross-Language Evaluation Forum (CLEF).
- Co-organizers of the Interactive Cross-Language Evaluation Forum (iCLEF).
- Co-organizers of the CLEF Cross-Language Question Answering track.

Natural Language processing Group, Technical university of Catalonia

C/ Jordi Girona, 1-3, edif. Omega, despatx 315, Barcelona, Spain 08034

Tel: +34 93 413 7856 Fax: +34 93 413 7833

Website: <http://www.lsi.upc.edu/~nlp/>

Contact: Núria Castell

Email: castell@lsi.upc.edu

Number of positions: 31

Research projects

CHIL: Computers in the Human Interaction Loop.

HOPS: Enabling an Intelligent Natural Language Based Hub for the Deployment of Advanced Semantically Enriched Multi-channel Mass-scale Online Public Services.

Meaning: Developing Multilingual Web-scale Language Technologies.

LC-STAR: Lexica and Corpora for Speech to Speech Translation Components.

FAME: Facilitating Agent in Multiculture Exchange.

NAMIC: News Agencies Multilingual Information Categorization.

EUROWORDNET: Building a multilingual wordnet with semantic relations between words

OpenTrad: Traducción Automática de Código Abierto para las lenguas del Estado Español

HERMES: Electronic Libraries with Multilingual Information Retrieval and Semantic Processing.

PETRA: Speech Interfaces for Advanced Unified Messaging Applications.

BASURDE: Spontaneous-Speech Dialogue System In Limited Domains.

ITEM: Textual Information Retrieval in a multilingual environment using NL Techniques.

RILE: Spanish Linguistic Resources Server.

Speech Processing Group (VEU), Technical University of Catalonia

Department of Signal Theory and Communications Building D5 - Campus Nord C/Jordi Girona, 1-3 08034 Barcelona

Tel: +34 93 401 64 40 Fax: +34 93 401 64 47

Website: <http://gps-tsc.upc.es/veu/>

Contact: José B. Mariño Acebal

Email: canton@gps.tsc.upc.edu

Number of positions: 28

Research projects

The Speech Processing Group (VEU) works on language engineering and speech processing. VEU main objective is to develop oral communication systems and its different research areas: speech coding, robust speech features extraction, text-to-speech conversion, speech recognition, speech-to-speech translation and oral dialogue. The main interest of this research field is to develop systems that will aim to overcome the linguistic barriers and enhance the information systems accessibility.

The Research Group on Signals, Telematic and Communications, University of Granada

Dpto. Teoría de la Señal, Telemática y Comunicaciones, ETSI Informática, Universidad de Granada, 18071 – Granada

Tel: 958-243283 Fax: 958-243230

Website: http://ceres.ugr.es/info_eng/

Contact: José Carlos Segura Luna

Email: segura@ugr.es

Number of positions: 15, unknown number of students assistants

Research projects

The Research Group on Signals, Telematic and Communications (Grupo de Investigación en Señales, Telemática y Comunicaciones, GSTC) is composed of fourteen researchers from the Department of Electronics and Computer Technology (DETC) of the University of Granada (UGR).

The group's members have the B.Sc. Degree in Physics (Electronics specialty) or in Computer Science. Twelve of them have a Ph.D. and the rest are Ph.D. students, taking part in the Doctorate Programme *Multimedia Technologies*. They also teach on the fields of Signal Processing and Coding, Data Communications and Transmission, Computer Networks, Robotics, and Electronic Communications.

The current research efforts are specially focussed on the following fields:

- Signals Processing;
- Continuous speech recognition in noisy environment;
- Text-to-speech Conversion;
- Speech Coding and Compression.

Organismes et associations

Catalan Association of Artificial Intelligence
<p>Asociació Catalana d'Intel·ligència Artificial Institut d'Investigació en Intel·ligència Artificial (IIIA) Campus UAB 08193 Bellaterra, Catalonia</p> <p>Tel: +34 93 580 9570 Fax: +34 93 580 9661</p> <p>Website: http://www.acia.org</p> <p>Contact: Carles Sierra</p> <p>Email: sierra@iiia.csic.es</p>
Mission
<p>The Catalan Association of Artificial Intelligence (Associació Catalana d'Intel·ligència Artificial) is a non profit association for the advancement of Artificial Intelligence (AI) in Catalan society. The main goal of ACIA is to support communication between the persons and organizations involved in AI and promote social, cultural, scientific, economic and governmental awareness on AI.</p>

États-Unis

Laboratoires

Brain and cognitive sciences, Rochester University
Meliora Hall, Box 270268, University of Rochester, Rochester, NY 14627-0268 Tel: (585) 275-1844 Fax: (585) 442-9216 Website: http://www.bcs.rochester.edu/ Contact: Kathy Corser Email: kcorser@bcs.rochester.edu Number of positions: 58
Research projects
A constellation of faculty and students studies the <i>nature of language processing and acquisition</i> . This research spans a broad set of topics, ranging from the perception of speech sounds and words, the production and comprehension of spoken and written language, and the acquisition of spoken languages by young children and adults, to investigations of the structure, processing, and acquisition of signed languages. The Center for Language Sciences, an interdisciplinary research unit associated with the department of Brain and Cognitive Sciences, brings together our own faculty, fellows and students with language scientists from the departments of Linguistics, Computer Science, and Philosophy.

Cambridge research Lab

Hewlett-Packard, HP Labs, one Comabridge Center, Cambridge, Massachusetts 02142-1612

Tel: (617) 551-7600 Fax: (617) 551-7650

Website: <http://www.hpl.hp.com/research/crl/>

Contact: Richard Zippel

Email: webmaster@hpl.hp.com

Number of positions: 35

Research projects

The Cambridge Research Laboratory (CRL) was founded in 1987 to advance the state of the art in core computing and human-computer interaction. Today our research includes activities in capture, coding, storage, indexing, retrieval, decoding and rendering of multimedia data.

Large-scale multimedia indexing - building technologies to allow easy access, navigation and organization of vast repositories of multimedia (audio, images, video). This work includes SpeechBot, which brings the power of web search to audio and video content.

Center for advanced information processing

CAIP Center at Rutgers University, CORE Building 706, Busch Campus, 96 Frelinghuysen Rd. Piscataway, NJ 08855-1390

Tel: (732) 445-3443

Website: <http://www.caip.rutgers.edu/>

Contact: David Gaut

Email: daut@ece.rutgers.edu

Number of positions: 20

Research projects

CAIP is an Advanced Technology Center (ATC) established in early 1985 and jointly supported by the New Jersey Commission on Science and Technology, Rutgers University, and industry membership.

Current Researchs

- VRLAB: Human-Machine Interface Laboratory (Burdea);
- Computational Vision Laboratory (Dana);
- Speech and Language Processing Laboratory (Dusan, Flanagan, Rabiner);
- Fluid Mechanics of Speech Sound Production (Krane, Wei);
- Layered Manufacturing Technology (Langrana);
- Warning and Indicator Systems Engineering Laboratory (Mammone);
- Multimedia Information System Laboratory (Marsic, Flanagan);
- RIUL: Robust Image Understanding Laboratory (Meer);
- BioRobotics Laboratory (Bouzit);
- TASSL: The Applied Software Systems Lab (Parashar);
- Human Computer Interaction (Tremaine).

Center for automated learning and discovery

Carnegie Mellon University, School of Computer Science, 5000 Forbes Avenue, Wean Hall 4609, Pittsburgh, PA 15213-3891

Tel: (412) 268-1299 Fax: (412) 268-3431

Website: <http://www.cald.cs.cmu.edu/about/index.html>

Contact: Tom Mitchell

Email: Tom.Mitchell@cmu.edu

Number of positions: 50

Research projects

The Center for Automated Learning and Discovery (CALD) is an academic department within Carnegie Mellon University's School of Computer Science. We focus on research and education in all areas of statistical machine learning.

What is Machine Learning Machine Learning is a scientific field addressing the question "How can we program systems to automatically learn and to improve with experience?" We study learning from many kinds of experience, such as learning to predict which medical patients will respond to which treatments, by analyzing experience captured in databases of online medical records. We also study mobile robots that learn how to successfully navigate based on experience they gather from sensors as they roam their environment, and computer aids for scientific discovery that combine initial scientific hypotheses with new experimental data to automatically produce refined scientific hypotheses that better fit observed data.

To tackle these problems we develop algorithms that discover general conjectures and knowledge from specific data and experience, based on sound statistical and computational principles. We also develop theories of learning processes that characterize the fundamental nature of the computations and experience sufficient for successful learning in machines and in humans.

CALD Research CALD has an energetic research effort focused on developing new statistical learning algorithms of general use, new foundational theories of learning and learnability, and new learning applications in areas ranging from robotics, to human brain imaging, to natural language text interpretation.

Human Language Technology Research institute

Department of Computer Science, University of Texas at Dallas, Richardson, TX 75083-0688

Tel: (972) 883-4654 Fax: (972) 883-2349

Website: <http://www.hlt.utdallas.edu/page.php?p=index>

Contact: Sandra Harabagiu

Email: sanda@hlt.utdallas.edu

Number of positions: 41

Research projects

The main goal of the Human Language Technology Research Institute (HLTRI) is to stimulate and foster research in the area of Human Language Technology, comprising Natural Language Processing (NLP) and Automatic Speech Recognition and Synthesis, and to increase the visibility and reputation of UTD as a place of excellence in HLT research. At HLTRI, we are set to build a first class research group by tackling some of the most important research problems in Human Language Technology. The HLTRI is part of the Erik Jonsson School of Engineering and Computer Science. The HLTRI provides a rich environment in which graduate students learn and work with faculty and technical staff on supported research for both government and corporate funders. The Institute comprises five language technology related research centers.

The Center for Basic Research in Natural Language Processing has the goal of developing tools and basic research in lexico-semantics, syntactic and semantic processing of texts and dialogues, in discourse processing as well as development of cognitive models of language processing.

The Center for Emerging Natural Language Applications has the goal to develop research techniques and to study the emerging areas of natural language applications, e.g. textual question answering, multi-document summarization, text mining and open-domain information extraction.

The Center for Search Engines and Web Technologies has the goal of developing and applying new methods for locating information on the Internet's World Wide Web.

The Center for Text Mining is involved in the analysis of general human text sources such as newspaper articles, web pages, usenet articles, and other publications. The goal is to build better machine knowledge bases from human text.

The InterVoice Center was made possible by a generous gift from InterVoice, Inc. Research in the InterVoice Center focuses on the integration of Syntactic and Semantic knowledge with current Automatic Speech Recognition techniques.

Center for Human-Computer Communication

CHCC, OGI School of Science & Engineering, 20000 NW Walker Rd., Beaverton, OR 97006

Tel: (503) 748-1248 Fax: (503) 748-1875

Website: <http://www.cse.ogi.edu/CHCC/>

Contact: Thad Davidson

Email: davisth@cse.ogi.edu

Number of positions: 22

Research projects

- Designing Next-generation Mobile Interfaces for Dynamic Conversational Speech.
- New Directions in Adaptive Processing for Multimodal Systems.
- Multimodal Interaction for Field and Mobile Environments.
- Multimodal Interaction for Wearable Augmented Reality.
- Multimodal Interaction for Virtual Environments.
- Robust Agent-based Systems Incorporating Teams of Communicating Agents (AAA, Agent-Talk, STAPLE).
- Interactive Multimodal Interfaces: Designing for Human PerformanceMultimodal.

Center for intelligent information retrieval

CIIR, Computer Science Department, University of Massachusetts, 140 Governors Drive, Amherst, MA 01003-9264

Tel: (413) 545-0463 Fax: (413) 545-1789

Website: <http://ciir.cs.umass.edu/>

Contact: W. Bruce Croft

Email: croft@cs.umass.edu

Research projects

The Center for Intelligent Information Retrieval, a National Science Foundation-created S/IUCRC Center, is one of the leading information retrieval research labs in the world. The CIIR develops tools that provide effective and efficient access to large, heterogeneous, distributed, text and multimedia databases. CIIR accomplishments include significant research advances in the areas of distributed information retrieval, information filtering, topic detection, multimedia indexing and retrieval, document image processing, terabyte collections, data mining, summarization, resource discovery, interfaces and visualization, and cross-lingual information retrieval. The Center for Intelligent Information Retrieval continues to support the emerging information infrastructure, both through research and technology transfer. The goal of the CIIR is to develop tools that provide effective and efficient access to large, heterogeneous, distributed, text and multimedia databases.

AQUAINT Project : This research project, Relevance Models and Answer Granularity for Question Answering, is an ARDA initiative under the Advanced Question and Answering for Intelligence (AQUAINT) program.

NSF Digital Government Project : This research project, A Language-Modeling Approach to Metadata for Cross-Database Linkage and Search, is a National Science Foundation sponsored initiative. The CIIR is working in collaboration with Carnegie Mellon University, the Library of Congress, Department of Commerce, U.S. Geological Survey, and R.I.S.C. on this project.

NSF IDM Mongrel Project: "Supporting Effective Access through User- and Topic-Based Language Models" is a research project in collaboration with Rutgers University and sponsored by the NSF IDM program.

NSF NSDL - Search and Browsing Support for NSDL On an NSF National Science, Mathematics, Engineering, and Technology Education Digital Library (NSDL) project, the CIIR worked with a team of institutions that are developing the technical capabilities and executing the organizational responsibilities of the core integration of the NSDL Program.

Online Music Retrieval and Search Project The OMRAS research project is a Digital Libraries II sponsored program under the international digital libraries initiative. The CIIR is working in collaboration with Kings College of London and Indiana University on this project.



Center for language and speech processing

Barton Hall, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218

Tel: (410) 516-4237 Fax: (410) 516-5050

Website: <http://www.clsp.jhu.edu/index.shtml>

Contact: Frederick Jelinek

Email: jelinek@jhu.edu

Number of positions: 34

Research projects

Its aim is to promote research and education in the science and technology of language and speech. Research is conducted by faculty, research scientists, and graduate students affiliated with six associated academic departments: biomedical engineering, cognitive science, computer science, electrical and computer engineering, mathematical sciences, and psychology. The research involves work in all aspects of the science and technology of language and speech, with fundamental studies under way in areas such as language modeling, natural language processing, neural auditory processing, acoustic processing, optimality theory, and language acquisition.

Projects

- PIRE: Investigation of Meaning Representations in Language Understanding for Machine Translation Systems Frederick Jelinek, Eugene Charniak, Jason Eisner, Mark Johnson, and Sanjeev Khudanpur.
- MALACH: Multilingual Access to Large Spoken Archives? William Byrne and Frederick Jelinek
- Clustering of Words and Class-Based Language Models? Ahmad Emami and Frederick Jelinek.
- Word Clustering Based on Latent Semantic Analysis? Yan Huang and Sanjeev Khudanpur.
- Exploiting Nonlocal and Syntactic Word Relationships in Language Models for Conversational Speech Recognition (STIMULATE)? Peng Xu and Frederick Jelinek.
- Acoustic Modeling for Conversational Speech Shankar Kumar, Asela Gunawardana, Vaibhava Goel, Veera Venkataramani, William Byrne.
- Minimum Risk Acoustic Clustering for Multilingual Acoustic Model Combination and Maximum Mutual Information Estimation of HMMs Stavros Tsakalidis and William Byrne.
- Pronunciation Modeling for Mandarin Causal Speech Veera Venkataramani and William Byrne.

Center for machine translation

LTI Research, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213

Tel: (412) 268-5480 Fax: (412) 268-5578

Website: <http://www.lti.cs.cmu.edu/Research/CMT-home.html>

Contact: Céline Carraux

Email: celine@cs.cmu.edu

Number of positions: 20

Research projects

There are currently a number of active projects at the LTI, in the fields of machine translation, speech, information retrieval, knowledge acquisition, educational applications and other related projects in SCS/CMU.

- Babylon: Two-way speech-to-speech translation on a handheld computer.
- JANUS Speech-to-Speech Translation of Spontaneous Conversational Dialogs in multiple languages primarily using an interlingua based approach.
- KANT Knowledge-based Machine Translation The KANT project, Knowledge-based, Accurate Translation for technical documentation .
- LingWear aims to develop wearable language support .
- RADD MT Rapidly-Adaptable, Data-Driven Machine Translation.
- CAMMIA: Dialogue Management System Multi-tasking dialog management system that can be used with automatic speech recognition and VoiceXML to provide mobile information access.
- Fluency Foreign language accent correction.
- FestVox: Building Synthetic Voices.
- LET'S GO!: A Spoken Dialog System For the General Public.
- Towards Communicating with Dolphins This project applies aspects of speech technology and machine learning to aid communication with dolphins.
- SUBLIME Speech and Language Based Information Management Environment.
- Speech Graffiti ? Universal Speech Interfaces.
- Adaptive Information Filtering Automatically monitoring a stream of documents (e.g., news stories, news groups, etc) to find just those stories that are interesting to you.

Center for spoken language research

CSLR, university of Colorado, 1777 Exposition Drive, Boulder, Colorado 80301

Tel: (303) 735-5150 Fax: (303) 735-5072

Website: <http://cslr.colorado.edu/index.html>

Contact: Catherine Boyer

Email: boyer@cslr.colorado.edu

Number of positions: 23

Research projects

Diverse Suite of NLP Demos Several natural language processing (NLP) tools have been developed at CSLR in research projects funded by ARDA and NSF. The "Diverse Suite of NLP Demos" introduces a number of those tools. The goal is to illustrate how various NLP technologies are working.

SONIC: Large Vocabulary Speech Recognition System SONIC is a large vocabulary speech recognition system used in a number of research projects at CSLR. The software is being made available for non-commercial use. Executables are provided for Linux (kernel 2.2/2.4), MS Windows, Sun Solaris, and Mac OS X.

NSF-ITR Reading Tutor Project The NSF-ITR Reading Tutor Project is a multi-university collaborative research effort to develop intelligent animated conversational agents-life like computer characters capable of natural face-to-face conversational interaction in specific task domains. Developing these agents requires basic research leading to new and improved language technologies in areas of computer speech recognition, natural language understanding, dialogue modeling, visual recognition of facial movements, expressions and gestures, and generation of natural and expressive auditory and visual speech by three-dimensional animated characters. The research advances and systems developed during the project will be evaluated in the context of language training applications that serve as test beds for research. Learning programs will be developed to enable English- and Spanish-speaking children with speech and reading difficulties to interact with animated agents to learn to speak, understand and read language.

CU Communicator Project CU Communicator is a conversational dialogue system that interacts with users over the telephone to provide information on air travel, hotel and rental car reservations. This system serves as a test bed for our research to improve the state of the art in spoken dialogue systems. This project is funded by the Defense Advanced Research Projects Agency (DARPA).

Center for Study of Language and Information

Stanford University, CSLI, Ventura Hall, 220 Panama Street, Stanford, CA, 94305-4101

Tel: (650) 723-0758 Fax: (650) 723-0758

Website: <http://www-csli.stanford.edu/>

Contact: Byron Reeves

Email: reeves@stanford.edu

Number of positions: 72

Research projects

CSLI is an interinstitutional laboratory. Founded in 1983 by researchers from Stanford, SRI International, and Xerox PARC, CSLI has since its inception promoted collaboration between industrial laboratories and academic departments. In recent years, this collaboration has expanded to include researchers from additional universities, laboratories, and companies, both within the immediate geographical vicinity and around the world. CSLI's Industrial Affiliates Program currently includes twenty corporate members, many of which send researchers to participate in research projects on site.

Interface Laboratory : These projects constitute a concerted effort at CSLI to address human/computer interface problems emerging in the world of telecommunications, information processing, and consumer electronics.

Research in the Cognitive Sciences : These interdisciplinary projects cover a broad range of topics in computer science, linguistics, logic and semantics, philosophy, psychology, and education.

Center of Excellence for Document Analysis and Recognition

CEDAR, State University of New York at Buffalo, 520 Lee Entrance, Suite 202, Amherst, NY 14228

Tel: (716) 645-6162 Fax: (716) 645-6176

Website: <http://www.cedar.buffalo.edu/>

Contact: Sargur N. Srihari

Email: srihari@cedar.buffalo.edu

Number of positions: 55

Research projects

From a computer science research and education perspective the work spans the areas of pattern recognition, machine learning, data mining, information retrieval and computational linguistics. There is a continuum from the analysis of scanned document images to areas such as text mining and information retrieval. There is also a continuum from analyzing documents for forensic purposes to the analysis of fingerprint images and other areas of biometrics.

Document analysis and recognition research has been ongoing at the University at Buffalo since 1978. The start of this program was a research grant received by Sargur Srihari from the National Science Foundation on "Contextual Algorithms for Text Recognition" in 1980. This led to the establishment of CEDAR in 1991. Areas of current research include: machine learning; information retrieval, information extraction; optical handwriting recognition; fingerprint identification; data mining, and more.

Computational Linguistics and Information Processing Laboratory

CLIP, University of Maryland, UMIACS, 3357 AV Williams Building, College Park, MD 20742

Tel: (301) 405-6444 Fax: (301) 314-2644

Website: <http://www.umiacs.umd.edu/research/CLIP/clip.htm>

Contact: Bonnie J. Dorr

Email: bonnie@cs.umd.edu

Number of positions: 32

Research projects

CLIP consists of two main entities: The Natural Language Group and the Database Group. The natural language group focuses on several areas of broadscale multilingual processing, e.g., machine translation, scalable translangual document detection, and cross-language information retrieval. The database group focuses on architectures for wide area computation with heterogeneous information servers, e.g., scientific discovery from biomolecular data sources.

- Computational Methods in Linguistics
- Computational Psycholinguistics Linguist's Search Engine Modeling Historical Change.
- Human Language Technology
- Divergence Unraveling (DUSTer) GALE Generation Heavy Hybrid MT (GHMT) Headline Generation Interlingual Annotation MALACH Statistical MT TIDES.
- Information Processing, Query Optimization, and Web Semantics
- Related Projects in Collaborating Laboratories
- Human-Computer Dialog Quality of Data (QoD) Metadata for WebSources Semantic Web.

Computational Linguistics at Stanford University

CLSU, Linguistics Dept, Bldg 460, 450 Serra Mall, Stanford University, Stanford, CA 94305-2150.

Tel: (650) 723-0924 Fax: (650) 723-5666

Website: <http://www.stanford.edu/dept/linguistics/compling/>

Contact: Dan Jurafsky

Email: jurafsky@stanford.edu

Number of positions: 21

Research projects

The computational linguistics program at Stanford is one of the oldest in the country, and offers a wide range of courses and research opportunities. It takes a very broad view of computational linguistics, encompassing both theoretical and applied studies, and covering diverse linguistic areas from computational phonology, morphology, and syntax, to computational semantics and pragmatics. Among the theoretical foci are models for parsing and learning grammatical structure, models of communication, conversation, and dialogue, and computational psycholinguistic models of human language comprehension and production. CLSU also works on key applications in human language technology such as multi-party dialogue systems, machine translation, sentence understanding, probabilistic parsing and tagging, speech recognition, speech synthesis, biomedical information extraction, grammar induction, word sense disambiguation, and automatic question answering.

While the department includes a wide variety of research projects and teams studying different things, it shares the ambitious goal of dealing with the complexity and the uncertainty of human language by integrating rich models of linguistic structure at every linguistic level with sophisticated modern probabilistic and statistical techniques.

Together with the Computer Science Department, the department houses a wide variety of research labs, reading groups, and informal workshops on computational linguistics, and also maintains close ties with the wide variety of companies working on computational linguistic applications in the surrounding Silicon Valley area.

Columbia Natural Language Processing Group

450 Computer Science Building, 1214 Amsterdam Avenue, 0401 / New York, New York
10027-7003

Tel: (212) 939-7000 Fax: (212) 666-0140

Website: <http://www1.cs.columbia.edu/nlp/index.htm>

Contact: Sandra Morris

Email: swm16@columbia.edu

Number of positions: 48

Research projects

AQUAINT is a question answering project which focuses on the issues involved with answering complex questions.

The goal of **CLiMB** (Computational Linguistics for Metadata Building) is to explore and develop computer-assisted strategies for extracting item-level metadata -- i.e., cataloging suitable for use in large-scale digital library collection projects -- from the body of existing scholarly monographic literature relating to those collections.

MAGIC is an intelligent multimedia presentation system for the medical domain. After a patient has heart surgery, the physicians in the operating room (OR) must inform the caregivers in the intensive care unit (ICU) what happened during the surgery in order to prepare for the patient when he/she arrives in the ICU. MAGIC replaces the OR physicians in this scenario by presenting similar information using coordinated speech and graphics.

Newsblaster is a system that helps users find the news that is of the most interest to them. The system automatically collects, clusters, categorizes, and summarizes news from several sites on the web (CNN, Reuters, Fox News, etc.) on a daily basis, and it provides users a user-friendly interface to browse the results.

PERSIVAL (PErsonalized Retrieval and Summarization of Image, Video And Language resources) aims to provide personalized access to a distributed patient care digital library. PERSIVAL is a joint research initiative between the fields of NLP, human-computer interaction, medical informatics, video processing, library and cognitive science.

TIDES is DARPA's Translingual Information Detection, Extraction and Summarization project. It consists of a practical multidocument information tracking and summarization system. The system is intended to be portable across languages and we are in the process of extending the system to handle Arabic, Japanese and Chinese. Technology designed in TIDES is used everyday in Newsblaster.

Computing research laboratory

Science Hall Building, University of New Mexico, NM 88003-8001

Tel: (505) 646-5466 Fax: (505) 646-6218

Website: <http://crl.nmsu.edu/index.html>

Contact: James R. Cowie

Email: jcowie@crl.nmsu.edu

Number of positions: 24

Research projects

AWARE - Discovering characteristics of habitable answering systems The purpose of this project is to investigate how question answering systems can best present results to a user in such a way that the user is aware of what question the system has actually answered and how it arrived at the answer. As part of this project we are investigating how users interact with MIT's START question answering system and Language Computer Corporation's Internet Answer question answering system. In addition, we are examining how people use referring expressions when interacting with natural language systems.

Computer-Mediated Multilingual Collaboration This project involves evaluation of a multilingual chat system in two dimensions: How well the users can interact with the system to accomplish their tasks; The quality of the translations produced in multilingual chats. To test for the latter case, the system was tested on English, Chinese, Japanese, Spanish, and Korean languages.

Gene Literature Exploration Environment This is a project in collaboration with the University of New Mexico Cancer Research and Treatment Center. The goal of this project is to speed up the process of searching volumes of data for help in analyzing new genetic research.

Interlingual Annotation of Multilingual Text Corpora CRL has been awarded an NSF project in collaboration with UMIACS (Univ. of Maryland, College Park), ISI (Univ. of Southern California), LTI (Carnegie-Mellon University), Columbia University, and Mitre Corporation.

Meaning Oriented Question Answering is CRL's project in collaboration with CoGenTex in Ithaca, NY, and ILIT (Institute for Language and Information Technologies) at UMBC (University of Maryland, Baltimore County). The goal of this project is to bring to bear CRL's ONTOSEM ontology and TMR (Text Meaning Representation) to enhance both the accuracy and user-friendliness of question/answering systems.

Biography Generator This project involves developing a multilingual system to generate personal biographies. Currently the system is being tested on English, Spanish and Russian languages.

Department of Cognitive and Linguistic Sciences, Brown University

Department of Cognitive and Linguistic Sciences, Box 1978, Brown University, Providence, RI 02912-1978

Tel: (401) 863-2616 Fax: (401) 863-2255

Website: <http://www.cog.brown.edu/>

Email: coglingInfo@brown.edu.

Number of positions: 78

Research projects

The faculty includes people with backgrounds in cognitive psychology, theoretical and computational linguistics, psycholinguistics, and computational neuroscience; many affiliated and related faculty are found in departments such as Applied Mathematics, Computer Science, Neuroscience, Philosophy, and Psychology. This configuration of faculty offers unusually strong opportunities for within- and cross-disciplinary graduate study. For example, it is possible to combine formal training in theoretical linguistics with experimental research in psycholinguistics, neurolinguistics, or language acquisition. Students apply neural network modeling techniques to empirical findings in perception, cognition, language, and speech. The vision group provides training in computational, psychophysical, and ecological approaches to perception and action, a combination unavailable elsewhere. The speech group encompasses experimental, developmental, neurolinguistic, and evolutionary perspectives on speech perception and production.

Main research areas

- Acquisition of Language.
- Computational Linguistics (Natural Language Processing).
- Neurolinguistics.
- Speech Perception and Production.
- Syntax and Semantics.

Haskins Laboratories

Haskins Laboratories, Yale University, 300 George Street, suite 900, New haven, CT 06511

Tel: (203) 865-6163 Fax: (203) 865-8963

Website: <http://www.haskins.yale.edu/index.html>

Contact: Carol A. Fowler

Email: carol.fowler@haskins.yale.edu

Number of positions: 122

Research projects

Research at Haskins Laboratories contributes to the advancement of speech technology. The current group of research projects being carried out at the Laboratories are mutually supportive of one another and combine to form a comprehensive research program with a single unifying focus: human communication by speech and reading.

Nature and Acquisition of the Speech Code and Reading. The overall goal of this program is to understand how the language apparatus, biologically specialized for speaking and listening, becomes adapted to reading and writing.

Sources of Sound in Speech. Explores kinematic coordination of articulation and the regulation of air pressure in speech production.

Perception and Production of Rhythmic Microstructure. Focuses on sensorimotor synchronization and rhythm production.

Links Between Production and Perception in Speech. Examines the link between production and perception in speech, using articulatory data, modeling and synthesis.

Task Dynamics: Goals and Time in Speech Production. Expands the account provided by the task dynamic model of temporal aspects of speech production.

Imitation: A Tool for Studying Speech Perception. Uses the demonstrated tendency for adults to imitate the speech they hear as a tool for exploring the nature of speech perception.

Dynamic Patterns of Speech Aerodynamics in Children. Explores the development of laryngeal control in children between the ages of four and ten years.

Human-Computer Interaction Institute

Human-Computer Interaction Institute (HCII) School of Computer Science Carnegie Mellon University 5000 Forbes Avenue Pittsburgh, PA 15213-3891

Tel: (412) 268-3830 Fax: (412) 268-1266

Website: <http://www.hcii.cmu.edu/index.html>

Contact: Anne Watzman

Email: aw16+@andrew.cmu.edu

Number of positions: 148

Research projects

Carnegie Mellon has expertise in areas relevant to human-computer interaction, including user-interface software tools, cognitive models, speech recognition, natural language understanding, computer graphics, gesture recognition, data visualization, intelligent agents, visual interface design, multimedia, computer-supported cooperative work, computer music and drama, intelligent tutors, technical writing, and the organizational and social impact of technology.

The Human-Computer Interaction Institute carries on this tradition by fostering and carrying out multidisciplinary research and education in human-computer interaction. The scope of its work includes designing and testing of new tools and technologies to support human activity and organization, and building theory in the field. The research includes empirical and analytic studies of behavior, and studies of behavior in groups and organizations to inform the design and evaluation of new technologies. Students from supporting departments find a rich source of research opportunities in the HCII; and the Institute also offers a framework for research and development alliances with industry.

USC Natural Language Group

USC information Sciences Institute, 4676 Admiralty Way, Suite 1001, Marina Del Rey, CA 90292

Tel: (310) 448-8731 Fax: (310) 823-6714

Website: <http://www.isi.edu/natural-language/nlp-at-isi.html>

Contact: Eduard Hovy

Email: hovy@isi.edu

Number of positions: 30

Research projects

Adgen: Advanced Generation for Question Answering.

AGILE-MT: Human Informed Statistical Machine Translation.

AQ2-Time: Temporal Awareness Algorithms for Natural Language Text.

AQ2-Wordnet: Extending WordNet for Question-Answering.

EDC: Energy Data Collection.

eRule: Language Processing Technology for Electronic Rulemaking.

eRulemaking: Language Processing Technology for Electronic Rulemaking.

GRAMMAR: Advanced Language Modeling for Machine Translation.

GrowOnto: Information Discovery in Digital Government: Self-Extending Topic Maps and Ontologies.

ICT-Chatter: coming soon.

IL-Annot: Interlingua Annotation.

ISI-NL: Semantic analysis and generation for autonomous agents.

ITR-LM: Translation Technology for Multilingual Language Modeling.

Nitrogen: Hybrid Statistical-Symbolic Natural Language Generation.

OmegaOntology: coming soon.....

PDP: A Maximum Entropy-Based Approach to Discourse Parsing.

International Computer Science institute, Speech Group

International Computer Science Institute 1947 Center Street. Suite 600 Berkeley, CA 94704

Tel: (510) 666-2900 Fax: (510) 666-2956

Website: <http://www.icsi.berkeley.edu/about/index.html>

Contact: Nelson Morgan

Email: isci@berkeley.edu

Number of positions: 46

Research projects

EARS Rich Transcription of Conversational Speech In collaboration with SRI and the University of Washington, the Speech group is working to generate readable transcriptions of conversational speech in multiple languages incorporating capitalization, punctuation, and speaker markers.

EARS Novel Approaches ICSI researchers are studying replacements of the standard spectral envelope as the speech representation of choice (typically with cepstral transformation).

Speech Processing for Meetings ICSI researchers seek to develop algorithms and systems for the recognition of speech from meetings, as well as methods for information retrieval and other applications that such recognition would make possible.

Speaker Identification This project is concerned with the discovery of highly speaker-characteristic behaviors ("speaker performances") for use in speaker recognition and related speech technologies. The intention is to move beyond the usual low-level short-term spectral features which dominate speaker recognition systems today, instead focusing on higher-level sources of speaker information, including idiosyncratic word usage and pronunciation, prosodic patterns, and vocal gestures. ?

Clarissa Clarissa is a fully voice-operated procedure browser, enabling astronauts to be more efficient with their hands and eyes and to give full attention to the task while they navigate through the procedure using spoken commands.

Regulus Regulus is an Open Source toolkit for construction of spoken command grammars. Grammar development is carried out mainly using example-based methods, which greatly simplifies the process.

Speech Technology for Developing Countries ICSI researchers are developing speech recognition technologies for "emerging regions". As part of this effort, they have developed simple recognizers for Tamil, a language spoken by over 50 million people in Southeast India. The system is one example of ICSI's capability to rapidly design and deploy low-cost speech prototypes using openly available technology.

SmartWeb The SmartWeb project, which is being led by the German Institute for Artificial Intelligence (DFKI), deals with access to semantic Web services on mobile devices such as mobile phones. Speech input and output are well suited to mobile devices and will be a major focus of SmartWeb. A major part of the project vision is the ability for a user to ask a question using a mobile device and immediately receive an answer based on information drawn from the Web.

Linguistic Grammars Online

Stanford University, CSLI, Ventura Hall, 220 Panama Street, Stanford, CA, 94305-4101

Tel: (650) 723-0758 Fax: (650) 723-0758

Website: <http://www-csli.stanford.edu/>

Contact: Byron Reeves

Email: reeves@stanford.edu

Number of positions: 32

Research projects

The CSLI LinGO Lab is committed to the development of linguistically precise grammars based on the HPSG framework, and general-purpose tools for use in grammar engineering, profiling, parsing and generation. Early work in the CSLI LinGO Lab focused on the construction of a general-purpose grammar of English in the form of the English Resource Grammar (or ERG), and on further development of the LKB grammar engineering system. The LKB was also used at CSLI as the testbed for a number of teaching grammars and smaller-scale grammars for other languages including Japanese and Spanish.

The ERG is the cornerstone of ongoing research on multiword expressions (MWEs) jointly funded by the NSF and NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation. It is also being deployed in the construction of a novel, *rich* and *dynamic* treebank for HPSG, the LinGO Redwoods initiative. We are separately bringing together our grammar writing experience across a variety of languages in devising a Grammar Matrix to aid in the development of broad-coverage, precision, implemented grammars for natural languages.

The CSLI LinGO Lab is one of the founding members of the international Deep Linguistic Processing with HPSG Initiative (DELPH-IN), a network of academic and commercial research institutions that target applications of precise, linguistic NLP. HPSG and Minimal Recursion Semantics (MRS) provide the common scientific framework for DELPH-IN members, and many of the LinGO resources are part of the DELPH-IN open-source repository.

In past projects, the ERG was used in the Verbmobil machine translation system, and in an NSF-funded project on computer-aided speech generation for people who cannot speak because of disability.

Maryland Information and Network Dynamics Lab (MIND)

Maryland information and network Dynamics, 8400 Baltimore Avenue, College Park, Maryland 20740

Tel: (301) 405-2525 Fax: (301) 405-7662

Website: <http://www.mindlab.umd.edu/index.shtml>

Contact: Ashok K. Agrawala

Email: agrawala@cs.umd.edu

Number of positions: 31

Research projects

The University of Maryland established the Maryland Information and Network Dynamics (MIND) Lab to provide a focal point for defining, developing, evaluating and deploying new information technologies.

The MIND Lab provides support for collaborative efforts and operational experimentation on systems of reasonable sizes by deploying research infrastructure and 'testbeds' that can be used for addressing the problems of scalable distributed heterogeneous systems.

The MIND Lab is proud to host the Semantic Web Agents Group, or MINDSWAP. The MINDSWAP group, headed by Dr. James Hendler, was formed in 2002 to investigate cutting edge technologies for the next generation of the world wide web. Some of their work includes ontology visualization and creation, photo markup, reasoning, and distributed trust systems.

MIT Media Laboratory

The Media Laboratory, Building E15, 77 Massachusetts Avenue, Cambridge, MA 02139-4307

Tel: (617) 253-0365

Website: <http://www.media.mit.edu/>

Contact: Alexandra Kahn

Number of positions: 140

Research projects

The Media Laboratory vision of "enabling technology for learning and expression by people and machines" emphasizes technologies that improve the quality of life in the digital age, and that assist people in constructing their own tools for expression. The Lab advocates a process that includes both imagination and realization, criticism and reflection.

Research at the Media Lab comprises interconnected developments in an unusual range of disciplines. Lab research is now exploring new frontiers, such as wireless, "viral" communications; wearable computing; machines capable of common-sense reasoning; new forms of artistic expression; and how children learn. These themes outline a future where the bits of the digital realm interact seamlessly with the atoms of our physical world, and where our machines not only respond to our commands, but also understand our emotions—a future where digital innovation becomes the domain of all.

Research Cluster on Language, Center for Cognitive Science, Rutgers State University

Rutgers Center for Cognitive Science (RuCCS), Center for Cognitive Science & Lab of Vision Research, Psych Bldg Addition, Busch Campus, Rutgers University - New Brunswick , 152 Frelinghuysen Road, Piscataway, NJ 08854-8020

Tel: (732) 445-0635 Fax: (732) 445-6715

Website: http://ruccs.rutgers.edu/ruccs/research_language.html

Contact: Sue Cosentino

Email: suecos@ruccs.rutgers.edu

Number of positions: 61

Research projects

The language group seeks to uncover the grammatical, computational, psychological and neural limitations on language comprehension that interact to constrain the architecture of the language processing system. Bruce Tesar (Linguistics, RuCCS) has adapted dynamic programming-based parsing techniques to work with grammatical theories based on Optimality Theory. Tesar's algorithms simultaneously inform our understanding of language processing and language learning, as the parsing algorithms play an important role in Tesar's models of language learning. Jerry Fodor (Philosophy, RuCCS) has argued that certain formal properties are necessary for any computational architecture underlying human linguistic processing, providing further constraints on proposed models. Empirical studies provide not only the behavioral data which the group seeks to explain, but also yield crucial evidence about the processing architecture and its limitations, and the cognitive structures underlying this knowledge. Karin Stromswold (Psychology, RuCCS) and Fodor contribute from this perspective by conducting psycholinguistic experiments that investigate the structure of lexical, syntactic, and semantic knowledge. Stromswold further explores the properties of the language processing architecture through the use of functional neuroimaging techniques (PET and MRI).

Robust Speech Processing Group Center for Spoken Language Research

Robust Speech Processing Group Center for Spoken Language Research, University of Colorado Boulder, Campus Box 594, Boulder, Colorado 80309-0594

Tel: (303) 735-5148 Fax: (303) 735-5072

Website: <http://cslr.colorado.edu/rspl/>

Contact: John H.L. Hansen

Email: John.Hansen@colorado.edu

Number of positions: 24

Research projects

- Robustness Issues in Phrase Spotting, Topic Search.
- Accent Classification and Dialect Analysis for Speech Recognition.
- Speech Analysis for Vocal Fold or Tract Pathology.
- Speech Enhancement / Speech Quality Assessment for normal/hearing impaired listeners.
- Recognition of Speech under Noisy, Stressful, and/or Lombard Effect conditions.
- Robustness issues in Human Computer Interfaces, Hands-Free Cellular, Voice Coding: GSM, CELP, etc.
- Speech Synthesis and Speaker Normalization.
- Classification, Assessment, and Synthesis of Speech Under Stress.

The Signal, Speech and Language Interpretation Lab at the University of Washington

Department of Electrical Engineering University of Washington Box 352500 Seattle, WA 98195-2500

Tel: (206) 543-2150 Fax: (206) 543-3842

Website: <http://ssli.ee.washington.edu/ssli/>

Contact: Mary Ostendorf

Email: mo@ee.washington.edu

Number of positions: 33

Research projects

Domain-dependent speech synthesis - combining prosody prediction (phrasing and emphasis) with unit selection, gradient models of prominence, adaptive intonation models.

Acoustic and pronunciation modeling for LVCSR - use of syllable and higher level linguistic structure, alternatives to HMMs.

Language modeling for LVCSR - dynamic dialog-dependent mixture models, training from sparse data.

Low-power and efficient software algorithms - optimize at the software level to reduce power consumption for speech systems running on portable devices.

Statistical modeling for information extraction from speech - identification of named entities in errorful data using phrase language models.

Speech Analysis and Interpretation Laboratory, University of Southern California

University of Southern California, 3740 McClintock Avenue, Room EEB430, Los Angeles, CA 90089-2564

Tel: (213) 740-6432 Fax: (213) 740-4651

Website: <http://sail.usc.edu/index.php>

Contact: Shrikanth S. Narayanan

Email: shri@sipi.usc.edu

Number of positions: 26

Research projects

This research group is devoted to theoretical issues and practical applications of:

- Speech and Language Processing & Automatic Speech Recognition
- Speech Production Modeling, Articulatory Acoustics, Speech Synthesis
- Biomedical Signal Processing and Modeling; Imaging & Instrumentation for Speech Research
- Human Machine Interaction; Multimodal-Multimedia Interfaces, Devices, and Systems.

Human-like Speech Processing New directions in speech recognition through a Multidisciplinary University Research Initiative (MURI).

Dynamics of Vocal Tract Shaping Novel methods for studying human speech production using realtime magnetic resonance imaging.

USC Transonic Solutions Enabling Human-Human Communication via Automated Two-way Speech-to-Speech Language Translation for critical domains such as medical triage.

Initial Target Languages: *English, Farsi.*

Content-based approach to Indexing, Query and Retrieval of Music Through signal processing and knowledge-based methods.

Spoken Language Interactions in Preschoolers: bridging linguistic, social, and technological challenges.

Robust Speech Processing for Immersive Interactions.

Automated Synthesis of Expressive Speech for Military Training.

Speech Technology and Research Laboratory

SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025-3493

Tel: (650) 859-2388 Fax: (650) 859-5984

Website: <http://www.speech.sri.com/>

Contact: Kristin Precoda

Email: precoda@speech.sri.com

Number of positions: 26

Research projects

- Speech in Noisy Environments (SPINE).
- Prosody for Dialog Systems.
- Large Vocabulary Conversational Speech Recognition.
- Meeting Recognition and Understanding.
- Hidden Word-Level Events in Spontaneous Speech.
- Information Extraction from Speech.
- The SRI Communicator.
- Speaker Recognition.
- Mapping MeetingsData Collection and Annotation.

Spoken Language Systems

Spoken Language Systems, MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar Street 32-G434, Cambridge, MA 02139 USA

Tel: (617) 253-3049 Fax: (617) 258-8642

Website: <http://groups.csail.mit.edu/sls//sls-blue-noflash.shtml>

Contact: Marcia Davidson

Email: marcia@csail.mit.edu

Number of positions: 25

Research projects

Since its formation in 1989, the Spoken Language Systems (SLS) group has focused its research on the creation of technology that enables humans to interact with computers using natural spoken language. In pursuit of this goal the SLS group is actively engaged in a number of research projects. Some of the current projects are:

- Conversational Systems for Human/Computer Interaction.
- Multilingual Research & Language Learning.
- Multimodal Processing and Interaction.
- Human Language Technology and Education.
- Rapid Conversational System Prototyping.
- Ubiquitous, Human-Centered Computing.

Core Technology Development

To support its research on spoken language systems for human/computer interaction, the SLS group has developed its own suite of core speech technologies. These technologies include:

- Speech recognition (SUMMIT).
- Natural language understanding (TINA).
- Dialogue modeling.
- Language generation (GENESIS).
- Speech synthesis (ENVOICE).

The core speech technologies can be integrated to create conversational systems using the GALAXY architecture for conversational speech systems. The creation of GALAXY has enabled developers to rapidly create conversational systems for a wide variety of applications.

Organismes de financement et associations

American association of artificial intelligence

American Association for Artificial Intelligence, 445 Burgess Drive, Suite 100, Menlo Park, California 94025

Tel: (650) 328-3123 Fax: (650) 321-4457

Website: <http://www.aaai.org/>

Contact: Carol McKenna Hamilton

Mission

American Association for Artificial Intelligence 445 Burgess Drive Suite 100 Menlo Park, California 94025

Founded in 1979, the American Association for Artificial Intelligence (AAAI) is a nonprofit scientific society devoted to advancing the scientific understanding of the mechanisms underlying thought and intelligent behavior and their embodiment in machines. One of AAAI's main fields of study is natural language processing. AAAI also aims to increase public understanding of artificial intelligence, improve the teaching and training of AI practitioners, and provide guidance for research planners and funders concerning the importance and potential of current AI developments and future directions.

Major AAAI activities include organizing and sponsoring conferences, symposia, and workshops, publishing a quarterly magazine for all members, publishing books, proceedings, and reports, and awarding grants, scholarships, and other honors.

Applied Voice Input/Output Society (AVIOS)

Applied Voice Input/Output Society (AVIOS), P.O. Box 20817 Sa Jose, CA 95160

Tel: (408) 323-1783 Fax: (408) 323-1782

Website: <http://www.avios.com/>

Contact: Peggie Johnson

Email: info@avios.org

Mission

AVIOS is a not-for-profit professional membership organization founded in 1981. For most of the years since then, the AVIOS annual conference has been the only forum dedicated to practical applications of advanced speech technology (going beyond basic research). The conference recently matured into AVIOS-SpeechTEK Spring Conference though a partnership with AmComm.

Having established this strong base, AVIOS is now emphasizing its role as an international professional membership organization.

Mission

AVIOS is a professional society dedicated to promoting the development and diffusion of real world applications of speech technology. AVIOS creates linkages between users, developers, and researchers to advance speech technology as the interface of choice. AVIOS provides opportunities for members to learn more about market and technical developments and to exchange informal communications. Through AVIOS, researchers discover what developers, marketers, and end users require in speech-automated products and services. Developers and marketers learn how to identify appropriate applications for speech technology and how to present them.

Association for machine translation of America

Priscilla Rasmussen, AMTA Focal Point, 3 Landmark Center, East Stroudsburg, PA 18301

Tel: (570) 476-8006 Fax: (570) 476-0860

Website: <http://www.amtaweb.org/>

Contact: Priscilla Rasmussen

Email: focalpoint@amtaweb.org

Mission

AMTA is an association dedicated to anyone interested in the translation of languages using computers in some way. This includes people with translation needs, commercial system developers, researchers, sponsors, and people studying, evaluating, and understanding the science of machine translation (MT) and educating the public on important scientific techniques and principles involved. It provides an opportunity for commercial developers, users--both professional translators and casual MT users--and research scientists to meet in a convivial atmosphere and share ideas, dreams, and hopes.

AMTA has members in Canada, Latin America, and the United States. It is the regional component of a worldwide network headed by the International Association for Machine Translation (IAMT). The two sister organizations are the Asian-Pacific Association for Machine Translation (AAMT) and the European Association for Machine Translation (EAMT). The IAMT Council comprises a balanced representation of all the regions. Individuals, institutions, and corporations may join any of the three regional associations and automatically become part of the IAMT network.

AMTA facilitates access by researchers to machine-readable corpora and cooperation on the exchange of formats and text-encoding conventions, as well as discussions about establishing reference criteria for the evaluation of the technology.

AMTA members are engaged in developing training materials and programs, and in maintaining lists of the latest technology and innovations available.

IEEE Signal Processing Society

IEEE Signal Processing Society, 445 Hoes Lane, Piscataway, NJ, 08854

Tel: (732) 562-3888 Fax: (732) 235-1627

Website: <http://www.ieee.org/organizations/society/sp/>

Contact: Alfred O. Hero

Email: sp.info@ieee.org

Mission

The Signal Processing Society's Field of Interest shall be the theory and application of filtering, coding, transmitting, estimating, detecting, analyzing, recognizing, synthesizing, recording, and reproducing signals by digital or analog devices or techniques. The term "signal" includes audio, video, speech, image, communication, geophysical, sonar, radar, medical, musical, and other signals.

The Signal Processing Society is an international organization whose purpose is to: advance and disseminate state-of-the-art scientific information and resources; educate the signal processing community; and provide a venue for people to interact and exchange ideas.

The Signal Processing Society is a dynamic organization that is the preeminent source of signal processing information and resources to a global community. We do this by: being a one-stop source of signal processing resources; providing a variety of high quality resources to a variety of users in formats customized to their interests; adapting to a rapidly changing technical community; and being intimately involved in the education of signal processing professionals at all levels.

Linguistic data Consortium

Linguistic Data Consortium, 3600 Market Street, Suite 810, Philadelphia, PA, 19104-2653

Tel: (215) 898-0464 Fax: (215) 573-2175

Website: <http://www ldc.upenn.edu/>

Contact: Mark Liberman

Email: myl@ldc.upenn.edu

Mission

The Linguistic Data Consortium is an open consortium of universities, companies and government research laboratories. It creates, collects and distributes speech and text databases, lexicons, and other resources for research and development purposes. The University of Pennsylvania is the LDC's host institution.

The LDC is involved in a number of projects to support language education, research and technology development.

ACE - In support of the Automatic Content Extraction Program, LDC develops text corpora in English, Chinese and Arabic annotated for entities, the relations among them and the events in which they participate.

GALE - LDC is developing integrated linguistic resources and related infrastructure to support language exploitation technologies within the DARPA GALE (Global Autonomous Language Exploitation) Program.

LCTL (Less Commonly Taught Languages) - LDC is creating and distributing linguistic resources including monolingual and parallel text, lexicons, encoding converters, word and sentence segmenters, morphological analyzers, named entity taggers, annotations, annotation infrastructure and specifications for a number of "less commonly taught languages". The focus in Year One of LCTL is on seven languages that have more than a million speakers but do not possess adequate resources for building human language technologies: Urdu, Thai, Hungarian, Bengali, Punjabi, Tamil and Yoruba.

TIDES - LDC is collecting up to an hour per day in each of the languages in which Voice of America broadcasts to support this DARPA project in Translingual Information Detection Extraction and Summarization. The following are subprojects in TIDES.

Extraction - Corpus creation to support extraction of entities, relations and events from text as a TIDES technology and in collaboration with the Automatic Content Extraction program.

Retrieval Group

100 Bureau Drive, Stop 8940, Gaithersburg, MD 20899-8940

Website: <http://www.itl.nist.gov/iaui/894.02/>

Contact: Ellen M. Voorhees

Email: ellen.voorhees@nist.gov

Mission

The Retrieval Group of the Information Access Division works with industry, academia and other government agencies to promote the use of more effective and efficient techniques for manipulating (largely) unstructured textual information, especially the browsing, searching, and presentation of that information.

Activities (grouped into 4 major areas):

- Encourage retrieval research involving large, unstructured text files by providing test collections and organizing the TREC conference and its proceedings
- Continue to create new test collections, focusing mainly on collections to support specific information retrieval sub-tasks such as cross-language retrieval and multimedia retrieval
- Develop better evaluation methodology for information access, including improved evaluation measures for comparing systems using test collections and new evaluation measures for interactive searching and browsing operations
- Enable faster development of prototype (commercial, academic and government) retrieval systems by distribution of a basic state-of-the-art search engine (the PRISE system) that includes implementation of a Z39.50 client/server.

Open Mind Initiative

Website: <http://www.openmind.org>

Contact: David G. Stork

Email: stork@openmind.org

Research projects

The Open Mind Initiative is a collaborative framework for developing "intelligent" software using the internet. Based on the traditional open source method, it supports domain experts (who provide algorithms), tool developers (who provides software infrastructure and tools) and non-specialist "netizens" (who contribute raw data).

After many decades of research, there are still very many tasks for which computers are far worse than humans: recognizing speech, reading printed or handwritten text, recognizing objects from their image, understanding scenes, making complex plans, summarizing a story, and so on. It is clear that if such software could ever approach human performance, it would be extremely useful. Even if such software merely enhanced or aided human cognition, it would be useful. Much early research in "artificial intelligence" concentrated on small "toy" problems, such as game playing. There is a growing realization that we now need information contained in very large data sets.

The principal goal of Open Mind is to develop "intelligent" software, in part by collecting such large data sets and providing an open infrastructure where different ideas can be tried out. The data and resulting software are made available to all.

Another goal is to educate the public as to the problems in cognitive science, computer science, pattern recognition and related fields.

Finlande

Laboratoires

From data to Knowledge research unit
<p>Department of Computer Science, P.O. Box 68, (Gustaf Hällströmin katu 2b), FIN-00014, UNIVERSITY OF HELSINKI</p> <p>Tel: +358 9 191 51280 Fax: +358 9 191 51120</p> <p>Website: http://www.cs.helsinki.fi/research/fdk/</p> <p>Contact: Esko Ukkonen</p> <p>Email: Esko.Ukkonen@cs.Helsinki.FI</p> <p>Number of positions: 58</p>
Research projects
<p>The From Data To Knowledge (FDK) research unit develops methods for forming useful knowledge from large masses of data. The unit operates in a multidisciplinary fashion, integrating in its research groups excellence in discrete algorithms, statistical techniques and application sciences.</p> <p>The major methodological tools of the research unit are combinatorial pattern matching and data mining. The combination of these two is unique in the world. The work combines conceptual advances, algorithmic, statistical and analytical methods, and empirical work: theory and practice go hand in hand.</p>

Helsinki Institute for Information Technology, Basic Research Unit

Helsinki Institute for Information Technology (HIIT), Basic Research Unit (BRU), PO Box 68 (Gustaf Hällströmin katu 2b), FI-00014 University of Helsinki, FINLAND

Tel: +358 0 1911 Fax: +358 9 191 51120

Website: http://www.cs.helsinki.fi/hiit_bru/

Contact: Ukkonen Esko

Email: Esko.Ukkonen@cs.helsinki.fi

Number of positions: 60

Research projects

The Basic Research Unit (BRU) focuses in long-term basic research in computer science. The mission of BRU is to do basic research in computer science in areas in which there are applications visible, either in other sciences or in industry.

The present research groups work on data analysis, adaptive computing systems and neuroinformatics. The unit works in close collaboration with application groups in other sciences (such as medical genetics, structural biology and neuroscience) and in industry. The Basic Research Unit was established in the beginning of 2002.

Laboratory of Acoustics and Audio Signal Processing

Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing,
Otakaari 5 A, 02150 Espoo, Finland

Tel: +358-9-451-2496 Fax: +358-9-460-224

Website: <http://www.acoustics.hut.fi/>

Contact: Paavo Alku

Email: Paavo.Alku@hut.fi

Number of positions: 35

Research projects

- Analysis and parameterization of voice production.
- Cerebral processing of speech sounds.
- Predictive algorithms for spectral modeling of speech.
- EU project: Improving Human Potential (IHP) research training network for Hearing. Organisation and Recognition of Speech in Europe (HOARSE).
- Kieliteknologian opetuksen verkosto (KIT)(National Language Technology network).
- Graduate School of Language Technology.
- Acoustical analysis of emotion in speech.
- Artificial Bandwidth Expansion of Telephone Speech.
- Acoustic analysis of occupational dysphonia.
- Parametric models of speech.
- STT Speech to Text (phonemic speech recognition).
- Auroral Acoustics.
- Avesound - automatic recognition of bird species by sounds they produce.

Brain functions evoked by speech and 3D sounds

Measurement of brain responses to different speech signals and spatial signals.

Emotions in Speech

Detection and modelling of the emotions embedded in the speaker voice quality.

Occupational Voice Research

Occupational voice research investigates voice characteristics of speech professionals with objective measures in different circumstances.

Voice Source Parametrisation

The parametrisation of features in glottal activity, especially in the glottal flow.

Multimodal Interfaces Group, Adaptive Informatics Research center

Helsinki University of Technology Laboratory of Computer and Information Science P.O. Box
5400 FI-02015 HUT FINLAND

Tel. +358-9-451 3267? fax +358-9-451 3277

Website: <http://www.cis.hut.fi/research/multimodal/>

Contact: Mikko Kurimo

Email: Mikko.Kurimo@hut.fi

Number of positions: 125

Research projects

Our research in the Multimodal Interfaces group aims at solving the problem of information explosion by fusing several of the available perceptual and user feedback modalities. The results will be applied in interactive interfaces, for example in content-based multimodal retrieval. The Multimodal Interfaces group coordinates our work on relevant subtopics, namely image retrieval, speech recognition, proactive information retrieval, and natural language processing.

Morpho project: the unsupervised discovery of morphemes and the morphology of a natural language as well as word segmentation. The discovered morphs (word segments) have been applied to improve Finnish speech recognition performance.

Semantic Computing Research Group

SCRG, Helsinki University of Technology (TKK) Media Technology Room 2540 P.O. Box 5500
FI-02015 TKKFinland

Tel: +358 9 451 3362 Fax: +358 9 451 3356

Website: <http://www.seco.tkk.fi/>

Contact: Eero Hyvönen

Email: eero.hyvonen@tkk.fi

Number of positions: 21

Research projects

Our research focus is on machine-processable semantics, i.e., we investigate techniques

For representing data and knowledge in such a way that machines can "understand" its meaning, and develop algorithmic methods for creating intelligent applications based on such representations.

This focus is related to many research and application fields of computer science, such as artificial intelligence and knowledge engineering, semantic web and web services, natural language processing, and information retrieval.

Our group is active at the Helsinki University of Technology, Media Technology, University of Helsinki, Dept. of Computer Science, and Helsinki Institute for Information Technology HIIT.

The Intelligent Web Services (IWebS) research project studies the possibilities of the Semantic Web and Web Services technologies in both annotating services and delivering relevant services to end-users.

Our goal is to create a basis for national Finnish ontologies on the Semantic Web. The work covers:

- research on ontology technology (uncertainty in ontologies, ontology versioning, distributed ontology development),
- creation of central national ontologies from controlled vocabularies, such as the Finnish Webthesaurus and MASA, the controlled vocabulary used in many Finnish Museums, and application of the results to a number of case problems.

We investigated in 2002-2004 from the information storage and retrieval viewpoints, how semantically rich data should be annotated with metadata in order to facilitate content based-information retrieval. Cultural artifact databases and photograph repositories are used as case studies.

France

Laboratories

Analyse et traitement informatique de la langue française (ATILF)

(Computerized Analysis and Processing of French)

ATILF - CNRS/Université Nancy 2, 44 avenue de la Libération, B.P. 30687, F 54063 NANCY CEDEX

Tel: +33 3 83 96 21 76 Fax: +33 3 83 97 24 56

Website: <http://www.atilf.fr/>

Contact: Jean-Marie Pierrel

Email: jean-marie.pierrel@atilf.fr

Number of positions: 94

Research Projects

The mixed research unit of ATILF (computerized analysis and processing of French) was established on January 1, 2001, resulting from the combination of the *Institut National de la Langue Française (INALF - CNRS)* (National Institute of French) and *LANDISCO (Langue Discours Cognition - Université Nancy 2 (Language, Discourse, Cognition))*. Associated with the *Département Sciences de l'Homme et de la Société du CNRS* (the CNRS's Department of Humanities and Society Sciences), it is also an interface laboratory with the *Département Sciences et Technologies de l'Information et de la Communication* (Department for Sciences, Information Technologies and Communications). It is a member of the CNRS Federation - *Institut de Linguistique Française et du consortium international* (Institute of French Linguistics and of the International Consortium) TEI (Text Encoding Initiative).

- Computational linguistics: activities of this orientation are based on defining, managing and using computational language resources when processing languages.
- Computational development of tools, bases and computerized resources, with a unified management of our XML-coded resources based on STELLA software developed at the laboratory, specifically the *Trésor de la Langue Française informatisé (TLFi)* (Computerized Treasure of French), as well as the text database *Frantext*.
- Language and metadata resources, maintenance and evolution with two main sites: lexicographical monitoring to ensure the required updating of the *Trésor de la Langue Française informatisé (TLFi)* (Computerized Treasure of French) and the enhancement of the text database *Frantext* according to scientific project requests.
- The automatic processing of language in order to offer resources (knowledge bases and analysis and processing tools) that can be integrated into language engineering chains.

***École nationale supérieure de télécommunications, Groupe
Information, Interaction, Intelligence***
**National Advanced School of Telecommunications, Information
Group, Interaction and Artificial Intelligence**

46 rue Barrault, 75634 Paris

Tel: +33 1 45 81 78 70 Fax: +33 1 45 81 31 19

Website: <http://www.infres.enst.fr/>

Contact: Nicolas Puech

Email: npuech@enst.fr

Number of positions: 20

Research Projects

Information processing systems have entered a phase where they are characterized by the need to process very large amounts of information and knowledge, which requires user-friendly and powerful interactions means, and intelligent automatic processing tools. In this context, the Group 3I, whose name is inspired from the GDR-PRC I3 that was established in 1997 and to which the group takes part, has focused its research activities on the following sectors:

Information: Search for information in large document bases; optimize access to multidimensional information; visualize large quantities of information.

Interaction: writing and speech recognition; human-machine dialogue; interacting with a virtual world.

Artificial Intelligence: this aspect characterizes qualitatively the processings mentioned above; the AI can be obtained either by techniques that are specific to the data handled, or by using a cognitive approach.

- Search and visualize information.
- Interact with information.
- Multidimensional, spatial and spatio-temporal databases.
- Corpus-based linguistics.
- Intelligence information
- Model language and communications strategies.

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur, Human-Machine Communication Department

LIMSI-CNRS BP 133, F-91403 ORSAY CEDEX (France)

Tel: +33 (0)1 69 85 80 80 Fax: +33 (0)1 69 85 80 88

Website: <http://www.limsi.fr/Recherche/CHMdpgb.html>

Contact: Philippe Tarroux

Email: philippe.tarroux@limsi.fr

Number of positions: 31

Research projects

The Human-Machine Communication Department gathers a set of competences which allows us to address the various communication modalities with different views. Those activities are related to spoken language processing (analysis, synthesis and perception), natural language processing (analysis, text comprehension and production, written dialog, semantic representations), computer vision, image analysis and synthesis, gesture analysis and synthesis, multimodal communication, human factors and studies of the cognitive processes. Collaborate to this department researchers from engineering sciences (computer science, signal processing and artificial intelligence) as well as from linguistics, cognitive psychology and specialists of socio-economics questions regarding to innovation.

Our research in **speech recognition** focuses on the segmentation and transcription of continuous speech as well as speaker and language identification, with the purpose of automatically annotating and structuring audio documents. Advances in speech recognition, which rely upon supporting research in acoustic-phonetic modeling, lexical modeling and language modeling, are undertaken in a multilingual context (English, French, German, Mandarin, Portuguese, Spanish, Arabic, etc.). The underlying approach aims to develop models and algorithms that jointly take into account the diverse sources of information so as to globally decode the audio signal.

Human-machine spoken dialog is a multi-faceted research area requiring us to model spontaneous speech, the communication process and to develop a dialog manager. This work has led to the realization of dialog systems for information access. The interaction can rely entirely on speech (via the telephone), or can be associated with other means of interaction, such as a touch screen (multimodal kiosk).

Laboratoire des sciences de l'information et des systèmes, INCOD
(Information and Systems Science Laboratories)

Domaine Universitaire de Saint-Jérôme, Avenue Escadrille Normandie-Niemen, 13397
MARSEILLE CEDEX 20

Tel: (33) 04 91 05 60 30 Fax: (33) 04 91 05 60 33

Website: <http://www.lsis.org>

Contact: Norbert Giambiasi

Email: norbert.giambiasi@lsis.org

Number of positions: 34

Research projects

The work of the INCOD team targets information and knowledge systems engineering, specifically the distributed, cooperative and security aspects. The team's research focuses on four major themes in which its researchers are involved in a non-proprietary way:

- Concept reuse of information systems;
- Model and simulation multi-agents to help decision-making;
- Search for and integrate heterogeneous information.

Laboratoire d'Automatique Documentaire et Linguistique

Automated Documents and Linguistics Laboratory

77454 Marne-la-Vallée, Université Marne-la-Vallée

Website: <http://ladl.univ-mlv.fr/>

Contact: M. Mathis

Email: mmathis@univ-mlv.fr

Number of positions: 20

Research projects

The language computing team focuses on processing natural language texts, emphasizing the use of precise and explicit data: dictionaries, grammars, with respect to approximating from uncertain data [Lap97b]. The applications targetted are numerous, but the most significant ones deal with information retrieval. The three aspects which currently interest us the most are the following:

Producing electronic dictionaries in languages other than those already possessing reliable tools;

Intermediate processings between lexicon and syntactical parsing to access information in large text bases: terminology recognition and indexation, ambiguity resolution, etc.

This work is based in the more general context of the RELEX linguistics and computational linguistics network that includes some twelve laboratories located mostly in Europe, which collaborate in building an inventory of precise and processable language information in machine processing, on the basis of methodological requirements: reproducibility, completeness and cumulativity. In this aggregate, our team plays a major role since its establishment in 1987 under the name of *Centre d'études et de recherches en informatique linguistique* (CERIL) (Centre of Studies and Research in Computational Linguistics), with the *Laboratoire d'automatique documentaire et linguistique* (LADL) of Paris 7, through the joint organization of conferences and thanks to cross-contributions when designing fundamental tools: dictionaries and software.

The RELEX network's global research program is simple. Physical language units are detailed in electronic dictionaries. As these units can be combined to form sentences and texts, the formal description of syntactical constraints is implemented in the form of lexicons-grammars. As the linguistic descriptions (grammatical, morphological, syntactical and phonetic) are meant to be used in computer applications, objectives are therefore very concrete: specify the forms that can be observed. The transition from language descriptions to computerized results is simplified by using finite state automatons for common support.

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur – Computer Sciences Laboratory for Mechanics and Engineering Sciences

LIMSI-CNRS BP 133, F-91403 ORSAY CEDEX (France)

Tel: (33) 1 69 85 80 85 Fax: (33) 1 69 85 80 88

Website: <http://www.limsi.fr/>

Contact: Patrick Le Quéré

Email: plq@limsi.fr

Number of positions: 180

Research projects

LIMSI is a research unit of the CNRS, associated with the *Pierre et Marie Curie (Paris-6)* and *Paris-Sud (Paris-11)* universities. This laboratory has some 120 permanent employees (researchers, teacher-researchers and ITA-IATOS) and some sixty doctorands. It leads multidisciplinary research projects in Mechanics, Energetics, Information Technology and Science and Communications, covering a wide disciplinary spectrum from thermodynamics to cognition, encompassing fluid mechanics, energetics, acoustic and voice synthesis, spoken language and text processing, vision, virtual reality. LIMSI strives to combine discipline excellence in all its scientific sectors and in developing multidisciplinary projects in partnership with other industrial and research organizations, be they French or foreign. It is specifically committed to training through research and to disseminating scientific and technical information.

Laboratoire de Linguistique Formelle
(Formal Linguistics Laboratory)

UMR 7110, Université de Paris 7, Tour centrale, Place Jussieu 75005 Paris

Tel: (+33)(0)1 57 27 57 94

Website: <http://www.llf.cnrs.fr/>

Contact: Alain Kihm

Email: alain.kihm@linguist.jussieu.fr

Number of positions: 67

Research projects

The *Laboratoire de Linguistique Formelle* (LLF) is a theoretical linguistics laboratory whose scientific mission is based on the systematic description of languages belonging to different topological families, on confronting separate linguistic theories, on interfaces between grammar components (syntax and phonology, syntax and morphology, syntax and semantics, syntax and pragmatics).

Every research operation (Morphophonemics and interfaces, Formal Syntax and semantics, Signifier units and Declaratory structures are linked to detailed descriptions of linguistic phenomena in a defined grammar framework. Empirical fields go from phonemes to utterances. The linguistic subsectors of the laboratory are the following: phonology, morphology, syntax, prosody, semantics (lexicon, sentence, discourse) and pragmatics (formal pragmatics and enunciation theory).

LORIA Laboratoire lorrain de recherche en informatique et ses applications (The Lorraine Laboratory of IT Research and its Applications)

LORIA - Campus Scientifique - BP 239 - 54506 Vandoeuvre-lès-Nancy Cedex

Tel: +33 3 83 59 30 00 Fax: +33 3 83 27 83 19

Website: <http://www.loria.fr/>

Contact: Yves Laprie

Email: yves.laprie@loria.fr

Number of positions: 450

Research projects

The scientific themes of the language and dialogue team are centred on models, techniques and resources for the short, medium and long-term implementation of man-machine dialogue systems with a significant language component. In this context, our activities are expanding in three complementary directions: the study of the fundamental mechanisms of communication in natural language, alone or accompanied with gestural designation (multimodal dialogue). This research is being undertaken in a cross-disciplinary context combining mainly linguistics and computer science; building in-use dialogue systems within industrial collaborations. This activity allows us to have a platform for experiments to validate the various models we design; the design of generic methods and tools for refined study of real dialogue situations obtained by transcribing experiments and simulations or by direct observation. This work is based on experience acquired over a number of years with standardization and manipulation of linguistic resources (in particular "corpus" resources).

The goal of the members of the Orpailleur group is to design knowledge-based systems for reasoning and solving problems in a given domain. Knowledge can have different forms and sources: it can be given explicitly by a specialist of the domain (an expert) or it may be extracted from various databases related to the domain. Knowledge is then used as a basis for reasoning purposes in order to solve problems in a given application domain. The Orpailleur research group is mainly interested in the following application domains: agriculture, analysis of scientific and technical texts, chemistry (organic synthesis planning), classification of temporal signals, medicine, iron and steel industry.

Laboratoire informatique d'Avignon

Avignon IT Laboratory

339 Chemin des Meinajariès – BP 1228 – 84911 AVIGNON Cedex 9

Tel: 04 90 84 35 09 Fax: 04 90 84 35 01

Website: <http://www.lia.univ-avignon.fr>

Contact: Renato De Mori

Email: marc.elbeze@lia.univ-avignon.fr

Number of positions: 66

Research projects

The major research orientations of the team are as follows:

- Design language models minimizing errors when generating hypotheses from an automated speech recognition system;
- Recognize speech automatically to access large electronic directories;
- Use dialogue and expanded queries in document research.

The LIA PAROLE (speech) team has undertaken work in speech and speaker recognition.

LIA Projects in Information Research

Question/Answer Engines: The SIAC research system can be used to find a zone within a document when answering a query, rather than looking through the entire document. This possibility, coupled with the LIA developed extraction module of given entities (F. Béchet) will allow the laboratory to take part in the next TREC -10 (2001) assessment campaign for question/answer engines (this problem will be addressed within the framework of F. Genet's thesis, co-directed by Messrs. El-Bèze, F. Béchet and P. Bellot).

Information Retrieval Engines. Lia intends to continue developing the SIAC system search engine to bring its performances up to the level of those obtained by the best systems at this time. To implement the principle that its systems must meet LIA established-criteria, we plan on taking part in future TREC and Amaryllis assessment campaigns.

Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen

(Caen IT, Image, Machine and Instrument Research Group)

6, Boulevard du Maréchal Juin, 14050 CAEN Cedex

Tel: +33 (0)2.31.56.74.35 Fax: +33 (0)2.31.56.73.30

Website: <http://www.greyc.unicaen.fr>

Contact: Bruno Cremilleux

Email: Bruno.Cremilleux@info.unicaen.fr

Number of positions: 147

Research projects

The "*Données, Document, Langue*" (Data, Document, Language) team is one of the seven GREYC CNRS UMR 6072 laboratory teams. It comprises 10 permanent employees, 17 doctorands and 1 research associate. The team aims to simplify the production, access and usage of electronic documents, and more generally data. Its work deals with, among others, geographic documents, language and knowledge extraction from databases. The team's scientific activities are structured around three themes (briefly described below). The team focuses on inter-theme projects as well as collaborations with other teams and laboratories. It initiates and develops collaborations between researchers stemming from different sectors and schools of thought around many facets including producing, representing, disseminating and using information.

The "*fouille de données et apprentissage*" (Data Search and Learning) group proposes interactive models to improve understanding of information contained in one document or in vast collections of documents. It is interested in both the methods (extraction algorithms) and multiple uses (e.g. data interpretation, class characterization, clustering) of obtaining results stemming from a knowledge-based extraction process.

Based on the premise that developing information and communications technologies leads to a paradox, as it makes the production and dissemination of electronic documents easier and easier, and yet renders information research and access more and more complex, the "*document électronique composite*" (Composite Electronic Document) Group studies the nature and representations of electronic documents, with a specific interest in geographic documents.

The "*langue pour le document*" (Document Language) Group is specifically interested in textual information, designing language models to research information, linked to more fundamental semantic work. The group's work comprises both theoretical work proposing computer models, specifically in semantics and rhetoric, and applicable aspects more directly linked to the automatic processing of languages, enabling us to consider linguistic material models.

***Groupe de traitement du langage parlé, Laboratoire d'Informatique
et de Mécanique pour les Sciences de l'Ingénieur***

**(Spoken Language Processing Group, IT and Mechanics
Laboratory for Engineering Sciences)**

LIMSI – CNRS, B.P. 133, 91403 ORSAY CEDEX, France

Tel: +33 1 69 85 80 80 Fax: +33 1 69 85 80 88

Website: <http://www.limsi.fr/tlp/>

Email: contact@limsi.fr

Number of positions: 30

Research projects

The research carried out in the Spoken Language Processing Group aims to understand speech communication processes and develop models to use in automatic speech processing. The problems we address concern modelling at multiple levels: acoustic, lexical, syntactical and semantic, as well as the communications process. This research area is inherently multidisciplinary, requiring competence in signal processing, acoustics, phonetics and phonology, linguistics and computer science. These research activities are validated by developing systems for automatic speech processing such as speech recognition, spoken language dialogue, and the indexation of audio and video documents.

- TC-STAR (Technology and Corpora for Speech to Speech Translation)
- CHIL (Computers in the Human Interaction Loop)
- HUMAINE: Network of Excellence (Emotions and human-machine interaction)
- EARS RT project (Rich transcription project)
- AUDIOSURF (RNTL) (Automatic Indexation and Intuitive Research in Audio Documents)
- STIC-SHS MIDL (Modelling to identify languages)
- AMITIES (Automated Multilingual Interaction with Information and Services)
- MEDIA-EVALDA (Technolanguage Project to assess dialogue in and out of context)
- Processing multilingual broadcast audio for information access
- LVCSR (Large Vocabulary Speech Recognition)

French Organizations and Associations

<p style="text-align: center;"><i>Association francophone de la communication parlée</i> (The Francophone Association of Spoken Communications)</p>
<p>AFCP, Université D'Avignon, 74 rue Louis Pasteur, 84029 Avignon cedex 1 Tel: +33 (0)4 90 16 25 00 Fax: +33 (0)4 90 16 25 10 Website: http://www.afcp-parole.org/ Contact: Jean-Francois Bonastre Email: contact-afcp@afcp-parole.org</p>
Mission
<p>The <i>Association Francophone de la Communication Parlée</i> (The Francophone Association of Spoken Communications) is a non-profit association (Act of July 1, 1901) devoted to supporting, developing, disseminating and promoting various spoken communications sciences disciplines in the Francophone community.</p> <p>Even though subjects may seem very diverse, few of them can be studied without calling upon several disciplines simultaneously: signal processing, probabilities, computer science, phonetics, linguistics and others. The association aims to bring together researchers in these disciplines, in order to foster exchanges and help train young researchers by organizing seminars and by publishing specialized work. A biennial meeting "Les Journées d'Etudes sur la Parole" (JEP) (Study Days on Speech) allows members to perform a synthesis of the community's work.</p> <p>To meet its objectives, the AFCP, more generally, employs the following means in its field of competence: fostering and facilitating scientific exchanges; organizing conferences, scientific days and exhibitions; organizing thematic courses, seminars and schools; publishing and helping to publish reviews and work; disseminating information to institutional stakeholders and the general public; promoting relations between the research community and industry; cooperating with other scientific organizations, as well as all actions that comply with its goals.</p>

Association pour la Recherche Cognitive
(The Cognitive Research Association)

ARCo, Université de Technologie de Compiègne, COSTECH, BP 319 - 60203 Compiègne
Cedex

Tel: 03 44 23 43 66 Fax: 03 44 23 52 12

Website: <http://www.utc.fr/arco/index.php>

Contact: Olivier Gapenne

Email: olivier.gapenne@utc.fr

Mission

The *Association pour la Recherche Cognitive* (ARCo) (Cognitive Research Association) was established in 1981, and aims to promote interdisciplinary research in cognition. It comprises a workgroup on the automatic processing of language semantics. Its members are active in fields linked to the humanities, engineering sciences and life sciences: cognitive psychology, linguistics, artificial intelligence, logic, neurosciences, sociology and philosophy.

Since its inception, the association actively takes part in debates that are essential to living and to the coherence of an interdisciplinary community like ours. It aims to maintain a plurality in disciplines: theoretical, experimental and applied approaches; schools, paradigms and themes. The association also strives to promote the integration of young or isolated researchers into our community; it also endeavours to bring about a European dimension to cognitive sciences. In recent years, scientific institutions have recognized Cognitive Sciences with specific research programs and DEAs. In this renewed landscape, the specific role of ARCo, in accordance with its association status governed by the Act 1901, is to be an open and transparent body now more than ever, in both scientific debates and its own administration. ARCo currently comprises some 350 public and private sector researchers.

Association pour le traitement automatique des langues
(The Association for the Automatic Processing of Languages)

ATALA? 45, rue d'Ulm 75230 PARIS CEDEX 5. - SIRET: 393 902 721 00017 - CCP 13 809
69 J PARIS

Website: <http://www.atala.org>

Contact: Jean Véronis

Email: atala@biomath.jussieu.fr

Mission

The ATALA acts as a centralizing organization that aims to promote the advancement of technologies related to the automatic processing of languages.

ATALA has been publishing since 1960 the international review *Traitement Automatique des Langues* (previously entitled *La Traduction Automatique* and then *TA Informations*), in collaboration with the CNRS. This review is now published three times per year, and is sent free of charge to members in good standing of ATALA. Its dissemination to non-members has been entrusted to Lavoisier, which has taken over operations of HERMÈS Science.

The TALN Conference was established in 1994 within the framework of the Natural Language Zone of GdR-PRC Human-Machine Communications. It was founded by Philippe Blache, who was responsible for the first three editions and who set the pattern. The organization of the TALN has been operating since 1997, and has been under the stewardship of ATALA since 1999. A Permanent Committee, created in 1998, gathers applications to organize future editions of the conference, and therefore ensure its long-term success. As of 1999, the RÉCITAL Conference (*Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues* – Meeting of IT Researcher Students to Automatically Process Languages) has been held as a student event of TALN.

Since 1960, ATALA has been organizing several Study Days every year that comprise a series of presentations, which can include “demonstrations”. They represent an opportunity for various stakeholders of the field (university members, industrialists, researchers and students) to meet and discuss. Each day is headed by an organizer, and has a theme which is devoted to a specific aspect of theoretical research or applications in Linguistic, IT or various fields of automatic processing of languages: human-machine communications, electronic dictionaries, automatic indexing, machine translation, speech recognition, speech synthesis, etc.

European Language Resources Association (ELRA)

55-57 rue Brillat Savarin, 75013 Paris, France

Tel: +33 1 43 13 33 33 Fax: +33 1 43 13 33 30

Website: <http://www.elra.info/fr/index.html>

Contact: Khalid Choukri

Email: choukri@elda.fr

Mission

ELRA (the European Language Resources Association) is the driving force to make available language resources for language engineering and to evaluate language engineering technologies.

In order to achieve this goal, ELRA is active in identifying, distributing, gathering, validating, standardizing, improving, promoting the production of language resources, supporting the infrastructure to perform evaluation campaigns and developing a scientific field of language resources and evaluation.

These activities are achieved through ELRA's operational body, ELDA (Evaluations & Language resources Distribution Agency).

ELDA is involved in a number of projects at the French, European and international levels. These projects address various issues related to Language Resources, i.e. LRs production, LRs validation, or LRs standardisation, as well as issues in relation with HLT Evaluation, through ELDA's participation in current campaigns or the set-up of new evaluation platforms.

ELDA is also very involved in projects promoting language technologies, and gathers facts and figures on the market.

Association Internationale de la communication parlée
(International Speech Communication Association)

Site Internet: <http://www.isca-speech.org>

Contact: Julia Hirschberg

Email: isca-president@isca-speech.org

Mission

ISCA is a non-profit organization. The purpose of the Association is to promote speech communications science and technology in an international world-wide context. The Association will achieve its purpose by:

- fostering scientific research and education;
- organizing conferences and exhibitions;
- organizing courses;
- publishing and promoting the dissemination of scientific papers;
- promoting the study of various languages;
- promoting scientific exchanges;
- collaborating with other existing organizations;
- promoting relations between research and industry;
- serving regional and global needs;

Réseau Francophone de l'Ingénierie de la Langue (FRANCIL)

(Francophone Network of Language Engineering)

Francil, LIMSI-CNRS, BP133 ? 91403 Orsay Cedex, France

Tel: (33) (1) 69-85-80-64/80-66 Fax: (33) (1) 69-85-80-88

Website: <http://www.limsi.fr/Recherche/FRANCIL/frcl.html>

Contact: Joseph Mariani

Email: mariani@limsi.fr

Number of positions:

Research projects

The Francophone Body for Higher Education and Research has implemented a program designed to reinforce its activities in Language Engineering, which is a part of electronic information processing. In this context, machine processing of language is a sector that is in full expansion, and comprises research and development regarding text parsing and generation as well as speech recognition, comprehension and synthesis. It includes applications on document management, human-machine communications, text editing and machine-aided translation. It also comprises industrial, economic, scientific and technological elements, in addition to a very specific cultural dimension. It is always preferable to master the language that is used to express the results of research, especially if this research is made on the language itself. Therefore, the French language and *La Francophonie* are very directly concerned with this issue.

The fundamental objectives of the FRANCIL Network are to:

Ensure the bond between Francophone researchers in a joint-development perspective;

- Help to disenclave researchers, laboratories and research teams of Francophone countries, specifically countries in the South;
- Foster cooperation between teams, university laboratories, *les grandes écoles* and major research centres, be they national, international, public or private;
- Contribute to the quality, influence and reputation of Francophone research;
- Foster science in French;
- Allow, as much as possible, researchers speaking French but not residing in Francophone countries to collaborate with their Francophone peers.

Technolangue Portal

ELDA (Agence pour l'évaluation et la distribution de ressources linguistiques) 55-57 rue Brilat-Savarin 75013 Paris

Tel: 01 43 13 33 33 Fax: 01 43 13 33 30

Site Internet: <http://www.technolangue.net>

Contact: Valérie Mapelli, ELDA

Email: info@technolangue.net

Mission

The Technolanguage Project concerns the field of computational linguistics (CL).

This field, also known as “Natural Language Processing” (NLP) “Language Engineering”, aims to provide technologies used in many applications processing written, spoken, unilingual or multilingual information.

Applications examples include machine translation, document parsing, information research and filtering, voice dictation, speech synthesis, human-machine dialogue in natural language, etc.

Technolangue is a joint project of the three technological research and innovation networks (RNRT, RNTL, RIAM), and is financed within an interdepartmental framework.

Its primary objective is to implement a permanent infrastructure to produce and disseminate language resources; to assess written and spoken language technologies; to participate in regional and international organizations of standardization and information monitoring in this field.

The Ministries concerned are *le Ministère chargé de la Recherche et des Nouvelles Technologies* (Research and New Technologies), *le Ministère délégué à l'Industrie* (Industry) et *le Ministère de la Culture et de la Communication* (Culture and Communications).

- The “Language Resources” Aspect aims to reinforce the infrastructure to produce, validate and disseminate language resources;
- The “Assessment” Aspect aims to create an assessment infrastructure that takes natural language processing technologies into account;
- The “Standards” Aspect aims to simplify access to information on standards developed for this field, as well as support French participation in their creation;
- The “Monitoring” Aspect aims to implement technology monitoring, collecting information on current events in language technologies and how they are used in various applications internationally.

Hong Kong

Laboratoires

Halliday Centre for Intelligent Applications of Language Studies

The Halliday Centre for Intelligent Applications of Language Studies, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

Tel: (852) 2788 8722 Fax: (852) 2788 8706

Website: hallidaycentre.cityu.edu.hk/

Contact: Dr. Jonathan Webster

Email: ctjjw@cityu.edu.hk

Number of positions: 27

Research projects

Metalinguistic modeling -- development of our metalinguistic resources for modelling language in context, with particular attention to the development of more explicit higher-level models that can support new computational implementations and new computational tools.

Inde

Laboratoires

AU-KBC Research Center, Madras Institute of Technology
<p>The AU-KBC Research Centre, Madras Institute of Technology, Anna University, Chromepet, Chennai (Madras)-600 044, India.</p> <p>Tel: (+91) 044 2223 2711/4885 Fax: (+91) 044 2223 1034</p> <p>Website: http://www.au-kbc.org/index.html</p> <p>Contact: M. Murali</p> <p>Email: mmurali@au-kbc.org</p> <p>Number of positions: 44</p>
Research projects
<p>The Natural Language Processing (NLP) group focuses on developing Tools, Technologies, Products and Systems to facilitate the use of computers and the Internet for day-to-day life. In the Indian context, finding ways of making e-information accessible non-English knowing persons constitutes a major aspect of such work. The group also works on building lexical resources such as dictionaries, WordNet and tagged corpora that are essential for researchers working on various areas of NLP. Both linguists and computer scientists are members of the NLP group.</p> <p>The group has several major long term research goals. These include:</p> <ul style="list-style-type: none">- Developing Machine Translation (MT) systems between Tamil and other languages particularly English and Hindi;- Building lexical resources in Tamil that are essential for researchers and developers- Developing basic tools for computational work in Tamil, such as morph analyzer, Part-Of-- Speech (POS) tagger etc.;- Application of NLP tools for Information Extraction from domain specific texts so as to build Information Extraction systems for various domains such as medicine, agriculture etc.;- The following is a gist of the achievements of the group in about two and half years since its inception in January 2000;- The Tamil-Hindi Machine Aided Translation (MAT) system has a performance in the range of 75%;- The group has to its credit a Tamil search engine - the first of its kind for any Indian language;- The state-of-the-art Tamil Morphological analyser can handle nearly 3.5 million word forms including compound words with more than 95% accuracy;- The named entity extraction module is a part of Information extraction for medical literature.

Center for Development of Advanced Computing

Centre for Development of Advanced Computing Pune University Campus Ganesh Khind
Pune - 411 007 India.

Tel: +91-20-2570-4100 Fax: +91-20-2569 4004

Website: <http://www.cdac.in/>

Contact: Shri S. Ramakrishnan

Research projects

C-DAC has been a pioneer in developing and proliferating the use of Indian languages on computers. This technology is now extended to include multimedia and multilingual computing solutions covering a wide range of applications such as publishing and printing, word processing, office application suites with language interfaces for popular third party softwares on various operating platforms, electronic mail, machine translation, language learning, video and television and multimedia content in Indian languages. These have been successfully commercialized.

The two groups of C-DAC involved in the development of multilingual technologies are:

Graphics and Intelligence based Script Technology (GIST) The GIST Group of CDAC facilitates use of Indian languages in IT. In its endeavor to stay abreast with technologies worldwide, GIST has been adopting the latest concepts to be able to stay tuned with the Internet enabled world.

Applied Artificial Intelligence (AAI)

The AAI Group of C-DAC is involved in a number of activities such as knowledge-based understanding systems, Machine Translation System, Information Extraction & Retrieval, Language Tutors, Web-based solutions include Chat Server, and many other strategic software developments and solutions.

Computer Vision and Pattern Recognition Unit, Indian Statistical Institute of Calcutta

COMPUTER VISION AND PATTERN RECOGNITION UNIT INDIAN STATISTICAL INSTITUTE? 203, B. T. ROAD, Calcutta – 700035 INDIA

Tel: (91) (33) 577 7694 Fax: (91) (33) 577 6680, 3035

Website: <http://www.isical.ac.in/~cvpr/>

Contact: Bidyut Baran Chaudhuri

Email: bbc@isical.ac.in

Number of positions: 20

Research projects

The major areas of on-going research in the Unit include Optical Character Recognition, Natural Language Processing, Pattern Recognition, Image Processing and Soft Computing tools.

The **CVPR** unit is proud of a pioneering development of an **Optical Character Recognition (OCR)** system for the combined script recognition of Bangla and Devnagari (Hindi), the most popular scripts of the Indian sub-continent. The system performs document tilt correction, script line, word and character segmentation, character recognition and finally, error correction. A major area of interest of this unit is the general problem of digital document analysis. Novel algorithms for skew estimation and correction have been developed. A system for the automatic detection and processing of mathematical expressions from documents is being designed. The issues of automatic layout analysis of both printed and handwritten documents involving segmentation, information and feature extraction, data compression, processing and understanding line drawings are under study while a system for the automatic extraction of meta-information from document images has already been developed.

There have been pioneering developments in the area of **Natural Language Processing** of Indian languages. A powerful Bangla spell-checker for automatic detection and correction of spelling errors in computer stored documents has been developed and extended to the spell checking of Hindi documents. Among other important developments are a talking dictionary and talking editor for Bangla text, a morphological processor for Bangla and a computer based Bangla speech synthesizer. One of the novel objectives is to develop computer aids for the handicapped such as the deaf and dumb as well as the blind. Among other studies, statistical analysis of a large corpus of Bangla text, analysis of human spelling error patterns as well as analysis of onomatopoeic Bangla words have been carried out. Work on developing a **tri-lingual computer dictionary** (Hindi, Bangla, English) is in full swing.

Irlande

Laboratoires

Digital Signal Processing, University College Dublin
Digital Signal Processing Research, Department of Electronic and Electrical Engineering, Engineering Building, Belfield, Dublin 4 Ireland. Tel: +353-1-7161909 Fax: +353-1-2830921 Website: http://dsp.ucd.ie/ Contact: Anthony Fagan Email: tony.fagan@ucd.ie Number of positions: 33
Research projects
<ul style="list-style-type: none">• Cross-Talk Cancellation for High Speed Data Communications over Twisted-Pair Copper;• Audio-Visual Person Verification and Identification;• Diagnosis of Vocal Fold Pathology using multi-modal signal analysis;• Audio-Visual Automatic Speech Recognition.

Information Management Group, Dublin College University

School of Computing, Dublin City University, Glasnevin, Dublin 9

Tel: +353 1 700 5984

Website: <http://www.computing.dcu.ie/research/information.html>

Contact: Stephen Blott

Email: Stephen.Blott@computing.dcu.ie

Number of positions: 29

Research projects

The Information Management Group has three major research themes: Digital Multimedia, Database Engineering and Interoperable Systems, and Educational Technologies. The Centre for Digital Video Processing (CDVP) researches and develops techniques and tools to automatically analyse and index digital video information, and allow content-based operations such as browsing, searching, and summarisation. The Database Engineering and Interoperable Systems researchers develop formal and informal models for constructing database systems, and construct interoperable layers between heterogeneous information systems. Information system engineering research focuses on models and methodologies for increasing the value of data in information-intensive and networked economies. The Educational Technologies research is currently focused on the Improvement of the use of ICT in higher education through Action Research.

Knowledge and Data Engineering Group, Trinity College Dublin

KDEG Department of Computer Science Trinity College Dublin

Tel: +353 1 6082091

Website: https://www.cs.tcd.ie/research_groups/kdeg/index.php

Contact: Vincent P. Wade

Email: Vincent.Wade@cs.tcd.ie

Number of positions: 31

Research projects

The group focuses on research issues and applications in Knowledge and Data Management. In particular the group is actively researching Adaptive (hypermedia) systems, intelligent management for communications and ubiquitous computing, semantic web and interoperability, web orchestration and business process automation.

In particular, KDEG has pioneered research in four focused application areas, namely:

- Telecommunications Services Management;
- Smart Space (Ubiquitous Computing) Management;
- Adaptive Hypermedia and eLearning;
- Health Informatics.

Language and Intelligence Group, Dublin City University

School of Computing, Dublin City University, Glasnevin, Dublin 9

Tel: +353 1 700 8929

Website: <http://www.computing.dcu.ie/research/language.html>

Contact: Dónal Fitzpatrick

Email: Dónal.Fitzpatrick@computing.dcu.ie

Number of positions:

Research projects

Speech and Language Processing, including Machine Translation, Speech Processing, Computational Models of Semantics, Treebanks, Formal Syntax, Digital Signal Processing, Computer-Assisted Language Learning, Probabilistic Natural Language Processing and Parsing.

Artificial Intelligence, including Artificial Minds, Computational Models of Cognition, Knowledge Representation, Human-Computer Interaction, Cognitive Science, The Origins of Intelligence, Neural Networks and Autonomous Agents.

National Center for Language Technologies

National Centre for Language Technology (NCLT), School of Computing, Dublin City University, Dublin 9, Ireland.

Tel: +353 - (0)1 - 7005074 Fax: +353 - (0)1 - 7005442

Website: <http://www.computing.dcu.ie/research/nclt/index.html>

Contact: Prof. Josef van Genabith

Email: nclt@computing.dcu.ie

Number of positions: 20

Research projects

Statistical and Rule-Based MT (SMT, RBMT), Example-Based MT (EBMT), Translation Memories (TMs), Boosting Existing MT Systems, Machine-Aided Translation (MAT), Computer-Aided Translation (CAT), Controlled Languages.

Integrating CL/NLP/HLT Technology into CALL, CALL for Endangered Languages, CALL for Primary School Environments, CALL for Remedial Learners.

Statistical and Rule-Based MT (SMT, RBMT), Example-Based MT (EBMT), Translation Memories (TMs), Boosting Existing MT Systems, Machine-Aided Translation (MAT), Computer-Aided Translation (CAT), Controlled Languages.

Automatic Feature-Structure Annotation Algorithms, Subcategorisation Frame Extraction, Wide-Coverage Robust Probabilistic Unification Grammar Acquisition, PCFG-Based LFG Approximation, HPSG Acquisition, Multilingual Treebank-Based Grammar Acquisition.

Discourse Representation Theory, Linear-Logic Based Semantics, Computation of Logical Forms from Treebanks, Open-Domain Question Answering Systems.

Speaker Characterisation, Audio Classification, Retrieval and Coding, Human Computer Interfaces (HCIs).

Multilingual Information Retrieval/Extraction.

TCD Computational Linguistics Group

Computational Linguistics Lab, O'Reilly Institute G27, Department of Computer Science, Trinity College, University of Dublin, Dublin 2 Ireland.

Tel: +353 (1) 608 2866 Fax: +353 (1) 677 2204

Website: <https://www.cs.tcd.ie/>

Contact: Carl Vogel

Email: carl.vogel@tcd.ie

Number of positions: 40

Research projects

The Computational Linguistics Group focuses on basic research on human language and reasoning. The group is interested in understanding and modeling human cognitive abilities, as well as in engineering applications that can be realized as the fruit of successful scientific research in cognitive science and philosophy of language.

Israël

Laboratoires

Department of Computer Science, Bar-Ilan University
Department of Computer Science, Bar-Ilan University, Ramat Gan, 52900, Israel. Tel: 972 - 3 - 531 - 8866 Fax: 972 - 3 - 736 - 0498 Website: http://www.cs.biu.ac.il/ Contact: Amihod Amir Email: amir@cs.biu.ac.il Number of positions: 41
Research projects
Machine Learning, Theory of Self-organization, Text classification. Data Mining, Text Mining, Machine Learning and Link Analysis. Multidimensional pattern matching as applied to text processing, computer vision and multimedia, computational biology.

Department of Computer Science, University of Haifa

Department of Computer Science, University of Haifa, Mount Carmel, 31905 Haifa, Israël

Tel: +972-4-8240259 Fax: +972-4-8249331

Website: <http://cs.haifa.ac.il/welcome.html>

Contact: Daphna Stern

Email: daphna@cs.haifa.ac.il

Number of positions: 44

Research projects

Active areas of research at the department include algorithms, computational complexity, discrete mathematics, numerical analysis, approximations, parallel computation, operating systems, compilers, computer graphics, computer vision, image processing, color vision, CAD, graph algorithms, artificial intelligence, neural networks, computational biology, computational linguistics and natural language processing.

Israël Institute of Technology, Computational Linguistics Group

Israel Institute of Technology, Haifa 32000, Israel

Tel: +972-4-829-3379

Website: <http://www.cs.technion.ac.il/~lcl/>

Contact: Ran El-Yaniv

Email: rani@cs.technion.ac.il

Number of positions: 35

Research projects

Probabilistic Morphological Analyzer for Hebrew Undotted Texts Morphological analysis of words in a text is the first stage of most natural language applications. Due to the rich morphology of the Hebrew language and the inadequacy of the undotted script which results in a great degree of morphological ambiguity, the problem has not yet found a satisfactory solution. We notice that the problem of morphological analysis of Hebrew texts is similar to the well-studied problem of part-of-speech tagging in English, and thus can apply some of the approaches used to solve that problem. The work results in a morphological analysis which is correct for about 96% of the words. This result approaches results reported for English probabilistic part-of-speech tagging. It does so by using a very small training corpus - 5000 words only, in contrast to million-word corpora used for English tagging.

Corpus Based Analysis of Hebrew The project's aim is to study ways to arrive at better tools for morphological and syntactic analysis of Hebrew using corpus based techniques. The project is part of a joint project conducted in collaboration with the corpus based NLP group of the Computer Science Institute of the Hebrew University, Jerusalem, headed by Prof. Eli Shamir.

Extensions and Implementations of Natural Logic: This project develops an inference system for natural language that is based on Natural Logic: a logic that works directly on syntactic representations with no intermediate translation to a more familiar logical formalism. - monotonic expressions in Categorical Grammar.

Semantics of Natural Language Temporal Questions and Interfaces to Temporal Database Systems: We are developing a natural language interface for temporal databases, based on a semantic treatment of temporal questions. Using the interface, the user may express questions in natural language and have them automatically translated and submitted to the database. The project is sponsored by FIRST, administered by the Israeli Academy of science.

Knowledge Center for Processing Hebrew

KCPH, Taub Building, Technion Israel Institute of Technology, Haifa 32000

Tel: +972-4-8294313 Fax: +972-4-8293900

Website: <http://www.mila.cs.technion.ac.il/>

Contact: Arnon Cohen

Email: mila@cs.technion.ac.il

Number of positions: 20

Research projects

The research and development activities at the center are lead by most of the Computational Linguistics researchers in Israel, in addition to several researchers from related research fields. It is our intention to establish a computational infrastructure for processing the Hebrew language in a variety of frontiers:

- A large repository of Modern Hebrew texts, which is partially morphologically and syntactically tagged.
- Standards for tagging and presentation of annotated texts in Hebrew.
- Multilingual semantic networks of words.
- Systems for analyzing syntax and morphology based upon linguistic rules and systems based on statistical models.
- Methods for performance evaluation of natural language processing systems for Hebrew.
- Software for classifying and ranking Hebrew texts.
- Computational Grammars that will provide the foundation for syntactic analysis.

Research students of all the principal researchers will also take part in the research and development. Furthermore, the resources that will be developed in the center will be made available for the academic institutes in Israel for teaching and research purposes. This way the center will also server as a catalyst for development of updated educational programs which will allow training of additional personnel in computational linguistics in Israël.

Italie

Laboratoires

Center for Scientific and Technological Research, Cognitive and Communication Technologies Division

ITC-irst Via Sommarive, 18 38050 Povo (Trento) Italy

Tel: (+39) 0461 314 444 Fax: (+39) 0461 302 040

Website: <http://www.itc.it/irst/Renderer.aspx?targetID=164>

Contact: Fabio Pianesi

Email: pianesi@itc.it

Number of positions: 20

Research projects

The Cognitive and Communication Technologies Division conducts research on linguistic processing in its endeavor to develop advanced interface and knowledge technologies. Currently under study are user-centered, multilingual, multimodal interfaces; and semantic, context-based knowledge technologies.

Research Goals

The objective of TCC is to develop an ideal, simplified model in which information stored in linguistic format is accessed, mined, and made available to people. This requires the study and development of technologies for:

- Human-human and human-computer interaction. Research emphasizes multimodality and multilinguality, as well as interaction with virtual and physical environments.
- Flexible information representation for end-users. This representation combines the contribution of different media, adapting to the user's characteristics as and the changing status of the interaction.

Automatic analysis of text such as documents, e-mails, and web-pages and automatic extraction of information from them.

Cognitive and Communication Technologies Division, ITC-Irst

ITC-irst Via Sommarive, 18 ? 38050 Povo (Trento) Italy

Tel: (+39) 0461 314 444 Fax: (+39) 0461 302 040

Website: <http://tcc.itc.it/>

Contact: Fabio Pianesi

Email: pianesi@itc.it

Number of positions: 47

Research projects

The Cognitive and Communication Technologies (TCC) division at ITC-Irst is a major European research group in areas such as Natural Language Processing, Human-Computer Interaction and Dialogue Systems, Multi-modality and Natural Language Generation, Production and Maintenance of Linguistic Resources, Linguistic Theory. Being part of ITC-Irst, TCC is a member of the European Network of Excellence in Natural Language and Speech (ELSNET).

Currently, our activities are organised into two main areas: b

- TEXTEC (Technologies for Text Processing) investigates oth basic and applied issues in text processing: ?Question Answering, Word Sense Disambiguation and Information Extraction. It also addresses the construction?and maintenance of linguistic resources: Lexica and Wordnets, Grammars, Corpora, etc.
- i3p (Intelligent and Interactive Information Presentation) is involved in the study and design of systems for presenting information to users, stressing such aspects as adaptivity and supports for interaction, with an emphasis on the study and optimisation of the interactions between language and other modalities (images,video, animations, etc.). Scenarios are varied, ranging from web- to mobile (PDA)-based information presentation.TCC is also active in the field of Machine Translation, and in particular speech-to-speech translation.

TCC is committed to both basic and applied research and has considerable ties with many major research centres all over the world. The division has regular and fruitful relationships with private companies, especially those operating in the Web world, telecommunications and multimedia publishing. In addition, TCC has a long record of participation in EU funded projects.

Department of computer and systems science, University of Rome

Via Eudossiana 18 00184 Roma

Tel: (39)-06-44585360 Fax: (39)-06-44585367

Website: <http://www.dis.uniroma1.it/research.html>

Contact: Tiziana Toni

Email: abhaya@ics.mq.edu.au

Number of positions: 127

Research projects

The Integrative Modelling, Simulation and Data Analysis group was established at DIS in 2004. The research takes advantage of methodologies coming from diverse disciplines such as large scale dynamical modelling, simulation and optimization. Current main research areas are: Embedded and Reactive Real-Time Systems in Automotive Management, Computational Optimization in System Biology, Image Processing and Computer Vision, Complex Algorithms and Data Mining.

DAMSO: The Laboratory of Data Analysis, Integrative Modelling, Simulation and Optimization is aimed at developing models and testing efficient algorithms for processing real world data coming from industrial and biosystems engineering. Person in charge: Professor Alberto De Santis. Location: DIS-Buonarroti.

Embedded and Reactive Real-Time Systems in Management. Embedded systems are electronic components integrated onto a physical plant. The most challenging embedded systems, called reactive real-time systems, control the behaviour of the plant so that models and control techniques as well as simulation tools are needed when the algorithms are implemented on a given HW-SW architecture.

Computational Optimization in Systems Biology. Computational optimization in systems biology is a research program that links biologists and engineers in a multi-disciplinary approach to the systematic analysis of large scale and complex dynamical biological phenomena. The main interest is on the analysis of the dynamic properties of interacting genes, metabolites and proteins in the cell.

Signal Processing and Computer Vision. Digital signal processing is concerned with the theoretical and practical aspects of representing and extracting information from real data; it has a significant impact in most of the scientific and technical areas of management sciences. Current applications of interest in the research group are within the areas of industrial engineering, man-machine interaction, medical technology, science of materials and mechanical structures fault diagnosis.

- *Cognitive Robotics and Reasoning about Actions*, i.e. the definition of methods and techniques for reasoning about actions, and for the design and the realization of embodied agents (in particular mobile robots) that are able to accomplish complex tasks in real environments.

Department of Information Engineering, University of Padova

Via Gradenigo 6/B, 35131, Padova, Italy

Tel: +39-049-827-7601 Fax: +39-049-827-7699

Website: <http://www.dei.unipd.it/wdyn/?IDsezione=2>

Contact: Paolo Tenti

Email: paolo.tenti@dei.unipd.it

Number of positions: 68

Research projects

Data retrieval : Information Retrieval (IR) is the science that addresses the efficient and effective gathering, indexing, organization and retrieval from billion document collections of all and only the multi-media data that are relevant to millions users' information needs. The activities of the research programme within this line of research are concerned with

- The retrieval of information contained in documents spread across networks: Networks provide access to collections of data which can be both structured and unstructured. In the latter case, if an IR service is available in a peer-to-peer (P2P) network, a user can search for relevant documents stored in peers connected to his own computer, mobile phone or PDA. A weighing model is studied and proposed to reduce network exploration and maximize retrieval performance.
- Evaluation of cross-language information retrieval: In this programme, stemming has been addressed in a multilingual context. Because of the presence of different languages in the documents of the collections, stemming implies a more complex work than in a classical mono-lingual context. Two language independent stemming methodologies – one based on Hidden Markov Models, another based on link-analysis -- have been developed to build stemmers without a-priori linguistic knowledge on the language morphology, but inferring it directly from the collection of documents.
- Automatic link generation, analysis and mining: As knowledge is also between and not only inside entities, the activities are concerned with the transformation of a set of documents into a hypertext, add semantic content Web links, provide effective retrieval, and mine graph topology from Web sites. Algorithms for link generation and analysis have been designed and experimented.

Artificial intelligence: The research focuses on two specific areas of artificial intelligence, namely temporal reasoning and natural language processing. In the area of temporal reasoning, the research aims at the development of automated systems capable of representing and reasoning about temporal knowledge in presence of uncertainty and vagueness. New models of integration of quantitative and qualitative temporal information affected by vagueness and uncertainty are investigated, based on the FCSP (Fuzzy Constraint Satisfaction Problem) paradigm. Realized systems are being applied to planning and scheduling applications, and to the medical domain for the recognition of diseases from the temporal patterns of patient symptoms. In the area of natural language processing, the research is oriented toward the design of statistical models and the development of specialized algorithms for the automatic syntactic analysis of natural language.

European Academy Bolzano, Institute for Specialised Communication and Multilingualism

EURAC research Viale Druso 1, 39100 Bozen/Bolzano - Italy

Tel: +39 0471 055 055 Fax: +39 0471 055 059

Website: <http://www.eurac.edu/Org/LanguageLaw/Multilingualism/index.htm>

International Directory of Language Technologies – March 2006

Contact: Marcello Soffritti

© Language Technologies Research Centre

Email: msoffritti@eurac.edu

Institute of Cognitive Sciences and Technologies

ISTC, V.le.Marx 00137 Rome Italy

Tel: +39 06 860 90 260 Fax: +39 06 82 47 37 15

Website: <http://www.istc.cnr.it/>

Contact: Albano Leoni

Email: leoni@ip.rm.cnr.it

Number of positions: 92

Research projects

The Institute of Cognitive Sciences and Technologies (ISTC) is the result of a fusion of various institutions such as: the former Institute of Psychology, the former Institute of Phonetics and Dialectology in Padova and some groups from Biomedical Technologies in Rome, LADSEB in Padova and from the Solid State Electronics group in Rome. It thus comprises various research groups, many of which tend to projects related to the field of language technologies, be it in the area of signal processing or datamining.

The research groups composing ICST are :

Cognitive Processes and New Technology Group, Division of Artificial Intelligence Cognitive Modelling and Interaction, Environmental Psychology Unit, Gesture and Language Laboratory (GaLL), Group for the Psychological Study of Interpersonal Relations, Human Development and Society Research Group, Laboratory on Agent-Based Social Simulation (LABSS), Laboratory for Applied Ontology (LOA), Laboratory of Artificial Life and Robotics - Stefano Nolfi, Language Development and Disorders Laboratory (LaDD), Neuropsychology of Language and Deafness Unit (NLS) - Virginia Volterra, Planning and Scheduling Team (PST), Sign Language Laboratory (SLL), Speech and Multimodal Communication Laboratory, T-cube: Trust, Theory and Technology Group, Unit of Cognitive Primatology and Primate Center, Laboratorio di Psicologia della Partecipazione Infantile.

Instituto di Elaborazione della Informazione

Pisa Research Area, Via Moruzzi, 1, 56124 , PISA

Tel: +39 050 3152876 Fax: +39 050 3152810

Website: <http://www.iei.pi.cnr.it/>

Contact: Piero Maestrini

Email: maestrini@iei.pi.cnr.it

Number of positions: 120

Research projects

The Institute conducts research in Computer Science, Information Technology and related application areas. In particular, the research activities focus on the sectors showed below. Most of the activity is performed within the framework of national and international research projects. Projects concerning language technologies touch the following areas :

- Modelling the retrieval of multimedia information (including text, image and video data) based on probabilistic terminological logic models;
- The design of high performance multimedia object stores including access methods in multilevel storage hierarchies and methods for multilingual access and querying on multimedia information bases;
- Formal specification and design methodologies and tools for the development of database applications;
- Interoperability and metadata for networked library systems.

Interactive Sensory System Division, SSI

Trento, Italy

Tel: +39-0461-314566 Fax: +39-0461-314591

Website: <http://ssi.itc.it/home.htm>

Contact: Gianni Lazzari

Email: lazzari@itc.it

Number of positions: 56

Research projects

SSI, aim at studying new methods and implementing systems for the acquisition, processing and organization of sensorial and environmental information . Innovative applications based on spoken language technology, vision, predictive and software modeling are our challenge.

CASPER - Comprehensive Automatic SPEech Recognition

The goal of this research line is to improve the technology involved in the extraction of linguistic information from audio data, coping with the issues that make automatic speech recognition in unconstrained domains still an open problem. Particular consideration is given to issues concerning robustness, genericity and efficiency of automatic transcription systems.

Projects FAME - Facilitating Agents in Multicultural Exchange TC-STAR - Technology and Corpora for Speech to Speech Translation PF-STAR - Preparing Future Speech Translation Research (EU project) PEACH - Personal Experience with Active Cultural Heritage WebFAQ - Web Flexible Access and Quality

DITELO - Dialogue over the TELEphOne

Dialogo su TELEfonO is related to speech recognizers for telephonic applications, with the main aim to include them in call-center systems (but not only). This means building acoustic and language models, and offering dialogue capabilities. **Projects** CoralRom - Integrated reference corpora for romance languages Homey - Home monitoring through intelligent dialog systems WILMA - Wireless Internet and Location Management Architecture project PF-STAR - Preparing Future Speech Translation Research (EU project).

HERMES - Cross-Language Information Processing

The Hermes research line aims at advancing and developing machine translation technology in the following application areas: -spoken language systems -document translation - information access **Project** TC-STAR - Technology and Corpora for Speech to Speech Translation PF-STAR - Preparing Future Speech Translation Research FAME - Facilitating Agent for Multicultural Exchange WebFAQ - Web: Flexible Access and Quality

SHINE - Speech Interaction in Noisy Environment

This research line has the goal of investigating on robustness in speech recognition and on acoustic scene analysis.

Currently, the main activities refer to the development of conversational voice interaction systems in the car environment and of distant-talking interaction in noisy and reverberant environment.

ISTI, Knowledge Discovery and Delivery Laboratory

ISTI • Area della ricerca CNR, via G. Moruzzi 1, 56124 PISA, Italy

Tel: +39 050 3152878 Fax: +39 050 3152811

Website: <http://www.isti.cnr.it/ResearchUnits/Labs/kdd-lab/>

Contact: Fosca Gianotti

Email: Fosca.Giannotti@isti.cnr.it

Number of positions: 27

Research projects

The central goal of the Lab is to develop new computational methods for the analysis of large and complicated data sets, i.e., to develop methods that help to extract knowledge from data and deliver it to the end users. The methods for constructing knowledge are based on algorithmic and statistical approaches, database techniques and machine learning methods. The research of the unit can be viewed as an intertwined combination of five research areas:

Knowledge Discovery Support Environment - Data Mining Query Languages.

Most knowledge-intensive data analysis applications require the combination of two kinds of activities: knowledge acquisition, and reasoning on the acquired knowledge according to the domain rules. Data mining techniques are an answer to the first issue, in that they extract from raw data implicit knowledge at a higher abstraction level. There is the need of knowledge discovery support environments and data mining query languages, capable of integrating knowledge extraction and knowledge manipulation to the purpose of developing vertical data analysis applications which incorporate data mining technology. In this context, the Lab is exploring several directions:

LDL-Mine is a logic database language (rule-based, SQL-compatible) that facilitates the integration of mining and querying by allowing a uniform treatment of extracted knowledge with domain-specific knowledge .

Frequent Pattern Discovery in Databases. A special research line is oriented to frequent pattern discovery. The aim is twofold. On one side we investigate how to design a highly expressive primitive for frequent pattern discovery in databases, which could be added to a relational query language. On the other side we investigate the optimization techniques that make such primitive efficient enough to be embedded into a database management system.

KDDML-MQL. MQL is an SQL-like algebraic language for the specification of knowledge extraction problems, which supports combination of various forms of knowledge by means of nested operators, as well as composition of steps of the knowledge discovery process. MQL is based on the lower-level environment KDDML (KDD Markup Language), which exploits XML for the representation of the extraction process and extracted patterns.

Istituto di Linguistica Computazionale del CNR

ILC-CNR? Istituto di Linguistica Computazionale del CNR Area della Ricerca CNR Via Giuseppe Moruzzi N° 156124, Pisa, ? Italy

Tel: [+39] 050 315 2872 Fax: [+39] 050 315 2834

Website: <http://www.ilc.cnr.it/indexflash.html>

Contact: Nicoletta Zamorani Calzolari

Email: direttore@ilc.cnr.it

Number of positions: 53

Research projects

The Istituto di Linguistica Computazionale del CNR (ILC-CNR) - working in the field of Computational Linguistics since 1967, when a Division of Computational Linguistics was formed at the Centro Nazionale Universitario di Calcolo Elettronico (CNUCE) - was founded as an independent Institute of the CNR in 1978. ILC-CNR has been one of the major promoters of the notion of language resources as the central component of the "linguistic infrastructure" (aware also of its cultural, economical and political implications), has coordinated the major initiatives relating to language resources and standardisation and has often been the promoter of new "paradigms" in the field. ILC-CNR has designed and built several types of corpora and lexicons and the respective ontologies, has developed a complete chain of tools for a robust processing of the Italian language, for the acquisition of information from corpora and for word-sense disambiguation and has developed technologies for several application domains (question answering, information retrieval, text mining, monolingual and multilingual terminology extraction, ontology acquisition and structuring, summarisation, filtering of Web documents, preservation of the cultural heritage through digital image processing and digital libraries techniques etc.).

The activities of ILC-CNR are articulated in 7 main research lines.

- Design of standards and building of computational language resources;
- Models and methods for the processing of natural languages and monolingual and multilingual prototypes application-oriented;
- Computational methods and tools for the humanistic research, with a particular care of linguistic and literary disciplines and of lexicography;
- Library material and Computational Philology;
- Linguistic Miner: a virtual observatory of contemporary Italian;
- Architecture of language technologies for the promotion of Italian in the knowledge society;
- Natural Language Processing and natural access to knowledge.

Knowledge Representation Research Center, Free University of Bozen-Bolzano

KRDB Research Centre Faculty of Computer Science Free University of Bozen-Bolzano
Piazza Domenicani, 3 I-39100 Bozen-Bolzano BZ, Italy

Tel: +39 0471 016 120 Fax: +39 0471 016 129

Website: <http://www.inf.unibz.it/krdb/>

Contact: Enrico Franconi

Email: franconi@inf.unibz.it

Number of positions: 22

Research projects

In recent years, knowledge and data base applications have progressively converged towards integrated technologies which try to overcome the limits of each single discipline. Research in Knowledge Representation (KR) originally concentrated around logic-based formalisms that are typically tuned to deal with relatively small knowledge bases, but provide powerful deduction services, and the language to structure information is highly expressive. For example, research on formal languages for ontologies was originated from KR, as well as research in computational semantics for natural language. In contrast, Information Systems and Database (DB) research mainly dealt with efficient storage and retrieval of powerful query languages, and with sharing and displaying large amounts of (multimedia) documents. However, data representations were relatively simple and flat, and reasoning over the structure and the content of the documents played only a minor role.

This distinction between the requirements in Knowledge Representation and Databases is vanishing rapidly. On the one hand, to be useful in realistic applications, a modern KR system must be able to handle large data sets, and to provide expressive query languages. This suggests that techniques developed in the DB area could be useful for KR systems. On the other hand, the information stored on the web, in digital libraries, and in data warehouses is now very complex and with deep semantic structures, thus requiring more intelligent modelling languages and methodologies, and reasoning services on those complex representations to support design, management, flexible access, and integration. Therefore, a great call for an integrated logic-based view of Knowledge Representation and Database technologies is emerging. KRDB technologies offer promising formalisms for solving several problems concerning Conceptual Data Modelling and Ontology Design, Intelligent Information Access and Query processing, Information Integration, Peer to Peer systems, Semistructured Data, Distributed and Web Information Systems, E-services, Computational Logic, Logic-based Computational Linguistics, and Bio-informatics.

The KRDB Research Centre at the Faculty of Computer Science of the Free University of Bozen-Bolzano was founded in 2002, and it aims at being an international centre of excellence in basic and applied research on KRDB technologies and at proposing to selected enterprises innovative ideas and technologies based on the research developed in the centre.

Laboratory fo Applied Ontology

Polo Tecnologico Via Solteri, 38 38100 Trento, ITALY

Tel: +39 0461 436641 Fax: +39 0461 435344

Website: <http://www.loa-cnr.it/index.html>

Contact: Nicola Guarino

Email: guarino@loa-cnr.it

Number of positions: 32

Research projects

The Laboratory for Applied Ontology (LOA) performs basic and applied research on the ontological foundations of conceptual modeling, exploring the role of ontologies in different fields, such as: knowledge representation, knowledge engineering, database design, information retrieval, natural language processing, and the semantic web. The group is characterized by a strong interdisciplinary approach that combines Computer Science, Philosophy and Linguistics, and relies on logic as a unifying paradigm. On the application side, special emphasis is given to the use of ontologies for electronic commerce, medical information systems, enterprise modeling, integration of lexical resources, and information access to the Web.

- ALgebraic Tools IN Formal Ontology (collaboration and exchange project)
- ILIKS: Interdisciplinary Laboratory on Interacting Knowledge Systems;
- MOSTRO, Modeling Security and Trust Relationships within Organizations;
- Metokis, Methodology and Tools Infrastructure for the Creation of Knowledge Units;
- SemanticMining Network of Excellence - Semantic Interoperability and Data Mining in Biomedicine.

Organismes et associations

Center for the Evaluation of Language and Communication Technologies

c/o BIC - Via dei Solteri, 38, 38100 Trento - Italy

Tel: +39 0461 405338 Fax: +39 0461 405372

Website: <http://www.celct.it/>

Contact: Amedeo Cappelli

Email: cappelli@celct.it

Number of positions:

Mission

- Studying and implementing evaluation procedures for a number of highly relevant HL-MCTs, e.g. multilingual question answering, speech-to-speech translation, cross-language information retrieval from multimedia archives, word sense Disambiguation, multimodal technologies (emotions in speech/language and synthetic agents interaction with new devices);
- Launching, organizing and promoting evaluation campaigns;
- Monitoring the evolution of HL-MCT, to timely address emerging needs in the field of evaluation;
- Editing and publishing guidelines for HL-MCT evaluation;
- Cooperating intensively with other scientific institutions and business partners;
- Providing services to companies operating in the field of HL-MCT.

FoLLI, Association of Logic, Language and Information

39100 Bolzano Italy

Tel: +33 (0)3 83 58 17 90 Fax: +33 (0)3 83 41 30 79

Website: <http://www.folli.org/>

Contact: Carlos Araces

Email: carlos.araces@loria.fr

Mission

FoLLI has been conceived as a meeting place between various existing fields. Very broadly, FoLLI's basic natural focus is on the phenomenon of information. Linguistics studies information structures, computer science studies mechanisms for processing information, and logic is the mathematical study of both representational and computational aspects. This general connection shows in many specific research lines. In substantial and flourishing areas like generalized quantifier theory, type theory, or dynamic semantics, it is hard to tell where one discipline begins and another ends. And new links of this kind keep emerging in our community.

The main activity of FoLLI is to act as a mother organization for the growing community in the fields of Logic, Language and Information. FoLLI provides a gathering point for members of the community, scientifically sponsors activities in the field, distributes relevant information, etc.

FoLLI is actively involved in a number of concrete international activities in the fields of Logic, Language and Information including. Among the most important ones, Folli organizes each year the European Summer School for Logic, Language and Information (ESSLLI), and publishes the Journal of Logic, Language and Information (JoLLI).

Tuscan Word Centre

The Tuscan Word Centre, Azienda Casanova Vellano 409, 51010 Pescia (PT), Italia

Tel: +39 055 295470

Website: <http://www.twc.it/>

Contact: John Sinclair

Email: jms@twc.it

Mission

The Tuscan Word Centre is a non-profit Association devoted to promoting the scientific study of language. TWC made its name by offering short, high-level courses and similar events for language researchers and workers in the language industries : A short intensive course on the theme: **The Case for a Corpus - Spoken and Written Corpora**. Ten acknowledged experts in corpus design and construction lead this packed four-day course. As always there is an emphasis on practical matters of corpus handling and retrieval software.

A Guided Seminar on the theme: **Dial-a-Corpus - Compiling Corpora from the Web**. A team of six experts of international standing lead this new venture, which aims to reach workable conclusions after thorough discussion and practical trials. Participants will have every opportunity to join the discussions and have attention paid to their own corpus needs.

TWC Courses concentrate on the use of electronic corpora for different purposes, including: translation, automatic or machine-aided, language processing, tagging, parsing etc., language teaching support, language learning assistance, lexicography and language reference

Other Activities of TWC: Advisory and consultancy services, Language processing software evaluation, Project design and management, Language product development, Organisation of academic and professional events, eg, conferences, seminars, workshops.

Resources TWC offers access to unique and growing resources for language and information, including: large corpora in English and many other European languages, state-of-the-art retrieval software, designed to meet EU standards, specially designed corpus processing tools for sophisticated analysis, range of computers and proprietary software, a panel of advisers with expertise in most aspects of language and information .

Japon

Laboratoires

Advanced Telecommunications Research Institute International ATR

ATR Spoken Language Communication Research Laboratories, 2-2-2 Hikaridai, "Keihanna Science City", Kyoto 619-0288 Japan

Website: www.slt.atr.co.jp

Contact: Selichi Yamamoto

Email: webmaster@slc.atr.jp

Number of positions: 30

Research projects

ATR is an independent corporation which conducts R&D, both basic and advanced, in the field of telecommunications. Research is commissioned by public and private sectors, or carried out with funds obtained through open competition. We are also actively engaged in applying our research results to industrial and commercial use.

- CHATR – Technology for recreating human speech;
- ATR Matrix – Multilingual automatic translation for information exchange;
- ATRCALL – Computer assisted language learning system;
- Semantic Map – Community Network Visualizer;
- Research of corpus-based language translation technologies;
- Construction of large-scale bilingual databases.

Computational Linguistics Laboratory

Computational Linguistics Laboratory, Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan

Tel: +81 743 72 5240 Fax: +81 743 72 5249

Website: cl.naist.jp

Contact: Yuji Matsumoto

Email: matsu@is.naist.jp

Number of positions: 36

Research projects

- Practical translation pattern acquisition from combined language resources;
- Robust ontology acquisition from machine-readable dictionaries;
- Information extraction and sentence classification applied to clinical trial MEDLINE abstracts;
- The Hinoki Treebank: A treebank for text understanding;
- Building a paraphrase corpus for speech translation.

Itaku Laboratory

Itakura Laboratory, Department of Information Electronics, Graduate School of Engineering,
Nagoya University, Furo-cho 1, Chikusa-ku, Nagoya, 464-8603 JAPAN

Tel: +81-52-789-3171 Fax: +81-52-789-3172

Website: <http://www.itakura.nuee.nagoya-u.ac.jp/>

Contact: Fumitada Itakura

Email: itakura@nuee.nagoya-u.ac.jp

Number of positions: 24

Research projects

Acoustic Signal and Speech Processing technologies in the context of human communications are main interests of this research group.

The following systems are being developed as a part of research.

- Acoustic Virtual Reality System;
- Large Vocabulary Continuous Speech Recognition System;
- Robust Speech Recognition System;
- Speech Signal Processing toward Robust Speech;
- Recognition System;
- Speaker Recognition System;
- Acoustic Scene Analysis;
- Audio Data hiding.

Kasuya Lab (Klab)

Utsunomiya University Faculty of Engineering, Graduate School of Engineering, 7-1-2, Yoto, Utsunomiya-city, Tochigi, 321-8585, JAPAN

Website: <http://www.utsunomiya-u.ac.jp/en/index.html>

Contact: Hideki Kasuya

Email: kasuya@klab.jp

Number of positions: 21

Research projects

- Automatic extraction of formant and voice source parameters based on an autoregressive with exogenous input (ARX) speech production model;
- Modeling of aperiodic nature of speech signals;
- Acoustic and articulatory dynamics to discriminate female and male voices;
- Acoustic modeling for spontaneous speech;
- Unification-based semantic analysis for robust understanding;
- Timing control of speech feedback for fluent man-machine communication;
- Database design and development;
- Linguistic/paralinguistic/extralinguistic analysis of casual conversation;
- Generation of prosodically-rich response.

Language Media Laboratory

Language Media Laboratory, Division of Media and Network Technologies
Graduate School of Information Science and Technology, Hokkaido University

Tel: +81 11 706 6534 Fax: +81 11 709 6277

Website: http://sig.media.eng.hokudai.ac.jp/index_e.html

Contact: Kenji Araki

Email: araki@media.eng.hokudai.ac.jp

Number of positions: 26

Research projects

Natural Language Processing

Inductive Learning Method (IL-NLP)

Syntactic Analysis

Morphological Analysis

Semantic Analysis

Machine Translation

Speech Translation

Spoken Dialogue Processing

Input Method for Mobile Phones

Speech Information Processing

Robust Spoken Language Processing

Multi-Modal Dialogue System

Dialogue Analysis

User Satisfaction

In-vehicle Information System

Knowledge Retrieval and Processing

Automatic Creation of Commonsense Database

Emotional Information Retrieval

Discovery of Utterances Illogicality

**National Institute of Advanced Industrial Science and Technology
(AIST)**

AIST Tokyo Headquarters 1-3-1, Kasumigaseki Chiyoda-ku, Tokyo 100-8921 Japan
AIST Tsukuba Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan

Tel: +81-29-862-6553 Fax: +81-29-862-6559

Website: www.aist.go.jp

Contact: Hiroaki Kojima

Email: speechweb@m.aist.go.jp

Number of positions: 3200

Research projects

- Robust Speech Signal Analysis, Feature Extraction, and Speech Separation;
- Speech Representation using a Universal Intermediate Phonetic Code System;
- Articulatory Feature-Based Speech Coding, and Speech Synthesis;
- Speech Recognition and Speech Retrieval based on the Universal Phonetic Codes.

Shikano Lab

Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0101, Japan

Tel: +81 743 72 5280 Fax: +81 743 72 5289

Website: <http://spalab.naist.jp/en/e-home.html>

Contact: Kiyohiro Shikano

Email: shikano@is.aist-nara.ac.jp

Number of positions: 42

Research projects

- Voice Conversion between Speakers, Languages, and Emotions
- Human-Robot Dialog System
- Speech Agent Guidance System
- Speech Command Recognition for wearable computer
- Unsupervised Speaker Adaptation in Noisy Environments

Tokyo Institute of Technology

Department of Computer Science, Graduate School of Information Science and Engineering
2-12-1 Oookayama, Meguro-ku, Tokyo 152-8552, Japan

Tel: +81-3-5734-3480 Fax: +81-3-5734-3480

Website: <http://www.furui.cs.titech.ac.jp/english/index.html>

Contact: Sadaoki Furui

Email: furui@furui.cs.titech.ac.jp

Number of positions: 51

Research projects

Speech Recognition

- Large vocabulary continuous speech recognition;;
- Spontaneous speech recognition;
- Robust speech recognition using prosodic information.

Speech synthesis

- Smooth and natural speech synthesis
- Speech synthesis with various prosody style

Spoken dialogue system

- Multimodal dialogue system for information retrieval
- Parallel computing-based spoken dialogue system.

Organisations et Associations

<p style="text-align: center;">Asian-Pacific Association for Machine Translation AAMT</p>
<p>c/o JEITA, Mitsui Sumitomo Kaijo Bldg., Annex, 3F, 3-11, Kanda-Surugadai, Chiyoda-ku, Tokyo 101-0062, JAPAN</p> <p>Tel: +81-3-3518-6418 Fax: +81-3-3518-6472</p> <p>Website: www.aamt.info</p> <p>Contact: Junichi Tsujii</p> <p>Email: aamt@aamt.info</p>
<p>Mission</p>
<p>The association is comprised of three entities: researchers, manufacturers, and users of machine translation systems. The association endeavors to develop machine translation technologies to expand the scope of effective global communications. For this purpose the association is engaged in machine translation system development, improvement, education, and publicity.</p>

Information Processing Society of Japan

Kagaku-kaikan (Chemistry Hall) 4F, 1-5 Kanda-Surugadai, Chiyoda-ku, Tokyo 101-0062
JAPAN

Tel: +81-3-3518-8374 Fax: +81-3-3518-8375

Website: <http://www.ipsj.or.jp/english/index.html>

Contact: Takashi Masuda

Email: through website mail only

Mission

Founded in April 1960, the IPSJ, Information Processing Society of Japan, has been a leading authority in technical areas of information processing and computer science for professionals and students. IPSJ will provide a leadership for sound evolution of the computer science and technology in an increasingly computerized society and will contribute to creation of new ideas to cope with the accountability for evolving information technology. Through its authoritative publications, conferences, and other activities, the IPSJ will play a critical role in the world for the global prosperity.

IPSJ promotes development of the arts, sciences, industry, and humanity through conducting various activities about information processing with computers and communications and providing resources for discipline and opportunities of cooperation with sister societies to members.

Pays Bas

Laboratoires

Computational Linguistics and AI Group, Tilburg University

Tilburg University, Faculty of Arts, Computational Linguistics & AI Group, P.O. Box 90153, 5000 LE Tilburg, Netherlands

Tel: 31 13 4663060 Fax: 31 13 4663110

Website: <http://let.uvt.nl/research/ti/indexE.htm>

Contact: Harry C. Bunt

Email: Harry.Bunt@uvt.nl

Number of positions: 29

Research projects

The research program of the Computational Linguistics and AI section, investigates the computational aspects of the acquisition, representation, processing and communication of knowledge and information expressed in natural language.

The programme is based on the view that language, knowledge, and information processing are most fruitfully studied in an integrated fashion. Many linguistic phenomena can be understood only by considering the contextual background knowledge and the dynamic information states of agents that produce and process language, while on the other hand many aspects of the knowledge acquisition, representation, and processing by intelligent agents are determined by language-based communication. The integrated approach to the study of language, knowledge and information and how they are related computationally takes shape in a number of projects, which fall in one or more of the following four areas and subprograms:

- Dialogue theory and human-computer communication
- Induction of linguistic knowledge
- Computational semantics and pragmatics
- Reasoning with knowledge and belief.

Mathematics and Computer Science? Faculty, University of Twente

University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science
Faculty. Office P.O. Box 217 7500 AE Enschede

Tel: +31 (0) 53 489 3701 Fax: +31 (0) 53 489 4571

Website: <http://www.ewi.utwente.nl/en/>

Email: BFD_EWI@ewi.utwente.nl

Research projects

Multi-media, Virtual Reality and Human Computer Interaction

Research in this field is conducted by a number of groups within the department of Computer Science. In the data management group there is a growing interest in multimedia. The attention for multimedia and information retrieval as practiced in Twente is unique in the Netherlands. Research is also directed towards developing ambient data management solutions for ambient intelligence environments, through a peer to peer based multi media database infrastructure for the support of person oriented, context aware access in distributed or ad hoc environments. Human computer interaction at the UT has always been aimed at multi modal interaction that is directed by the (im)possibilities of the environment. The starting point in this for the computer is multimodal representation. This research is linked to research in the field of social user interfaces and an interest in the possibilities of speech and language that is also present in multi media retrieval research.

Institute for Logic, Language and Computation

ILLC Office University of Amsterdam, Plantage Muidersgracht 24 1018 TV Amsterdam, Netherlands

Tel: +31 20 525 6051 Fax: +31 20 525 5206

Website: <http://www.illc.uva.nl/>

Contact: Ingrid Van Loon

Email: ingridillc.uva.nl

Number of positions: 42

Research projects

Many broad flows of information drive the modern technological world. It is a challenge for contemporary science to provide a deeper understanding, and where possible, enhance existing practice. Indeed, in the course of this century, information has become a crucial theme for scientific studies across many disciplines.

Encoding, transmission and comprehension of information are the central topics of research at the Institute for Logic, Language and Computation (ILLC) of the University of Amsterdam. The broader context in which ILLC sees itself is that of an upcoming information science or 'informatics', which is concerned with information flow in natural and formal languages, as well as many other means of communication, including music and images of various kinds.

Research at ILLC aims at developing logical systems that can handle this rich variety of information, making use of insights across such disciplines as linguistics, computer science, cognitive science, and artificial intelligence. Additional methods are actively pursued as well, whenever relevant, ranging from statistics to argumentation theory.

In addition to its specific research goals, ILLC aims at overcoming traditional borderlines between faculties and disciplines, and serves as a rallying point for information scientists across computer science, linguistics, philosophy, or social sciences. Moreover, the institute propagates exact logical standards of semantic clarity, algorithmic perspicuity, and increasingly also efficient computability.

The resulting view of information science transcends the boundaries of the university. ILLC is also committed to dissemination of its results into the broader world of general education, vocational training, and industrial research.

Dutch Research School for Information and Knowledge Systems

SIKS, Utrecht University, Institute for information and computing sciences, P. O. Box 80.089,
3508 TB Utrecht

Tel: 030-253-4083/1454 Fax: 030-251-3791

Website: <http://www.cs.uu.nl/siks/>

Contact: Dr. R.J.C.M. Starmans

Email: address:office@siks.nl

Number of positions: 270

Research projects

SIKS is the Dutch Research School for Information and Knowledge systems. It was founded in 1996 by researchers in the field of Artificial Intelligence, Databases & Information Systems and Software Engineering. SIKS is an interuniversity research school that comprises 12 research groups in which currently over 270 researchers are active, including 120 Ph.D-students.

The mission statement of SIKS is:

- to perform high-level fundamental and applied research in the field of information and computing science, more particularly in the field of information and knowledge systems;
- to organise a high-quality four-year educational program for its Ph.D. students, employed at 10 different Universities in the Netherlands or at leading companies in the field of ICT.
- to facilitate and stimulate co-operation and communication between our members (Ph.D. students, research fellows, senior research fellows and associated members) and between the School and its stakeholders, including leading (industrial) companies in the field of ICT.

The research program concerns the study of information systems and knowledge systems (IKS) from the perspectives of the following scientific areas:

- database systems, as studied in computing science
- information systems, as viewed from their areas of application
- knowledge engineering, as studied in artificial intelligence
- software engineering, as it is applied to IKS
- theoretical computer science and applied logic, as a foundation of IKS.

Human Computer Studies Laboratory, University of Amsterdam

Human-Computer Studies Laboratory, Matrix I, Kruislaan 419, 1098 VA, Amsterdam

Tel: +31 (0)20 888 4689 Fax: +31 (0)20 525 6896

Website: <http://hcs.science.uva.nl/index.html>

Contact: Pieter Willems Adriaan

Email: pietera@science.uva.nl

Number of positions: 31

Research projects

GARP Knowledge Articulation and Communication Tools (for reasoning about system behaviour)

HOPS is a three-year project focused on the deployment of advanced ICT "voice-enabled front-end public platforms" in Europe permitting access for European citizens to their nearest Public Administration.

METIS Knowledge management: from smart employees to smart organizations.

IRIS Information Retrieval and Information, Universtiy of Nijmegen

Room A4021 (Floor #4) Toernooiveld 1 6525 ED Nijmegen the Netherlands

Tel: +31 (0)24 365 34 56 Fax: +31 (0)24 365 33 56

Website: <http://osiris.cs.kun.nl/iris/web-docs/>

Contact: Nicole el Moustakim

Email: nicolem@cs.ru.nl

Number of positions: 27

Research projects

Pronir In the NWO PRONIR project, in collaboration with the University of Tilburg profile based retrieval of heterogeneous networked researches is being researched. The contribution of IRIS is to develop a generic model for webresources, and to develop a general theory for an effective brokering function. ? Contact person: Erik Proper. ?

Retrieval Models The most well-known information broker on the internet is the search engine. Several approaches have been proposed, based on some method to represent the contents of documents, a retrieval language and a matching strategy. IRIS tries to develop new models for applications beyond the search engine. Contact person: Theo van der Weide.

The Leiden University Centre for Linguistics

Tel: +31 (0) 71 527 2105

Website: <http://www.ulcl.leidenuniv.nl/>

Contact: Vincent Van Heuven

Email: v.j.j.p.van.heuven@let.leidenuniv.nl

Number of positions: 135

Research projects

Acquisition: The (morpho)syntactic development of children with a cochlear implant (NWO-Vidi)

Amerindian: Giving them back their languages: The endangered Amerindian languages of the Guianas (NWO-Endangered Languages)

Bantu: The phrasal phonology of Bantu languages (NWO-Veni)

Bantu: Word order and morphological marking in Bantu (NWO-Programme)

Degrees across Categories (NWO-Vidi)

English: The codifiers and the English language: tracing the norms of Standard English (NWO-Vici)

Dutch: Dutch as a construction language (NWO-Vidi)

Ethiopia: The morpho-syntax of two modal categories in Omotic languages of south-west Ethiopia (NWO-Endangered Languages)

Europe and Eastern Asia: Spinoza-programme of Frederik Kortlandt (Spinoza)

Fataluku: The Fataluku Language Project (NWO-Endangered Languages)

Hittite: Connectives in Hittite texts (NWO-Veni)

Indo-European: The Indo-European Etymological Dictionary (IED) (NWO, Brill)

Indonesia: Linguistic Variation in Eastern Indonesia: The Alor and Pantar Project (NWO-VIMP)

Syntactic Change: How and Why? (NWO-Vidi)

The Typological Database Project (NWO, participating universities)

Utrecht Institute of Linguistics OTS

Trans 10 3512 JK UTRECHT The Netherlands

Tel: +31 30 253 6006 Fax: +31 30 253 6000

Website: <http://www-uilots.let.uu.nl/>

Contact: Martin Everaert

Email: Martin.Everaert@let.uu.nl

Number of positions: 52

Research projects

Aim of the UiL OTS is to develop scientific expertise in the systems underlying language, speech and their use. Crucial to this end is the investigation of the structure of the human language faculty, and the way it is embedded in man's general cognitive capacity. Research is conducted on the basis of the understanding that properties of the computational system of language cannot be derived from functional considerations of the system of use/interpretation: systems of use/interpretation and communication are consistent with many possible natural and non-natural languages, and thus cannot explain why human language has the specific properties it has been found to have. On the other hand, it is a crucial fact about human language that it can be used to argue, communicate, think, etc. If some formal analysis of the computational system turns out to be inconsistent or incongruent with basic facts of language use and interpretation, this cannot be the correct analysis, since the actual sentences of human language can be used and interpreted for such purpose. Capturing correctly the interface between the formal system and the systems of interpretation and use is, therefore, a crucial criterion of adequacy for any linguistic theory. Similar considerations pertain to the relation between linguistic form and its expression. Linguistic form is compatible with several modes of expression. On the other hand, given some mode of expression, a crucial task is to investigate its properties and to show precisely what constraints it poses on the parts of the linguistic system it interfaces with.

The basic research of UiL OTS is divided into seven groups. Added to these is the special research programme Language in Use.

- Syntax and Semantics
- Morphology and Phonology
- Computational Linguistics and Logic
- Phonetics
- Development of Language Systems
- Language Use
- Experimental Psycholinguistics

Organimes et associations

Acoustical Society of the Netherlands

NAG-secretariat P.O.Box 480 6200 AL Maastricht, The Netherlands

Tel: +31 43 3470956 Fax: +31 43 3438781

Website: <http://www.nag-acoustics.nl/>

Contact: Ingrid Bonfrere

Email: secr@nag-acoustics.nl

Mission

The "NAG" is an acronym of "Nederlands akoestisch Genootschap" (the Acoustical Society of the Netherlands). Its 500 members are from the Dutch acoustical industries, universities and research institutes. Our main objective is to encourage the development and the applications of the science of acoustics and acoustical techniques.

The NAG tries to achieve its objectives, among others, by :

- Organising scientific meetings, approximately every 3 months.
- Publication of a journal with the proceedings of each scientific meeting
- Publication of the NAG-almanak, including a list of all society members and sponsors.

European Network of Excellence in Human Language Technologies

ELSNET, Trans 10, 3512 JK UTRECHT, The Netherlands

Tel: +31 30 253 6050 Fax: +31 30 253 6000

Website: <http://www.elsnet.org>

Contact: Steven Krauwer

Email: steven.krauwer@let.uu.nl

Mission

ELSNET is the European Network of Excellence in Human Language Technologies. Its main objective is to advance human language technologies in a broad sense by bringing together Europe's key players in research, development, integration or deployment in the field of language and speech technology and neighbouring areas. The network's role is to offer an environment that allows for optimal exploitation of the available human and intellectual resources in order to advance the field. This environment comprises a number of structures (committees, special interest groups), actions (summer schools, workshops) and services (web site, email lists, newsletter, information dissemination, knowledge brokerage).

Royaume Uni

Laboratoires

Center for communication Interface Research

Centre for Communication Interface Research The School of Engineering and Electronics The University of Edinburgh King's Buildings, Mayfield Road Edinburgh, EH9 3JL

Tel: (+44) 131 650 2783 Fax: (+44) 131 650 2784

Website: <http://www.ccir.ed.ac.uk/index.html>

Contact: Mervyn Jack

Email: maj@ccir.ed.ac.uk

Number of positions: 20

Research projects

The Centre for Communication Interface Research (CCIR) is part of the School of Engineering and Electronics at the University of Edinburgh.

CCIR's research deals with people interacting with automated systems

CCIR undertakes research in usability engineering and dialogue engineering for speech recognition services (vCommerce), for Internet services (eCommerce) and for mobile services (mCommerce).

Usability engineering is the discipline of creating human-computer interfaces which are comfortable and satisfying for the user whilst being efficient in their operation.

Dialogue engineering is the application of scientific principles to the design, construction and maintenance of human-computer interfaces.

CCIR's commercial sponsors support research into the design and experimental evaluation of human-computer interfaces. CCIR also undertakes more general studies of human factor aspects of human-computer interface design.

The Center for Speech Technology Research

The Centre for Speech Technology Research University of Edinburgh 2 Buccleuch Place
EDINBURGH, EH8 9LW, UK

Tel: +44 (0) 131 650 4434 Fax: +44 (0) 131 650 6626

Website: <http://www.cstr.ed.ac.uk/people/>

Contact: Steve Renals

Email: S.Renals@ed.ac.uk

Number of positions: 37

Research projects

AMI: Augmented Multiparty Interaction

AMI is an EU Integrated Project about computer enhanced multi-modal interaction in the context of meetings

BARKS

The BARKS project is exploring the use of switching linear dynamic models for automatic speech recognition.

Cougar

The Cougar project investigates using an articulatory(-like) domain for the calculation of join costs in conjunction with smoothing unit transitions in unit selection speech synthesis.

ESLASR

ESLASR aims to improve the quality of automatic speech recognition using loosely-coupled HMMs with articulatory-acoustic features.

Espresso

Novel acoustic models for ASR

Expressions

Expressions aims to improve the quality of prosody and intonation for unit selection speech synthesis.

The Festival speech synthesis system

The Edinburgh Speech Tools

Speech tools is a set of core libraries used by Festival and various other applications

SVitchboard

Small vocabulary tasks from Switchboard

TESS: Testing Evaluation of Speech Synthesis

TESS is a project designed to investigate the psychoacoustic processes underlying human auditory evaluation of synthetic speech

Center for vision, Speech and Signal Processing

Center for vision, Speech and Signal Processing, School of Electronics & Physical Sciences,
University of Surrey, Guildford, GU2 7XH, Surrey, UK

Tel: +44 (0) 1483 686030 Fax: +44 (0) 1483 686031

Website: <http://www.ee.surrey.ac.uk/CVSSP/index.php>

Contact: Joseph Kittler

Email: J.Kittler@surrey.ac.uk

Number of positions: 78

Research projects

The Centre for Vision, Speech and Signal Processing is one of the major research centres of the School of Electronics and Physical Sciences of the University of Surrey. Its aim is to advance the state of the art in multimedia signal processing and computer vision, with a focus on image, video and audio applications.

The Multimedia Signal Processing and Interpretation research group clusters research interests in signal processing and sensory data analysis. The focus of research extends far beyond the traditional signal processing tasks of noise reduction and information compression for communication and storage. The primary aim of our endeavour is to develop techniques that enable machines to understand and act on the information content of the data (e.g. speech recognition). The representation of the content we attempt to extract must therefore emphasize the semantic information conveyed by the data, as it has to serve as a basis for decision-making and interpretation. Our work is also distinguished by exploiting the complementarity of multimedia and multimodal data and the synergy derived from its joint processing and analysis.

Pattern Recognition & AI The key technology of multimedia signal interpretation. The Centre's predictive validation approach to modelling pattern classes help solve abnormality and fault detection.

Automated Analysis of Audio-Visual Material We have been investigating several approaches to the automated interpretation of sports material, as an example of the analysis of highly structured multimedia content.

Speech and Audio Processing Activities in this area include time-frequency analysis and adaptive filtering of audio signals, such as periodic-aperiodic decomposition of speech, and approaches to recognizing speech based on articulatory gestures.

Biometric Personal Identity Authentication ID authentication combining face image, voice characteristics and lip dynamics. Such technology is required for access control, smart applications and security systems.

Natural Language and Information Processing Group

University of Cambridge, Computer Laboratory, William Gatesw Building, 15 JJ Thomson Avenue, Cambridge, CB3 0FD, UK

Tel: +44 (0) 1223 763500 Fax: +44 (0) 1223 334678

Website: <http://www.cl.cam.ac.uk/Research/NL/>

Contact: Caroline Matthews

Email: cb210@cl.cam.ac.uk

Number of positions: 20

Research projects

Robust Parsing Technology

- Robust statistical parsing (RASP)
- Automated grammar/lexicon induction
- Disambiguation
- Semantic analysis

Constraint-Based Processing

- LKB system and English Resource Grammar
- Underspecified compositional semantics
- Representation issues (defaults, lexical rules, multilinguality)
- Lexical acquisition
- Alvey Natural Language Tools.

Text Summarisation

- Robust discourse analysis (Argumentative Zoning)
- Types of discourse structure and their roles
- Context factor effects
- DARPA/NIST TREC programme

Rhetorically-Motivated Search and Navigation

- Skim-reading electronic texts; navigation to similar texts
- Use of citation information for search

Flexible and Robust Text Generation and Regeneration

- Text simplification
- Regeneration of coherent summary text
- Reporting results from information extraction

Language Acquisition and Evolution

- Acquisition of categorial grammar
- Evolutionary simulation of language development and change

System Evaluation

- Methodology for NLIP system assessment (role of task context)
- Summary evaluation by extrinsic experiments
- Summary evaluation by content units (factoids)

Text Mining and Bioinformatics

- Automatic curation of scientific literature
- Parsing and information extraction

Natural Language Processing and AI

The Faculty of Mathematics and Computing, The Open University, Walton Hall, Milton Keynes, Buckinghamshire, MK7 6AA, UK.

Tel: 01908 652892 Fax: 01908 652140

Website: <http://mcs.open.ac.uk/>

Contact: Anne De Roeck

Email: a.deroeck@open.ac.uk

Number of positions: 21

Research projects

Artificial Intelligence (AI) research within the Computing department falls into two broad areas. We are investigating how computers can be used to model cognitive processes, and how such models can be applied to areas such as planning. Our research in natural language processing is concerned with finding the patterns in bodies of real natural language texts and the interpretation of text and diagrams.

In particular, we are attempting to develop answers to the following questions:

What are the necessary conditions for sentient awareness in artificial entities, such as robots?
How can we develop non-sentential (i.e., analogical, diagrammatic, or connectionist) representations for reasoning about action and planning?
How can computers understand diagrams?
What patterns can be detected in bodies of real text?
How can concepts be distilled from textual data?
How can we model the interaction between different layers of representation in an intelligent system?

Many of our members are part of the Open University Natural Language Processing Group.

Projects

CLinical E-science Framework

Designing adaptive infORMation exTRaction from text for KnOwledge Management.

European Learning Grid Infrastructure based on GRID technologies

HALO

Network for Euro-Mediterranean LAnguage Resources

Natural Language Processing Research Group

NLP Research Group, Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield S1 4DP, UK

Tel: +44 (0)114 222 1814 Fax: +44 (0)114 222 1810

Website: <http://nlp.shef.ac.uk/>

Contact: Yorick Wilks

Email: y.wilks@dcs.shef.ac.uk

Number of positions: 45

Research projects

The Sheffield NLP group has been in existence ten years and is one of the largest and best known in the UK; we enter international competitions on best computer conversationalist, best question-answerer etc. (usually US competitions) and have won a number of these. Our major emphases in research are on the use of coded representations of meaning content, belief and knowledge, on Machine Learning techniques to derive our data from sources like the web, and on the provision of software architectures to underpin NLP research. Our GATE architecture has been installed at over 400 sites world-wide. We also see processing language on computers as a major route to understanding how the mind works, the traditional goal of Artificial Intelligence. We see NLP as the main way of using and coping with the world wide web, which means bringing intelligent machines and knowledge, including all scientific knowledge, into contact with people through conversation technology, as well as being a principal contributor to the future of electronic games and entertainment.

Institute for Communication and Collaborative Systems

School of Informatics, Appleton Tower, Crichton Street, Edinburgh, EH8 9LE, Scotland, UK

Tel: +44 131 650 2690 Fax: +44 131 651 1426

Website: <http://www.iccs.inf.ed.ac.uk/>

Contact: Johanna Moore

Email: j.Moore@ed.ac.uk

Number of positions: 213

Research projects

The Institute for Communicating and Collaborative Systems (ICCS) is dedicated to the pursuit of basic research into the nature of communication among humans and between humans and machines using text, speech, and graphics, and the design of interactive dialogue systems, using computational and algorithmic approaches, with applications including natural language processing, information retrieval and presentation, education, musical analysis, and instruction. ICCS is part of the School of Informatics at the University of Edinburgh.

The research areas covered by ICCS are listed below, together with the staff members that work in these areas:

Annotation and markup technology: Jean Carletta, Ewan Klein, Oliver Lemon, Henry Thompson, Richard Tobin

Biomedical natural language processing: Claire Grover, Ewan Klein, Bonnie Webber

Computational psycholinguistics: Frank Keller, Jon Oberlander

Computational semantics: Ewan Klein, Mirella Lapata, Alex Lascarides, Oliver Lemon, Colin Matheson, Bonnie Webber

Conversational agents: Oliver Lemon, Hiroshi Shimodaira, Mark Steedman

Discourse and dialogue: Jean Carletta, Alex Lascarides, Oliver Lemon, Colin Matheson, Johanna Moore, Bonnie Webber

Educational technology: Johanna Moore, Helen Pain, Keith Stenning

Graphical communication: John Lee, Jon Oberlander, Keith Stenning

Human reasoning: Keith Stenning, Jon Oberlander

Information extraction: Claire Grover, Ewan Klein, Bonnie Webber

School of Computer Science, University of Manchester, IMG, NLE

School of Computer Science, University of Manchester, Kilburn Building, Oxford Road, Manchester, M13 9PL, United Kingdom

Tel: +44 161 275 6154 Fax: +44 161 275 6204

Website: <http://www.cs.manchester.ac.uk/>

Contact: Carole Goble

Email: carole@cs.man.ac.uk

Number of positions: 39

Research projects

The Information Management Group The Information Management Group (IMG) is run by Prof. Carole GOBLE and Prof. Norman PATON, and conducts research into the design, development and use of data and knowledge management systems. Challenging applications motivate and validate our research, in particular the Semantic Web and e-Science. The group currently comprises 5 academic staff, 17 research staff and 13 research students. The current IMG research programme is supported by grants with a total value of around 3 million ukp from a wide range of sources, including the UK Research Councils (EPSRC, BBSRC), the EU, the Wellcome Trust, US DARPA, the DTI and industry. The group has strong industrial links with major companies including AstraZeneca, GlaxoSmithKline, Millennium Pharmaceuticals, IBM, Oracle and Sun Microsystems, and SMEs such as geneticXchange. In addition to collaborating with established companies, the group has been connected with the establishment of two start-up companies: Sagitus Solutions and Network Inference (which recently acquired \$4 million in Venture Capital from Nokia Ventures). IMG plays a leading role in standardisation and industrial forums, as leading members of the W3C Semantic Web activity and co-chairs of working groups of the pharmaceutical/biotech industrial forum I3C and the Global Grid Forum. Members of the group sit on a wide range of UK and international government and community funding and policy forming committees.

Natural Language Engineering Language engineering: building software tools to support people doing useful things with natural language text. The MultiFlora project aims to provide proof of concept that document information extraction can be improved by the analysis of multiple parallel texts. Applied to botanical taxon descriptions, we believe this technique has the potential to be a useful tool in biodiversity informatics. Multiflora aims to (i) ground the work in a structured domain model, capitalising on the expertise in Medical Informatics at Manchester, and (ii) extend taxonomic coverage, contributing significantly to biodiversity research.

School of computing, Robert Gordon University

School of computing, Robert Gordon University, Schoolhill, Aberdeen, AB10 1FR, Scotland, UK

Website: <http://jabber.comp.rgu.ac.uk/research/aboutus.php>

Contact: Susan Crow

Email: S.Craw@comp.rgu.ac.uk

Number of positions: 27

Research projects

- Adaptive Web Agents
- Algorithms for distributed constraint solving
- Extending Reasoning about Constrained Taxonomies
- Easing Knowledge Acquisition for Case Based Design
- Knowledge Discovery from Text
- RECOP: Representing and Reformulating Constraint and Optimisation Problems
- Smart Web Technologies Centre
- Supporting Federated Information Sharing Communities
- HORUS: Hieroglyphic Optical Recognition and Update System
- ProfileSkim - An Intelligent Document Browser
- Sentinel - Knowledge sharing tools for email
- Smart Web Technologies Centre

Computing Science at Aberdeen

Department of Computing Science, University of Aberdeen, Aberdeen AB24 3UE, Scotland

Tel: +44 (0)1224 272295 Fax: +44 (0)1224 273422

Website: <http://www.csd.abdn.ac.uk/>

Contact: Sandra Edwards

Email: sedwards@csd.abdn.ac.uk

Number of positions: 84

Research projects

With a long-standing reputation in Artificial Intelligence and Databases, the Department of Computing Science is internationally recognised for its work in Knowledge and Information Management. Much of the Department's current research looks at the problems of managing knowledge and information in the context of the Internet and World Wide Web. A key problem is helping humans make decisions and take actions when faced with a vast amount of available data and information.

The Department is involved in several international collaborative projects and is a member of the EPSRC-funded Interdisciplinary Research Collaboration (IRC) in Advanced Knowledge Technologies. A recent feature is a number of funded projects involving fundamental work in E-Science and connected to the development of the Semantic Web. Our main current research themes are listed below. The activities in these themes are inter-related, and many members of the department contribute to more than one theme.

Intelligent Software Agents We are examining how to create networks of autonomous software entities that can act on a human's behalf in a complex environment, acquiring information, learning, and interacting with other agents. Example application: e-business, building robust electronic marketplaces of agents that buy and sell products and services.

Knowledge Technologies We are developing the processes of acquiring, modelling, using, and re-using knowledge. Example application: e-engineering in the aerospace industry, making the accumulated expertise of 100s of engineers available to a designer working on a jet engine, by modelling and re-using meaningful knowledge from a company's intranet.

Information Interpretation & Communication We are researching techniques for extracting knowledge from data streams, identifying significant events, and communicating with users via natural language generation and/or graphics. Example applications: e-healthcare, detecting patient emergencies from real-time data in an intensive care unit, to help medical staff take appropriate action; bioinformatics, modelling and visualising protein interactions in 3D.

Language and computation group

Department of Computer Science University of Essex Wivenhoe Park Colchester CO4 3SQ
United Kingdom

Tel: +44 1206 872770 Fax: +44 1206 872788

Website: <http://cswww.essex.ac.uk/LAC/>

Email: genoffice@essex.ac.uk

Number of positions: 20

Research projects

The research by the members of the Language and Computation group covers most areas of Computational Linguistics and Natural Language Engineering.

Grammars and Parsing: research focuses on Constraint-Based grammars formalisms (Arnold, Borsley, Sadler, Spencer) and statistical and symbolic parsing (Arnold, Borsley)

Semantics and semantic interpretation: our research in this area includes work on the logical foundations of semantics (Fox), on psychologically motivated computational models of semantic processing (Poesio), on vector-based models of lexical and text representation, used, e.g., in anaphora resolution (Poesio).

Applications: some of the areas of interest include generation, e.g., for dynamic web page generation (Poesio), information retrieval and web search (Kruschwitz, Robinson), machine translation (Arnold, Sadler), spoken dialogue systems (Kruschwitz, Poesio).

Projects

- The acquisition of lexical and ontological knowledge
- Intelligent web search;
- Reliability statistics for corpus annotation; formal semantics and semantics-prosody interaction; compositional semantics below the word level; focus; coordination; temporal quantification;
- Formal semantics, property theory, plurals, underspecification.

National Center for Text Mining

University of Manchester, Oxford Road, Manchester, UK, M13 9P

Tel: +44 (0)161 275 6100

Website: <http://www.nactem.ac.uk/>

Contact: Julia Chruszcz

Email: julia.chruszcz@manchester.ac.uk

Number of positions: 53

Research projects

The National Centre for Text Mining is the first publicly-funded text mining centre in the world. We provide text mining services in response to the requirements of the UK academic community. We also make significant contributions to the text mining research community, both nationally and internationally.

Arabic WordNet - aims to develop a version of WordNet for Arabic, following the development process of **Princeton WordNet** and **Euro WordNet**.

Bio-MITA - aims to support biological knowledge discovery by means of text-based term association mining.

BOOTStrep - to contribute to the improvement of resources such as ontologies and dictionaries, required for text mining in the biomedical domain.

JST Project - investigates the use of machine learning techniques for the development of clustering and classification algorithms.

Manchester Centre for Integrative Systems Biology - will pioneer the development of new technologies in Systems Biology, based in part on the use of text mining tools.

ProFClass-TM - investigates the use automatic text-classification to assist in the assignment of proteins to functional categories.

Natural Language and Computational Linguistics

Department of Informatics, University of Sussex, Falmer, Brighton BN1 9QH UK

Tel: +44 1273 678029 Fax: +44 1273 877873

Website: <http://www.informatics.susx.ac.uk/research/nlp/>

Contact: John A. Carroll

Email: J.A.Carroll@sussex.ac.uk

Number of positions: 24

Research projects

COGENT: Controlled Generation of Text: In this project, we are developing reflective techniques for controlling wide-coverage generation effectively. The project aims to deliver a substantial and novel resource to support future research in this area, and practical implementations of wide-coverage controllable generators.

Natural Habitats: This project is investigating how NLP techniques can help make service composition a possibility for non-technical users, focusing on the development of an interactive service composition tool that uses a natural language interface. ?

Ranking Word Senses for Disambiguation: Models and Applications? In this project we aim to develop novel ways of estimating the frequency distributions of senses of words from raw (unannotated) text.

Past projects

MEANING: Developing Multilingual Web-scale Language Technologies. **DEEP THOUGHT:** Hybrid Deep and Shallow Methods for Knowledge-Intensive Information Extraction. This project investigated methods for combining robust shallow methods for language analysis with deep semantic processing. **LUCY:** The project develops an electronic database of structurally analysed modern written English. **PSET:** Practical Simplification of English Text. **CHRISTINE:** The CHRISTINE Corpus comprises a socially-representative annotated sample of current spontaneous speech to create resources for studying structure in present-day British language. **LEXSYS:** Analysis of Naturally-occurring English Text with Stochastic Lexicalized Grammars. The project developed a robust wide-coverage parsing system for English text, exploiting a combination of statistical techniques involving online corpora, inheritance hierarchies for imposing structure on NLP data, and lexicalised grammars. **POLYLEX:** The project developed an inheritance-based trilingual lexicon for the core vocabulary of Dutch, English and German using inheritance networks to share information across the languages at all levels of linguistic description. **SPARKLE:** Shallow Parsing for Acquisition of Lexical Knowledge. **SUSANNE:** Surface and Underlying Structural Analysis of Natural English.

Oxford Computational Linguistics Group

Oxford University Computing Laboratory Wolfson Building Parks Road Oxford,

OX1 3QD UK

Tel: +44 (0)1865 283520

Website: <http://www.clg.ox.ac.uk/>

Contact: Stephen Clark

Email: Stephen.Clark@comlab.ox.ac.uk

Number of positions: 20

Research projects

The group's interests cover a wide range of active research areas in Computational Linguistics and Language Engineering, from theory to applications, including:

Statistical Parsing

- Probability models
- Statistical parsing with Combinatory Categorical Grammar (The CCG site)
- Wide-coverage semantic analysis

Statistical Tagging

- Maximum Entropy taggers
- Part-of-speech tagging, named-entity recognition, CCG supertagging

Machine Translation

- MT by statistical parsing (2005 JHU Workshop)
- Example-Based MT

Knowledge Acquisition

- Learning domain theories from text
- Estimating selectional preference probabilities using WordNet

Large-Scale Grammar and Lexicon Development

- Developing large-scale LFG grammars for a wide range of languages
- The ParGram Project

Semantic representations for natural languages

- First and higher order logics for NL semantic representation
- Inference methods for these representations
- Semantic composition for LFG using linear-logic

Natural Language for Knowledge Representation

UCL Department of Phonetics and Linguistics

Department of Phonetics and Linguistics, University College London, Gower Street, London, WC1E 6BT, United Kingdom

Tel: +44 (0)20 7679 7172 Fax: + 44 20 7679 3262)

Website: <http://www.phon.ucl.ac.uk/index.html>

Contact: Stefanie D. Anyadi

Email: stefanie@ling.ucl.ac.uk

Number of positions: 36

Research projects

The focus of our research in speech recognition has been how knowledge of human linguistic processing can aid the development of machine recognition of speech. This has two main threads: the application of traditional linguistic analysis within the decoder, and the analysis of machine and human recognition performance.

Recent work on **linguistic analysis** has looked at incorporating morphological analysis into the speech recognition decoder. Morphological analysis allows for smaller pronunciation lexicons, while at the same time increasing the dissimilarity of pronunciations. In combination with a word-level language model, we have shown that this approach can lead to improved word accuracy. See the article by Huckvale & Fang (2001).

Work on **human-machine comparisons** has investigated the psycholinguistic aspects of human morphological processing. Interesting parallels can be drawn between our morph recogniser and the results of priming experiments on human listeners. The most recent work here has been an investigation of how information processing accounts of perception can be informed by knowledge of pattern recognition systems engineering (see **seminar**).

Current work by Gordon Hunter and Mark Huckvale is looking at the **statistical properties of words in dialogue turns**. The hope is both to develop superior language models specifically for dialogue and also to find relationships between statistical models and conversational analysis.

In our **prosody work** we have been concerned with how to represent the prosodic structure of text in a hierarchical phonological representation. This has involved studying prosodic phrasing (breaking text into intonational phrases) and the categorisation and assignment of pitch accents within the phrase. Our work on the phonetic interpretation of these structures involves modelling of fundamental frequency contours and predicting the durations of syllabic constituents as a function of the segmental content and the phrase context.

Future work will look at how intonation is used to express the information structure of sentences. This will be of use in *concept to speech* systems in which the text to be spoken is also generated by the computer system.

University Center for Computer Corpus Research on Language

InfoLab 21, South Drive, Lancaster University, Lancaster, LA1 4WA UK.

Tel: +44 1524 510357 Fax: +44 1524 510492

Website: <http://www.comp.lancs.ac.uk/computing/research/ucrel/>

Contact: Paul Rayson

Email: paul@comp.lancs.ac.uk

Number of positions: 27

Research projects

UCREL (the University Centre for Computer Corpus Research on Language) is a research centre of Lancaster University. It draws upon the expertise of the Department of Linguistics and Modern English Language and the Department of Computing. For more than two decades, we have led the way in an approach to natural language processing that is based upon information derived from large bodies of naturally-occurring text. These bodies of text are stored on the computer and are known as corpora (sg. corpus). The vast majority of UCREL's work is carried out within this corpus-based paradigm. The corpora are used to derive empirical knowledge about language, which can supplement, and frequently supplant, information from reference sources and introspection (Leech, 1991; 1992).

Because they are well suited to quantitative analysis, corpora can provide information about the relative frequencies of many aspects of language. These frequencies can then be employed in probabilistic analysis techniques, which are another major feature of UCREL's work. UCREL's work is very much focussed on practical outcomes. We have engaged in corpus-based research contributing to such practical applications as: speech synthesis, speech recognition, machine-aided translation dictionary publishing, social survey interview analysis, computer-aided language teaching

Our work focusses on:

English - we were a leading partner in the British National Corpus consortium and are now exploiting the BNC to arrive at new, data-grounded analyses of present-day British speech and writing. We are also involved in corpus-based work on the historical development of the English language, as well as on learner English.

Modern foreign languages - we have built, annotated, and exploited corpora of modern languages such as French and Spanish, and we are presently involved (in collaboration with the University of Lodz) in producing a major corpus of contemporary Polish.

Minority, endangered, and ancient languages - we have pioneered corpus work on non-indigenous minority languages in the UK (e.g. Chinese, Hindi, Punjabi), and we are now extending this work to European indigenous minority languages. We have also carried out computer-aided linguistic research on ancient languages such as Latin.

Suède

Laboratoires

Department of Linguistics, Lund University

Center for Languages and Literature, Linguistics/Language Technology, Lund University, Helgonabacken 12, Sweden

Tel: +46 +46 222 04 38 Fax: +46 +46 222 42 10

Website: <http://www.ling.lu.se/index.html?language=English>

Contact: Joost van de Weijer

Email: studierektor@ling.lu.se

Number of positions: 42

Research projects

The Department of Linguistics at Lund University has two sections: Phonetics and General Linguistics, both of which have undergraduate as well as graduate programmes. Phonetics was established first, with its first professorial chair already in 1950. General Linguistics followed in 1969. In 1983 a professorial chair and a section in Child Language Research was established at the department. It was replaced in 1998 by a professorial chair in Language Acquisition.

Today the department consists of approximately 50 employees, including teachers, researchers, doctoral students and technical/administrative personnel.

The department has a broad research and teaching profile. The study of different aspects of speech and language processing constitutes one major area of interest for the department. Within this area, focus is on the following themes: production- and perception-based studies of prosody including speech technological applications, language acquisition, including cognitive aspects of the acquisition process, the reading process, comparative studies of speaking and writing and the cognitive processes governing them. Another profile area represented at the department is language typology and the description of less well-known languages as well as minority languages. Focus lies on the study of languages in Asia, in particular South-East Asia and the Caucasus.

Department of Linguistics, Stockholm University

CL Group, Department of Linguistics, Stockholm University, S-106 09 Stockholm, Sweden

Tel: +468 16 20 00 Fax: +468 15 53 89

Website: <http://www.ling.su.se/eng/>

Contact: Kristina Nilsson

Email: kristina.nilsson@ling.su.se

Number of positions: 84

Research projects

Parallel Treebanks As part of our efforts to collect and to build the resources for building Natural Language Processing systems for Swedish we are contributing to the initiative for building a Swedish treebank, and we are experimenting with parallel English-German-Swedish treebanks. We are interested in minimizing the manual effort in compiling a treebank by automatic treebank deepening and annotation projection, we work on the transfer of guidelines, methods and tools from English and German to Swedish, and we are building editing, visualisation and search tools for parallel treebanks

Extracting Ontologies from Parallel Corpora There are many situations where it is of great importance for two agents (human or artificial) to know that they are talking about the same thing. Misunderstandings can be caused by the agents speaking different languages or by them using the same word(s) to mean different things (homonymy/polysemy). To solve this problem, one can make use of an *ontology*, which, put informally, associates the terms from a domain with nodes with unique meanings, ordered in a hierarchical structure. Here is another definition of *ontology* (from Chandrasekaran et al., 1999):?

Coreference Resolution for Information Access Coreference Resolution can be defined as the identification of *identity* between entities in text, both entities identified by named entity recognition and anaphoric references to those entities. Thus, by coreference resolution, information scattered across a text (or if across a collection of texts, by cross-document coreference resolution) can be associated with the entities to which it refers.

Ontology-based Product Name Classification The aim of this project is to develop a system that, given a text with company and product information, suggests one or more nodes in a product classification scheme (the Common Procurement Vocabulary, CPV).

Stockholm University Information System (SUiS) SUiS is a cross-language ontology-based information system which allows English and Swedish searches across all documents indexed by Google in the Stockholm University domain.

Graduate School of Language Technology

Graduate School of Language Technology, Faculty of Arts, Göteborg University, Box 200 SE-405 30, GÖTEBORG

Tel: +46 31 - 773 1000

Website: <http://www.gslt.hum.gu.se/>

Contact: Robin Cooper

Email: director@gslt.hum.gu.se

Number of positions: 83

Research projects

The national Graduate School of Language Technology (GSLT) is hosted by Faculty of Arts, Göteborg University, and is a collaboration between leading centres in language technology in Sweden. The school aims to integrate research on speech and language and to provide a sound basis in both theoretical foundations and applications oriented research.

- Cross-language information retrieval for the languages Amharic, English and Swedish;
- Analysis of feedback phenomena in human-human and human-machine communication;
- Labeling semantic relations and entities.

Investigating if and how some automated textlinguistic methods can give more relevant hits in information retrieval, and give coherent summaries that are more query and user adapted than those usually given in information systems.

Semantic interpretation and speech recognition, speech interfaces for dynamic environments.

Multi-sensory Natural Language User Interface between a human user and an Intelligent Service Robot.

Improving the vector model obtained by LSI, using Linguistic information that could be extracted automatically from raw textual data

Investigations into the relationships between document structure, markup and genre, in which the relationship between document structure and markup is carried out on the basis of memory-based learning techniques and the relationship between document structure and genre on the basis of clustering and machine learning techniques.

A knowledge-based model of pronunciation variation in central standard Swedish built up from phonological rules.

Multilingual and cross-cultural spoken dialogue systems.

Automatic indexing.

Human Computer Interaction and Language Engineering Laboratory (HUMBLE), Swedish Institute of Computer Science

HUMBLE, SICS, Uppsala Science Park, SE-751 83 Uppsala, Sweden

Tel: +46 8 633 15 00 Fax: +46 8 751 72 30

Website: <http://www.sics.se/humle/>

Contact: Jussi Karlgren

Email: jussi@sics.se

Number of positions: 20

Research projects

Projects employ language technology to deepen the system's understanding of some aspects of text. We study the topical structure of text itself, mainly using information extraction technology. We also study the build-up of semantic knowledge using statistical algorithms in cooperation with the ARC laboratory.

Similarly, we stress the importance of systematically tailoring tools to task and workplace, and of providing tools to aid human professional competence. To this end we use and develop methods for non-obtrusive knowledge gathering in a workplace environment.

We currently focus on some concrete research issues: information refinement, collaborative information retrieval, multi-lingual information retrieval, and thesaurus management for keyword assignment.

Interaction and Presentation Laboratory

Interaction and Presentation Laboratory, Department of Numerical Analysis and Computing Science Royal Institute of Technology S-100 44 Stockholm, Sweden

Tel: +46 8 790 62 80 Fax: +46 8 10 24 77

Website: <http://www.nada.kth.se/iplab/>

Contact: Ulla-Britt Lindqvist

Email: ubl@nada.kth.se

Number of positions: 43

Research projects

IPLab was established in 1985 at the Department for Numerical Analysis and Computer Science (NADA), the Royal Institute of Technology (KTH), and is based on an interdisciplinary group of researchers and research students in computing science, linguistics, psychology, sociology and design, with common interest in human computer interaction. The following are IPLab's current research projects.

Writing: Computer support for writing and reading processes

In this part of IPLab, interdisciplinary research is being conducted on computer-supported writing and reading, as well as automatic text summarization and generation.

CSCW: Computer-supported collaborative work

This research concerns aspects of the design and use of groupware applications and computer-mediated communication systems.

OOP: Object-oriented programming for distributed systems

This research focuses on design patterns for interactive and distributed applications, especially within the CSCW area.

HRI: Human-robot interaction

This area deals with the interaction between humans and autonomous service robots.

CL: Communication and Learning

This research concerns fundamental aspects of human communication and learning, and how technology is used in different learning contexts.

PUI: Perceptual User Interfaces

This area focuses on the design of user interfaces with perceptual capabilities, i.e., interfaces that enable the user to communicate with computers via speech and body or face gestures.

UCSD: User-centered System Development

This part of IPLab focuses on User-centered System development.

Natural Language Processing Laboratory, Linköping University

Department of Computer and Information Science, Linköping University, S-583 81 Linköping, SWEDEN

Tel: +46-13-28 24 22 Fax: +46-13-14 22 31

Website: <http://www.ida.liu.se/labs/nlplab/>

Contact: Lars Ahrenberg

Email: lah@ida.liu.se

Number of positions: 20

Research projects

- Development of Generic Resources for Language Technology (LATOS);
- Multimodal Interaction for Information Services (MIFIS) ;
- Multimodal Dialogue Systems for Industrial Applications (MDS) ;
- Corpus-Based Machine Translation (KOMA).

Speech and Communications Technology Group

Royal Institute of Technology (KTH) Dep. of Speech, Music and Hearing, Institutionen för Tal, Musik och Hörsel, Speech and Communications Tcehnology Group, Lindstedtsvägen 24 SE-100 44 STOCKHOLM, SWEDEN

Tel: +46 (8) 790 7879 Fax: +46 (8) 790 7854

Website: <http://www.speech.kth.se/speech/>

Contact: Björn Granström

Email: bjorn@speech.kth.se

Number of positions: 33

Research projects

- Speaker Recognition in Telephony Project homepage.
- Biometrics-Based Recognition of People over the Internet.
- Spoken Language Interaction in Telecommunication.
- Computers in the Human Interaction Loop.
- Dynamic Universal Mobility for Adaptive Speech Interfaces.
- A System for Teaching Spoken Dialogue Systems Technology.
- HP Voice Web Initiative.
- Synthesised talking face derived from speech for hard of hearing users of voice channels.
- Multimodal speech synthesis Project homepage.
- Multimodal and multisensory interfaces for interaction with musculo-skeletal models.

Suisse

Laboratoires

Artificial intelligence Laboratory, Federal Institute of Technology in Lausanne

Artificial Intelligence Laboratory, Institute of Core Computing Science, School of Computer and Communication Science, Swiss Federal Institute of Technology, IN (Ecublens), CH-1015 Lausanne (Switzerland)

Tel: (+41 21) 693-2735 Fax: (+41 21) 693-5225

Website: <http://liawww.epfl.ch/>

Contact: Boi Faltings

Email: Boi.Faltings@epfl.ch

Number of positions: 20

Research projects

The research in Natural Language Processing at the LIA focuses on text-mining (knowledge extraction out of textual data), automatic production of syntactic tools and evaluation of NLP tools. Our text-mining methods are based on techniques developed for information retrieval using a Distributional Semantic approach. In such methods, semantic proximities are derived from co-frequency matrices computed on large textual corpora. Different similarity measures are used to characterize the proximity between queries and documents which are represented in a unified way as projections in a high-dimensional vector space of pertinent terms. Methods for automatic production of syntactic tools aim to implement probabilistic techniques and models operating on textual corpora (raw or annotated texts) in order to adapt various generic algorithms to specific applications: part-of-speech tagging, speech recognition, information retrieval, etc...

- Automatic Structuring of Textual Data: Applications to Text-Mining
- INSPECT - Integration of acoustic and advanced linguistic models into speech understanding systems;
- STING: Evaluation of scientific & technological innovation and progress in Europe through patents;
- NLP-FPGA - Hardware NLP coprocessor;
- INFOVOX - Interactive Voice Servers for Advanced Computer Telephony Applications;
- EXTRACT: Automated Information Extraction from Classified Newspaper Advertisements;
- GRACE - part-of-speech tagging evaluation;
- ELSE - Evaluation of Language and Speech engineering;
- Data-oriented probabilistic syntactic analysis
- Distributional Semantics: application to retrieval from large textual bases;
- ISIS - Design of Advanced Vocal Information Servers;

- Industrial tools for Natural Language Processing.

Department of Linguistics, University of Geneva, Language Technology Laboratory

2, rue de Candolle, CH - 1211 Genève 4 (Landolt, 7e étage, bureau L 705)

Tel: 022 379 73 63 Fax: 022 379 79 31

Website: <http://www.latl.unige.ch/>

Contact: Eric Wehrli

Email: Eric.Wehrli@lettres.unige.ch

Number of positions: 20

Research projects

The work undertaken at the LATL in recent years has basically focused on problems linked to natural language syntax processing, both from the perspective of pure research and of its application to typical language industries issues. On the one hand, it has led to the development of a cognitive model of parsing (Walther 1997), and, on the other, to the creation of syntactic parsers of English, German and French, based on Chomskyan models of formal linguistics. These parsing tools and their lexical databases were the starting points for several projects of written/spoken language processing prototypes and systems.

IPS (Interactive Parsing System)

The IPS Project aims at developing interactive parsers for various languages, based on the linguistic model of GB theory. The goals of project IPS are at the same time theoretical and practical. Project FipsVox (Gaudinat & Wehrli 1997) aims at developing a text-to-speech synthesizer. What makes it different from most of the other systems of French synthesis is, above all, its linguistic base, i.e. the (F)IPS parser. Besides its practical objectives (text reading in MS-Word or HTML formats), this project also has theoretical concerns, and tends in particular to show the importance of linguistic parameters for speech processing.

Indeed, despite the considerable progress recently made in speech synthesis, current systems still have major shortcomings, essentially in the field of prosody (intonation, duration, etc.), although also in the processing of homographs and liaison. These shortcomings, which are largely responsible for the rather unnatural sound of synthetic voice, can hamper the understanding of an utterance and thus put place considerable constraints on practical applications of speech synthesizers.

Institute of Computational Linguistics, University of Zurich

interthurerstr. 190, CH-8057 Zurich, Switzerland
University of Zurich, Institute of Computational Linguistics

Tel: (+41)-44-635-4331 Fax: (+41)-44-635-6809

Website:

Contact: Manfred Klenner

Email: klenner@cl.unizh.ch

Number of positions: 20

Research projects

Computational semantics as unifying theme

The main focus of research in our group is on *computational semantics*, i.e. on methods of computing the meaning of natural language utterances. This is a task with as much theoretical as practical interest. In theoretical terms, computer models of language understanding might shed light on the question of how humans understand language input and, given that language understanding is a very central component of human intelligence, it would probably tell us a lot about intelligence in general.

Of more interest to us are practical considerations, i.e. applications of such methods. One potential application would be systems that answer arbitrary questions (phrased in natural language) over texts, so-called "*Questions Answering Systems*". Types of text for which such systems would be extremely useful are encyclopedias, technical manuals, business reports, legal documents, test and service reports, but also newspapers, news feed messages and even historical records.

The ideal Questions Answering System would understand the texts completely and with all the implicit information contained in them, understand the question (complete with the situation in which it was uttered), and derive all the (and only the) the correct answers from the textual information. Clearly, this kind of text understanding system is way beyond of what is possible in the foreseeable future. We need simpler, more manageable, tasks which, nevertheless, are interesting both in theoretical and practical terms.

Organismes et associations

European Association for Machine Translation

EAMT 40, boulevard de Pont-d'Arve? CH-1221 Genève, Switzerland

Fax: +41 22 379 8689

Website: <http://www.eamt.org/>

Contact: Bente Maegaard

Email: bente@cst.dk

Mission

The European Association for Machine Translation (EAMT) is an organization that serves the growing community of people interested in MT and translation tools, including users, developers, and researchers of this increasingly viable technology.

Together with the AMTA and the AAMT, the EAMT publishes a newsletter, MTNI, and organizes workshops and conferences, such as the bi-annual MT Summit. Under the auspices of the IAMT, it also compiles listings of companies and products which are distributed free or at nominal cost to its members. The EAMT also maintains a mailing list, mt-list@eamt.org, as a public forum for the discussion of translation technology.

European Association for Speech, Signal and Image Processing

EURASIP European Association for Signal, Speech, and Image Processing, EPFL-STI-LTS
Station 11 CH-1015 Lausanne Switzerland

Website: <http://www.urasip.org>

Contact: Peter Grant

Email: info@eurasip.org

Mission

The European Association for Signal, Speech and Image Processing (EURASIP) was founded on 1 September 1978 to improve communication between groups and individuals that work within the multidisciplinary, fast growing field of Signal Processing in Europe and elsewhere, and to exchange and disseminate information in the field all over the world. The association exists to further the efforts of researchers by providing a learned and professional platform for dissemination and discussion of all aspects of signal processing.

Areas of Interest

- Continuous and discrete time signal theory;
- Applications of signal processing;
- Systems and technology;
- Speech communication;
- Image processing and communication.