Scientific report of the
Research Institute for Linguistics,
Hungarian Academy of Sciences,
for 2016
I. Main duties of the research unit in 2016

Basic research


Applied research


Innovation and development in language and speech technology


II. Outstanding research and other results in 2016

a) Outstanding research results
Basic research

Theoretical linguistics

“Momentum” research group – Division of Labor in the Grammar of Scope: Experiments examining the interpretation of contextualized sentences revealed a double dissociation between the logical scope of universal quantifiers and their Focus or Given information structural status. This provides empirical evidence that in the domain of linguistic competence, the interpretation of logical scope and the interpretation of information structural roles exhibit a modular organization, despite some previous claims in the literature to the contrary. In another subproject carried out in a research co-operation it was demonstrated that a recent biclausal analysis of sentences involving contrastive left dislocation is unable to account for a number of properties of the construction, and should be analysed instead in terms of two alternative mono-clausal structures. Comprehensive Grammar of Hungarian: In the research aiming at a theory-independent grammatical description of Hungarian, the first two volumes of a series expected to be eight volumes altogether - a comprehensive study of noun phrases - was ready for publication by the end of the year and was submitted to the publisher. Analogy-based, laboratory and formal phonology and morphology: Variation in the inflexional system of Lovari was described in an analogy-based framework. Computer-based tools for morphological analysis and for prosodic stylisation were developed. They found that English /l/ shows a similar pattern to other glides in this language. Prosodic experiments showed that the duration ratio of vowels decrease with increasing speech rate, and that word stress is only measurable acoustically if a sentence accent is simultaneously present on the syllable. According to a corpus-based study, self-directed questions do not carry prosodic features that are typical for real questions. In collaboration with an external partner, a morphological annotation system was developed for the output of the Hungarian morphology analyser software, which can be applied both for computational and linguistic processing. The grammar and pragmatics of interrogatives: the grammar and pragmatics of interrogatives: a dilemma for distinguishing V-to-I movement from Verb Second in subordinate clauses was pointed out and reflected on. An account was developed for the absence of infinitival polar interrogatives in German. Semantico-pragmatic principles for classifying the ‘bias profiles’ for polar interrogatives were proposed. Semantic distinctions between different forms of negative interrogatives, polar interrogatives containing embedded declaratives and the particle ugye in Hungarian were investigated. Novel observations were made about the structure of the topic field of Hungarian clauses. Research on formal grammar, semantics and pragmatics: The manuscript of the monograph titled Dependency and Directionality has been submitted to Cambridge University Press, it is expected to be published in 2017. Two papers on syntactic theory have been accepted for publication in leading international journals (Linguistic Inquiry, Syntax). Discussing the so-called one-dimensional model of syntax, under development, it has been shown that this new model is capable of accounting for two major asymmetry effects of symmetric syntax models (the direction of movement operations, and the directedness of head-complement adjacency).

Research on historical linguistics

MTA “Momentum” Research Group for Computational Latin Dialectology: Cca 6,500 digital data forms were added to the online database (“Computerized Historical Database of the Imperial Age”; http://lldb.elte.hu), as well as the database, software also underwent significant developments. Linguistic and dialectological data evaluation were also continued, and the results of evaluation were published. The theoretical background for the dialectology of Latin inscriptions and some subtasks in the processing of the sound history of the Latin dialects (e.g. approaches to b-w fusions from the aspects of dialectology and contact linguistics) were elaborated. The analysis of the so-called curse tablets proceeded, and further curse tablets
were published. The characteristic features of the Alps-Danube-Adriatic region of the Roman Empire have also been processed, including the checking of the data originating in this region. In May, a successful project conference was organised. The Herman József Award for Young Researchers (for 2015) was granted to a member of this team. Hungarian diachronic generative syntax: Ugric relics have been identified in Old Hungarian (e.g., the determiner-like use of possessive agreement; evidential paradigms reanalysed as perfect tenses). The grammaticalization of való into a functional element, as well as cyclic grammaticalization processes involving Hungarian postpositions, adverbial case endings and verbal particles, as well as the changing distribution of definiteness-marking modifiers and determiners in the noun phrase have been investigated. Normalization of the text of the Jordánszky Codex was finished, together with the first version of the corpus of Bible translations. All normalized texts are accessible via the search page of the Old Hungarian Concordance (http://omagyarkorpusz.nytud.hu/hu-search.html). Additionally, a new corpus query device was developed that can list parallel Bible translations from different diachronic stages and different languages. Hungarian historical linguistics: There was a significant increase in the proportion of disambiguated texts in the corpus containing informal texts (the size of the texts that went through disambiguation amounts to 2.5 million characters). The following linguistic variables were investigated using a variationist approach: a) competing word order patterns in negative sentences in three historical periods; b) grammatical encoding of temporal continuity; c) patterns expressing conditional and concessive subordination; d) history of relativizers on the basis of different types of sources. Historical pragmatics: Specific types of strategies of language use (threats and incriminations) were analysed in texts of witch trials. Anatolian linguistics and general Indo-European studies: investigations focused on questions of epigraphy, etymology, historical phonology, contact linguistics, semantics, lexicology, as well as onomastics concerning several Anatolian languages including Luwian, Carian, Sidetic, Hittite etc. In addition, the editing work of the electronic etymological dictionary of the languages of Asia Minor continued.

Uralic linguistics
Developing of the chapters on the Khanty language, phonology, and the characteristics of the Ob-Ugric languages continued, and was partially completed, for The Oxford Guide to the Uralic Languages. Syntactic research: Udmurt, Khanty and Nenet databases of both older (20th century) texts and contemporary recordings have been set up, transcriptions underwent standardization, translation, and morphological annotation. Various manifestations of the SOV → SVO change have been shown. The research on Tundra Nenets concentrated on certain aspects of the word order of the language, in particular on the following: the basic word order of the declaratives, the informational structural role of the elements in clauses with non-canonical OS orders, as well as the order of the locational constructions. In the latter case, the effects of the Russian–Nenets language contact were sharply focused on. Postverbal elements in Khanty dialects were analysed in texts from different periods, dialects, and genres, including the potential motivations behind this syntactic phenomenon, i.e. inner development or Russian influence. The investigation of Udmurt non-finite clauses included the following topics/research questions: adjunct control, agreement, case-marking of the subject as well aimed at tearing apart the non-finite clauses and deverbal compounds (partly from a comparative perspective). Sociolinguistic studies: Saami studies on bilingualism continued with in-depth interviews and their transcription. The analysis of minority identities in three arctic cities focused on exploring the peculiarities, similarities and differences in the linguistic landscape of each research destination. Anthropological linguistics: The investigations in the language of Mansi heroic epic have been completed. Typology: the lexico-semantic investigations of the question words in the Ob-Ugric languages were added to the former
similar research concerning the Samoyedic languages. **Lexicology:** Tendencies in lexical borrowings were researched using the materials produced by methods of language technology.

**Applied research**

**Experimental linguistics**

*The grammar of quantifiers and linguistic relativism:* A volume of studies on linguistic and cognitive aspects of quantification has been edited, due to come out in 2017 as a part of Springer’s *Studies in Theoretical Psycholinguistics* series. New hypotheses have been proved about children’s construal of quantifiers. The different theories concerning the semantics of structural focus in Hungarian have been evaluated, and the semantic relationship between *csak* ‘only’ and structural focus has been clarified in the light of new experiments. A series of testing has been launched about the formation of the concepts of space and time in small children, and the relation between the two concepts. **Neurolinguistics:** In the research of analysing spontaneous speech of aphasic people, the characteristics of the point of view’s organization, and aphasia-type-dependent and -independent phenomena, with particular attention to egocentricity, deictic relations, the pattern of appearance of exophoric and endophoric references were explored. The results show that in questions of referring to persons, several factors are independent of the type and the severity of aphasia, but other factors — e.g. the type of appearance of egocentric expressions — had increased in direct proportion to the seriousness of aphasia. These research results might be useful for aphasia therapies. The diagnostic test method, based on spontaneous speech analysis, continued to detect Alzheimer’s disease and mild cognitive impairment (MCI). In collaboration with the Department of Artificial Intelligence Research Group (University of Szeged) software was developed to automate analyses and evaluations, which has a nearly 83% precision rate in comparison to the manual measurements. International trial testing of examination procedure continued in research cooperation with American, German and Spanish research groups. The results of the morphological and syntactic analysis in spontaneous speech transcripts of MCI persons suggest that primarily morphological characteristics are those which can help to distinguish healthy control individuals from people with MCI. **Psycholinguistics:** Psycholinguistic and psychoacoustic experiments have been prepared within the study of neuro-cognitive predictors of early language development. These experiments will serve as the output measures of child development in the research. The study of possessive morphology in children with language disorders showed that plurality in the context of possessive markers makes the production of word forms difficult. Children with language disorders perform below age controls in the production of complex word forms with possessive suffixes, and with specific error types. The development of the pragmatic competence (irony and implicature comprehension) in 4 to 8 year-old children correlates with cognitive flexibility and specific aspects of ToM ability. Children with ASD show delayed development. Data from two experiments investigating the exhaustive reading of Hungarian preverbal focus have lent support to the view that both the structure and the mental processes of construal associated with it strongly depend on the number of alternatives that can be chosen in the experiment task. This result suggests the context-dependence of the given structure. Preliminary results of research on new approaches to the acquisition of recursion by Hungarian children show that the experimental method under pilot testing (the so-called ‘coloring task’) may be suitable for testing children, and that the meaning of postpositions may affect the processing difficulty of the given structure. A study of perspective taking and egocentrism in 3 to 7 year-old children found that meta-representational ability has a strong effect on perspective taking and shows a negative correlation with egocentrism and performance. Work has started on the legacy of Zita Réger, some 1500 cassettes of sound-recordings, to develop a corpus out of it. The digitalization of the whole of the sound material
(recorded Hungarian and Romani speech of children and adults) has been completed. The digitalized sound files have been initially mapped for content, language, communicative genre, recording time, and technical parameters. Anthropological linguistics: A monograph titled Pragmatic role and social gender – Curses and vows in the language use and linguistic ideologies of Transylvanian Roma communities has summarized the results of two and a half years of anthropological linguistic fieldwork in native Romani communities. Another paper has been completed, too, titled ‘Curse, gender and teasing in Romani child-directed speech.’

Research on bi- and multilingualism
Sign language research: There were important results in the development of the sign language corpus processing, improvement of the annotation system, further development of the sign language online dictionary, as well as continuous processing of data and materials collected during the SIGNificant Chance TÁMOP project and the publication of project results. In connection with this, as a result of a previous methodological research a study was prepared titled Towards the language conscious school. In the frame of a new methodological research grant awarded in 2016, the LanguageChance Educational Research Group was established, the partner network was set up and expanded, and an education methodological workshop was organized. Development of sign language assessment methods and trial tests was planned for measuring Deaf children’s literacy, writing and reading comprehension. The experimental preparation of these trial tests and methods would be unique and pioneering in Hungary, and would be a significant development concerning Deaf children’s equal opportunities. Handshape perception research was carried out among Hungarian Sign Language users, the results were evaluated. 90 hand shapes were identified which may be considered as possible phonemes. The results of this research are significant scientific contributions not only to strengthening the scientific background of the sign language dictionary but to international aspects as well. Deaf colleagues contributed to the creation of a corpus research supporting system by which spoken language effects and limits of spoken language frames can be minimalized by providing exclusively visual annotation, it is also a good solution for the reduction of the time and efforts spent on annotation. Another aim of the created system is to support increasing searchability and accessibility of the Hungarian Sign Language Corpus. One period of the research on bilingualism, especially on sustainable bilingualism within minority communities ended with a publication of an international journal article. In addition, research projects were carried out in the fields of sociolinguistics, cognitive linguistics, stylistics and discourse analysis on actual social issues (e.g., migration and language shift, multilingual communities, discourses of ageing, language and identity, internet language use of minorities, linguistic expression of emotions, drug consumption and language use); dissertation, studies and conference presentations were prepared on the results.

Phonetics
Phonetic investigations confirmed language-specific patterns of vowels’ phrase-final lengthening, temporal and spectral properties of stop consonants, phonetic characteristics of co-articulation in consonant clusters, and specific behaviour of reduction, temporal and glottal processes in spontaneous speech. A new analysis was conducted concerning the inherent speech planning processes (silent pauses, lengthenings), and the results were interpreted as the interrelations of speech planning and execution of spontaneous utterances. In addition, disfluency phenomena were examined in spontaneous speech of speakers who lived 40 years ago. Temporal analysis of word stems and suffixes provided new results in language acquisition across various ages. Interrelations of individual speech patterns and some physical parameters were identified for developing speaker’s profile depending on speech. An automatic system was established for the purposes of speech recognition and text extracting in
spontaneous speech. A doctoral dissertation has been successfully defended. Various papers were published in national and international journals and edited books.

**Lexicology-lexicography**

Comprehensive Dictionary of Hungarian: The revising and standardising of the entries of headwords beginning with di-ek were finished, Vol. 6. (Di–ek) was published. The web version of the so-far published volumes was made available at [http://nagyszotar.nyttud.hu/index.html](http://nagyszotar.nyttud.hu/index.html). The work on the entries of the next volume (7. El–ez) progressed. In the reporting year, approximately 1200 new entries from the material beginning with di−ez were written and 1100 entries were finished (including revision, standardization and morphological coding). 70 percent of the list of headwords beginning with f and 60 percent of the h- and k-lists were completed. The OTKA project of the New Hungarian Etymological Dictionary was finished. Entries of headwords beginning with a and b were completed and made available. Both of these dictionaries have the utmost importance in the exploration of the present and the past of our cultural heritage and of the Hungarian language and also in providing access to these, both for public and academic audiences.

**Language consultancy**

In 2016 the Language Consultancy Service answered 7000 questions regarding orthography, language use and name giving. Colleagues have held several lectures and published articles in the fields related to orthography, language consultancy and name giving.

**Innovation and development in linguistics**

**Language technology**

Language technology research group: As a result of extensive co-operation the e-magyar digital research infrastructure has been completed, which integrates the text processing tools so far developed within Hungarian language technology into one unified toolchain. The system contains a new tokeniser and morphological analyser, and several other text processing tools based on these. The system is accessible both online and in the graphic environment of the GATE language processing framework. Furthermore, the source code of the modules is also open access. Any given Hungarian text may be analysed by the entire text processing toolchain on the website [http://e-magyar.hu](http://e-magyar.hu). e-magyar can serve as a common platform for developments in language technology in the future. Besides serving researchers working on Hungarian texts or speech the system is equally targeted at lay people and industrial applications using language technology, while having a potential use in education, as well. Research has been continuous in several corpus-building projects concerning Hungarian translations of the New Testament, small Uralic languages and their syntax, as well as multimodal communication. The Hungarian National Corpus has reached 1 billion parsed tokens. An important publication has been published proposing an approach of translational equivalence via conditional probability. There is an ongoing development and troubleshooting of the helyesiras.mta.hu website based on user feedback. The automatic suggestions offered by the respective tools have been updated according to the changes to the 12th Edition of the Orthography Regulations. Several automatic methods of creating bilingual dictionaries have been tested in the Finnish-Hungarian joint OTKA project “Language Technological Support of Creating Online Content in Finno-Ugric Communities”. Every method tested proved successful in producing proto-dictionaries with several hundreds of vocabulary items for different language pairs. Manual validation and evaluation is in progress. Research Group for Mathematical Linguistics: wordsim, a machine-learning based system for measuring lexical similarity has been set up and tested in an international competition, where it has performed with a higher precision than similar systems have reached so far. A Speech Archive has been
created within the e-magyar system, together with related open source speech technology software supporting research in speech technology. Work on word representations acquired by neural networks (word embeddings) with a focus on polysemy has continued, as well.

**Speech technology**

The development of the multi-functional BEA spontaneous speech database (consisting of over 365 hours of recorded speech material of 459 informants) is noteworthy both nationally and internationally. In 2016, 65 new recordings were made, while speech sound level annotations of 36 subjects’ recordings were completed. The overview table of the database, which provides for easy search in the database, has been in continuous expansion. In 2016, it was a great leap forward that the files of the BEA large spontaneous speech database were shared with researchers. The open access platform of the database is ready and can be accessed from the following homepage: [http://fonetika.nytud.hu](http://fonetika.nytud.hu).

**b) Science and society**

The institute functions as an independent academic information center, the main instrument of which is its website, receiving a monthly average of 57,000 visits. Its web pages offer several publications of potential interest for the wider public too, for reading or download, while the NyelvészInfo electronic mailing list sends out regular newsletters on academic events, conferences, funding calls, and other current topics, to 570 external addresses. The institute is also in charge of editing the contents of Nyelvész Portál [Linguist Portal] ([http://www.nytud.hu/portal/index.html](http://www.nytud.hu/portal/index.html)), an independently run website partially associated with the institute, too.

The **library**, as a specialized library of linguistics, offers various scholarly services. Its main duty is to support the work of the institute’s researchers, and takes care of collecting the domestic and international literature needed for the ongoing projects. It also organizes and coordinates the electronic publication of the research results (studies, databases) of the institute. In 2016, a significant amount of resources was obtained, and the digitization of older paper-based materials was going on.

**Publicly accessible linguistic databases**

Reflecting the intense interest in language use by the general public, several language resources, developed in the department, continue to receive considerable public attention (the Hungarian National Corpus, for instance, has nearly 11,000 registered users, who fielded about 85,000 queries in 2016). These resources are digital repositories of Hungarian linguistic cultural heritage, serving as reference points not only for scientific research but also for public awareness by all those interested in their native language. Our uniformly encoded historical corpora are unique language resources that preserve Hungarian linguistic heritage. The major publicly accessible databases are the following:

- **The Comprehensive Dictionary of Hungarian** ([http://nagyszotar.nytud.hu/index.html](http://nagyszotar.nytud.hu/index.html)): the entries of the volumes of the dictionary published to date (vols. I–VI.), with a state-of-the-art query interface and auxiliary functions. From early 2017 on, a total of 15,567 entries with some 50,000 meanings, and about 170,000 example sentences will become available.
- **The Hungarian National Corpus** ([http://clara.nytud.hu/mnsz2-dev/](http://clara.nytud.hu/mnsz2-dev/)): a representative sample of written present-day Hungarian. In the fall of 2016, after incorporating a large amount of texts from the media, its size reached one billion words.
- **Uralonet** ([http://www.uralonet.nytud.hu](http://www.uralonet.nytud.hu)): an online etymological database of the Uralic languages for educational and research purposes.
- **Hungarian Generative Diachronic Syntax – Old Hungarian Corpus** ([http://omagyarkorpusz.nytud.hu](http://omagyarkorpusz.nytud.hu)): a unified corpus of all old Hungarian texts and codices containing more than 2.2 million running words.
- **The Hungarian Historical Corpus** ([http://clara.nytud.hu/mtsz](http://clara.nytud.hu/mtsz)): a 30-million-word corpus of historical texts (1772–2010) with an online query interface. In 2016 it was moved to a new corpus handling system, and was given a highly functional user interface.
- **The list of sources and authors cited in The Comprehensive Dictionary of Hungarian** ([http://nszt.nytud.hu/forr.php](http://nszt.nytud.hu/forr.php)): a list of about 38,000 items.
- **BEA Hungarian Spontaneous Speech Database** ([http://www.nytud.hu/adatb/bea/](http://www.nytud.hu/adatb/bea/)): a multifunctional spontaneous speech database covering the speech of present-day Hungarian native speakers from Budapest, with over 365 hours of recordings, from 459 informants.
- **The Budapest Sociolinguistic Interview (BUSZI)**: ([http://buszi.nytud.hu/](http://buszi.nytud.hu/)): the materials of a large-scale tape-recorded survey of the Budapest variety of Hungarian, carried out on a representative sample of 251 Budapest residents, originally yielding 600 hours of recordings, of which almost 100 hours of annotated material is available for research.

**Non-profit services and advisory activities targeting the general public**

Complementing the already existing telephone and e-mail public services, there are two additional online interfaces, in accordance with current requirements. The **Orthography Consultancy Portal** ([helyesiras.mta.hu](http://helyesiras.mta.hu)), which has been in operation for more than two years now and was voted the best Hungarian orthography advisor service by a professional survey, has become the central source for acquiring orthographical information in Hungarian. In 2016, it gave advice on over 1.8 million occasions and the number of returning visitors exceeded 850 thousand. The permanent exhibition of historical phonetics has given a guided tour for the general public about 80 times in 2016.

**Popularizing science**

The Institute is in continuous contact with important strata of Hungarian society in a variety of ways and via a variety of channels. Researchers of the Institute organized three important events of communicating their achievements to the wide public in 2016. As part of the program of the **Hungarian Science Festival**, researchers of the Momentum Research Group for Computational Latin Dialectology held two sessions of talks in the main building of the H.A.S., titled “From spellings to dialects – in the wake of the regional varieties of the Latin language”, and “A Computerized Historical Database of the Imperial Age”, respectively (November 8th and 17th). Under the auspices of the **Researchers’ Night 2016** program series, a public lecture was given on the psycholinguistic applications of eye-movement tracking, and the significance of its results, titled “Eye-movement tracking: Insight into the mental processes of comprehending utterances.”

Researchers at the institute actively participate (both individually, and as organized by the institute) in **popularizing science**, as well as in clearing up linguistic misconceptions and prejudices concerning language, which are widely spread in society. Lectures popularizing science and aimed at the general public have also been given. Numerous educational linguistic articles have been published at the web portal **Language and Science** ([http://www.nyest.hu](http://www.nyest.hu)), many of which contain new scientific results, formulated in a way which is understandable for the widest possible audience, so that the public could get a grasp on the various subfields of linguistics, their accepted methodology, and their results. Apart from this, they also published a number of popularizing articles in various outlets (e.g., about typological research, which generally gets little coverage in popularizing material, as well as language history, diachronic change, and revitalization). They actively participated in
preparing a broadcast on language development disorders (in the series Szonda ‘probe’ of Radio Kossuth, [http://nava.hu/id/2436746/](http://nava.hu/id/2436746/)), and a film on specific language impairment ([https://www.youtube.com/watch?v=bulnEJSsfA4&t=196s](https://www.youtube.com/watch?v=bulnEJSsfA4&t=196s)), connected to the Raising Awareness of Language Learning Impairments (RALLI) international campaign. One researcher served as consulting expert and anchorman of the popularizing series Szószátyár ‘Verbose’ of Klubradio. The institute’s Research Center of Multilingualism is constantly present in the media, actively publishing news, webpages, tweets, and facebook notes.

Within a framework of cooperation with the School for the Hard-of-Hearing, in compliance with the agenda set up in 2015, three popularizing lectures were held in the 2015/2016 academic year for teachers of the school. In the program series of Finno-ugric Capitals of Culture, or organized by schools, several talks have been given. On the occasion of the Day of Uralic Peoples a quiz game was conducted for secondary school pupils. In Croatia, four lectures were given on the situation of Hungarian spoken in Croatia. A course was held on language education in schools at the 24th Bolyai Summer Academy, Kolozsvár (Cluj-Napoca), titled “The modern educator – changing roles and skills, methods of change.” One researcher gave a talk on request from the Federation of Mother-Tongue Cultivation at the 23rd National Junior Mother-Tongue Cultivation Camp (in the Museum of the Hungarian Language, Sátoraljújhely–Széphalom), titled “A concise history of Hungarian dictionaries”, while three others gave popularizing talks at the Újbuda Pedagogy Days, and another one at the Scientific Café program of Csoódák Palotája (Palace of Wonders), titled “Colorfully popularizing linguistics.” They participated in the selection, as well as in the training of the Hungarian team for the International Linguistics Olympiad. Several researchers are members of the bureaux of the Kalevala (Hungarian–Finnish) Friendship Society and the Reguly Society, which aim to popularize the Finnish language and culture, as well as the languages and cultures of the Uralic peoples, respectively. One other researcher put her experience gained in dealing with endangered languages to good use in supporting the work and activities of the Soroksár Local Council of the Ethnic German Minority (continuing the ‘Schwabe’ tradition, organizing and carrying out revitalizing programs).

**Science and society**

The direct benefits of sign language research are the increase of the educational and labour market opportunities of the Deaf community, the support of their better access to scientific results, the support of their involvement in scientific and public life; in the long term the creation of equal opportunities for Deaf children, and providing their access to language, information and sufficient cognitive development.

The subject-pedagogical, methodological research is directly related to the exploration of the actual and urgent problems, and the actual and socially debated features of Hungarian education, and to problem solution. Studying the language and language-related issues, the exploration of the latest international scientific results and their involvement in education, as well as the direct practical studying of practices and problems of Hungarian education may increase success in several areas of education and in several learning groups. The significance of the Languag-E-Chance project launched in 2016 is that although its primary focus is on the education of disadvantaged, marginalized – in terms of access to science as well – groups of language users such as Deaf and Romani bilingual children, yet – by creating the language conscious school approach, destroying boundaries between disciplines and building on increasing awareness of the cognitive, emotional and social advantages of multilingualism – it wishes to create such methods and educational aids (e.g., e-learning materials, Knowledge platform) which increase the educational success of children belonging to the majority as well, and which can be effectively used in all educational institutions.
The research on the varieties of Romani makes great contributions to answering one of the most salient questions of contemporary Hungarian society: the integration of the Roma. By fieldwork done in Hungary we can gather Romani linguistic data hitherto unknown to the scientific community. Such data may shed light on the independence and diversity of the Romani language and by exposing these to the public may facilitate the decrease of prejudice in the society towards the Roma. Moreover, the investigation of the interactional patterns in Roma communities can enlighten us on the kind of linguistic input surrounding the children growing up in those communities. The knowledge thus gained can then be utilized in teacher training, and the education of Roma children, facilitating a more successful intercultural communication.

The CDI-3-H language screening instrument based on parental report that has been developed and normalised in 2015 has been introduced to clinical practice within a joint model program cooperating with Eötvös University Gusztáv Bárczi Faculty of Special Education and the Pedagogical Services Centre in Budapest. The screening tool is now used systematically in several districts by speech and language therapists as an inclusionary instrument to preventive language therapy.

It is an important indicator of the acknowledgment of achievement on behalf of the society that a researcher of the institute was awarded a Prima Junior Prize in the category of ‘Science’ by the Prima Primissima Foundation.

III. A presentation of national and international R&D relations in 2016

National relations
In addition to the many ongoing cooperative projects dating back to earlier years, new links were established in 2016 with the following partner institutions within Hungary:

- Cooperation has started with the CARTOUR Touristic Services Llc. and TRAVELWEB Informatics and Commercial Services Llc. within the framework of the GINOP 2.1.1 project. The cooperation aims at creating an intelligent tourist recommendation system. Language technological tools will aid finding relevant tourist websites, storing their information in a structured database, as well as offering a coherent package to the client.

- A joint model program has been started with Eötvös University Gusztáv Bárczi Faculty of Special Education and the Pedagogical Services Centre (FPSZ) in Budapest aiming to introduce systematic language screening for 3-year-old children in the practice of speech and language therapists. Screening is followed by preventive language intervention in order to prevent or reduce severe language disorders and their serious impacts on learning and socialization. Researchers in the RIL–HAS and ELTE are jointly working on development and longitudinal study of screening and effects of intervention. The expert group regularly informs the Klebelsberg Institutional Centre about the results.

- As a result of joint work within the Hungarian language technology community the e-magyar project has been completed. The work has been coordinated by the Research Institute for Linguistics, with the following partners: University of Szeged, Institute for Computer Science and Control – HAS, Pázmány Péter Catholic University, AITIA International, Inc. and Morphologic Llc. The infrastructure that has been created as a result of the efforts is going to serve industrial partners, the language technology community, researchers and students untrained in programming but interested in language and text/speech processing, as well as the general public.

- Co-operation started with the Institute of Cognitive Neurology and Psychology of HAS within the framework of an ERC Advanced Grant. The topic of the planned research: Multimodal alignment in face-to-face communication. The aim of the research is to
analyse the listener’s reactions and strategies to the speakers’ verbal utterances using various objective measurements.

- Two members of the Phonetics Department participate in the NKFIH-K-120234 project entitled ‘Temporal characteristics of children’s utterances using speech corpus’. The aim is to analyse articulation and speech tempi as well as durations of speech sounds in children’s spontaneous speech.

- Co-operation started with the research group of Lendület (‘Momentum’) Programme entitled Lingual Articulation (at the Phonetics Department of ELTE University). Analysis will provide new data concerning the measurable properties of the tongue movements during articulation using ultra sound method.

- A new cooperation was launched with the Institute of Information Technology of the University of Szeged for converting morphologically analysed Old Hungarian texts.

- Cooperation agreements were made with primary and secondary schools and with the Institute of Hungarian Literature and Cultural Studies of Eötvös Loránd University related to the HAS Language-E-Chance Educational Research project.

**Higher education**

The institute has been operating (jointly with Eötvös Loránd University) the undergraduate and graduate training programs in theoretical linguistics for 26 years now, providing for the courses, tutoring, and thesis supervising both at MA and PhD levels. Apart from this, the institute’s researchers taught courses in 2016 in the following departments and programs: Eötvös Loránd University, Faculty of Humanities: Department of Chinese Studies, Department of Hungarian Historical Linguistics, Sociolinguistics and Dialectology, Department of Present-Day Hungarian, Department of Phonetics, Latin Department, the Applied Linguistics Doctoral Program, and the Doctoral School of Linguistics; Péter Pázmány Catholic University, Faculty of Humanities: Department of Hungarian Linguistics, and the Linguistics Doctoral School; Eötvös Loránd University, Bárcezi Gusztáv Faculty of Special Education: Speech Therapy Expert Group; Péter Pázmány Catholic University, Faculty of Humanities: Department of Hungarian Linguistics, Linguistics Doctoral School; Péter Pázmány Catholic University, Faculty of Information Technology; University of Szeged, Faculty of Humanities: Institute of English and American Studies, Department of General Linguistics, Department of Finno-Ugric Linguistics, as well as the Theoretical Linguistics and Uralic Studies programs of the Linguistics Doctoral School; Károli Gáspár University of the Reformed Church, Faculty of Humanities: Department of Hungarian Linguistics; Semmelweis University, Center for Language Communication; and Strossmayer University of Osijek. Two researchers served as heads of doctoral schools.

The researchers of the institute actively participate in various academic forums in Hungary: they presented nearly 100 lectures at over 50 national conferences and other events, and worked in 27 academic and scholarly committees and organizations.

**International relations**

The institute has collaborated with numerous foreign institutions, and establishes more and more international contacts each year. In 2016, the following new cooperations have been launched:

- The Institute has joined the CLARIN ERIC organisation as the Hungarian coordinator of the European research infrastructure network ‘CLARIN’ (Common Language Resources and Technology Infrastructure). CLARIN’s mission is to build a network of research infrastructure that supports researchers in the humanities and social sciences with
language technological tools and data, as well as with knowledge sharing. The Hungarian HUN-CLARIN network comprises altogether seven universities and research institutes of the Hungarian Academy of Sciences. Through joining the organisation HUN-CLARIN partners will be able to become involved in the building of a European level infrastructure and to adapt cutting-edge standardised tools that may later on be made available to Hungarian researchers.

- The “Momentum” Research Group for Computational Latin Dialectology started cooperation with the Department of Greek, Latin, Romance Languages and Semitic Languages of the University of Barcelona, and the Center for Classical Studies of the University of Lisbon.

- A new cooperative research project was launched with Tartu University (Estonia) titled Finno-Ugric Languages - Contact effects, investigating the syntax of Uralic languages.

- New collaboration started within the framework of the project titled Ob-Ugric Database: analysed text corpora and dictionaries for less described Ob-Ugric dialects with the Ludwig Maximilian University of Munich, for transcribing to IPA, and morphologically analysing Khanty texts from Surgut.

- Various research groups of the institute took part in the SIREN project of the Linguistic Data Consortium, preparing several chapters of a handbook.

- A researchers’ exchange began with the Institute of Linguistics of Göttingen University in the area of experimental approaches to logical scope and information structure.

- Cooperation has started between researchers of the Uppsala University and the Budapest University of Technology and Economics, which has led to the submission of a joint "COST" project proposal aiming at using language technological methods in decrypting historical ciphers, as well as at creating a structured database of historical ciphers.

- A conference was organized, and a journal issue edited, in collaboration with researchers of the Universities of Hamburg and Tromsø, who are founding members of the ACL Special Interest Group on Uralic Languages.

- The editing of The Oxford Guide to the Uralic Languages has been ongoing in cooperation with researchers of the of the Institut für EVSL University of Vienna, and the Faculty of Language and Literature of the Ludwig Maximilian University, Munich.

- Cooperation agreement was made with Finnish primary school and university (Jyväskylä) related to the HAS LanguageChance Educational Research project.

- The second phase of Danube Region cooperation titled Mind your Language! – Fostering Language-driven heritage in the Danube Region started involving several universities and municipalities as new partners.

- Cooperation was undertaken with Stockholm University by the Mercator LearnMe project.

### Organization of academic events

**Events organized by the institute:**

13th Old World Conference in Phonology (Budapest, January); 10th Doctoral Students’ Conference in Applied Linguistics (Budapest, February); First International Workshop on Computational Latin Dialectology (Budapest, April); Beszédkutatás [Speech Research] 2016 (Budapest, October); Hungarian Science Festival 2016: “From misspellings to dialects” – in the wake of the regional varieties of the Latin language, and Computerized Historical Database of the Imperial Age; (Budapest, MTA Main Building, November); a monthly lecture series in semantics (The Round Table of Hungarian Semantics) was organized with the participation of experts from Hungary and abroad, as well as local students.

**Events co-organized by the institute:**

Second International Workshop on Computational Linguistics for Uralic Languages (Szeged, January); Uralic Syntax Days (Budapest, May); Conference of the European Child Language
Disorders Group (Budapest, May); Budapest–Potsdam–Lund Linguistics Colloquium (Budapest, July); SEAS130 Jubilee Conference (Budapest, November).

One researcher participated in the organizing committee of the Hungarian Linguistics Olympiad as vice-chairman. Another researcher participated in the preparatory work for the Conference at the 10th Day of Dictionaries (“Lexicography and dictionary use in the age of modern technologies”), as secretary of the Working Committee of Lexicography.


Guest lecturers in the institute: During the year, several foreign researchers visited the institute, many of whom also gave lectures: Paul Smolensky (Johns Hopkins University), Geraldine Legendre (Johns Hopkins University), Susan Gal (University of Chicago), Jef Verschueren (University of Antwerp), Pavel Caha (Masaryk Egyetem, Brno), Nikolaus Ritt (Universität Wien), Oliver Bott (University of Tübingen), Janina Rado (Goethe Universität Frankfurt), Silvio Cruschina (University of Vienna), Anna Kocher (University of Vienna), Jaklin Kornfilit (Syracuse University), Lucia Satinská (L. Štúr Institute of Linguistics, Slovak Academy of Sciences), Robert M. Vago (Queens College & Graduate Center, CUNY), Farrell Ackerman (University of California San Diego), Hans Kamp (Universität Stuttgart), Nate Charlow (University of Toronto), Fredrik Haraldsen (SASPRO/Marie Curie Fellow, Slovak Academy of Sciences), Laurence White (Plymouth University), Ilan Kernerman (K Dictionaries, Tel Aviv), Petra Wagner (Universität Bielefeld), Gregory Grefenstette (INRIA TAO, Université Paris Sud), Anna-Christina Boell (University of Göttingen), Thórhallur Eythórsson (University of Iceland), Jóhannes Gísli Jónsson (University of Iceland), Swantje Tönnis (University of Göttingen), Julia Bacskai-Atkari (Universität Potsdam).

Researchers of the institute gave invited lectures at various universities and research centers abroad, among others at the Universities of Barcelona, Lisbon, Uppsala, Verona, and Charles University of Prague. Moreover they taught courses to international audiences:

- LOT Winter School of the Netherlands Graduate School, Tilburg University: one-week advanced seminar on syntactic structure-building and filler-gap relations.
The EGG (Eastern Summer School in Generative Grammar), Lagodekhi, Georgia: one-week advanced seminar titled ‘From GB to Minimalism’, and another one-week advanced seminar on syntactic structure-building and filler-gap relations.

University of Bucharest: advanced minicourse titled Agree(ment).

At the Summer School of the University of Vienna (Gesprochene Sprache in der Romania: Von der Theorie zur Empirie) on the following subjects: vowel production, automatic segmentation, Praat script writing, analysis using the R statistical software.

At the Göttingen Spirit Summer School on Complex Clauses: a course on minor and special sentence types.

At the University of Wuppertal and the University of Cambridge: advanced seminars on embedded clauses containing second-position verbs in Germanic languages.

About the building of the Computerized Historical Database of the Roman Imperial Age for students at both undergraduate and graduate level in Brno and Barcelona.

Researchers of the institute participated in 91 international conferences and other academic events in 26 different countries, with 168 talks and posters.

IV. Brief summary of national and international research proposals, winning in 2016

National proposals


- NKFIH PD 121386: “Where is the Result? Decomposing the argument structure of Hungarian resultatives and motion predicates”. 01.12.2016–30.11.2019. Total support: 15.264 million HUF. Aim: To investigate verb phrases expressing result- and goal-oriented events in Hungarian and to give a detailed syntactic analysis of their argument structure. This research will thereby contribute to the study of argument structure and the syntactic configurations involved in the expression of complex events cross-linguistically.


- János Bolyai Research Fellowship of HAS (BO/00699/16/1). Topic: Perception of phonemic length of Hungarian consonants. Support: cca. 4.5 million HUF. Aim: Investigation of phonemic length of Hungarian stop consonants based on perceptual experiments (depending on the listeners’ age, gender, phonetic context, and types of geminate consonants). Results (i) shed light on the effect of the analysed acoustic phonetic parameters on the identification of phonological categories; (ii) provide an opportunity to build phonological models. Empirical data may validate the phonological categories of length, furthermore, confirm the language-specific correlations between acoustics and perception. This research contributes to the mapping of universals in speech sound perception.

- János Bolyai Research Fellowship of HAS (BO/00749/16/1). Topic: Automatic profiling of speakers. Support: cca. 4.5 million HUF. Aim: This research is an attempt to develop an automatic estimation of physical conditions of speakers based on the acoustic-phonetic
features of their speech. The basic research question is how speaker’s age, height and weight can be estimated automatically on the basis of their speech parameters using machine learning techniques. The research results will contribute to the knowledge about representation of human physical condition in speech acoustics, on the one hand, and to profiling of an unknown person based on his/her speech for forensic purposes, on the other.

- **New National Program for Excellence Scholarship of the Ministry of Human Resources (ÚNKP-16-3).** Topic: Age-dependent temporal patterns of speech. Support: 3.5 million HUF. The research aims to describe the temporal factors of spontaneous speech across the life span (children, adolescents, adults, and seniors, altogether 100 speakers; ages between 3 and 90 years). Analysis focuses on temporal patterns of spontaneous utterances focusing on articulation and speech rate as well as some parts of segmental structures in relation to ages. Findings will be relevant and useful in phonetics, psycholinguistics including atypical language development, speech technology, and speech therapy.

- **HAS bilateral academic mobility project NKM-39/2016:** “Contact induced change in Finno-Ugric languages”. 01.01.2016–31.12.2018. Financial support for 2016: 900,000 HUF. Aim: Co-operation with the University of Tartu. It is connected to the NKFIH project “Uralic syntax changing in an asymmetrical contact situation”, and joint Estonian-Hungarian publications are expected.

- **HAS Guest Scientist Scheme VK-004/2016:** “Cognitive Approaches to the Type-Mood-Force-Nexus: Review and Preview”. Support: 7 million HUF. Aims: (i) to review and systematize cognitive approaches to sentence types, sentence moods, and illocutionary forces, (ii) to sketch a research agenda and prepare for empirical work in the above area that pays particular attention to the study of sign languages.

- **NKA 201102/2757: NKA 201102/2757:** From the establishing charter of the Tihany Abbey to the Káldi Bible. Financial support: 500,000 HUF. Goal: the publication of a popularizing book with a wealth of knowledge, original texts and facsimiles, reflecting the results of current research.

- **HAS President’s budget 2016, Sz-005/2016:** Support for the continuation of the SIGNificant Chance project. Grant: cca. 14.5 million HUF.

- **HAS, EU-08/2016:** European Union and national research institute grants I. Aim: H2020, EACEA, Danube Program preparation, CEF Telecom grant. Grant: 8.8 million HUF.

- **HAS, EU-25/2016:** European Union and national research institute grants II. Aim: Danube Program preparation, CEF Telecom grant. Grant: 25 million HUF.

**International proposal**

COST 16105 EnetCollect (European Network for Combining Language Learning with Crowdsourcing Techniques): The consortium members, the institute among them, successfully applied for this grant. The goal of this interdisciplinary project is to explore the possibilities of a novel community-based approach to assembling study materials in foreign language teaching, and to work out sustainable methods of producing study materials and textbooks. Project period: March, 2017 – March 2021.

**V. List of important publications in 2016**


Gósy, Márta, Irene Vogel, Viola Váradi: Temporal patterns of error repairs and appropriateness repairs in Hungarian. PHONETICIAN 113: pp. 61-78. (2016) URL

Gyarmathy, Dorottya: Izé ‘stuff’: a corpus-based study on a curious Hungarian filler. PHONETICIAN 113: pp. 82-101. (2016) URL


Kas Bence (ed.): “Szavad ne feledd!”: Tanulmányok Bánréti Zoltán tiszteletére. [“Remember your word!”: Studies in honor of Zoltán Bánréti] Budapest: RIL HAS, 2016. 381 p. ISBN: 9789639074644 containing: Babarczy Anna, Balázs Andrea: A kognitív kontroll és a preverbalis fókusz értelmezése [Cognitive control and the interpretation of preverbal focus]; Borbély Anna: Kétgyakorlat a kétgyakorlat értelmében [Bilinguals on bilingualism]; Gósy Mária, Beke András, Gyarmathy Dorottya, Neuberger Tilda, Auszmann Anita: Magánhangzók (a, e, u, i) mintázata elsődleges progresszív afáziában [Patterns of vowels (a, e, u, i) in primary
Neuberger, Tilda: Perception of consonant length opposition in Hungarian stop consonants.
PHONETICIAN 113: pp. 6-23. (2016) URL
Rebrus, Péter, Péter Szigetvári: Diminutives: Exceptions to Harmonic Uniformity.
Rebrus, Péter, Miklós Törkenczy: Types and degrees of vowel neutrality.
LINGUISTICA 56:(1) pp. 239-252. (2016) DOI: 10.4312/linguistica.56.1.239-252
Simon, Zsolt: Wer war Allumari, König von Malatya?
ANATOLICA 42: pp. 67-76. (2016) DOI: 10.2143/ANA.42.0.3178201
Surányi, Balázs: Discourse configurationality.
In: Caroline Féry,, Shinichiro Ishihara (eds.) Oxford Handbook of Information Structure.