



EESTI
RAKENDUSLINGVISTIKA
ÜHING ESTONIAN ASSOCIATION OF
APPLIED LINGUISTICS

TEESID ABSTRACTS

17. rakenduslingvistika kevadkonverents
EESTI VABARIIK 100:
RAKENDUSLINGVISTIKA VAATENURK
19.-20. aprill 2018, Tallinn

EAAL 17th Annual Conference
ESTONIA 100:
THE PERSPECTIVE OF APPLIED LINGUISTICS
April 19-20, 2018, Tallinn, Estonia

Sisukord

PLENAARETTEKANDED / KEYNOTE SPEECHES.....	3
RIITTA PYYKKÖ (University of Turku) - Multilingualism as a strength and a challenge. State and level of the Finnish language reserve and its development needs	3
LARS BORIN (Språkbanken, University of Gothenburg, Sweden) - Language corpora: Territory or map?	4
PIRKKO NUOLIJÄRVI (Helsinki) - Sotsiolingvistika perspektiive ja ülesandeid ühiskonnas.....	5
FRED KARLSSON (University of Helsinki) - How is immigration changing our Nordic linguistic landscapes?	6
RAKENDUSLINGVISTIKA I / APPLIED LINGUISTICS I	7
Using Graphs to Study Human Language	7
Eesti morfoloogia kirjeldus lõplike muundurite abil	8
„Eesti keele süntaksi“ lihtlauseehituse terminid.....	9
Tölkijate kirjutamisprotsess ja veebiotsingud ekraanipildi salvestiste põhjal	10
KEELETEHNOLOGIA JA LEKSIKOGRAAFIA / LANGUAGE TECHNOLOGY AND LEXICOGRAPHY.....	11
Edeneb takistusteta – grammatiline või leksikalne kollokatsioon?	11
EKI uus sõnastikusüsteem EKILEX: kollokatsioonide juhtum	12
+ DEMO: EKI uus sõnastikusüsteem EKILEX ja kasutajaliides SõNAVEEB	13
etLex ja etSkELL: uued ressursid eesti keele õppijatele ja õpetajatele	14
DEMO/POSTER: Lingvistilise metsa kaaren: KORP pärtingusüsteem abiks keeleuurimisel	15
RAKENDUSLINGVISTIKA II / APPLIED LINGUISTICS II	17
Stiimulite rollist sõna-assotsiatsioonide testides	17
Kreeka- ja ladinakeelsed nimed ning eesti keel	18
Eesnimede reguleerimine ja selle mõju	19
Organisatsiooniline neuroteadus kui sümbiootiline alus sõjaliste juhtide interaktsiooni analüüsimeks	20
LASTEEKELE TÖÖTUBA / WORKSHOP ON CHILD LANGUAGE	21
The distribution of parts-of-speech in mother-child interaction of Estonian and Lithuanian	21
Acquisition of Russian noun declension system in 5-year-age children with language impairment	22
Oral narrative comprehension skills in Russian 7-years-age children with language impairment	23
The acquisition of politeness strategies in Lithuanian	24
Comparative analysis of early acquisition of noun word-formation in Estonian and Russian	25
Derivation of verbs, adjectives and nouns in Finnish child language.....	26
Eesti keelt emakeelena omandavate 6-aastaste laste sõnavara seosed kasvukeskkonnaga	27
TEISE KEELE (L2) TÖÖTUBA / WORKSHOP ON SLA.....	28
Digging deeper in complexity – SLA results based on the DEMfad-model	28
Eesti keele kui teise keele kommunikatiivset arengut saatev leksikalgrammatiline areng B1- ja B2-tasemel ...	29
LAK-õpe vene õppekeelega põhikooli järgses gümnaasiumis: keele- ja aineõpetajate koostöö koolijuhtide ja õpetajate vaatevinklist.....	30
Eesti keele kui teise keele innovatiivse kursuse „Juhan“ metoodilised ja lingvistilised põhimõtted ning osalejate tagasiside	31
Vague language in English L2: a focus on Lithuanian learner English	32
Learner English versus English as a lingua franca – a theoretical comparison and implications for a corpus-based comparison of spoken language	33
The Genitive Alternation in Estonian ESL Learners	34
Effects of Transparency in Estonian EFL Learners' Idiom Comprehension	35
SOTSIOLINGVISTIKA TÖÖTUBA / WORKSHOP ON SOCIOLINGUISTICS.....	36
The Imitation Game: the effect of linguistic normativity in membership categorization.....	36
Multilingualism and the definition of language	37
Research of features of russian outside of Russia by means of integrum	38
The use of the english language on the signs in Tallinn old town 2013–2017.....	39
Preliminary study of Estonian-Russian bilinguals' colour associations	40
The analysis of different registers of spoken Lithuanian: lexical and grammatical features	41
The Spatial Organization of Language: Creating, and sustaining, intentional language ecologies in Estonia's dual-immersion kindergartens	42
Explaining English-Estonian code-copying: the role of different CMC genees	43

PLENAARETTEKANDED / KEYNOTE SPEECHES

RIITTA PYYKKÖ (University of Turku) - Multilingualism as a strength and a challenge. State and level of the Finnish language reserve and its development needs

→ **19.4.2018 10:10-10:55 Eesti Keele Instituut / Institute of the Estonian Language**

The Ministry of Education and Culture set up an investigation in February 2017, into the current state of the Finnish language reserve and its development needs. Although language proficiency needs in the society are complex, language learning has become more one-dimensional; most schoolchildren and students study only English in addition to the national languages. On the other hand, the number of immigrants has increased which affects language training needs but also increases the language reserve. Decision-making on language education has been fragmented: solutions offered apply to a single level of education and are implemented as separate projects.

The investigation used a variety of materials from statistical data to surveys and interviews and its aim is to provide a basis for drafting a national language strategy and making decisions concerning language education policy. In this presentation, the author of this investigation will give an overview of the state of language learning and language proficiency in Finland, as well as measures proposed for developing Finland's national language reserve.

LARS BORIN (Språkbanken, University of Gothenburg, Sweden) - Language corpora: Territory or map?

→ **19.4.2018 17:00-17:50** Eesti Keele Instituut / Institute of the Estonian Language

The adage “A map is not the territory” is commonly attributed to Polish-American philosopher of language Alfred Korzybski. Those of us who work with text corpora are forever faced with the question of what exactly constitutes the territory in linguistic studies. Since the mid-1990s the Association for Computational Linguistics has arranged a series of annual conferences entitled Empirical Methods in Natural Language Processing. For a few years, until the turn of the millennium, these conferences were colocated with the Workshop on Very Large Corpora series which had started in the early 1990s, but which was discontinued around the year 2000. Apparently the sense in the CL community is that corpora (in a wide sense) constitute empirical language data, i.e., they are in fact the territory, and our task as practitioners of language technology is simply to provide various kinds of maps of this territory.

But languages are interestingly different from real physical territories, and many intricate methodological questions arise when we consider how we can use, e.g., a corpus collection – such as the close to 15 billion words of Swedish text available in Språkbanken (the Swedish Language Bank) through its corpus infrastructure Korp <<https://spraakbanken.gu.se/korp/>> – to conduct research on the present and past linguistic landscape of Sweden, and in particular what the consequences can be of treating the corpus collection as the territory to be mapped (rather than, say, as a kind of map in itself).

PIRKKO NUOLIJÄRVI (Helsinki) - Sotsiolingvistika perspektiive ja ülesandeid ühiskonnas

→ **20.4.2018 12:00-12:45 Tallinna Ülikool / Tallinn University, Mare-218**

Sotsiolingvistika on keeleteaduse eriala, mis arutleb keelekasutuse ja sotsiaalse struktuuri suhete üle. Uurimisobjektiks on keelekollektiivi variatsioon, keele varieerumine ja muutumine, keele funktsioonid näiteks oma identiteedi moodustumises, suhtumine keeltesse ja arusaamat keeltest ja variantidest ning mitmekeelsus. Ajalooline sotsiolingvistika käsitleb keele muutumist pikemas perspektiivis ja kirjaliku materjali põhjal, kuna taas nüüdiskeelele keskenduv sotsiolingvistiline uurimine läheneb muutumisele kõigepealt sünkrooniliselt või lühemal ajajärgul suulise ja kirjaliku materjali valguses. Diakroonilise ja sünkroonilise uurimise piir ei ole nagunii järsk ega kategooriline, sest keelekasutuse uurimine ja teatud ajahetkel arutelu muutuste üle vajab andmeid varasemate ajajärkude keelekasutuse kohta.

Ettekanne käsitleb keelt ühiskonna arengu ja sotsiaalsete muutustega raamides ning arutleb, millisteks perioodideks sotsiolingvistilise uurimise võib jaotada. Eesmärgiks on anda ülevaade, kuidas perioodid ja uurimissuunad erinevad ja kuidas need teiselt poolt langevad kokku. Arutelu lähtub sotsiolingvistilisest variatsioonianalüüsist ja läheneb etnolingvistika erialale, mis on nüüd üldine suund mitmes ühiskonnas. Ettekanne toob ka esile, milliseid ülesandeid sotsiolingvistil on avalikus keelepoliitilises ja ühiskondlikus arutelus.

Kirjandus

Coupland, Nikolas 2007. Style: Language, variation and identity. Cambridge: Cambridge University Press.

Coupland, Nikolas 2009. Dialects, standards and social change. – Maegaard, Marie, Gregersen, Frans, Quist, Pia & Jørgensen, J. Normann (toim.). 2009. Language attitudes, standardization and language change. Oslo: Novus, 27–49.

Eckert, Penelope 2012. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. Annual review of anthropology 41, 87–100.

Nuolijärvi, Pirkko. 2000. Sosiolinguistikka kielentutkimuksen kentässä. – Sajavaara, Kari & Piirainen-Marsh, Arja (toim.) Kieli, diskurssi & yhteisö. Soveltavan kielentutkimuksen teoriaa ja käytäntöä 2. Soveltavan kielentutkimuksen keskus. Jyväskylä: Jyväskylän yliopisto, 13–37.

FRED KARLSSON (University of Helsinki) - How is immigration changing our Nordic linguistic landscapes?

→ **20.4.2018 15:25-16:15** Tallinna Ülikool / Tallinn University, Mare-218

Linguistic ecology is the study of languages in relation to one another and to various social factors. It is also known as language ecology or ecolinguistics. Peter Mühlhäusler has said that “[t]he ecological metaphor is action oriented. It shifts the attention from linguists being players of academic language games to becoming shop stewards for linguistic diversity, and to addressing moral, economic and other 'non-linguistic' issues.”

In this spirit I shall discuss the new language situations arising in the Nordic countries (especially Finland, Sweden and Estonia) due to the rapidly growing influx of migrants from all around the globe during the past 25 years. The domestic ensembles of different languages are much more heterogeneous than is currently thought and known, both among policy makers, administrators and professional linguists.

Questions to be discussed include these:

- What is the current situation of language registration (and how could it be improved)?
- How many and what are the home languages (HLs) of the migrants?
- What are the future prospects of the HLs in their new surroundings?
- What could be a domestic language taxonomy including traditional languages and HLs?
- What could be realistic goals for domestic language policies concerning HLs?

RAKENDUSLINGVISTIKA I / APPLIED LINGUISTICS I

Using Graphs to Study Human Language

Tõnu Tamme (University of Tartu, Institute of Computer Science), Leo Võhandu (Tallinna Tehnikaülikool)

It is well-known that social networks can be studied with graphs. The same methods can be used to study human language. There are various links between language objects such as syntactic, semantic, or adjacency relations (Cong 2014). One can also find closeness relations between the works of different writers, or direct links between the acts of law. Properties of graphs have been used to study the structure of Estonian Wordnet (Lohk 2015).

One can study various graph properties such as vertex connectivity, lack of cycles, or degree distribution. Graphs can have one or several centers, but there can be graphs without any centers. We are also interested whether the graph under consideration has some statistical properties such as small-world or scale-free properties. Word graphs have been used to show the closeness of Slavic languages but also to examine relatedness of English and Chinese languages.

We use the graph-theoretic approach to compare the works of Estonian poets. We study how the graph presentation of their creation reflects changes in their writing style.

References

Jin Cong, Haitao Liu 2014. Approaching human language with complex networks. Physics of Life Reviews 11 (2014) 598–618.

Ahti Lohk 2015. A System of Test Patterns to Check and Validate the Semantic Hierarchies of Wordnet-type Dictionaries. PhD thesis, TTU.

Eesti morfoloogia kirjeldus lõplike muundurite abil

Heiki-Jaan Kaalep (Tartu Ülikool)

Eesti keele morfoloogia kirjeldamine lõplike muundurite (finite state transducers, edaspidi FST) abil sai alguse üle 20 aasta tagasi (Uibo 1998). Teoreetilistest ja metodoloogilistest kaalutlustest lähtudes peaks just lõplike muundurite kasutamine olema parim viis, et mingi keele morfoloogilist analüsaatorit või süntesaatorit teha, s.t. parim viis morfoloogia-alaste teadmiste esitamiseks arvutilingvistika ja keeletehnoloogia valdkonnas. Eesti keele puhul näeme praegu siiski olukorda, kus kasutust leib hoopis mingis programmeerimiskeeltes realiseeritud keelspetsiifiline tarkvara, olgu selleks siis EKI või Filosofti loodu (s.h. viimast kasutav EstNLTK). Selle põhjuseks võib olla, et senised FST-põhised käsitlused (Pruulmann-Vengerfeldt 2010) ei hõlma piisava kvaliteediga mõningaid praktiliselt seisukohalt olulisi aspekte, eeskätt sõnastikust puuduvate sõnade töötlemist, s.h. liitsõnamoodustust.

Ettekanne kirjeldab viimase kahe aasta jooksul tehtud tööd uue FST-põhise morfoloogiakirjelduse loomisel. Olulisemad aspektid on järgmised:

- Morfotaktika ja morfonoloogia on selgelt eristatud.
- Morfotaktikat kirjeldatakse muuttüüpide kaudu, mis on realiseeritud tüvede ja jätkuleksikonide abil.
- Morfonoloogiat kirjeldatakse kahe-tasemeliste reeglite abil.
- Paralleelvormide olemasolu korral on alati märgitud, milline neist on levinuim ja millised on harvemad.
- Ka levinumad ja süstemaatilisemad mitte-normikohased vormid on esindatud, seejuures vastava märgendiga.
- Muuttüüpide süsteem erineb mõneti kõigist varasematest.
- Astmevahelduslikkuse võimalikkuse määrab muuttüüp, selle realiseerumise konkreetse sõna puhul aga morfonoloogiline reegel. Näiteks kott, president ja seminar kuuluvad samasse astmevaheldust võimaldavasse tüipi, seminar puhul astmevaheldus aga ei realiseeru, sest sõna häälikkosseis seda ei võimalda.
- Mitte-produktiivseid häälikuvhaheldusi esindavad tähed on sõna leksikaalses esituses tähistatud erimärkidega, produktiivseid esindavad tähed aga tavaortograafias, näiteks leiB1 (leib-leiva) ja kott (kott-koti). Sel moel on tänapäeva keele seisukohalt ebareeglipäraselt muutuvatele sõnadele rakenduvad kahe-tasemelised morfonoloogiareeglid leksikaalselt piiratud.
- Kirjeldus on leksikonipõhine, järgides seejuures põhimõtteid, mida on avaldanud (Kaalep 2012), (Kaalep 2015)
- Kirjeldus hõlmab kogu sõnavara, mis on Filosofti analüsaatori leksikonis, aga ka sõnamoodustust – tuletust ja liitsõnamoodustust.

Viited

- Kaalep, H.-J. 2012. Käänamissüsteemi seaduspärasused. Keel ja Kirjandus, 6/2012, 418–449.
- Kaalep, H.-J. 2015. Eesti verbi vormistik. Keel ja Kirjandus, 1/2015, 1–15.
- Pruulmann-Vengerfeldt, J. 2010. Praktiline lõplikel automaatidel põhinev eesti keele morfoloogiakirjeldus. Magistritöö. TÜ ATI, Tartu.
- Uibo, H. 1998. Kahetasemeline morfoloogiamudel ja eesti keel. Keel ja Kirjandus 1/1998, 13–21.

„Eesti keele süntaksi“ lihtlauseehituse terminid

Peep Nemvalts (Tallinna Ülikool)

„Eesti keele varamu“ sarja III köitenä ilmunud süntaksiülevaade on saatesõna järgi kirjutatud selleks, „et keeleõpetusel, keelekorraldusel ja ka keeleuurijail endil oleks, millele oma töös tugineda. Raamat on seega eelkõige möeldud nende valdkondadega tegelejatele, samuti filoloogiaüliõpilastele.“ (EKS 2017: 5) Vaatlen selles teoses tarvitatud keskset lihtlauseehituse terminivara kahest lähtekohast: 1) kas ja kuidas on terminid muutunud võrreldes eelmises suures süntaksiköites „Eesti keele grammatika II“ (EKG II 1993) kasutatuga (nt preteeritum, vrd varasem imperfekt; 2) kui otstarbekas on valitud termin tähistatava mõiste jaoks ning kogu käsitletavat mõistestikku silmas pidades.

Kirjandus

EKG II 1993 = Mati Erelt, Reet Kasik, Helle Metslang, Henno Rajandi, Kristiina Ross, Henn Saari, Kaja Tael, Silvi Vare. Eesti keele grammatika II. Süntaks. Lisa: kiri. Trükki toimetanud Mati Erelt peatoimetajana, Tiiu Erelt, Henn Saari ja Ülle Viks. Tallinn: Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.

EKS 2017 = Eesti keele süntaks. Eesti keele varamu III. Toimetanud Mati Erelt ja Helle Metslang. Tartu: Tartu Ülikooli Kirjastus.

Tõlkijate kirjutamisprotsess ja veebiotsingud ekraanipildi salvestiste põhjal

Marju Taukar (Tallinna Ülikool)

Tõlkimise õpetamisel ja üleüldse tõlkimise olemuse mõtestamisel annab väärthuslikku infot tõlkimisprotsessi uurimine. Nii varem kasutuses olnud valjustimõtlemisprotokollid kui ka küsitlused annavad vaid vahendatud infot. Näiteks on meetodeid võrdlevatest uurimustest selgunud, et tõlkija ei pruugi mõnda tekkinud probleemi hiljem mäletada või isegi mitte märgata. Lisaks on esimese puhul oluliseks minuseks ebaloomulik sekkumine tõlkimisprotsessi. Ekraanipildi, klaviatuuritegevuste ning silmade liikumise salvestamine on vahetu viis, et näha, mida tõlkija täpselt teeb ning mis talle probleeme võib valmistada. Neist kolmest kõige kergemini teostatav (kuigi analüüs mõttes kõige ajamahukam) viis tõlkimisprotsessi uurida, on vaadelda tõlkija tegevusi ekraanosalvestuse abil. Sedasi on võimalik leida vastuseid küsimustele tõlkija töökorralduse kohta (nt milline on dokumentide paigutus ekraanil, kas tõlgitakse lähteteksti sees jne), tõlkimisprotsessi etappide kohta (lugemine, kirjutamine, toimetamine), tekstiloome viisiide kohta (kas tõlgitakse jooksvalt või tehakse vahepeal otsinguid, millal tehakse parandusi jne) ning tõlkija veebiotsingute kohta. Ettekandes käsitletav uurimismaterjal koguti koos tudengite Tallinna Ülikooli ELU kursuse projekti raames ning sisaldas nii kirjalikke kui suulisi tõlkeit. Keskendun praegu vaid kirjaliku tölke protsessile. Uurimismaterjaliks on 4 tõlkija tõlkimisprotsessiaegsed ekraanipildi salvestised. Käsitlen tõlkijate töökorraldust, ajalist jaotust kirjutamise ja otsingute vahel ning veebiotsinguid.

Tulemustest on näha, et nii tõlkijate kasutatavad allikad, otsinguviiisid kui üldine tööprotsess erineb suurel määral. Kuna osalejate taust ja tõlkimisalane kogemus olid erinevad, võiks uurimust korrrata, suurendades kõiki gruppe ja kogudes ka täpsemaid taustaandmeid. Nii oleks võimalik näha, mis eristab alustavate ja kogenud tõlkijate tõlkimisprotsessi ja kuidas seda tõhusamaks muuta. Samuti võiks ekraanipildi salvestamise ja analüüs kaasata õppeprotsessi, sest nii saavad nii juhendaja kui õppija nii enda kui teiste tõlkimisprotsessi kohta infot, mida vaid teksti põhjal ei saa.

KEELETEHNOLOGIA JA LEKSIKOOGRAAFIA / LANGUAGE TECHNOLOGY AND LEXICOGRAPHY

Edeneb takistusteta – grammatical line või leksikaalne kollokatsioon?

Geda Paulsen (Eesti Keele Instituut)

Kollokatsioonide analüüsimes on tavaks jagada koosesinevad sõnad kollokatsiooni liikmete sõnaliigikuuluvuse põhjal kahte põhitüpi: grammaticalists ja leksikaalseteks kollokatsioonideks (vt Nattinger 1988, Howarth 1988). Grammatilised kollokatsioonid on tüüpiliselt kahe täistähendusliku sõna (nimisõna, verbi või adjektiivi) ühendid mittetäistähendusliku sõnaga (adverbi, adpositiooni või näiteks infinitiivtarindiga jne), leksikaalseks kollokatsiooniks peetakse aga kahe täistähendusliku sõna sagedat koosesinemist.

Eesti keele sõnaliikide vaheliste piiride hägususe tõttu ei ole üksühene ka grammaticaliste ja leksikaalsete kollokatsioonide jaotus. Ühe hajusate piiridega rühma moodustavad adverbilaadsed noomenid ehk nimisõnad, mille tähendus näitab abstraheerumise märke ning mis kaugenevad nii semantiliselt kui süntaktiliselt sõna algsest, konkreetsest funktsionist. Niisugused võimalikult grammatiseerumisprotsessis olevad sõnad äratavad mitmeid küsimusi. Millistes kollokatsioonides adverbilaadsed nimisõnad esinevad ja milline on (verbi)kollokaadi mõju nendele noomenitele? Milliseid tähendusnüansse nad väljendavad? Millised käänded on potentsiaalsete uute adverbitüüpide allikad ning kuidas mõjutab käändesemantika potentsiaalset adverbi? Ning kollokatsioonisõnaraamatu koostamise seisukohalt kõige olulisem küsimus – kuidas peaks adverbilaadseid nimisõnu sõnaraamatu kasutajale esitama, leksikaalsete või grammaticaliste kollokatsioonidena, ehk siis nimisõnade või määrsõnadena?

Ettekande eesmärk on vaadelda lähemalt mõningaid eesti keele kollokatsioonisõnaraamatu koostamise käigus esile tulnud adverbilaadseid nimisõnu ning püüda leida lahendusi hägusate sõnaliigipiiride probleemile leksikograafilise analüüsiga vaatenurgast. Keskendun erikäändelistele adverbilaadsetele noomenitele ehk nimisõnadele, mida leksikograafilise tarkvara Sketch Engine ja eesti keele seletav sõnaraamat käsitlevad nimisõnadena, ehk verbikollokaatidele saavas (eluajaks ametisse nimetama), rajavas (ärvahetamiseni sarnanema), olevas (lõppakordina tulema), ilmaütlevas (takistusteta sujuma) ja kaasaütlevas käändes (tõusude-mõõnadega kulgema).

Kirjandus

Eesti keele seletav sõnaraamat 2009. <http://www.eki.ee/dict/ekss/index.cgi>.

Howarth P. 1998. Phraseology and Second Language Proficiency, Applied Linguistics XIX.
Nattinger, J. R. 1988. Some current trends in vocabulary. In R. Carter & M. McCarthy, (Eds) (1988), Vocabulary and language teaching. London/New York: Longman.

Sketch Engine, <https://www.sketchengine.co.uk/auth/corpora/>

EKI uus sõnastikusüsteem EKILEX: kollokatsioonide juhtum

Arvi Tavast (Tripledev), Jelena Kallas (Eesti Keele Instituut), Kristina Koppel (Eesti Keele Instituut), Margit Langemets (Eesti Keele Instituut), Geda Paulsen (Eesti Keele Instituut), Maria Tuulik (Eesti Keele Instituut)

Eesti keele kollokatsioonisõnastikku (Kallas jt 2015) on EKI-s koostatud alates 2014. aastast EKI sõnastikusüsteemis EELEX (Langemets jt 2006, Jürviste jt 2011). Sõnastik on plaanitud ilmuma 2018. a sügisel. Alates 2017. a on EKI-s arendatud uut, ühtsel andmemudelil põhinevat sõnastikusüsteemi EKILEX ning sellele toetuvat avalikku kasutajaliidest Sõnaveeb. Paralleelselt toimetamisega alustasime 2018. aastal selle sõnakogu importimist uude süsteemi.

EKILEXi üks olulisi eesmärke on parandada EKI sõnakogude ühtsust: praegu on eri sõnastikes laiall igat liiki infot, kohati dubleerivalt ja kohati vastuoluliselt. Kasutaja peab kannatlikult naveerima paljudes eri sõnastikes, et oma küsimusele vastust leida. Pikas perspektiivis oleks ideaalvariant n-ö üksainus EKI agregaatorilaadne portaal, mis annaks infot sõnade, mitte sõnastike kohta. Kollokatsioonid on tüüpiline näide infoliigist, mida praegu esitatakse paljudes sõnastikes (nt seletavas sõnaraamatus, ŶS-is, kakskeelsetes sõnastikus). Seetõttu on sellise info importimise lahendamine ülivajalik. Lisaks on EKILEXi jaoks oluline nende andmete struktureerimise ja kvantifitseerimise potentsiaal.

Kollokatsiooni võib vaadelda harjumusliku seosena kahe (või enama) sõna vahel. EELEXis on kollokatsioon esitatud koosnevana märksõnast ja kollokaadist, millele mõnikord on lisatud ka kontekst. Sõnade jaotus nende kategooriate vahel põhineb sagedusel ja seose tugevuse sel (statistik LogDice): märksõnadeks on ligikaudu 10 000 sagedamat sõna Eesti keele ühendkorpuuses 2013. Kollokaatideks on nii korpusest leitud kõrge koosesinemissagedusega sõnad kui ka käsitsi lisatud kõrge esilduvusega sõnad. Konteksti tugevam ja nõrgem seos võib sõltuda märksõnast (joonis 1).

TÜHI 1

/verest/ tühjaks jooksma

VERI

verest /tühjaks/ jooksma

JOOKSMA 2

/verest, veest/ tühjaks jooksma

Joonis 1. Näide sama kollokatsiooni esitusest eri märksõnade juures (märksõna järel on tähindusnumber, poolpaksus kirjas on kollokaat, kaldsulgudes on kontekst)

EKILEXis hoiamo andmeid maksimaalselt struktureerituna, st iga andmeelemendi kohta peaks olema võimalikult täpselt teada, mida see endast kujutab. Kollokatsioonisõnastiku importimisel oleme kohtunud kahe olukorraga, milles esimene on lihtne (kuigi mitte välimatult lihtasti lahendatav), aga teine tundub elevat huvitav ka teoreetiliselt.

- Kollokaadid ja kontekstid on tundmatud tähejadad, mitte tähindusega sõnad: ei ole teada, millist "tühi" ja "jooksma" tähindust on mõeldud (joonis 1). Seega pole teada, millise sõnaga millises tähinduses see kollokatsiooniseos luua.

- Jaotus märksõnaks, kollokaadiks ja kontekstiks. Nagu jooniselt 1 näha, saab sama kollokatsiooni esitada mitmel eri viisil, mis köik paistavad omamoodi mõistlikud.

Kolmikjaotuse asemel seotuse skaala? Kollokatsioon rohkem kui kahe sõna vahel?

Rekursiivsed kollokatsioonid (kollokaat on ise kollokatsioon)?

Ettekandes kirjeldame selle hetke töövarianti, kuidas oleme mõelnud imporditavaid andmeid struktureerida, mis sellega paremaks läheb ja millised probleemid tekivad.

Kirjandus

Jürviste, M., Kallas, J., Langemets, M., Tuulik M. & Viik, Ü. (2011). Extending the functions of the EELEX dictionary writing system using the example of the Basic Estonian Dictionary. – I. Kosem & K. Kosem (eds.) eLexicography in the 21st Century: New Applications for New Users, Proceedings of eLex 2011. Ljubljana: Trojina, Institute for Applied Slovenian Studies, 106–112.

Kallas, J., Koppel, K. Tuulik, M. (2015). Korpusleksikograafia uued võimalused eesti keele kollokatsioonisõnastiku näitel. – Eesti Rakenduslingvistika Ühingu aastaraamat 11, 75–94.

Langemets, M.; Loopmann, A. & Viik, Ü. (2006). The IEL dictionary management system of Estonian. – G.-M. De Schryver (ed.) DWS 2006: Proceedings of the 20 Fourth International Workshop on Dictionary Writing Systems: Pre-EURALEX workshop: Fourth International Workshop on Dictionary Writing System. Turin: University of Turin, 11–16.

+ DEMO: EKI uus sõnastikusüsteem EKILEX ja kasutajaliides SÕNAVEEB

Arvi Tavast (Tripledev), Jelena Kallas (Eesti Keele Instituut), Kristina Koppel (Eesti Keele Instituut), Margit Langemets (Eesti Keele Instituut), Geda Paulsen (Eesti Keele Instituut), Maria Tuulik (Eesti Keele Instituut)

etLex ja etSkELL: uued ressursid eesti keele õppijatele ja õpetajatele

Jelena Kallas, Kristina Koppel (Eesti Keele Instituut), Katrin Tsepelina (Tartu Ülikool)

Tutvustame ettekandes kaht uut ressurssi, mis on abiks eesti keelt teise keele või võõrkeelena õppijatele ja õpetajatele.

Esmalt tutvustame tööriista etLex, mis hindab eesti keele eri kleeoskustasemete sõnavara ja teksti leksikaalset keerukust. etLex on loodud abistama peamiselt kleeõpetajaid ja testide koostajaid õppematerjalide koostamisel. Tööriist analüüsib tekste eesti keele A1–C1 aktivse ja passiivse sõnavara loendite põhjal. Aktiivse sõnavara loendite koostamisel olid aluseks seni ilmunud eri kleeoskustasemete sõnavaraloendid, Eesti Keele Instituudis koostatud ja koostamisel olevate õppesõnastike materjalid ja eesti keele korpused. Passiivse sõnavara loendid koostati Eesti Keeleressursside Keskuse ja Eesti Keele Instituudi 2017. a projekti „Eesti keele õpikute korpus A1–C1“ käigus. Kokku on etLexis määratud ligikaudu 13 000 sõna tasemekohasust. etLexi eeskujudeks on inglise English Profile (Capel 2010, 2012), rootsi SVALex (François et al. 2016) ja prantsuse FLELex (François et al. 2014).

Teiseks tutvustame etSkELLi, mis kuulub automaatselt koostatud kleeõppakeskkondade SkELL (Sketch Engine for Language Learners, Baisa ja Suchomel 2014) sarja. Õppakeskkonna aluseks on spetsiaalne eesti keele õppekorpus, mis sisaldab kleeõppijale sobiva leksikaalse ja grammatilise keerukusega lauseid (25 mln lauset, 250 mln sõna). Laused on välja valitud 2017. a loodud eesti keele ühendkorpusest (1,6 mld sõna) tööriista GDEX (Good Dictionary Example, Kilgarriff jt 2008) abil (Koppel 2017).

etSkELL võimaldab ligipääsu kleeandmestikule kolmel moel:

- 1) näitelaused ehk otsitav sõna oma tavapärases kontekstis;
- 2) seotud sõnad ehk kollokatsioonid;
- 3) sarnased sõnad ehk sõnad, mis esinevad otsitava sõnaga sarnastes kontekstides.

Kirjandus

Baisa, V., Suchomel, V. 2014. SkELL: Web Interface for English Language Learning. – Eighth Workshop on Recent Advances in Slavonic Natural Language Processing. Brno: Tribun EU, 63–70.

Capel, A. 2010. A1–B2 vocabulary: insights and issues arising from the English Profile Wordlists project. English Profile Journal, 1(1):1–11.

Capel, A. 2012. Completing the English Vocabulary Profile: C1 and C2 vocabulary. English Profile Journal, 3:1–14.

Kilgarriff, A., Husák, M., McAdam, K., Rundell, M., Rychlý, P. 2008. GDEX: Automatically finding good dictionary examples in a corpus. – Proceedings of the 13th EURALEX International Congress, 425–432.

Koppel, K. 2017. Heade näitelausete automaattuvastamine eesti keele õppesõnastike jaoks. Eesti Rakenduslingvistika Ühingu aastaraamat, 13, 53–71.

François, T., Volodina, E., Pilán, I., Tack, A. 2016. SVALex: a CEFR-graded Lexical Resource for Swedish Foreign and Second Language Learners. Proceedings of LREC 2016, 213–219.

François, T., Gala, N., Watrin, P., Fairon, C. 2014. FLELex: a graded lexical resource for French foreign learners. Proceedings of LREC 2014, 3766–3773.

DEMO/POSTER: Lingvistilise metsa kaaren: KORP pärtingusüsteem abiks keeleuurimisel

Olga Gerassimenko, Neeme Kahusk, Kadri Vider (Tartu Ülikool)

Eesti Keeleressursside Keskuses on aktiivselt arendatud mitmeid teadlaste uurimistööd hõlbustavaid veebiteenuseid. KORP (<https://korp.keeleressursid.ee/>) on tööriist paindlike pärtingute tegemiseks tekstikorpustest.

Lisaks varasemale poolele miljonile käsitsiühestatud sõnele (Gerassimenko, Kahusk 2017) on nüüd võimalik pärda korpusi ligikaudu poole miljardi sõne ulatuses. KORPi lisatud eesti keele korpused on automaatsest lausestatud, morfoloogiliselt analüüsitud ja ühestatud.

Automaatse ühestamise töttu võib esineda ka mitmeseid ja valesti ühestatud analüüse.

Morfoloogilise ühestamise täpsus on ligikaudu 93–98% (täpsemalt vt. Veski ja Liba 2008), ühestamata sõnalike ca 3,6%. Lisaks on eesti keele korpuutes leitud ka osalausete piirid, osalauseteks jagamise täpsus on ligikaudu 96% ja saagis ligikaudu 95% (Kaalep ja Muischnek 2012).

Tänapäevase korpuuse otsingusse koondatud ressursid võib jagada kahte suuremasse rühma:

- (1) Veebikorpus 2013 ja (2) TÜ Koondkorpus.

Veebikorpus 2013 on SketchEngine firma poolt automaatselt veebist korjatud eestikeelsete veeblehtede korpus. Korpuuses on ligikaudu 310 miljonit sõnet ligikaudu 21 miljonis lauses. Korpuuses puudub täpsem ajateave, kuid ükski tekst ei ole pärít hilisemast ajast kui jaanuar 2013. Korpus on jagatud kümneks alaosaks suurustega 20–50 miljonit sõnet. Osadeks jagamine on olnud üsna meelevaldne, selles pole järgitud ei ajalist ega žanrilist jaotust, isegi ühelt ja samalt domeenilt pärít tekstdid võivad olla sattunud erinevatesse alaosadesse. Selle allkorpuuse lähem kirjeldus on leitav EKRK registrist aadressilt <http://doi.org/10.15155/1-00-0000-0000-0000-0012EL>.

Koondkorpus koosneb ilukirjanduse, teaduskirjanduse, ajakirjanduse, seaduste ja Riigikogu protokollide allosadest, milledest mahukaim on ajakirjanduse (ca 152 miljonit sõne) ja väikseim Eesti seaduste (ca 2 miljonit sõne) osa. Tänapäevases korpuuses on võimalik alaosi valida ka ajaskaalal, samuti otsida lisaks muudele tunnustele ka ajaväljendeid.

Võru- ja vadjaeksed tekstdid, mis on esitatud eraldi vahetehtedel, on märgendamata. Võrukeelne aines pärineb rootsi keeletehnoloogiakeskuse Giellatekno allalaetud võru vikipeediast (vt <http://giellatekno.uit.no/cgi/index.vro.eng.html>) ja vadjaekelne aines Kristian Kankaise töödeldud vadja keele õpikogumikust "Sõnakopittõja" (Heinsoo 2015) ja vadja keele sõnaraamatu näitelausetest (2013).

Peatselt ilmub EKRK KORPi pealehele ka praegu eraldi aadressil kasutatav EKI keeleõpikute korpus, kus leiab eestikeelseid näitelauseid köigi ELI keeletasemete (A1-C1) kohta.

Allkorpuuse täpsem kirjeldus on leitav EKRK registrist aadressilt <http://doi.org/10.15155/3-00-0000-0000-0000-06ADEL>.

KORP võimaldab esitada konkordantsipärtinguid lihtpärtinguna (pärting üksiku sõna või sõnaosa kaupa), laiendatud pärtinguna (metaandmete, grammatiliste kategooriga ja sõnajärjendite liht- ja liitpärting) ja pärtinguna edasijõudnutele (pärting CQP avaldistena). KORPi pärtinguliides salvestab pärtingute ajalugu ja võimaldab varasemaid pärtinguid lihtsa vaevaga taasesitada. Uus funktsioon "võrdle" võimaldab ülevaatlikult kõrvutada kahe salvestatud pärtingu tulemusi. Graafiline pärtinguliides on intuitiivne ja mugav, kasutatav nii kogenud uurijale kui ka keelehuvilisele. Pärtingutulemuste esituse graafiline mugavus annab

kiiresti ülevaate, millistes alamkorpustes otsitav nähtus esineb, ning võimaldab hõlpsalt liikuda eri alamkorpuste konkordantside ja päringustatistika vahet. Statistikat esitab KORP nii absoluutarvudes kui suhtarvudes (X otsitavat nähtust miljoni sõne kohta); statistikat saab koostada ja korraстada eri grammatiske kategooriate ja metaandmete alusel.

KORPi arendamist ja eestikeelsete korpuse lisamist toetatakse ERFi projektist "Eesti Keeleressursside Keskuse (EKRK) ühendatud sisuotsing" (2014-2020.4.01.16-0134) tegevuse alt "Riikliku tähtsusega teaduse infrastruktuuri toetamine teekaardi alusel".

RAKENDUSLINGVISTIKA II / APPLIED LINGUISTICS II

Stiimulite rollist sõna-assotsiatsioonide testides

Ene Vainik (Eesti Keele Instituut)

Sõna-assotsiatsiooniks nimetatakse sõnalisele stiimulile antavat spontaanset sõnalist vastust (Krasnõh 2001). Vastuste seas on tavaiselt märkimisväärne kokkulangevus, mida juba meetodi loomisest saadik on nimetatud assotsiatsiooninormiks (Kent ja Rosanoff 1910; Postman ja Keppel 1970). Vastuste varieerumist vastajate lõikes on seletatud individuaalsete ja grupikäitumise erinevustega (Nelson jt 2012; Fitzpatrick jt 2013).

Eesti keele assotsiatsioonisõnastiku koostamisel ilmnes, et stiimulsõnu endid iseloomustab küllalt suur erinevus neile antavate vastuste üksmeeles (5—75 %). On sõnu, millele enamik inimesi vastab sarnaselt (nt *garaaž, vasak*), kuid on ka selliseid, mille puhul inimesed kalduvad vastama väga erinevalt (nt *aus, hobune, jalgratas*).

Ettekandes võrreldakse kõige üksmeelsemate ja hajuvamate vastustega stiimulite gruppide ja vaadeldakse mõningaid stiimulit iseloomustavaid tegureid, mis võiks mõjutada seda, kas sõna saab katsetes pigem üksmeelseid või hajuvaid vastuseid (nt sõnaliik, sõna pikkus, morfoloogiline komplekssus, sagedus korpuses, polüseemia, semantiline kuuluvus, emotsionaalne valents, osalemise kindlaksujunenud semantilistes suhetes).

Senistes uurimustes on pakutud, et vastuste üksmeelt vs hajuvust võiks mõjutada sõna sagedus, sõnaliik, omandamise iga, ettekujutatavus ning osalemine ennustatvates seostes (Fitzpatrick jt 2013).

Kirjandus

Fitzpatrick, Tess Playfoot, David Wray, Alison Wright Margareth J. 2013. Establishing the reliability of word association data for investigating individual and group differences. – Applied Linguistics 36(1), 23-50. doi:10.1093/applin/amt020

Kent, Grace H.; Aaron. J. Rosanoff 1910. A study of association in insanity. – American Journal of Insanity 67(1-2), 37-96.

Krasnõh, V. 2001. Osnovõ psihholingvistiki i teorii kommunikatsii. Moskva: Gnosis.

Nelson, Douglas. L.; McEvoy, Cathy. L.; Dennis, Simon 2012. What is free association and what does it measure? – Memory & Cognition 28(6), 887-899.

Postman, Leo; Keppel Geoffrey (eds)1970. Norms of Word Association. Elsevier.

Kreeka- ja ladinakeelsed nimed ning eesti keel

Kaarina Rein (Tartu Ülikool)

Vanakreeka ning ladina keele hääl dust mõjutavad tänapäeval nii traditsioonid, rahvuskeeled kui ka keele- ja kultuurikontaktid. Sama kehtib ka kreeka- ja ladinakeelsele nimede kohta, sest Eestis kasutatakse neid teisiti kui mujal maailmas. Sellest, et ladina keelt hääl dab iga rahvas iseenda keele reeglite järgi, kirjutas juba Johannes Aavik oma ideede päevikus „Ideepe” 1940ndatel aastatel. Erinevad kontekstid mõjutavad nii kreeka kui ladina keele kasutust tänapäevalgi. Kuna Eestis on ka käesoleval sajandil nii kreeka- kui ladinakeelseid nimesid ja pealkirju kasutusele võetud, võib siangi tähdada eripärasid nii valikus, hääl damisel, röhutamisel kui ka mugandamisel. Ka nimede tõlkimisel ladina keelde võib tähdada eesti keele eripäradest tingitud vigu.

Ettekanne vaatleb nii eesti kui ka teiste keelte mõju Eestis 20. ja 21. sajandil kasutusele võetud kreeka- ja ladinakeelsetes nimedes ja pealkirjades.

Johannes Aavik „Ideepe. Johannes Aaviku ideede päevik.” Väljaande koostaja ja peatoimetaja Helgi Vihma. Johannes Aaviku Selts. Tallinn: TEA kirjastus, 2010.

Līva Bodniece, Audronė Kučinskienė, Maria-Kristiina Lotman „Kreeka pärisnimed antiikkirjanduse õpikutes Baltimaade näitel.” – Keel ja Kirjandus nr 4/2017, 291–304.

Peeter Päll „Võõrnimede hääl dus.” – Keel ja Kirjandus nr 10/2000, 737–738.

Kaarina Rein „Kreeka nimed eestikeelsetes tekstis ja kõnes.” – Eesti Rakenduslingvistika Ühingu aastaraamat nr 13/2017. Tallinn, 157–169.

Eesnimede reguleerimine ja selle mõju

Annika Hussar, Tiina Rüütmaa (Tallinna Ülikool)

Eesti ja Ungari eesnimekasutust on reguleeritud väga erinevalt. Mõneti võib meie eesnimetavasid vaadata kui ühe reguleerimisskaala kaht äärmist punkti, kus Eesti nimekasutust iseloomustab peaaegu täielik vabadus ja Ungari kasutust üsna täpne reguleeritus.

Ungaris hakati nimekasutust reguleerima samal ajajärgul, kui nende nimevara oluliselt uuenes, s.o 19. sajandil. Loodi riiklik nimeregister ning võeti vastu nimeseadus, mis astus jõusse 1895. Registrisse oli lubatud kanda ainult ungari kirjaviisis nimesid (nt János), muid ettekirjutusi esialgu polnud. 1971. aastal ilmus esimene eesnimeraadamat, sellest ajast alates võis kasutada ainult nimeraamatus esindatud nimesid. Praegu kehtiva 2010. aastal vastu võetud seaduse järgi lubatud eesnimed on kantud nimeloendisse, mida haldab Ungari teaduste akadeemia. Kui nimi ametlikust loendist puudub, võivad vanemad siiski pöörduda teaduste akadeemia poole palvega nimi registrisse kanda.

Eesti nimevara analoogne uuenemine sai hoo sisse parajagu 100 aastat tagasi, kõrgaeg saabus 1930. aastatel. Ka muus mõttes moderniseerunud nimevalikusse laenati teiste rahvaste nimesid, nende kõrval levisid aga omakeelsed nimed. Uusi ja vanu unustatud nimesid pakuti u 3000, populaarseks sai neist mõnikümmend, ülejäänud nimed esinesid üksikkordi.

Eesnimede kasutust pole oluliselt reguleeritud, Eesti isikunimeseadus on pärit alles sellest sajandist, see hakkas kehtima 2005. aastal.

Vaatleme oma ettekandes selliste traditsioonide mõju kummagi maa eesnimekasutusele, toome näiteid ka teiste maade nimeseadustest.

Organisatsiooniline neuroteadus kui sümbiootiline alus sõjaliste juhtide interaktsiooni analüüsimiseks

Silvi Tenjes (Eesti Kaitseväe Ühendatud Õppeasutused)

Ametisuhtlus sõjalistel juhtidel sisaldab peale keeleväljendite ka mittekeelelisi suhtlusviise. Vahetu suhtlus on olemuslikult multimodaalne e mitmeviisiline protsess, kus osaleb mitu suhtlusmodaalsust samaaegselt. Töö pakub uusi teoreetilisi lahendusi interaktsiooni multimodaalsuse rakendamisele sõjaliste juhtide suhtuskäitumise analüüsimsel. Töös teadvustatakse mitmeviisilise suhtluse rolli ja tähtsust juhtidele olulistes situatsioonides ning esitatakse organisatsioonilise neuroteaduse (ON 2015) seisukohti organisatsiooni juhtimise, mitmeviisilise suhtluse ja grupitöö käsitlemiseks.

Organisatsiooniline neuroteadus on 21. sajandil arenenud valdkond. Nagu uuel valdkonnal ikka, on sel mitmeid erinevaid definitsioone. Vaatamata erinevustele on definitsioonide ühisosas kolm elementi: 1) analüüsi aju-tasand; 2) analüüsi organisatsiooniline tasand ning 3) interaktsionid aju-uuringute ja organisatsiooniliste uuringute vahel. Organisatsioonilise käitumise uurimise aspektid keskenduvad indiviidi ja gruvi analüüsile, millesse kaasatakse neuroteaduse teooriad ja meetodid. Näiteks neuroloogiliste uuringute tulemusel on leitud ühisosa grupimõtlemise, jagatud mentaalsete mudelite, mälu ja jagatud juhtimise teemadega (Waldman jt 2015: 284). Neuroteaduses tuntud peegelneuronite käsitlused lisavad teavet arutellu, kas mentaalsetel kujutuspiltidel on keeeline või pildiline alus (Willingham, Dunn 2003). Teoreetiliste mudelite hulgas on eeliskasutuses keelt, kultuuri ja kognitiivsust hõlmav mudel (Volk jt 2014). Töö on alanud projekti uuemate teoreetiliste seisukohtade tutvustus ja analüüsitud materjal pärineb selle projekti raamprojektis osalemisel kogutud andmetest.

LASTEKEELE TÖÖTUBA / WORKSHOP ON CHILD LANGUAGE

The distribution of parts-of-speech in mother-child interaction of Estonian and Lithuanian

Ineta Dabašinskienė (Vytautas Magnus University), Reili Argus (Tallinn University)

Children acquire language system very fast and rather easy due to their sensitivity to lexical and grammatical structures and their distribution (Behrens 2003). Additionally, a stable distribution of linguistic structures in the input could help children to recognize these structures (Bates & MacWhinney 1987). We are aware of the fact that in many languages children start with nouns followed by verbs, therefore the distribution of the main parts-of-speech could signal important information about the syntactic structure of the early child speech. We have to admit that availability of the large modern corpora (Child Language Data Exchange System, CHILDES, MacWhinney 2000) could provide access to explore these questions on the data of many languages, however, the studies in this area are very limited (see Behrens, 2003 for German, Li and Fang 2011 for English, Argus, Kõrgesaar 2014 for Estonian).

The primary aim of this study is to investigate the parts-of-speech frequency distribution of children and their input in two typologically different languages, Estonian and Lithuanian. The analysis focuses on the parts-of-speech, and includes nouns, verbs, adjectives, adverbs, and pronouns, as these are the main elements constructing the syntactic relations of the utterance. The Estonian and Lithuanian corpora have been studied and included data of two pairs of different sex children. Both corpora contain mainly dialogues between mothers and children. The participants were engaged in normal every day activities with their mothers. The main findings include the analysis from the data of approximately one year (from the age of 1;7 to 2;8) and demonstrate an interesting results: the high degree of stability in distribution of the parts-of-speech in the mother' data and in the children's later data. The amount of nouns decreases moderately during the later period of observation while similar tendency can not be observed in the input.

References

- Argus, Reili; Kõrgesaar Helen 2014. Sõnaliigid eesti lapse kõnes ja lapsele suunatud kõnes. – Eesti Rakenduslingvistika Ühingu Aastaraamat, 10, 37–53.
- Bates, E.; MacWhinney, B. (1987). Language universals, individual variation, and the competition model. In MacWhinney, B. (Ed.), Mechanisms of language acquisition. Hillsdale, NJ: Erlbaum
- Behrens, Heike 2006. The input-output relationship in first language acquisition. *Language and Cognitive Processes*, 21 (1/2/3), 2–24.
- Li, Hanhong; Fang, Alex Chengyu 2011. Word frequency of the CHILDES corpus: Another perspective of child language features. – The International Computer Archive of Modern and Medieval English (ICAME) Journal 35, 95–116.
- MacWhinney, Brian 2000. The CHILDES project: tools for analyzing talk. 3rd edition. Mahwah, NJ: Laurence Erlbaum Associates.

Acquisition of Russian noun declension system in 5-year-age children with language impairment

Ingrida Balčiūnienė (Vytautas Magnus University), Anastasya N. Radaeva (Saint-Petersburg State Pediatric Medical University)

An acquisition of the morphological system of the native language is always a long, multilevel, and multi-stage process that causes difficulties (depending on the typological and genealogical features of the given language) not only for language-impaired but also for typically-developing (Voeikova, Dressler 2002, Stephany, Voeikova 2009) children. A vast number of longitudinal studies in Russian language acquisition have highlighted the main difficulties in typically-developing children at the first stages of morphology acquisition (Gvozdev 1949, Dobrova 2003, Voeikova 2011, Gagarina 2008, Ceytlin 2000, 2009, 2012); however, we still lack experimental studies (especially in language-impaired children) which would shed light on the morphology acquisition during the later stages of language development. Hence, the presentation deals with typically-developing (TD) vs. primarily language-impaired (PLI) Russian children's skills in noun declination. We aimed at developing and piloting a new diagnostic tool for the acquisition of Russian noun declension system – The Test for Russian Noun Declination (Balčiūnienė, Radaeva 2016). The test was developed on the theoretical basis of the Natural morphology (Dressler, Karpf 1995) and allows us to assess not only the acquired skills of noun declination but also so-called 'zone of proximal development' (Vygotsky 1962). The structure of the test is based on the elicited (but still autonomous) declination of eight animate nouns (four words and four non-words) which represent different declension types.

The test was piloted in a sample of Russian 5-years-age children: an experimental group consisted of 20 PLI children, while a control group consisted of 20 TD peers. Each subject was tested individually; the procedure was audio-recorded; then, individual scores were estimated and submitted for statistical analysis.

Results of the experiment evidenced statistically significant ($p \leq 0.05$) differences between the groups in a declination of the most difficult (i.e. non-transparent and unmarked) non-word Pyn'. Also, the PLI children struggled with selecting the proper inflections for the medium difficult (i.e. transparent and rather marked) non-words Ponk and Vunalo but only in the genitive and dative case forms ($p \leq 0.05$). And neither TD or PLI children had difficulties with a declination of the simplest (i.e. transparent and marked) non-word Vanaka. Additionally, an error analysis revealed statistically significant ($p \leq 0.05$) differences between the groups in the distribution of error types. E.g., the TD children used more incorrect inflections in comparison to the PLI peers; and the PLI children tended to omit the inflection or to repeat the given non-word in the nominative case form.

Oral narrative comprehension skills in Russian 7-years-age children with language impairment

Ingrida Balčiūnienė (Vytautas Magnus University), Elena V. Zeman (Saint-Petersburg State Pediatric Medical University, St. Petersburg)

Introduction. At the stage of transition from a preschool to a school age, children already should have mastered sufficient skills for both narrative comprehension and production. During school activities, a vast number of narratives are usually employed, and, thus, a child who struggles with narrative comprehension/production falls into the at-risk group.

Narrative comprehension/production ability requires complex cognitive and linguistic skills (Kornev, Balčiūnienė 2017) and, thus, even the mild intelligence/language impairment might lead to dramatic decrease in the learning outcomes. Unfortunately, in Russia, a considerable number of children with the mild primary language impairment (literally, ‘general underdevelopment of speech, the third level’) attend mass schools and do not receive a speech/language therapy; therefore, their difficulties in narrative comprehension presumably have a negative impact on their learning outcomes. However, narrative comprehension at the school age still lacks experimental and large-sample studies.

Methodology. The purpose of our study was to assess the comprehension of the oral narrative by Russian primary school age children. The experimental group consisted of 7-years-age primarily language-impaired (PLI) children (N=20); the control group consisted of typically-developing (TD) peers (N=20). To assess the comprehension of oral narrative, the Russian version (Kornev, Levkovskaya 2016) of the Test for Listening Text Comprehension (Levorato, Roch 2007) was employed. The methodology allowed us to evaluate the children’s ability: a) to extract and to memorize the basic information from the narrative text, b) to draw conclusions based on the narrative, and c) to understand explicit, figurative, and implicit meanings of the sentences.

Results. As it was expected, the PLI children were scored lower than the TD peers in all the measures of narrative comprehension. However, statistical analysis confirmed only one significant ($p = 0.02$) difference between the groups, i.e. a comprehension of the figurative and implicit meaning of the sentences. The results will serve as a basis for methodological guidelines for narrative intervention in children with language impairment at the primary school age.

References

Kornev A.N., Balčiūnienė I. 2017. Fictional narrative as a window to discourse development: A psycholinguistic approach. In: Badio J. (Ed.) Events and Narratives in Language, 171-188. Peter Lang.

Levkovskaya A.N. 2016. Narusheniya ponimaniya rechi u detej s pervichnymi ekspressivnymi narusheiyami rechi [Speech comprehension impairments in children with primary expressive language impairment]. St. Petersburg: SPbGPMU.

Levorato M.C., Roch M. 2007. Valutare la comprensione del testo orale: Il TOR 3-8 [Evaluation of listening comprehension: The TOR 3-8]. Florence: Organizzazioni Speciali.

The acquisition of politeness strategies in Lithuanian

Viktoria Kavaliauskaite (Vytautas Magnus University, Lithuania)

Acquisition of politeness is an integral part of communicative competence as it reveals when children are able to reach their communicative goals effectively by differentiating linguistic expressions according to situation and interlocutor they are speaking to. In this paper expressions and development of linguistic politeness in Lithuanian are analyzed. So far, the acquisition of politeness in Lithuanian have been studied only little, thus the present investigation contributes to politeness studies.

The main aim of this paper is to discuss the results of a pilot study while applying role-play method to elicit expressions of politeness in the speech of 4–7 years old Lithuanian children. A small group of children were asked to play with each other in the kindergarten setting according to given social situation, which varied according to several variables: social status and social distance of the interlocutors. The role-play contained well known social situations, such as a visit to the doctor, café, shop, school, family's everyday life and birthday celebration. Children's conversations were recorded and transcribed. The utterances were coded for politeness markers, e.g., politeness phrases, direct and indirect requests, mitigating devices, additionally, other remarks were included. In all the investigated age groups stylistic variation was observed. The main finding show that children used an informal register when negotiated game plot, whereas a formal register with more politeness markers was introduced when acting in the specific role. The results are discussed and analyzed in relation to children's age.

Comparative analysis of early acquisition of noun word-formation in Estonian and Russian

Viktorya Kazakovskaya (Russian Academy of Sciences), Reili Argus (Tallinn University)

The report aims to study the emergence and early development of noun affixation and compounding in two morphemically rich but typologically different languages, Estonian (more a ‘compound language’) and Russian (more an ‘affix language’). The general goals are to compare the early-stage development features of both word-formation systems (Argus & Kazakovskaya 2013, 2018). Our main research question is whether the typological character of the languages influences the acquisition of word-formation, or whether this process depends only on the morphological features of each of the methods mentioned.

The results are based on corpus analysis of recorded spontaneous child speech (CS) of four typically developing Estonian-speaking and Russian-speaking monolingual children (1;5–3;0) who are from middle SES families (CHILDES). The recording length is 57 hours, which covers 11958 noun tokens. The longitudinal data of CS was analysed according to the first occurrence of any kind of derivatives, their word formation characteristics (viz. belonging to a definite pattern, semantics of affixes etc.), diversity (lemmas), productivity (types), frequency (tokens), productivity and morphosemantic transparency.

Comparing the results obtained, one can state that: a) in Estonian, compounding is an earlier process, whereas in Russian, affixation comes earlier; b) in Estonian, compounding becomes productive more quickly, whilst in Russian, affixation does; c) in both languages within compounding, productive and morphosemantically transparent patterns are acquired earlier than non-productive and opaque ones. Thus, morphological features of both languages emerge more prominently in the acquisition of compounds, whereas the typological nature of the languages under observation reveals more with affixation.

The research of the Russian data was carried out with the financial support of the Russian National Foundation (grant 14-18-03668).

References

- Argus, Reili, Kazakovskaya, Victoria V. 2013. Acquisition of compounds in Estonian and Russian: frequency, productivity, transparency and simplicity effect. *Eesti Rakenduslingvistika Uhingu Aastaraamat* 9, 23–42. DOI: 10.5128/ERYa9.02.
- Argus, Reili, Kazakovskaya, Victoria V. 2018. Acquisition and usage of noun derivations in Estonian and Russian. *Eesti Rakenduslingvistika Uhingu Aastaraamat* (forthcoming)

Derivation of verbs, adjectives and nouns in Finnish child language

Klaus Laalo (University of Tampere)

The paper is continuation to the author's last year's paper in ERY-EAAL 2017 and presents a more detailed analysis of the derivation types in Finnish child language.

The most important derivation suffixes of verbs, nouns and adjectives found in the recorded child language data are presented, both child speech (CS) and child-directed speech (CDS) are taken into account. The recorded data is supplemented with diary data including especially a rich variety of neologisms; these are important in evaluating the onset of the productive use of various derivative elements and the productivity of different derivational types.

Among the most productive derivational elements are for VERBS

1. TTA = transitivizer: causative etc. (lentää 'fly' – lennättää 'make to fly', tippua 'fall' – tiputtaa 'drop') - the base can be a verb, a noun (apu 'the help' – auttaa 'to help', pelko 'fear' – pelottaa 'to fear'), an interjection (kaak – kaakattaa 'cackle, cluck', kop – koputtaa 'knock') etc.
2. U(+) = productive already in the first recordings in both CS and CDS; intransitivizer: reflexive, passive
3. ELE/ILE = frequentative for ADJECTIVES -inen = productive already in the first recording of Mari, emergence in the 1;9 recording of Tomi and productive cumulatively counted in the 1;10 recording, clearly productive in the 1;11 recording
-vA a little later, others clearly later for NOUNS - agent (-jA), instrument (-in) etc. derivational types

Eesti keelt emakeelena omandavate 6-aastaste laste sõnavara seosed kasvukeskkonnaga

Andra Kütt (Tallinna Ülikool)

Laiemal pinnal on kasutuspõhisest keeleomandamisteoorias olulised eeskõige suhtluskeskkond ja laps-täiskasvanu suhtlustasand, mis pakub lapsele keeleliseks arenguks vajaliku (Tomasello 2003: 21–23). Sotsiaalmajandusliku staatuse (ing socioeconomic status) arvestamine keelelise arengu juures juhindub arvamusest, et areng on kujundatud erinevate kogemuste kaudu ning kõrgema sotsiaalmajandusliku staatusega peredel on paremad ressursid, et tagada lapsele keskkond, mis keelelist arengut toetab (Rowe, Rodgers 1997). Tihti on sotsiaalmajandusliku staatuse mõju hinnatud vaid ühe komponendi kaudu, milleks on kõige sagedamini ema haridus (vt nt Hart, Risley 1995). Haridustaset on seostatud kõige enam ka köne loomuse ja stiiliga, mis omakorda mõjutab oluliselt lapse sõnavara aregut (Williamns 1999). On leitud, et kõrgema sotsiaalmajandusliku staatusega vanemate lapsed edestavad madalamana sotsiaalmajandusliku staatusega vanemate lapsi mitme keelelise näitaja poolest, sealhulgas näiteks lausetekste pikkus, erinevate sõnaliikide kasutamine ja köne hulk, mis mõjutab nii keele omandamist kui ka keelelisi oskusi esimesse kooliastmesse sisenevamisel (Ginsborg 2006). Keeleüksuste kasutussagedust on täiskasvanud keelekasutajate puhul peetud oluliseks keele igal tasandil. Lähtudes kasutuspõhisest lähenemisest, loovad lapsed oma grammatika fonoloogilis-leksikalsetest jadadest ehk sõna(vormi)dest, mida nad kuulevad, mitte ei analüüsidi sisendkeelt abstraktsetest sünnipärastest kategooriatest lähtuvalt (Lieven 2010: 2547). Eesti keele kohta ei leidu palju andmeid selle kohta, kuidas on seotud erinev kasvukeskkond ja keeleline areng (vt nt Kütt 2015). Samuti ei ole palju uuritud eesti keele sõnaliigilist jaotumist ja esinemissagedust igapäevavestlustes (vt nt Argus, Kõrgesaar 2014).

Uurimuse eesmärk on vaadelda, kuidas suhtestub spontaanses igapäevavestluses lapse ja täiskasvanu vahel lapse köne sõnavara suurus ja sõnaliigiline jaotumine lapsevanema köne vastavate näitajatega. Üldjoontes kasutuspõhise lähenemise analüüsimalterjaliks on nelja kõrgharidusega ema pooleteisttunnised lindistused (kokku 6 tundi keelematerjali).

Uurimus töi välja, et lapse köne peegeldab tugevasti lapsele suunatud könet mitmel keeletasandil ka ühtsel alusel valitud (ema haridustaseme järgi) valimis. Lapsele suunatud keel aitab lapsel omandada ka peamiste sõnaliikide esinemissageduse, mida näitab ilmekalt nimisõnade ja tegusõnade esinemise osakaal ja selle otsene seos lapsele suunatud keelega.

TEISE KEELE (L2) TÖÖTUBA / WORKSHOP ON SLA

Digging deeper in complexity – SLA results based on the DEMfad-model

Maisa Martin (University of Jyväskylä), Annekatrin Kaivapalu (University of Turku)

Second language development is often described by three dimensions: Complexity, Accuracy and Fluency (CAF). This has improved the comparability of studies conducted in different languages and learning contexts. There are, however, many problems and discrepancies in defining and measuring the three dimensions. Complexity, in particular, has turned out to develop along different lines in structurally different languages: the traditional measures do not always capture the essential features of development.

At the start of a research project called Cefling in 2007 (later followed by Topling and other projects, see: <https://www.jyu.fi/hytk/fi/laitokset/kivi/tutkimus/hankkeet/paattyneet-tutkimushankkeet/topling/en>) a model called DEMfad, designed for looking at written data was developed (see Author 1 et al. 2010 in *Eurosla Monograph Series 1*, 2010). Each Domain is described from Emergence to Mastery along three trajectories: frequency, accuracy and distribution. There is a loose connection to CAF but for written data, instead of fluency, the frequency of the occurrences of the given domain per 1000 words was deemed to be a better indicator of development. We also did not want to limit ourselves to traditional measures of complexity and chose to name the third dimension distribution. This has allowed researchers using the model to concentrate on different definitions and views of complexity. To date the DEMfad model has been used in numerous studies of three corpora of learner Finnish and one of learner Estonian. It has also been applied to the study of cross-linguistic influence.

In our presentation we will discuss some findings which might help in further refining the construct of complexity. The strong focus on subordinate clauses in traditional --measures, for instance, has been questioned as it does not seem to differentiate adjacent CEFR levels in the target languages. Other findings are due to using construction as a unit of analysis. In addition to traditional statistical measures of complexity, the growth in this area has been described by qualitative methods, particularly tracing the number and variety of lexical items in the given constructions. This has also raised a question about a fourth dimension, or a subcategory of complexity, which might be called abstractness. This refers to the growth in the use of constructions in increasingly non-concrete ways. Our presentation will contain some examples of such development, together with a discussion on tentative pros and cons of postulating this dimension.

Eesti keele kui teise keele kommunikatiivset arengut saatev leksikaalgrammatiline areng B1- ja B2-tasemel

Mare Kitsnik (Tallinna Ülikool)

Räägin oma ettekandes eesti keele kui teise keele kommunikatiivse oskuse arenguga kaasaskäivist leksikaalgrammatilise arengu protsessist. Lähemalt käsitlen Euroopa keeleõppe raamdokumendi B1- ja B2-taset eristavate tingiva köneviisi ja modaalverb võima konstruktsioone. Näitan, kuidas muutub nende esinemissagedus, morfoloogiline, süntaktiline, leksikaalne ja funktsionaalne jaotuvus ning täpsus tasemetel vahel. Olen oma uurimustes lähtunud uudsest teise keele arengu põhimõttest: keelekonstruktsoonid on vormi-tähenduse-suhtlusfunktsiooni terviküksused, mis arenevad autentse suhluse käigus. Uurimistulemustel on rakenduslik väärthus keeleõppemetoodika jaoks. Et keelekonstruktsoonid saaksid õppija keeles areneda, peab õppija saama kuulda-näha palju autentset ja endale huvitavat keelekasutust ja tal peab olema võimalus ise turvalises õhkkonnas ning motiveerivates situatsioonides keele kasutamist katsetada. Õpetajad saavad õppijaid toetada, kui nad tunnevad õppijakeele arengu seaduspärasusi ning teavad, milline keelekasutus on mis tasemele iseloomulik. Õppijate keelelises arengus tuleb värtustada lisaks täpsusele ka keerukust ja sujuvust. Vigu tuleb parandada ettevaatlikult, sest kui õppija jõuab konstruktsooni iseseisva produtseerimise faasi, toob see alguses kaasa ebatäpsuste kasvu, mida ei tohi liigselt takistada.

LAK-õpe vene õppekeelega põhikooli järgses gümnaasiumis: keele- ja aineõpetajate koostöö koolijuhtide ja õpetajate vaatevinklist

Eduard Odinets (Tallinna Ülikool)

Lõimitud aine- ja keeleõpe (LAK-õpe) kui mitme erineva õppekorralduslahenduse katusmõiste toetab ühelt poolt eesti keele kui teise keele omandamist ja teiselt poolt eestikeelset aineõpet. LAK-õpe on sätestatud õppetöö korraldamise võimalusena gümnaasiumi riiklikus õppekavas. Vene õppekeelega põhikoolide järgsete gümnaasiumide üleminekul (osalisele) eestikeelsele aineõppele on eeldatud LAK-õppe kasutamist.

LAK-õppe üks tunnusjooni on keele- ja aineõpetajate koostöö, mille õnnestumine on lõimitud aine- ja keeleõppe edukuse üheks eelduseks. Vene õppekeelega koolides korraldatava eestikeelse õppe varasemad uuringud (Metslang jt 2013; Kello jt 2011; Ugur, Raudvassar 2011; Masso, Kello 2010) on näidanud, et koostöö lõimitud aine- ja keeleõppe raames on vähene, juhusliku iseloomuga, koostöövõimalusi ei kasutata piisavalt ning uurijate tehtud soovitustega ei arvestata.

Ettekande esimene osa annab ülevaate keele- ja aineõpetajate koostöö olemusest LAK-õppes, selle korraldamisest ja arendamisest. Samuti ülevaate mõningatest Eesti (Tartu Ülikool, Tallinna Ülikool) ja teiste riikide uuringutest õpetajate koostööst, selle õnnestumise eeldustest ja ebaõnnestumise põhjustest (Davison 2006, Kong 2014).

Ettekande teises osas esitan enda läbiviidud 22 intervjuu tulemusi. Intervjuud on tehtud koolijuhtide ja õpetajatega neis koolides, kus rakendatakse LAK-õpet. Ettekandes antav ülevaade keskendub koolijuhtide ja õpetajate arusaamale koostööst LAK-õppe raames ning väljakutsetele koostöö arendamisel. Intervjuude tulemused selgitavad olukorda vene õppekeelega põhikooli järgses gümnaasiumis ning pakuvad vajalikke samme koostöö korraldamisel. Samuti on võimalik võrrelda varasemate uuringute ja läbiviidud intervjuude järeldused.

Eesti keele kui teise keele innovatiivse kursuse “Juhan” metoodilised ja lingvistilised põhimõtted ning osalejate tagasiside

Mare Kitsnik (Tallinna Ülikool)

Käsitlen oma ettekandes täiskasvanute eesti keele kui teise keele uudse suhtluskursuse “Juhan” (autorid Aleksei Razin ja Mare Kitsnik, Game Club) sisu ja läbiviimise põhimõtteid. Lähemalt on vaatluse all mitteformaalne õpikeskkond, õpperühma psühholoogilise õhkkonna kujundamine, läbiv mängulisuse printsipi, improvisatsiooni suur osakaal, keelelise keerukuse ja sujuvuse arendamine, keelevigade tegemise lubamine jm. Lisaks analüüsini kursuse läbinud õppurite tagasisidet kursuse kohta.

Vague language in English L2: a focus on Lithuanian learner English

Jurate Ruzaitė (Vytautas Magnus University, Lithuania)

The present paper aims to account for the use of vague language (VL) in argumentative essays written by advanced Lithuanian learners of English. The study focuses on two main categories of VL: general extenders (GEs), e.g. and so on, etc. /etcetera, and or so, and vague quantifiers, e.g. some, many, a lot of, and a little. So far, research on Lithuanian learner English has mostly focused on linking words, writer positioning, hedging, and multi-word clusters, and VL has not been addressed yet. However, recent research on VL has shown that it constitutes an important part of pragmatic language competence and thus should be addressed in language teaching in a systematic way.

The present research, first of all, aims to assess how extensively Lithuanian learners of English use VL in comparison to native speakers. Second, to explain why there are differences in the frequency of VL, quantitative corpus data is interpreted in view of the main functions that VL performs in non-native speakers' essays. Finally, the frequency of VL is interpreted with regard to the formality of the items under investigation.

This study is based on the Lithuanian component of the International Corpus of Learner English (LICLE; 244,746 words in total), which contains argumentative essays written by Lithuanian undergraduate students majoring in English Philology. Learner data is compared to native speaker language by resorting to the sub-corpus of texts for the discipline "English" in the British Academic Written English corpus (BAWE-English; 458,780 words in total). The AntConc software was applied to process the data.

The findings indicate that there are some major differences in the frequency of VL in English L1 and L2, both VL categories being more frequent in English L2. The most striking differences appear in the use of vague quantifiers. The overall frequency of quantifiers in English L2 is five times as high as that in L1 (89 occurrences vs. 18 occurrences respectively; f/10,000). All quantifiers under investigation are more frequent in English L2, and some is the most frequent item of all of them (30 occurrences). Interestingly, four out of top five quantifiers in English L2 are those referring to large quantities: much, many, a lot of, and majority. In contrast, mitigating quantifiers, which refer to small quantities, e.g. a bit and a little, are very rare. The overall frequency of GEs in both databases is considerably lower than that of vague quantifiers, but in English L2 they are significantly more frequent than in L1 (10 occurrences vs. 1 occurrence respectively).

In general, the results point to the overuse of VL in Lithuanian learner English, which suggests that Lithuanian learners may lack competence in register differentiation. Another possible interpretation of the differences observed in this study is that English L2 speakers perhaps have a different argumentation style.

Learner English versus English as a lingua franca – a theoretical comparison and implications for a corpus-based comparison of spoken language

Merli Kirsimäe (University of Tartu)

Research into English as a lingua franca (ELF) has been carried out for already a few decades, which has helped develop the paradigm into an established field of research. Nevertheless, amongst all the analyses of language use, descriptions of possibly emerging features, and theoretical discussion, the debate over the difference between learner (or inter-) language and ELF has not really been settled. Nor have the two been compared, to my knowledge, using empirical linguistic data. The lack of such analysis and thus, the possibility of it as a future direction for research was also outlined in Mauranen, Carey & Ranta (2015: 412). Gilquin (2015: 119) also referred to the possibility of including ELF in a wider framework which would allow us to draw comparisons between learner English(es), New Englishes and other varieties of English.

The aim of the research project that I will introduce in my presentation is to compile and then compare two corpora of spoken language – a learner language and an ELF corpus. Since data collection is still ongoing, the presentation will have a more theoretical focus on the similarities and differences between learner language and ELF, with hopefully a few comparative examples from the data thus far gathered. The linguistic aspect that the research project will concentrate on is lexicogrammatical features, more specifically, the use of prepositions, verbs of high semantic generality (e.g. make, take, get), and the existential there is/are/was/were.

So far, ELF researchers have argued that the difference between ELF and learner language fundamentally lies in context – learner language being the language use that occurs in a teaching/learning setting and ELF occurring in any real-life situation where English is the most convenient communication tool at hand for people who do not share a native language (e.g. Mauranen 2011). Historically, learner language research has also been more concerned with errors, whereas ELF research has from the very beginning tried to avoid labelling anything as an ‘error’ but rather as a ‘feature’ or ‘innovation’ (Cogo & Dewey 2012).

Whether the difference in context between learner English and ELF also manifests in the use of language in any way, is still unclear. To this end, the current research project hopes to benefit from corpus research. With both corpora, the compilation process is still at an early stage – the Estonian Spoken Corpus of English as a Lingua Franca (ESCELF) currently consists of approximately 20,000 words of transcribed speech with more recordings done and underway. The Tartu Corpus of Estonian Learner English (TCELE) currently only features written text, but some recordings of spoken language have recently been conducted as well and more have been planned. While the outcome of the analysis is very difficult to predict, I do expect there to be considerable overlaps in the features in learner language and ELF, which is something that Mauranen (2011) also predicted.

The Genitive Alternation in Estonian ESL Learners

Jane Klavan (University of Tartu)

It can be said that one of the central goals of learner language research is to “uncover factors of ‘foreign-soundingness’” (Granger 1996: 43). In Granger’s opinion, over- or underuse make up the “foreign-soundingness even in the absence of downright errors” (2004: 132). Our aim is to pinpoint the over- or underuse of the genitive constructions in Estonian learner English. By identifying any systematic distributional differences between the Estonian ESL learners and native speakers, we are able to zoom in on potentially problematic areas.

The paper takes a look at the genitive alternation. The focus is on the parallel use of the s-genitive (example 1) and the of-genitive (example 2):

- (1) s-genitive: The squirrel’sPOSSESSOR nestPOSSESSED
- (2) of-genitive: The nestPOSSESSED of the squirrelPOSSESSOR

Countless previous studies on the genitive alternation carried out with native speaker data have showed that the alternation between the two genitive constructions is not free. The studies have identified a number of semantic and morpho-syntactic factors that influence the choice between the two constructions. There have also been a few studies looking at the genitive alternation in non-native English. The non-nativelike tendency to overuse the s-genitive has been referred to as a marker of fossilization (e.g. Olsen 1999). A more recent study by Gries & Wulff (2013) fitted a logistic regression model to 2,986 attestations of the genitive alternation in Chinese and German ESL learners and the ICE-GB that were coded for 12 factors. They found that learners rely heavily on processing-related factors and that learners’ choices are differentially modulated by their L1.

In order to test whether animacy and weight, the two prime suspects that have been found to drive native speakers’ choice between the two constructions, have the same effect on the choices made by Estonian learners of English we adopted an experimental design developed by Rosenbach (2005). Rosenbach (2005) carried out a forced choice task, where subjects (39 monolingual native speakers of American English) were presented with short text passages (41 in total) adapted from authentic crime fiction novels that elicited contexts for genitive constructions. Apart from weight and animacy other factors were controlled for. The task of the subjects is to choose as spontaneously as possible to use either the s-genitive or the of-genitive in the given contexts.

Rosenbach’s (2005) results with show that animacy and weight are independent factors and that animacy is potentially more important than weight. By conducting a similar study with Estonian learners of English we aim to compare whether the distribution of the results is similar to that found in Rosenbach (2005). We are currently in the final stages of gathering the data (~40-50 participants). Differently from Rosenbach (2005) we will use mixed-effects logistic regression modelling to analyse the results.

Effects of Transparency in Estonian EFL Learners' Idiom Comprehension

Rita Anita Forssten (Tallinn University)

Transparency is a tool with the help of which the figurative meaning of an idiom can be derived from its literal meaning. The greater the transparency is, the easier it should be to infer the figurative meaning of the idiom. L2 learners tend to resort to this tool more frequently than L1 speakers (e.g. Abel, 2003). However, not all idioms are transparent: idioms constitute a continuum of transparency with, at one end, almost perfect lucidity and, at the other end, complete opacity. Furthermore, one factor that contributes to idiom comprehensibility is imagery. For example, Steinel, Hulstijn and Steinel (2007) have indicated that imagery, which is related to idiom decomposability, seems to be particularly essential in L2 idiom processing.

Since the research on Estonian EFL learners' idiom comprehension is scarce, this presentation introduces an attempt to investigate Estonian native speakers' ability to comprehend English idioms and the effect of idiom transparency in that process. With that goal in mind, an English idiom comprehension test was given to a group of 41 Estonian sixth graders. The test consists of 21 idioms categorized as transparent, semitransparent or opaque. Whereas the transparent idioms (like tell tales out of school) are phrases in which at least one element preserves its literal meaning and the semitransparent idioms (like a whale of a) phrases that contain a crystal-clear metaphor (or are otherwise clearly imageable), the opaque idioms (like between wind and water) do not contain elements that would facilitate understanding.

To bring clarity to EFL idiom teaching, the main objective of the study was to find out whether the assumed order of idiom intelligibility (1. transparent, 2. semitransparent and 3. opaque) is realised among Estonian learners of English. The results suggest that there is a clear difference between the comprehensibility of transparent and opaque idioms –in favour of the former. Moreover, the intelligibility of semitransparent idioms seems to position itself nicely between that of the two other idiom categories, even though to somewhat closer to opaque idioms than transparent ones.

References

- Abel, B. (2003). English idioms in the first language and second language lexicon: A dual representation approach. *Second Language Research*, 19(4), 329-358.
- Steinel, M. P., Hulstijn, J. H., & Steinel, W. (2007). Second Language Idiom Learning in a Paired- Associate paradigm: Effects of Direction of Learning, Direction of Testing, Idiom Imageability, and Idiom Transparency. *Studies in Second Language Acquisition*, 29(3), 449-484.

SOTSIOLINGVISTIKA TÖÖTUBA / WORKSHOP ON SOCIOLINGUISTICS

The Imitation Game: the effect of linguistic normativity in membership categorization

Daria Bahtina (Finnish Centre of Excellence in Research on Intersubjectivity in Interaction), Anna Branets (Tallinn University)

Categorization is a primary cognitive mechanism that supports our understanding of the world around us, from objects to cultural constructs such as social groups (Tomasello & Call 1997).

While social scientists agree that the process of categorization of self and others as members of social groups is fundamental to the construction of identity, there are competing interpretations of its nature, from fixed essentialist approaches to fluid constructivist. One solution to approach the question without preconceptions is to talk about identity as skilled practical accomplishments in ordinary social actions, or categories of practice (for instance, Barth 1969; Bourdieu 1977; Brubaker & Cooper 2000). The Imitation Game – an experimental paradigm to investigate how people distinguish members of their own social group from pretenders (Collins et al. 2017) – allows to tease out the performative displays of identities as they exist in reality. The overarching aim of the experiment is to advance our understanding of the complex processes of social exclusion and inclusion. In this particular paper we focus on linguistic normativity as one of the fundamental mechanisms used for categorization. Different types of linguistic proficiency will be compared and discussed in relation to their reliability for group membership categorization. The degree of socialization into culturally diverse environments, composed of exposure to multi-ethnic contexts and positive attitudes towards language and culture of a social group in question, has a positive effect both on interactive efficiency (Bahtina-Jantsikene 2013) and on linguistic comprehension (Branets, Bahtina-Jantsikene, & Verschik, to appear). In this paper we explore the implications of these non-linguistic factors on the quality of language-based judgements performed in the experiments.

The Imitation Game is played by members of two distinct social groups that are organized into computer-mediated interactions, the aim of which is either to demonstrate one's actual belonging to a social group (in the role of a true representative), to imitate such belonging (a pretender), and to tell apart participants engaging in the previous two activities (a judge) through a series of questions and answers. All questions, answers, judgements and comments on the process of judging are automatically recorded by the software. In our multiplayer set-up, each role is played by a team of three participants from the same social group, which provides video-taped data of the in-group discussions and negotiations. The presented materials are collected in a course of experiments between ethnic Estonians and well-integrated Russian-speakers from Estonia. The emerging results point to the difference between grammar-perfect and native-like but colloquial language use on perceived belonging to a group. We will also illustrate some challenges of self-identification for members that exercise multiple identities in their everyday life.

Multilingualism and the definition of language

Ad Backus (Tilburg University, Netherlands)

A usage-based approach views language as an emergent dynamic system, grounded in social interaction and affected by cognitive processes. Under this approach, an individual's linguistic knowledge emerges from and is constantly updated by her linguistic experience, and this dynamic competence in turn co-determines further language use. The mind categorizes the language use it experiences through usage and exposure by comparing units that are used with past usage events in terms of similarity and dissimilarity. The resulting linguistic knowledge is organized as a network that enables the representation of both overt ('specific') and abstract ('schematic') linguistic units in a unified system.

When languages are in contact, elements of both languages appear together, often in asymmetrical fashion. The usage-based approach accounts for bilingual and monolingual language use in the same way. It links the occurrence of individual items from any language, whether lexical or grammatical, to their usage frequencies and perceptual salience, and thus to the ease with which they are activated and retrieved during language use even if the language being spoken is 'the other one'. Grammatical interference is further stimulated by perceived similarity of two structures across the contact languages. Under this approach 'structural' similarity is the result of the cognitive comparison carried out by an individual speaker on formal or semantic dimensions, or both. Ultimately, speakers select what is easiest to activate.

The dynamic character of the system presupposes that the representations underlying linguistic items gain in robustness or deteriorate. One determining factor is obviously frequency of use, but also important is salience, which is much more difficult to model owing to its unsystematic nature. Activation is easier for high frequency elements, as recurrence strengthens entrenchment, and entrenchment makes retrieval easy. However, intentional selection of particular units also plays an important role in human communication.

Modifications in an individual's lexicon and grammar may spread as innovations to other speakers and change the community's usage patterns. A usage-based approach holds that patterns of language mixing result from synchronic selection processes, and borrowing is the diachronic effect of these processes. As a result, synchronic mixing and interference, and diachronic lexical and grammatical borrowing, are driven by the same cognitive and social mechanisms.

This paper sketches a usage-based approach to language contact phenomena as a unified theoretical framework capable of integrating existing explanations and of yielding new insights into the nature of language and linguistic competence.

Research of features of russian outside of Russia by means of integrum

Inna Adamson (Tallinn University)

In report are provided results of application of a database system Integrum for a study of features of Russian outside Russia. The goal of a research was identification of the facts and features of Russian of Estonia. The main object of a research is Russian in Estonia particularly lexical features including features of lexical semantics . During the work with Integrum were applied various methods. One of them is a method of formal search in the set parameters. The method of research has been applied also to identification of assessment language material in Latvia. The reason of such choice was also the that a number of the facts are similar to the Russian in Estonia. In this work is provided Information on time of penetration of required units into language, and frequency data.

The use of the english language on the signs in Tallinn old town 2013–2017

Marleen Kedars (Tallinn University)

The linguistic landscape (LL) study has gone through a major development during the last 10 years but it is still exceptional that it has been studied diachronically (Pavlenko 2010). LL is mostly explored in the urban environments where the commercial areas can be found (Gorter, Cenoz 2017) and this is the reason why this study has been conducted in Tallinn Old Town that is a place with a big tourist flow and small tourism businesses.

The main outcome of the research is that the LL in Tallinn Old Town is multilingual,multimodal, also symbolic and informative. A large number of visual and printed texts and also different languages are in use. English as a foreign language is most frequently usedwhich shows that English is used for a fetishized purpose (Kelly-Holmes 2014).

Linguistic fetish refers to the phenomenon of using languages for symbolic (fetishised) rather than utility purposes incommercial texts (Kelly-Holmes 2014: 135). English is often associated with modernity, internationalism, technological advancement (Gorter 2013: 202). English “sounds sexier, cooler and more attractive” and from that comes the term “power of English” (Kasanga 2010: 186–187).

References

- GORTER, Durk. 2013. Annual Review of Applied Linguistics. Cambridge University Press.33, 190–212.
- GORTER, Durk; CENOZ, Jasone. 2017. Linguistic Landscape and Multilingualism. Language Awareness and Multilingualism. Third Edition, 233–245.
- KASANGA, Luanga A. 2010. Streetwise English and French advertising in multilingual DR Congo: symbolism, modernity, and cosmopolitan identity. 206. 181–205.
- KELLY-HOLMES, Helen. 2014. Linguistic fetish: The sociolinguistics of visual multilingualism. Visual Communication, 135–151.
- PAVLENKO, Aneta. (2010. Linguistic landscape of Kyiv, Ukraine: A diachronic study. InShohamy, E., Barni, M. & E. Ben Rafael (eds.) Linguistic landscape in the city. Bristol, UK: Multilingual Matters, 133–150.

Preliminary study of Estonian-Russian bilinguals' colour associations

Olga Loitšenko (Tallinn University)

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, želtyj, sinij, goluboj, koričnevyj, fioletovyj, rozovyj, oranževyj, seryj, birjuzovyj, sirenevyyj, lilovyj, beževyj. The colour terms birjuzovyj and sirenevyyj 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 põhivärvimed. Keel ja Kirjandus, 12, 797–808.
- Paramei, G., Gribers, Y. & Mylonas, D. (2017). An online colour naming experiment in Russian using Munsell colour samples. Color Research and Application, 1–17.

The analysis of different registers of spoken Lithuanian: lexical and grammatical features

Laura Kamandulytė-Merfeldienė, Ingrida Balčiūnienė (Vytautas Magnus University)

The aim of the study is to present lexical and grammatical features typical for different registers of modern Spoken Lithuanian. The study is based on a data of the Corpus of Spoken Lithuanian (<http://donelaitis.vdu.lt/sakyties-kalbos-tekstynas/>) which was developed in the framework of national projects coordinated by Vytautas Magnus University and supported by the Lithuanian State Science and Studies Foundation (grant No. L-12/2008) and by the Research Council of Lithuania (grants No. LIT-9-11, No. LIP-085/2016). Currently, the Corpus consists of 256 conversations (373,073 grammatically annotated word forms) involving 1086 speakers (659 women, 427 men). The age of the speakers range from 3 to 81. The largest group of informants belong to the age range from 19 to 44.

The Corpus data were classified into different registers of spontaneous conversations and prepared conversations. The mostly frequent lemmas, the distribution of parts of speech and the main functions of the sentences in different registers of spoken language (spontaneous private conversations and prepared public conversations) were analysed. The results will be discussed in the presentation.

The Spatial Organization of Language: Creating, and sustaining, intentional language ecologies in Estonia's dual-immersion kindergartens

Kara Brown (University of South Carolina)

How is school space reimagined with the adoption of new language programs and policies? In this presentation, I examine the spatial organization of language in classrooms and across kindergartens in the context of a new national initiative in Estonia to develop two-way immersion (TWI). The TWI policy, rolled out in the 2015-2016 academic year, showcases the ways schools, even at the pre-primary level, can be reimagined as spaces that intentionally bring together populations and expand learning opportunities in new ways. An essential aspect of the TWI policy is the cultivation of a school-based language ecology (Haugen, 1972) conducive to dual-language learning; I label this planned and directed environment, which has a particular genealogy, as the intentional language ecology. But, the TWI policy is also nested within an existing language environment, which largely remains stable across longer periods—the sustained language ecology. In this presentation, I develop both concepts of the intentional and sustained language ecology by drawing on evidence from an ongoing qualitative research project focused on the TWI initiative in Estonia. Drawing particularly on post-colonial theories (Annus, 2018; Millei et al., 2017), I focus on the ways the intentional language ecology points to deliberate efforts to transform kindergarten space with the intent to distribute language-learning opportunities across the Russian- and Estonian-speaking populations more equitably and effectively. I consider, in particular, the ways space is sustained and transformed at the city, school, and classroom levels.

References

- Annus, E. (2018) Soviet postcolonial studies: a view from the Baltic borderlands. London: Routledge.
- Haugen, E. (1972). The ecology of language: Essays by Einar Haugen. Stanford: Stanford University Press.
- Millei, Z., Silova, I. & Piattoeva, N. (2017) Towards decolonizing childhood and knowledge production.
- In Silova, I., Millei, Z. & Piattoeva, N. (eds.). Childhood and schooling in (Post) socialist societies: memories of everyday life. New York: Palgrave Macmillan.

Explaining English-Estonian code-copying: the role of different CMC genes

Anna Verschik, Helin Kask (Tallinn University)

So far, the research on English-Estonian language contacts has been focusing on what gets copied (in the terms of Johanson's code-copying framework) and what factors condition different degrees of copying (global, corresponds to borrowing; selective, corresponds to structural change; mixed, combination thereof). We view our data from a comparative perspective and try to establish whether and how genre of computer-mediated communication (CMC) affects copying. We analyze code-copying in Estonian blogs (275,263 words), Facebook conversations (22,601 words) and vlogs (3 hours) and look whether attributes such as synchronous/asynchronous, monologic/dialogic character, limited/unlimited space influence copying.

The data suggests that vlog has the highest number of copies in general, as it is more spontaneous and less "permanent" unlike written genres (blog, Facebook posts). Whatever the genre is, global copies prevail everywhere and selective copies are more characteristic of vlogs. We will provide an explanation for the prevalence of global copies and try to explain the low number of selective copying, as well as varying share of selective/mixed copies across genres. Apparently, there is a continuum of implicit norms but all of the considered genres are constructed as more or less multilingual, hence, there is no apparent need to avoid explicit other language items. At the same time, a low number of selective copies may have a cognitive explanation (non-lexical impact appears at a later stage of contact-induced language change).

References

- Johanson, Lars 2002. Contact-induced change in a code-copying framework. – *Language change: The interplay of internal, external and extra-linguistic factors*. Eds. M. C. Jones, E. Esch. (*Contributions to the Sociology of Language*, 86.) Berlin: Mouton de Gruyter, pp 285–313.
- Kask, Helin. 2016. English-Estonian copying in Estonian blogs. *Philologia Estonica Tallinnensis* 1, 80-101.
- Verschik, Anna and Kask, Helin. To appear. Estonian-English code alternation in fashion blogs: structure, norms and meaning. In Heiko Marten and Sanita Lazdina (eds.), *Multilingualism in the Baltic States: Societal Discourses, Language Policies and Contact Phenomena*. Palgrave.

Konverentsi toetavad

Eesti Keele Instituut
Haridus- ja Teadusministeerium (Hasartmängumaksu Nõukogu)
Tallinna Ülikool
Eesti Keeleressursside Keskus <https://keeleressursid.ee/et/>
Euroopa regionaalarengu fond



Eesti Keele Instituut



Eesti
tuleviku heaks



HARIDUS- JA
TEADUSMINISTERIUM



Euroopa Liit
Euroopa
Regionaalarengu Fond



TALLINNA ÜLIKOO

