

**ANNUAL REPORT OF THE YEAR 2006
AND RESEARCH OBJECTIVES FOR THE YEARS 2007, 2008
OF THE RESEARCH INSTITUTE FOR LINGUISTICS,
THE HUNGARIAN ACADEMY OF SCIENCES**

I. Main objectives of the Institute for the year of 2006

1) *Lexicographic research*

Compiling the volume of the entries of letter A and the volume of bibliographic databases and supplementary material of the Comprehensive Dictionary of Hungarian (A magyar nyelv nagyszótára).

2) *Research on the history of Hungarian*

Historical Syntax of Hungarian: seven further chapters have been drafted. The critical edition of the Keszthely Codex has been published. The Érsekújvár Codex is being prepared for publishing.

3) *Research on language technology*

- Successful participation in the following projects: “Hungarian-English Machine Translation”, “Ontology Building” and “Natural Dialogue Modelling”, as well as in the new NKFP/Jedlik project starting in 2006.

- Accomplishing 75% of the verbal Hungarian WordNet database – approx 7000 synsets – and 75% of the nominal and adjectival WordNet together with the consortium partners in the framework of the GVOP 3.1.1-2004-05-0191/3.0 project. Developing an anaphor resolution algorithm, necessary for the illustration of the practical usage of the Hungarian WordNet through an Information Retrieval System.

- Working out a system of approx. 2000 conceptual nodes in the top ontology of the “Hungarian Unified Ontology“. Fine-tuning of the psychological subset of the top ontology's network according to the needs encountered in the Jedlik project.

- Tightening the professional relationship with Marcus Kracht's research group.

- Active participation in carrying out the “CLEF 2006” international contest of information-retrieval.

- Preparation of the revised version of the Hungarian National Corpus with improved accessibility. Revision of the query- and user interface and completion of the prototype version.

- Further developments of existing language technology tools used in the preparation of the revised version of the Hungarian National Corpus, as well as for applications of document-classification tasks and for shallow semantic analysis.

- Research in language technology and computational linguistics: corpus-based investigation of infinitival constructions, syntactic analysis of the Hungarian language with finite-state tools, the use of machine learning algorithms in the automated learning of Hungarian morphosyntax.

4) *Research on Normative Linguistics*

The *examination of linguistic attitudes concerning codified linguistic forms* has been continued. Within the framework of *Geolinguistic investigation of Hungarian language variations*, in cooperation with the Department of Historical Linguistics, Sociolinguistics and Dialectology of Loránd Eötvös University and the Institute for Musicology of the Hungarian Academy of Sciences, the completion of a digital sound corpus of Hungarian dialects has been continued. Besides, 1182 CDs (754 hour's long) have been prepared in order to be used as educational and research material in audio and written form. Dissemination of knowledge about language, as well as consultancy and providing technical expertise have been performed within the framework of public relations.

5) *Finno-Ugric and Uralistic research*

- Several *computer programs and databases* have been created:

Uralic and Permic Etymological Database (UPEDb) completed with the Proto-Permic reconstructed lexemes from A to K. UEDb: The results of the publications in 2001–2005 have been included in the database; two significant collections containing numerous reconstructed data (by Janhunen, Juha and by Sammallahti, Pekka) have also been added. The database has been increased by cca. 8,000 data. Atlas of Udmurt Dialects (AUD) and the Historical Phonological Database of Hungarian have been completed. The root inventories of Komi and Udmurt have been created. The development of morphological analysers for Nenets and Mari has also started. The analyser for Nganasan has also progressed.

Permistics: The research of Permic–Hungarian parallel changes has been finished.

- *Studies on Ugric languages:*

Within the framework of Hungarian Historical Phonology preparatory survey of the (mainly personal) conditions of a comprehensive research project has been launched. The research of Hungarian syllable history continues to be carried out. A brief survey of the present-day Khanty language of journalism and a summary of the linguistic situation of Synya Khanty (the latter in English) have been written. As a summary of the researches on the phonotactics of Ob-Ugric languages, a survey of two meetings and a conference paper have been written. Translation of the monography on Ob-Ugric Syllable history has been continued.

- *Samoyedology*

Negation in Samoyedic Languages, revision of Nganasan grammar, analogy-based approaches to phonology and morphology

- *Typology*

As a contribution to the monograph concentrating on the characterization of Uralic languages from a typological point of view, a survey of the languages in question has been finished. The typological analyses of conditional sentences in the Uralic languages continued in 2006.

6) Orientalistic research

One of the earliest literary remains of the old Osmanli has been published, the *Ferec bād eş-şiddeşş* in two volumes. Editing of an important *Festschrift* (Mélanges en l'honneur d'Elizabeth Zachariadu, Wiesbaden, Harrassowitz, 346 p.). Publishing the second volume of the *Fontes Orientales (Usāma ibn Munqidz's Memoirs from the times of the Crusades)*. The accomplishment of the third volume of the *Old Egyptian Etymological Dictionary*.

7) Sociolinguistic research

- Work on the Budapest Sociolinguistic Interview project (*Budapesti Szociolingvisztikai Interjú – BUSZI-2*) continues in 2006–2009 with as sponsored by OTKA. The guided conversations of 25 interviews have been checked for the second time, as a result of which the searchable version is now available. The programs necessary for the comparison of the dBase and FOX Pro files have been completed, and the data-clarification and data-correction was carried out. As part of the construction of the final version of the test data, the manual of “A Finalised List of the Items in BSI-2” was created.

- Analyses of the data gathered for the Longitudinal Study of Language Shift project and the Dimensions of Linguistic Otherness project have resulted in several conference presentations articles.

- Within the framework of the Hungarian Sign Language project a computer program was designed and tested to gather sentence-comprehension data from users of HSL.

- The final phase of editing was completed for the 23 Hungarian articles for *Lexicon Grammaticorum*, to be published in 2007.

- Data-checking and data-correction were carried out for the real-time questionnaire study *Nézetek-B2*.

- The *Dimensions of Linguistic Otherness: Prospects of Maintenance and Revitalization of Minority Languages* (acronym: DILING) project began on 1 April 2006, funded by the EU

Specific Support Action. Participants include linguists in the Czech Republic, Hungary, Moldavia, Rumania, Serbia, Slovakia, Slovenia and Ukraine.

- The study of Hungarian linguicism (i.e. social discrimination of Hungarians by Hungarians based on language use) has resulted in a novel, theoretically well-based explanation of the country-wide empirical data used.

8) Research on phonetics

- Phonetic research was planned to focus on three main topics: (i) a phonetic analysis of coarticulation processes, (ii) the investigation of the prosodic structures of read texts and of the archived recorded speech material, and (iii) research in speech perception development and self-monitoring.

1. The acoustic-phonetic analysis concerned the properties of neighbouring speech sounds and word final (utterance final) consonants, the acoustic consequences of the articulation of consonant clusters, the specific nature of tremulants, and the effect of nasal context on vowels' acoustic structures. The duration and formant frequency changes of the speech sounds involved in the coarticulation processes were defined on the basis of more than 25,000 measured data. The acoustic-phonetic parameters of the Hungarian tremulants were investigated focusing on their quality patterns, on their voicing character, and on the occurrence of neutral vowels around the consonant in question. The analysis of vowel-vowel combinations is a novelty, the data obtained partly supported, partly disconfirmed previous phonological statements concerning these processes. The effect of nasal consonants showed large context-specific and inter-speaker differences. The first corpus of the acoustic structures of consonant-consonant clusters was created and finalized.

2. The research on the time structure of narrators' speech was concluded, the results were published as a chapter of a book (entitled „*Hangidőtartamok és időszerkezeti elemek a magyar beszédben*”/Speech sound durations and durational elements of Hungarian speech, Academic Publishing House). The phonetic investigation based on the saved files of the Hegedűs Archive began in 2006. The phonetic typology together with the actual text content was the input to the annotation process. The suprasegmental analysis (articulation and speech tempi, pauses, speech melody measurements) was based on 26 speech samples with ritual contents (about 30 minutes) and another set of 60-minute spontaneous narratives from the same speakers. Comparisons were carried out depending on speech styles. The various disfluency phenomena were analyzed using the spontaneous speech materials of the Hegedűs Archives in order to compare these findings to those of present-day spontaneous speech.

3. The speech perception investigations concerned the self-monitoring mechanism on the one hand, and speech perception development, on the other. The effects that speakers' disfluencies have on the listener are rather complex. The goal of the experiments was to describe the correction process and determine its efficiency. Various types of disfluency were tested with nine-year-old children, young adults, and older subjects. The results show that the time span of the corrective process depends upon the type of disfluency, the context, and the listener's age. The higher operational level the production error involves the more time is required for correcting it and the corrections are poorer than at lower operational levels.

In typically developing children, no dissociation is assumed between speaking and speech perception. However, observational data show that there might be some hidden impairment in their speech perception abilities. The test results from a total of 600 children were analyzed. The data confirmed a particular interrelation among the individual perceptual processes: Development can be accounted for in terms of a decrease in interconnections among the various types of processing. Children with relatively poor language skills often display higher levels of disfluency. The present analysis showed that the Hungarian-speaking speech-

impaired children showed more types of and more frequent disfluencies due to the rich morphology of Hungarian. The acoustic-phonetic characteristics of Hungarian dysarthric speech is a new research field where the segmental and suprasegmental levels of articulation of three patients' speech were studied.

9) Research on experimental linguistics

Neurolinguistic research

- In the research related to the model of mental grammar an objective was the construction of an argument system pertaining to the cooperation and common restrictions of the syntactic, morphological and phonological subsystems and certain memory subsystems in the linguistic processing and production performance of healthy and aphasic subjects.
- Preparation and clinical testing of test materials for the systematic analysis of sentence comprehension in aphasic patients. Investigation of the relationship between sentence comprehension performance and verbal working memory capacity.
- Performing psychophysiological tests and tests on reaction time investigating the effect of word order and the increase in the temporal distance of linguistic elements undergoing agreement on the perception of correctness of tense and person/number agreement.
- In the research of context-sensitive phonological strategies the objectives were discovering syllable organisation processes, gathering spontaneous speech data containing phonological deviances and building an analysed corpus from the data.

Psycholinguistic research

- Investigating grammaticality judgment ability and its relations to the parameters of processing in child groups of typical language development and groups with language development problems. Testing the ability to recognise syntactic functions by analogy in healthy adults.
- Starting experimental research on incomplete questions (silent predicates).

Sociolinguistic research

- Starting the EU project DILING (*Dimensions of Linguistic Otherness: A nyelvi másság dimenziói*) and coordinating efforts spanning eight countries. Organising an international conference under the aegis of the project.

Research on languages spoken by Roma communities

- Description of the syntactic structure of Boyash sentences. Studying the possibilities for language shift and language maintenance in Boyash-speaking communities, preparing education material. Surveying the linguistic-political situation of the Roma minority in Hungary. Description of linguistic ideologies in a Transsylvanian Roma community

II. Outstanding research results of the year 2006

1) Language technology

Successful completion of the project, called "Unified Hungarian Ontology". During the year the in-depth description of six different lexical domains (so-called relational microtheory) was completed, and several hundred lexical units were described in a formalized way (see more at <http://ontologia.hu/meo/docs/microtheo/>). By the middle of the year the compilation of the spoken language material of the Carpathian-Basin Hungarian Minority Corpus was accomplished with a working web access (http://corpus.nytud.hu/kmmk/beszelt_nyelv.html). With the support of the "Presidential Grant" altogether 20.000 Hungarian records (instead of the planned 6000) were translated. In the framework of the project "The Building of the Hungarian EuroWordNet and its Application in Information Retrieval Systems" 2000 verbal synonym sets and approx. 90% of the adjectival WordNet were completed. Approx. 470 million words were preprocessed for the purposes of the revised version of the HNC, and the

prototype of the new web query interface was also completed. In cooperation with the Institute for Psychology a corpus of narrative reminiscences concerning national and ethnic identity (1.6 million words) was compiled in the framework of the Jedlik project. Additionally, morphological dictionaries and surface analyzing tools necessary for the analysis have been developed, as well.

2) Normative linguistics

- As result of the project, called: “*Examination of linguistic attitudes concerning codified linguistic forms*”: two conference presentation and three studies has been completed. Fieldwork, carried out by means of questionnaires, were amended with sociolinguistic interviews. Data-processing is to be finished by 2007.

- As a result of the applied linguistical workshops, organized in 2004 and 2005, a study has been compiled.

- Within the framework of a consortium project together with Eötvös Loránd University, the digital version of dialectical sounds, recorded in 1960, has been prepared. Phonological transcription of the chosen sound-recordings was started. More information can be found on the following homepage: <http://geolingua.elte.hu>

3) The history of Hungarian

The critical edition of the Keszthelyi Codex (1522) has been published, as well as a comprehensive monograph on the early written linguistic material of Hungarian from 930 to 1590. The Érsekújvár Codex is being prepared for publishing. Historical Syntax of Hungarian: seven further chapters have been drafted: Sentences expressing wishes; Macro-structure and length of heterogeneous complex sentences; Definiteness and topic-comment structure; Verbal phrases with objects; Modifier phrases of quality; Word-level conjunctive coordination; Subject subordinate clauses.

4) Lexicography

The first two volumes of the Comprehensive Dictionary of Hungarian were completed and published. The entries of the letter C and the selection of the headwords of the letter Cs were completed; the listing of the headwords of the letter D is in progress. The unified bibliographical database was revised and completed.

5) Finno-Ugric and Uralic

The main goal of the year was to form a connection with further branches of linguistics. This ambition, among others, is reflected in the list of conference participations. Beyond the events where the Department of Finno-Ugristics was represented, a number of conferences were organized:

- The conference held with the title *Ob-Ugric peoples in the 21st century* (the proceedings were published on CD) aimed at surveying the tasks and goals of studying these two endangered languages and cultures.
- The first *Workshop and Round-table on Samoyedology* was organized in cooperation with the Research Institute for Linguistics, Russian Academy. The goal of this event was to gather the most prominent representatives of Samoyedology, as well as to sum up the most important researches, methods and results of these days. It was intended to be the opening event of a series held in different places.
- The conference summarizing the results of the project entitled "Complex Database of Uralic Languages", as opposed to the above mentioned events focusing on endangered cultures, aimed at presenting the various types of methods in documenting and describing languages.

- The 5th event of the series entitled *Budapest Uralic Workshop, Workshop on Word Classes* was devoted to investigating word classes in different theoretical frames, from either synchronic or diachronic point of view.

6) *Phonetics*

Considering the actual practical use of research, the results of a series of experiments that were carried out with three groups of (i) 150 mentally retarded (MR), (ii) 150 learning disabled (LD) and (iii) 150 typically developed control (C) first-, second- and third-grade schoolchildren seem to be remarkable. Eight subprocesses of the GMP speech perception diagnostic procedure were used in the experiments (all children were tested individually). Results confirmed that the development of the speech perception mechanisms of the subjects of the three groups showed divergent tendencies.

The videodemonstration on Farkas Kempelen and the reconstruction of his speaking machine was awarded by the 151st Congress of the Acoustical Society of America (first prize) (see in Physics Today: www.physicstoday.org).

7) *Experimental linguistics*

Neurolinguistics

- The project entitled *The hypothesis of weakened syntax in sentence production* has been carried out under the leadership of Zoltán Bánréti. Sentence repetition tests were carried out involving subjects with agrammatic aphasia. Results showed that subjects systematically dropped syntactic markers of information structure contained in the sentences, such as definite and indefinite articles and tense markers. However, the information context of the test did not encourage dropping these endings. Based on our results we have concluded that in an aphasiac disorder syntax disruption does not necessarily involve the dropping of rules/endings. A condition that can be characterised with the term weakened syntax is thus produced. Structure building operations may still be carried out in this weakened syntactic system but only by abnormally overusing mental resources. This induces compensation strategies in the subjects where structural markers related to information structure are dropped, erroneously supposing that they can be deduced from context.

- Katalin Szentkúti-Kiss's research carried the title *Aphasic sentence comprehension*. She has found that based on comprehension data of Hungarian aphasics the validity of hypotheses accounting for agrammatic comprehension solely by the primary disruption of subsystems responsible for producing structural representation can be questioned. Existing results have been interpreted in a theoretical framework where the restricted activation of information stored in the lexical representation of the verb and the disturbed operation of syntactic-semantic mapping processes, rather than restrictions of structural operations is supposed to be in the background of agrammatic comprehension.

- *Investigations in electrophysiology* have been coordinated by Éva Mészáros. The results emphasise the significance of memory-related processes in the judgment of congruence of number/person agreement in healthy subjects. In the experiments on the sensitivity for tense-time adverb congruence the effect of word order on grammaticality judgments of aphasics also points toward that performance is not determined by the primary deterioration of morphological-structural agreement operations. Interpretation of patterns discernible in response times and correct responses was carried out applying an interaction-type model of sentence processing based on capacity.

- Judit Szépe has studied the syllable organisation operations of phonological mistakes, errors and linguistic deviances related to old age and aphasia. She has found that these organisation operations follow the principle of simplification, meaning that their phonological information content is always lower than that of their counterparts following rules. The diverse syllable-

construction deviances are governed by only a few wide-scope processes of structure reordering, structure repetition, and syllable optimality increasing/lowering processes that are based on rule scope modification, that is, on rule extension and rule constraining in addition to the two former operation types. The former includes cases of extending sonority hierarchy, while the latter accounts for the cases of producing malformed syllable types/syllable match.

Psycholinguistics

- The research team lead by Ágnes Lukács investigated grammaticality judgments on sentences with agreement errors in children diagnosed with language development disorder and in a control group with matched receptive vocabulary. Performance was significantly dependent on sentence processing characteristics. Although the performance level of the groups was identical, accompanying factors showed difference. In the group diagnosed with language development disorder (as opposed to the control group) there was a close correlation between the size of vocabulary and verbal working memory and error detection rate, in addition to the clearly discernible effect of grammatical knowledge (or the deficits therein). According to the analysis, grammaticality judgments cannot be regarded as clear measures of grammatical competence.

- Ilona Kassai edited a bilingual volume on the information structure of sentences in which she published the results of a French-Hungarian joint research project. Research on so-called interrupted questions containing silent predicates has begun. The research, also lead by Ms. Kassai, is based on a wide-scope empirical database.

Sociolinguistics

- Lead by Csilla Bartha, the EU-funded project *Dimensions of Linguistic Otherness*, targeted towards surveying the situation of linguistic minorities in eight countries of East-Central Europe has started successfully. As the first milestone of the project an international conference was organised. The conference was held in Budapest, October 25-27. The comparative research project on language shift in six Hungarian minority groups, started with NKFP support, was carried on, also under the leadership of Csilla Bartha. Results were made available in multiple publications appearing in Hungary and abroad and conference presentations in Hungarian as well as in foreign languages. In 2006, work was finished on the corpus of the Hungarian language spoken in the Carpathian basin. A project surveying linguistic, legal, educational political conditions of Hungarian minority schools in the Carpathian basin has also been finished together with a sociolinguistic survey investigating bilinguality among Hungarians living in the Mura basin and the effectiveness of bilingual education in Slovenia applying questionnaire and interview sampling with a pool of 140 people and subsequent statistical analysis. Csilla Bartha is editing a volume of linguistic studies entitled *Roma communities and their languages in the Carpathian basin*.

Research on languages spoken by Roma communities

- Anna Orsós and László Kálmán are preparing a work on Boyash sentence grammar. In addition to various word structures they put down descriptive rules concerning special and yes/no questions. A classification of Boyash compound sentences based on grammatic-semantic relations of clauses has also been finished.

Orsós Anna has completed her PhD dissertation dealing with the sociolinguistic and linguistic/educational political situation of the Boyash language. The aim of fundamental research on the Boyash linguistic community has been the investigation of opportunities for language shift and language maintenance, as well as the production of supplementary material applicable in minority education and community building.

- Andrea Szalai has finished a study (in press) on some aspects of territorial and social variability of the Romani language and on the diversity of the Roma minority in Hungary. In

the research project "Oath and gender in the linguistic ideology of a Transylvanian Roma community" Andrea Szalai's PhD dissertation is not far from being completed. The dissertation investigates gender aspects of Romani linguistic ideology concerning oaths. Csilla Bartha, Anna Orsós, and Andrea Szalai have completed a selected bibliography on Romani and Boyash linguistics.

III. Collaborations and joint projects in Hungary

- Joint project work at the Department of Language Technology: Cooperation with MorphoLogic Ltd, the Institute of Informatics of the University of Szeged, with the Media Educational and Research Centre of the Budapest University of Technology and Economics encompass three large projects (Hungarian-English machine translation, ontology-constructing (NKFP) and WordNet constructing (GVOP) projects). Several colleagues work in tight working relationship with the Media Educational and Research Centre on the development of free language technological tools and on publications (two joint papers in the proceedings of the Hungarian Computational Linguistics Conference 2006, two joint papers of the conferences EACL 2006, LREC 2006).

- Participation in the most important event of computational linguistics in Hungary, at the third Hungarian Computational Linguistics Conference. Altogether ten papers or posters were presented on the basis of research results.

- Research is currently carried out in the framework of four projects. The most significant one is the joint project with MorphoLogic Ltd and the Institute of Informatics of the University of Szeged, aiming at developing a Hungarian-English machine translation programme. Another notable project is the NKFP project aiming at building an upper level ontology, called "Hungarian Unified Ontology" ("MEO"), which has brought together all of Hungary's experts working on this sub-field of language technology. The GVOP project "The Building of the Hungarian EuroWordNet and its Application in Information Retrieval Systems" is also a significant project. A new NKFP project (Jedlik) has just started in the reporting year, in cooperation with the Institute for Psychology of the Hungarian Academy of Sciences, MorphoLogic Ltd and the Institute of Informatics of the University of Szeged.

- In the year 2006 several conferences / lecture series have been (co-)organised by colleagues from the Department of Language Technology:

* On 20 February the Department hosted a highly successful public meeting of the "Hungarian Association of Applied Linguists and Language Teachers", with the title "*Corpora, language processing tools and their applications*". Department members gave a first insight into the basics of computational language processing in seven talks (<http://corpus.nytud.hu/manye/>).

* One colleague had a leading role in organising the conference "László Antal and today's linguistics", with the aim of popularising both the structuralist tradition of today's linguistics and computational linguistics in a wider circle – also among university students. Six from the 13 invited speakers are members of the Research Institute for Linguistics (István Kenesei, András Komlósy, László Kálmán, Péter Rebrus, Miklós Törkenczy, Ferenc Kiefer, Tamás Váradi). See also: <https://eotvos.elte.hu/wiki/Angol/antalkonferencia>

* Co-organisation of the first students' conference of the Anglo-American Studies Workshop of József Eötvös Collegium (ELTE József Eötvös Collegium, 12 May 2006)

- Cooperation with the Terminological Innovation Center (TermIK) has started in the reporting year.

- Within the framework of NKFP applications, cooperation was made with the Department of Historical Linguistics, Sociolinguistics and Dialectology of Loránd Eötvös University at the Department of Normative Linguistics.

- Ongoing funded project: OTKA K 60403 (2006–2009) for the Budapest Sociolinguistic Interview

- Completed project: NKFP 5/044/2002

- Collaboration with the Research Institute of Rumanians in Hungary; the Research Institute of Slovaks in Hungary; the Department of Hungarian Linguistics, Eötvös Loránd University, Budapest; the Doctoral School of Linguistics, University of Szeged.

- The publication of the second volume of the Schmidt Éva Series (Nyelv, nyelvjárás, írásbeliség, irodalom = Language, Dialect, Literacy, Literature) was sponsored by Nemzeti Kulturális Alap (NKA, The National Cultural Fund of Hungary, NCFH). The preparatory work of the following volume has also advanced. With the financial help of NCFH, by the end of 2007, the digitalization of the archives will be finished.

- Ngasan morphological analyser: The team has gone on preparing the fieldwork (OTKA project) on the Taymyr Peninsula, as well as working out a questionnaire (for phonological studies).

- As a contribution to the monography concentrating on the characterization of Uralic languages from a typological point of view, a survey of the languages in question has been finished. In cooperation with the Finno-Ugric Institute of the University of Vienna, as well as the Finno-Ugric Department of Eötvös Loránd University, the establishment of a team for creating a typological database, as well as organizing a conference held in 2008, have started (cf. <http://www.univie.ac.at/urtypol>).

- At the Department of Lexicology the *Comprehensive Dictionary of Hungarian* is compiled in close cooperation with the research team of ELTE BTK TKI.

- There are two projects with the Department of Phonetics supported by OTKA and another one supported by NKFP.

- Joint research projects were carried out at the Experimental Linguistics Department together with the Developmental Psychophysiology group of the HAS Institute for Psychology, with the Cognitive Science group of SZTE and with the General Linguistics Department of DTE BTK.

Hungarian grants

Members of the department take part in work under grants OTKA T047381, OTKA K60403 and in the NKFP project *Nyelvbányász (Language Miner)*.

Activities in Hungarian scientific life

Department members were active in the following societies:

Nyelvtudományi Társaság (Linguistic Society)
Magyar Alkalmazott Nyelvészek Egyesülete (Society of Hungarian Applied Linguists)
Nemzetközi Magyar Filológiai Társaság (International Society of Hungarian Philology) (as board member)
Magyar Fonetikai, Foniátriai és Logopédiai Társaság (Hungarian Society of Phonetics, Phoniatory and Speech Therapy).
Editorial board of the journal *Nyelvtudományi Közlemények*
One of the researchers at the Department of Phonetics worked in the Presidency of the Hungarian Academy of Sciences as an elected representative of the field. The researchers worked in the organization of the National Kazinczy Contest of Fair Pronunciation and one of them served as its president.

Participation at scientific events in Hungary

Staff members delivered talks at the following conferences:

- 14. Élőnyelvi Konferencia (14th Conference on the Spoken Language), Bük, 2006. október 9.
- XVI. Magyar Alkalmazott Nyelvészeti Kongresszus (16th Hungarian Congress on Applied Linguistics), Szent István Egyetem, Gödöllő, 2006. április 11.
- VIII. Dunaújvárosi Nemzetközi Alkalmazott Nyelvészeti, Nyelvvizsgáztatási és Medicinális Lingvisztikai Konferencia (8th Conference in Dunaújváros on Applied Linguistics, Language Testing and Therapeutic Linguistics), Dunaújvárosi Főiskola Nyelvi Intézete, 2006. május 5.
- X. Jubileumi Alzheimer-kór Konferencia (10th Conference on Alzheimer Disease), SzTE és Szegedi Neurológiai Klinika, Szeged, 2006. szeptember 22.
- Kommunikációelmélet (Communication Theory), SzTE Médiatudományi Intézete, Szeged, 2006. november 17.
- A Magyar Fonetikai, Foniátriai és Logopédiai Társaság XXIX. Országos Szakmai Konferenciája (29th Conference of the Hungarian Society of Phonetics, Phoniatory and Speech Therapy), Hódmezővásárhely
- A magyarországi beások nyelvi közössége (The Boyash language community in Hungary). 2006. december 18. Pécs

National and international conferences and workshops organised by the Institute:

- The conference held with the title *Ob-Ugric peoples in the 21st century* (the proceedings were published on CD) aimed at surveying the tasks and goals of studying these two endangered languages and cultures.
- The first *Workshop and Round-table on Samoyedology* was organized in cooperation with the Research Institute for Linguistics, Russian Academy. The goal of this event was to gather the most prominent representatives of Samoyedology, as well as to sum up the most important researches, methods and results of these days. It was intended to be the opening event of a series held in different places.
- The conference summarizing the results of the project entitled "Complex Database of Uralic Languages", as opposed to the above mentioned events focusing on endangered cultures, aimed at presenting the various types of methods in documenting and describing languages.
- The 5th event of the series entitled *Budapest Uralic Workshop, Workshop on Word Classes* was devoted to investigating word classes in different theoretical frames, from either synchronic or diachronic point of view.

- On 20 February the Department of Language Technology hosted a highly successful public meeting of the “Hungarian Association of Applied Linguists and Language Teachers”, with the title “*Corpora, language processing tools and their applications*”. Department members gave a first insight into the basics of computational language processing in seven talks (<http://corpus.nytud.hu/manye/>).

- On 30-31 October the department hosted an international meeting of the CLARIN (Common Language Resources and Technology Infrastructure) initiative, with the participation of thirty guests. This project is one of the priority actions of the 7th Framework Programme. Tamás Váradi fulfils the role of managing coordinator in the project, while the Institute serves as a regional center.

Maintaining the national linguistic information centre:

The electronic mailing system *Nyelvészinfo* (Linguist-info) operated by the Institute sends up-to-date information to 700 addresses about scientific programmes, grants and other topics. The official website of the Institute has been visited more than 512.000 times in 2006, which is 232.000 more than in the year 2005.

The *Phonetics (Fonetika)* website contains interactive databases in English and in Hungarian, sound archives, educational and scientific materials.

The website *Neuroinfo* makes the latest English and Hungarian language publications and test results accessible.

The stock of inventory of the Institute Library has grown by 514 items. The electronic services have been improved (multilingual catalogue, repertory). During the year the Library has had around 1840 personal and 3535 online visitors.

The permanent exposition *The history of Hungarian experimental speech research* has been improved, as well.

The linguistic counselling and the public service of the Institute have been working continually in the following forms:

- Disseminating knowledge about language (for eg.: program, called: “Let’s take us!” produced by Hungarian Public Radio, and the program “Hungarians”, produced by Danube Television/Channel Autonomy)
- Writing articles for the review, titled: “Our sweet mother tongue”
- 394 expert’s reports was written out of these 349 was in connection with registration of birth and marriage
- Operating a help-line on three days of the week, and answering questions through telephone-calls, and emails, most of which are concerned with grammatical standards, spelling, name-giving, name change, or other domains of language use.

- The permanent exhibition *The history of experimental phonetics in Hungary*, is permanently open to the public, provides an opportunity to get acquainted with, and at the same time preserves the history of speech research.

Editorial work:

- The editorial boards of the following journals are to be found in the Institute: *Nyelvtudományi Közlemények, Acta Linguistica Hungarica, Finnugor Világ.*

- The following periodicals are being co-edited by colleagues of the Institute: *Magyar Nyelvőr*, *Modern Nyelvoktatás*, *Beszédgyógyítás*, *Alkalmazott Nyelvtudomány*.
- The results of basic research on speech are published yearly in Hungarian in the series of collections of papers edited at the department of Phonetics under the title *Beszéd kutatás* [Speech Research].

Participation in higher education:

Besides maintaining the Theoretical Linguistics MA and PhD programmes, the colleagues of the Institute participate in higher education in the following departments:

ELTE – Department for Contemporary Hungarian, Department for Hungarian language history, sociolinguistics and dialectology, Department of Chinese Studies, Interpreter and Translator Training Institute; Péter Pázmány Catholic University – Department for Hungarian Linguistics, Department for German Studies, Department for General Linguistics; University Szeged – Department for General Linguistics, Department for Hungarian Linguistics, Department for Finno-Ugric Studies, School of English and American Studies, Cognitive Science Group of the Department of Psychology; University of Miskolc – Department for Hungarian Linguistics; University Pécs – Linguistics Department; Budapest University of Technology and Economics – Department of Cognitive Sciences, PhD programme of the Psychology Department. One colleague of the Institute is a teacher in the Erasmus College and is member of the College Council.

Gergely Bottyán was involved in leading of the Anglo-American Studies Workshop of Eötvös Collegium (designing the course schedule and coordinating the work of the teachers); Bálint Sass has taught at the Faculty of Information Technology of Péter Pázmány Catholic University; Viktor Nagy has taught at the Department of Applied Linguistics and at the Department of Theoretical Linguistics of Eötvös Loránd University; Csaba Oravecz has taught at the Department of Theoretical Linguistics of Eötvös Loránd University and at the doctoral programme of the Department, as well as at the BA programme of the Faculty of Humanities; Károly Varasdi has taught at the Department of Theoretical Linguistics of Eötvös Loránd University and at the Department of Theoretical Linguistics of Péter Pázmány Catholic University; Tamás Váradí has taught at the Department of English Linguistics and at the doctoral programme of the Interpreter and Translator Training Centre of Eötvös Loránd University. Károly Gerstner – PPKE BTK Department of Hungarian Linguistics (History of Hungarian, Dialectology, Historical Lexicography) Gábor Kemény – University of Miskolc, Department of Hungarian Linguistics, lecturer (Stylistics), College for Communication, Budapest (Stylistic practices)

All members of the Experimental Linguistic Department were active in higher education, teaching first and foremost in the Theoretical Linguistics MA and PhD programmes of ELTE/HAS. Zoltán Bánréti is Chair of both the PhD programme in Theoretical Linguistics and the department itself. He lead the work of preparing new detailed curricula for BA and MA in Theoretical Linguistics.

Members of the Department were also teaching at the following higher-education institutions: ELTE Bárczi Gusztáv Gyógypedagógiai Főiskolai Kar, BA and university program, a BME Doctoral School in Cognitive Psychology, Theoretical Linguistics Workshop of PPKE BTK Doctoral School in Language and Literature, PTE BTK French and Linguistics departments, PTE BTK Department of Romology and Sociology of Education, ELTE BTK Department of Modern Hungarian Language, SzTE PhD Programme in English Applied Linguistics. Targeted professional courses were taught to members of the National Family Protection Service.

IV. International relations and projects:

International research projects

- A successful project has begun in 2006 with the participation of eight countries under the title *Dimensions of Linguistic Otherness: Prospects of Maintenance and Revitalization of Minority Languages*. The two-year sociolinguistic project was submitted to the call of the 6th framework programme of the European Commission, within the 7th priority. Csilla Bartha from the Department of Experimental Linguistics is the project coordinator.

- There is an accepted common research at the Department of Phonetics together with the Phonetic Department of Zagreb University (Croatia) starting in 2007.

- Colleagues: Zoltán Bánréti, Éva Mészáros, Katalin Szentkúti-Kiss and Judit Szépe have developed the detailed research plan for a new European project entitled *Data and theories in neurolinguistics* together with the Clinical Linguistics Department of the University of Reading.

- Andrea Szalai is participating in the project no. SAL - 029013, entitled *Society and Lifestyles: Towards Enhancing Social Harmonisation through Knowledge of Subcultural Communities*.

Activities in international scientific life

- The first *Workshop and Round-table on Samoyedology* was organized in cooperation with the Research Institute for Linguistics, Russian Academy. The goal of this event was to gather the most prominent representatives of Samoyedology, as well as to sum up the most important researches, methods and results of these days. It was intended to be the opening event of a series held in different places.

- Bartha Csilla Bartha is member of the International Pragmatics Association, is Hungarian research expert of the UNESCO project *Multilingualism in the world*, is also an expert of *Mercator Education Network* in the field of minority languages, bilingual and minority education, and is on the international advisory board for the book series *Multilingualism and Linguistic Diversity*.

- As a member of the Scientific Council of Centre Interuniversitaire d'Études Hongroises, Sorbonne Nouvelle, Paris 3, Ilona Kassai is actively taking part in shaping the scientific profile of the institution.

- One of the members of the Phonetic Department works as the Secretary General of the International Society of Phonetic Sciences. Two researchers are members of the editorial boards of three international journals.

- The Department of Language Technology took a leading role in a competition of computational linguistic applications organized by the European association "Cross-Linguistic Evaluation Forum" (CLEF), for which the department prepared the linguistic infrastructure, and performed the evaluation procedure, as well. As a result of the cooperation, 30 new topic descriptions were created, which are available to the European research community for doing any other type of research, too.

- Participation in the work of the INTEX/NOOJ software developing community. Colleagues presented their achievements in the development of the Hungarian module in two talks at the annual INTEX conference.

- The department of Language Technology continues to cultivate close connections with the Institute for Parallel Processing of the Linguistic Modeling Department at the Bulgarian Academy of Sciences, where the Clark System for Corpora Development is being developed under the supervision of Kiril Simov.

- It is an important achievement in the further expansion of international scientific relationships that this year Tamás Váradi has been appointed the coordinator of international affairs in the Social Sciences and the Humanities workgroup of ESFRI (European Strategy Forum on Research Infrastructures).
- Several members of the department of Language Technology took part in the organization of CESCL (Central European Student Conference in Linguistics), hosted by the Institute on 29-31 May 2006 (see <http://www.nytud.hu/cescl/>). One colleague, Károly Varasdi, was co-organiser of the IXth Logic and Language International Conference (Besenyőtelek, 24-26 August, 2006).
- Miklós Kontra served on the editorial boards of *Journal of Sociolinguistics*, *Language Policy*, and two book series (for Multilingual Matters and Benjamins). He was also invited to contribute a chapter to *Sociolinguistics around the World: A Handbook* (edited by Martin J. Ball and Nicole Müller, Routledge).
- Preparation of the annual Hungarian bibliography for the European yearbook *Sociolinguistica*.
- In the field of aphasia phonology Judit Szépe has been in cooperation with the Montréal research lab founded by Jean-Luc Nespoulous.
- Ágnes Lukács has been engaged in joint research with Lawrence B. Leonard at Purdue University (US).
- As a contribution to the monography concentrating on the characterization of Uralic languages from a typological point of view, a survey of the languages in question has been finished. In cooperation with the Finno-Ugric Institute of the University of Vienna, as well as the Finno-Ugric Department of Eötvös Loránd University, the establishment of a team for creating a typological database, as well as organizing a conference held in 2008, have started (cf. <http://www.univie.ac.at/urtypol>).
- Ágnes Lukács is to teach a course on language development disorders at Babes-Bolyai University, Cluj, Romania.
- The department of Language Technology won an international grant coordinated by the XEROX Research Center, in which tools for multilingual search in digital libraries are to be developed in cooperation with one Hungarian and nine international partners (Cross-language Access to Catalogues And On-line libraries - CACAO).

VII. Research objectives of the Departments of the Institute for 2007

VII.1) Department of Language Technology:

- Successful participation in the project developing the Hungarian-English machine translation system, and in the NKFP project on computational content analysis (Jedlik).
- Accomplishing the whole of the planned Hungarian WordNet database in the framework of the GVOP project.
- Participating in the work of the "CLEF 2007" international contest of information-retrieval.
- Accomplishing the revised version of the Hungarian National Corpus, with increasing the effectiveness of access. Launching the new version of the query- and user interface. Completing the development of the existing language technological tools for the analysis of the revised version of the HNC, for applications of document-classification tasks and for shallow semantic analysis.
- Finalizing the proposal CLARIN.EU, and starting the preparatory work of the project.
- Starting the work in the framework of the project "CACAO" (Cross-language Access to Catalogues And On-line libraries), that was granted support in 2006.
- Starting the compilation of an up-to-date spoken language corpus according to international standards, in cooperation with the Department of Phonetics.
- Research in corpus linguistics and computational linguistics: syntactic analysis of the Hungarian language with finite-state tools, machine learning algorithms in the automated learning of Hungarian morph syntax, research on verbal argument structure.

VII.2) Department of Normative Linguistics

- In addition to current research topics:
- Examination of linguistic attitudes concerning codified linguistic forms: assimilation of questionnaires is to be continued, preparing and evaluating interviews. Writing articles and studies about the results.
- Participation in the consortium project, called: "Geolinguistic investigation of Hungarian language variations"
- Preparation of new research topics:
- In cooperation with the Gramma Language Office (Dunaszerdahely, Szlovákia) problems and tasks of language management is to be examined. The research would be broadened towards the Carpathian Basin. In connection with it, a conference is to be held in 2007.
- Fulfillment of public relations, providing language consultancy, expert opinions and disseminating knowledge about language

VII.3) Department of Finno-Ugric Studies

All research projects are to be continued.

Computer /electronic databases:

- Developing UPEDb (lexemes from l- to ž-), finishing the Proto-Permic database, as well as checking the data in it.
- UEDb: working out the system of searching in the so-called commenting part of the entries.
- Translating the reconstructed meanings into Russian and Hungarian.
- Analysing the doublets and triplets among the reconstructed forms of the Uralic Etymological Dictionary, summarizing the results of the analysis in the form of a study.
- Finishing the first versions of the Komi and Udmurt morphological analysers
- Developing the Nganasan morphological analyzer

Permistics

- Elaborating the maps of AUD
- Studies in Ugric languages

- Hungarian Historical Phonology: Continuing the description of the syllable, organizing a team of post-graduate / PhD students, and other researchers. Defining the aims and tasks, planning a schedule.
- Ob-Ugric languages: continuing the investigations into the present-day language of journalism in some Northern Khanty dialects – checking the results with a native speaker.
- Analyzing the syntactic and typological background of the peculiarities of the Northern Khanty neologisms.
- Finding the resources for publishing the 4th volume of the Schmidt Éva Series containing a selection of the individual songs of the collection 1980-1982 (in cooperation with the Institute for Musicology, Hungarian Academy of Sciences).
- Writing a study concerning phonology, and another one on certain syntactical phenomena of the Ob-Ugric languages.

Samoyedology

- Approaching the negation in Samoyedic languages with the methods of typology.
- Elaborating questionnaires for the fieldwork on the Taymyr Peninsula.
- Collecting data on the Taymyr Peninsula to test the analogy-based approaches to phonology and morphology
- Finishing the revision of Nganasan grammar.

Typology

- Participation in the organization of the conference on typology in Vienna – presenting papers
- Writing the chapter dealing with conditional sentences for the monography *The Typology of Uralic Languages*.

Writing coursebooks

- Finishing the manuscript of the following volume of the series *Magyar Mozaik (Hungarian Mosaic)*

Conference

- Organizing the following (6th) event of the series *Budapest Uralic Workshop* with the title *Tense in the Uralic Languages*.

VII.4) Department of Oriental Studies

- The 27-28 volumes of the *Turkologischer Anzeiger* will be brought out. Our Turkologists will prepare for publication in Turkish and English the Catalogue of the Turkish manuscripts being in the Oriental Library of the Hungarian Academy of Sciences.
- The third volume of *Fontes Orientales* will be published.
- J. Fehér's book entitled *Emptiness of Persons. Madhyamaka Buddhist Analysis* will also appear this year. Research is carried out in historical pragmatics, on a corpus of old Chinese texts (2007: Discourse and (Im)politeness in the vernacular Chinese institutional discourse, 2008: Theory and practice of power in traditional Chinese criminal investigations).
- Huba Bartos starts working on a monograph treating the syntax of quantificational expressions in Mandarin Chinese.

VII.5) Department of Sociolinguistics

- The guided conversations of another 25 BSI interviews will be checked for the second time. The second checking of the test-data will continue. As far as possible, missing data will be recovered by listening to the audio files. The software to handle the entire database of the BSI-2 project will be developed and tested.
- Continuing projects: *The longitudinal study of language shift; Dimensions of Linguistic Otherness: Prospects of Maintenance and Revitalization of Minority Languages; Hungarian Sign Language; Nézetek-B2 (=real-time changes in Budapest); Bilingual Education for the Bilingual Hungarian Communities in the Carpathian Basin; Gypsy studies*

- Project to be completed: *Lexicon Grammaticorum: Who's Who in the History of World Linguistics*

VII.6) Department of Lexicology and Lexicography

- Completing the list of headwords of the letter D, followed by the selection of headwords of the letters E and É.
- The edition of 75-80% of the entries of the letter B will be completed, as well as the compilation of the entries of the letter Cs. The work on the entries of the letter D commences. Simultaneously with the completion of the entries, the consulting editors will supervise the completed entries. The morphologic coding of these words will also be carried out.
- Department research fellows will continue their individual work in higher education and in public scholarly life.
- The work on an orthographical dictionary for physics will be completed (with Attila Mártonfi and physicists as contributors).
- Compilation of a handbook of historical linguistics.
- Commence of the preparatory work of the Hungarian version of EWUng.

VII.7) Department of Historical Linguistics and Dialectology

- Research on Middle Hungarian syntax includes the following topics in 2007: affirmative sentences; the structure of the clause: multi-unit clause structure I.; topic-comment structure of the sentence in the 16-17. centuries; verbal phrases with indirect objects; modifier phrases of quality; word-level coordination; subordinate clauses of location.
- In a joint-project with the National Széchényi Library, the Department provides (and if necessary, rewrites) the material of already published critical editions for the on-line publication of *Old Hungarian Texts*.
- In cooperation with the Sermones project, completed at Eötvös Loránd University, the literal transcription for the on-line critical edition of the Érdy Codex. Annotation of parts of the Érdy Codex in preparation for creating a syntactically annotated database.
- Plans for preparing two not easily accessible 16. century Bible translations for reprint publication, complete with prefaces (The New Testament translated by Heltai, Gáspár and his team, and The New Testament translated by Félegyházi, Tamás), to be carried out if funds are available.

VII.8) Phonetics Department

The analysis will continue concerning the acoustic-phonetic outcomes of the coarticulation processes in spontaneous speech. The aim is to investigate new phonetic and phonological interrelations. The final goal is to describe the phonetic behavior of all the coarticulation processes of Hungarian speech. A new research area in this field is to analyse the F1 and F2 interrelations in various contexts regarding inter- and intrasubject differences as well.

The research of prosodic structures aims at developing a processing model based on actual measurements of various read styles of speech on the one hand, and using the Hegedűs Archive, on the other. The common research with the Phonetic Department of Zagreb University on the disfluencies starts in 2007. The analysis of disfluency phenomena incorporates neuro-psychologic methods as well. There is a continuous research about the perceptual development of kindergarten and schoolchildren with various types of speech defects.

There is a new topic at the Department of Phonetics the aim of which is to elaborate a spontaneous speech data base (BEA) which does not exist among the Hungarian corpora. The goal of the data base is to provide a research possibility for spontaneous speech and also for applied speech research (e. g. forensic phonetics). The final corpus will contain the speech samples of 500 Hungarian native speakers. There is a starting development of a pronunciation

dictionary on the internet based on the previous speech technology knowledge of the Department.

The Phonetics Department organizes its biennial national Speech Research Conference in November.

VII.9) Department of Experimental Linguistics

1. Neurolinguistic research

Participants: Zoltán Bánréti, Éva Mészáros, Katalin Szentkúti-Kiss, Judit Szépe, and 3 PhD students.

1.1. The theory of grammar in the light of language pathology data: research on mental grammar includes efforts for describing the relationship between subsystems of structure-building operations and certain memory subsystems, primarily in structures of tense and verb – object agreement. Based on test data from aphasic and healthy subjects the aim is constructing a model relating to principles of universal grammar on the one hand, and accounting for functions of memory subsystems on the other.

1.2. Performing joint physiological research with the HAS Institute for Psychology (on number-person agreement, tense agreements, and the agreement of the definite object marker). Extending research activities to subjects with agrammatic aphasia. Results gathered from these physiological investigations, complemented by data from reaction time experiments already completed, can give a clearer picture of healthy and damaged sentence analysis processes.

1.3. Further analysis of the processing of temporal relations of events will be commenced with investigating compound sentences where the event time of the main clause should be determined relative to that of the time clause. Method applied: ordering pictures corresponding to event times expressed in the target sentence. We have developed grammaticality decision tasks for testing information processing related to sentence aspect.

1.4. Research related to aphasic sentence comprehension will include the offline testing of the comprehension of semantically reversible transitive sentences (applying the method of sentence-picture matching) in case of different predicate types, with canonical and non-canonical thematic role order variants of three different construction types: (i) topic-predicate (ii) multiple topics, (iii) topicless structure. Discovery and comparative analysis of comprehension patterns of sentences containing *agentive* and *psychological* verbs in groups with agrammatic and Wernicke's aphasia.

1.5. Continued gathering of data containing deviances of phonological sequence organisation from aphasic subjects, lapses, and old-age mistakes. Of the operations of phonological sequence organisation strategies, we aim to discover the linear reordering processes operating on segment feature assignment in the studied deviance types.

Building a corpus based primarily on data from paraphasia, lapses, mistakes of healthy subjects and dementia patients in old age and, secondly, on data of healthy and pathological language acquisition processes.

2. Psycholinguistic research

Participants: Ágnes Lukács and Bence Kas.

2.1. Investigating traits of language development disorder specific to Hungarian, focussing on new areas such as the adaptation of word order tests applied in aphasia research for development tests, conducting the tests and interpreting their results.

2.2. Under OTKA grant *Analogy-based formal approach to phonology and morphology*, well-formedness judgments related to different forms of defective stems will be tested in a psycholinguistic experiment.

3. Speech research:

Participants: Ilona Kassai and PhD students

3.1. Analysis of the linguistic material gathered by questionnaire survey (240 questionnaires) on "silent" predicates in incomplete questions, started in 2006 under the project MTA/CNRS, will be continued.

3.2. In order to give a more exact definition of sonority, one of the key concepts in phonology by providing an adequate universal cross-language scale, similarities and differences between sonority scales applied for different languages will be examined. The aim is to produce a scale more uniform and more detailed than the one used today, relying on objective acoustic measurements and somewhat less objective psychoacoustic data.

4. Sociolinguistic research

Participants: Csilla Bartha and members of the Department of Sociolinguistics

4.1. Under the European Union project DILING (Dimensions of Linguistic Otherness) -- FP6 SSA the first version of a volume will be produced, containing country review studies on the situation of linguistic minorities of eight participating countries, as well as theoretical studies by international experts. At the end of the first report period a progress report concerning professional, financial, legal, and ethical topics will be prepared for the European Commission.

Organisation of a two-day DILING workshop in June 2007 in Pécs (ICML 11 pre-conference workshop).

4.2. Under the project no. KFP 5/126/2001, *A nyelvi masszázs dimenziói: a kisebbségi nyelvek megőrzésének lehetőségei (Dimensions of linguistic otherness: possibilities for preserving minority languages)* research will be carried out on topics such as the relationship of identity and language and the role of linguistic attitudes and ideologies in language shift. Results will be publicised as written studies and presentations at international conferences.

4.3. Preparing a theoretical study on the relation of minorities, bilingual education programs, and language maintenance, offering an outlook on international practices.

5. Research on languages spoken by Roma communities

Participants: Csilla Bartha, Anna Orsós, and Andrea Szalai

5.1. Editing the volume *Roma communities and their languages in the Carpathian basin* (Budapest: Nemzeti Tankönyvkiadó)

5.2. Extending existing chapters of the Boyash sentence grammar and writing new ones. Extensions: Types of the simple sentence: Attributive sentences. Sentences of existence. Possessive sentences. Interrogative sentences. The yes/no question. Chapters to be written: Compound sentences: The initial section of the sentence (conjunctions, topic). The subjunctive in main and subordinate clauses. Relative clauses. Clauses with the conjunctive "kã". Conditional sentences. Sentence formulas. Greetings, oaths, curses, etc. Vocatives. Formulas with communicative purposes.

5.3. Completing and submitting the doctoral dissertation "Oath and gender in the linguistic ideology of a Transsylvanian Roma community".

5.4. Research cooperation with Susan Gal (Chicago University) and Michael Stewart (University College of London) in the following two topics: 1. *Pragmatic value and gender*, 2. *Language ideologies in statistical discourse: the case of the Gypsy minority in Hungary*

5.5. Preparing selective bibliographies of Hungarian and international literature on Romani linguistics and of Hungarian literature on Boyash linguistics.

VIII. Research objectives of the Departments of the Institute for 2008

VIII.1) Department of Language Technology:

An important goal for the year 2008 is the usage of the linguistic resources at the department, in the following forms:

- * argument-lexicon, based on the verbal subcategorisation frame table that has been developed at the department for the past recent years, with frequency data on the arguments
- * frequency dictionary based on the revised and improved version of the HNC

Maintenance of ongoing projects:

- * Successful completion of the NKFP / Jedlik project on computational content analysis
- * Setting up a regional centre for linguistic resources and technologies in the framework of the CLARIN project

VIII.2) Department of Normative Linguistics

- Follow up of current research projects
- New research topics are to be started.
- Language consultancy, based on the concept of language-management, is to be made more efficient and extended with the survey of official interactions. Planned phases of research:
 1. Empirical surveys (data-processing by questionnaires, preparing interviews)
 2. Data induction
 3. Preparing suggestions on the basis of results
 4. Passing suggestions to different offices, authorities

The whole research is planned to be carried out together with the Gramma Language Office. Involving Hungarian language offices outside Hungary's borders, the research is to be extended to the whole Carpathian Basin.

VIII.3) Department of Finno-Ugric

Computational projects

- Connecting our databases, simplifying the searching systems / browsers.
- Finishing the morphological analysers for the two Permic languages, and Nganasan.
- The text collected and proceeded for the analysers will be accessible via internet. The same applies to the corpora of the future morphological analysers, too.

Studies in endangered languages

- In the case of Nganasan and Khanty, collecting the necessary amount and type of materials to be described/ morphological analysers etc. needs fieldwork, which, taking into consideration that these languages are spoken in Siberia, proves to be a respectable undertaking.
- Financial resources permitting, in the case of Khanty, the following interests of fieldwork can be outlined: collecting poorly documented or unknown genres of children's folklore, language variations spoken by children, as well as documenting the language of the most archaic groups (e.g. reindeer herders). This material could also be used for the purposes of language revitalization.
- In the case of Nganasan, the questionnaires aim at clearing morphological and syntactical questions.

Typology

- The typological descriptions of the Uralic languages seem to appear independently nowadays. The growing interest towards the typology of Uralic languages suggests that the individual researches should be coordinated in order to make the work even more effective. The creation of a concise Uralic, typological database will probably take several years. The conception, however, should be worked out in the next 1-2 years. The members of the Finno-

Ugric Department intend to take part not only in the preparatory work but also in the actual typological description of the languages.

VIII.4) Department of Sociolinguistics

- Analyses will be made and studies will be written using the *Budapest Sociolinguistic Interview (BSI-2)* data. Envisioned topics include: the variability of the long vowels *í, ú,* and *ű*; contextual styles; variation in vowel harmony; etc. Results will be presented at the 15th Hungarian Sociolinguistic Conference, to be organized by the department.

- Continuing projects: *The longitudinal study of language shift; Dimensions of Linguistic Otherness: Prospects of Maintenance and Revitalization of Minority Languages; Hungarian Sign Language; Nézetek-B2* (=real-time changes in Budapest); *Bilingual Education for the Bilingual Hungarian Communities in the Carpathian Basin; Gypsy studies*

VIII.5) Department of Lexicology and Lexicography

- The selection of headwords of the letter E and É will be finished, the selection of headwords of the letter F commences. The edition of the entries of the letter B will be completed, followed by the general edition and the making-up of this volume, and in late 2008 Volume 3: B of the Comprehensive Dictionary of Hungarian will be published.

- The compilation, edition and the supervision by the consulting editors of the C-Dzs entries proceeds, as well as the morphologic coding of the words.

- Department research fellows will continue their individual work in higher education and in public scholarly life.

- The work on the Hungarian version of EWUng proceeds.

VIII.6) Department of Historical Linguistics and Dialectology

- Historical Syntax of Hungarian: drafting of further chapters, starting the editing process of the present volume on Middle Hungarian.

- In cooperation with the Sermones project, completed at Eötvös Loránd University, preparing the text of the Érdy Codex for the implementation of a lexical search programme.

- Research on the pragmatic aspect of Middle Hungarian grammatical changes.

- Research on the lexicographic modelling of etymological relationships.

VIII.7) Phonetics Department

The investigations started and/or continued in 2007 will be carried out in 2008. There will be investigations that will end next year within the coarticulation area and also concerning the prosodic investigations of speech styles while there will be others that will definitely take a longer time across several following years (like the building of the spontaneous speech data base, BEA). The acoustic-phonetic analysis of specific speech defects (like dysarthria) is planned to start in 2008. The book series entitled *Beszéd kutatás* [Speech Research] is also planned to be published.

VIII.8) Department of Experimental Linguistics

1. Neurolinguistic research

Participants: Zoltán Bánréti, Éva Mészáros, Katalin Szentkúti-Kiss, Judit Szépe, and 3 PhD students.

1.1. Writing a comprehensive study summarizing the characteristic features of morphosyntactic agreement operations of Hungarian agrammatic aphasics based on the results of production and sentence comprehension test conducted so far. Building on empirical generalizations, conclusions relating to the theory of grammar will be drawn.

1.2. Investigating the comprehension of Hungarian transitive sentences containing non-accusative predicates and causative verbs in case of three sentence construction types: (i)

topic-predicate, (ii) multiple-topic, (iii) topicless structures, with both canonical and non-canonical thematic role order sentence variants.

1.3. Prolonged gathering and processing of corpus data containing deviances of phonological sequence organisation with the intention of producing a structured database, using primarily data from paraphasia, lapses, mistakes of healthy and demential subjects in old age and, secondly, data of healthy and pathological language acquisition processes.

1.4. Working out the categorization of information narrowing operations manifested in phonological mistakes/lapses, and performing quantitative investigations on a lapse corpus of spontaneous speech.

1.5. Developing therapeutic material for aphasia patients: preparation of a systematic package for practicing noun/verb retrieval promoting both individual and institutional therapy of aphasia. The pictorial test/practice material will be recorded on digital storage media. Financial resources for the project are to be provided by research grants.

2. Psycholinguistic research:

Participants: Ágnes Lukács and Bence Kas.

2.1. Concluding the tests on language development disorders. Evaluating and summarising the results. Psycholinguistic analysis of analogy processes operable in linguistic knowledge

3. Sociolinguistic research

Participants: Csilla Bartha and members of the Department of Sociolinguistics

3.1. Continuing collective projects and individual research work started in 2007 on the following topics:

- Research of language shift and linguistic revitalization in East-Central Europe.
- Continuing the project DILING, organising the second DILING workshop (to be held in the Ukraine). - Preparing policy recommendations for individual countries and for the entire region.
- Forms of language shift and language maintenance in minority communities (comparative research).
- The theory and practice of language shift, language maintenance and linguistic revitalization.
- Forms of social intercourse, bilinguality, minority and majority attitudes, prejudices in linguistic communities living in Hungary.
- Approaches to classifications of bilinguality
- Ideologies related to bilinguality.
- Bilinguality and language shift in the Mura region (synthesis of research results)
- Paradigms of sociolinguistics: quantitative and qualitative methods.
- Sociolinguistic interpretations of style.

4. Research on languages spoken by Roma communities

Participants: Csilla Bartha, Anna Orsós, and Andrea Szalai

4.1. Boyash linguistics: the Boyash descriptive grammar (which also includes information on phonetics/phonology, morphology and syntax) to be published in 2008 is an unparalleled venture not only of Hungarian but international linguistic research. Impact analysis of the published volume, starting work on designing a new corpus that is indispensable for Boyash language teaching as well as for extending the Boyash vocabulary to better reflect current usage and also for appearing on the international scene.

4.2. Romani linguistics: starting the topic *Romani discourse on legal and political matters. Discourse positions, oath, responsibility* The aim of the research is analysing linguistic ideologies in the legal context of the community, with a special regard to producing and sharing responsibility by means of communicative interaction.

4.3. Linguistic variables and social barriers: the planned project focuses on an understudied issue of Romani linguistics, intradialect variation. Speakers do not only perceive linguistic variations of their dialect but also identify them, producing and maintaining social barriers by emphasising or ignoring differences.

5. *Speech research*

Participants: Ilona Kassai and PhD students

5.1. Investigating the operation of exchange, insertion, deletion, shortening and lengthening of sounds (reflected in the database of the Hungarian Dialect Vocabulary) as general principles valid for young children acquiring language as well as for elderly speakers and aphasic patients. Word structure reordering methods of dialect speakers, such as deletions, insertions, metatheses, can be conceived as unconsciously performed tests of the psychological reality of phoneme connections, which may give clues for the empirical characterisation of one of the key concepts in language description, well-formedness.