

19. rakenduslingvistika kevadkonverents

**KEELE MÕJU:
ANDMETEST TÕENDATUD
TEADMUSENI**

16.-17. juuni 2022, Tallinn

TEESID / ABSTRACTS

EAAL 19th Annual Conference

**INFLUENCE OF THE LANGUAGE:
FROM DATA TO CONTENT-RICH
KNOWLEDGE**

June 16-17, 2022, Tallinn, Estonia



EESTI
RAKENDUSLINGVISTIKA
ÜHING
ESTONIAN ASSOCIATION OF
APPLIED LINGUISTICS



Sisukord/Contents

PLENAARETTEKANDED / KEYNOTE TALKS	4
Digital Dictionary Database for Slovenian: unstructured, semi-structured and structured data in modern lexicography	4
Simon Krek (Jožef Stefan Institute, Slovenia)	4
The Making and Breaking of Classification Models in Linguistics: A Multimethod Perspective on Alternations	4
Jane Klavan (University of Tartu, Estonia)	4
Arzt, der – männliche Person, die ... Ärztin, die – weibliche Person, die ... Women and men in current German-language meaning dictionaries	5
Kathrin Kunkel-Razum (Dudenredaktion, Germany)	5
Assessing morphological complexity: theoretical, methodological and educational issues	5
Gabriele Pallotti (University of Modena and Reggio Emilia, Italy).....	5
ETTEKANDED/TALKS	6
Specific idiomatic phrases in Russian in Estonia	6
Inna Adamson, Mari Uusküla (Tallinn University, Estonia)	6
Estonian static spatial relations by Chinese learners: from support to containment and beyond	6
Agu Bleive (University of Tartu)	6
Vanade eestikeelsete tekstide normaliseerimiskatsed	7
Gerth Jaanimäe (Tartu Ülikool)	7
Estonian National Corpus 2021 - the largest collection of Estonian language data	7
Kristina Koppel, Jelena Kallas (Eesti Keele Instituut)	7
Monitor corpus of Slovene and automatic text categorization	8
Iztok Kosem (Jožef Stefan Institute, Slovenia)	8
Drillharjutused häädustreeningus – olemus, funktsioonid, tähtsus	8
Einar Kraut (vabakutseline kõne- ja häädusõpetaja)	8
Mida kriisis võib ja saab? Koroonakriis ja modaalverbid meediatekstis	9
Marge Käsper (Tartu Ülikool)	9
(Koroona)kriisi kaardistamine diskursustes	10
Liina Maurer (Tartu Ülikool)	10
Vigaste ja korrektsete lausete paralleelkorpuse loomine edasi-tagasi masintõlke abil ..	11
Martin Mõtus, Kais Allkivi-Metsoja (Tallinna Ülikool)	11
Mis teadust teed? Eesti teaduskeel eestikeelse doktorandi töös	12
Peep Nemvalts, Eve-Liis Roosmaa, Triin Roosalu, Helena Lemendik (Tallinna Ülikool) ..	12
Adjektiiv või mitte? Prototüüpse adjektiivi korpusprofiil	13
Geda Paulsen, Maria Tuulik, Ene Vainik, Ahti Lohk (Eesti Keele Instituut).....	13
Empiirilised andmed semantikas: kui juhuslik peaks olema polüseemia uurimisel kasutatav korpusvalim?	14
Aimi Pikksaar (Tartu Ülikool).....	14
Verbirektsioon, keelekorraldus ja tegelik kasutus	14
Lydia Risberg, Liina Lindström (Eesti Keele Instituut, Tartu Ülikool).....	14
The Importance of Accessibility in Digital Communication Higher Education Curricula	15
Anastassia Zabrodskaja (Tallinn University), Jon Switters (Funka Nu AB)	15
Nõudlik verbivorm koroonapiirangusildil	16
Ilona Tragel, Külli Habicht (Tartu Ülikool)	16
Usefulness of various online tools for determining the CEFR level of English reading texts	17
Ülle Türk (Tartu Ülikool)	17
Siberian Ingrian Finnish Talking Dictionary	17
Ivan Ubaleht (Omsk State Technical University)	17
Prototyping Instructions Where Prosody and Pragmatics are Combined in Order to Improve EFL Advanced Learners' Language Production	19

Margarita Vul (Tallinn University)	19
WORKSHOP ON MULTILINGUALISM AND MULTILINGUAL ACQUISITION	20
„Magus mesilik minevik“ – Eestist lisraeli emigreerunud juutide keeltest	20
Kristel Algvere (Tartu Ülikool)	20
Family language policy research meets code-switching research: a case study on Russian-Estonian code-switching	20
Reili Argus, Anna Verschik (Tallinn University)	20
Estonian-English child code-mixing in the case of fairly even input situation	21
Piret Baird (Tallinn University).....	21
Eesti noorte keeleline oskus reflekteerida poliitilist raamistamist	22
Yvonne Bindrim (University of Greifswald, Germany)	22
Ethnolects of Russian in performative use: issues of accents, attitudes, and identities of young non-native speakers	22
Kapitolina Fedorova (Tallinn University), Vlada Baranova (Institute for Linguistic Studies, Russian Academy of Sciences)	22
Quantitative dialectology of necessity structures in Finnish Romani	23
Kimmo Granqvist (University of Helsinki)	23
The Minority of Estonian Tatars: Ways and Challenges of Maintaining Tatar	24
Maria Iqbal (University of Tartu).....	24
Comprehension and Production of Nouns and Verbs in Lithuanian TD children and children with DLD	25
Laura Kamandulytė-Merfeldienė (Vytautas Magnus University, Lithuania).....	25
The commodification of identity among the 3rd generation Russian speakers of Estonia	26
Eva Katona (Eötvös Loránd Tudományegyetem, Hungary)	26
Vene emakeele üliõpilaste keelelisest ja kultuurilisest kohanemisest	26
Birute Klaas-Lang, Kadri Koreinik, Kerttu Rozentalde (Tartu Ülikool)	26
Pilguheit eesti juutuberite sõnavarasse: for real, mis on nende vibe?	27
Mari-Liis Korkus, Annika Pant, Aive Mandel, Kristiina Praakli, Elisabeth Kaukonen, Liina Lindström (Tartu Ülikool).....	27
Methodological issues of the Non-word repetition test: recorded stimuli vs. live stimuli	28
Egle Krivickaite-Leisiene (Vytautas Magnus University, Lithuania)	28
The Role of Vodograi Sunday School of the Ukrainian Language and Culture in Transmission of the Ukrainian Language and Cultural Identity in Ida-Viru County	29
Elvira Küün (University of Tartu)	29
„Vahepeal ma mõtlen, et miks on vaja asendada mingid sõnad inglise keelega“ – inglise keele kasutusest sisuloomes Eesti juutuuberite vaatest	29
Kristiina Praakli, Liina Lindström, Aive Mandel, Mari-Liis Korkus, Annika Pant, Elisabeth Kaukonen (Tartu Ülikool)	29
The grammar(s) of code-mixing: A corpus-based approach to multilingual first language acquisition	30
Antje Endesfelder Quick (University of Leipzig), Stefan Hartmann (University of Düsseldorf), Nikolas Koch (Ludwig Maximilian University of Munich).....	30
Linguistic Landscape of a Modern Post-Soviet Capital- Case of Chisinau, Republic of Moldova	31
Ana Vasiliiev (Eötvös Loránd Tudományegyetem, Hungary)	31
Kvantitatiivne sissevaade inglise keele kasutusse Eesti teismeliste ja täiskasvanute suulistes vestlustes	32
Virve-Anneli Vihman, Kristiina Praakli, Maarja-Liisa Pilvik, Mari-Liis Korkus (Tartu Ülikool)	32
Negotiation of identity and Lithuanian language in the 2nd and 3rd generations of migrants from Lithuania to Kazakhstan	32
Loreta Vilkiene (Vilnius University, Lithuania).....	32



PLENAARETTEKANDED / KEYNOTE TALKS

Digital Dictionary Database for Slovenian: unstructured, semi-structured and structured data in modern lexicography

Simon Krek (Jožef Stefan Institute, Slovenia)

Lexicography in 21st century is faced with many challenges, from the opportunities arising from the increasingly versed Artificial Intelligence, to the democratisation of knowledge through open collaboration and citizen science projects. However, it seems that the very basic *raison d'être* of lexicography persists - to offer as much knowledge about the vocabulary of a particular language, and its interaction with other languages, as possible. The real challenge in the beginning of the 21st century is how to organise this knowledge which comes in different forms, as unstructured data (text, corpora), semi-structured data (XML, JSON etc.) or structured data (relational databases). Traditionally, from the beginnings of their use in 1980s, lexicography considered corpora as detached from lexicographic description, as a source consulted at a given point in time, not intrinsically connected with lexicographic resources, with such a connection ultimately enabling dynamic (semi-)automatic monitoring of language use in lexicographic fashion (identification of lexical units, sense division, explanation, exemplification etc).

Now, with the disappearance of book-driven organisation of lexicographic data, there is no need for compartmentalisation of data types. The new challenge is how to monitor, what to monitor, how to store lexicographically organised information, and how to visualise it. The organisation of Digital Dictionary Database (DDDS) for Slovenian tries to provide an answer for some of these challenges. In the talk, we will present DDDS data model, our tools for monitoring language use in real time, and visualisation of various types of data: monolingual data, bi- or multilingual data, corpus data, multi-modal data etc.

The Making and Breaking of Classification Models in Linguistics: A Multimethod Perspective on Alternations

Jane Klavan (University of Tartu, Estonia)

In my talk I focus on combining various linguistic methods to find out what we can infer about linguistic variation from the patterns and structures we see in the language data. The case study I present comes from the Estonian language and pertains to the alternation between the exterior locative cases and the corresponding adpositions. I will first fit a machine classifier (mixed-effects logistic regression) to the corpus data in order to see which variables contribute significantly to the model fit.

The second part of the talk looks at the results of linguistic experiments (a forced choice task and an acceptability rating task) carried out on the same phenomenon. The discussion of the talk highlights some of the pros and cons of combining corpus-linguistic and experimental methods for the study of alternations.



Arzt, der – männliche Person, die ... Ärztin, die – weibliche Person, die ... Women and men in current German-language meaning dictionaries

Kathrin Kunkel-Razum (Dudenredaktion, Germany)

The topic of gender in language has been highly controversial in Germany, especially in 2021.

In addition to whether to mark gender in German, the discussion was about how to do it. Among other things, the editorial team of the German dictionary Duden was accused of abolishing the generic use of the masculine form after redefining the masculine personal designations on its online dictionary Duden online.

In this talk, I analyze some current German dictionaries (Duden online, DWDS, Austrian Dictionary) with focus on the way definitions are phrased and the way women, men and people who do not see themselves as man or women are represented. This involves keywords (masculine/feminine/neutral form?) in the respective dictionary and the definition of their meaning.

Moreover, particular emphasis is placed on the linguistic examples in order to obtain an answer to the question of whether social reality is adequately reflected in dictionaries.

Assessing morphological complexity: theoretical, methodological and educational issues

Gabriele Pallotti (University of Modena and Reggio Emilia, Italy)

In recent years several studies have discussed the construct of morphological complexity, proposing theoretical and operational definitions and measuring its development over time or its variation across tasks.

The talk will first review these definitions and measures, critically assessing their merits and shortcomings, with an emphasis on the comparative dimension and cross-linguistic applicability. Some studies will be presented that have used the Morphological Complexity Index (MCI, Pallotti 2015; Pallotti & Brezina 2015) to assess both native and non-native speakers' texts.

This will be followed by a discussion on how the assessment of morphological complexity should not be performed in and for itself, but always with an eye on the functional consequences of more or less complex discourse. This has several implications for language teaching and assessment, which should not just aim for 'more complexity' but rather consider its optimal, appropriate levels in real-life communication.



ETTEKANDED/TALKS

Specific idiomatic phrases in Russian in Estonia

Inna Adamson, Mari Uusküla (Tallinn University, Estonia)

Previous research has shown that in modern Estonia there are phrases that have local specifics (Zabrodskaja 2009; Verschik 2010; Külmoja 2013; Kostandi 2013, 2018 and others). These include, firstly, nominations-phrases of the social sphere, and, secondly, phraseologisms that came from quotes and can be used interspersed, reflecting the realities of Estonia. This is usually a translation borrowing.

The first type includes such phrases as *semjnyj vrach* (a general practitioner reflecting its Estonian equivalent “*perearst*” ‘family doctor’), *opornije lico* (a support person), *dorozhnoje strahovanije* (a motor liability insurance from Estonian “*liikluskindlustus*” ‘vehicle insurance’), etc., to the second found in publications of various times: *Gruziny I osetiny Estonii: davajte zhit' družno!* (“Georgians and Ossetians of Estonia: let's live (together) friendly!”); *Nu, musornyj volk, pogodi!* (“Well, garbage wolf, wait a minute!”) and others. Such phrases should be studied due to the discrepancy in the discourse of different countries. These language units have previously been studied (e.g. Adamson 2009, 2010, 2013, 2018, 2020) but further research is needed to understand the functioning of these phrases. This kind of language material entails the difficulty of translating and explanation of meaning.

Estonian static spatial relations by Chinese learners: from support to containment and beyond

Agu Bleive (University of Tartu)

Both Chinese and Estonian distinguish between containment and support relations and accordingly have means to express those distinctions. But there are some substantial differences in how situations are seen by the speakers of those languages as containment or support relations. For teaching purposes it is important to know how Chinese native speakers acquire the Estonian static spatial relations.

This research is the first attempt to find out how those differences influence how Chinese speakers acquire Estonian static spatial relations. For this purpose a picture description task was carried out by 22 Chinese native speakers. Every participant described 40 pictures with different kinds of static spatial relations from the topological relations' picture series by Bowerman and Pederson (1992).

Despite the fact that Chinese learners of Estonian have the means to express support relations in their native language, it turns out that it is difficult to fully acquire them in Estonian. The fact that in Chinese there is only one kind of universal support relation and the dominance of the postposition 上 *shang* ‘on, on top of’ to express this relation (Zhang 2013, Bleive 2022) and the fact that in Estonian the support relations can be divided into three groups (simple support, support by attachment and support by attachment to the tip), seems to influence how Chinese learners of Estonian describe support relations in Estonian.

The main finding of the picture description task was that the Chinese learners of Estonian didn't use support through attachment relation (postposition *küljes*) and support by the tip relation (postposition *otsas*) almost at all. Instead, they used simple support relation (adessive case or postposition *peal*) as in their native language, which was in stark contrast with the native Estonian usage. Unexpectedly, expressing containment relations didn't pose such a problem to Chinese learners of Estonian.



Another important finding of this research was that Chinese learners of Estonian had a higher probability of using native-like logic in Estonian in choosing between support and containment relations while describing the pictures when they used case suffixes and lower probability when they used adpositions. Considering all that, there are a few factors that should be taken into account when teaching the usage of Estonian locative adpositions and suffixes to native Chinese speakers, which will be discussed in detail in the presentation.

Vanade eestikeelsete tekstide normaliseerimiskatsed

Gerth Jaanimäe (Tartu Ülikool)

Ajaloolised tekstid pakuvad huvi nii keeleteadlastele, ajaloolastele, perekonnaloo uurijaile kui ka teistele, kes oma töös digiarhiive kasutavad. Selleks, et nendest märksõnu otsida või neid tänapäeva eesti keele jaoks väljatöötatud analüüsivahendite abil uurida, on vaja need tänapäevasele kirjakujule teisendada, ehk teisisõnu normaliseerida. See aga hõlmab endas mitmesuguseid probleeme. Nendeks on enamasti varieeruvused kirjaandis, kuna õigekirjareeglid ei olnud sageli niisuguste tekstide kirjutamise ajaks välja kujunenud või ei olnud need eriti ranged. Peale selle esineb kirjutistes ka murdelisi jooni. Sageli on probleemiks masinõppealgoritmide õpetamiseks kasutatavate andmete vähesus.

Lisaks tekitab eesti keel oma muutevormide rohkusega veelgi mitmeid probleeme. Peamine nendest on, et erinevad sõnavormid tekitavad andmetesse automaatselt rohkem varieeruvust. Erisusi tekitab ka põhja- ja lõunaeesti keele paralleelne kasutus, mis tähendab, et masinõppealgoritmide treenimiseks on tekstid tarvis üksteisest eristada. Kolmandaks probleemiks on vormihomonüümia, ehk üks ja sama kirjakuju võib olla vastavuses nii mõnele vanale kui ka tänapäevasele sõnale. Näiteks viitab sõna pesnud lõunaeesti keeles peksmisele.

Selles ettekandes tutvustatavas töös on uuritavaks ja normaliseerimiskatsete läbiviimiseks kasutatud andmestikku Rahvusarhiivis rahvahanke käigus sisestatud vallakohtuprotokollid¹, mis annavad hea ülevaate 19. sajandi elust, olust ja keelekasutusest. Normaliseerimismeetodina kasutati tähemärgipõhist statistilist masintõlget.

Meetodi ideeks on käsitleda vana ja uut kirjaviisi eraldi keeltena, tähemärgipõhine masintõlge aga tähendab, et lausete ja fraasimustrite asemel tõlgitakse sõnu ja täheühendeid. Sõna kaupa tõlkimisel võib aga tekkida eelnevalt kirjeldatud vormihomonüümia probleem. Selle lahendamiseks katsetati tõlkimiseks mitte sõnu, vaid sõnapaare.

Proovimaks lahendada andmete vähesuse probleemi kasutati n-ö hõbestandardit. Kuna tänapäevasele kujule vanale kujule tekstide teisendamine on lihtsam kui vastupidine protsess, tekitati väikest hulka käsitsi konstrueeritud reegleid kasutades kunstlikult andmeid juurde. Ettekandes tutvustan lühidalt vallakohtuprotokollide normaliseerimise problemaatikat, rakendatud normaliseerimismeetodit ja katsetatud lisandusi ülalkirjeldatud probleemide lahendamiseks. Lisaks tutvustan ka esialgseid tulemusi.

Estonian National Corpus 2021 - the largest collection of Estonian language data

Kristina Koppel, Jelena Kallas (Eesti Keele Instituut)

The paper presents the new version of the Estonian National Corpus: the Estonian National

¹ <https://www.ra.ee/vallakohtud/>



Corpus 2021 (ENC 2021), which was made available at the beginning of 2022. The corpus was developed in collaboration between the Institute of the Estonian Language and the software company Lexical Computing Ltd. ENC 2021 is the fourth in the series and the largest, containing 2.9 billion tokens.

In our presentation, we outline the process of crawling the web, the process of cleaning and post-processing the crawled data, and the methodology for classifying web texts into genres and topics. We will also describe subcorpora of ENC 2021. ENC 2021 is the first in the series to be tagged both syntactically and morphologically. Multi-level tagging allows for more complex queries and the analysis of the morphosyntactic properties of words and constructions. In the second part of the presentation, we will introduce new opportunities for corpus analysis in Sketch Engine, a corpus query system developed by Lexical Computing Ltd. These show genres and topics in Word Sketches, Diachronic Analysis and Trends, and also allow Term Extraction. In addition, using Word Sketches has become more user-friendly, with grammatical relations now having names in Estonian.

Monitor corpus of Slovene and automatic text categorization

Iztok Kosem (Jožef Stefan Institute, Slovenia)

The infrastructure for the Slovenian language has been steadily developed over the years, especially as far as corpora are concerned. One of the sorely missing resources has been a monitor corpus, and a recent Slovenian project titled SLED (Monitor Corpus for Slovene and related language resources), which is funded by the Ministry of Culture of the Republic of Slovenia, will address this gap. This will enable lexicographers and researchers to include up-to-date information in their works or analyses.

The monitor corpus of Slovene called Trendi will be launched in late May 2022, and will complement the Gigafida 2.0 reference corpus of written Slovene, which means it will cover the period from where the Gigafida coverage stops (currently from 2019 onwards). A pipeline for text extraction has been developed and includes various annotation processes as well as certain statistic calculations (e.g. lemma list). At the moment, 110 different sources provide material for the corpus, with the size of daily text collection being around 6 million words (approx. 150–170 million words per month). In the presentation, we will present the corpus contents, as well as offer a demo of the corpus in the KonText and NoSketchEngine tools on CLARIN.SI.

Another important aspect of the SLED project is the development of a tool for automatic text categorization. The tool will be trained on representative manually selected data from a selection of 110 sources. We have prepared a list of 11 categories, consulting resources such as IPTC (<https://iptc.org/>) and corpora, including the corpus of Estonian. The tool will probably not be ready by the time of the conference, thus we will focus on presenting the category selection process, and the mapping to source pages, with special emphasis on more influential news sites. The plans are to integrate the tool in the Trendi pipeline, as well as to use it in categorizing texts in other corpora of Slovene, in particular the Gigafida reference corpus.

Drillharjutused häädustreeningus – olemus, funktsioonid, tähtsus

Einar Kraut (vabakutseline kõne- ja häädusõpetaja)

Häädusõppe kontrastiiv-dünaamiline meetod käsitab häädussüsteemi motooriliselt programmeeritud liigutuskäskude hierarhilise struktuurina. Igas keeles on selle struktuuri ülesehitus, rääkimata põhielementidest (invariantseid häädusžestid), radikaalselt ainulaadne – häälstatud välja tooma just selle keele jaoks olulisi fonoloogilisi opositsioone.



Aktsenti, seda traditsioonilise keeleõppe tavalist kaasnähtu põhjustab õppijale emakeelest omaseks saanud koordinaatsiooni- ja -harjumuste süsteem. Isegi juhul, kui piisava täpsusega on ära õpitud uue hääldustehnika nõuded kõigi oluliste üksikvõtete tasemel, ei pruugi neist vormuda ühtset mootorilis-koordinatoorset tervikut – siduskõnes jätkub visalt emakeelse kõnega kinnistunud mootorikaprogrammi mõju.

Et hääldusõpe tõesti kannaks vilja, peab see kõigepealt keskenduma sihtkeele jaoks oluliste baasliigutuste ja elundihoiakute selgitamisele ning praktilisele selgeksõpetamisele. Siin kasutab dünaamiline treening iga liigutusvõtte jaoks vastavaid võimlemisharjutusi, mille mootoriline loogika järgib sihtkeele häälduslaadi nõudeid ega olene õppijate emakeelest. Kontrastiivsed rõhuasetused eri harjutuste tähtsustamisel (võimalikult ka nendega töötamise järjekorra osas) määrab taustkeele häälduslaadi oma dünaamilise hierarhiaga – ükski emakeelest tuttav liigutusvõtte ei tohi pääseda uut, kujundatavat liigutustervikut häirima (vt Kraut 2000.). Seega on baasliigutused ja nende korralik omandamine hääldusõppe aluslüli – selle edukuse kindlustab ühelt poolt võimalikult selgena eksponeeritud kontrast taustkeele liigutusvõtetega ning teiselt poolt hästi juhendatud, hoolikas taktil-auditivne enesekontroll.

Drilli ülesanne on arendada eelnevalt äraõpitud dünaamikavõtted mootorilise automatismi tasemeni – seisundini, mis vabastaks aktiivse tähelepanu juba uutele ülesannetele (järgmise, keerukama raskusastme hääldusvõtte või siis tegelik kõnesuhtlus). Arvestades mootorikasüsteemi hierahilist ülesehitust ja sellest lähtuvat treeningulooikat, on tähtis, et vähemalt osaline automatism saaks välja töötatud õpiprotsessi igas ettevõetavas järgus (sellest oleneb järgmiste õpilülide tulemuslikkus). On tähtis, et juba seegi, nii-öelda esmaastme drill töötaks võimalikult üldhaaravana, süsteemi parajasti käsilolevat lõiku täielikult katvana – õppija peab saama füüsilise kogemuse antud eritlustaseme fonotaktiliste võimaluste ja nõuete süsteemsest dünaamilisest ühtsusest! Sellise treeninguga läbitakse samm-sammult kõik sihtkeele häälduses olulised teemavaldkonnad (eesti keele puhul rõhu teravuse/lõdvõttuse kontrast, alalõua-keele-huulte koostöö, keelekäikude ulatuvus, keele vabahoid konsonante eristades, lõokrõhu ja palataliseerimise tehnika, rütmi- ja meloodilise rõhu eristus). Kõige nõudlikum ja vastutusrikkam on drill viimasel, kokkuvõtval treeninguastmel, kus eelnevalt omandatud üksikvõttelised automatismid tuleb sulandada juba ühtseks optimaalseks, vastastikku toetavaks tervikuks – orgaanilise, funktsionaalselt toimiva siduskõne tasemel.

Ettekandes tutvustan õpiprotsessi eri järkudes sobivate drilliharjutuste alusnõudeid ja tähtsamaid koostamispehõimõtteid. Kuna õppijale tähendab drill järjekindlat, igapäevast tööd endaga, tuleb juttu ka motivatsiooniraskustest ja psühholoogilistest võtetest eduka iseseisva tööritmi kujundamisel.

Kirjandus

- Kraut, E. 2000. Eesti keele hääldamine. Käsiraamat harjutuste ja helinäidetega. TEA, Tallinn.

Mida kriisis võib ja saab? Koroonakriis ja modaalverbid meediatekstis

Marge Käsper (Tartu Ülikool)

Ettekanne analüüsib modaalverbide kasutust prantsuskeelses meediakorpuses, mis hõlmab kõiki koroonakriisi esimesel, 2020. aastal, Prantsusmaa suurimates ajalehtedes Le Monde ja Le Figaro ilmunud artikleid, kus kasutati märksõna „Covid-19“. Tekstikorpust on kogutud projekti PRG934 „Kriisi tavalisuse kujutamine: diskursus, kirjandus ja pilt“ raames, uurimaks kriisi mõistega seonduvaid tähendusvälju pandeemiaatingimustes.

Kroonviirusega võitlemiseks kehtestatud karmidest piirangutest johtuvalt on koroonakriisi



kajastamisel üheks läbivaks küsimuseks kujunenud pandeemiat ohjavate meetmete enam või vähem range kohustuslikkus. Selles kontekstis võtab ettekanne vaatluse alla modaalverbide keelilise kategooria, mis keeles ühe olulisema vahendina võimalikkuse ja vajalikkuse tähendussuhteid väljendavad (Auwera ja Plungian 1998). Prantsuse keeles kahest peamisest modaalverbist väljendab vajalikkust peamiselt *devoir* ('pidama') ning võimalikkust *pouvoir* ('saama, võima, suutma'), aga nii nende verbide kasutuskontekstid kui tõlkevõimalused on mõistagi laiemad (Käsper 2016 jt).

Kriisis selgelt esiplaanil oleva kohustuslikkust väljendava deontilise modaalsuse kõrval keskendub ettekanne vastukaaluks just võimalikkuse-modaalsuse(t)ele, mida väljendab modaalverb *pouvoir*. Selle verbi tõlkevasteid on autor varasemalt eesti keeles uurinud ilukirjanduslike tõlgete paralleelcorpuse põhjal, konkreetsete oludega seotud suur meediakorpus ja lisandunud leksikomeetriselised vahendid (SketchEngine, Hyperbase) võimaldavad testida ja täpsemalt uurida selles uurimuses väljatoodud mõjureid modaalsusetüüpide määramisel (eitus, määruste semantika, jms). Modaalverbide tähenduspotentsiaali kriisimeetmete kirjeldamisel täheldasime juba ühele kriisimeetmele – kaugtöele – pühendatud uurimuses (ilmumas), täpsemalt modaalverbidele keskenduv uurimus võimaldab kriisimodaalsusi laiemalt kirjeldada.

Kirjandus

- Käsper, Marge (2016). 'Võima', 'saama' ja 'suutma' prantsuse modaalverbi 'pouvoir' vastetena. Võimalikkuse määratlemine ja rõhutamine eituses. Eesti ja soome-ugri keeleteaduse ajakiri = Journal of Estonian and Finno-Ugric Linguistics, 7 (2), 75–92. DOI: 10.12697/jeful.2016.7.2.04.
- Van der Auwera, Johan and Vladimir Plungian (1998) "Modality's semantic map". Linguistic Typology 2, 79–124.

(Koroona)kriisi kaardistamine diskursustes

Liina Maurer (Tartu Ülikool)

Prantsuse ajalehtedes *Le Monde* ja *Le Figaro* mainiti sõna 'kriis' (*crise*) 2020. aastal rohkem kui 36 000 korda – me elame kriiside ajastul. Et täpsemalt uurida märksõna kriis kasutust pandeemiaaegses prantsuse meedias, kogusime kokku kõik artiklid aastast 2020, mis sisaldasid otsisõna Covid-19. Materjal näitab, et kriis ei tähendanud ainult koroonapandeemiat, vaid kõike, mida koroonaviirus mõjutas. Uurimuse eesmärk on seega leida meediadiskursusest selgitust, milline on kriis, milles me elame: kuidas seda tajutakse, kuidas sellest kui probleemist kõneldakse ja kuidas sellega emotsionaalselt harjutakse. Toetudes Mayaffre'i (2008), Veniard'i (2013) jt diskursiivsele semantikale, lähtub uurimus mõttest, et sõna kasutus diskursuses defineerib selle sotsiaalse tähenduse.

Nagu on rõhutanud juba kvalitatiivse diskursuseanalüüsi puhul Moirand (1992), annab mõtestatult interpreteeritava tulemise materjali mitmeplaaneline omavaheline seostamine (vt ka Marling & Käsper 2021). Kvantitatiivsete meetodite abil oleme saanud ajalise ülevaate koroonakriisi arengust ja haripunktidest: millal ja kui palju mainiti sõna 'kriis'. Tulemustest lähtuvalt vaatame andmetesse sisse ja teeme erinevaid "lõikeid", kasutades leksikomeetrisi meetodeid analüüsima, milliseid valdkondi kriis riivas ja mida tegelikult kriis erinevatel ajahetkedel tähendas.

Kirjandus

- Marling, Raili; Käsper, Marge (2021). Communicating Covid-19: Framing Science and



Affect in U.S., French and Estonian Traditional Media. *ESSACHESS – Journal for Communication Studies*, 14 (2), 15–32. DOI: 10.21409/z3xp-m289.

- Mayaffre, Damon (2008). De l'occurrence à l'isotopie. Les co-occurrences en lexicométrie. – *Syntaxe et Sémantique*, 9, 53–72. DOI: 10.3917/ss.009.0053.
- Moirand, Sophie (1992). Des choix méthodologiques pour une linguistique de discours comparative, *Langages*, 105, 28–41. DOI: 10.3406/lgge.1992.1622.
- Veniard, Marie (2013). La nomination des événements dans la presse. Essai de sémantique discursive. Besançon: Presses universitaires de Franche-Comté.

Vigaste ja korrektsete lausete paralleelkorpuse loomine edasi-tagasi masintõlke abil

Martin Mõtus, Kais Allkivi-Metsoja (Tallinna Ülikool)

Üks viis masinõppepõhise grammatikakontrollija treenimiseks on õpetada arvutit tõlkima vigaseid lauseid korrektseteks lauseteks. Selleks kasutatakse eelkõige tehisnärvivõrkudele tuginevat ühekeelset masintõlget, mis eeldab ülimahukaid paralleellausete korpusi. Keelevigu ja nende parandusi sisaldavate korpuste vähesuse tõttu lisatakse niisuguste keeleandmete loomiseks korrektsetesse tekstidesse sünteetilisi vigu.

Paralleellausete genereerimiseks on ühe meetodina rakendatud edasi-tagasi masintõlget (ingl round-trip translation), mis on andnud häid tulemusi inglise keele grammatikavigade parandamisel (Lichtarge *et al.*, 2019). Edasi-tagasi masintõlke käigus tõlgitakse lähtekeelne lause esmalt sihtkeelde ja seejärel tagasi lähtekeelde.

Eestikeelne vigaste ja korrektsete lausete paralleelkorpus on genereeritud 2019. aasta eesti keele ühendkorpuse alusel. Edasi-tagasi-masintõlkeks on kasutatud TartuNLP Neurotõlget ning sihtkeeltena inglise, soome, saksa, läti, leedu ja vene keelt. Tulenevalt keelte ja tõlkemudelite erinevustest varieeruvad ka tõlke käigus moodustatud paralleellaused. Edasi-tagasi tõlke tulemusel esineb muuhulgas sõna- ja fraasiasendusi, vormiasendusi, kokku- ja lahkukirjutuse ning sõnajärje muudatusi, ka sõnade kustutamist ja lisamist. Tuleb arvestada, et kõik lausetes tehtud muudatused ei ole tingimata vead.

Kuna tõlke kaudu paralleellauseid genereerides ei teki täheortograafia vigu, siis on masintõlke väljundisse lisatud nelja tüüpi sünteesvead: tähtede kustutused, lisamised, asendused ja nihutused. Nende sagedus põhineb Eesti vahekeele ehk õppijakeele korpuse A2–C1-taseme tekstidel, milles on märgendatud eri liiki keelevead, järgides M2-formaati (Dahlmeier & Ng, 2012). Täpsemalt on veasünteesis arvesse võetud tähevigude sisaldavate lausete ja sõnade keskmist osakaalu nelja keeleoskustaseme lõikes, erinevat tüüpi tähevigude sagedusjaotust ja ulatust ehk seda, kui mitut tähte need hõlmavad.

Ettekandes antakse ülevaade muudatustest, mis tekivad eestikeelseid lauseid eri keelte kaudu edasi-tagasi tõlkides. Samuti tutvustatakse esialgseid veaparanduse tulemusi, mis on saadud eesti keele grammatikakontrollijat uudse keeleressursi alusel treenides. Korrektori tulemuslikkust hinnatakse võrdluses käsitsi lisatud ekspertparandustega, lähtudes automaatparanduste täpsusest ja saagisest.

Kirjandus

- Dahlmeier, Daniel; Ng, Hwee Tou 2012. Better Evaluation for Grammatical Error Correction.
- – Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 568–572.



- Lichtarge, Jared; Alberti, Chris; Kumar, Shankar; Shazeer, Noam; Parmar, Niki; Tong, Simon 2019. Corpora Generation for Grammatical Error Correction. – Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 3291–3301.

Mis teadust teed? Eesti teaduskeel eestikeelse doktorandi töös

Peep Nemvalts, Eve-Liis Roosmaa, Triin Roosalu, Helena Lemendik (Tallinna Ülikool)

Alates Eesti kõrghariduse õigusraami muutmisest 1.9.2019 on enamikus õppekavades vähemalt osa õpet inglise keeles. See võimaldab välisõppejõudude loenguid ja töötab paremat hariduse kvaliteeti, kuid vähenev eestikeelne teadussuhtlus toob lisamotivatsiooni ingliskeelseks teadustööks ning lisaraskusi teadmiste sisulisel omandamisel. Ülikoolide ja teaduse rahvusvahelistumise ja konkurentsi kõrval pälvib enam tähelepanu ka eesti teaduskeel.

Mida see suundumus näitab doktorandide² hoiakute ja kogemuste kohta eesti teaduskeele ja inglise keeles avaldamise suhtes?

Selgitamaks välja praeguste doktorandide arvamused teaduskommunikatsiooni keelte kohta korraldasime 2022. aasta algul üle-eestilise küsitlusuuringu (N = 101), mis on kavandatud 2012. aasta uuringu (N = 240, vt Roosmaa jt, 2014) kordusena ja lubab jälgida muutusi. 2012. aasta uuringus ilmnes, et eesti keeles teaduse tegemine valmistab algajatele teadlastele raskusi, kuna doktorandid, eriti loodus- ja reaalteadustes, ei leia sobivaid termineid ning kardavad, et töö ei jõua lugejani. Enam kui kolmandik tunnistas eestikeelsete teadustekstide loomisel probleeme lauseehitusega. Mõõndi, et teadustöö kirjutamine on emakeeles lihtsam, loogilisem ja nüansirikkam. Tulemused näitasid, et doktorand, kes osaleb eestikeelses teadussuhtluses mitmel moel (loeb, kirjutab, osaleb konverentsidel), hindab oma keeleoskust mõnevõrra paremaks – aga pigem on need inimteaduste doktorandid, kelle uurimistöö on sagedamini seotud Eesti eripära avamisega, nii on eestikeelne teadussuhtlus valdkonnas laiem. Nagu võõrkeeles, vajavad doktorandid süsteemset tuge akadeemilise eneseväljenduse arendamisel ka eesti keeles. Võrreldes 2012. aastaga on aastaks 2022 kahanenud doktoriõppes õppijate arv (vastavalt 3044 ja 2353), kuid kasvanud on võõrkeelse õppe osakaal: 96% vs. 20% õppis täielikult eestikeelses õppekavas. Muutus ka doktorite osakaal õppekeele järgi: 2012/13 lõpetas õpingud edukalt 233 doktorandi, neist eestikeelse õppekava 25% ning eesti- ja ingliskeelse õppekava 73%; 2020/21 olid need näitajad vastavalt 222, 1% ja 96%.

Kui 2012. aasta uuringust ilmnemise doktorandide kõhklused eesti keeles publitseerida, siis milliseid uusi eesti teaduskeele toetusviise on ülikoolid kasutusele võtnud? Kas doktorandide kogemused eestikeelse teaduskommunikatsiooniga on muutunud rikkalikumaks ja enesekindlamaks? Käsitleme doktorandide kujuneva eestikeelse kogukonnana ingliskeelses teadusruumis, lähtudes Berry (1997) kultuuridevaheliste strateegiatega käsitusest. Institutsioonide kihilisuse mudeli (Scott 1995) toel analüüsime kõrgharidus- ja teadussüsteemi rahvusvahelistuva institutsioonina, mis innustab või pärsib teaduskeele valikut.

Kirjandus

- Berry, J. W. 1997. Immigration, acculturation and adaptation – Applied Psychology: An International Review, 46, 5–68.
- Roosmaa, Eve-Liis; Roosalu, Triin; Nemvalts, Peep 2014. Doktorantide teadustöö keele

² Eelistame süsteemseid ladina gerundiivist pärinevaid termineid doktorand, magistrand, 'doktori-, magistratöö tegija' tähisena, nagu eksaminand (ld *examinandus*) on 'eksamitegija' ja diplomand 'diplomitöö tegija v kaitsja'.



valikutest. – Ülikool ja keelevahetus. Tartu Ülikooli ajaloo küsimusi XLII, lk 37–52.

- Scott, W. Richard. 1995. *Institutions and organizations*. Sage Publications.

Adjektiiv või mitte? Prototüüpse adjektiivi korpusprofiil

Geda Paulsen, Maria Tuulik, Ene Vainik, Ahti Lohk (Eesti Keele Instituut)

Sõnaliigiline mitmesus on üks eesti keele sõnavara kategoriseerimise keerdküsimusi muuhulgas ka leksikograafia vaatenurgast. Eesti leksikograafide uuring (Paulsen jt 2019) osutas, et üks sõnaliigi määramisel enim probleeme valmistavaid sõnaliike on adjektiiv ja sellega piirnevad klassid. Mitmese sõnaliigitõlgendusega sõnade ja vormide analüüs (Vainik, Paulsen ja Lohk 2020) näitab, et sõnaliigiliselt paindlik adjektiiv võib seostuda lisaks ka verbidega (armunud, austatud; siduv, lõöv; rääkimata, värvimata), nimisõnadega (vaimulik, pull, räbal), pronoomenitega (samane, teistsugune) ja adverbidega (kiivas, laokil, lömmis). Leksikograafi käsutuses on üha mahukamad tekstikorpused, mille põhjal saab teha järeldusi (potentsiaalsete) märksõnade kasutuse kohta. Andmestiku kiiremaks ja objektiivsemaks hindamiseks läheb tarvis selle automaatset eeltötlust (vt nt Langemets jt 2021: 760). Tekib küsimus, kuidas rakendada korpusmaterjali sõnaliigipiiride ehk sõnade morfoloogilise, süntaktilise ja semantilise käitumise tuvastamiseks? Kuidas aidata leksikograafil piiritleda näiteks adjektiivi?

Ettekandes tutvustame ühe eeltötlusvõimalusena sõnaliigi esilduvate morfosüntaktiliste struktuurimallide rakendamist korpusanalüüsil. Lähtume adjektiivile omistatavatest morfosüntaktilistest tunnustest (atributiivne funktsioon, ühildumine, võrrete moodustamine jne) ning rakendame neid prototüüpse adjektiivi profiili loomiseks. Oleme välja töötanud nn adjektiiviparameetrid, mida testisime 12 erineva leksikaalse rühma (N=10) peal, võttes aluseks eesti keele ühendkorpuse 2019 andmestiku (Tuulik jt 2022). Uuringu tulemuste põhjal saab järeldada, et parameetrite abil on võimalik eristada adjektiive teistest sõnaliikidest.

Käesolevas uuringus otsime vastust küsimusele, kust läheb piir, kus prototüüpne adjektiiv hakkab omandama mingi teiseks sõnaliigi funktsioone. Prototüüpse adjektiivi lävendväärtuste väljaselgitamiseks testime parima eristusvõimega parameetrite põhjal eelmise uurimusega võrreldes suuremat ja esinduslikumat valimit (N=100) prototüüpsetest adjektiividest. Neid tulemusi võrdleme sõnaliigilise staatuse mõttes kriitilisima rühmaga – verbi partitsiivvormidega. Määrame kindlaks testrühmade profiilid ning suhestame eri rühmad üksteisega eukleidilise kauguse ning koosinussarnasuse meetodikaga.

Kirjandus

- Langemets, Margit, Koppel, Kristina, Kallas, Jelena, Tavast, Arvi 2021. Sõnastikukogust keeleportaaliiks. – *Keel ja Kirjandus* 8–9, 754–769.
- Paulsen, Geda; Vainik, Ene; Tuulik, Maria; Lohk, Ahti 2019: *The Lexicographer's Voice: Word Classes in the Digital Era*. Kosem, I., Zingano Kuhn, T., Correia, M., Ferreria, J. P., Jansen, M., Pereira, I., Kallas, J., Jakubíček, M., Krek, S. & Tiberius, C. (toim.). *Proceedings of the eLex 2019 conference*, 1–3 October 2019, Sintra, Portugal, lk 434–452. Brno: Lexical Computing CZ, s.r.o.
- Tuulik, Maria; Vainik, Ene; Lohk, Ahti; Paulsen, Geda 2022. Kuidas ära tunda adjektiivi? Korpuskäitumise mustrite analüüs. – *Eesti Rakenduslingvistika Ühingu Aastaraamat*.
- Vainik, Ene; Paulsen, Geda; Lohk, Ahti. 2020. A typology of lexical ambiforms in Estonian. Z. Gavriilidou, M. Mitsiaki & A. Fliatouras (toim.) *Proceedings of XIX*



Empiirilised andmed semantikas: kui juhuslik peaks olema polüseemia uurimisel kasutatav korpusvalim?

Aimi Pikksaar (Tartu Ülikool)

Distributiivse semantika nime all tuntud lähenemine sõnatähendustele toetub seisukohale, et sõnade tähendusi on võimalik määratleda nende kasutuskontekstide kaudu loomulikus keelekasutuses (Harris 1954). Sellest ideest inspireerituna on keele automaattöötuse vahendite loojad juba paarkümmend aastat proovinud leida tekstikorpustest automaatselt sõnade tähendusi (vt nt Pantel ja Lin 2002). Mõned aastad tagasi pakkusid ka Leuveni ülikooli juures töötavad kognitiivsed keeleteadlased välja oma meetodi polüseemse sõna tähenduste tuvastamiseks kasutuskontekstide põhjal (Heylen jt 2015). Teemat on kognitiivlingvistilises suunas edasi arendanud nt Hilpert ja Saavedra (2020) ning hiljuti Leuveni ülikoolis doktoritöö kaitsnud Mariana Montes (2021).

Toetudes mainitud uurimustele ja enda paariaastasele kogemusele distributiivsete mudelite loomisel ja treenimisel, otsin ettekandes vastuseid küsimustele, kas ühekordne juhuvalim korpusest on piisavalt representatiivne andmekogum polüseemia uurimiseks ning mida hakata analüüsiprotsessis peale juhuvalimisse sattunud „müraga“ ehk mitterelevantsete andmetega. Samuti arutlen kahe printsiipiaalselt erineva võimaluse üle, kuidas ja mille jaoks kasutada korpusandmeid polüseemia uurimisel: 1) uuritava sõna tähendusliigenduse väljaselgitamiseks, 2) uuritava sõna kohta loodud introspektsioonipõhis(t)ele tähendusliigendus(t)ele empiirilise kinnituse leidmiseks. Need kaks lähenemist toetuvad korpusvalimi moodustamisel (sh eriti just juhuslikkust puudutavas osas) täiesti erinevatele põhimõtetele, mida ma ettekandes tutvustan polüseemsete omadussõnade analüüsimise näitel.

Kirjandus

- Harris, Z. 1954. Distributional structure. – *Word*, 10 (2–3), 146–162.
- Heylen, K.; Wielfaert, Th.; Speelman, D.; Geeraerts, D. 2015. Monitoring polysemy: Word space models as a tool for large-scale lexical semantic analysis. – *Lingua*, 157, 153–172.
- Hilpert, M.; Saavedra, D. C. 2020. Using token-based semantic vector spaces for corpus-linguistic analyses: From practical applications to tests of theoretical claims. – *Corpus Linguistics and Linguistic Theory*, 16(2), 393–424.
- Montes, M. 2021. Cloudspotting: Visual analytics for distributional semantics. PhD thesis. Leuven.
- Pantel, P.; Lin, D. 2002. Discovering word senses from text. – *KDD'02: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, 613–619. Edmonton, Alberta, Canada: ACM Press.

Verbirektsioon, keelekorraldus ja tegelik kasutus

Lydia Risberg, Liina Lindström (Eesti Keele Instituut, Tartu Ülikool)

Rektsioon ehk sõltumine on lauses esinev alistav seos, milles põhisõna tähendus tingib laiendi



vormi (Rätsep 1978: 64, EKG II: 8). Verbi rektsoonilisi laiendeid võib pidada verbi argumentstruktuuri osaks: verbi tähendusest sõltub osaliste olemasolu ja markeerimine lauses. Taolisi verbi juurde kuuluvaid obliikvakäändes argumente on eesti grammatikates nimetatud sõltuvusmäärusteks ehk rektsooniadverbiaalideks (vt Veismann, Erelt, Metslang 2017).

Mitmed verbid võimaldavad tähenduslikult sarnaseid argumente, mis on markeeritud erinevalt, nt „teadma [mida?]” või „[millest?]”. Osa neist on keelekorralduslikust aspektist osutunud probleemiks, sest verb võimaldab enam kui üht rektsoonimalli, ent nende tähenduserinevused pole selged. Preskriptiivsed lähenemised on eelistanud vaid üht variantidest, nt „sarnanema [millega?]”, mitte „[millele?]”. Eesti rektsoonikäsitlustes ongi teemat kirjeldatud pigem preskriptiivselt (vt Mäearu 2011; Pool 2019 [1996]; Erelt 2019 [2006]: 35–41; EKK 2020: 443–448). Nenditakse, et „mõnel juhul on õige rektsooni valimisega raskusi” (Erelt 2019: 35), aga ei täpsustata, kellel ja miks raskused esinevad.

Nii mõnedki tüüpvead, tänapäevaks „rasked kohad” on aga ette kirjutatud juba sajand tagasi, kui ühist eesti kirjakeelt alles loodi (nt Muuk, Tedre 1930; Aavik 1936). Verbi rektsooni fikseerides eelistati üht teisele, ilma selge teadmista selle kohta, mis kõnelejate jaoks on tavalisem või loomulikum. Nüüd, sajand hiljem, on mahukate keelekorpuste najal võimalik vaadelda, kuidas tollased normingud on keeles vastu võetud ning millised neist on tänapäeva keeles kasutusel. Allkivi (2012) on leidnud, et sagedasti kasutatav ja loomulikuna tajutav rektsoon võiks ka ametlikus kirjakeeles mööndud olla.

Ettekandes vaatleme eesti keele ühendkorpuse 2021 põhjal mõningaid eesti keele verbirektsoone, mis on keelekasutajatele probleeme valmistanud. Vaatleme lähemalt verbide „sarnanema” ja „sarnane/identne/paralleelne/analoogne olema” ning baseeruma/põhinema/rajanema/toetuma” juurde kuuluvaid laiendeid, nende vormistust ja vormistuse varieerumist. Selgitame välja, mil määral nende verbide juurde kuuluvad laiendid jälgivad varem esitatud soovitusi. Kui esineb olulisi erinevusi soovitude ning tegeliku keelekasutuse vahel, selgitame põhjusi, millest erisused on tingitud, ning hindame soovitude ajakohasust tänapäeval.

The Importance of Accessibility in Digital Communication Higher Education Curricula

Anastassia Zabrodskaia (Tallinn University), Jon Switters (Funka Nu AB)

According to the World Bank, 15% of the world's population experience some form of disability. This, coupled with a shift towards the provision and consumption of information in digital formats, has increased the need for accessible content publishing. The onset of the COVID-19 pandemic has made this need is even greater than before. It is essential that information is accessible and reaches all citizens, including People with Disabilities, older adults and immigrants, leaving no one behind.

Accessible content publishing means the generation of information that can be used by people with the widest user needs, characteristics and capabilities. As of the 23rd of September 2020, the EU's Web Accessibility Directive obliges public sector websites and apps to comply with accessibility requirements. Despite the implementation of this Directive in national legislation across the EU, most digital content is still not accessible to all users. In the near future, this will increase the demand for communication professionals who possess accessibility skills. However, there is not yet a corresponding supply of skilled professionals to meet this demand as accessibility is usually not included within the communication programme curricula in higher education.

The ADORE project aims to respond to this need by educating university teaching and training staff in communication disciplines on how to publish content in an accessible way



and how to provide a more inclusive teaching experience. The project will do this by:

- carrying out a gap analysis on accessibility training needs analysis for university teaching and training staff in fields related to communication and main issues faced by end user groups when accessing content online.
- mapping university curricula in fields related to communication where accessibility training can be implemented.
- co-creating a needs-based accessibility training toolkit for university teaching and training staff.
- developing a set of personas representing professional profiles in the field of accessible communication.
- providing recommendations on the inclusion of accessibility knowledge and training materials in European university curricula.

The aim of this presentation is to introduce the ADORE project and highlight the unique, multidimensional and interdisciplinary work we carry out scrutinizing the accessibility in digital communication higher education curricula. The project “Accessibility in Digital Communication Higher Education Curricula” (ADORE) is supported by Erasmus+ and the European Solidarity Corps Agency under the Erasmus+ programme. Partners: Tallinn University (Leader), Funka, INUK, Paris Lodron University of Salzburg, University of Maribor (Partners).

This project has been funded with support from the European Commission Erasmus+ programme. This publication [communication] reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Nõudlik verbivorm koroonapiirangusildil

Ilona Tragel, Külli Habicht (Tartu Ülikool)

Piirangusilt on viirusdiskursuse asünkroonne suhtlussündmus (vt nt Jones (toim) 2021; Tragel, Habicht 2021), mis eeldab lugejalt teatud käitumist. Selles sündmuses, nagu igasuguses suhtluses, on autor (sildi koostaja), sõnum (sildi tekst) ja vastuvõtja (lugeja). Silditeksti koostades teeb autor kaht tüüpi valikuid: (1) mida silditekstis eksplitsiitselt väljendada ja mida mitte ja (2) kuidas silditekstis piiranguid ja muud teavet väljendada.

Ettekandes esitatakse esialgseid uurimistulemusi silditekstide isikukasutusest enesekehtestamise eesmärgil ning sellest, kuidas diskursusekogukonna liikmed ennast isikukategooriat puudutavate valikute abil diskursuse ja üksteisega seostavad (vt nt Fairclough 2003).

Materjal on 900 2020.–2021. aasta koroonapiirangute ukseildi teksti, millest on automaatanalüüsiga otsitud isiku ja kõneviisi markereid ning nendele lisaks spetsiifilisi võimukasutusele viitavaid lemmasid, nn võimusõnu (vt ka Tragel, Tomson 2022).

Analüüsi esialgsed tulemused näitavad, et isikuliste asesõnade ja tegusõnade kasutamisel tehtud grammatilise isiku valikud (nt meie keskuses hoiad vahet) peegeldavad sildi autori võimupositsiooni sildi lugeja suhtes. Selles peegeldub diskursuserolli seos isikukategooriaga. Spetsiifiliste keelevahenditena pakuvad huvi indikatiivi vormid, mida kasutatakse piirangusildidel direktiivses funktsioonis (nt hoiad vahet).



Kirjandus

- Fairclough, Norman 2003. *Analysing Discourse. Textual Analysis for Social Research*. London, New York: Routledge.
- Jones, R. (ed.) 2021. *Viral Discourse. (Elements in Applied Linguistics)*. Cambridge: Cambridge University Press. doi:10.1017/9781108986465.
- Trigel, Ilona, Külli Habicht 2021. Erakorralise aja sildikeel. – *Oma Keel*, 1, 56–65.
- Trigel, Ilona, Kairit Tomson 2022. Tanel ja Jüri kiusavad: piirangute põhjuste väljendusvahendid koroonapiirangute siltidel. *Õiguskeel* 2.

Usefulness of various online tools for determining the CEFR level of English reading texts

Ülle Türk (Tartu Ülikool)

When choosing reading texts for language teaching or testing purposes, the question often asked is how to determine whether a text is of suitable difficulty for a particular group of learners or test takers. The Common European Framework of Reference for Languages: Learning, Teaching, Assessment includes several reading comprehension scales (for overall reading comprehension, for reading correspondence, reading for orientation, reading for information and argument, reading instructions), but the information that they provide about text difficulty can be considered vague at best. At lower levels (A1 and A2), phrases like 'very short and simple texts' and 'short and simple texts' are used; at higher levels (C1 and C2), 'lengthy, complex texts' and 'long and complex texts' are mentioned. However, what makes a text simple or complex is not clearly explained. In the United States, on the other hand, there is a long tradition of measuring readability of reading texts. However, the formula used have been developed for those for whom English is their first language, not for English language learners and the scores are not related to the CEFR levels.

In order to help both language teachers and test developers to determine the CEFR level of reading texts, several online tools have been developed and made freely available. The question is how valid, reliable and user friendly such tools are. Thus the presentation will compare three such tools (CEFR Checker, Text Inspector and Text Analyzer), the statistical information they provide about the texts analysed and will compare the CEFR levels as determined by the tools with the qualitative analyses of experts in order to answer the question whether there is a free tool that can be considered valid and reliable enough to be used for estimating the CEFR levels of the reading texts in the Estonian Year 9 and Year 12 English examinations.

Siberian Ingrian Finnish Talking Dictionary

Ivan Ubaleht (Omsk State Technical University)

The Siberian Ingrian Finnish Language is a Lower Luga Ingrian Finnish – Lower Luga Ingrian (Izhorian) mixed language. This language is used by the descendants of the settlers from the Lower Luga area (the Rosona river area, to be exact). This language has been existing in Siberia for over 200 years. D.V. Sidorkevich researched and documented Siberian Ingrian Finnish in 2008–2014 [1,2]. N.V. Kuznetsova [3] and M.Z. Muslimov continue to research this language to the present.

We started to create the open Siberian Ingrian Finnish speech corpus in 2019. This corpus



based on our own audio data from our expeditions and telephone interviews with speakers [4 subsection 3.1]. Approximately 5 hours of audio data and part of annotations are available on GitHub and licensed under a Creative Commons Attribution 4.0 license (CC BY 4.0) [5]. A talking dictionary is the first step to address the creation of the speech corpus for the Siberian Ingrian Finnish language. Talking dictionaries are used to revitalizing and linguistic research of endangered languages [6-8]. At this moment, we have achieved the following results:

- a software tool for parsing annotation files (we use ELAN media annotation tool [9] to annotate speech data) and feeding a database was created [10];
- the structure of the database has been developed and this database has been filled [11,12];
- the web application of the talking dictionary had been created.

The source code of the web application is open and available in GitHub [13]. At this moment, the first version of the talking dictionary is available via Internet: <http://lexeme.net/sif>. The first version of Siberian Ingrian Finnish talking dictionary contains about 300 words and 200 phrases. We plan to use the talking dictionary and the speech corpus for the language revitalization of Siberian Ingrian Finnish.

References

- Sidorkevich, D.V. (2014). *Yazyk ingermanlandskih pereselentsev v Sibiri (struktura, dialektnye osobennosti, kontaktnye yavleniya)*, [Язык ингерманландских переселенцев в Сибири (структура, диалектные особенности, контактные явления)], Diss. ILIRAN.
- Sidorkevich, D.V. (2011). On domains of adessive-allative in Siberian Ingrian Finnish, In *Proceedings of Institute for Linguistic Studies Vol. 7*, pp. 575–607.
- Kuznetsova, N. (2016). Evolution of the non-initial vocalic length contrast across the Finnic varieties of Ingria and adjacent areas, *Linguistica Uralica*, Vol. 52, pp. 1–25.
- Ubaleht, I. (2021, March). Lexeme: The Concept of System and the Creation of Speech Corpora for Two Endangered Languages. In *Proceedings of the Workshop on Computational Methods for Endangered Languages Vol. 2*, pp. 20–23.
- The repository of Siberian Ingrian Finnish speech corpus:
<https://github.com/ubaleht/SiberianIngrianFinnish>
- David K. Harrison, and Gregory DS Anderson. (2006). Tuvan talking dictionary. Retrieved from <http://tuvan.talkingdictionary.org>
- David K. Harrison, *et al.* (2019). Zapotec language activism and Talking Dictionaries. In *Electronic Lexicography in the 21st Century: Smart Lexicography: Proceedings of eLex 2019 Conference*, pp. 31–50.
- Moira Saltzman. (2017). Jejueo talking dictionary: A collaborative online database for language revitalization. In *Proceedings of the 2nd Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pp. 122–129.
- Peter Wittenburg, Hennie Brugman, Albert Russel, Alexander Klassmann, and Han Sloetjes. (2006). ELAN: A Professional Framework for Multimodality Research. In *Proceedings of Language Resource and Evaluation 2006*, pp. 1556–1559.
- The tool for processing ELAN annotation files:
<https://github.com/ubaleht/LexemeELAN>
- The script of the create the scheme of the database for the talking dictionary:
<https://github.com/ubaleht/SiberianIngrianFinnish/tree/master/SpeechDatabase/Scheme>
- Data for the web-application of the talking dictionary:



<https://github.com/ubaleht/SiberianIngrianFinnish/tree/master/SpeechDatabase/Data>

- The source code of the web application: <https://github.com/ubaleht/Lexeme>
-

Prototyping Instructions Where Prosody and Pragmatics are Combined in Order to Improve EFL Advanced Learners' Language Production

Margarita Vul (Tallinn University)

The current research proceeds from the idea that wrong pronunciation as well as accents act as distractors during conversational events which, in some contexts, can be so brief that it becomes crucial to guarantee intelligibility. In the theoretical part it was revealed that linguistic transfer phenomenon contributes greatly to accentedness in EFL learners.

The current research focuses on both segmental (articulatory features of phonetic segments) and suprasegmental (stress, pitch, length) features in improving phonetic realisation in EFL learners in Estonia. The aim of the research was to develop a prototype of instructions to stop negative transfer in EFL learners in Estonia. This was achieved through the Design-Based Research approach which allows empirical work with practical problems of human learning.

The data for prototyping the instructions was collected through observation and assessment of EFL advanced learners of different sociolinguistic backgrounds present in Estonia to identify the areas of speech which are significantly affected by linguistic transfer as well as to describe the degree of prosodic fossilisation. 10 instructions combining prosody and pragmatics were created and evaluated using three formative evaluation methods with an expert-teacher and a phonetician. The research question "What characteristics should be emphasized in instructions for improving oral production of EFL learners in Estonia?", was fully answered by providing clear guidelines on the specifics of voice pitch, stress, duration and speech tempo in Standard English.

The research also showed that stopping transfer and rewiring common speech patterns in EFL learners in Estonia would require a great deal of learners' metacognition. It was also revealed that teaching prosody and pragmatics would require the ability of the teacher to demonstrate to the learners the differences between L1 and the target language phonation as well as introduce the aspects of speech that need to be rewired for improving their intelligibility. The study's product has a potential to solve practical problems with generating effective intervention and contribute to the body of knowledge in the field.



WORKSHOP ON MULTILINGUALISM AND MULTILINGUAL ACQUISITION

„Magus mesilik minevik“ – Eestist Iisraeli emigreerunud juutide keeltest

Kristel Algvere (Tartu Ülikool)

Ettekandes tuleb juttu juudi päritolu inimeste keelelistest valikutest ja kuidas on toimunud põlvkondlik keele edasikandumine. See põhineb empiirilistel andmetel, mille kogusin TÜ usuteaduskonna bakalaureusetöö jaoks, ent mis on seotud minu doktoriõpingutega. Peamisteks uurimisküsimusteks on: kas ja kuidas on ühest keelekeskkonnast teise minek mõjutanud informantide keelevalikuid. Analüüsin ennekõike püha ja argise, sümboolse ja instrumentaalse funktsiooni vahetamist nende keelevalikutest.

Andmed on kogutud poolstruktureeritud veebiintervjuudega, mis toimusid ajavahemikus 22.03.–12.04.2022. Informantide valimi koostamiseks kasutasin statistiliselt mittetõenäosusliku kihistatud valimi põhimõtteid (Trost 1986). Lähtusin uuritavate keelekogemust mõjutavatest teguritest: aeg, sugu, vanus ja keeleoskus (sh eesti, uusheebrea).

Õnnestus intervjuuerida viit informanti. Andmeid analüüsisin kvalitatiivse sisuanalüüsiga. Keele ja religiooni teemade uurimine on uus, sel sajandil sotsiolingvistika vaatevälja jõudnud teema (Darquennes, Vanderbussche 2011). Judaism ja keeled on läbi aegade olnud tihedalt seotud; heebrea, aramea ja kreeka keel on olnud pühade tekstide keeled (Spolsky 2003).

Eestis Iisraeli emigreerunud inimeste keelekasutust ei ole varem uuritud. Tegu on Eesti mõttes huvitava etnolingvistilise rühmaga. Ettekandes annan lühikese ülevaate Eesti juudi kogukonna kujunemisloost (vt nt Weiss-Wendt 2017). 1972. aastast kuni 2015. aastani on Eestist emigreerunud 1868 inimest, kõige enam nõukogude korra kokkuvarisemise järel: ajavahemikus 1990–1992. Minu informandid on Iisraeli läinud enne seda. Leidsin, et informandid on eesti keele alles hoidnud ja oluliseks peetakse ka seoseid Eestiga. Vähem huvitavad ei ole ka informantide endi arusaamad ja kirjeldused uus- ja vanaheebrea keele kohta.

Kirjandus

- Darquennes, J., Vanderbussche, W. 2011. Language and religion as a sociolinguistic field of study: some introductory notes. *Sociolinguistica*, 25, 1–11.
- Spolsky, B. 2003. Religion as a site of Language Contact. *Annual Review of Applied Linguistics*, 23. Cambridge University Press, 81–94.
- Trost, J. E. 1986. Statistically nonrepresentative stratified sampling: A sampling technique for qualitative studies. – *Qualitative Sociology*, No. 9, 54–57.
- Weiss-Wendt, A. 2017. *On the Margins. Essays on the History of Jews in Estonia*. Budapest: CEU Press.

Family language policy research meets code-switching research: a case study on Russian-Estonian code-switching

Reili Argus, Anna Verschik (Tallinn University)

The paper analyses FLP in a family where both parents are ethnic Russians but mother speaks Estonian to their two children. Based on Spolsky (2004), FLP consists of three components: language ideology, language management, and language practices. As a rule, FLP studies concentrate on parent's goals and attitudes, and semi-structured interviews are the standard data collection method. In contrast, our goal is to both interview parents and to analyse actual child-directed speech in order to find out whether there are differences between beliefs,



measures taken for the children's bilingual development, and naturalistic language use. Both parents are ethnic Russians; the father speaks Russian and the mother speaks Estonian to the children.

The data was collected in two ways: a semi-structured interview with the parents (1.5 h) and 6 h of recordings of family conversations.

The findings are as follows. There are discrepancies between the declared ideologies, management and practices. The declared policy is strict OPOL and, as the father rendered it, purism because of the concern that the children will be confused otherwise. During the interview, it was the father who switched between Estonian, Russian and English. Language management in the family is guided by practical considerations, i.e., not by the wish to balance the use of both languages: if it is easier to do things in either of languages, it is done so.

The number of turns in the mother's speech in one hour recording is 539, of which 50 % are child-directed. Of all child-directed turns 7 % are in Russian and 8 % contain either insertional code-switching or alternational switching within one turn. There are only 6 instances of code-switching in father's speech. Although the parents declared that they speak Russian to each other, the mother switched to Estonian from time to time while addressing the father. Sometimes, the mother repeats the same turn in the other language in child-directed speech. In general terms, the results tell us that in family language policy research, triangulation of statements obtained from interviews (ideology), practical steps (management) and actual linguistic behaviour (practices) is highly desirable.

Despite the declared OPOL policy, it is in fact impossible to avoid code-switching.

References

- Spolsky, Bernard. 2012. Family language policy – the critical domain. *Journal of Multilingual and Multicultural Development* 33 (1), 3–11.

Estonian-English child code-mixing in the case of fairly even input situation

Piret Baird (Tallinn University)

Studying bilingual children's code-mixing in the light of usage-based theory has become more popular in recent years (Gaskins *et al.*, 2019; Quick *et al.*, 2020). Most studies with bilingual children involve families that use OPOL or ML@H strategies. As a result of these strategies the child usually receives significantly more input in one language. The current case study focuses on a 2-year-old (2;4–2;10, 36h of data) Estonian-English bilingual whose language input is more even between her two languages as the family separates languages by time rather than by place or person. How does a balanced language input influence MLU, language proportions and code-switching?

The study will calculate child's MLU in each of her two languages and in code-mixed utterances. Previous studies have shown that the MLU is longer in code-mixed utterances (Quick *et al.*, 2018, 2020). Also monolingual and code-mixed utterance proportions will be calculated to see if they reflect the input situation as other studies have shown (Allen *et al.*, 2002; Quick *et al.*, 2018, 2020). Additionally, sibling code-mixing rate will be looked at. As a final element the utterance complexities will be studied to find differences between the two languages and code-mixed utterances. The results indicate that the proportion of code-mixed utterances is higher than in previous studies, MLU for code-mixed utterances is longer and they are more complex.



References

- Allen, S., Genesee, F., Fish, S., & Crago, M. B. (2002). Patterns of code mixing in English-Inuktitut bilinguals.
 - Gaskins, D., Frick, M., Palola, E., & Endesfelder Quick, A. (2019). Towards a usage-based model of early code-switching: Evidence from three language pairs. *Applied Linguistics Review*, 0(0). <https://doi.org/10.1515/applirev-2019-0030>.
 - Quick, A. E., Gaskins, D., Bailleul, O., Frick, M., & Palola, E. (2020). A gateway to complexity: A cross-linguistic comparison of child bilingual speech. *International Journal of Bilingualism*, 136700692095672. <https://doi.org/10.1177/1367006920956721>
 - Quick, A. E., Lieven, E., Backus, A., & Tomasello, M. (2018). Constructively combining languages: The use of code-mixing in German-English bilingual child language acquisition. *Linguistic Approaches to Bilingualism*, 8(3), 393–409. <https://doi.org/10.1075/lab.17008.qui>
-

Eesti noorte keeleline oskus reflekteerida poliitilist raamistamist

Yvonne Bindrim (University of Greifswald, Germany)

Ühiskondlikud toimijad ja organisatsioonid levitavad oma arusaamu erinevate kanalite kaudu. Selleks kasutavad nad muuhulgas intervjuusid ja sotsiaalmeediat. Toimijate kasutatav keel on professionaliseeritud, st sõnade ja väljendite valik ei ole juhuslik, vaid planeeritud ja sõnum nõnda strateegiliselt raamistatud. Raamistamine on informatsiooni seostamine teatud keelekasutusega, mis mõjutab informatsiooni tõlgendamist. Strateegilise raamistamise eesmärk on mõjutada avalikku arvamust soovitud viisil. Selleks kasutatakse metafoore või valitud mõisteid ja väljendeid.

Varasemad uuringud on näidanud, et kuigi raamistamine üldiselt toimib, ei mõju see kõigile ühtemoodi. Lisaks erinevad eksperimentides kasutatud raamid tegelikkuses ettetulevatest raamidest.

Strateegilise raamistamise mõju avalikule arvamusele ei tohi alahinnata. Projektis käsitletakse küsimust, kas Eesti noored reflekteerivad teatud ühiskondlike toimijate keelekasutuse üle, millised sotsiaaldemograafilised tegurid mõjutavad keelelise refleksiooni määra ja kas teatud teemade vahel valitseb vastastikune sõltuvus.

Retseptiivset ja produktiivset keelekäitumist uuritakse kirjaliku küsitluse abil. Osalejate vastustest saab tuletada nende keelelise refleksiooni oskust ning seda mõjutavaid sotsiaaldemograafilisi tegureid.

Ethnolects of Russian in performative use: issues of accents, attitudes, and identities of young non-native speakers

Kapitolina Fedorova (Tallinn University), Vlada Baranova (Russian Academy of Sciences)

Mass labour migration to Russian megalopolises from the territory of the former Soviet Union and from Asian countries is gradually transforming their ethnic and linguistic composition turning them into global cities. New communities of speakers are forming in main Russian cities which include ethnically mixed young non-native Russians who understand their diversity and can distinguish different ethnic styles of speaking. The so called crossing identity (Rampton 1995)



is a sociolinguistics practice observed nowadays in many different global cities around the globe (Benor 2010); in some cases, it follows the emergence of ethnolect, like it was with Kiezdeutsch in Germany (Wiese 2009; Wiese, Rehbein 2016), or other more or less stable linguistic variant.

The paper is based on the study of data from video blogs created by young non-native speakers of Russian who aim at performative use of different accents and ethnic styles of speaking. Both the speech of video bloggers and comments from their audience under the videos are analyzed to reveal different types of discourses regarding multilingualism circulating in Russian society. The study focuses on different perspectives native and non-native speakers of Russian have regarding ethnolects of Russian and its linguistic features: while for Russian speakers they are related with alienation and / or mocking of non-native speakers, ethnolects speakers themselves reclaim them via performative aesthetic use and employ in creating more subtle distinctions and identities.

References

- Benor, S.B., 2010. Ethnolinguistic repertoire: shifting the analytic focus in language and ethnicity. // *Journal of Sociolinguistics* 14 (2), 159–183.
- Rampton, B., 1995. *Crossing: Language and Ethnicity Among Adolescents*. Longman, London.
- Wiese, H., 2009. Grammatical innovation in multiethnic urban Europe: new linguistic practices among adolescents // *Lingua* 119, 782–806.
- Wiese, H., Rehbein, I., 2016. Coherence in new urban dialects: A case study // *Lingua* 172–173, 45–61.

Quantitative dialectology of necessive structures in Finnish Romani

Kimmo Granqvist (University of Helsinki)

I will discuss the morphosyntax of the necessive structures of Finnish Romani, to which little attention has been paid in the study of Romani grammar (e.g. Granqvist 2007; Granqvist & Laitinen, forthcoming). Focus will be on competing types of necessive structures, in which show the forms of the necessive verbs, the case of the subject and the presence/absence of person inflection in the necessive verb vary. Other differences concern the complementizer and the form of the verb complement.

I will examine the necessive structures of Finnish Romani from the point of view of their geographical distribution, comparing the outcomes to earlier observations in studies on Finnish Romani dialectology (Granqvist 2017). Methodological starting points are in corpus-based studies of dialect syntax (Grieve 2009; Szmrecsanyi 2013). I will apply Classification Algorithms to Romani data, such as random forest (Breiman 2001) and decision trees (Hothorn, Hornik & Zeileis 2006).

The data from the project “Finnish Romani and other northern dialects of the Romani language in the Baltic Sea region” (2013–2017) comprise samples of Finnish Romani from 25 localities in Finland. The data was initially collected for the purposes of qualitative research, but it also allows for a quantitative examination of some language phenomena.

References

- Breiman, Leo. 2001. Random Forests. *Machine Learning* 45(1), 5–32.



- Granqvist, Kimmo. 2007. Suomen romanin äänne- ja muotorakenne. Suomen Itämaisen Seuran Suomenkielisiä julkaisuja 36. Kotimaisten kielten tutkimuskeskuksen julkaisuja 145. Helsinki: Yliopistopaino.
- Granqvist, Kimmo. 2017. Finnish Romani and its dialectology. Teoksessa Kirill Kozhanov, Mikhail Oslon & Dieter W. Halwachs. (toim.). Das amen godi pala Lev Čerenkov. Romani historia, čhib taj kultura, 221–237. Graz: GLM.
- Granqvist, Kimmo & Saarni Laitinen. Tulossa. Descriptive Grammar of Finnish Romani. München: Lincom Europa.
- Grieve, Jack. 2009. A Corpus-based Regional Dialect Survey of Grammatical Variation in Written Standard American English. Ph.D. dissertation, Northern Arizona University.
- Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15(3), 651-674.
- Szmrecsanyi, Benedikt. 2013. Grammatical Variation in British English Dialects: A Study in Corpus-based Dialectometry (Studies in English Language). Cambridge: Cambridge University Press.

The Minority of Estonian Tatars: Ways and Challenges of Maintaining Tatar

Maria Iqbal (University of Tartu)

This presentation focuses on Family Language Policy (FLP) of Estonian Tatars (see also Iqbal 2021), a small and rather understudied minority living in Estonia (*et al* Rannut, Rannut, 2010; Klaas, 2015; Iqbal, 2021). According to Statistics Estonia, the total number of Tatar-identifying residents is 1,937 (2021). In particular, its scope is on parents' strategies for the maintenance of Tatar and whether language environment, either Russian or Estonian-dominant -- is shaping parents' decision when choosing a school.

The study is informed by Bernard Spolsky's (2004) FLP model which includes three independent but interconnected components: language practices, language beliefs or ideologies, and language management. Between Jan 2019 and Jan 2022 I recorded semi-structured in-depth interviews with 14 informants from 8 families (*ca* 8 hours), who were born or living in Ida-Virumaa (dominantly Russian spoken area (Ehala *et al.* 2019: 499)), Harju and Tartu counties. An inquiry displayed the Russian- or the Estonian-dominating settings shape informants' decisions. However, when choosing a medium of instruction for their children, Estonian MOI were preferred over Russian as those are believed to increase better career opportunities in Estonia.

Apart from those two languages, the Tatar language is an important vehicle that sustains connection to the Tatar culture. Informants shared that they used to attend or are currently attending Tatar summer and Tatar Sunday schools, and participate in events organised in Tatar. Some families speak only Tatar at home and some speak Tatar, Estonian and/or Russian mix. Many parents brought up their need to take additional Estonian language courses in order to better their skills.

References

- Ehala, M.; Koreinik, K.; Praakli, K.; Siiner, M. (2014). Kuidas uurida keele kestlikkust? *Keel ja Kirjandus*, 7, 1–19, 489–508.
- Iqbal, M. (2021). Estonian Tatar Families Navigating Between Four Languages. *Tehlikedeki Diller Dergisi*, 11 (19), 288–308.
- Klaas, M. (2015). The role of language in (re)creating the tatar diaspora identity: the



case of the Estonian Tatars. *Journal of Ethnology and Folkloristics* 9 (1): 3–19.

- Rannut, M.; Rannut, Ü. (2010). Russification of Non-Estonian Pupils in Tallinn. *Eesti Rakenduslingvistika Ühingu aastaraamat* 6: 243–259.
- Spolsky, B. (2004). *Language policy*. Cambridge University Press.

Comprehension and Production of Nouns and Verbs in Lithuanian TD children and children with DLD

Laura Kamandulytė-Merfeldienė (Vytautas Magnus University, Lithuania)

Previous study conducted on Estonian-Russian bilinguals' colour vocabulary (Loitšenko 2018) showed that linguistic convergence caused for certain colour terms to appear more frequently in the list and naming tasks. The aim of present study is to compare bilinguals' associations between the Estonian-dominant and Russian-dominant bilingual groups in their strong and weak languages. The comparative analysis will show whether the colour associations are language restricted or not.

Bilingual groups are determined using the BLP (Bilingual Language Profile) questionnaire (Birdsong, Gertken, & Amengual, 2012). The questionnaire helps to determine the more dominating language based on the self-reported data.

The procedures are conducted via Skype, mobile phone application (i.e. Whatsapp) or face to face in the form of an interview. The researcher explains how the experiment is conducted and starts naming the colour terms. This kind of approach allows the participants name as many associations as they want even in their non-dominating language. The questionnaire-based studies that research associations are limited to a written form, which means that bilinguals who lack a writing skill in their less dominant language cannot express their thoughts freely.

The test is carried out in the participants' preferred language first and then two weeks later in the second language. The colour terms named in Russian are: belyj, černyj, krasnyj, zelenyj, želtij, sinij, goluboj, koričnevij, fioletovij, rozovij, oranževij, seryj, birjuzovij, sirenevij, lilovij, beževij. The colour terms birjuzovij and sirenevij have been added based on the recent study on Russian basic colour terms (Paramei, Gribers, Mylonas 2017). The Estonian colour terms named are: valge, must, punane, kollane, roheline, sinine, pruun, hall, lilla, oranž, roosa, helesinine, türkiissinine, violetne, beež. All of the basic terms (Sutrop 1995) were selected with an addition of helesinine, türkiissinine, violetne and beež.

The expected outcome is 1) linguistically similar colour terms will cause the bilinguals' colour associations to converge or 2) colour associations are not language restricted.

References

- Birdsong, D., Gertken, L.M., & Amengual, M. (2012) *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. COERLL, University of Texas at Austin. Web. 20 Jan. 2012. <<https://sites.la.utexas.edu/bilingual/>>
- Loitšenko, O. (2018). In: Lindsay MacDonald, Carole P. Biggam, & Galina Paramei (Ed.).
- *Colour Studies: Vision, Cognition, and Language*. Amsterdam/New York: John Benjamins [to appear].
- Sutrop, U. (1995). Eesti keele põhivärvinimed. *Keel ja Kirjandus*, 12, 797–808.
- Paramei, G., Griber, Y. & Mylonas, D. (2017). An online colour naming experiment in Russian using Munsell colour samples. *Color Research and Application*, pp. 1–17.



The commodification of identity among the 3rd generation Russian speakers of Estonia

Eva Katona (Eötvös Loránd Tudományegyetem, Hungary)

After the restoration of Estonia's independence, a third of the country's population has not ceased consisting of Russian-speakers (residents with Russian, Ukrainian, Jewish, Belarusian and other ethnic backgrounds) who had predominantly arrived to the country during the Soviet era, and therefore never identified themselves with Estonia.

The "Russian-speaking" identity soon became popular because of its unifying power and political neutrality. Today, however, the practice of claiming to be "an Estonian" is becoming more and more widespread among the 3rd generation of the Russian-speaking minority. Among members of certain subgroups of this minority being a Russian or a Russian-speaker is a label only pertaining to language skills.

According to Monica Heller, the commodification of language and identity means that the primary marker of ethnic identity is transformed into a measurable knowledge and commodity that shapes the labor market and social expectations, giving individuals with certain language skills an advantage.

Members of the younger generation are aware that their Russian skills are an advantage in the labor market. Therefore, even if they identify themselves as "an Estonian", they do not hide their Russian language skills, because they consider and emphasize them as a linguistic capital.

In our research 11 out of the 30 interviewed young people (between the ages of 18 and 34) pointed out that the knowledge of Russian gives them advantage over others in their everyday life and in the world of work as they are trilingual (Russian, Estonian, English), unlike members of the Estonian- and English-speaking, bilingual (and sometimes monolingual) majority society. Even young people who have little contact with Russia, the Russian culture, and for whom being the descendant of a Russian family does not play a role in their identity construction, they nevertheless see the Russian language proficiency itself as a positive trait that defines them. In our conference presentation, we illustrate the value of Russian identity and Russian language skills for the members of the researched group by showing and analysing quotes taken from our on one-on-one emic interviews recorded in the cities of Tallinn, Tartu, Narva and Jõhvi during 2020.

Vene emakeelega üliõpilaste keelilisest ja kultuurilisest kohanemisest

Birute Klaas-Lang, Kadri Koreinik, Kerttu Rozentalde (Tartu Ülikool)

Tartu Ülikooli üliõpilastest umbes 10% on oma õpitee jooksul lõpetanud Eestis venekeelse haridusasutuse (põhikooli) ja sooritanud ka eesti keele teise keelena riigieksami. Riigieksami tulemused on sellel sihtgrupil väga head, kuid ometi on vene emakeelega üliõpilastel raskused eestikeelsetel õppekavadel õppimise ja ülikooli kultuuriga kohanemisel.

Need üliõpilased peavad läbi tegema mitu paralleelset kohanemist: koolist ülikooli, vene dominantkeelega suhtluskeskkonnast (Ida-Virumaa, aga ka Tallinn) mitmekeelsesesse ülikooli ja eesti dominantkeelega Tartu linna (Ehala & Koreinik 2021), eesti-vene kakskeelsest õppeasutusest eesti-inglise õppekeelega ülikooli (Soler & Rozentalde 2021, Klaas-Lang 2021). Eelnevast keelekogemusest erinev uus keele- ja kultuurikeskkond surub peale oma diskursuse ja reeglid (Busch 2017). Asumine ühest kooliastmest teise või gümnaasiumist ülikooli võib olla võrreldav geograafilise rändesündmusega, mis nõuab indiviidilt teatavat (keelelist) säilenõtkust (vt nt Dawson, Pooley 2013).

Oma ettekandes esitleme käimas olevat uuringut, mille käigus oleme kogunud vene emakeelega Tartu Ülikooli üliõpilaste keelilise ja kultuurilise kohanemise lugusid. Oleme läbi



viinud 26 umbes 60-minutilist poolstruktureeritud intervjuud, sihtgrupiks erinevate valdkondade - Medicina, Humaniora, Socialia, Realia et naturalia – teise või kolmanda aasta üliõpilased.

Uuringu üldeesmärk on mõista, kuidas venekeelse põhikooli ja kakskeelse gümnaasiumi lõpetanud üliõpilaste keelelised repertuaarid, ühiskondlikud ja kultuurinormid, mis reguleerivad keeleloomet ja retseptiooni (Gumperz 1982, Blommaert, Backus 2013, Koreinik & Klaas-Lang 2021), on seotud nende toimetulekuga/kohanemisega akadeemilises keskkonnas. Uuringu rakenduslik eesmärk on osutada võimalikele kitsakohtadele üldhariduskoolide eesti keele teise keelena ja kultuuripädevuste arendamises, mis ei pruugi anda ka riigieksami hästi ja väga hästi sooritanud üliõpilastele piisava ettevalmistuse akadeemilises keskkonnas hakkama saamiseks. Ühtlasi on Tartu Ülikooli soov uuringu tulemusi kasutada selleks, et pakkuda muu emakeelega tudengitele tuge võimalikult kiireks ja sujuvaks lõiminguks eestikeelses keskkonda ja rahvusvahelisse organisatsioonikultuuri.

Kirjandus

- Blommaert, J. & Backus, A. (2013). Superdiverse repertoires and the individual. *Current Challenges for Educational Studies*. – Ingrid de Saint-Georges, Jean Jacques Weber (eds), *Multilingualism and Multimodality*. Rotterdam: Sense, 11–32.
- Busch, B. (2017). Expanding the Notion of the Linguistic Repertoire: On the Concept of *Spracherleben* – The Lived Experience of Language. – *Applied Linguistics* 38 (3): 340–358.
- Dawson, M. & J. A. Pooley. (2013). Resilience: the role of optimism, perceived parental autonomy support and perceived social support in first year university students. *Journal of Education and Training Studies* 1 (2) 38–4.
- Ehala, M. & K. Koreinik. (2021). Patterns of individual multilingualism in Estonia. *Journal of Baltic Studies* 52 (1), 85–102.
- Gumperz, J. (1982). *Discourse strategies*. Studies in interactional sociolinguistics. Cambridge: Cambridge University Press.
- Klaas-Lang, B. (2021). Üks ülikool, mitu keelt. *Akadeemia*, 9, 1571–1598.
- Koreinik, K. & Klaas-Lang, B. (2021). Linguistic repertoires and teacher resilience: Russian-dominant bilingual teachers in Estonian-medium schools. – Anna Verschik (Ed.). *Acta Universitatis Tallinnensis. Humaniora* (212–239). Tallinn: TLÜ Press.
- Soler, J. & Rozenthal, K. (2021). The Englishization of higher education in Estonia and Latvia: Actors, positionings, and linguistic tensions. In R. Wilkinson & R. Gabriëls (eds.) *The Englishization of Higher Education in Europe*. Amsterdam University Press, pp. 57–75.

Pilguheit eesti juutuberite sõnavarasse: for real, mis on nende vibe?

Mari-Liis Korkus, Annika Pant, Aive Mandel, Kristiina Praakli, Elisabeth Kaukonen, Liina Lindström (Tartu Ülikool)

Pilguheit eesti juutuberite sõnavarasse: for real, mis on nende vibe? Keeleteemadest on just noortekeel avalikkuse pidevas huviorbiidis: noorte kaasabil tutvustatakse uute sõnade tähendusi (vt Klaas 2022), viidatakse eri vanusrühmade keelekasutuserinevustele (nt Mikk 2011) ja peegeldatakse nende arusaamu keelekasutuse kohta (vt Areen 2022). Eesti keeleteaduses kogub noortekeele uurimine alles hoogu. Üleilmastumise ja digikultuuri kontekstis pälvib üha rohkem tähelepanu inglise keele roll noorte keelelistes valikutes. Selle taustal kerkib küsimus, millist mõju avaldab eesti-inglise kontaktsituatsioon noorte sõnavarale



laiemalt.

Ettekandes heidame pilgu eesti juutuuberite sõnavarale, fookusega inglise päritolu ainesel nende keelepruugis. Andmed kogusime kaheksa eestikeelse YouTube'is tegutseva sisulooja videotest (kokku ligi 80 videot erinevatest žanritest). Videotes tuvastasime kõik ingliskeelsed sõnad, väljendid ja lausungid ning koondasime need nn juutuuberite sõnastikku. Kogutud ainesitikus analüüsime sisuloojate leksikaalseid valikuid, tuginedes laenatavuse hierarhia mudelile (vt lähemalt Field 2002, Verschik 2007). Tutvustame juutuuberite andmetel põhinevat sõnaliikide laenatavuse skaalat, mille kaudu selgitame integreeritavate ingliskeelsete elementide atraktiivsust. Ühtlasi pakume välja, et eesti juutuuberite keelekasutust võib nimetada eraldi keelevariandiks, sisuloojate registriks.

Kirjandus

- Aree 2022. Kitsad teksad põlevad, zoomer'id tulevad: 14 noort vastavad, mis neile meeldib ja mis mitte. – Eesti Ekspress, 19.01.2022. <https://ekspress.delfi.ee/artikkel/95660937/kitsad-teksad-polevad-em-zoomer-em-id-tulevad-14-noort-vastavad-mis-neile-meeldib-ja-mis-mitte> (05.05.2022).
- Field, Fredric 2002. Linguistic Borrowing in Bilingual Context. Amsterdam, Philadelphia: John Benjamins.
- Klaats, Erika 2022. Kas tead, mida tähendab „seeni peale jätma“, „toptier luud“ või „deemmidesse slaidima“? – Maaleht, 19.01.2022. <https://maaleht.delfi.ee/artikkel/95671241/kas-tead-mida-tahendab-seeni-peale-jatma-toptier-luud-voi-deemmidesse-slaidima> (05.05.2022).
- Mikk, Mihkel-Markus 2011. Spiik estonlisch? – Öhtuleht, 12.03.2011. <https://www.ohtuleht.ee/418016/spiik-estonlisch> (09.05.2022).
- Verschik, Anna 2007. Keelekontaktid, laenatavus ja verbi kopeerimine eestivene keelevariandis. – Keel ja Kirjandus 5, 357–377.

Methodological issues of the Non-word repetition test: recorded stimuli vs. live stimuli

Egle Krivickaite-Leisiene (Vytautas Magnus University, Lithuania)

This study aims to discuss methodological issues of the non-word repetition test and its results. Experiments in which auditory and visual information has been presented have shown that vision strongly influences speech perception (Massaro, Bosseler 2006). Visual information obtained from a talker's face can improve speech intelligibility over that obtained with the presentation of only auditory information (Massaro *et al.* 1995). The use of recorded stimuli ensures uniformity of input, eliminating variations in rate, pitch, volume, and other phonetic and auditory features; it also restrains the tester's influence on children's performance. Recorded stimuli eliminate any visual cues. Comparison of live vs. recorded stimuli is needed for evaluating the possible effect of variations on the accuracy of response as well as the comparability and validity of the two methods (Chiat, Roy 2007). The study presents two methods of non-word repetition test: 1. live stimuli; 2 recorded stimuli. The sample of the study consists of 100 typically developing children (2 age groups: 4;00–4;11 and 5;00–5;11).

The research showed that observing the visual articulation helped the children to identify the non-words: statistically, the live stimuli were repeated significantly better than the recorded stimuli ($p=0.005$). The word structure (word length and syllable structure) was found to be an important indicator: both live and recorded two syllable non-words were repeated similarly,



whereas the live stimuli of three- and four-syllable non-words statistically were repeated significantly better than the recorded stimuli; the live stimuli non-words with consonant clusters were repeated statistically significantly better than the recorded stimuli.

The results indicate that the methodology of the non-word repetition test has an effect on the outcome of both children and adult tests.

References

- Chiat Sh., Roy P. 2007: The Preschool Repetition Test: An Evaluation of Performance in Typically Developing and Clinically Referred Children. *Journal of Speech, Language, and Hearing Research* 50, 429–443.
- Massaro D. W., Cohen M. M., Smeele P. M. T. 1995: Cross-linguistic comparisons in the integration of visual and auditory speech. *Memory & Cognition* 23 (1), 113–131.
- Massaro D. W., Bosseler A. 2006: Read my lips. The importance of the face in a computer-animated tutor for vocabulary learning by children with autism. *Autism* 10 (5), 495–510.

The Role of Vodograi Sunday School of the Ukrainian Language and Culture in Transmission of the Ukrainian Language and Cultural Identity in Ida-Viru County

Elvira Küün (University of Tartu)

The aim of the study was to ascertain in detail, by way of four case studies, the trends of family language policy (FLP) applied in the families of the children with Ukrainian roots studying in the Vodograi Sunday School of the Ukrainian Language and Culture in Sillamäe, Ida-Viru County, Estonia. Examination of FLP in relevant families was based on Bernard Spolsky's (2007) model of family language policy, according to which FLP consists in language practices utilized by families, language management tools internal and external to the family, and language ideologies. Another objective was to investigate the role of the Vodograi Sunday School of the Ukrainian Language and Culture in Sillamäe in relevant FLP developments. The research method used was a semi-structured interview and interviewees included the mothers of the families studied. It became clear that the Ukrainian Sunday School is also seen as a supporting pillar of Ukrainian cultural identity.

Reference

- Spolsky, B. 2007. Towards a Theory of Language Policy – Working Papers in Educational Linguistics (WPEL): Special Issue on Language Policy, 22(1);
https://repository.upenn.edu/wpel/vol22/iss1/1/?utm_source=repository.upenn.edu%2Fwpel%2Fvol22%2Fiss1%2F1&utm_medium=PDF&utm_campaign=PDFCoverPages

„Vahepeal ma mõtlen, et miks on vaja asendada mingid sõnad inglise keelega“ – inglise keele kasutusest sisuloomes Eesti juutuuberite vaatest

Kristiina Praakli, Liina Lindström, Aive Mandel, Mari-Liis Korkus, Annika Pant, Elisabeth Kaukonen (Tartu Ülikool)

Osalusmeedia platvormidel on noorte meediatarbimises üha suurem tähtsus ja tähendus (vt nt Siibak 2020). Üks Eesti noorte seas atraktiivsemaid keskkondi on YouTube, mis võimaldab



igapäevaelu olla nii sisu looja kui ka teiste loodust osa saaja. Meediatarbimisega seotud rutiinide ja tavade kõrval kujundavad osalusmeedia keskkonnad otseselt ka noorte keelekasutust.

Keskseim roll selles lasub noorte seas enim jälgitud sisuloojatel kui mõjuisikutel; nende keelelistest valikutest aga inglise keelel. Seejuures pole juutuuberite keelevalikutel tähtis kaal mitte ainult noorte keelepruugi, vaid ka keeltega seotud hoiakute ja väärtuste mõjutamisel ja kujundamisel (eesti õpilaste keelehoiakutest vt Tammemägi, Ehala 2012).

Ettekandes võtame vaatluse alla Eesti sisuloojate arvamused ja seisukohad inglise keele kasutamise kohta – vaatleme, millistes seostes ja milliste rõhuasetustega inglise keelest kõneldakse. Keskendume kolmele põhiküsimusele: kuivõrd teadlikud on sisuloojad oma keelelistest praktikatest; millisena nähakse inglise keele rolli ja positsiooni eestikeelses sisuloomes; milline tähendus on sisuloojate igapäevapraktikates eesti keelel? Ettekandes kasutatavad tekstilised andmed pärinevad kaheksa sisuloojaga tehtud poolstruktureeritud individuaal- ja paarisintervjuudest (maht ca 6 tundi; salvestatud kevadtalvel 2022). Andmete lähivaatluseks kasutame kvalitatiivset sisuanalüüsi.

Analüüsile tuginevalt võib öelda, et sisuloojate inglise keele kasutuses nähtub teatav pragmatism: inglise keele kasutuse tingib suuresti inglise keele positsioon sisuloojate üleilmse n-ö töökeelena, olles selles tegevuses loomulik suhtluskeel. See aga ei pisenda kuidagi eesti keele integratiivset tähtsust sisuloojate enesemääratluses ja keelelistes praktikates.

Kirjandus

- Siibak, Andra 2020. Digipõlvkonnast sotsiaalmeedia põlvkonnaks: põlvkondlikku enesemääratlust kujundavad trendid Eesti noorte internetikasutuses. – *Methis Studia humaniora Estonica* 26, 17–34. 10.7592/methis.v21i26.16908.
- Tammemägi, Anni; Ehala, Martin 2012. Eesti õpilaste keelehoiakud 2011. aastal. – *Keel ja Kirjandus* 4, 241–260.

The grammar(s) of code-mixing: A corpus-based approach to multilingual first language acquisition

Antje Endesfelder Quick (University of Leipzig), Stefan Hartmann (University of Düsseldorf), Nikolas Koch (Ludwig Maximilian University of Munich)

Code-switching or code-mixing, i.e. the use of several languages in one speech event, is a characteristic feature of multilingual language use, with far-reaching pragmatic and sociolinguistic implications (see e.g. Gardner-Chloros 2009). Code-mixing also occurs prominently in multilingual first language acquisition, where it has traditionally been treated from a more formal perspective (see e.g. Cantone 2009). From a usage-based perspective, according to which fixed chunks like *What's this?* and frame-and-slot patterns like *[What's X?]* can be considered the basic building blocks of language (Arnon & Christiansen, 2017; Tomasello 2003), we argue that early code-mixing can be accounted for in terms of entrenched and productive patterns at various levels of abstraction. To test this hypothesis, we use the traceback method that has been applied successfully to monolingual acquisition in previous research (see Koch 2019, Hartmann *et al.* 2021).

Drawing on three corpora of German-English bilingual children aged 2 to 4 years, we show that code-mixed utterances like *I see kelle 'I see a laddle'* can be traced back to recurrent patterns both in the child's own previous language use and in the child's input (Quick *et al.* 2018, 2021). This suggests that both in monolingual and in multilingual language acquisition, children draw heavily on multiword chunks that are strongly dependent on the patterns a child encounters in his or her input.



References

- Ambridge, Ben & Elena Lieven. 2011. *Child Language Acquisition: Contrasting Theoretical Approaches*. Cambridge: Cambridge University Press.
- Arnon, Inbal & Morten H. Christiansen. 2017. More Than Words: The Role of Multiword Sequences in Language Learning and Use. *Topics in Cognitive Science* 9(3). 542–551.
- Cantone, Katja Francesca. 2007. Code-switching in bilingual children. (*Studies in Theoretical Psycholinguistics* 37). Dordrecht: Springer.
- Gardner-Chloros, Penelope. 2009. *Code-switching*. Cambridge: Cambridge University Press.
- Hartmann, Stefan, Nikolas Koch & Antje Endesfelder Quick. 2021. The traceback method in child language acquisition research: identifying patterns in early speech. *Language and Cognition* 13(2). 227–253.
- Koch, Nikolas. 2019. Schemata im Erstspracherwerb: eine Traceback-Studie für das Deutsche. (*Linguistik, Impulse & Tendenzen* 80). Berlin, Boston: De Gruyter.
- Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge and London: Harvard University Press.
- Quick, Antje Endesfelder, Ad Backus & Elena Lieven. 2018. Partially schematic constructions as engines of development: Evidence from German-English bilingual acquisition. In Eline Zenner, Ad Backus & Esme Winter-Froemel (eds.), *Cognitive Contact Linguistics*, 279– 304. Berlin, Boston: De Gruyter.
- Quick, Antje Endesfelder, Ad Backus & Elena Lieven. 2021. Entrenchment effects in code- mixing: individual differences in German-English bilingual children. *Cognitive Linguistics* 0(0). doi:10.1515/cog-2020-0036.

Linguistic Landscape of a Modern Post-Soviet Capital - Case of Chisinau, Republic of Moldova

Ana Vasiliev (Eötvös Loránd Tudományegyetem, Hungary)

The study of language in multiethnic regions catches the attention of scientists from many different perspectives and methods of studying. Since the beginning of the XXI century, a new method has been applied for research: linguistic landscape. The study of linguistic landscapes may consider the hierarchy of languages motivated by political, administrative, and economic reasons, as reflected in the graphic design and order of inscriptions.

Since 2020, in the Republic of Moldova has been adopted a new law that regulates the placement and use of advertisements, billboards and notices. According to the law, advertising shall be disseminated in Moldovan/Romanian or, at the request of the advertiser, in other languages in accordance with the Constitution of the Republic of Moldova, the Law on functioning of languages on the territory of the Republic of Moldova, as well as international treaties to which the Republic of Moldova is a party. Despite the law, majority of advertisements in the Moldovan capital are still printed in two languages – Moldovan/Romanian or Russian. In this research, we will present the linguistic landscape of Chisinau by analyzing governmental advertisements, commercial billboards, and private notices. The choice fell on the capital of Moldova since it is a multiethnic city, home for Moldovans/Romanians, Russians, Ukrainians, Bulgarians, Gagauz, and other nationalities.

The relevance of the topic is explained by the fact that the Republic of Moldova aspires to join the European Union; in 2018 Chisinau became one of the country's pilot municipalities for the



application of the European Charter for Regional or Minority Languages, which the country has adopted, but has not yet ratified. The new signs must indicate public and cultural institutions, monuments and memorials, parks and major streets, university buildings, the railway station, and the airport with trilingual inscriptions in Romanian, a local minority language (Russian, Ukrainian, Bulgarian, German, or Yiddish) and English. In order to provide an analysis, we will use both quantitative and qualitative methods. The investigation of the linguistic landscape will help identify the dominant language in the Moldovan capital, understand relations between linguistic communities and the linguistic situation in Chisinau.

Kvantitatiivne sissevaade inglise keele kasutusse Eesti teismeliste ja täiskasvanute suulistes vestlustes

Virve-Anneli Vihman, Kristiina Praakli, Maarja-Liisa Pilvik, Mari-Liis Korkus (Tartu Ülikool)

Viimaste aastakümnete jooksul on inglise keele kui dominantse maailmakeele roll ja prestiiž ka Eestis oluliselt kasvanud. Eriti esilduvateks peetakse inglise keele mõjusid noorte keeles, kusjuures avalikkuses on seda kujutatud eeskätt negatiivse tendentsina ning ohuna eesti keele kestlikkusele. Teisalt peegeldub just noorte keelises käitumises suur osa keele varieerumisest ja keelemuutustest, samuti kohalikud ja üleilmsed ühiskondlikud protsessid (nt meediastumine, digikultuuri tähtsus, ajaveetmisviiside muutumine, suhtlusvõrgustike laienemine). Ometi puuduvad süstemaatilised empiirilised uuringud selle kohta, kas ja kuivõrd Eesti noorte inglise keele kasutus täiskasvanud kõnelejate omast tegelikkuses erineb.

Ettekandes lähtume keelekorpuste sagedusandmetest ning kõrvutame noorte inglise keele kasutust täiskasvanud kõnelejate omaga, et uurida, kas erinevused keelekasutuses ilmnevad ka kvantitatiivses analüüsis. Selleks kasutame suuliste vestluste andmeid projekti „Teismeliste keel Eestis“ (TeKE) raames loodavast noortekeele korpusest ning eesti keele spontaanse kõne foneetilisest korpusest. Keskendume põhiliselt sõnavaralistele ilmingutele ning võrdleme nii üksikkõnelejate kui ka soo ja vanuse põhjal moodustuvate kõnelejagruppide ingliskeelsete sõnade ja lausungite osakaalu, ingliskeelsete sõnade, sõnapaaride ja -kolmikute sagedust ja levikut, ingliskeelsete sõnade paare eestikeelsete sõnadega ning eesti- ja ingliskeelsete sõnade tüübi-sõne suhet (type-token ratio). Nimetatud näitajad võimaldavad ühelt poolt kvantitatiivselt iseloomustada eri aspekte koodivahetuse intensiivsusest ning teiselt poolt tuvastada ka koodivahetuse funktsioone (näiteks temaatilisi erinevusi ning sagedasi lausungisiseseid ja lausungitevahelisi üleminekuid ühelt keelelt teisele).

Meie esialgsed tulemused näitavad, et inglise keele osakaal noorte keelekasutuses on küll suurem kui täiskasvanutel, ent jääb keskmiselt 3–4 protsendi vahele ning varieerub oluliselt üksikkõnelejati. Noorte ingliskeelses repertuaaris domineerivad üksikud sagedased fraasid (nt oh my god) ning eelistatud ajaveetmisviisidega (arvutimängud ning sotsiaalmeedia) seotud sõnavara (nt bot, kick, bomb; snap, selfie, subscriber).

Negotiation of identity and Lithuanian language in the 2nd and 3rd generations of migrants from Lithuania to Kazakhstan

Loreta Vilkiene (Vilnius University, Lithuania)

This paper focuses on the aspect of formulaicity in Estonian-English-Japanese multilingual communication, within the framework of usage-based approach to contact-induced language change (Backus 2015). Applying a usage-based approach enables the analysis of



morphosyntax and lexicon under the same terminological framework (Verschik 2019). Code-copying and code-alternation cases are analyzed, utilizing Johanson's code-copying framework (2002), as it is compatible with usage-based approach.

The combination of these languages is rare, with few users (Estonians who speak Japanese). The data consists of synchronous Facebook messages, with about 10,000 tokens in the corpus. Qualitative analysis is applied for observing changes, their entrenchment and conventionalization. I concentrate on Japanese-Estonian and Japanese-English copying in cases where elements of all three languages are present.

The paper examines whether specific code-copying instances could be considered formulaic: sequences of words or other elements that are or appear to be prefabricated, not subjected to generation or analysis at the time of use (Namba 2010). Reasons for code-copying are varied, for example semantic specificity, idiomaticity, relevance for organisation of discourse, or a pragmatic gap. As copied units are strongly marked out as different, they might appear to be formulaic, and a number of formulaic sequences have a fillable gap that in itself is not formulaic (Namba 2010). A bottom-up approach allows to examine whether these copies are produced holistically, or whether an analytic construction process has occurred.

References

- Backus, A. (2015). A usage-based approach to code-switching: a need to reconcile structure and function. In G. Stell, & K. Yakpo (Eds.) *Code-switching between structural and sociolinguistic perspectives*, 19–37. Berlin: de Gruyter.
- Johanson, L. (2002). Contact-induced linguistic change in a code-copying framework. In M. C. Jones, E. Esch (Eds.) *Language change: The interplay of internal, external and extra-linguistic factors*, 285–313. Berlin: de Gruyter.
- Namba, K. (2010). Formulaicity in Code-switching: Criteria for Identifying Formulaic Sequences. *Perspectives on Formulaic Language: Acquisition and Communication*, in D. Wood (ed.), London: Continuum.
- Verschik, A. (2019). English-Estonian code-copying in blogs: combining contact linguistic and cognitive approach. In Zenner, E.; Backus, A.; Winter-Froemel, E. (Eds.) *Cognitive Contact Linguistics*, 51–80. de Gruyter.

Konverentsi korraldavad:

Eesti Rakenduslingvistika Ühing ja Eesti Keele Instituut

Konverentsi toetab:

Haridus- ja Teadusministeerium