VOLUME 36:2023

FINNISH JOURNAL OF LINGUISTICS

EDITED BY MARIANN BERNHARDT, TIINA EILITTÄ, EMMI LAHTI,
LAURI MARJAMÄKI, OLLI O. SILVENNOINEN

Finnish Journal of Linguistics is published by the Linguistic Association of Finland (one issue per year).

Notes for Contributors

Policy: *Finnish Journal of Linguistics* welcomes unpublished original works from authors of all nationalities and theoretical persuasions. Every manuscript is reviewed by at least two anonymous referees. In addition to full-length articles, the journal also accepts short (3–9 pages) 'squibs' as well as book reviews.

Language of Publication: Contributions should be written in English, French, German, Finnish, or Swedish. If the article is not written in the native language of the author, the language should be checked by a professional in that language.

Style Sheet: *Finnish Journal of Linguistics* follows the Generic Style Rules for Linguistics, with complementary house rules, which can be found at https://journal.fi/finjol/about/submissions.

Abstracts: Abstracts of the published papers are included in *Linguistics Abstracts* and *Cambridge Scientific Abstracts*. The published papers are included in *EBSCO Communication & Mass Media Complete*. *Finnish Journal of Linguistics* is also indexed in the *MLA Bibliography*.

Editors' Addresses (2023):

Mariann Bernhardt, Department of Finnish, Finno-Ugrian and Scandinavian Studies, P. O. Box 3, FI-00014 University of Helsinki, Finland

Tiina Eilittä, Research Unit for Languages and Literature, English, Faculty of Humanities, P. O. Box 1000, FI-90014 University of Oulu, Finland

Emmi Lahti, Department of Finnish, Finno-Ugrian and Scandinavian Studies, P. O. Box 3, FI-00014 University of Helsinki, Finland

Lauri Marjamäki, Department of Languages, FI-00014 University of Helsinki, Finland

Olli O. Silvennoinen, Department of Languages, FI-00014 University of Helsinki, Finland

Editors' E-mail: sky-journal(at)helsinki(dot)fi

Publisher:

The Linguistic Association of Finland
Suomalais-ugrilainen kielentutkimus
Hämeenkatu 1
FI-20014 Turun yliopisto
Finland
http://www.linguistics.fi, http://journal.fi/finjol

The Linguistic Association of Finland was founded in 1977 to promote linguistic research in Finland by offering a forum for the discussion and dissemination of research in linguistics, both in Finland and abroad. Membership is open to anyone interested in linguistics. The membership fee in 2022 was EUR 30 (EUR 15 for students and unemployed members).

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

Finnish Journal of Linguistics
36

Suomen kielitieteellisen yhdistyksen aikakauskirja
Tidskrift för den Språkvetenskapliga föreningen i Finland
Journal of the Linguistic Association of Finland

Editors:

Mariann Bernhardt
Lauri Marjamäki
Tiina Eilittä
Olli O. Silvennoinen
Emmi Lahti

Layout:
Tommi Nieminen

Advisory editorial board:

Werner Abraham
University of Vienna

Arto Anttila
Stanford University

Kimmo Granqvist
Södertörn University

Auli Hakulinen
University of Helsinki

Jaakko Hämeen-Anttila
University of Edinburgh

Martin Haspelmath
Max Planck Institute for Evolutionary Anthropology

Marja-Liisa Helasvuo
University of Turku

Anders Holmberg
Newcastle University

Tuomas Huumo
University of Turku

Juhani Härmä
University of Helsinki

Fred Karlsson
University of Helsinki

Leena Kolehmainen
University of Turku

Meri Larjavaara
Åbo Akademi University

Jaakko Leino
University of Helsinki

Matti Miestamo
University of Helsinki

Urpo Nikanne
Åbo Akademi University

Martti Nyman
University of Turku

Krista Ojutkangas
University of Turku

Mirja Saari
University of Helsinki

Ulla Tuomarla
University of Helsinki

Maria Vilkuna
University of Helsinki

Jussi Ylikoski
University of Turku, Sámi University of Applied Sciences

Jan-Ola Östman
University of Helsinki

The editors wish to acknowledge the invaluable work of the advisory editorial board and to express their gratitude to those leaving the board this year, to those continuing to serve in the board as well as to those now joining the board.

2023

ISSN: 2814-4376

Contents

Reviewers of *Finnish Journal of Linguistics* 36 (2023)

Luobbal Sámmol Sámmol Ánte (University of Oulu), Silvana Deilen (Johannes Gutenberg University Mainz), Valerie Guérin (no affiliation), Juha Kuokkala (University of Helsinki), Olli Kuparinen (Tampere University), Liina Lindström (University of Tartu), Anneli Pajunen (Tampere University), Michael Riessler (University of Eastern Finland), Iker Salaberri Izko (University of Basque Country), Laura Schiffl (Technical University of Munich), Jari Sivonen (University of Oulu), Riitta-Liisa Valijärvi (Uppsala University), Max Wahlström (University of Helsinki)

The editors are most grateful to all the scholars who have acted as reviewers for *Finnish Journal of Linguistics* in 2023, including those who wish to remain anonymous and are therefore not listed here.

Reading Easy Language texts written by public authorities: Evidence from eye tracking

Carina Frondén
University of Helsinki

Johanna K. Kaakinen
University of Turku

Abstract

Previous research has shown that word length, frequency and word repetition influence word reading times (Rayner 1998; 2009). Guidelines for Easy Language advise writers to use frequent and short words, and to repeat words instead of using synonyms. However, some of these guidelines are based on research that has been misinterpreted, simplified, or is outdated (Wengelin 2015), and studies focusing on effects of word length, frequency and word repetition among adult readers in the Easy Swedish target group are lacking. This eye-tracking study investigated the reading of Easy Language texts written by public authorities, as well as the effects of word length, frequency, and word repetition on readers in a day centre for people with intellectual disabilities. The results showed significant effects for word length and frequency in all readers. In addition, the effects were significantly greater in the target group than in the control group. The effects for word repetition were not as clear, affecting only one of the reading measures. Furthermore, the study revealed poor comprehension rates in the target group, i.e., when asked, they were not able to reproduce the main contents of the texts. The significantly greater effects of word length and frequency suggest that the related Easy Language guidelines are valid for this group of readers. The poor comprehension rates indicate that the texts were too difficult for these readers.

Keywords: Easy Language, Easy Swedish, easy-to-read texts, eye tracking, public authority communication, lättläst

1 Introduction

Following democratic values and the aim to increase inclusion, the publication of Easy Language texts has grown considerably over recent decades (e.g., Lindholm & Vanhatalo 2021). As services become increasingly digitized, more

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

Finnish Journal of Linguistics 36 (2023), 7–36
DOI: 10.61197/fjl.126062

and more information and public services are only available online. However, the EU Web Accessibility Directive (2016/2102) encourages providing people with disabilities with better web accessibility, but it neglects aspects of accessible language. Readers in the heterogeneous target group of Easy Language, e.g., people with intellectual disabilities, are especially vulnerable. Despite the variety of needs and challenges in this group of readers, public authorities tend to publish only one Easy Language text version (instead of adapting different texts or materials to meet different needs within the Easy Language target group), hoping it will be suitable for all the readers in this group (e.g., Forsberg 2014).

Easy Language has been defined as an "easy-to-understand variety" (Maaβ 2020: 12), but conceptualizations and definitions have varied (Arle & Frondén 2022).[1] The Swedish equivalent term lättläst has been defined as "broadly controlled natural language, […] a subset of natural languages obtained by restricting the grammar and vocabulary in order to reduce or eliminate ambiguity and complexity" (Heimann Mühlenbock 2013: 22). However, the term usually refers to texts "that are, or should be, comprehensible for struggling readers" (Arle & Frondén 2022). The terms Easy to Read (easy-to-read texts) and Easy-Read have been used to describe the same concept (see e.g., Karreman et al. 2007; Fajardo et al. 2014; Sutherland & Isherwood 2016; Arle & Frondén 2022). In this study, the term Easy Language is used as a translation of the Swedish terms *lättläst* and *lätt språk*, when referring to written language, spoken language, and signed language; and Easy Swedish when specifically referring to the Swedish language.[2] Although Easy Language texts can differ greatly, the usual characteristics of such texts are reduced text complexity on the word and sentence level, a simple and airy layout, and reduced content load (Heimann Mühlenbock 2013; Maaβ 2020; Arle & Frondén 2022). Despite a reduced content load, however, Easy Language texts can still be longer than

[1] Also internationally, the conceptualizations of Easy Language and closely related concepts, such as Plain Language, Accessible Language, Accessible Communication, and Comprehensibility, have differed, (e.g., Maaβ 2020; Moonen 2021; Hansen-Schirra et al. 2021). For a discussion on the conceptualizations, see Arle & Frondén (2022).

[2] The Swedish term *lättläst* and the equivalent Finnish term *selkokieli* have been translated into *Easy Language* in linguistic research (e.g., Bohman 2021; Leskelä 2021; O'Donnell & Ramdén 2021; Arle & Frondén 2022). A comparison of guidelines for English, Swedish and Finnish languages show great similarities, i.e., the same or similar advice (e.g., LL-Center; MTM; Sundin 2007; Lundberg & Reichenberg 2008; IFLA 2010; Österlund 2011; Selkokeskus 2022). The German term *Leichte Sprache* has also been translated into Easy Language (e.g., Hansen-Schirra & Maaß 2020; Pappert & Bock 2020; Schiffl 2020; Borghardt et al. 2021; Hansen-Schirra et al. 2021).

the original texts (e.g., Maaβ 2020).

Sets of guidelines (e.g., LL-Center; MTM) for writing such texts are provided by and for professionals working with Easy Language texts (e.g., Bohman 2021: 544–547; Lindholm & Vanhatalo 2021: 12, 15–16, 18). However, some of these guidelines are based on research that has been misinterpreted, simplified, or outdated (Wengelin 2015; Arle & Frondén 2022). Having scrutinized the psycholinguistic research literature, Wengelin (2015) found, for example, that the use of the passive voice[3] (except for the reversible passive voice) does not in itself cause poorer comprehension. It can, however, result in a more complex sentence structure (Wengelin 2015). Updated information about the validity of the guidelines for Easy Language is thus needed. In word-level guidelines for Easy Swedish (see Table 1) it is stated that writers should use short and frequent words and repeat words instead of using synonyms. This is supported by previous eye-movement research that has shown that longer words and infrequent words attract longer fixations, and that the repetition of words in the same text attracts shorter fixations in both adult and developing readers (e.g., Rayner 1998; 2009). However, studies focusing on such effects on adult readers in the Easy Swedish target group are lacking, and it remains unclear what actually is easy to comprehend for different readers in the Easy Language target group (Sutherland & Isherwood 2016; Arle & Frondén 2022). Despite the heterogeneity of the target group, guidelines for Easy Swedish are often written with the whole group in mind, and not specific subgroups. Furthermore, producers of Easy Swedish texts seldom carry out reception surveys, so we have very little information on the reception of Easy Swedish texts (Domeij & Spetz 2014).

This eye-tracking reading study at a day centre for people with intellectual disabilities in Finland examines the reading of Easy Language texts written by public authorities. It also investigates how word length, frequency, and repetition influence fixations and whether their effects on these readers and non-disabled readers differ. If these effects are significantly greater among readers with disabilities, the results add validity to the related word-level guidelines for Easy Language. Methodologically, eye-tracking measures were combined with open-ended questions posed during the eye-tracking experiment (e.g., Gutermuth 2020). Eye tracking was chosen because it is especially suited to examining reading-related processes as it provides detailed information on the time-course of processing (Rayner 1998).

[3] The guidelines for Easy Swedish often advise writers to avoid the passive voice (e.g., LL-Center).

Table 1. Word-level guidelines for Easy Swedish

Guidelines	References
undvik långa ord / välj korta ord 'avoid long words / choose short words'	LL-Center; MTM; Sundin 2007; Lundberg & Reichenberg 2008.
undvik obekanta ord 'avoid uncommon words'	LL-Center; MTM; Sundin 2007; Lundberg & Reichenberg 2008; Österlund 2011.
repetera ord i stället för att använda synonymer 'repeat words instead of using synonyms'	MTM; Sundin 2007; Lundberg & Reichenberg 2008; Österlund 2011.

The research questions were: 1) How do word length, word frequency and word repetition influence fixation times and how do these effects differ in the target group and the control group? 2) Were the Easy Language texts comprehensible to the participants, i.e., when asked, were they able to reproduce the contents of the texts?

1.1 Easy Swedish: An emerging research field

Research on Easy Language is an internationally growing field, but most linguistic research has focused on the German and English languages (e.g., Chinn & Homeyard 2017; Sutherland & Isherwood 2016; Gutermuth 2020; Hansen-Schirra & Maaß 2020; Pappert & Bock 2020). Although research interest in other languages is now growing (e.g., Lindholm & Vanhatalo 2021), research on Easy Swedish remains scarce, and the focus has mainly been on written language. A meta-narrative review by Arle & Frondén (2022) showed that the conceptualization of Easy Language also differs depending on the studied material, modality, aims, and discipline. Easy Language novels have been studied from the perspectives of reading experience and pedagogical utility (e.g., Nordenstam & Olin-Scheller 2018), whereas Easy Language information texts have been examined from perspectives of accessibility (e.g., Domeij & Spetz 2014; Forsberg 2014). A few descriptive articles provide some historical and contemporary practical information on Easy Language in Sweden (Bohman 2021; O'Donnell & Ramdén 2021). Heimann Mühlenbock (2013), based on a comparison of Easy Language texts in the LäSBarT corpus of different genres with other texts, created the *SVIFT* text complexity model.

In addition to surface-level features, the model also includes vocabulary load, sentence structure, idea density, and human interest. Heimann Mühlenbock's study showed that mean word length and lemma variation index was highly relevant when separating Easy Language texts from other texts (Heimann Mühlenbock 2013: 150–151). However, the need for more research on Easy Language and the reception of Easy Language texts has been stressed both internationally and in the context of the Swedish language (e.g., Sutherland & Isherwood 2016: 297, 307–308; Moonen 2021: 393; Arle & Frondén 2022).

1.2 Effects of word length, frequency, and repetition on word reading

The word-level variables chosen for this study are based on Easy Language writing advice that recommend short words or advise against long words (LL-Center; MTM; Sundin 2007: 132; Österlund 2011: 11–12). Writers are also advised to use "common words" (Sundin 2007: 132; Lundberg & Reichenberg 2008: 64; Österlund 2011: 11) or to avoid "uncommon words" (LL-Center). The guidelines also include recommendations to repeat words instead of using synonyms (MTM; Sundin 2007: 124–125, 147; Lundberg & Reichenberg 2008: 74–75; Österlund 2011: 15).

Previous research demonstrates that word length and frequency influence the duration of the reader's gaze on words (Rayner 1998; 2009). Infrequent words are read with longer gaze duration than words of high frequency, even when other factors are controlled for (e.g., Inhoff & Rayner 1986; Rayner & Duffy 1986). Other studies have shown that frequency affects word recognition (Forster & Chambers 1973; Balota & Chumbley 1984; Schilling et al. 1998; Hyönä & Kaakinen 2019). Hyönä & Olson (1995) showed that both non-dyslexic and dyslexic readers have longer fixations and more regressions when reading long, low-frequency words. However, they found no significant difference between these groups in terms of the effects of word length and frequency. Similarly, a study of Easy German by Schiffl (2020; 2021) that examined effects of word frequency, word length and repetition among adult German readers with cognitive impairments and readers without impairments, found no significant difference between these groups. Joseph et al. (2009) presented stronger word length effects in gaze duration and refixation probability among children than adults. These differences were the greatest in refixation behaviour. In contrast, Tiffin-Richards & Schroeder (2015) found generally greater effects for word length and frequency for children than for adults.

Several studies have shown that the repetition of texts and words influence eye movements in reading (Rayner 1998). In a study by Hyönä & Niemi (1990), rereading a text resulted in decreased fixation duration. A similar study by Inhoff et al. (1993) showed that fixation durations decreased when passages of text were reread. Raney & Rayner (1995) demonstrated that fixation times on both high- and low-frequency words decreased when the words were encountered several times within a passage. However, this effect was more prominent for low-frequency words. All the above-mentioned studies used unimpaired readers. A study by Fajardo et al. (2014) of comprehension of Easy Language texts among students with intellectual disability showed that word length and frequency had no effect on comprehension, but that the number of coreferences (including repetition of the same noun, repetition of the word stem, and repetition of shared word steam allowing word category variation) had a significant effect on literal comprehension. Their analysis was based on reading comprehension questions to measure comprehension on the literal and inferential levels (for a further discussion of the different levels of comprehension, see e.g., Kintsch & Dijk 1978) and did not include eye-tracking measures.

1.3 Eye movements among adult readers with intellectual disabilities

Compared to good readers, disabled readers show longer average fixation durations, make more regressions, read more slowly, and make shorter saccades (Rayner 1983). Reichle et al. (2013) presented a review of studies on the eye movements of readers of different skills or ages, showing differences between skilled and non-skilled readers in fixation times, reading times, saccade length, and regressions. Non-skilled readers also tend to skip words less frequently (Joseph & Blythe 2011: 9). The perceptual span of non-skilled readers is smaller than that of proficient readers (Hyönä & Kaakinen 2019: 240; Schiffl 2021). In a study by Schiffl (2021), readers with cognitive impairments displayed reading patterns similar to those of children, i.e., longer reading times, and more numerous and shorter saccades. According to Rayner (1983), different types of dyslexia display different eye movement behaviour during reading (Rayner 1983: 167–171).

Very few studies have studied reading of Easy Language texts using eye-tracking methodology. A study by Gutermuth (2020) compared comprehension of Easy German texts, Plain Language texts, and standard language texts with different readers in the Easy Language target group, showing longer

reading times for more complex texts. Some practical reflections on conducting eye-tracking studies in the Easy Language target group has been provided: Borghardt et al. (2021) present background information and recommendations for metadata and test battery, Deilen & Schiffl (2020) provide practical advice concerning difficulties related to this specific group of readers. Like Gutermuth (2020), Deilen (2021) and Schiffl (2020; 2021) have conducted eye-tracking experiments using German-speaking readers from the Easy Language target group. Deilen (2021) focuses on compound words; Schiffl (2020; 2021) in turn examines the effects of word length, frequency and repetition, and the long-term learning effects of infrequent words.

2 Method

2.1 Participants

Eleven participants (8 women and 3 men) were recruited from a day centre for persons with intellectual disabilities (target group). As the original purpose of this study was to examine whether this type of experiment using longer texts could be conducted with this group of readers (pilot study), only a small group was recruited. All were native speakers of Swedish, living in Finland. The inclusion criteria for this group were affiliation to this specific type of institution and age between 18 and 65. No diagnosis information was collected. Due to calibration failure, eye-tracking data were only gathered from seven of the participants. The control group consisted of eight university students (6 women and 2 men), and all were native speakers of Swedish and aged between 18 and 35.

The language proficiency measure (§ 2.4) scores showed variations both between and within the groups (Table 2). This was the case for both the reading comprehension test (§ 2.4) and the decoding test (§ 2.4). One of the participants in the control group had reading difficulties – this information was only provided to the researcher in the middle of the experiment – which explains the significant variation in this group. Because the size of the control group remained smaller than expected, the data on this participant was still included in the analyses.

The study was conducted in accordance with the ethical principles of the Finnish National Board on Research Integrity, TENK (2019). Each participant gave their written consent to participation, and the Ethical Review Board of the University of Helsinki approved the study.

Table 2. Baseline test scores presented in mean (M) and standard deviation (SD) values per group.

	Target group M	Target group SD	Control group M	Control group SD
Text comprehension test*	5.64	3.2	15.13	4.54
Decoding test	43.45	25.08	101.25	10.17

*theoretical max/min=20/0

2.2 Apparatus

Eye movements were recorded monocularly using EyeLink Portable Duo (SR Research, Canada) at 500 Hz sampling frequency. A chin-and-forehead rest installed 58 cm in front of the screen was used to minimise head movements. The stimuli were presented on a 17.3″ Asus ROG G752V laptop screen (refresh rate of 120 Hz, resolution 1920 × 1080).

2.3 Materials

Each participant silently read two informative texts comprising nine pages in total, written in Easy Swedish, on a computer screen (font: Courier New, font size: 20). The texts had been published online by the Finnish Tax Government Authority on the Inkomstregistret website[4]. Both texts were defined at the beginning as an Easy Language (*lättläst*) text. Text A contained general information on the national income register and Text B contained information on proxies for dealing with issues in the income register (Table 3).

The word frequency estimates were retrieved from the LäsBart[5] corpus of The Swedish Language Bank. Lemma frequency ratings were used.

[4] The text materials are available at https://www.vero.fi/sv/inkomstregistret/om-oss/inkomstregistret-lättläst/.

[5] This corpus contains Easy Language texts of different genres, and text from children's fiction, published in Sweden (Heimann Mühlenbock 2013). According to Balota et al. (2004: 494), subtitle corpus frequency estimates outperform the same type of estimates from book corpuses. However, the LäsBart frequencies were compared and found to resemble those of subtitle-based corpuses.

Table 3. OVIX, length, and distribution of word length, frequency, and word repetition in Text A and Text B.

	Text A	Text B
Length in words	328	263
Word length, M	6.54	5.65
Word length, min	1	1
Word length, max	31	25
Frequency, M	3.66	3.77
Frequency, min	0	0
Frequency, max	4.46	4.46
Word repetition, M	6.05	6.51
Word repetition, min	1	1
Word repetition, max	22	22
OVIX	44.5	36

2.4 Language proficiency measures

An adapted version of the *LS Klassdiagnoser Läsförståelse I* reading comprehension test was used to assess the participants' reading comprehension. The LS Klassdiagnoser test is a diagnostic test for qualitative diagnosis of reading and writing disabilities, adapted for a Swedish-speaking context in Finland. As an adaptation of the original reading comprehension test was necessary from both an ethical and a practical perspective, only the first half of the test was used. The same adaptation was used for both groups. Thus, the participant's scores are not comparable to national scores but can nevertheless provide a comparison between groups and participants. To assess decoding while reading, the *LS Klassdiagnoser Avläsning nonsensord* readout test was used. This test is also part of the LS Klassdiagnoser diagnostic test described above.

2.5 Pre-questionnaire

Prior to testing, the participants answered an open-ended pre-questionnaire that tested for possible previous knowledge of the content of the texts (Appendix B).

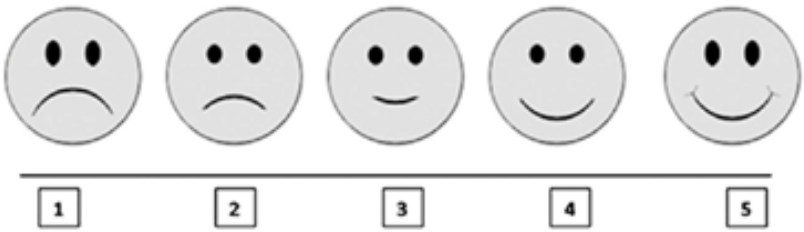


Figure 1. Evaluation scale

2.6 Text comprehension and evaluation tasks

After reading each text, the participants answered an open-ended free recall text comprehension question ("What do you remember of what you just read?", see Gutermuth 2020). This question format was chosen due to difficulties related to intellectual disabilities that might arise when working with this group of readers (for a further discussion, see e. g., Fajardo et al. 2014; Sutherland & Isherwood 2016; Gutermuth 2020). The questions were answered orally and recorded. The answers were marked down in the ethnographic material consisting of field notes on the participants' answers. The participants also evaluated how difficult and how interesting they found the text ("How difficult/interesting do you think the text was, on a scale of 1–5?").

A picture of a 1–5 scale was used to facilitate communication concerning both questions (see Figure 1). These evaluations were included in the ethnographic material.

2.7 Procedure

Testing took place in a room at the hosting institution (target group) and in the eye-tracking lab of the university (control group). The participants were tested individually and instructed orally following the same scheme. They were informed that the purpose of the research was to scrutinise Easy Language texts and that they were about to read two such texts. Each participant signed an informed consent form before experimentation. Subsequently, the eye-tracker was set up and each participant was calibrated using a three-point

calibration screen. They were instructed to "Read as well as you can, so that you understand.", and were informed that the experiment consisted of two texts divided onto several pages each. Reading time was not restricted – the participants were allowed to read at their own pace. Advancement in the text was moderated by the researcher. The participants were instructed to inform the researcher when they had finished reading a page, and the researcher then changed the page. This procedure was chosen to allow the participants to fully concentrate on the reading task and not be distracted by technical and motor challenges. Returning to a previous page was not possible. Prior to the experiment, a practice trial was run, consisting of two text pages and a short break that was used to clarify the task if needed. Half of the participants read Text B first. Those who failed the calibration process (§ 2.1) read both the texts in one word document on the 14″ screen of a laptop computer (Yoga 530-14IKB). The researcher moderated advancement in the text. All the participants answered the follow-up question, also those whose eye-tracking data we could not include due to calibration failure.

2.8 Qualitative analysis

The answers to the follow-up recall question were transcribed and analysed using qualitative content analysis (e.g., Graneheim et al. 2017; Krippendorff 2019). The approach was inductive, consisting of only a concrete analysis of the manifest content, with a low abstraction level and a low interpretation degree (Graneheim et al. 2017: 30–31). The answers were then analysed using an adapted version of the coding scheme used by Gutermuth (2020: 154). In this scheme, the content of the recall was given a number (0, 0.5, 1) according to the information included (see Table 4). Although the free recall question answers provide limited information about the participants' text comprehension, the term comprehension is used when referring to the free recall coding scheme results (cf. Gutermuth 2020).

2.9 Statistical analysis

Eye-tracking data were analysed by linear mixed-effects models (LMM) using the *lme4* package (Bates et al. 2015) in the R statistical software (RStudio version 1.4.1103; R Core Team 2020). The dependent variables (the different eye movement measures: dwell time, first run dwell time, and regression path duration, see § 3.1 for a detailed description of the measures) were log-

Table 4. Comprehension rates of the free recall coding scheme (for comparison, see Gutermuth 2020: 154).

Code	Comprehension	Evaluation Criteria
1	Good overall comprehension	Reproduction of the main content, reproduction of relevant information.
0.5	Partial comprehension	Reproduction of some relevant details but no reproduction of the main content.
0	Poor or no comprehension	No reproduction of the main content or relevant information.

transformed before the analyses. The target group (dummy coded: control group as baseline) and word length (centred), frequency (centred), and repetition were added to the models as fixed effects variables, one at a time. Due to the high correlation between word length and frequency ($r = -0.72$), these were separated into different models. Random intercepts for participants and words were included in the random part of the models.

Observations exceeding three standard deviations from the grand mean were excluded from the analyses. This resulted in the exclusion of 1.92% of data for dwell time, 1.85% for the first run dwell time, and 1.3% of the regression path duration. A statistical significance of .05 was indicated by values of [t or z] > 1.96 (Baayen 2008). For the sake of brevity, only significant effects are reported in the text. The final models are reported in Appendix A Tables 1–3. The dataset and the analysis code are available at https://osf.io/dcjgk/.[6]

3 Results

3.1 Pre-processing of data

Three different reading measures were computed for all the words in the texts from the eye movement data: dwell time, first run dwell time, and the regression path duration. Dwell time is the summed duration of all the fixations landing on a word. First run dwell time is the summed duration of fixations landing on the word during its first-pass reading. Regression path duration is the summed duration of fixations calculated from when the word is first fixated

[6] Last updated 2023-01-31.

on until the time the reader moves on to the next word. These measures were chosen because they provide information on the different stages of processing. The first run dwell time informs us of initial processing, whereas the dwell time and the regression path duration provide information on later processing, such as integrating the word meaning into the sentence context (Rayner 1998; 2009).

Every word in the text was assigned an area of interest (AOI). Due to calibration issues, in some of the gaze measurements, the fixations were positioned slightly off the rows and were systematically moved to the right row. In some cases, the calibration had drifted to the right. In cases where it was completely clear how the calibration had drifted, fixations were systematically moved to the left so that the first fixation was on the first word. However, the words affected by this drift or obscurity were removed from the data before the word-level analysis was conducted. Accurate field notes were made on the drifting, so that the drifted fixations could be later moved to the right line.

3.2 Effects of word length, frequency, and repetition

As expected, longer words attracted longer fixation durations. In addition, the effect of word length was more prominent in the target group. The effect of word length was observed in all three fixation measures: dwell time ($\beta = 0.29$, 95% CI [0.26, 0.32], $t = 19.41$), first run dwell time ($\beta = 0.18$, 95% CI [0.14, 0.21], $t = 10.84$), and regression path duration ($\beta = 0.20$, 95% CI [0.15, 0.25], $t = 8.03$). An interaction between group and word length was observed in all three measures: dwell time ($\beta = 0.09$, 95% CI [0.06, 0.12], $t = 5.77$), first run dwell time ($\beta = 0.07$, 95% CI [0.04, 0.11], $t = 4.48$), and regression path duration ($\beta = 0.16$, 95% CI [0.10, 0.22], $t = 5.12$). This means that the word length effect was greater in the target group than in the control group (see Figure 2).

The models with lemma frequency as a predictor showed that as the frequency of a word increased, the duration of the gaze on the word decreased. This effect was more prominent in the target group than in the control group. The effect of lemma frequency was revealed for dwell time ($\beta = -0.25$, 95% CI [−0.29, −0.21], $t = -12.14$), first run dwell time ($\beta = -0.15$, 95% CI [−0.19, −0.11], $t = -8.01$), and regression path duration ($\beta = -0.15$, 95% CI [−0.21, −0.10], $t = -5.47$). An interaction between group and frequency was observed in all three measures: dwell time ($\beta = -0.07$, 95% CI [−0.10, −0.04], $t = -4.50$), first run dwell time ($\beta = -0.06$, 95% CI [−0.09, −0.02],

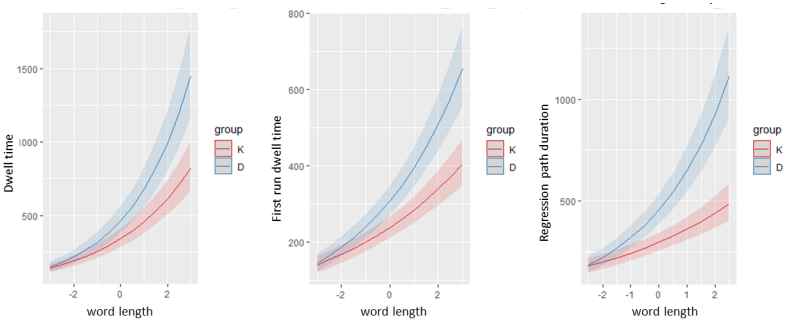


Figure 2. Effects of word length and interactions between group and word length for dwell time, first run dwell time, and regression path duration (D=target group, K=control group)

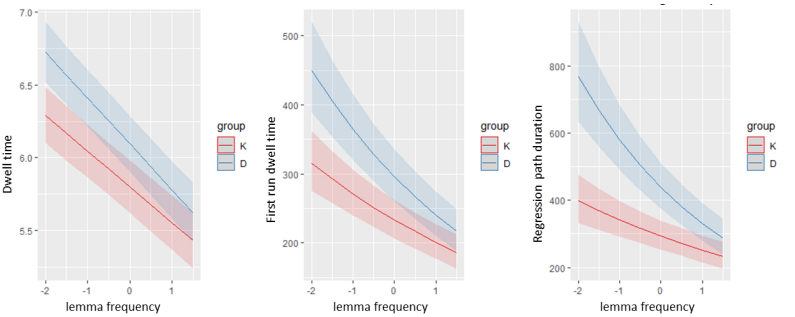


Figure 3. Effects of lemma frequency and interactions between group and frequency for dwell time, first run dwell time, and regression path duration (D=target group, K=control group)

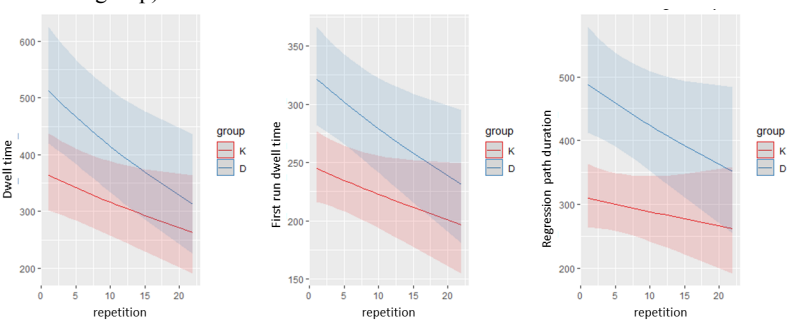


Figure 4. Effects of word repetition and interactions between group and repetition for dwell time, first run dwell time, and regression path duration (D=target group, K=control group)

$t = -3.37$), and regression path duration ($\beta = -0.13$, 95% CI [−0.19, −0.07], $t = -4.03$). This means that the effects of lemma frequency were greater in the target group (see Figure 3).

Word repetition affected only some measures. As the number of repetitions of the word within the text increased, the duration of the gaze on the word decreased. However, the repetition effect was only significant for dwell time ($\beta = -0.02$, 95% CI [−0.03, −0.00], $t = -2.1$), not for first run dwell time or regression path duration. An interaction between group and repetition was also only observed during dwell time ($\beta = -0.01$, 95% CI [−0.01, −0.00], $t = -2.76$). Hence, the greater effect in the target group was only observed in the dwell time fixations, the other two reading measures showed similar effects in both groups (see Figure 4).

In addition, the readers in the target group read more slowly overall, as indicated by an effect of group in all measures (Appendix A). A complete presentation of the model results can be found in Appendix A.

3.3 "This was not exactly easy to read"

The free recall question comprehension rates were considerably low in the target group; this was the case for both texts (Table 5). None of the participants in this group were able to reproduce the main content of the texts, even when they had some previous knowledge of the text subject. Although three of the participants gave an accurate description of a proxy in the pre-questionnaire, they were unable to reproduce the content of the text on proxies (Text B). One of them could reproduce fragments of the content (Transcript 2). In contrast, the control group participants were able to reproduce the main content and recalled relevant information (Table 5).

As shown in § 2.1, although the decoding and reading comprehension test results showed variance both between the participants in the target group and between the groups, the results of these tests had no apparent connection to the results of the free recall task in the target group. The participants who scored better in the baseline tests did not necessarily reproduce more, or more accurate information from the texts.

How difficult the texts were rated by the participants also had no apparent connection to the results of the free recall task (Table 5). The texts were rated as slightly more difficult in the target group, but also as more interesting. Seven in the target group and five in the control group rated the text they read first as more difficult. One of the participants commented that the texts contained

Table 5. Results of follow-up questions provided in mean (M) and standard deviation (SD) values for each text in the target group and in the control group.

TEXT A	Target group M	Target group SD	Control group M	Control group SD
Prequestionnaire	0	0	0.81	0.35
Comprehension*	0.19	0.24	0.94	0.17
Difficulty rating	3.43	1.05	2.13	1.27
Interest rating	3.71	0.88	3	0.71
TEXT B				
Prequestionnaire	0.38	0.48	0.88	0.33
Comprehension*	0.06	0.17	1	0
Difficulty rating	3.08	0.61	2.44	0.77
Interest rating	3.71	0.88	2.57	0.86

* 0=poor or no comprehension, 0.5=partial comprehension, 1=good overall comprehension

difficult words, and another stated: "This was not exactly easy to read". The highest interest ratings were received from the participants with the lowest comprehension rates, except for one from the target group, who recalled partial information from Text A.

Most participants recalled some details but no other content of the texts. For example, one of them recalled the words *proxy*, *private persons* and *company*, but did not recall the connection between these or the roles of these in the text (Transcript 1). Another participant recalled only the name *FPA*,[7] the words registration and wage information, and the abbreviation *LL*, which was part of the name *LL-Center* mentioned at the beginning of the text. A third participant recalled that the texts contained information about money ("it's about money"), paying bills, and "having access to" something but did not recall to what. Five participants recalled the name *FPA*.

[7] Folkpensionsanstalten (Kela) is The Social Insurance Institution of Finland. This name is probably familiar to many in the Easy Language target group as it provides social security coverage for Finnish residents and offers social security benefits such as family benefits, health insurance, rehabilitation, basic unemployment security, basic social assistance, and disability benefits. (https://www.fpa.fi/)

(1) PARTICIPANT 2. R=researcher, P=participant.

01 R: Vad kommer du ihåg från det som du just läste?
'What do you remember of what you just read?'

02 P: (.) om (.) det där (.) fullmakt och sånt (...) till privater personer och (.) företag och (...)
'(.) of (.) that (.) proxy and such (...) for private persons and (.) company and (...)

03 R: mm (...)

04 P: Sen var där nån länk också.
'Then there was a link too.'

Five of the participants were able to recall some relevant details of the content. Three answers contained accurate information on proxies following the free recall question on Text B. One explained that it was dealing with things "instead of someone", another recalled that the text contained information on "if you want to give a proxy to someone". Three participants recalled partial information from Text A. One seemed to comprehend one of the aims of the income register: that the income information is gathered without the employees having to take measures. Another recalled, "they can see there what your wage is and whatnot".

(2) PARTICIPANT 8. R=researcher, P=participant.

01 R: Vad kommer du ihåg av det som du just läste? (.)
'What do you remember of what you just read?'

02 P: Jo det var så att man sku ge dehär fullmakt åt en annan som har någo ärende att man kan hjälpa nån annan med fullmakter å dehär inkomster å dehär för fullmakt skaffar man också för någo sånt förstod jag [...]
'Yes, it was so that you would give this proxy to another who has to take care of something that you can help someone else with proxies and this income and that for proxy you also get or something like that I understood'

10 P: (?) någo, va någo om dehär förvaltningsregister eller dehär någo FPA o dehä me Januari 2021 som man sen sku börja me att lämna in dehä me re re register å löner å de som hjälper att andra kan göra sina ärenden dehär me pension och löner de får man via såna här löneregister å så får man fara på FPA å på någo info (?) löneregister, någo sånt förstod jag med det här nu.

> '(?) something, it was something about this administration register or this FPA and in January 2021 that you then start to hand in this with the re- re- register and wages and it then helps others take care of their affairs with pensions and wages that you get through these wage registers and then you can go to FPA and some info (?) wage register, something like that I understood of this now'

The participants who read the papers in one word document did not perform better on the free recall task, with the exception of one, who was able to reproduce some relevant details from both texts. This participant explained that a proxy can be given to another person who has to deal with some matter (Transcript 2). However, the main point of the text – how to use proxies when handling matters in the electronic income register, was not reproduced. A similar answer was given regarding Text A, showing that the participant had understood that the text had something to do with wages and a register and that the register collects information on wages, but apparently did not fully understand the content. The participant also noted some relevant details: "FPA" and the date, January 2021.

4 Discussion

The purpose of the present study was to investigate the effects of word length, frequency and repetition on eye movements, and whether these differed in the target group and the control group. Another aim was to examine reading comprehension by asking the participants what they recalled of the texts. The results showed significantly greater word length and frequency effects and poor comprehension rates in the target group.

The eye-tracking results and statistical analysis showed significant word length and frequency effects for all measures. That is, high frequency words and short words were read faster than low frequency and long words. These results are in line with previous eye-tracking research (Rayner 1998; 2009). The effects were also significantly more prominent in the target group. This provides new knowledge about the Easy Language group: long and infrequent words seemed to have a greater effect on the processing time of the readers in this group than on that of the non-impaired readers: the longer and more infrequent the word, the longer the fixations. These results differ from those of the eye-tracking study of Finnish language texts by Hyönä & Olson (1995), which showed similar effects of word length and frequency on non-dyslexic

and dyslexic readers. Similarly, an eye-tracking study by Schiffl (2020; 2021) demonstrated no greater word length, frequency, or word repetition effects among readers with cognitive impairments than among other readers. However, the number of participants was greater in the German study ($n = 30$), and the group definition was slightly different ("readers with cognitive impairments"). The differences could also derive from dissimilarities between, for example, the levels of text complexity in the Swedish and German texts used in the experiments, or from the fact that Schiffl used sentences instead of texts. The results of this study are yet in line with previous research showing generally greater effects of word length and frequency for children compared to adults (Tiffin-Richards & Schroeder 2015). Similar to the present study, Tiffany and Schroeder used reader-appropriate frequency estimates.

The results regarding word repetition were not as clear. The repetition effect was only significant for the dwell time measure. Moreover, the greater effect in the target group was only observed in the dwell time fixations. The more prominent word length, frequency, and (partial) repetition effects in the target group indicate that infrequent words, long words, and words that are repeated less often within the text can cause greater difficulties in processing words for members of this group of readers. It is thus likely that these characteristics cause more reading problems in this group of readers. This adds validity to the Easy Language guidelines, which state that writers should use frequent and short words and repeat words instead of using synonyms.

Nevertheless, as the sample of this study was quite small, and the inclusion criteria included affiliation with the day centre for people with intellectual disabilities and certain diagnoses were not accessed, the results are not generalizable to readers with intellectual disabilities in general. Furthermore, the Easy Language target group is a heterogeneous group of readers with different needs, qualifications and reading abilities, with variations even in diagnoses (e.g., Heimann Mühlenbock 2013: 18–19; Forsberg 2014: 33, 39–40; Arle & Frondén 2022). The target group also displayed longer fixation times overall, which is in line with previous findings regarding English language texts, that have shown that disabled readers, compared to good readers, had longer average fixation durations (Rayner 1983; Reichle et al. 2013). The baseline test scores for both reading comprehension and decoding varied both between the groups and within the target group, which is also in line with previous research (e.g., Schiffl 2020). This underpins the expressed heterogeneity of the Easy Language target group (for a further discussion, see Arle & Frondén 2022).

The application of eye-tracking methodology in this subgroup of the Easy Language target group was rather complicated. As described in § 2, the experimental design underwent several adaptations in order to meet the needs and difficulties of the readers in the target group. For example, the instructions and the reading task had to be very simple, so that they were not too difficult for the readers to comprehend. We also noticed that the readers seemed to be afraid of underperforming and eager to meet the researchers' expectations. Consequently, designing and implementing the experiment without creating a sense of failure among the participants was important, and resulted in the simple experiment design described in § 2. Despite this, calibration issues were experienced, resulting in loss of data. These experiences are in line with those in previous eye-tracking studies that included readers with disabilities (see e.g., Gutermuth 2020).

The free recall task showed poor comprehension rates in the target group. The low rates indicate that these texts were too difficult for the readers in the target group. The longer fixation times on long, infrequent, and partly less repeated words reflect problems in comprehending individual words. Despite this, although Text A contained slightly longer words and less frequent words in mean values (Table 5), the comprehension rates were slightly higher than those for Text B (Table 5). Even though this difference was not statistically significant, this finding indicates that text comprehension is also influenced by factors other than word frequency and word length, such as the syntactic and semantic complexity of the text, or the familiarity of the topic discussed in the text. The details remembered by the readers in the target group also did not have a clear connection to either frequency, word length or repetition. This study revealed significant effects of word length and frequency on fixations while reading, and some effect for word repetition. The study by Fajardo et al. (2014) showing no effect for word length and frequency on comprehension but significant effects for the number of coreferences on literal comprehension, studied reading comprehension on both the literal and inferential levels. However, as the present study only studied reading fixations and free recall response, it did not examine the effects of word length, frequency, and repetition on different levels of comprehension. Furthermore, this study included only repetition of the same noun in the analysis and excluded repetition of the word stem and shared word steam allowing word category variation (cf. Fajardo et al. 2014). Inclusion of the latter two would possibly have resulted in greater effects for repetition.

The answers to the recall question show that the readers in the target group mostly remembered irrelevant details. Some details were relevant, but very few of the readers recalled any of the main content. The answers did not reflect a deeper integration of information or comprehension of the texts and what this information could be used for. However, the free recall answers provide merely an indication about their comprehension (e.g., Gutermuth 2020: 151–152). Based on the answers of the participants in the examples (1–2), it is difficult to determine whether the lack of information in the answer was due to poor comprehension, issues related to expression difficulties, or other factors. Nevertheless, as this study did not examine further cognitive and language processing, and no neuropsychological test battery was included in the baseline test (cf. Gutermuth 2020; Pappert & Bock 2020; Schiffl 2020; Borghardt et al. 2021), this issue was not investigated any further. However, the robust effect of word frequency on eye fixation times indicated that the readers were indeed processing the meaning of the words in the text, and not simply gazing at the text mindlessly (see e.g., Reichle et al. 2010). One of the participants in the target group who read the texts without the eye-tracking camera read the text aloud. This participant read every word correctly but was later unable to reproduce the content of the texts.

The reading strategies fostered by the instructions before reading (to read as well as they could) might also have generated different reading strategies – and thereby the free recall question responses – than in a real-life situation, in which public authority texts are probably read for a purpose that derives from an actual problem or situation. The reading and production context of public authority texts usually have specific aims for both producer and reader (e.g., Forsberg 2014). In this case, the reading strategy was perhaps not equal to those in real-life situations and might thereby have had negative effects on the recall task. However, in terms of the participants' disabilities, this procedure was chosen to keep the experiment as simple as possible, and the free recall procedure provides yet an indication of poor comprehension in the target group.

The large impact of the word length and frequency effects in the target group supports the guidelines' recommendation that writers should use shorter and more frequent words. The higher fixation times indicated that the readers struggled with these word characteristics.

5 Limitations of the study and future research

Other practical details that differed from natural reading situations (e.g., the layout of the original texts) in addition to those described previously, were bigger font size and line spacing, and the way in which the text was presented on several pages in the eye-tracking experiment. In a real-life situation, the original texts would be presented on a web page as linear texts. The focus of this study was three specific word-level characteristics: word length, frequency, and repetition. The study contained no multimodal aspects, although the recommendation to add pictures that support the text is included in the Easy Swedish guidelines (e.g., LL-Center; MTM), and the use of pictures could affect reading comprehension. However, the original format of the chosen text also contained no pictures. From a perspective of readability and comprehensibility, it should also be noted that many factors interact and that a great range of different aspects affect the level of readability and comprehensibility (for a further discussion, see e.g., Wengelin 2011; 2015).

The small sample of this study is perhaps its greatest limitation. The small number of participants in both in the target group and the control group, in combination with the calibration problems mentioned in § 3.1, resulted in a small data size. As recalibration noticeably led to increased stress among the participants, this was not an option. Despite this, the LMM analyses generated significant effects, which strongly demonstrates the examined word-level effects (word length, frequency, and repetition) in this group. However, future studies should examine these effects with a larger group of readers.

Word length was calculated in terms of numbers of letters; this could also have been calculated in number of syllables or morphemes (cf. e.g., Hyönä & Pollatsek 1998). Future studies could add these. For frequency estimates, lemma frequencies were used. However, as morpheme frequency can also affect word identification (Reichle & Perfetti 2003), it may also be beneficial to study morpheme frequency effects among disabled readers. A study by Valtasalmi (2022) examined lexical knowledge of adults with intellectual disabilities. The results of the study showed that despite high frequency of the words included in the task possibly contributed to correct responses, these readers also knew low-frequency words if the words were familiar from "everyday language" (Valtasalmi 2022). As this study used some of the Easy Language writing guidelines as its basis, and the formulation of "common words" was understood as *frequent* words, the aspect of individual familiarity was ignored. However, this aspect affects processing (i.e., words

that are familiar to a certain reader are processed faster) and familiarity is not always linked to frequency, as a word that is familiar to one reader may be unfamiliar to another (Gernsbacher 1984: 275–277). This aspect could perhaps be included in a future study. In future studies, reading comprehension could also be examined at different levels of comprehension to further examine the effect of Easy Language and the validity of the related guidelines (see e.g., Kintsch & Dijk 1978; Kintsch 1994; Fajardo et al. 2014). Future studies could also be designed to reflect more of the ordinary reading strategies used when reading public authority information, for example, answering a specific question that influences the reader's everyday life. Instead of being instructed to read as well as they can so that they understand, the readers could be given a practical question to answer or a practical problem to solve (e.g., "you have to apply for a subsidy" or "you are going to vote in a public election, how should you do this?"). However, this may complicate the experiment and the instructions too much for some readers in the target group. Future studies could also examine the competencies that influence the reading abilities of this group of readers. A text analysis of the studied texts, as well as a comparison of Easy Language authority texts in different countries, was also excluded from this study, and would be a welcome focus in future linguistic research.

References

Arle, Solveig & Frondén, Carina. 2022. Bringing order to chaos: Research on Easy Swedish. *Journal of Nordic Linguistics* 45(2). 167–193. https://doi.org/10.1017/S0332586522000105

Baayen, Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Balota, David A. & Chumbley, James I. 1984. Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human perception and performance* 10(3). 340–357. https://doi.org/10.1037//0096-1523.10.3.340

Balota, David A. & Cortese, Michael J. & Sergent-Marshall, Susan D. & Spieler, Daniel H. & Yap, Melvin J. 2004. Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General* 133(2). 283–316. https://doi.org/10.1037/0096-3445.133.2.283

Bates, Douglas & Mächler, Martin & Bolker, Ben & Walker, Steve. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01

Bohman, Ulla. 2021. Easy Language in Sweden. In Lindholm, Camilla & Vanhatalo, Ulla (eds.), *Handbook of easy languages in europe*, 527–571. Frank & Timme.

Borghardt, Liv & Deilen, Silvana & Fuchs, Julia & Gros, Anne-Kathrin & Hansen-Schirra, Silvia & Nagels, Arne & Schiffl, Laura & Sommer, Johanna. 2021. Neuroscientific research on the processing of Easy Language. *Frontiers in communication* 6. https://doi.org/10.3389/fcomm.2021.698044

Chinn, Deborah & Homeyard, Claire. 2017. Easy read and accessible information for people with intellectual disabilities: Is it worth it? A meta-narrative literature review. *Health Expectations* 20(6). 1189–1200. https://doi.org/10.1111/hex.12520

Deilen, Silvana. 2021. *Optische Gliederung von Komposita in Leichter Sprache: Blickbewegungsstudien zum Einfluss visueller, morphologischer und semantischer Faktoren auf die Verarbeitung deutscher Substantivkomposita*. (Doctoral dissertation).

Deilen, Silvana & Schiffl, Laura. 2020. Using eye-tracking to evaluate language processing in the easy language target group. In Hansen-Schirra, Silvia & Maaß, Christine (eds.), *Easy language research: Text and user perspectives*, 273–282. Frank & Timme.

Domeij, Richard & Spetz, Jennie. 2014. Begripliga myndigheter på nätet. In Lindström, Jan & Henricson, Sofie & Huhtala, Anne & Kukkonen, Pirjo & Lehti-Eklund, Hanna & Lindholm, Camilla (eds.), *Svenskans beskrivning 33. Förhandlingar vid Trettiotredje sammankomsten för svenskans beskrivning, Helsingfors den 15–17 maj 2013*, 52–62. University of Helsinki.

Fajardo, Inmaculada & Ávila, Vicenta & Ferrer, Antonio & Tavares, Gema & Gómez, Marcos & Hernández, Ana. 2014. Easy-to-read texts for students with intellectual disability: linguistic factors affecting comprehension. *Journal of applied research in intellectual disabilities* 27(3). 212–225. https://doi.org/10.1111/jar.12065

Forsberg, Camilla. 2014. Myndigheten som pedagog: Om myndigheters lättlästa texter. In Persson, Anders & Johansson, Roger (eds.), *Vetenskapliga perspektiv på lärande, undervisning och utbildning i olika institutionella sammanhang: Utbildningsvetenskaplig forskning vid Lunds universitet*, 25–43. Lund University.

Forster, Kenneth I. & Chambers, Susan M. 1973. Lexical access and naming time. *Journal of verbal learning and verbal behavior* 12(6). 627–635. https://doi.org/10.1016/S0022-5371(73)80042-8

Gernsbacher, Morton A. 1984. Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General* 113(2). 256–281. https://doi.org/10.1037//0096-3445.113.2.256

Graneheim, Ulla H. & Lindgren, Britt-Marie & Lundman, Berit. 2017. Methodological challenges in qualitative content analysis: A discussion paper. *Nurse Education Today* 56. 29–34. https://doi.org/10.1016/j.nedt.2017.06.002

Gutermuth, Silke. 2020. *Leichte Sprache für alle? Eine zielgruppenorientierte Rezeptionsstudie zu Leichter und Einfacher Sprache*. (Doctoral dissertation).

Hansen-Schirra, Silvia & Abels, Katja & Signer, Sarah & Maaß, Christiane. 2021. *The dictionary of accessible communication*. Frank & Timme.

Hansen-Schirra, Silvia & Maaß, Christine (eds.). 2020. *Easy language research: Text and user perspectives*. Frank & Timme.

Heimann Mühlenbock, Katarina. 2013. *I see what you mean: Assessing readability for specific target groups*. Språkbanken, Department of Swedish, University of Gothenburg. (Doctoral dissertation).

Hyönä, Jukka & Kaakinen, Johanna K. 2019. Eve movements during reading. In Klein, Christoph & Ettinger, Ulrich (eds.), *Eye movement research: An introduction to its scientific foundations and applications*, 239–274. Springer Nature Switzerland.

Hyönä, Jukka & Niemi, Pekka. 1990. Eye movements during repeated reading of a text. *Acta Psychologica* 73(3). 259–280.

Hyönä, Jukka & Olson, Richard K. 1995. Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21. https://doi.org/10.1037/0278-7393.21.6.1430

Hyönä, Jukka & Pollatsek, Alexander. 1998. Reading Finnish compound words: Eye fixations are affected by component morphemes. *Journal of Experimental Psychology: Human Perception and Performance* 24(6). 1612–1627. https://doi.org/10.1037//0096-1523.24.6.1612

IFLA [International Federation of Library Associations and Institutions]. 2010. *Guidelines for easy-to-read materials*. (IFLA Professional Reports 120).

Inhoff, Albrecht Werner & Rayner, Keith. 1986. Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception and Psychophysics* 40. 431–439. https://doi.org/10.3758/BF03208203

Inhoff, Albrecht Werner & Topolski, Richard & Vitu, Francoise & O'Regan, Kevin J. 1993. Attention demands during reading and the occurrence of brief (express) fixations. *Perception & Psychophysics* 54(6). 814–123.

Joseph, Holly S. S. L. & Blythe, Hazel I. 2011. Children's eye movements during reading. In Liversedge, Simon & Gilchrist, Iain & Everling, Stefan (eds.), *The Oxford handbook of eye movements*, 643–662. Oxford University Press.

Joseph, Holly S. S. L. & Liversedge, Simon P. & Blythe, Hazel I. & White, Sarah J. & Rayner, Keith. 2009. Word length and landing position effects during reading in children and adults. *Vision Research* 49. 2078–2086. https://doi.org/10.1016/j.visres.2009.05.015

Karreman, Joyce & Van Der Geest, Thea & Buursink, Esmee. 2007. Accessible website content guidelines for users with intellectual disabilities. *Journal of applied research in intellectual disabilities* 20(6). 510–518. https://doi.org/10.1111/j.1468-3148.2006.00353.x

Kintsch, Walter. 1994. Text comprehension, memory, and learning. *The American Psychologist* 49(4). 294–303. https://doi.org/10.1037//0003-066x.49.4.294

Kintsch, Walter & Dijk, Teun A. van. 1978. Toward a model of text comprehension and production. *Psychological Review* 85(5). 363–394. https://doi.org/10.1037/0033-295X.85.5.363

Krippendorff, Klaus. 2019. *Content analysis: An introduction to its methodology*. 4th edn. SAGE.

Leskelä, Leealaura. 2021. Easy language in finland. In Lindholm, Camilla & Vanhatalo, Ulla (eds.), *Handbook of Easy Languages in Europe*, 149–190. Frank & Timme.

Lindholm, Camilla & Vanhatalo, Ulla. 2021. Introduction. In Lindholm, Camilla & Vanhatalo, Ulla (eds.), *Handbook of Easy Languages in Europe*, 11–26. Frank & Timme.

LL-Center. n. d. *Om lätt språk och lättläst*. (https://www.ll-center.fi/svenska/om_lattlast/om_lattlast2/). (Accessed 2021-04-27).

Lundberg, Ingvar & Reichenberg, Monica. 2008. *Vad är lättläst?* Specialpedagogiska skolmyndigheten.

Maaβ, Christiane. 2020. *Easy Language – Plain Language – Easy Language Plus: Balancing comprehensibility and acceptability*. Frank & Timme.

Moonen, Xavier. 2021. Easy Language in the Netherlands. In Lindholm, Camilla & Vanhatalo, Ulla (eds.), *Handbook of Easy Languages in Europe*, 345–370. Frank & Timme.

MTM [Myndigheten för tillgängliga medier]. n. d. *Att skriva lättläst*. (https://www.mtm.se/var-verksamhet/lattlast/att-skriva-lattlast/). (Accessed 2021-03-08).

Nordenstam, Anna & Olin-Scheller, Christina. 2018. Lättläst: En demokratifråga. *Utbildning och Demokrati* 27(3). 35–52. https://doi.org/10.48059/uod.v27i3.1107

O'Donnell, Maria & Ramdén, Maria. 2021. Addition to Easy Language in Sweden. In Lindholm, Camilla & Vanhatalo, Ulla (eds.), *Handbook of Easy Languages in Europe*, 527–571. Frank & Timme.

Österlund, Maria. 2011. *Så här skriver du lättläst*. LL-Center / FDUV.

Pappert, Sandra & Bock, Bettina M. 2020. Easy-to-Read German put to the test: Do adults with intellectual disability or functional illiteracy benefit from compound segmentation? *Reading & writing* 33(5). 1105–1131. https://doi.org/10.1007/s11145-019-09995-y

R Core Team. 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. (https://www.r-project.org/). (Accessed 2023-11-15).

Raney, Gary E. & Rayner, Keith. 1995. Word frequency effects and eye movements during two readings of a text. *Canadian Journal of Experimental Psychology* 49(2). 151–173. https://doi.org/10.1037/1196-1961.49.2.151

Rayner, Keith. 1983. Eye movements, perceptual span, and reading disability. *Annals of Dyslexia* 33. 163–173. https://doi.org/10.1007/BF02648003

— 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124(3). 372–422. https://doi.org/10.1037/0033-2909.124.3.372
— 2009. Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology* 62(8). 1457–1506. https://doi.org/10.1080/17470210902816461
Rayner, Keith & Duffy, Susan A. 1986. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition* 14. 191–201. https://doi.org/10.3758/BF03197692
Reichle, Erik D. & Liversedge, Simon P. & Drieghe, Denis & Blythe, Hazel I. & Joseph, Holly S. S. L & White, Sarah J. & Rayner, Keith. 2013. Using E-Z Reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review* 33. 110–149. https://doi.org/10.1016/j.dr.2013.03.001
Reichle, Erik D. & Perfetti, Charles A. 2003. Morphology in word identification: A word-experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading* 7(3). 219–237. https://doi.org/10.1207/S1532799XSSR0703_2
Reichle, Erik D. & Reineberg, Andrew E. & Schooler, Jonathan W. 2010. Eye movements during mindless reading. *Psychological Science* 21(9). 1300–1310. https://doi.org/10.1177/0956797610378686
Schiffl, Laura. 2020. Hierarchies in lexical complexity: Do effects of word frequency, word length and repetition exist for the visual word processing of people with cognitive impairments? In Hansen-Schirra, Silvia & Maaß, Christine (eds.), *Easy Language research: Text and user perspectives*, 227–240. Frank & Timme.
— 2021. Lexical complexity in German Easy Language: Effects of length, frequency and repetition on the visual word processing of the target group. (Presented at the 1st International Easy Language Day Conference (IELD), 27–28 May 2021).
Schilling, Hildur E. H. & Rayner, Keith & Chumbley, James I. 1998. Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition* 26(6). 1270–1281. https://doi.org/10.3758/BF03201199
Selkokeskus. 2022. *Selkokielen mittari 2.0*. Selkokeskus / Kehitysvammaliitto.
Sundin, Maria. 2007. *Lättläst: Så funkar det*. Jure förlag.
Sutherland, Rebekah Joy & Isherwood, Tom. 2016. The evidence for Easy-Read for people with intellectual disabilities: A systematic literature review. *Journal of policy and practice in intellectual disabilities* 13(4). 297–310. https://doi.org/10.1111/jppi.12201
TENK [Forskningsetiska delegationen]. 2019. *Etiska principer för humanforskning och etikprövning inom humanvetenskaperna i Finland*. (Forskningsetiska delegationens anvisningar).

Tiffin-Richards, Simon P. & Schroeder, Sascha. 2015. Word length and frequency effects on children's eye movements during silent reading. *Vision Research (Oxford)* 113(Pt A). 33–43. https://doi.org/10.1016/j.visres.2015.05.008

Valtasalmi, Idastiina. 2022. Lukutaitoisten kehitysvammaisten aikuisten sanamerkitysten hallinta. *Finnish Journal of Linguistics* 34. 301–332.

Wengelin, Åsa. 2011. Forskning om myndighetstexters begriplighet: Från praktiken till labbet och tillbaka igen? In *Juridisk sprog i Norden: Rapport fra en nordisk konference om klarsprog Reykjavík 11.–12. oktober 2011*, 91–100.

— 2015. Mot en evidensbaserad språkvård? En kritisk granskning av några svenska klarspråksråd i ljuset av forskning om läsbarhet och språkbearbetning. *Sakprosa* 7(2). https://doi.org/10.5617/sakprosa.983

Appendix A Final models

A.1 Models' outputs with lemma frequency as predictor

Dwell time			
Predictors	Estimates	CI	p
(Intercept)	5.80	5.62…5.98	< 0.001
Group [D]	0.29	0.04…0.55	0.025
Lemma frequency	−0.25	−0.29…−0.21	< 0.001
Group [D] × lemma frequency	−0.07	−0.10…−0.04	< 0.001
Random Effects			
σ^2		0.34	
τ00 Word		0.06	
τ00 Participant		0.06	
ICC		0.27	
N Participant		15	
N Word		231	
Observations		5895	
Marginal R^2 / Conditional R^2		0.173 / 0.394	

First run dwell time			
Predictors	Estimates	CI	p
(Intercept)	5.45	5.34…5.57	< 0.001
Group [D]	0.24	0.07…0.40	0.005
Lemma frequency	−0.15	−0.19…−0.11	< 0.001
Group [D] × lemma frequency	−0.06	−0.09…−0.02	< 0.001
Random Effects			
σ^2		0.29	
τ00 Word		0.04	
τ00 Participant		0.03	
ICC		0.19	
N Participant		15	
N Word		230	
Observations		4100	
Marginal R^2 / Conditional R^2		0.111 / 0.277	

Regression path duration			
Predictors	Estimates	CI	p
(Intercept)	5.68	5.54…5.83	< 0.001
Group [D]	0.40	0.19…0.61	< 0.001
Lemma frequency	−0.15	−0.21…−0.10	< 0.001
Group [D] × lemma frequency	−0.13	−0.19…−0.07	< 0.001
Random Effects			
σ^2		0.48	
τ00 Word		0.04	
τ00 Participant		0.04	
ICC		0.14	
N Participant		15	
N Word		210	
Observations		1991	
Marginal R^2 / Conditional R^2		0.131 / 0.257	

A.2 Models' outputs with word length as predictor

Dwell time			
Predictors	Estimates	CI	p
(Intercept)	5.83	5.65…6.01	< 0.001
Group [D]	0.31	0.04…0.57	0.023
Word length	0.29	0.26…0.32	< 0.001
Group [D] × word length	0.09	0.06…0.12	< 0.001
Random Effects			
σ^2		0.34	
τ00 Word		0.02	
τ00 Participant		0.07	
ICC		0.21	
N Participant		15	
N Word		245	
Observations		6332	
Marginal R^2 / Conditional R^2		0.231 / 0.396	

First run dwell time			
Predictors	Estimates	CI	p
(Intercept)	5.47	5.36…5.59	< 0.001
Group [D]	0.23	0.09…0.42	0.003
Word length	0.18	0.14…0.21	< 0.001
Group [D] × word length	0.07	0.04…0.11	< 0.001
Random Effects			
σ^2		0.30	
τ00 Word		0.03	
τ00 Participant		0.03	
ICC		0.15	
N Participant		15	
N Word		244	
Observations		4471	
Marginal R^2 / Conditional R^2		0.143 / 0.273	

Regression path duration			
Predictors	Estimates	CI	p
(Intercept)	5.69	5.54…5.84	< 0.001
Group [D]	0.43	0.21…0.64	< 0.001
Word length	0.20	0.15…0.25	< 0.001
Group [D] × word length	0.16	0.10…0.22	< 0.001
Random Effects			
σ^2		0.48	
τ00 Word		0.02	
τ00 Participant		0.04	
ICC		0.12	
N Participant		5	
N Word		221	
Observations		2104	
Marginal R^2 / Conditional R^2		0.178 / 0.276	

A.3 Models' outputs with repetition as predictor

Dwell time			
Predictors	Estimates	CI	p
(Intercept)	5.91	5.72…6.10	< 0.001
Group [D]	0.35	0.09…0.61	0.008
Repetition	−0.02	−0.03… −0.00	< 0.035
Group [D] × repetition	−0.01	−0.01… −0.00	< 0.006
Random Effects			
σ^2		0.34	
τ00 Word		0.14	
τ00 Participant		0.06	
ICC		0.37	
N Participant		15	
N Word		245	
Observations		6332	
Marginal R^2 / Conditional R^2		0.057 / 0.405	

First run dwell time			
Predictors	Estimates	CI	p
(Intercept)	5.51	5.38…5.64	< 0.001
Group [D]	0.28	0.11…0.45	0.001
Repetition	−0.01	−0.02… −0.00	< 0.072
Group [D] × repetition	−0.01	−0.01… −0.00	< 0.111
Random Effects			
σ^2		0.30	
τ00 Word		0.07	
τ00 Participant		0.03	
ICC		0.25	
N Participant		15	
N Word		244	
Observations		4471	
Marginal R^2 / Conditional R^2		0.047 / 0.285	

Regression path duration			
Predictors	Estimates	CI	p
(Intercept)	5.74	5.54…5.83	< 0.001
Group [D]	0.46	0.19…0.61	< 0.001
Repetition	−0.01	−0.02… −0.01	< 0.316
Group [D] × repetition	−0.01	−0.02… −0.00	< 0.206
Random Effects			
σ^2		0.48	
τ00 Word		0.10	
τ00 Participant		0.04	
ICC		0.23	
N Participant		15	
N Word		221	
Observations		2104	
Marginal R^2 / Conditional R^2		0.069 / 0.280	

Appendix B Open-ended pre-questionnaire

1. Vad är inkomstregistret? [‘What is the income register?’]
2. Vad är en fullmakt? [‘What is a proxy?’]

Contact information:

Carina Frondén
Scandinavian Languages
University of Helsinki
e-mail: carina(dot)fronden(at)helsinki(dot)fi

Johanna K. Kaakinen
Psychology
University of Turku
e-mail: johanna(dot)kaakinen(at)utu(dot)fi

Towards a typology of demonstrative verbs

Don Killian
University of Helsinki

Ekaterina Gruzdeva
University of Helsinki

Abstract

Demonstratives have traditionally been recognized as a semantic category in which different members are classified as belonging to certain lexical or syntactic categories. Current research suggests there are at least 7 different distinct lexical categories: determiners, pronouns, adverbs, non-verbal predicators, verbs, adpositions, and articles. This study looks at one of the aforementioned category of demonstratives, demonstrative verbs, based on a sample of 101 languages with demonstrative verbs out of a total of 1182 languages examined. We present a typological classification of demonstrative verbs based on semantics, an exploration on the morphosyntactic properties of demonstrative verbs and their use in different pragmatic functions, and vast illustrative data in support of our analysis.

Keywords: demonstrative verb, deixis, demonstrative, verb, adverb, predication, grammaticalization, word class

1 Introduction and background

Demonstratives are a semantic class of deictic expressions which serve to focus joint attention onto a referent in the surrounding situation or unfolding discourse (Diessel 2012). According to Hanks (1992: 47), the basic communicative function of deictic forms is "to individuate or single out objects of reference or address in terms of their relation to the current interactive context in which the utterance occurs". Himmelmann (1996: 210) suggests the following criterion for identification of demonstratives: "the element must be in a paradigmatic relation to elements which […] locate the entity referred to on a distance scale: as proximal, distal, etc."

Current data from around 1200 languages shows no exception to the generalization that all languages have demonstratives; the language with the simplest

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

Finnish Journal of Linguistics 36 (2023), 37–92
DOI: 10.61197/fjl.126939

Table 1. Classes of demonstratives

PoS	Function
determiner	used in apposition to a noun, e.g. *I like **this** book.*
adverbs	*He read the book **here**.*
pronoun	used to replace a noun, e.g. *I like **this**.*
non-verbal predicator	used in non-verbal clauses, e.g. ***Here-is/This-is** John.*
verbs	used for verbal heads of predicate, e.g. *The book **is-here**.*
articles	*I saw **the-here** dog.*
adpositions	*I bought soda **in-there** the store.*

demonstrative system known so far is Nimboran, where the single deictic *ndie* may mean variously 'this', 'that', 'these', 'those' (May 1997). Nimboran, however, has extremely complex spatial marking on the verb, and even verbal tense marking has allomorphs depending on whether the action takes place in the current location or not (ibid.).

Demonstratives have traditionally been classified as belonging to certain lexical or syntactic categories, with the vast majority of the research focused on demonstrative pronouns and demonstrative determiners. Diessel (1999) points out, however, that demonstratives can also fill other syntactic slots like adverbs and identifiers, such as English *here* and *there*, and Welsh *dyma* 'here-is/this-is' and *dyna* 'there-is/that-is'. Current research suggests at least 7 different distinct lexical categories of demonstratives, seen in Table 1; additional and more fine-grained distinctions may also be needed for the categories of determiners and adverbs, categories which contain under-researched semantic types such as manner, quality, quantity, and degree.

In comparison to such established categories as demonstrative pronouns, determiners and adverbs, demonstrative verbs have seen little research or awareness. One of the reasons is their referential meaning, which generally points out events rather than entities. For instance, Mosel (2004: 150) affirms with respect to demonstrative verbs in Samoan "Although both situational and discourse deictic uses can be observed, they [demonstrative verbs] cannot be classified as demonstratives because they do not refer to entities, but express how something is done or what someone thinks or says". Other reasons pertain to the frequent irregular behavior of demonstrative verbs compared to other verbs in that language, such as heavy restrictions on TAM marking. Even if a category of demonstrative verb is attested in a language, the demonstrative

verbs can lose deictic oppositions, cease to be used with deictic reference to an action, or can be grammaticalized away from their original deictic meaning. Finally, there is a frequent mismatch between morphosyntactic category and morphosyntactic function: although demonstrative verbs can function as verbal heads of predicates, they frequently occur in derived adverbial function. All these processes can make detection and analysis of demonstrative verbs a rather challenging task.

As a distinct morphosyntactic category, demonstrative verbs became known in the typological literature published in English[1] from Dixon (2003), where the author suggests a differentiation between nominal, adverbial and verbal demonstratives, based primarily on syntactic properties. Verbal demonstratives 'do like this/that', with a deictic reference to an action, were acknowledged in two languages, Bouman Fijian (Oceanic) and Dyirbal (Pama-Nyungan). In a later publication, Dixon (2010) retains the same classification strategy as Dixon (2003), but adds one additional language, Mapuche (Mapudungun).

Hagège (2008) is a study looking primarily at interrogative verbs such as Classical Mongolian (Mongolic) *je-ji* 'do what', but Hagège does mention the close relationship that interrogative verbs share with what he calls *deictic verbs* (what we call *demonstrative verbs*); as an example, comparable to Classical Mongolian *je-ji* 'do what' he finds *e-ji* 'do this' and *te-ji* 'do that' (Hagège 2008: 20). It should be noted, however, that demonstrative verbs had been identified as a separate morphosyntactic category in Mongolic languages already earlier, e.g. Poppe (1937) as well as the subsequent translation into English, Poppe (1964), and Rassadin (1991: 96).

More recently, Gruzdeva (2013) looks at demonstrative verbs, with a variety of semantic categories of verbs such as 'do like this/that', 'be like this/that', 'be this/that size', and 'be here/there'. Guérin (2015) focuses on demonstrative manner verbs in eighteen languages, although she also acknowledges that other types of demonstrative verbs exist, namely what she views as spatial and locational. Breunesse (2019) discusses demonstrative verbs in detail, focusing on three languages: Abui (Alor-Pantar), Musqueam (Salish), and Neverver (Oceanic). Moyse-Faurie (2019) examines demonstrative verbs and their role in expressing similarity, comparison of equality, and manner in Polynesian languages.

[1] In his introductory book on general linguistics published in Russian, Maslov (1975: 218) suggested recognizing the category of demonstrative verbs ("местоглаголия"), or the broader category of demonstrative predicatives ("местопредикативы").

Demonstrative verbs are also distinguished in some grammatical descriptions of individual languages. Table 2 shows the ontological categories of demonstratives in Siar (Oceanic) and includes the singular and plural forms of two demonstrative verbs, which are translated as 'be here' and 'be there'.

In the same vein as Guérin (2015) and Breunesse (2019), we claim that the category of demonstrative verb fits semantically and pragmatically with other ontological categories of demonstratives, sharing the same deictic features, and performing similar functions. Based on form, morphological properties, and syntactic functions, demonstrative verbs form a distinct category belonging to the part of speech of verb. The most important verb-like properties include the ability to function as a clausal predicate without a copula, and inflecting for some of the different verbal categories such as tense, aspect, mood, and voice.

Both Guérin (2015) and Breunesse (2019) attempt to contrast demonstrative verbs with what they call demonstrative identifiers, relabeled here as predicative demonstratives. While we agree with both authors that a distinction exists, and indeed many languages show both verbal as well as non-verbal predicative demonstratives, some clarification must be made on the details. Predicative demonstratives are non-verbal predicators, heads of predicates which do not allow for the morphological marking of tense, aspect, or modality (Killian 2022b). Predicative demonstratives do frequently show person indexation, however. There is generally little functional overlap between verbal and non-verbal demonstratives, aside from some specific cases of ambiguity between certain subtypes of predicative demonstratives and demonstrative verbs, discussed further in the corresponding sections (cf. §§ 3.1 and 3.4.1).

Demonstratives are known to be used both in deictic and non-deictic functions. Deictic uses include exophoric (including the so-called *deixis am Phantasma*, following Himmelmann 1996) and discourse deictic, both of which are prolific with demonstrative verbs. Non-deictic use includes tracking (anaphoric and cataphoric reference), empathetic (what we call expressive), and recognitional, and are also attested within the class of demonstrative verbs. Figure 1 shows the different functions of demonstratives as presented in Levinson (2018).

The main goals of this article are to present a semantic typology of demonstrative verbs, to study their morphosyntactic properties, to explore their use in different pragmatic functions, and to provide vast illustrative data in support of the presented analysis. Comparing to previous works, the current study expands considerably both on the number of languages as well as the classification into different categories. The typology presented here is based

Table 2. Siar demonstratives, adapted from Frowein (2011: 333)

Function	Root	Demonstrative determiner-pronoun		Demonstrative locative-existential verb		Demonstrative adverb	
		Singular	Plural	Singular	Plural	Locative	Allative
Proximal	a	da 'this'	na 'these'	ada 'be here (sg.)'	ana 'be here (pl.)'	ta 'here'	kata 'hither'
Indexical	è	dè 'that'	nè 'those'	adè 'be there (sg.)'	anè 'be there (pl.)'	tè 'there'	katè 'thither'
Anaphoric	ing	ding 'that'	ning 'those'	ading 'be there (sg.)'	aning 'be there (pl.)'	ting 'there'	kating 'thither'
Clockwise	óng	dóng 'that'	nóng 'those'	adóng 'be there (sg.)'	anóng 'be there (pl.)'	tóng 'there'	katóng 'thither'
Counter-clockwise	im	dim 'that'	nim 'those'	adim 'be there (sg.)'	anim 'be there (pl.)'	tim 'there'	katim 'thither'
Upward	-(i)sai	disai 'that'	nisai 'those'	adisai 'be there (sg.)'	anisai 'be there (pl.)'	sai 'there'	kasai 'thither'

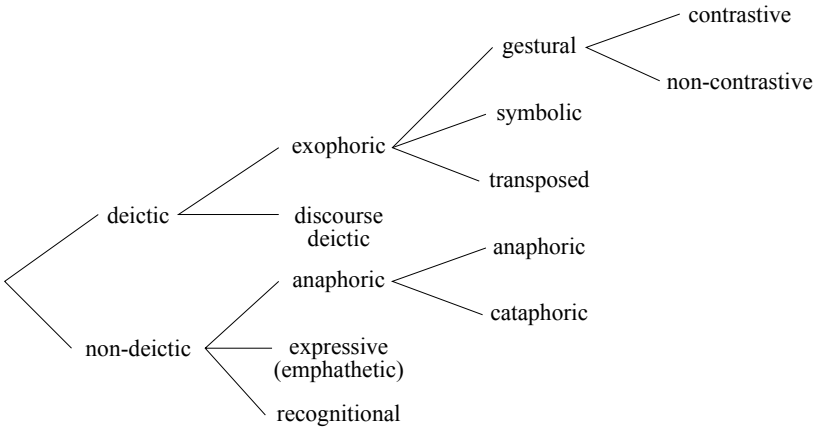


Figure 1. Distinct uses of demonstratives (Levinson 2018)

on 101 languages with demonstrative verbs (Appendix 1), out of a total of 1182 languages examined (Appendix 2).[2] The total sample has some bias, as the focus was on collecting as many descriptions of demonstrative systems as possible; the 101 languages included here as the focus of study are the languages with demonstrative verbs found out of the 1182 total.

Table 3 shows the total number of sampled languages and language families. Note that the total number of families is not simply a sum of the families for each macroarea. Families may be spoken in multiple macroareas and thus be counted multiple times. Afro-Asiatic for instance is counted in Africa as well as Eurasia, and treated as a family for both macroareas. The total, however, is rather the total of all families without macroareal partition, so Afro-Asiatic would only be counted once. Furthermore, the following non-genealogical "families" on Glottolog were excluded here: Unclassifiable, Pidgin, Unattested, Artificial Language, Mixed Language, Speech Register, and Sign Language.[3]

Despite the bias, the total sample is still reasonably diverse; using a chi-

[2] The appendices "Appendix 1. Languages with demonstrative verbs" and "Appendix 2. Total language sample" are available at https://doi.org/10.61197/fjl.126939

[3] Note that although pidgins are excluded from the sample, creoles are included. Sign languages are excluded only due to practical considerations, and more research on demonstrative use in sign languages is desperately needed.

Table 3. Total number of sampled languages, families, and coverage (share of families sampled) in each macroarea

	Afr.	Eur.	Papunes.	Austr.	N. Am.	S. Am.	Total
Languages	478	170	301	60	81	92	1182
Families (sample/total)	28/53	28/38	63/126	16/33	36/75	57/110	228/421
Coverage	53%	74%	50%	48%	48%	52%	54%

square goodness of fit for number of top-level families in the sample compared to the total number of top-level families found on Glottolog returns a p-value of 0.5778. The greatest deviations from the expected are found in Eurasia and Papunesia; the number of sampled families in Eurasia is 8 more than expected, and for Papunesia it is 3 less than expected. Although the sample is not perfect (a sample with perfect proportions would result in the p-value 1.0), a p-value of 0.5778 is far above the threshold of 0.05, and the proportion of languages per macroarea is not statistically different.

The article starts with the overview of morphosyntactic properties of demonstrative verbs in § 2. § 3 forms the core of the article, and presents the semantic classification of demonstrative verbs based on their use in exophoric function. Discourse uses of demonstrative verbs are addressed in § 4, whereas various types of non-deictic uses are explored in § 5. The final conclusions are drawn in § 6.

2 The morphosyntax of demonstrative verbs

In this section we explore the morphosyntactic properties of demonstrative verbs, focusing on their form (§ 2.1) and function (§ 2.2).

2.1 Morphosyntactic form and behavior

Demonstrative verbs are typically at least bi-morphemic, composed of a deictic combined with the verb root (base). In Nêlêmwa for instance, the verb root *shum(a)* 'be like' combines with four different deictic markers: *winy* (proximal), *wena* (medial), *weli* (mentioned previously in discourse), and *bai* (known from shared experience) (Bril 2004: 107).

Crow (Siouan) works similarly, in that demonstrative roots "[…] combine with the locative verb *la* 'be at' to derive locative verbs meaning 'be here, be

there', etc. These are stative verbs, and they are inflected as ordinary statives […]" (Graczyk 2007: 84).

(1) Crow (Graczyk 2007: 84)

awaxaawé ***íilakaa-la-k***
mountains **DST-LOX.V-DECL**

'The mountains are way over there.'

Demonstrative verbs can sometimes be derived further. Locative verbs in Crow for instance can combine with the direct causative to form verbs meaning 'situate or locate oneself here, there'.

(2) Crow (Graczyk 2007: 86)

kootdák ***éehkoo-n-n-aa-lak*** *chiláakshilak*
all.right **DST-LOX.V-2SG-CAUS-COND** tomorrow
baa-w-asshihk-aát-boo-k *he-k* *huu-k*
INDEF-1SG-consider-APPROX-1PI-DECL say-DECL say.PL-DECL

' "All right, if you take that spot over there, tomorrow we'll consider the matter", she said.'

Halkomelem (Salish) also can derive the verbs *ʔí* 'be here' and *lí* 'be there' to create the verbs *xʷeʔí* 'come here, get here; arrive' and *xʷelí* 'go there, get there'.

Verbs can also be derived by other means. In many languages with demonstrative verbs, deictic roots show partial ambiguity between adverb and verb. In Reta (Alor-Pantar) for instance, there are a number of different demonstrative verb types, including several verbs indicating location, as well as three verbs indicating deictic comparison of amount, size, and height. Verbs indicating location show great flexibility in their functions; in (3) and (4), *gi'e* '(be) here' is used predicatively, but in (5) it is used in more of an adverbial sense.

(3) Reta (Willemsen 2021: 207)

meleng *gang* ***gi-'e***
yesterday 3SG.NOM **PROX-LOX**

'Yesterday he was here.'

(4) Reta (Willemsen 2021: 183)

na-vaal ***gi-'e***
1SG.POS-child **PROX-LOX**

'I have children (lit.: my children exist [proximally]).'

(5) Reta (Willemsen 2021: 166)

boma anu ***gi-'e*** *matee jia*
old one **PROX-LOX** stand placed

'A man is standing here.'

In Balantak (Celebic), adverbial demonstratives in the allative ('thither') may function verbally. As verbs, they take verbal morphology such as aspect and mood in the same way as other verbs do:

(6) Balantak (van den Berg & Busenitz 2012: 191)

noko daa k<um>aan, kai ***no-mbaa-tu'u-na-mo*** *na laigan-na*
after.R finish INTR-eat 1PE **R-ALL.R-DST-3SG-PFV** LOC house-3SG

'When we had finished eating, we went back to the house'

In Northern Subanen it is possible to simply add verbal affixes such as realis markers onto locative adverbial demonstrative bases directly, to make deictic motion and placement verbs, without any other overt morphology needed (Daguman 2004).

(7) Northern Subanen (Daguman 2004: 221)

mə-ditu=ita *sə g=binaal*
ITR.AG.I-DST=1PI.ABS OBL SCM=farm

'Let us go to the farm.'

(8) Northern Subanen (Daguman 2001: 4)

pi-diaʔ-u *sə k-siluŋ su kandiŋ*
CAUS.AG.R-MD-1SG.ERG OBL SCM-below ABS goat

'I placed the goat under the house.'

Table 4. Buryat Demonstrative verb forms (Poppe 1960: 74)

ii-/ii-ge-	'do like this'
ii-že/iige-že	'doing like this' (imperfective gerund)
ii-ge-ed	'having done like this' (perfective gerund)
tii-/tii-ge-	'do like that'
tii-že/tii-ge-že-	'doing like that'
tii-ge-ed	'having done like that'

In Toqabaqita (Oceanic), such lexical flexibility can even occur at the phrasal level with the phrase *quna qeri* 'this/that way' (Lichtenberk 2008). More rarely, this can also be with a different demonstrative than *qeri*. *Quna qeri* is treated as a lexical unit, a verb taking the appropriate subject markers in the same way other verbs do, and in (9) it also shows a type of verbal reduplication indicating iteration.

(9) Toqabaqita (Lichtenberk 2008: 139–140)

qe ***quu-quna qeri*** *qana gwau-na,* *ma* *ifu-na*
3SG.NFUT **RDP-MANNER this** PREP head-3SG.PERS and hair-3SG.PERS
ka *katu* *na=mai* *labaa*
3SG.SEQ hinder PERF=VENT there

'He kept doing like this with his head [the speaker jerks his head several times one way to demonstrate what he was doing], and his hair held fast there'

According to their morphosyntactic behavior, demonstrative verbs may be grouped into a few categories. First, demonstrative verbs may occur as regular lexical verbs in the language, with complete and regular inflection, as in Buryat (Poppe 1960: 74), seen in Table 4.

However, morphosyntactic regularity of demonstrative verbs appears to be more of an exception than a rule, which can be explained by the specific semantics of these lexical items (discussed further in § 2.2 as well as throughout the article more generally). Demonstrative verbs frequently display features of atypical verbal behavior compared to other verbs in a given language, such as constraints on marking of certain morphosyntactic categories, including TAM, voice, and number.

In Halkomelem (Salish) for instance, mentioned previously, there are two locative-existential demonstrative verbs, *ʔí* 'be here' and *lí* 'be there'. These verbs can be inflected for subject, future tense, or subjunctive mood, but past and interrogative suffixes must be attached to a preceding auxiliary, and no continuative, imperative, participle, passive, or pluralizing inflection is possible (Galloway 1977: 350).

Icari Dargwa (Nakh-Daghestanian) has a series of localizing demonstrative verbs, which also show restricted number of forms. Using the proximal *le=b* 'be here' as an example, only the following forms are possible: present tense *le=w–da* (first person), *le=w–di* (second person), *le=w* (third person); participle *le=w-ci*, and converb *le=w-li*. Any other forms are replaced by forms of the verb *=ū* 'to exist' (Sumbatova & Mutalov 2003: 145).

2.2 Morphosyntactic function

In terms of syntactic functions, demonstrative verbs fill a number of roles. Prototypically as verbs, they function as main predicators of a clause, as in (10).

(10) Siar (Frowein 2011: 453)

é	*Pasta*	***a-d-óng***	*ma*	*an*	*piu*	*i*	*tur*	*tar*
ART	pastor	**LOX.V-DEM.SG-CLK**	TRANS	at	ground	3SG	stand	PERF
k-i	*kòlòng*	*laulau*	*tar*					
FOC-3SG	terrified	bad	PERF					

'The Pastor was there outside, he was terribly afraid.'

Other examples come from Mauwake (Madang), Northern Subanen (Greater Central Philippine), Barok (Oceanic), and Korean (Koreanic).

(11) Mauwake (Berghäll 2015: 172)

aa,	*o*	*koora*	***fan-e-k***	*a*
INTJ	3SG	house	**LOX.V.PRX-PST-3SG**	INTJ

'Ah, his house is here.'

(12) Northern Subanen (Daguman 2004: 290)

məkpəd-ditu *masiʔ g=balay g=bədiŋ kiin*
ITR.AG.HAB.I-there MIR SCM=house SCM=cat that

'I'm surprised to know that that cat regularly goes to (our) house.'

(13) Barok (Du 2010: 99)

i=bo ***bi∼biringaan***
3SG.SM=HAB **RDP∼PQ.V.DST**

'He habitually does it like that (making faces).'

(14) Korean (Chingduang Yurayong, p. c.)

A: na socwu-pota maykcwu-lul te cohahay.
A: 1SG rice_wine-COMP beer-ACC more like.PRS.IND
B: na-to ***kulay.***
B: 1SG-also **PQ.V.MED**

'A: I prefer beer to rice wine. B: Me too. (lit: I am also like that.)'

Demonstrative verbs seem to commonly occur in other types of constructions than heads of predicates, however. In Musqueam, demonstrative verbs can be nominalized, functioning as arguments of predicates.

(15) Musqueam (Suttles 2004: 428)

wəl-híθ k'ʷə ***nə-s-ʔí*** *ʔə tən'a*
already-last.long ART **my-NMLZ-LOX.V.PRX** OBL this

'My being here has lasted long.'

One very common construction type that demonstrative verbs occur in is multi-verb predicate constructions, or "serial verbs". In such constructions, they are structurally verbs, but they often show distinct adverb-like uses. Kratochvíl (2007: 103) mentions that in Abui for instance, locative-existential demonstrative verbs primarily occur in serial verb constructions, and only rarely are they inflected for aspect or person. In such constructions they are very adverb-like, both in form as well as semantically.

(16) Neve'ei (Musgrave 2007: 128)

utnen i-rong ***Ø-menenang*** *i-mera' i-ngang*
when 3SG.R-hear **3SG.R-PQ.V.DST** 3SG.R-INCP 3SG.REAL-laugh

'When he heard it like that, then he laughed.'

(17) Kavalan (Jiang 2009: 3)

nayau=ti *ya bai-bai-ta m-Rimazuq*
PQ.V.MD=PFV INTJ RDP-grandmother-1PI.GEN AF-foolish

'Our ancestors were foolish like that.'

(18) Siar (Frowein 2011: 366)

é tata ***á-d-óng*** *ma is*
ART daddy **LOX.V-DEM.SG-CLK** TRANS return

'Daddy was there now returning.'

Rundi (Bantu) is a language in which demonstrative verbs marginally occur, only in the most restricted sense. Demonstrative verbs do not occur as heads of simple predicates at all, instead occurring only as modifiers of a head in a complex predicate. However, they nonetheless have been considered verbs by previous researchers, as they take verbal agreement marking, seen in (19) below.[4]

(19) Rundi (Rodegem 1967: 78)

ba-vuga ***bá-tyo***
CL2-speak **CL2-PQ.V.DST**

'They speak like that.'

Languages may also use derived forms of demonstrative verbs to function as (adverbial) modifiers, as converbs which function somewhat similarly to the serial verb construction. In Urarina (isolate) for instance, the verb *nitoania* 'be like that' overwhelmingly occurs with the "participle" suffix *-ĩ*, and rarely with other types of inflection (Olawsky 2006: 798–799).

[4] Translated from the original French, with glosses added.

(20) Urarina (Olawsky 2006: 798)

nitoaneĩ *hetau=te ɲukwana ha-ɯre-rehete ɲoaelɯ katɛa-ɯrɯ*
PQ.V.DST.PTCP HRS=FOC 3=field make-PL.HAB:3 earlier man-PL
aherɨ+ku-teri ke kufwihja-ĩ
stone+ASC-axe INST fell-PTCP

'The ancient people made their fields like that, felling [trees] with stone axes.'

Korean demonstrative verbs show polysemy between 'to say' and 'to do'. To specify the meaning of 'do' rather than 'say', speakers can use derived converbs together with the verb *ha* 'to do'.

(21) Korean (Chingduang Yurayong, p. c.)

wuli-nun manna-l sikan-i eps-umyen ***ileh-key***
1PL-TOP meet-PTCP.FUT time-NOM not_exist-COND **PQ.V.PROX-ADV**
ha-ca
do-HORT

'If we don't have time to meet, let's do like this (as I suggested earlier).'

Demonstrative verbs are also frequently used in various types of clause linking. Demonstrative verbs in Nuaulu (Nunusaku) are used as relativizers for relative clauses.

(22) Nuaulu (Bolton 1990: 165)

hiti tihu ***a-po*** *pam rei roe*
lift water **LOX.V-DOWN** pot this up

'Lift up the water that is in the pot.'

In Aguaruna (Chicham), demonstrative verbs are commonly used in clause linking, with the demonstrative verb anaphorically referring to the previous clause, indicating a temporal or causal relationship between it and the following clause.

(23) Aguaruna (Overall 2007: 410)

waamaki *hu-hu-ki-ta-humi* *waha-a-u*
quickly take-1SG.OBJ-TRF-IMP-2PL call-IPFV-REL
nuni-taĩ-ʃakama *antu-ka-tʃa-aha-u*
PQ.V.DST-SBD:1/3:DS-CONCESS listen-INTS-NEG-PL-REL
a-ina-wa-i
COP-PL:IPFV-3-DECL

'He was calling out, "quickly take me away!" Although he did that, they didn't listen.'

In Yurakaré (Yuracaré) and Nivkh (Amuric), converbs derived from demonstrative verbs are used as sentence connectives which can express a resultant relationship between the two events.

(24) Yurakaré (van Gijn 2006: 285)

na *ishete* *buybu* *ka-n-dyuju-ø=ti* *lash* ***achama*** *lëtta* *dia* *nish*
DEM agouti word 3SG-IO-tell-3=DS then **PQ.V.DST** one day NEG
wita-ø=ya
arrive.SG-3=NVR

'When the agouti told him the news, he did not come home for a day.'

(25) Nivkh (Amur) (Gruzdeva 2020: 54)

luvr *me-qr+ŋi-ra* ***hoʁo-r*** *kʰrəz-ra*
spoon two-CLF+eat-COORD:3SG **PQ.V.DST-CVB:3SG** be.full-COORD:3SG

'[He] ate two spoons, then [he] was full.'

Identification demonstrative verbs in Makalero (Timor-Kasar) are also used for clause linking, in particular adverbial "reason" clauses.

(26) Makalero (Huber 2011: 119)

uere'=konai=ni *ani* *esperensia* *ere* *la'ane'* *konta* *[...]*
IDT.V.ADDR=CSQ=LNK 1SG experience DEM.PRX various tell

'that is why I tell these experiences'

For more information on demonstratives functioning as clause connectors, see Diessel & Breunesse (2020) and Guérin & Aiton (2018).

3 Semantic categorization of demonstrative verbs

In this section, we propose a semantic classification of demonstrative verbs based on their use in an exophoric function. The notion *exophoric* refers to entities physically present in the speech situation, located in the speaker's sphere of perception (Halliday & Hasan 1976: 57–76). Demonstrative verbs frequently indicate events (or locations) rather than participants, so the "entity" in this case is the event that is present in the speech situation. Being used exophorically, demonstratives serve a language-internal function — they focus the hearer's attention on entities in the speech situation and are characterized by a deictic reference to an activity, either actual or mimicked (Dixon 2003: 72). In the course of discussion, note that we do not typologize spatial oppositions found in demonstrative verbs. Levinson (2018: 19) points out that "[...] proximity is an elastic notion, and according to each language, it has different extents depending on multiple pragmatic factors". Too few grammatical descriptions take into account all the myriad differences that play a role in spatial oppositions, and how they are used in different contexts. The distinction between person-oriented and distance-oriented systems has been frequently discussed in cross-linguistic research on demonstratives (see e.g. Diessel 1999), but it remains unclear as to what extent three-term person-oriented systems for instance have been misanalyzed as egocentric distance-oriented systems.

Furthermore, many of the systems discussed here are rather heterogeneous, and defy easy categorization into a simple distance extending out from the ego. Nuaulu for instance creates existential demonstrative verbs by combining the root *wai* with a variety of clitics indicating proximity and direction: *mai* 'here', *kua* 'around here', *nau* 'seaward', *noi* 'unspecified direction', *pani* 'across', *poe* 'down', *ria* 'inland', *roe* 'up', and *hae* 'on' (Bolton 1990). Comparing this type of complex deictic system involving topographic or elevational notes with other types of complex systems is not trivial.

From a semantic point of view, demonstrative verbs can be classified into three major categories: locative-existential demonstrative verbs (§ 3.1), processive-qualitative demonstrative verbs (§ 3.2), and movement and placement demonstrative verbs (§ 3.3). Additionally, there are some minor categories

referring to identification, size, and speech (§ 3.4). In some cases, it is difficult to assign demonstrative verbs to a certain class, as will be shown in § 3.5.

3.1 Locative-existential demonstrative verbs

The first major semantic class is that of locative-existential demonstrative verbs, used to indicate the location of a referent relative to the deictic center. They serve to localize a participant or event in a certain space, establishing a figure-ground relationship. Such verbs are typically translated into English as 'be, exist here/there'.

Examples from Mauwake and Siar are given below.

(27) Mauwake (Berghäll 2015: 172)

No	*ikiw-e,*	*irak-owa*	*maneka*	***fan-e-k***	*a*
2SG	go-IMP.2SG	fight-NMLZ	big	**LOX.V-PST-3SG**	INTJ

'Go (home), the big war is here.'

(28) Siar (Frowein 2011: 232)

ép	*kirai*	*n-a*	*ép*	*lakman*	***a-d-óng***	*sén*	*an*
ART	time	DEM.SG-PRX	ART	village	**LOX.V-DEM.SG-CLK**	EMPH	at

Kingén
Kingén

'That time the village was further north, at Kingén.'

Existence and location share a strong semantic relationship. Creissels (2019: 38) explains that "[...] the semantic relationship between existence (in the usual sense of 'being an element of the world') and location follows from the fact that, for concrete entities (but only for concrete entities!), *X is an element of the world* is equivalent to *X can be found somewhere in the world*". Although Guérin (2015) was uncertain of the validity of this semantic type, locative-existential demonstratives in this study are well represented, with approximately 40 languages in the database showing locative-existential demonstrative verbs. Let us consider several examples.

In Momu (Baibai-Fas), there are two locative-existential verbs distinguishing proximal and distal usages. These locative-existential verbs not only express the relationship between the figure and the ground, but also include

Table 5. Lumun demonstratives (Smits 2017: 392)

C-εí	'be here (near speaker)'	ɛn-C-í	'this, these (near speaker)'
C-êrík	'be there (near addressee)'	ɛn-C-ərík	'that, those (near addressee)'
C-éɽê	'be over there (away from speaker and addressee)'	ɛn-C-əɽê	'that, those (away from both, but in sight)'

a further deictic specification of the figure to the deictic center, typically the position of the speaker (Blake 2007: 31).

(29) Momu (Blake 2007: 30)

mi teBu Australia ***yaiwo***
mother 1SG.POSS Australia **LOX.V.DST.AN.3SG**

'My mother is in Australia.'

(30) Momu (Blake 2007: 33)

kaf anu fiki ***nouwo***
cup DEM.PRX near **LOX.V.PRX.INAN.3SG**

'This cup is near me.'

In Lumun (Talodi) the case is similar. There are three demonstrative verbs of location, which parallel adnominal demonstratives in structure and semantics (Smits 2017: 392); see Table 5, and (31) for an example of their use.

(31) Lumun (Smits 2017: 393)

aɽık nú̱ṯ̱ətruk n-əɽɛk ***n-éɽê***
come pigs C-some **C-LOX.V.DST**

'Come, there are some pigs over there!'

The language Thao (Western Plains Austronesian) also shows a series of demonstrative verbs for expressing location: *inay* 'be here', *iutu* and *isa(há)y* 'be there', *isu(hú)y* 'be over there (visible)', and *itusi/itantu* 'be over there (not visible)' (Wang 2004: 303). Furthermore, derivational morphology on

locative-existential demonstrative verbs is possible, such as *mu-* to express movement such as come/go, or *pi-*/ *pu-* to convey the idea of putting something somewhere, i.e. *inay* 'be here', *mu-nay* 'come here', and *pi-nay* 'put here'.

(32) Thao (Blust 2003: 399, cited in Wang 2004: 303)

rumfaz	***in-inay***	*marfaz*	*makitnahazish=iza*
bird	**PFV-LOX.V.PROX**	fly	go.gradually.away=already

'The birds were here, but have flown off into the distance.'

Biak (South Halmahera-West New Guinea) has two sets of demonstrative verbs, one which carries a meaning of locative-existential, and a second for identification. Both sets of verbs allow for a number of different demonstrative form combinations, including directional marking, anaphoric marking, and topographic reference.

(33) Biak (Steinhauer 2005: 817)

mko-is-ya-m-ra
2P-LOX.V-MD-VEN-SEA

'You (PL) are on your way towards the sea (where I am).'

In Makalero (Timor-Kisar), the addressee-centered verb *ue'* 'be there (near addressee)' is the most widely used demonstrative verb, including being used for general reference and existential predication.

(34) Makalero (Huber 2011: 182)

Huma'	*ni'isi*	*uari*	***ue'***
soul	simultaneous	still	**LOX.V.ADDR**

'Ghosts really exist.'

Eskimo-Aleut languages show a large number of verbs involving demonstratives, most of which deal with location and movement. In Central Alaskan Yupik (Eskimoan) for instance there is a basic stative verb for being in a location, *+m(i)t- / +n(i)t-*, which combines with locative nouns and adverbial demonstrative roots. Another stative derivation is the verb *-qsiɣ-* 'to be in the

direction', which follows a similar pattern to *+m(i)t- / +n(i)t-* in that it also largely occurs with locative nouns or adverbial demonstrative roots.[5]

(35) Central Alaskan Yupik (Miyaoka 2012: 368)

ma-a-nt-ukut
PRX-EXPND-LOX.V-IND.1PL.LOC

'We are here.'

(36) Central Alaskan Yupik (Miyaoka 2012: 368)

aci-qsig-aa	*May'a-m*	*eni-i*
AREA.BELOW-BE.IN.DIRECTION-IND.3SG	Mayaq-GEN	house-ABS.3SG

'Mayaq's house is far below.'

A subtype of exophoric use exists known as *Deixis am Phantasma*, or *imaginary deixis*, following Himmelmann (1996: 222). In such instances, the perspective shifts from the utterance situation to the narrated situation. Although this use has not been studied in depth for demonstrative verbs, this perspective shift is also possible, seen in the Momu examples below. In the case of Momu, Blake specifies that the question was posed outside Antonia's house, and Speaker B was not close to Antonia.

(37) Momu (Blake 2007: 31)

A:	*Antonia*	*ai-ta?*	*B:*	***naiwo,***	*fi*	*pə*
A:	Antonia	stay-3SG.INTERR	B:	**LOX.V.PRX.AN.3SG**	water	go.SG

momu
not

A: 'Is Antonia there?' B: 'She's there (lit.: here), she hasn't gone to the water yet.'

Locative-existential demonstrative verbs are sometimes used to indicate possession, as is common for regular lexical verbs. Such use fits what has Stassen has labeled as a locational possessive (Stassen 2009). In Dargwa (Dagestanian), Puyuma (Austronesian), Makalero, and Kambera (Bima-Lembata) for

[5] No glosses were provided in the original for (36).

instance, locative-existential verbs are described as being used in possessive predication constructions, with the possessor often in a genitive or locative case.

(38) Dargwa (Icari) (Sumbatova & Mutalov 2003: 146)

di-la *mašin* ***te=b***
1-GEN car **LOX.V.DST=NEUT.SG**

'I have a car.'

(39) Kambera (Klamer 1998: 150)

Ningu *kabela* *lai* *nyuna*
LOX.V.PRX machete LOC he

'He has a machete.' (lit.: 'There is a machete with him.')

Nuaulu also may use demonstrative verbs to mark general existential predication as well as predicative possession.

(40) Nuaulu (Bolton 1990: 104)

Sona *penne* *a-ro* *mainae?*
sago.paste cold **LOX.V-ABOVE** much

'Is there a lot of cold sago paste?'

(41) Nuaulu (Bolton 1990: 108)

Ami *rua-ma* *mani* *akama* ***wa-n***
we self-1PE our religion **LOX.V-PRX**

'We have our own religion.'

Locative-existential demonstrative verbs are also sometimes used as markers of ongoing activity, giving a durative or progressive reading. In Crow (Siouan) for instance, there is a construction where *koolá* is suffixed to another verb, to refer to ongoing activity (Graczyk 2007).

(42) Crow (Graczyk 2007: 85)

is-bilaxpáake *baaík-shii-ak* *ihch-iiwaaíaschili-k*
POSS.3-people things-say-SS REFL-sell-DECL

huu-koola-k
say.PL-LOX.V.DST-DECL

'[Plenty Coups] people were saying things; they kept saying that he had sold himself (sold out).'

Similar constructions are found in Dargwa, where the "locative copulas" are used for progressive or durative meanings, e.g. in Ashti Dargwa (Belyaev 2012):[6]

(43) Dargwa (Ashti) (Belyaev 2012: 196)

uᶜq'-uᶜn ***ti-w***
[M]go.IPFV-CVB **LOX.V.DST-M[3]**

'(He) is coming' (removed from the speaker, at the moment).

(44) Dargwa (Ashti) (Belyaev 2012: 196)

uᶜq'-uᶜn ***li-w***
[M]go.IPFV-CVB **LOX.V.PRX-M[3]**

'(He) is coming' (next to the speaker, at the moment).

Belyaev mentions that locative copulas in combination with other verbs specify that the act has a longer duration; such uses are mirrored in Blagar, where "their [demonstrative verbs] function is not only to localise the referent of the rest of the (part of the) predicate to which they belong, but also to express that that event is of some duration" (Steinhauer 1991).

Locative-existential demonstrative verbs overlap heavily in function with non-verbal predicative localizing demonstratives, discussed in Killian (2022b). Both categories appear to show a preference for restricting TAM marking, but may allow argument indexation or other verbal forms such as converbs. This semantic category appears to be the most difficult to draw conclusions

[6] Translated from the original Russian.

on differentiating verbal from non-verbal predicative demonstratives, and will require careful research.

Tagalog is an interesting case which shows at least 3 semantic types of demonstrative predicators, 2 non-verbal and 1 verbal, and historical evidence suggests two additional uses (existential and placement) which have now largely left the language. Table 6 shows demonstrative verbs and non-verbal predicative demonstratives in Tagalog. All forms except the presentative are derived from the oblique "*sa*" form, and it is interesting to note the part of speech ambiguity and variation among these derived demonstratives.

Oblique forms with meanings of 'here, there' are also included despite showing little synchronic evidence for a predicative nature, as historically they also allowed for verbal morphology, e.g. *d<um>oon* 'take (your) place there, situate (yourself) there'. Demonstrative verbs of placement ('put (t)here') have been largely lost, but survive in earlier descriptions. The distinction between 'near speaker' and 'near speaker and hearer' has also been lost in Standard Tagalog, but is still used in other dialects.

Tagalog presents an interesting case of grammaticalization, which leads one to wonder: was the non-verbal localizer originally verbal, and subsequently lost its verbal properties? Or did the non-verbal predicator derive directly from a verb? In either case, it's highly suggestive of the fact that predicators which express static location generally disprefer TAM marking, suggesting their non-verbal nature.

One distinction between verbal and non-verbal localizing/locative-existential demonstratives does appear to be the semantic extension towards existence that demonstrative verbs frequently show. Non-verbal localizers do not seem to allow for the expression of pure existence such as 'ghosts exist' (with the possible exception of archaic Tagalog seen in Table 6), nor do they allow for bounded existence, such as 'Lions exist in Africa'. Furthermore, whether any language with localizing demonstratives is able to make the distinction between what Creissels labels as *inverse locationals* (there is a book on the table) and *plain locationals* (the book is on the table) is currently unclear.

Locative-existential demonstrative verbs, in contrast, show more flexibility in their constructions; Makalero for instance was shown in (34) using demonstrative verbs for pure existence.

Table 6. Tagalog demonstrative forms

	Oblique	Presentative (nv)	Localizer (nv)	(Existential) (nv)	Movement (v)	(Placement) (v)
Near speaker	dine / rine	ere	nandine / narine	(mayrine / maydine)	parine	(magrine)
Near speaker / hearer	dito / rito	(h)eto	nandito / narito	(mayrito / maydito)	parito	(magrito)
Near hearer	diyan / riyan	(h)ayan	nandiyan / nariyan	(mayriyan / maydiyan)	pariyan	(magriyan)
Remote	doon / roon	(h)ayon	nandoon / naroon	mayroon / maydoon	paroon	(magroon)

3.2 Processive-qualitative demonstrative verbs

Processive and qualitative demonstrative verbs refer to a process or quality with reference to the deictic center. Both meanings are frequently marked by the same demonstrative verb, and depending on the context can be translated as 'do like this/that' (processive) or 'be like this/that' (qualitative). Even for languages which have two different forms, it can be difficult to differentiate. Korean has two distinct series of demonstrative verbs, processive verbs 이러다 *ileta*, 그러다 *kuleta*, 저러다 *celeta* 'do (like) this/that', and qualitative verbs 이렇다 *ilehta*, 그렇다 *kulehta*, 저렇다 *celehta* 'be like this/that'. However, in practice the difference seems to be often blurred. Due to phonological reasons, the verbs show overlap in some of their conjugations, and it can be difficult to tell whether a form like the declarative 그래 *kulay* comes from the processive 그러다 *kuleta* or the qualitative 그렇다 *kulehta*.[7]

Processive-qualitative demonstratives are among the most researched of demonstrative verbs, discussed for example in Dixon (2003), and in Guérin (2015) under the label of manner demonstrative verbs. Around 50 languages in the sample show processive-qualitative demonstrative verbs.

Processive-qualitative demonstrative verbs may directly refer to a concrete action, seen in examples below from Dyirbal (45) and Vaeakau-Taumako (47). They may also refer to a gesture mimicking an action, as in example (46). In both cases the demonstrative verbs are used in a context typical of other proximal demonstratives.

(45) Dyirbal (Dixon 2003: 102)

bala	*baja!*	*ŋaja*	***yalama-n***	*baja-n*
there:N	chew:IMP	1SG	**PQ.V.PRX-TR.NFUT**	chew-NFUT

'Chew it [the spear grass]! I'm chewing (it) like this.'

(46) Dyirbal (Dixon 2003: 101)

ginya-ginya	*garrgal*	*yuba-n*	***yalama-n***
this:N-this:N	arm	put.down-NFUT	**PQ.V.PRX-TR.NFUT**

'These arms were put down (stretched out) like this.' (narrator mimes what was done)

[7] We thank 허진 for their kind assistance with understanding Korean demonstrative verbs.

(47) Vaeakau-Taumako (Næss & Hovdhaugen 2011: 128)

Noho ***phe-nē****!*
sit **PQ.V-PRX**

'Sit like this!'

Processive-qualitative demonstrative verbs can also express comparison of equality, seen in East Uvean (Oceanic) (48).

(48) East Uvean (Moyse-Faurie 2019: 145)

Kua ***fēia*** *te lahi o te tamasi'i*
PFV **PQ.V.DIST** SPC tall POSS SPC child

'The child is that tall' (with a movement of the hand to indicate the size)

The following examples in Nêlêmwa (Oceanic) show the proximal and distal processive-qualitative verbs functioning as main predicators; note that Nêlêmwa has two additional demonstrative anaphoric processive-quality verbs.[8]

(49) Nêlêmwa (Bril 2002: 285)

co ***shu-mwela*** *i na me da?*
2SG **PQ.V-DST** REL 1SG SUBORD what?

'Why are you doing that to me? (why are you behaving like that?)'

(50) Nêlêmwa (Bril 2002: 285)

shu-mwiny *wany bai io i iyulî*
PQ.V-PRX boat that.ANPH FUT 3SG buy.TR

'It will be like this the boat he will buy.'

There are two processive-qualitative verbs in Eibela (Bosavi), which may be used exophorically as well as in discourse.

[8] Translated here from the original French.

(51) Eibela (Aiton 2016: 287)

sɛ:-ja:	*to:bo:*	*oma:ni:-jɛ:*	*o:*	*go:la:*	***wo:ga:***	*kɛi*
bank-ABS	all	blood-LOC	lake	pool	**PQ.V.ANPH:PST**	ASSER

'All the sand was like an entire lake of blood.'

Kavalan (East Formosan) also has processive-qualitative verbs. Such verbs in Kavalan frequently show adverbial use.

(52) Kavalan (Jiang 2006: 120)

mana	***nayau-an-su***	*biyat-ku*	*zin-na*	*sunis*	*'nay*
why	**PQ.V.MD-LF-2SG.GEN**	frog-1SG.GEN	say-3SG.GEN	child	that

'The child said, "Why did you do that to my frog?" '

(53) Kavalan (Jiang 2009: 3)

nayau-an-ta	*k<um>tun*
PQ.V.MD-LF-1PL.GEN	<AF>chop

'We chop (banana trees) like that.'

Similar adverbial uses of processive-qualitative demonstrative verbs are also found in Kalmyk (Mongolic).[9]

(54) Kalmyk (Bläsing 2003: 239)

en	*ködlmsh-ig*	***iig-j***	*ke-x*	*kerg-tä*
this	work-ACC	**PQ.V.PRX-CVB.IPFV**	do-PTCP.FUT	work-ASC

'This work has to be done in this way.'

3.3 Movement demonstrative verbs and placement demonstrative verbs

Demonstrative verbs of placement and demonstrative verbs of movement refer to the corresponding processes which take place in relation to the deictic center, and can be conventionally translated as 'move here/there' or 'place here/there'.

[9] No glosses were provided for the Kalmyk examples in the original.

Table 7. Aklanon verbs of motion (de la Cruz & Zorc 1968: 116)

Root	Meaning	Verb	Meaning
iya	near speaker	ariya	'come here' (to speaker)
una?	near addressee	aruna?	'go to addressee'
inha?	near addressee, far from speaker	arinha?	'go there'
idto	removed from both	adto	'go (away)'

Unlike locative-existential verbs and processive-qualitative verbs, movement demonstrative verbs and placement demonstrative verbs show no functional overlap. The two types of demonstrative verbs are rather grouped together in a single section, simply because they frequently occur together in the same language. Out of 15 languages in the database with verbs of either movement or placement, 11 languages have derivations both for movement as well as placement; 4 languages have movement without placement, and 2 languages have placement without movement.

Demonstrative verbs of movement or placement are rarer than the other two types of demonstrative verbs, and more restricted. They are more commonly found in languages of the Philippines, Taiwan, and Indonesia. Note that for practical reasons, we exclude any languages as having demonstrative verbs of placement or movement if they show no clear morphological connection to demonstratives of other categories.

In many Bisayan languages of the Philippines for instance, basic verbs of coming and going are derived from demonstrative based with a *ʔa-* or *ka-* prefix, e.g. Hiligaynon *karí* 'come here (near speaker)', *karáʔ* 'go there (to addressee)', *kádto* 'go yonder (removed from both)' (Zorc 1977). Table 7 and example (55) illustrate such forms in the Bisaya language Aklanon. In Aklanon, the prefix to derive verbs of motion is *(a)d/r-*, *d* sometimes alternating with *r*.

(55) Aklanon (de la Cruz & Zorc 1968: 116)

pa-ar-iya-ha	*gid*	*imáw*
CAUS.OF-MOVE-PROX-IMP	EMPH	3SG

'Make him come here.'

In many languages movement and placement demonstrative verbs are derived from demonstrative verbs of location. Thao was previously mentioned in § 3.1

as showing derivational forms for placement and movement with demonstrative verbs, seen in Example (56).

(56) Thao (Wang 2004: 171)

Numa ***pu-sáy-in*** *baruku, lhimpania'anin,* ***pu-tu-an*** *shnir,*
then **PUT.V-DST-TR** bowl mix.vegetables **PUT.V-MD-TR** soup
qtilha numa suksuk.
salt and ginger

'Then (we would) put them in a bowl, mix (them) with vegetables, and put soup, salt and ginger in (them).'

Takivatan Bunun makes use of similar derivations, such as *pa-* 'dynamic causative marker', *mun-* 'go (to)', and *pan-* 'arrive'.

(57) Takivatan Bunun (Austronesian) (de Busser 2009: 338)

pa-ʔiti *aipun*
CAUS.DYN-PROX.LOC DEM.SG-MD

'He was sent here [to understand how we Bunun live].'

(58) Takivatan Bunun (de Busser 2009: 493–494)

tuqas istun ***ita-ʔa*** ***pan-maka-ʔita-in***
older.sibling 3SG.MD **LOC.DIST-SUBORD** **ITIN-ROUTE-LOC.DIST-PFV**

'Her older sister was also there, and she had come over.'

Biak shows a largely compositional demonstrative system in which there are 3 speaker-based distances which combine with positional/directional and motion markers. The same deictic roots underlie the demonstrative verbal system as well as the demonstrative pronominal system.

(59) Biak (Mofu 2005: 255)

inai be-y-dya ***i-is-ne***
daughter POSS-3SG-DET.SG **3SG-LOX.V-PRX**

'His/her daughter is here.'

Table 8. Kambera demonstrative verbs

Root	Meaning	Verb	Meaning
ni	near speaker	ni-ngu	'be here' (at speaker)
na	near addressee	na-ngu	'move towards addressee'
nàmu	motion towards	nàmu-ng	'move towards speaker'
nàhu	motion from	nàhu-ng	'move away from/go past'

(60) Biak (Mofu 2005: 51)

i-is-ne-m-pur
3SG-LOX.V-VENT-BACK

'He/She/It is moving towards my back.'

Kambera (Central Malayo-Polynesian) has four deictic verbs, one static locative-existential verb and three verbs of movement (Table 8). Note that there is no distal equivalent to *ningu*, nor does there appear to be a verb originating from demonstrative roots with the meaning of 'come to speaker'.

Only two languages, Makassarese and Nivkh, are described as showing demonstrative verbs of placement, without movement, seen in Makassarese (61).

(61) Makassarese (Jukes 2006: 197)

ku=pa-anjoreng=i *anu=nna* *ri* *balla'=na*
1SG=CAUS-DST.LOC=3 INDEF=POSS.3 PREP house=POSS.3

'I put his stuff there in his house.'

Central Alaskan Yupik, was mentioned previously in § 3.1 as having some verbs which derive from demonstratives, and two verbs of location were mentioned. There are additional verbs related to placement and movement, derived from demonstrative stems, shown in Table 9, and an example seen in example (62).

(62) Central Alaskan Yupik (Miyaoka 2012: 379)

uk-a-var-tuq
VENT-EXPND-MOVE.FORWARD-IND.3SG

'He is moving this way.'

Table 9. Central Alaskan Yupik demonstrative verbs (Miyaoka 2012)

-(q)vaɣ-	'to move/put forward'
(t)muɣc- and +viɣc-	'to go to/toward'
xuiɣ- ∼ +kuiɣ-	'to go through, by way of'
ŋu-	'to be'
n(i̇)t- (or +m(i̇)t-)	'to be at'
ŋiɣ-	'to move over/through'

In their semantics, demonstrative verbs of movement are close to non-demonstrative motion verbs with meanings of 'come' and 'go'; this comparison is rather problematic, however, as there is no current consensus on whether any universality exists for basic meanings of 'come' and 'go' (cf. Wilkins & Hill 1995).

Nakazawa (2007) for instance defines verbs such as 'come' and 'go' as 'MOVE TOWARD a point which is the location of the speaker' and 'MOVE TOWARD a point which is not the location of the speaker', respectively. Such definitions already point towards spatial deixis, so distinguishing 'come' from 'come here' on a cross-linguistic level may not even be possible. Types of variation such as the addressee effect (shifting the ground or deictic center to the addressee) are not universal. Further research is required on whether any cross-linguistic generalizations could be made to distinguish demonstrative verbs of movement from other basic motion verbs.

Aklanon was mentioned at the beginning of this section as having demonstrative verbs of motion with deictic direction. In the related language Cebuano, in contrast, despite superficial similarity, verbs appear to have generalized the meaning to a large extent, and no longer retain an inherent location. It is fairly common to use adverbial demonstratives immediately after the verb, e.g. *ari diri* 'come here', or *adto didto* 'go there'. Furthermore, in some dialects, *(ng)adto* has actually generalized to the point that it can mean any direction, including towards the speaker, e.g. *moadto ka diri* 'go here' instead of *moari ka diri* 'come here'. Cebuano thus does not have demonstrative verbs of motion, synchronically.

Although more research is needed in this area, we nonetheless present languages we are aware of with demonstrative verbs of this type, with the caveat that some languages may end up being like Cebuano, having generalized and grammaticalized away from the demonstrative meaning.

3.4 Minor classes of demonstrative verbs

Minor classes of demonstrative verbs, which are attested only in a few languages, include identification (§ 3.4.1), dimension (§ 3.4.2), and speech (§ 3.4.3).

3.4.1 Identification demonstrative verbs

Identification demonstrative verbs are used to identify a referent in the speech situation and can be conventionally translated as 'be this/that one'. The prototypical use of such is when the predicate is not a property but rather an entity, expressed with a spatial deictic in an equational predication. Verbal demonstratives of identification are rare, occuring in only three languages in the database: Biak, Ju|'hoan (Kxa), and Makalero; they are also found more marginally in Kokota (Oceanic). Despite the rarity, these lexical items do nevertheless behave as verbs.

The following examples (63) and (64) demonstrate the use of identification demonstrative verbs as main predicators focusing attention on a referent in the immediate situation.

(63) Makalero (Huber 2011: 184)

Dotor ini leu ere=ni ***ere'***
doctor 1PE call PRX.DEM=CTR **IDT.V.PRX**

'The doctor that called us is this one.'

(64) Biak (van den Heuvel 2006: 313)

Piet, ***aw-s-i-yás-ya****?*
Piet, **2SG-PRED-SPC-UP-MD**?

'Piet, are you the one (lit.: that one) up there?'

Makalero (Timor-Kisar) shows a series of identification demonstrative verbs which are derived regularly from adnominal demonstratives by means of adding a word-final glottal stop, e.g. *ere* 'this' (near speaker), *ere'* 'be this' (near speaker).

(65) Makalero (Huber 2011: 184)

Lapis ere=ni ani-isi' ei-isi'=ini ***umere'***
pencil PRX.DEM=CTR 1SG-belong 2SG-belong=CTR **IDT.V.DST**

'This pencil is mine, yours is that one.'

Biak has two types of demonstrative verbs, one for locative-existential constructions mentioned previously, and a second formed from pronouns, used for identification.

(66) Biak (van den Heuvel 2006: 382)

i-ne ido, nk-ór ve vrim=ya ***is-i-ne***
3SG.SPC-PRX THM 1PE-call as tuber=3SG.SPC **3SG.PRED-IDNT-PRX**

'As for this, what we call *vrim* is this here.'

(67) Biak (Mofu 2005: 69)

Roi be-fandun kaku-ya ***i-so-ine***
thing REL-need indeed-DET.SG **3SG-IDNT-PRX.SG**

'The thing that you really need is this.'

Ju|'hoan (and to a lesser extent other Ju varieties) shows similar types of verbal demonstratives of identification.

(68) Ju|'hoan (Dickens 2005: 49)

jù ***hè***
person.1 **PRX 1\3**

'This is a person.'

(69) |'U (Lionnet 2014: 190)

me n|ee ti ***e***
1SG head IPFV **PRX**

'This is my head.'

However, the most frequent use of demonstrative identification verbs does not appear to be that of a main predicate. This appears to be the case in both Ju|'hoan as well as Makalero at least; whether identification demonstrative verbs have other uses in Biak remains unclear. In Ju|'hoan, verbal demonstratives appear to be most commonly used as adnominal modifiers in a relative clause construction; note that Ju|'hoan does not have distinct exophoric adnominal demonstratives.

(70) Ju|'hoan (Pratchett 2017: 160)

tjù=à ***kè*** *g|àʊ̈h*
house.3=REL PRX.3 strong

'This house is strong.' (lit. the house that is this one is strong)

(71) Ju|'hoan (Pratchett 2017: 77)

!'ú=à ***kè*** *ó* *!'u* *jä̀qn*
bone.3=REL **PRX.3** COP bone.3 nice

'This bone is a nice bone.' (lit. the bone that is this one is a nice bone)

In Makalero, the most frequent use of the identification verbs is not exophoric, but rather as an argument, to recapitulate an object mentioned in an earlier clause (Huber 2011).

(72) Makalero (Huber 2011: 185)

Meestri sa'a kerek uatu ki=atanana ma'u ere=si
teacher thing write day ATTR=first come PROX.DEM=LNK
meestri ue-kerek-ini=ni taure-fani' ani
teacher LOX.V.ADDR:RED-write-do:BD=LNK which:RED-be.like 1SG
uere' *me'e mi-kerek*
IDT.V.ADDR able along:RED-write

'What the teacher wrote down, (even) on the first day, I was able to copy that.'

Kokota, the remaining language to be mentioned here, shows structural parallels in having an existential verb *au* combined with demonstrative enclitics to

indicate identification.[10] This construction is more commonly used, however, as a single-word subordinating clause, e.g. *t-au-ao* 'that which is this' and *t-au-are* 'that which are those'.

Clausal demonstratives in Kokota function both as modifiers (73) as well as arguments (74).

(73) Kokota (Palmer 2009: 70)

ḡ-a	*la*	*hod-i*	*gai*	*kala-na*	*gazu*
GENR-1.SBJ	go	take-TR	1PE	leaf-POSS.3SG	wood

t-au-na
SUBORD-exist-PRX.SG

'We take the leaves of the tree which is this one.'

(74) Kokota (Palmer 2009: 71)

ara-hi	*a-ti-ke*	*fufunu-di*	*bo*	***t-au-de***
1SG-EMPH	1.SBJ-NEG-PFV	begin-3PL.OBJ	CTR	**SUBORD-exist-PRX.PL**

'I didn't start these (arguments) (lit: the ones which are these).'

The fact that there are only several languages with identification demonstrative verbs can be explained by the strong tendency to express the relation of identity by non-verbal predicators, which were already mentioned in connection with locative-existential demonstrative verbs.

Our database includes 40 languages with non-verbal predicators used in a similar fashion to identification verbs. In general, non-verbal demonstrative identifiers occur more frequently in equational predicates to identify or present a referent; they also occur as independent sentences by themselves, or function adverbially in verbal clauses. As non-verbal elements they do not usually allow for the morphological marking of tense, aspect, or modality; however, functioning as predicates, they do show agreement marking. Identification demonstrative verbs on the other hand do not seem to be used presentatively, occur less frequently in equational predicates, and do not show adverbial usage in verbal predicates.

[10] Note that demonstrative enclitics appear to be possible to cliticize to any verb, not just *au*, and is the primary reason that Kokota is only a marginal example.

3.4.2 Dimensional demonstrative verbs

Dimensional demonstrative verbs refer to a physical property of the referent, such as size, height, or amount.

Demonstrative verbs referring to the size of a referent occur in four languages in our database, Nivkh, Reta, Hidatsa (Siouan), and Mapuche (Araucanian), and can be translated as 'be this size' or 'be that size'.

(75) Nivkh (East Sakhalin) (Gruzdeva 2020: 43)

cʰi *tʰana+gi+xiz-d=ara?* ***Tuna-d-lu****?*
you:SG what.size+footwear+wear-IND-INTERR **SZ.V.PRX-IND-INTERR**
Tuna-d-lu*?*
SZ.V.PRX-IND-INTERR

'What size is your footwear? Like this? [Or] like this?'

(76) Reta (Willemsen 2021: 178)

geng *'anga* ***po-vaang***
3SG.ACC DIST **BELOW-SIZE.V**

'He there, he's as large as that down yonder.'

(77) Hidatsa (Park 2012: 405)

múá ***hirihgá-ahgá-c***
fish **SZ.V.PRX-RDPL-DECL**

'The fish was about this size.'

(78) Mapuche (Smeets 2008: 321)

iñché *ñi* *tasa* *eymi* *mi* *tasa* ***fante-n-üy***
1SG POSS.1SG cup 2SG POSS.2SG cup **AMT.V.PRX-VRBLZ-IND.3**

'My cup is as big as your cup here.'

Demonstrative verbs of amount refer to the amount of a referent, translated as 'be this much/many' or 'be that much/many'. They occur as distinct forms in only two languages, Hidatsa and Reta (79–80); in Mapuche, size and amount are expressed by the same verb (cf. 78 and 81).

(79) Hidatsa (Park 2012: 405)

mii-siʔawí-ʔa-c
1.STAT-AMT.V.DST-PL-DECL

'There are that many of us.'

(80) Hidatsa (Park 2012: 445)

ii-siʔawí-ʔa-c
INST-AMT.V.DST-PL-DECL

'There are that many!' (speaker indicating the number by using his fingers)

(81) Mapuche (Smeets 2008: 69)

fante-n-mu *ngilla-n*
AMT.V-VRBLZ-INST buy-IND.1SG

'I bought [it] for this much.'

Demonstrative verbs of height are known to occur in one language, Reta. No examples were given in Willemsen (2021).

3.4.3 Speech demonstrative verbs

Demonstrative verbs of speech occur in two languages, Crow and Motuna (Bougainville). Crow has two demonstrative verbs of speech, freely translated as 'say this' or 'say that'. Demonstrative verbs of speech in Crow are primarily used in discourse structuring, discussed further in § 4.

(82) Crow (Graczyk 2007: 87)

íiwahkoo-shii-ak *hinne* *póopahta-chia-sh* *kuu-ák*
that-say-SS this owl-white-DET give-SS

'That is what he said, this White Owl, and he gave it to him.'

Motuna is described as having two demonstrative verbs of speech, but it is unclear whether these verbs can be used in exophoric function. More on Motuna is discussed in § 4.

3.5 Problematic cases

There is at least one case where it is difficult to determine how to categorize the semantics of the demonstrative verb. In Puyuma, demonstrative verbs appear to be ambiguous between location (83), quality (84), and process (85).

(83) Puyuma (Teng 2008: 109)

kaDu=la *na palakuan na ne-nem-a*
LOX.V.DST=PFV DEF.NOM men's.house DEF.NOM RDP-six-NPRS

'There are six men's houses already.'

(84) Puyuma (Teng 2008: 109)

an ***kaDu*** *i, ta=sabung-ay=mu*
if **<INTR>LOX.V.DST** TOP, 1PI.GEN=compensate-TR2=2PL.NOM

'If that is the case, we would compensate you.'

(85) Puyuma (Teng 2008: 109)

an ***kaDini*** *pa-ra-ragan=ta=Diyan i,*
if **<INTR>LOX.V.PRX** CAUS-RDP-erect=1PI.NOM=IPFV TOP
puar=ta i, ka-ra-ruwa Da sa-buLan maku
slow=1PI.NOM TOP ka-RDP-can ID.OBL one-moon tag

'If we build like this, we are slow, maybe it takes a month.'

4 Demonstrative verbs in discourse deictic function

Demonstratives carrying a discourse deictic function do not refer to a location or entity, but rather to a segment of discourse adjacent to the speech act. They help the listener orient themself in the ongoing discourse by linking the clause in which they are embedded to the propositions to which they refer (Diessel 1999; Levinson 1983: 83).

Discourse deictic functions are very common with demonstrative verbs, and demonstrative verbs also frequently either function as or even grammaticalize into sentence connectives. Different semantic groups of demonstrative verbs are typically responsible for certain discourse deictic functions, but there can

be language-specific differences, particularly when demonstrative verbs have multiple meanings. Additionally, marking of direct speech shows considerable diversity in employing different types of demonstrative verbs.

In Unua (Oceanic), proximal processive-qualitative demonstrative verbs are used cataphorically for discourse structuring, signalling that there is more to come.

(86) Unua (Pearce 2015: 568)

jirvaren nge i-vra ***i-mre-n:***
story PRX 3SG-go **3SG-PQ.V-PRX**

'This story goes like this:'

(87) Unua (Pearce 2015: 3)

nebo nga m-i-mo-rav-i ***i-mre-n:***
song COMP REL-3SG-CONT-take-TR **3SG-PQ.V-PRX**

'The song that she was singing went like this:'

Processive-qualitative demonstrative verbs can also be used to summarize descriptions of actions or events, most commonly in the addressee-proximal form or the distal form, as in Vaeakau-Taumako in (88).

(88) Vaeakau-Taumako (Næss & Hovdhaugen 2011: 128)

Ha-hano na a langi e tolu lhatu=no hangaota
RDP-GO.SG DEM.MD COLL day GENR three 3PL=IPVF fish
ai ***phe-nā*** *na*
OBL **PQ.V-MED** DEM.MED

'Then for three days they fished like that (= in a manner previously described).'

Identification demonstrative verbs in Makalero can also be used in discourse, to indicate both the beginning as well as the end of a block of text. In contrast to Unua, however, Makalero appears to use the addressee-proximal form for cataphoric reference, and speaker-proximal form for anaphoric reference.

(89) Makalero (Huber 2011: 127)

[...] *ki-rata=ni* *hai* *hau* ***uere'***
[...] 3:POSS-tell=CTR NSIT all **IDT.V.ADDR**

'the story is that (one).'

(90) Makalero (Huber 2011: 275)

Rata *ki-hau* *hai'=ini* ***ere'***
story 3:POSS-all finished=CTR **IDT.V.PRX**

'This is the end of the story.'

In Icari and Sanzhi Dargwa, locative-existential verbs may be used to express a type of thetic presentational, introducing a referent into the discourse. It is uncertain how spatially marked distance correlates with this use, however.

(91) Dargwa (Icari) (Sumbatova & Mutalov 2003: 147)

le=w *niš̄a-la* *š̄a=w* *ca* *palluq'al*
LOX.V=M we-GEN village:INESS=M one fortune-teller

'There is a fortune-teller in our village.'

(92) Dargwa (Sanzhi) (Forker 2020: 291)

hext:u-b *hin-na* *k'arant'* ***k'e-b=de***
there.UP-NEUT water-GEN spring **LOX.V.UP-NEUT=PST**

'There was a spring up there.'

(93) Dargwa (Sanzhi) (Forker 2020: 315)

Napisat *ka-r-ils-na-l-le-b*
Napisat DOWN-F-sleep-PTCP.LOC-OBL-LOC-N
ka-r-ils-na-b ***te-b=de*** *maʿlʔuʿn*
DOWN-F-sleep-PTCP.LOC-N **LOX.V.DST-NEUT=PST** snake

'Where Napisat was sleeping there was a snake.'

According to Daguman (2004), demonstrative verbs of motion in Northern Subanen signal a transition from one part of the discourse to the next.

(94) Northern Subanen (Daguman 2004: 224)

nandaw ***ditu=ita*** *sə g=ʔuna g=bahin*
now **DST.LOC.ADV=1PI.ABS** OBL SCM=first SCM=part

'Now let's *ditu* (lit:go there) to the first part [...]'

Demonstrative verbs of speech in Crow verbs refer to chunks of discourse occurring either previously or immediately after.

(95) Crow (Graczyk 2007: 87)

íiwahkoo-shii-ak *hinne póopahta-chia-sh kuu-ák*
that-say-SS this owl-white-DET give-SS

'That is what he said, this White Owl, and he gave it to him.'

(96) Crow (Graczyk 2007: 87)

John kuss ***hilía-shee-k*** *sáapa-ss da-luú-o-?*
John GOAL **this-say-DECL** what-GOAL 2A-come.PL-PL-INTERR

'John said this to them, "why did you come?" ' (Jn 3:7)

Motuna is described as having two demonstrative verbs of speech, *tiwo=tiik-* 'to say that way' and *tiwo=tiih-* 'to say that way to' (Onishi 1994). There does not appear to be a spatial distinction, with only the distal form possible. Examples are limited, but they suggest primarily being used anaphorically, for direct speech preceding the demonstrative verb.[11]

(97) Motuna (Onishi 1994: 282)

["..."] ***tiwo=tiih-ku*** *roki*
["..."] **say.to.that.way.3O.3A-GEN.DS** just
mihw-or-u-ng
move.vigorously-MD.3SG-RMPST-M

'After he said to it (the leaf) that way, it just moved vigorously.'

[11] A reviewer rightly questions the inclusion of Motuna, as *tiwo* is simply a manner adverb, and *tiik* is translated as the verb to 'say'. Motuna does not appear to have specific morphology distinct to these specific constructions, and the roots are underived. We do not have enough data to conclusively decide whether Motuna should really be included or not. We tentatively include it as a marginal instance of demonstrative verbs, pending further evidence.

Demonstrative verbs more generally are also often used in connection with direct speech, either as the main predicator, or with another verb such as 'say' or 'do'. Deictic expressions of manner, quality, and degree are well attested as sources of direct speech, or *quotative*, markers (cf. Guérin 2015; Güldemann 2008; König 2015; König & Umbach 2018). Part of speech does not appear particularly relevant to the use of such forms to mark direct speech; both adverbial as well as verbal examples are easily found.

Güldemann mentions in Bengali for instance, that the phrase *emon kora* 'do thus' is employed for direct speech (Güldemann 2008: 321).

(98) Bengali (Wurff 1996; cited in Güldemann 2008: 321)

Se emon korlo: a re baba eta ki
he thus did oh, dear father, what is this?

'He was like: "Oh my, what is this?" '

Parallel constructions are found with processive-qualitative demonstrative verbs. In Vaeakau-Taumako, the distal form of the demonstrative verb *phe-* is used for introducing direct speech.

(99) Vaeakau-Taumako (Næss & Hovdhaugen 2011: 128)

Ko au na, ko-i ***phe-lā*** *ange po e: a*
INCP come DEM.MD INCP-3SG **PQ.V-DST** go.along COMP hey COLL
kaikai ko=no tuna-a napo ke a:
food 2SG=IPFV cook-TR DEM.MD COMP HOR

'He came and (said): "The food you are cooking, what is it for?" '

In Wala, the verb *'uri 'e* '(be/do) like this' is predominantly used in combination with a speech act verb (Lovegren et al. 2015: 160), cf (100). In rare instances, it also occurs as a main verb reporting the contents of a book, song, or inscription (101).

(100) Wala (Lovegren et al. 2015: 160)

ma daulu ka rii ***'uri 'e*** *"Te alo-e 'are!"*
and 3PC SEQ shout **PQ.V.PROX** one spirit-INDEF.PERS thing

'And they shouted, saying, "It is a spirit!" ' (Mt 14:26)

(101) Wala (Lovegren et al. 2015: 160)

ma gera ka nu-li-a te fe nuu fa'alu la 'e
and 3PL SEQ sing-TR-3.OBJ INDEF.SPC CLF song be.new DST 3SG
'uri 'e *"'o totolia sake-na-la buka"*
PQ.V.PROX 2SG be.able take-NMLZ-3.PERS book

'And they sung a new song, saying, "you can take the book" ' (Rv 5:9)

Processive and qualitative verbs are not the only type of demonstrative verb to function as markers of direct speech. In at least one instance, locative verbs show similar functions. The anaphoric locative verb *koolá* in Crow is described as being used to introduce direct speech. It is interesting to note that the locative verb appears to be preferred for marking direct speech in Crow (102), despite the existence of both processive-qualitative verbs as well as demonstrative verbs of speech.[12]

(102) Crow (Graczyk 2007: 85)

Bill huua-sh ***koolá-(a)k*** *'dáa-h óolapi-h' he-k*
Bill say.PL-DET **ANPH.LOX.V-SS** go-IMP find-IMP say-DECL

'Bill it was "go find it", he said.'

No information is available on whether other languages can use demonstrative locative-existential verbs as quotatives, or whether other semantic types such as identification or movement can be used to mark direct speech. Buchstaller & van Alphen (2012) do mention that apart from some Germanic languages (English use of *go* as a quotative is widely described in the literature), motion verbs are only rarely attested source constructions for quotatives, and the only language they were aware of with clear evidence of a motion verb grammaticalizing into a quotative outside of Indo-European is Dongolawi. This suggests that movement demonstrative verbs at least may not easily carry this function.

5 Demonstrative verbs in non-deictic function

Non-deictic functions for demonstrative verbs include tracking (anaphoric and cataphoric), expressive, and recognitional. Very little is known about

[12] Despite the non-idiomatic translation in English, we chose to keep this example, as we could find no other examples of *koolá* in the original source.

such functions for demonstrative verbs in any language. Few languages with demonstrative verbs have enough texts or descriptions to draw conclusions on tracking use for instance, and almost no grammatical description investigates recognitional or expressive uses.

Sporadic limited evidence does give hints that demonstrative verbs may be used in an anaphoric sense, keeping track of preceding discourse events, though in general non-textual anaphoric and cataphoric tracking reference, including events, actions, or locations, is rare. In the following example in Hidatsa, for instance, the demonstrative verb of amount directly refers to the amount mentioned by the addressee immediately before.

(103) Hidatsa (Park 2012: 478)

íiwasi	*dáàwiʔ?*	*– gihxú-haa-biragá-c*
price	how.much-INTERR	– five-ADV-ten-DECL

– garu-siʔawí-hdaa?
– LIM-AMT.V.DST-LIM

'How much did it cost? – It was fifty. – Only that much!?'

A similar function is seen in Eibela, although (104) is somewhat open to interpretation whether it should be anaphoric or exophoric.

(104) Eibela (Aiton 2016: 153)

A:	*gɛː*	*haːnɛ*	*wɛɸɛːniː*	*aːneː=jaː?*		*B:*	*ɛː*
A:	2SG	water	fetch:PURP	go:PST=INTERR.NPRES?		B:	yes

woːgaː
PQ.V.ANPH:PST

'A: Did you fetch the water? B: Yes, I did that.'

When testing possible uses for such functions with native speakers of Cebuano and Tagalog, we discovered that Cebuano does not include overt deictic reference in its verbs of movement. Tagalog, however, does seem to be able to use demonstrative verbs of movement even in an anaphoric sense. In Example (105), the destination of 'go there' is coreferential with the location mentioned in the immediately preceding sentence.

(105) Tagalog (Stan Pineda, p.c.)

nawala ng kasintahan ko pitaka niya sa hotel
lost INDR girlfriend 1SG.INDR purse 3SG.INDR OBL hotel
kagabi. ***p<um>a-roon*** *siya para hanapin*
last.night. **<AF>-CAUS-DST.LOC** 3SG.DIR for search

'My girlfriend lost her purse in a hotel last night. She went there to look for it.'

For one language, Nivkh, it is overtly stated that demonstrative verbs are not used for tracking (Gruzdeva 2020). Note that this does not include discourse deictic reference.

A second non-deictic function of demonstratives is called expressive, also known as empathetic, emotional, or affective. Expressive use of demonstratives serves to express speaker attitude and psychological proximity. Expressive use of demonstrative verbs is rarely described, but some evidence is known. In Nivkh for instance, the proximal qualitative verb *təmra-* (East Sakhalin), *toʁa-* (Amur) is used expressively.

(106) Nivkh (Amur) (Gruzdeva 2020: 42)

toʁa-ɟ *ha-re* *u:məv+ɲivx*
PQ.V.PROX-NMLZ PQ.V.DST-EVID:DIR be.brave+man
toŋgur *kins-tox* *va-řa-ɟ*
MANNER.PROX:3SG devil-DAT fight-ITER:3SG-IND
ha-re.
PQ.V.DST-EVID:DIR

'Such (by quality) a brave man fights with a devil in this way.'

Moyse-Faurie (2019: 147) mentions that verbal processive-qualitative demonstratives in Polynesian can occur as nominal determiners, which "expresses the surprise of the speaker".

(107) East Uvean (Moyse-Faurie 2019: 148)

Ko taku hoki sio ki te me'a ***fēnei***
PRED POSS.1SG just see to SPC thing **PQ.V.PROX**

'It is the first time I have seen such a thing.'

In Abui, medial qualitative verbs combine with medial locative-existential verbs to express disapproval of a situation (108). Demonstratives of all types in Abui show extremely complex expressive functions, including additional uses such as modality and evidentiality (cf. Kratochvíl 2007: 107).

(108) Abui (Kratochvíl 2007: 373)

no-ma	***wa***	*naha*
1SG.REC-LOX.V.PRX	**PQ.MD.CNT**	NEG

'I do not want it so.'

Similar functions have also been described for other Alor-Pantar languages. Steinhauer for instance mentions that in Blagar, demonstrative locative-existential verbs not only have different sets depending on visibility and duration, but that they can also indicate evidentiality and feelings. "They [the *ʔaʔa* series of demonstrative verbs] indicate sure knowledge by the speaker about the existence of the event" and the *ʔaʔe* series of demonstrative verbs can express "indignation and/or amazement" (Steinhauer 1991). In Reta, two demonstrative verbs indicating location and existence, *gi'e* 'be here' and *a'e* 'be there', are used to express epistemic modality, with the choice between them based on the relevance of the propositional content with regard to the respective interlocutors (Willemsen 2021).

The final non-deictic function looked at here is recognitional. Recognitional use of demonstratives does not rely on the surrounding context of a demonstrative the way exophoric and discourse-deictic uses do; rather, it relies on shared knowledge between the speaker and the addressee. Recognitional use of demonstratives has had little discussion in the literature, with notable exceptions such as Diessel (1999) and Himmelmann (1996). Diessel (1999) furthermore states that recognitional demonstratives are restricted to adnominal environments. Although no clear example of recognitional use has been found for demonstrative verbs, the data in the grammatical descriptions has also been too limited to draw any firm conclusions. More data is desperately needed on whether demonstrative verbs truly do not show any recognitional functions.

6 Conclusion

This article has discussed the grammatical and semantic features of demonstrative verbs, and their uses in different pragmatic contexts. It was confirmed that

within a general class of demonstratives, it is possible to postulate a separate category of demonstrative verbs, with its own distinct semantic and morphosyntactic properties. We have shown that demonstrative verbs can be classified into three major semantic groups of locative-existential, processive-qualitative and movement and placement verbs, as a well as several minor types. Demonstrative verbs can be used in various pragmatic functions, though there is still a lot of questions to be answered in this area.

The mismatch between the morphosyntactic category of demonstrative verbs and their derived adverbial function is a particularly interesting area worthy of further study, not only for demonstratives specifically, but also fitting in a larger discussion in typologically-oriented research. Various form-function combinations of verb and adverb were discussed throughout this article. Verbal affixes can be used directly on some adverbial roots in languages like Northern Subanen; conversely, in languages like Urarina and Abui, demonstrative verbs are nearly always used in adverbial function rather than predicative. And in many languages, it is not even always possible to determine the lexical category of a given root. Moyse-Faurie (2019: 144) comments that demonstrative verbs in Polynesian languages are "[...] always polyfunctional, and can occur as verbal predicates, adverbs, connectives, quotative and optative verbal markers, and nominal modifiers". While the general connection between adverb and verb is abundantly clear when it comes to demonstrative verbs, the question on how to handle polyfunctional lexical roots when looking at lexical classes cross-linguistically remains a difficult question to grapple.

The study expands considerably on earlier investigations of demonstrative verbs, both in terms of the number of languages, as well as in terms of the scope covered by the analysis. We must admit that gaps in the data prevent conclusive evidence on a number of different topics, however. Due to practical considerations, an investigation into details of the spatial distinctions was not possible, and would be a fruitful area for future research.

We also note that nearly half of the languages in the study come from Austronesian languages. Although we did not conduct any statistical analysis, it is nonetheless highly unlikely to be a coincidence that so many languages with demonstrative verbs are Austronesian. Killian (2022a) notes a similar unexpectedly high frequency with adpositional demonstratives. One can only speculate that it is perhaps the part-of-speech flexibility many Austronesian languages show, combined with the rich collection of constructions devoted to space and direction, that has led to such a phenomenon.

Finally, we hope that the research presented here will spur on more detailed descriptive research in individual languages.

Acknowledgments

We are grateful to all language speakers, specialists, and researchers who have helped us over the years. We also thank Harald Hammarström for his help in supplying linguistic material, without which this work would not have been possible. Finally, we are grateful for the helpful comments provided by two anonymous reviewers on an earlier draft of this article. Any errors remain our own.

Funding for this research was provided in different periods by the Kone foundation and the Finnish cultural foundation, and their support is greatly appreciated.

Abbreviations

1	first person
2	second person
3	third person
ABS	absolutive
ABSENT	absent
ACC	accusative
ADDR	addressee-proximal
AF	actor focus
AG	agent
ALL	allative
AMT.V	demonstrative verb of amount
AN	animate
APPL	applicative
APPROX	approximative
ART	article
ASC	associative
ASSER	assertive
ATTR	attributive
AV	actor voice
BACK	back
BD	bound form
CAUS	causative

CL	class
CLF	classifier
CLK	clockwise
CNT	continuative
COLL	collective
COMP	complementizer
CONCESS	concessive
COND	conditional
CONT	continuous
COP	copula
CSQ	consequantial
CTR	contrastive
CVB	converb
DAT	dative
DECL	declarative
DET	determiner
DIR	direct
DS	different subject
DST	distal
DYN	dynamic
EMPH	emphatic
ERG	ergative
EVID	evidential
EXPND	expander
FUT	future
GEN	genitive
GENR	general tense-aspect-mood marker
HAB	habitual
HOR	hortative
HRS	hearsay
I	irrealis
IDT.V	identification demonstrative verb
IMP	imperative
INCP	inceptive
IND	indicative
INDEF	indefinite
INDR	indirect
INESS	inessive
INST	instrumental
INTERR	interrogative
INTJ	interjection

INTR	intransitive
INTS	intensifier
IO	indirect object
IPFV	imperfective
ITER	iterative
ITIN	itinerary locative
LF	locative focus
LIM	limitive
LNK	linker
LOC	locative
LOCZR	localizer
LOX.V	demonstrative verb, locative-existential
M	masculine
MANNER	manner adverb
MD	medial
MIR	mirative
MV.V	demonstrative verb of movement
NEG	negation
NEUT	neuter
NFUT	non-future
NMLZ	nominalizer
NOM	nominative
NPRS	non-personal
NPRES	non-present
NSIT	new situation
NVR	non-veridical
OBJ	object
OBL	oblique
OF	object focus
PC	paucal
PE	plural exclusive
PERF	perfect
PERS	personal
PI	plural inclusive
PL	plural
POSS	possessive
PQ.V	demonstrative verb, processive-qualitative
PRED	predicative
PRX	proximal
PST	past
PTCP	participle

PURP	purposive
PUT.V	demonstrative verb of placement
R	realis
RDP	reduplication
REC	recipient
REFL	reflexive
REL	relativizer
RMPST	remote past
ROUTE	route
SBD	non-temporal subordinator
SBJ	subject
SCM	syntactic category for nominal
SEA	seawards
SEQ	sequential
SG	singular
SM	subject marker
SPC	specific
STAT	stative
SUBORD	subordinator
SZ.V	demonstrative verb of size
THM	theme
TOP	topic
TR	transitive
TRANS	event transition
TRF	transferred action
UP	higher elevation
VENT	ventive
VRBLZ	verbalizer

References

Aiton, Grant. 2016. *A grammar of Eibela: A language of the Western Province, Papua New Guinea*. James Cook University. (Doctoral dissertation).

Belyaev, Oleg. 2012. Aspektual'no-temporal'naja sistema aštynskogo darginskogo. *Acta Linguistica Petropolitana* 8(2). 181–227.

Berghäll, Liisa. 2015. *A grammar of Mauwake*. Language Science Press.

Blake, Fiona. 2007. *Spatial reference in Momu*. University of Sydney. (Master's thesis).

Bläsing, Uwe. 2003. Kalmuck. In Janhunen, Juha (ed.), *The Mongolic languages*, 229–247. Routledge.

Blust, Robert. 2003. *Thao dictionary*. (Language and Linguistics Monograph Series). Academia Sinica, Institute of Ethnology.
Bolton, Rosemary Ann. 1990. *A preliminary description of Nuaulu phonology and grammar*. University of Texas at Arlington. (Master's thesis).
Breunesse, Merlijn. 2019. *Demonstratives in space and discourse: A synchronic and diachronic analysis*. University of Jena. (Doctoral dissertation).
Bril, Isabelle. 2002. *Le nêlêmwa (Nouvelle-Calédonie): Analyse syntactique et sémantique*. Vol. 16. (Langues et Cultures du Pacifique). Peeters.
— 2004. Deixis in Nêlêmwa. In Senft, Gunter (ed.), *Deixis and demonstratives in Oceanic languages*, 99–128. Pacific Linguistics.
Buchstaller, Isabelle & van Alphen, Ingrid. 2012. Preface: Introductory remarks on new and old quotatives. In Buchstaller, Isabelle & van Alphen, Ingrid (eds.), *Quotatives: Cross-linguistic and cross-disciplinary perspectives*, xi–xxx. John Benjamins.
Creissels, Denis. 2019. Inverse-locational predication in typological perspective. *Italian Journal of Linguistics* 31(2). 37–106.
Daguman, Josephine. 2001. Functions of locatives in Northern Subanen expository and hortatory discourse. *Studies in Philippine Languages and Cultures* 12(1). 1–29.
— 2004. *A grammar of northern Subanen*. La Trobe University. (Doctoral dissertation).
de Busser, Rik. 2009. *Towards a grammar of Takivatan Bunun: Selected topics*. La Trobe University. (Doctoral dissertation).
de la Cruz, Beato & Zorc, R. David Paul. 1968. *A study of the Aklanon dialect, volume 1: Grammar*. Peace Corps.
Dickens, Patrick J. 2005. *A concise grammar of Ju/'hoan*. Rüdiger Köppe.
Diessel, Holger. 1999. *Demonstratives: Form, function and grammaticalization*. John Benjamins.
— 2012. Deixis and demonstratives. In Maienborn, Claudia & von Heusinger, Klaus & Portner, Paul (eds.), *An international handbook of natural language meaning*, 2407–2431. Vol. 3. Mouton de Gruyter.
Diessel, Holger & Breunesse, Merlijn. 2020. A typology of demonstrative clause linkers. In Næss, Åshild & Margetts, Anna & Treis, Yvonne (eds.), *Demonstratives in discourse*, 305–340. (Topics at the Grammar-Discourse Interface 6). Language Science Press.
Dixon, R. M. W. 2003. Demonstratives: A cross-linguistic typology. *Studies in Language* 27(1). 61–112.
— 2010. *Basic Linguistic Theory, vol. 2: Grammatical topics*. Oxford University Press.
Du, Jingyi. 2010. *Towards a grammar of the Usen dialect of the Barok language New Ireland, Papua New Guinea*. La Trobe University. (Doctoral dissertation).
Forker, Diana. 2020. *A grammar of Sanzhi*. Language Science Press.
Frowein, Friedel Martin. 2011. *A grammar of Siar, an Oceanic language of New Ireland Province*. La Trobe University. (Doctoral dissertation).

Galloway, Brent Douglas. 1977. *A grammar of Chilliwack Halkomelem*. University of California at Berkeley. (Doctoral dissertation).
Graczyk, Randolph. 2007. *A grammar of Crow [Apsaalooke Alilaau]*. (Studies in the native languages of the Americas). University of Nebraska Press.
Gruzdeva, Ekaterina. 2013. Demonstrativnye glagoly v jazykah mira [Demonstrative verbs in the languages of the world]. In *Trudy Instituta lingvističeskih issledovanij RAN, T. IX. Č. 3: Sbornik statej k 60-letiju Evgenija Vasil'eviča Golovko*, 51–80. Vol. IX: 3. Acta Linguistica Petropolitana.
— 2020. Demonstratives in Nivkh: a semantic and pragmatic analysis. *Studia Orientalia Electronica* 8(1). 1–60.
Guérin, Valérie. 2015. Demonstrative verbs: A typology of verbal manner deixis. *Linguistic Typology* 19(2). 141–199.
Guérin, Valérie & Aiton, Grant. 2018. Bridging constructions in typological perspective. In Guérin, Valérie (ed.), *Bridging constructions*, 1–44. (Studies in Diversity Linguistics 24). Language Science Press.
Güldemann, Tom. 2008. *Quotative indexes in African languages: A synchronic and diachronic survey*. Mouton de Gruyter.
Hagège, Claude. 2008. Towards a typology of interrogative verbs. *Linguistic Typology* 12(1).
Halliday, Michael & Hasan, Ruqaiya. 1976. *Cohesion in English*. Longman.
Hanks, William F. 1992. The indexical ground of deictic reference. In Duranti, A. & Goodwin, C. (eds.), *Rethinking context*. Cambridge University Press.
Himmelmann, Nikolaus. 1996. Demonstratives in narrative discourse: A taxonomy of universal uses. In Fox, Barbara (ed.), *Studies in anaphora*, 205–254. Vol. 33. John Benjamins.
Huber, Juliette. 2011. *A grammar of Makalero: A Papuan language of East Timor*. Vrije Universiteit Amsterdam. (Doctoral dissertation).
Jiang, Haowen. 2006. *Spatial conceptualizations in Kavalan*. University of Taipei. (Master's thesis).
— 2009. Three dichotomies in the Kavalan demonstrative system. (Handout of a presentation at the Eleventh International Conference on Austronesian Linguistics. Aussois, France.).
Jukes, Anthony. 2006. *Makassarese (basa Mangkasara'): A description of an Austronesian language of South Sulawesi*. University of Melbourne. (Doctoral dissertation).
Killian, Don. 2022a. Deictic marking in adpositions. *Te Reo* 65(1). 1–41.
— 2022b. Towards a typology of predicative demonstratives. *Linguistic Typology* 26(1). 1–41.
Klamer, Marian. 1998. *A grammar of Kambera*. Vol. 18. (Mouton Grammar Library). Mouton de Gruyter.

König, Ekkehard. 2015. The deictic identification of similarity. In Treis, Yvonne & Vanhoeve, Martine (eds.), *Similative and equative constructions: A cross-linguistic perspective*. John Benjamins.

König, Ekkehard & Umbach, Carla. 2018. Demonstratives of manner, of quality and of degree. In Coniglio, Marco & Murphy, Andrew & Schlachter, Eva & Veenstra, Tonjes (eds.), *Atypical demonstratives: Syntax, semantics and pragmatics*, 285–328. De Gruyter.

Kratochvíl, František. 2007. *A grammar of Abui: A Papuan language of Alor*. Rijksuniversiteit te Leiden. (Doctoral dissertation).

Levinson, Stephen C. 1983. *Pragmatics*. (Cambridge Textbooks in Linguistics). Cambridge University Press.

— 2018. Introduction: Demonstratives: Patterns in diversity. In Levinson, Stephen C. & Cutfield, Sarah & Dunn, Michael J. & Enfield, N. J. & Meira, Sérgio (eds.), *Demonstratives in cross-linguistic perspective*, 1–42. Cambridge University Press.

Lichtenberk, Frantisek. 2008. *A grammar of Toqabaqita*. Vol. 42. (Mouton Grammar Library). Mouton de Gruyter.

Lionnet, Florian. 2014. Demonstrative and relative constructions in Ju: A diachronic account. In Güldemann, Tom & Fehn, Anne-Maria (eds.), *Beyond 'Khoisan': Historical relations in the Kalahari Basin*, 181–207. John Benjamins.

Lovegren, Jesse & Mitchell, Alice & Nakagawa, Natsuko. 2015. *The Wala language of Malaita, Solomon Islands*. Vol. 3. (Studies in the Languages of Island Melanesia). Asia-Pacific Linguistics.

Maslov, Yu. S. 1975. *Vvedeniye v yazykoznaniye* [Introduction to general linguistics]. Vysshaya shkola.

May, Kevin. 1997. *A study of the Nimboran language: Phonology, morphology, and phrase structure*. La Trobe University. (Master's thesis).

Miyaoka, Osahito. 2012. *A grammar of Central Alaskan Yupik: An Eskimo language*. Mouton de Gruyter.

Mofu, Suriel Semuel. 2005. *Biak morphosyntax*. University of Oxford. (Master's thesis).

Mosel, Ulrike. 2004. Deixis in Samoan. In Senft, Gunter (ed.), *Deixis and demonstratives in Oceanic languages*, 99–128. Pacific Linguistics.

Moyse-Faurie, Claire. 2019. Similarity, comparison of equality and verbal manner demonstratives in Polynesian languages. *Faits de Langues* 50(1). 135–156.

Musgrave, Jill. 2007. *A grammar of Neve'ei, Vanuatu*. Vol. 587. (Pacific Linguistics). Australian National University.

Næss, Åshild & Hovdhaugen, Even. 2011. *A grammar of Vaeakau-Taumako*. Vol. 52. (Mouton Grammar Library). Mouton de Gruyter.

Nakazawa, Tsuneko. 2007. A typology of the ground of deictic motion verbs as path-conflating verbs: The speaker, the addressee, and beyond. *Poznań Studies in Contemporary Linguistics* 43(2). 59–82.

Olawsky, Knut. 2006. *A grammar of Urarina*. Mouton de Gruyter.

Onishi, Masayuki. 1994. *A grammar of Motuna (Bougainville, Papua New Guinea)*. Australian National University. (Doctoral dissertation).
Overall, Simon. 2007. *A grammar of Aguaruna*. La Trobe University. (Doctoral dissertation).
Palmer, Bill. 2009. *Kokoto grammar*. University of Hawaii Press.
Park, Indrek. 2012. *A grammar of Hidatsa*. Indiana University. (Doctoral dissertation).
Pearce, Elizabeth. 2015. *A grammar of Unua*. Vol. 647. (Pacific Linguistics). De Gruyter Mouton.
Poppe, N. N. 1937. *Grammatika pis'menno-mongol'skogo yazyka* [A grammar of written Mongolian]. Izdatel'stvo Akademii nauk SSSR.
Poppe, Nicholas. 1964. *Grammar of written Mongolian*. Vol. 1. (Porta linguarum orientalium). Harrassowitz.
Poppe, Nicholas N. 1960. *Buriat grammar*. Vol. 2. (Indiana University Publications: Uralic and Altaic Series). Indiana University.
Pratchett, Lee. 2017. *Dialectal diversity in Southeastern Ju (Kx'a) and a documentation of Groot Laagte ǂKx'ao‖'ae*. Humboldt University of Berlin. (Doctoral dissertation).
Rassadin, B. I. 1991. Mestoimeniya v mongol'skix yazykax [Pronouns in Mongolic languages]. In Burayev, I. D. (ed.), *Voprosy grammatiki mongol'skix yazykov* [Grammatical issues in Mongolic languages], 89–99. Nauka.
Rodegem, Firmin M. 1967. *Précis de grammaire rundi*. Vol. 10. (Travaux de l'Université de Bujumbura, A, Faculté de Philosophie et Lettres). Éditions Scientifiques, E. Story-Scienta.
Smeets, Ineke. 2008. *A grammar of Mapuche*. Mouton de Gruyter.
Smits, Heleen. 2017. *A grammar of Lumun*. Rijksuniversiteit te Leiden. (Doctoral dissertation).
Stassen, Leon. 2009. *Predicative possession*. Oxford University Press.
Steinhauer, Hein. 1991. Demonstratives in the Blagar language of Dolap (Pura, Alor, Indonesia). In Dutton, Tom (ed.), *Papers in Papuan linguistics*, 177–221. Pacific Linguistics.
— 2005. Biak. In Adelaar, Alexander & Himmelmann, Nikolaus (eds.), *The Austronesian languages of Asia and Madagascar*, 793–823. (Routledge Language Family Series). Routledge.
Sumbatova, Nina & Mutalov, Rasul. 2003. *A grammar of Icari Dargwa*. Lincom.
Suttles, Wayne. 2004. *Musqueam reference grammar*. (First Nations Languages). University of British Columbia Press.
Teng, Stacy Fang-Ching. 2008. *A reference grammar of Puyuma, an Austronesian language of Taiwan*. Vol. 595. (Pacific Linguistics). Australian National University.
van den Berg, René & Busenitz, Robert L. 2012. *A grammar of Balantak, a language of Eastern Sulawesi*. Vol. 40. (SIL e-Books). SIL International.
van den Heuvel, Wilco. 2006. *Biak: Description of an Austronesian language of Papua*. Vrije Universiteit Amsterdam. (Doctoral dissertation).

van Gijn, Erik. 2006. *A grammar of Yurakaré*. Radboud Universiteit Nijmegen. (Doctoral dissertation).
Wang, Shan-Shan. 2004. *An ergative view of Thao syntax*. University of Hawai'i at Mānoa. (Doctoral dissertation).
Wilkins, David P. & Hill, Deborah. 1995. When 'go' means 'come': Questioning the basicness of basic motion verbs. *Cognitive Linguistics* 6(2–3). 209–259.
Willemsen, Jeroen. 2021. *A grammar of Reta*. Aarhus University. (Doctoral dissertation).
Zorc, R. David Paul. 1977. *The Bisayan dialects of the Philippines: Subgrouping and reconstruction*. Vol. 44. (Pacific Linguistics: Series C). Australian National University.

The appendices "Appendix 1. Languages with demonstrative verbs" and "Appendix 2. Total language sample" are available at https://doi.org/10.61197/fjl.126939.

Contact information:

Don Killian
University of Helsinki
e-mail: donald(dot)killian(at)helsinki(dot)fi

Ekaterina Gruzdeva
University of Helsinki
e-mail: ekaterina(dot)gruzdeva(at)helsinki(dot)fi

Syntax in transition: Emergence of differential argument marking in Aanaar Saami

Jukka Mettovaara
University of Oulu

Abstract

This paper is an investigation of a language contact phenomenon currently taking place in Aanaar (Inari) Saami, an indigenous minority language of Finland. Aanaar Saami people have been in contact with Finnish speakers for centuries, so much so that the language community has become bilingual. This has resulted in the borrowing of both numerous Finnish loanwords and even some syntactic constructions into Aanaar Saami and seems to herald a larger change in the language. The present study focuses on a type of syntactic change called *differential argument marking*, which is examined in three Aanaar Saami clause types: transitive clauses, existential clauses, and passive verb clauses. Finnish exhibits complex argument marking, characterized by a total–partial distinction, whereas traditional Aanaar Saami does not have differential argument marking in subjects or objects. However, new Aanaar Saami shows multiple emergent types of differential argument marking that vary between individual speakers and, while clearly influenced by Finnish, do not always mirror their Finnish equivalents one-to-one. This, and the observation that differential argument marking is non-existent in older language materials, suggests that the phenomenon is recent Finnish influence.

Keywords: Aanaar Saami, Finnish, differential argument marking, pattern replication, syntax, language contact

1 Introduction

The syntactic object in many Uralic languages is known for having a distinction between the marked and unmarked object. This is reflected in case marking so that marked objects are assigned the accusative case and unmarked ones the nominative. By and large, the motivation for whether an object is marked or unmarked can be said to lie in definiteness or, in the case of the Finnic

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

Finnish Journal of Linguistics 36 (2023), 93–129
DOI: 10.61197/fjl.113917

subgroup, partiality vs. totality (Havas 2008). At the same time, Finnic languages exhibit similarly conditioned variation also in subject marking, where the distinction between totality and partiality motivates the selection of either the nominative or non-nominative subject in some clause types. Together, the fluctuation of object and subject marking under certain conditions is labelled *differential argument marking* (DAM).

In this paper,[1] I examine the emergence of DAM as an instance of contact-induced language change in contemporary Aanaar Saami, an indigenous minority language spoken by an estimated 400 people mainly around lake Aanaar (Finnish *Inari*) in Northern Finland. The aim is to show that modern Aanaar Saami is in a stage of syntactic restructuring, whereby it is developing a new grammatical feature.

Saami languages constitute a subgroup within the Uralic languages, and their speakers are spread across four states: Norway, Sweden, Finland, and Russia. Saami languages are not majority languages in any of the areas where they are spoken, and virtually every member of the Saami speech community also masters their respective majority and state language (Aikio et al. 2015). The effects of intense contact with the majority languages are widely recognized among Saami language researchers, teachers, and other language workers, but systematic descriptions of results, especially ones concerning grammatical influence, are still mostly lacking (Mettovaara & Ylikoski forthcoming), save for some single studies, such as Rießler (2007) who investigates grammatical borrowings from Russian into Kildin Saami.

The impetus for this study came from my personal observations as a learner and teacher of Aanaar Saami as well as discussions with other language teachers and researchers who have noticed that L2 learners and younger L1 speakers of Aanaar Saami and North Saami frequently exhibit contact-induced changes in their language. One of these changes is the variation in argument marking that appears to follow the model of the complex Finnish system of argument marking (e.g. R. Magga & Pulska 2019).

This paper answers the following questions:

1. What types of DAM are attested in new Aanaar Saami?
2. What conditions appear to trigger the non-canonical argument marking?
3. What is the syntactic pivot in Finnish after which the new Aanaar Saami DAM is modelled?

[1] This work was funded by a grant from the Finnish Cultural Foundation.

The structure of the paper is as follows: after the introduction, § 2 presents the sociological and linguistic context wherein Aanaar Saami exists, as well as a short comparison of Aanaar Saami and Finnish. § 3 outlines the theoretical framework, namely argument marking and pattern replication. § 4 describes the data used in this study. § 5 investigates DAM as found in three clause types in new Aanaar Saami: a) transitive clauses, b) existential clauses, and c) passive verb clauses. In § 6, the features of DAM observed in the preceding section will be summarized and motivations for its emergence in Aanaar Saami will be discussed. The main results of this study are that new Aanaar Saami is beginning to show clear signs of Finnish influence in the way syntactic arguments are marked, in an attempt to reproduce the Finnish totality vs. partiality distinction in subjects and objects.

2 Background

The purpose of this section is to describe the nature of Aanaar Saami–Finnish language contacts and illustrate the significant degree of grammatical similarity that already exists between the languages which facilitates further convergence. I provide some background information on Aanaar Saami and its sociological context as well as a short comparison of Aanaar Saami and Finnish grammatical systems and nominal morphosyntax.

2.1 The sociolinguistic situation of Aanaar Saami

Based on a handful of phonetic and morphosyntactic criteria, Aanaar Saami has been traditionally classified as the westernmost member of the eastern Saami subgroup (Korhonen 1964; Sammallahti 1998). The bifurcate division has become the de facto standard, even though it has been called into question and criticized since its introduction. The problematique especially concerns Aanaar Saami, since based on lexical criteria, it could be included in the western group or even form its own dialect group (Rydving 2013; Tillinger 2014).

All Saami languages have been in contact of varying intensity with their neighbouring languages – Norwegian and Swedish in (north)western Fennoscandia, Finnish in the north, and Russian and Karelian in the east – for a long time (Laakso 2010: 600–601; Kittilä & Ylikoski 2018: 470). As a result, the different Saami languages have become dissimilar from each other in terms of vocabulary and, to an extent, grammatical structures. At the same time, they have also grown closer to their respective majority languages. The

closing of the borders between Norway, Sweden, and the then Grand Duchy of Finland in the 19th century and Finland's independence from Soviet Russia in 1917 shaped the geopolitical reality where the Saami live to this day. North Saami is a prime example of this divide, its traditional dialectal differences partly replaced by topolects of three regions, that is, those of the Finnish, Swedish, and Norwegian sides of the border (Aikio et al. 2015).

Aanaar Saami speakers have had long-standing contacts with both North and Skolt Saami speakers and Finns, and they have reportedly already spoken Finnish well by 1830s (Lehtola 2012: 41). Aanaar Saami is spoken only in Finland, and virtually all speakers are bilingual in Finnish, another Uralic language but genealogically originating in the Finnic subgroup, so it is natural that Finnish is the main source of loanwords and grammatical influence. This is evident already in the earliest attestations of Aanaar Saami. Due to large-scale societal changes and an intense language shift to Finnish in the latter part of the 20th century, the influence of Finnish became even more prominent. The situation was most dire in 1997, when there were only four or five speakers under 30 years old and the language domains had become very limited, prompting concern that Aanaar Saami would likely not be passed on to another generation (M. Morottaja 1996). However, successful revitalization efforts starting from the 1980s managed to halt the language shift and later even reverse it (Olthuis et al. 2013; Pasanen 2015).

The different types of Finnish grammatical and lexical transfer are often viewed as characteristic of learner language in Aanaar Saami and something to be corrected. For example, in his manual of Aanaar Saami, M. Morottaja (2007: 32, 34, 40–50) highlights some features he considers to be a result of Finnish influence in Aanaar Saami, such as confusing the cases of subjects and objects, difficulties in distinguishing close synonyms in cases where Finnish only has one corresponding translation and using predicative and attributive adjective forms interchangeably.[2]

In 2020, the number of Aanaar Saami speakers was estimated at 450, and the speaker community is even expected to grow. There are around 20 families with children who speak Aanaar Saami as the main language at home and many more where it is one of the languages spoken. There are 25 children

[2] The distinction between the predicative and attributive forms of adjectives is common to all Saami languages (Rießler 2016). However, even before the modern-day Finnish influence, there seems to have been language-internal variation that has been simply left out of the current standardized norm (Müller 2017). It is therefore debatable to what extent the apparent confusion between forms stems from Finnish.

in language nests, an immersion-based form of day-care, and around 80 children in total have participated in language nests since they were founded in 1997. Some of these participants now have children of their own to whom they speak Aanaar Saami. The language is taught in primary and secondary schools, where it is used as the language of instruction in some subjects, as well as at the University of Oulu (Pasanen 2020).

An interesting characteristic of Aanaar Saami revitalization in the realm of language revitalization programmes has been the principle of inclusion or ethnic neutrality. From the start, the Aanaar Saami language society *Anarâškielâ servi* has aimed the language nests not only at ethnic Aanaar Saami but also at children of any background; motivation to learn the language and willingness to commit to the language community have been considered more important than ethnicity. This has led to a situation where a considerable number of Aanaar Saami speakers are people who have learned the language in adulthood (*new speakers*) and people who have learned the revitalized language in childhood, typically in a language nest (*neospeakers*)[3] (Pasanen 2015: 315 ff., 341–342).

2.2 General morphosyntactic properties of Aanaar Saami and Finnish

Genealogically, the Finnic and Saami languages are classified as "distantly related" to each other, but in typological terms they are closest to one another within the Uralic language family. They share many morphological features, such as consonant gradation, primarily suffixal morphology, inflectional categories (person, tense, and mood system in verbs; cases and number in nouns), and a propensity to use derivation as a means to generate new lexemes. They share many syntactic features as well, such as the basic constituent order SVO[4] and clause combining strategies (Koponen 2022: 111). The most notable differences in morphology are the existence of the dual in most varieties of Saami and the average number of productive nominal cases: 6–9 in Saami vs. 8–18 in Finnic languages. In syntax, one of the main points of divergence is the marking of subjects and objects. This will be discussed in detail in § 3.1.

Finnic and Saami nominals are inflected in cases that can be roughly

[3] On the typology of speakers of endangered languages, see Grinevald & Bert (2011: 49–52).

[4] Some Saami languages exhibit SOV or vacillate between SVO and SOV (Ylikoski 2022b: 143). Nevertheless, based on personal knowledge, I agree with Valtonen et al. (2022: 192) and P. Morottaja & Toivonen (in preparation) that the basic word order in Aanaar Saami is SVO, even though there are no systematic corpus studies to support this statement.

grouped into four classes: 1) core grammatical cases (cases of subject, object, and possessor); 2) concrete and abstract local cases; 3) other cases that denote concomitance, lack, or means, and the like; and 4) other cases, including marginal and non-productive case forms, some of which could be classified as adverbial suffixes. An example of a noun paradigm in Aanaar Saami and Finnish is shown in Table 1.[5] Both language groups have one to three sets of three local cases that prototypically denote motion to(wards), motion (away) from, and residing in a location or state. Saami languages have only one set of all-around local cases, whereas Finnic languages commonly have two: inner (inessive, elative, illative) and outer (adessive, ablative, allative). In North Saami and other Saami languages to the east, the local cases denoting motion from and residing in have merged into one case called locative: Aanaar Saami *kuátá-n* [hut-ILL] 'to the hut' : *kuáđi-st* [hut-LOC] 'in/from the hut' vs. South Saami *gåata-n* [hut-ILL] 'to the hut' : *gåete-ste* [hut-ELA] 'from the hut' : *gåete-sne* [hut-INE] 'in/at the hut'.

There are some remarks to be made here. Firstly, despite sharing the same name, Aanaar Saami and Finnish partitives have very different functions. The Finnish partitive is a core grammatical case used to mark syntactic objects and subjects and nouns in quantifier phrases with numerals higher than 'two'. It is also used to indicate the complement of certain adpositions. In contrast, the Aanaar Saami partitive is a marginal case: its most common use is to mark nouns in quantifier phrases with numerals higher than 6, and even in this function it has begun to lose ground to the genitive case. Secondly, the Aanaar Saami comitative and abessive are common cases used to mark concomitance and the lack thereof respectively, whereas their Finnish namesakes are marginal cases that are mostly replaced by adpositional phrases.

To summarize, Aanaar Saami and Finnish share several structural features, which is of great help to learners of one language who already have command of the other. Often the number of similarities also enables the use of "morpheme-for-morpheme" translation when operating between languages: just replacing the units of lexical and grammatical substance in a Finnish sentence with their Aanaar Saami equivalents is likely to produce a perfectly acceptable Aanaar Saami sentence.

[5] Note that the order of Finnish cases does not follow the standard used in most grammars, as the table aims to illustrate the rough functional equivalences between cases.

Table 1. Example paradigms of the Aanaar Saami and Finnish words for 'hut'

Aanaar Saami			Finnish		
	singular	plural		singular	plural
Nominative	*kuáti*	*kuáđih*	Nominative	*kota*	*kodat*
Genitive	*kuáđi*	*koođij*	Genitive	*kodan*	*kotien*
Accusative	*kuáđi*	*koođijd*			
Illative ('to')	*kuátán*	*koođijd*	Illative ('into')	*kotaan*	*kotiin*
			Allative ('to, onto')	*kodalle*	*kodille*
Locative ('in, at; from')	*kuáđist*	*koođijn*	Inessive ('in')	*kodassa*	*kodissa*
			Elative ('from')	*kodasta*	*kodista*
			Adessive ('on, at')	*kodalla*	*kodilla*
			Ablative ('from')	*kodalta*	*kodilta*
Comitative ('with')	*koođijn*	*kođijguin*	Comitative ('accompanied by')	*kotine-* + poss. suffix	
			Instructive ('by means of')	–	*kodin*
Abessive ('without')	*kuáđittáá*	*kođijttáá*	Abessive ('without')	*kodatta*	*koditta*
Essive ('as')	*kuáttin*		Essive ('as')	*kotana*	*kotina*
			Translative ('[turn] into')	*kodaksi*	*kodiksi*
Partitive	*kuáttid*		Partitive	*kotaa*	*kotia*

3 Defining the phenomena

This section delineates the main topics of this paper, namely argument marking and contact-induced pattern replication. Both Finnic and traditional Saami argument marking systems are described succinctly, and the older system is used as a baseline to which the DAM in new Aanaar Saami is compared later in § 5. In this paper, *traditional Aanaar Saami* is understood to be a form of the language as described in earlier (prescriptive) grammars and grammar sketches such as those by P. Morottaja & Olthuis (2022), Valtonen et al. (2022), and – as far as case syntax is concerned – Bartens (1972). As of yet, there are no comprehensive descriptive grammars of Aanaar Saami.

3.1 Argument marking in Finnic and Saami languages

In languages with case marking, the marking of core arguments A(gent of a transitive verb), S(ubject of an intransitive verb), and P(atient/object of a transitive verb) can be motivated by not only syntactic factors but also semantic and pragmatic considerations. For example, the canonical case of S/A in nominative-accusative languages is the nominative, but in some languages S/A may be encoded by another case, as determined by factors such as agentivity, volitionality, and information structure (Malchukov & Spencer 2008; Seržant 2016: 137–138).

The alignment of case marking in both Finnic and Saami languages is nominative-accusative: canonically, the nominative encodes both transitive and intransitive subjects, whereas the accusative encodes objects of transitive verbs. This system is muddled in Finnic languages by the complex DAM system, where the main parameter of alternation in objects is between the total and partial object, and in subjects between the nominative and non-nominative subject. DAM in Finnish subsumes both *differential subject* and *object marking* (DSM and DOM), whose characteristics partly overlap.

Finnish assigns cases to its arguments in too intricate a way to be described fully here; suffice to say that the major parameter in Finnic DAM is based on meaning: both subject and object arguments in the partitive can be characterized in broad terms as indefinite, atelic, unbounded in quantity and aspect, and less prominent on a discourse-pragmatic level than arguments in the nominative/accusative (Kiparsky 2001; Huumo 2003). Within DOM, verb semantics or certain syntactic environments mandate the use of either the total or partial object and, within non-pronominal total objects, the selection of either the

Table 2. Object cases in Finnish

		Total		Partial
SG	NOM	*Syö **voileipä**!* eat.IMP.2SG sandwich 'Eat the sandwich!'	PTV	*Syö **voileipä-ä**!* eat.IMP.2SG sandwich-PTV 'Eat some sandwich!'
	GEN	*Syöt **voileivä-n**.* eat.2SG sandwich-GEN 'You will eat the sandwich.'		*Et syö **voileipä-ä**.* NEG.2SG eat.CNG sandwich-PTV 'You don't eat ∼ aren't eating a ∼ the sandwich.'
PL	NOM	*Syötkö **voileivä-t**?* eat.2SG.Q sandwich-PL.NOM 'Will you eat the sandwiches?'	PTV	*Syötkö **voileip-i-ä**?* eat.2SG.Q sandwich-PL-PTV 'Are you eating (the) sandwiches?'
PRON	ACC	*Näet **minu-t**.* see.2SG 1SG-ACC 'You see me.'	PTV	*Etkö näe **minu-a**?* NEG.2SG.Q see.CNG 1SG-PTV 'Don't you see me?'

nominative or genitive, also titled 0-*accusative* and *n-accusative*, respectively (Vainikka & Brattico 2011). For instance, negated verbs generally only accept partial objects, and the use of the total object would be ungrammatical even if the object was intended as definite. 1st and 2nd person imperatives can accept both total and partial objects, but a total object must be in the nominative.

The cases of singular and plural total and partial objects in Finnish are summarized in Table 2. Personal pronouns are set apart from other nouns because they behave differently in terms of case assignment, in that they have a special total object suffix *-t*, for example *minu-t* [1SG-ACC].[6]

The situation is similar within DSM, in that the clausal subject can be in either the nominative or partitive. However, the domains of partitive subjects are more limited, as they generally appear only in intransitive clauses. Partitive subjects in transitive clauses are uncommon or atypical (A. Hakulinen et al. 2004: § 916, § 919). In a subtype of the intransitive clause, the *existen-*

[6] In Finnish grammar tradition, the case in question is frequently referred to as the "accusative", which can at other times be used as a catch-all nomenclature for all non-partitive object cases (e.g. L. Hakulinen 1961: 342). Therefore, because of its ambiguity, the term is problematic both in the context of this paper and in Finnish grammar. Furthermore, in Saami languages, personal pronouns do not differ from other nominals in this respect, and "accusative" refers to the object case form of all nouns.

Table 3. Subject cases in Finnish

		Nominative		Non-nominative
SG	NOM	***Vesi*** *on lasissa.* water be.3SG glass.INE 'The water is in the glass.'	PTV	***Vet-tä*** *on lasissa.* water-PTV be.3SG glass.INE 'There's some water in the glass.'
		Poika *lähtee.* boy leave.3SG 'The boy leaves ∼ is leaving.'	GEN	***Poja-n*** *täytyy lähteä.* boy-GEN must.3SG leave.INF 'The boy must leave.'
PL	NOM	***Koira-t*** *juoksevat ulkona.* dog-PL.NOM run.3PL outside 'The dogs are running outside.'	PTV	***Koir-i-a*** *juoksee ulkona.* dog-PL-PTV run.3SG outside 'There are dogs running outside.'
		Poja-t *lähtevät.* boy-PL.NOM leave.3PL 'The boys leave ∼ are leaving.'	GEN	***Poik-i-en*** *täytyy lähteä.* boy-PL-GEN must.3SG leave.INF 'The boys must leave.'
PRON	NOM	*Onneksi minulla on* ***talo****.* luckily 1SG.ADE be.3SG house 'Luckily I have a house.'	ACC	*Onneksi minulla on* ***sinu-t****.* luckily 1SG.ADE be.3SG 2SG-ACC 'Luckily I have you.'

tial clause (see § 5.2.1), the subject is most often assigned the partitive in the presence of negation, similarly to negated objects in transitive clauses. Finnish genitive subjects are restricted to a few special clause types, and their equivalents in Aanaar Saami are mostly left outside the scope of this paper. The cases of singular and plural subjects in Finnish are summarized in Table 3.

Saami languages at large do not exhibit DAM (Beronka 1940: 134 ff.; Bartens 1972: 19, 30).[7] In fact, they serve as a conspicuously pure example of a prototypical nominative-accusative language (Kittilä & Ylikoski 2018: 458–461). For example, traditional Aanaar Saami (vis-à-vis new Aanaar Saami) employs straightforward argument marking and agreement: nominal subjects are in the nominative case, nominal objects are in the accusative case, and the verb agrees with the subject in number (singular, dual, plural) and in person.

[7] The only exception among the Saami languages is South Saami, where DOM conditioned on definiteness is attested in nominal plural objects: definite objects appear in the accusative and indefinite ones in the nominative, for example *Læjsa* ***aahk-ide*** *damta* [PN grandmother-PL.ACC know.3SG] 'Lisa knows the grandmothers (that we just mentioned)' vs. *Læjsa* ***aahka-h*** *damta* [PN grandmother-PL.NOM know.3SG] 'Lisa knows grandmothers' (Kroik 2016).

Subject NPs prototypically employ the nominative plural suffix *-h* to denote non-singular number; there are no distinct dual case suffixes.

Predicate verbs generally agree with the subject NP in number and person in both Finnish and Saami. However, there are a few caveats. First of all, Finnish non-nominative subjects do not trigger agreement, which causes the verb to appear in default agreement, that is, the 3rd person singular (Huumo 2003: 462), as can be seen in Table 3. In traditional Aanaar Saami, the subject NPs, despite appearing in the nominative case, may trigger only partial verbal agreement. This refers to a type of agreement whereby the verb appears in the 3rd person singular form with singular subjects and in the 3rd person plural form with dual and plural subjects. The boundary appears to lie between animate and inanimate subjects, so that animate subjects, particularly with specific human referents, trigger full agreement, whereas inanimate subjects generally trigger partial agreement, as in example (1a). Subject NPs denoting animals, sometimes even humans, accept both full and partial agreement, as seen in (1b) (Toivonen 2007):

(1) a. *Riddoost láá kyehti keeđgi.*
beach.LOC be.3PL two rock.GEN

'On the beach are two rocks.'

b. *Táálust lava ∼ láá kyehti ulmuu.*
house.LOC be.3DU ∼ be.3PL two person.GEN

'There are two people in the house.' (Toivonen 2007: 230–231)

3.2 DAM as pattern replication

In this paper, the emerging DAM in Aanaar Saami is examined in the framework of contact-induced structural replication, specifically *pattern replication* or *PAT*. PAT refers to the replication of language structures from a model language to a replica language so that the replica language's internal grammatical structures are reorganized without the replication or borrowing of *phonological matter* or *MAT* (e.g. Matras & Sakel 2007; Sakel 2007).

In other words, the contact-induced changes in Saami argument marking do not involve borrowing the shape of the Finnish genitive (*-n*) or partitive case morphemes (*-A*, *-tA*) but merely the abstract pattern of DAM, which is then mapped onto existing Aanaar Saami structures. Employing Matras &

Sakel's (2007) concept of *pivot-matching*, I examine the data from new Aanaar Saami and identify the syntactic pivot structure in Finnish that is used in the replication of DAM in Aanaar Saami.

Considering that Finnic DAM is very complex and typologically quite unique, the transfer of such a system to another language may seem unlikely. However, it turns out that contact-induced DAM systems are attested around the world, for example in Arawakan and Basque (Mardale & Karatsareas 2020). In fact, pattern replication in general has been observed to be rather unconstrained: all domains of language structures and use can be affected, which can lead to a high degree of interlingual structural convergence (Heine & Kuteva 2005: 261).

4 Data

The data is a convenience sample and comprises a little under 300 sentences, in which I have detected a deviation from the canonical Aanaar Saami syntax rules governing the subject and the object.[8] It has been collected from different sources, mostly from younger language users approximately 20–30 years old, and it includes both spoken and written language.

The upper-case code in parentheses is used in the example sentences to indicate the source type. The spoken language portion of the data consists of interviews from a field trip to Aanaar in February 2020 (INTER) and broadcasts of Yle Sápmi, the Finnish Broadcasting Company's Saami-language radio and television (YLE). The data from written sources includes sentences from manuscripts of both literary prose (LIT) and scientific/scholarly texts (SCI), Aanaar Saami Wikipedia articles (WIKI), and the International Sámi corpus (SIKOR). Any examples found from other studies focusing on Finnish influence on Aanaar Saami are indicated by a standard citation.

The examples in this paper have been pseudonymized and/or slightly edited if the original sentence contains personal names or other details that might reveal the person's identity. The Aanaar Saami community is very small, and even little details may be enough to divulge this information. The intention here is not to draw attention to any individual speaker but to examine an ongoing process of language change on a systemic level.

[8] In some types of Aanaar Saami odd-syllable noun stems (for example *puttâl* 'bottle'), the genitive/accusative singular is syncretic with the nominative singular. I have excluded such cases, since it is impossible to determine which form is the intended one.

5 DAM in new Aanaar Saami

As discussed above in § 2.2, Aanaar Saami and Finnish resemble each other morphosyntactically to a great extent. Since there exist no in-depth treatises of Aanaar Saami syntax beyond some (preliminary) grammar sketches, it is therefore practical to adopt the description of Finnish clause types as the starting point. Another solution would be to apply existing analyses of North Saami syntax to Aanaar Saami (e. g. O. H. Magga 1978; Sammallahti 2005), but since we are already dealing with two languages throughout the paper, introducing grammatical analyses of a third seems superfluous.

5.1 Transitive clauses

Transitive clauses in both Finnish and Aanaar Saami are divalent and with the unmarked word order SVO, meaning the sentence-initial position is occupied by the syntactic subject. The basic order can vary for syntactic and information structural reasons – for example, the syntactic object may be fronted to the sentence-initial position – but usually the case marking of nominal clausal elements differentiates the parts of speech, such as *Matti syö kalaa* [PN eat.3SG fish.PTV] 'Matti eats fish' (neutral) vs. *Kalaa Matti syö* [fish.PTV PN eat.3SG] 'It's fish that Matti eats' (fronted object) vs. *Syökö Matti kalaa?* [eat.3SG.Q PN fish.PTV] 'Does Matti eat fish?' (interrogative clause) in Finnish.

The parameters of Finnish DOM were discussed earlier in § 3.1. In traditional Aanaar Saami, DOM is not attested, and therefore factors such as definiteness, boundedness, or telicity had no bearing on the case of the object, which was always accusative, as in (2a). In the Finnish translation in (2b), both nominative and partitive objects are possible, although in this instance the nominative seems more probable, because the number of tickets is more reasonably understood as bounded rather than indeterminate:

(2) a. *Karttâvetteđ uástiđ uđđâ liipuid.*
end_up.2PL buy.INF new.ATTR ticket.PL.ACC

'You will have to buy new tickets.' [SIKOR]

b. *Joudutte ostamaan uudet liput ∼ uusia*
end_up.2PL buy.INF new.PL.NOM ticket.PL.NOM ∼ new.PL.PTV
lippuja.
ticket.PL.PTV

'You will have to buy new tickets ∼ some new tickets.'

Based on the observations of Olthuis (2018), it was to be expected that at least one type of non-canonical object, total objects in the nominative plural, appears in the data. Even though my data lends itself primarily to qualitative analysis, it does seem that these types of non-canonical objects are most common. Most of the occurrences of nominative plural objects can be characterized as telic or bounded. The boundedness can be intrinsic to the meaning of the verb, as in example (5), or a property of the entire clause, as in (6). Sometimes the boundedness of the situation is further reinforced by an adverbial phrase that denotes the endpoint of the movement or the resulting state, such as *ucce pinon* in (3) and *oovtâ sajan* in (4):

(3) *N lâi čokkim M* ***reeivâ-h*** *ucce pinon.*
PN be.PST.3SG gather.PST.PTCP PN.GEN letter-PL.NOM small stack.ILL

'N had gathered [all of] M's letters in a small stack.' [LIT]

(4) *Stuárráámus ulmen lii nuurrâđ* ***nomâttâsa-h*** *oovtâ*
big.sprl purpose.ess be.3sg collect.inf name-pl.nom one.gen
sajan.
place.ill

'The main purpose is to collect [all] the names in one place.' [SCI]

(5) *Kiđđuv kyeddim ääigi niŋálâsah* […] *hilgoh*
in.spring calving.GEN time.GEN female.PL.NOM […] abandon.3PL
oovdeb ive ***vyesi-h****.*
previous year.GEN calf-PL.NOM

'In the spring at calving time cows […] abandon the previous year's calves.' [WIKI]

(6) *Suoi laiđijn* ***pyerá-h*** *šalde paijeeln.*
3DU lead.PST.3DU bicycle-PL.NOM bridge.GEN over

'The two of them walked the bicycles over the bridge.' [LIT]

Total objects are not restricted to semantically telic situations but are also used with atelic verbs denoting static states, such as *uáiniđ* 'see', *tubdâđ* 'know, be familiar with; feel' and *mušteđ* 'remember', as in examples (7–9). Since such verbs in Finnish also take total objects and have thus been called *quasi-resultative* (e.g. Huumo 2001; Kiparsky 2001: 19, 31), this provides support to the hypothesis that the emerging system of DOM in Aanaar Saami is being copied more or less wholesale from Finnish.

(7) *Mut jooskâi ko ooinij M tuođâlii* ***muáđu-h****.*
but stop.PST.3SG when see.PST.3SG PN.GEN serious.GEN face-PL.NOM

'But [s/he] stopped when [s/he] saw M's serious face.' [LIT]

(8) *Nubeh tobdeh* ***poccuu-h*** *peljimeerhâ keežild.*
other.PL.NOM know.3PL reindeer-PL.NOM earmark.GEN because.of

'Others know the reindeers based on the earmark.' [SCI]

(9) *Tun kuittâg muštáh jieijâs* ***poccuu-h****.*
2SG however remember.2SG own reindeer-PL.NOM

'However, you remember your own reindeer.' [SCI]

DOM in the new Aanaar Saami data is prevalent in plural objects but less so in singular ones. The reason for this is likely morphological: as illustrated earlier in Table 1, the Aanaar Saami genitive and accusative are identical in form in the singular.[9] Thus, functionally both the Finnish genitive and partitive can correspond to the Aanaar Saami genitive-accusative case, which means that by default, DOM may not be possible in Aanaar Saami in environments where the choice in the Finnish equivalent sentence is between the genitive and partitive, as exemplified in (10).

[9] The only words where SG. GEN ≠ SG. ACC are the interrogative pronoun *kii* 'who' and the demonstrative pronouns such as *taat* 'this': *kiä-n* [who-GEN], *taa-n* [this-GEN] : *kiä-m* [who-ACC], *taa-m* [this-ACC]. However, there are only a few pronominal objects in the data in such syntactic environments where this type of DOM could appear and, contrary to expectation, some of them appear in the nominative singular. I have attributed these to slips of the tongue.

(10) a. *Mummo luki lehde-n/lehte-ä.*
grandma read.PST.3SG newspaper-GEN/newspaper-PTV

'Grandma read the [whole] newspaper/Grandma was reading the newspaper.'

b. *Ákku luuvâi loostâ.*
grandma read.PST.3SG newspaper.ACC

'Grandma read ∼ was reading the newspaper.' (constructed)

Furthermore, objects in the nominative singular in Finnish are restricted to syntactic environments where the verb does not fully agree with the subject, such as impersonal/passive verbs and necessive constructions,[10] whereas regular finite clauses have genitive total objects (Vainikka & Brattico 2011). Therefore, it stands to reason that nominative singular objects are less common overall. Aanaar Saami passive verbs, however, are not included in this category, because they are in fact derived intransitive verbs whose subject canonically corresponds to the object of the respective transitive verbs, that is, they exhibit object promotion: *Kumppi porá saavzâ* [wolf eat.3SG sheep.ACC] 'The wolf eats a ∼ the sheep' (transitive, active) → *Savzâ porroo* [sheep eat.PASS.3SG] 'The sheep is eaten' (intransitive, passive). This contrasts with Finnish, where the subject is omitted with passive verbs, but the object is not promoted. This is explored in more detail in § 5.2.2.

What we do find are some examples of necessive constructions that seem to have been structured similarly to Finnish, where many such modal verbs and constructions are monopersonal: the verb is inflected in 3rd person singular only, with the semantic subject appearing in the genitive case. This Finnish influence on Saami necessive constructions may actually go back further in time, as similar use of genitive subjects has been attested in written Aanaar Saami in mid-1950s (see Bartens 1972: 55) as well as in some western dialects of North Saami (Valtonen 2017: 215). In examples (11a) and (12a), the object is in the nominative case, and the semantic subject, if it is overt, is in the genitive; compare the Finnish translations in (11b) and (12b):

[10] These are constructions that have a monopersonal verbal expression with a modal meaning as the predicate and the main verb in an infinitive form. The sentence-initial position may be occupied by the semantic subject in the genitive case (Jaakola 2004: 258 f.), for example *minu-n täyty-y lähte-ä* [1SG-GEN must-3SG leave-INF] 'I must leave'.

(11) a. ***Suu*** *koolgâi* *luoihâttiđ* ***tuŋkki*** *kuálásteijest.*
3SG.GEN must.PST.3SG borrow.INF jack fisherman.LOC

‘S/he had to borrow a jack from the fisherman.’ (Seipiharju 2021: 21)

b. ***Hänen*** *täytyi* *lainata* ***tunkki*** *kalastajalta.*
3SG.GEN must.PST.3SG borrow.INF jack fisherman.ABL

‘id.’

(12) a. *Koolgij* *ain* *väldiđ* ***kiinii*** *fáárun.*
must.PST.3SG always take.INF someone along.ILL

‘One always had to take someone [else] along.’ [YLE]

b. *Täytyi* *aina* *ottaa* ***joku*** *mukaan.*
must.PST.3SG always take.INF someone along.ILL

‘id.’

Often the nominative singular objects in the data appear in syntactic environments where either the genitive or partitive – the default object cases – are expected even in Finnish. However, a large proportion of these seem to arise from either uncertainty about the word stem type or sometimes just a sporadic error. For example, words such as *ceelhâ* ‘sentence, clause’, *peerâ* ‘family’, and *čunoi* ‘sand’ have apparently been mistaken for regular even-syllable or odd-syllable stems and inflected as such: ***celhâ* ∼ *ceelhâ* : *ceelhâ*, ***peeŗâ* : *peerâ*, and ***čunoi* : *čunoi* [NOM : GEN/ACC]. In traditional Aanaar Saami, these words belong to a class of contracted noun stems that exhibit the strong consonant grade and a vowel change *-â*, *-oi* > *-uu-* in the oblique stem, such as in *ceelhâ* : *celkkuu*, *peerâ* : *perruu*, *čunoi* : *čunnuu*. At any rate, most of these cases can be attributed to confusion in morphophonology, not syntax, so they are outside the scope of this paper.

5.2 Intransitive clauses

Intransitive clauses in Aanaar Saami can be divided into multiple subtypes, which is why I will treat each type separately. The types of intransitive clauses examined in this paper are 1) existential clauses and 2) intransitive clauses containing a passive verb. In traditional Aanaar Saami, intransitive clauses

are monovalent and their verbs generally agree with the subject in number and person (but see the remark on full vs. partial agreement in § 3.1).

I begin by delineating the features of each clause type as they appear in existing grammars and other grammatical descriptions of traditional Aanaar Saami. After that, I will provide examples from newer data and highlight the differences.

5.2.1 Existential clauses

Existential clauses are defined here in the vein of Haspelmath (2021) as constructions whose purpose is to introduce new referents into the discourse. The new referent or *existent* is the subject, even though for a subject it is often non-prototypical: it does not serve as the starting point of the predication or actively participate in the situation, and it is often indefinite and unbounded. For this reason, its full subjecthood has often been questioned in analyses of Finnish, and the NP is given another label, such as *internal subject*, *e-theme*, or *e-subject*, to differentiate it from the canonical subject (Kiparsky 2001; Huumo 2003; A. Hakulinen et al. 2004: § 923).

Using Haspelmath's (2021) terminology, the prototypical existential clause construction in both Aanaar Saami and Finnish can be described as *existent-postposing*. In Aanaar Saami, the unmarked word order in existential clauses is AVS, where A is an adverbial, V is a copula, and S is the syntactic subject, for example *Kárbá-st lii kandâ* [boat-LOC be.3SG boy] 'There is a boy in the boat'. The adverbial is typically a noun in the locative case, but it may also be a locational adverb such as *olgon* 'outside' or *tääbbin* 'here', or a postpositional phrase such as *viäsu tyehin* [house.GEN behind] 'behind the house'. Contrary to typical adverbials of place in other clause types, the ones in existential clauses can be characterized as obligatory, that is, they are diagnostic members of the existential construction, although they can be omitted if inferable from context, as in (13a) and (14).

In terms of syntactic structure, possessive clauses are identical to existential clauses, but they have a possessor instead of location as the adverbial, for example *Kaandâ-st lii käärbis* [boy-LOC be.3SG boat] 'The boy has a boat'. For the purposes of this paper, possessive clauses are subsumed under existential clauses in Aanaar Saami.

Finnish existential clauses are structurally very similar to Aanaar Saami ones: The sentence-initial position is occupied by an adverbial that denotes location, either concrete or metaphorical, and the syntactic subject comes after

the copula, for example *Piha-lla on poro* [yard-ADE be.3SG reindeer] 'There is a reindeer in the yard'. One of the peculiarities of the Finnish existential sentence is the possibility of a partitive subject when the subject is indefinite or negated. The subject does not trigger agreement in the verb, which is always in the 3rd person singular even in the presence of a nominative subject (A. Hakulinen et al. 2004: § 893–894).

Judging from the new Aanaar Saami data, the use of partitive subjects in Finnish existential clauses seems to have paved the way for the possibility of accusative subjects in Aanaar Saami. Non-canonical subjects of existential clauses alone account for about half of all non-canonical subjects and objects in the data. I have examined the clauses based on a few parameters: word order; sentence-initial position; the predicate verb's lexeme, polarity, and number agreement; and the number and case of the e-subject and its possible adjuncts.

Since word order is one of the diagnostic criteria in discerning existential clauses from what Sammallahti (2005: 205) labels as *localizing clauses*[11] in North Saami, it is to be expected that most clauses in the data have the prototypical order AVS. Non-prototypical word order arises in situations such as in (13a), where the e-subject has been moved to the sentence-initial position because it is definite. However, despite the subject being definite, it still appears in the accusative. This structure has very likely been copied from Finnish, where negation usually triggers a partitive e-subject (see 13b).

(13) a. ***Mielhi*** *ij* *innig* *lah.*
milk.ACC NEG.3SG anymore be.CNG

'The milk is gone', "The milk is no more" (Olthuis 2018)

b. ***Maito-a*** *ei* *enää* *ole.*
milk-PTV NEG.3SG anymore be.CNG

'id.'

The most common predicate verb is the generic copula *leđe*, although there are others too, such as *kavnuđ* 'be found', *eellið* 'live', *aassâđ* 'reside', and *puáttið* 'come', as in examples (14–16). When used existentially, these verbs

[11] Compare existential *Kuáđist lâi ákku* [tent.LOC be.PST.3SG grandmother] 'There was a grandmother in the tent' vs. localizing *Ákku lâi kuáđist* [grandmother be.PST.3SG tent.LOC] 'Grandmother was in a ∼ the tent'.

undergo some semantic bleaching and function like the copula: they express either existence or coming into existence.

(14) *Kal mij tuáivup ete* ***puátá*** *lase párná-id teikkâ pärni.*
indeed 1PL hope.1PL COMP come.3SG more child-PL.ACC or child.

'We do hope to have more children or a child.' [YLE]

(15) *Kuáđist* ***kavnui*** *meid puáris liiti-jd.*
tent.LOC find.PASS.PST.3SG also old.ATTR dish-PL.ACC

'There were also old dishes in the tent.' (Olthuis 2018)

(16) *Jaavrijn* ***iälá*** *maaŋga-h ereslágán-eh iälán-eh.*
lake.PL.LOC live.3SG many-PL.NOM different-PL.NOM organism-PL.NOM

'There are many kinds of organisms living in the lakes.' [WIKI]

The examples (14–16) also display another sign of Finnish influence, namely the emergence of default agreement. Following Toivonen (2007), Aanaar Saami should have no default agreement and the expected agreement type with plural non-singular subjects should be partial, that is, the verb should appear in the 3rd person plural. Nonetheless, new Aanaar Saami shows many cases of predicate verbs defaulting to the 3rd person singular even in the presence of a plural subject. Even though Aanaar Saami conjugation does include the dual number, only personal pronouns have special dual forms. Also, as remarked by P. Morottaja & Toivonen (in preparation), dual agreement in existential and possessive constructions does not sound natural, and even when it occurs, its conditions are more restricted than those of the singular and plural (see § 3.1).

However, plural agreement is also not uncommon in the new data, where the copula in the 3rd person plural co-occurs with a non-canonical subject (= accusative plural), as evidenced by examples (17–19). This combination of plural verbal agreement and non-canonical subject might be considered a hybrid between Finnish and traditional Aanaar Saami existential constructions.

(17) *Mist* ***lää*** *eenâb* ***spesiaaltábáhtusâ-id***.
1PL.LOC be.3PL more special.event-PL.ACC

‘We have more special events.’ (Seipiharju 2022: 37)

(18) *škoovlâst* ***lää*** *ennuv* ***skipár-ijd***
school.LOC be.3PL many friend-PL.ACC

‘There are many friends at school’ [INTER]

(19) *motomijn Laapi jaavrijn* ***lää*** ***šaapšâ-id***
some.PL.LOC Lapland.GEN lake.PL.LOC be.3PL lavaret-PL.ACC

‘In some of Lapland’s lakes there are lavarets (Coregonus lavaretus)’ [WIKI]

An inverse hybrid type can be seen in (16) and (20–23), where the 3rd person singular copula occurs with a plural e-subject in the nominative. In other words, there is no number agreement, but the e-subject appears in the canonical case. A nominative e-subject may sometimes occur with verbal negation, as in (22–23), which would not be possible in Finnish.[12]

(20) *ko* ***lii*** *kielâpiervâlist lamaš ennuv* ***nuora-h***
when be.3SG language.nest.LOC be.PST.PTCP many young-PL.NOM
já ***párnáá-h*** *tääl*
and child-PL.NOM now

‘When there have been many youngsters and children in the language nest now’ [INTER]

(21) *Sust* ***lâi*** *segis* ***vuopta-h***.
3SG.LOC be.PST.3SG thin.ATTR hair-PL.NOM

‘S/he had thin hair.’ [LIT]

[12] Strictly speaking, Finnish does allow nominative e-subjects in negated clauses under certain conditions, whereby the focus of the negation is not the subject; its existence is not negated but instead the proposition is that it is located in a particular place or has a certain quality (A. Hakulinen et al. 2004: § 918). However, the Aanaar Saami data does not easily lend itself to such an interpretation.

(22) ***ij*** ***leh*** *lamaš* *taggaar* *tego* ***sárnumohtâvuođa-h***
NEG.3SG be.CNG be.PST.PTCP such like speaking.context-PL.NOM

'There have not been, like, such opportunities to speak [Aanaar Saami]' [INTER]

(23) *Taan* *uásist* ***ij*** ***lah*** *pennui* ***nooma-h***.
this.GEN part.LOC NEG.3SG be.CNG dog.PL.GEN name-PL.NOM

'There are no dogs' names in this section.' (Olthuis 2018)

5.2.2 Passive verb clauses

In this paper, the passive in Aanaar Saami is understood in a narrow sense, encompassing only verbs formed with the derivational suffix *-u-* (: *-o-* ∼ *-u-* : *-uvvo-*). This is the most productive means of turning active verbs into passive ones in Aanaar Saami, for example *puurrâ-đ* [eat-INF] 'to eat' → *purr-u-đ* [eat-PASS-INF] 'to be eaten' : *porr-oo* [eat-PASS.3SG], *vuolgât-iđ* [send-INF] 'to send' → *vuolgâtt-u-đ* [send-PASS-INF] 'to be sent' : *vuolgâtt-uvvoo* [send-PASS.3SG]. There are other derivational suffixes that form verbs with passive or related meanings, but they are not nearly as frequent or productive and may carry a collateral meaning, such as adversative or automative. Furthermore, verbs derived with the *u*-suffix tend to act as functionally equivalent to the Finnish passive forms.

Siewierska (2013) defines a construction as passive if it fulfils the following conditions:

i. it contrasts with another constuction [sic], the active;

ii. the subject of the active corresponds to a non-obligatory oblique phrase of the passive or is not overtly expressed;

iii. the subject of the passive, if there is one, corresponds to the direct object of the active;

iv. the construction is pragmatically restricted relative to the active;

v. the construction displays some special morphological marking of the verb.

Of these properties, the traditional Aanaar Saami passive displays all: morphologically, passive verbs are always secondary vis-à-vis the active and are derived from active verbs with a derivational morpheme. The derived *u*-verbs behave inflectionally as any other verb in that they have a complete paradigm, and in terms of semantics, some of them have additional lexicalized meanings, for example *tiettuđ* 'be known; be visible; appear' ← *tiettiđ* 'know'. In terms of syntax, the direct object of the active verb appears canonically as the subject of the passive, or if the active verb is intransitive, the passive is avalent. The passive verb fully agrees with its syntactic subject. The subject of the active verb is often omitted altogether, or it may appear as an adjunct in the illative or, rarely, in the locative (Bartens 1972: 22–23, 92, 127).

In terms of argument marking, the Finnish passive differs from the Aanaar Saami passive in that passivization in Finnish does not promote the direct object (P) of the active into a subject (S) but merely removes the A argument. For these reasons, the Finnish passive has also been called the *4th person* or *impersonal* (Blevins 2003; Kelomäki 2019).

Examples (24–26) illustrate the use of the *u*-passive in traditional Aanaar Saami. They show, as expected, agreement with the syntactic subject as well as omission of the semantic agent (24–25) or its inclusion as an illative oblique (26):

(24) *teehin* *lii* ***vuord-u-m*** *reŋgâ*
here.ILL be.3SG wait-PASS-PST.PTCP farmhand

'A farmhand has been expected here.' (Itkonen & Laitinen 1992: 149)

(25) *Kii* *talle* *lâi* *pappân* *Anarist* *ko* *tun*
who then be.PST.3SG priest.ESS Aanaar.LOC when 2SG
vihk-oj-ih*?*
marry-PASS-PST.2SG

'Who was the priest in Aanaar at the time when you were married?' (Itkonen & Laitinen 1992: 183)

(26) *sun* *masa* ***porr-oo*** *tooid* *kuobžááid*
3SG almost eat-PASS.3SG DEM.PL.ILL bear.PL.ILL

'He is almost eaten by the bears' (Itkonen & Laitinen 1992: 51)

In light of the new data, Aanaar Saami *u*-passives appear as a distinctly written language form, since there is only one attestation of a spoken passive verb. This is not an expected result, so there is a need to investigate the frequency of passive verbs in spoken language more thoroughly.

At any rate, new Aanaar Saami data shows some clear instances of DSM with passive verbs which may also take the accusative case instead of the canonical nominative. This also seems to be motivated by the Finnish usage: if the sentences were translated into Finnish, partitive objects would be expected in all examples. Accusative subjects with passive verbs appear in various syntactic environments, such as in affirmative and negative clauses and both preverbally and postverbally, and the verbs often exhibit the default 3rd person singular agreement even in the presence of plural subjects (see 27–29), although this is not always the case, as shown in (31).

(27) *já muu postâloován* ***lii*** ***mááláj-u-m*** ***poccu-id***
and 1SG.GEN mailbox.ILL be.3SG paint-PASS-PTCP reindeer-PL.ACC

‘And reindeer have been painted on my mailbox’ [LIT]

(28) *Sálttáá* ***ja***[***a***]***vri-jd*** ***kočod-uvvoo*** *sälttijävrin.*
salty.ATTR lake-PL.ACC call-PASS.3SG salt.lake.ESS

‘Salty lakes are called salt lakes.’ [WIKI]

(29) ***Nomâttâsâ-id*** ***ij*** ***lah*** *ovdil* ***tutk-u-m****.*
name-PL.ACC NEG.3SG be.CNG before study-PASS-PST.PTCP

‘Names have not been studied before.’ [SCI]

(30) *Ko lii saahâ sämikielâst,* ***ij*** ***uáivild-uu***
when be.3SG speech Saami.language.LOC NEG.3SG mean-PASS.CNG
tuše ***oovtâ*** ***kielâ****.*
only one.ACC language.ACC

‘When we are talking about Saami language, we do not mean just one language.’ [SIKOR]

(31) ***Algu-id*** ***vuolgâtt-uvvoj-eh*** *täärhib* *valmâštâlmân.*
motion-PL.ACC send-PASS-3PL accurate.CMPR preparation.ILL

'Motions are sent to a more detailed preparation.' [SIKOR]

Unfortunately, what we do not find in the data are subjects other than the 3rd person, so it remains uncertain whether 1st and 2nd person subjects would still trigger agreement with passive verbs. It also appears that the nominative subjects in the passive have become limited to the same contexts where nominative objects are used in Finnish, that is, in bounded situations, as in (32):

(32) ***Táálu-h*** *lâi* *huksej-u-m* *aaibâs* *luodâ* *roobdân*
house-PL.NOM be.3SG build-PASS-PTCP just road.GEN edge.ILL

'The houses were built right on the edge of the road' [LIT]

6 Discussion

Above, I have described the argument marking systems of Finnish and traditional Aanaar Saami and used them as a baseline to analyse the more recently collected Aanaar Saami data. It shows that a novel, Finnish-influenced argument marking system is beginning to emerge in modern Aanaar Saami. In this section, I will firstly summarize the observations of DOM, then those of DSM, and finally I will examine the DAM system altogether.

6.1 DOM

Many of the conditions of DOM that can be inferred from the data are as expected. In most cases, the motivation for choosing the non-canonical nominative object instead of the canonical accusative in the plural can be readily traced back to the same conditions that apply in the model language of Finnish, namely totality and telicity: the action is carried out upon all the referents, such as in examples (3–5), or the action reaches its endpoint, such as in (6).

Morphologically speaking, the non-canonical objects mostly appear in the plural. The explanation is that this is where the maximal differentiation of the total and partial object in Aanaar Saami is possible. In Finnish, total plural objects appear in the nominative plural whereas partial objects take the partitive

plural. The functionally closest case to the Finnish partitive plural in Aanaar Saami is the accusative plural, so it is selected as its equivalent.[13] However, in the singular the situation differs: Finnish singular partial objects do also appear in the partitive, but by default, total objects are assigned the genitive, whereas nominative total objects are restricted to specific syntactic environments (see § 3.1). Thus, the Finnish genitive should correspond to the Aanaar Saami genitive and the Finnish partitive to the Aanaar Saami accusative, but in Aanaar Saami the genitive and accusative are always syncretic in the singular forms of nominals except for few pronouns. This leads to a situation in many clause types where the Finnish singular total and partial object case both equate to the Aanaar Saami genitive-accusative singular, making it impossible to render the total vs. partial distinction.

Unfortunately, I have not been able to find examples of 2nd person imperative verbs with semantically singular total objects. In this instance, a nominative object would be expected in Finnish (see § 3.1), so such examples might have provided a clue as to whether totality plays a role in the singular in new Aanaar Saami as well.

As explained in § 5.1, there are many singular nominative objects in the data that do not conform to the Finnish model of DOM. These cases may therefore be better interpreted as confusion in inflectional types. Aanaar Saami has a typologically complex morphophonology, but any emerging variation therein should be studied separately.

6.2 DSM

The conditions of DSM in the new data have certain expected features as well. Accusative subjects are especially common in existential clauses whose main purpose is to predicate the existence of a referent in a location or someone's possession. In these types of clauses, predicate verbs are often semantically void, and thus it is no surprise that the copula is the most common choice. There are a handful other verbs too, but their meanings also range from existence to coming into existence.

[13] Coincidentally, the Aanaar Saami accusative plural suffix *-jd* and the Finnish partitive plural suffix *-i-tA* are etymological cognates, both going back to Finno-Saamic **-j-tA* [PL-PTV] (Sammallahti 1998: 68). Kittilä & Ylikoski (2018: 476) suggest that the common origin may play a role in why bilingual Saami-Finnish speakers and Finnish learners of Saami "often tend to equate the Saami genitive-accusative with the Finnish partitive".

The mapping of Finnish subjects onto Aanaar Saami results in a different configuration than in that of objects: genitive subjects appear in a very limited number of constructions and are indifferent with regard to totality vs. partiality, so the problem of case syncretism is not present here.

It is therefore somewhat puzzling as to why around 75% of all existential clauses in the data have plural e-subjects. The reason for this could be a bias in data collection or that, for some reason, singular partial subjects are not as frequent overall. When they do appear, they are usually under negation, as in (13a) or (33).

(33) *Sust* *ij* *lamaš* ***perruu***.
3SG.LOC NEG.3SG be.PST.CNG family.ACC

'S/he did not have a family.' (Seipiharju 2021: 29)

The subjects of Aanaar Saami passive verbs prove to be an interesting case. I have chosen to treat them as subjects based on their syntactic analysis in traditional Aanaar Saami despite the fact that, in the model language of Finnish, the NP arguments of passive verbs cannot easily be considered subjects (cf. Ylikoski 2022a). However, Finnish singular total objects of passive verbs cannot appear in the genitive but instead work syntactically in the same way as objects with 1st and 2nd person imperatives and in necessive constructions, in that they are assigned the nominative case. This means that they align better with the subjects of existential clauses in their behaviour. Nevertheless, taking semantics into account, I find there is reason to argue that it becomes unclear which syntactic role the NPs of passive verbs should be assigned in new Aanaar Saami.

Even though my focus has been on case forms, some remarks concerning verb agreement can be made as well. In Aanaar Saami, when the subject or object of a clause is a quantifier phrase of the type [cardinal numeral + noun], the numeral overrides the default syntactic case and assigns either the genitive singular or partitive to the noun being quantified: *ulmuu-h* [person-PL.NOM] 'people' vs. *kulmâ* ***ulmuu*** [three person.GEN] 'three people', *čiččâm* ***olmožid*** [seven person-PTV] 'seven people'. Thus, the resulting quantifier phrases are semantically non-singular but contain no overt dual/plural marker in the noun. (Nelson & Toivonen 2000.) In traditional Aanaar Saami, the predicate verb agrees with the semantic number, for example *Tobbeen* ***lää*** *kulmâ ulmuu* ∼ *čiččâm olmožid* [there be.3PL three person.GEN ∼ seven person.PTV]

'There are three ∼ seven people there', whereas in Finnish, these types of quantifier phrases typically do not agree with the predicate verb, so the verb is inflected in the 3rd person singular. This model seems to be giving rise to similar non-agreement in new Aanaar Saami data as exemplified in (34). The non-agreement is most common in existential clauses, but it is occasionally found in other types of clauses too, even when an overt plural marker is present (35). The common colloquial Finnish usage of singular verb forms with plural subjects in the 3rd person is very likely the reason why non-agreement is spreading to other types of clauses besides existential ones.

(34) *Suomâst* ***lii*** *ohtsis* *suullân* *188 000* *jävrid*
Finland.LOC be.3SG altogether around 188,000 lake.PTV

'In Finland there are altogether around 188,000 lakes' [WIKI]

(35) *Masa* *puohah* ***sá[á]rnui*** *sämikielâ.*
almost everyone.PL.NOM speak.PST.3SG Saami_language.ACC

'Almost everyone spoke Saami.' [LIT]

6.3 Explaining DAM in Aanaar Saami

If an Aanaar Saami speaker were to adopt both DOM and DSM from Finnish simultaneously, it begs the question of why it seems to be easier to find DSM than DOM in the new data. One of the reasons is very likely the syncretism of Aanaar Saami genitive and accusative singular described above, but I propose that this is also due to typological factors. The semantics and syntax of non-canonical subjects make them inherently non-prototypical: they are non-agentive, typically indefinite, and, with passive verbs, semantic patients or themes. Also, the default constituent order of existential clauses differs from the one in normal clauses, because the subject is positioned after the verb. On the other hand, non-canonical objects do not resemble subjects in other respects than perhaps definiteness, and transitive clauses usually already contain another, more subject-like constituent.

In sum, it appears that the emerging DAM in new Aanaar Saami can be adequately explained as pattern replication modelled after the Finnish total–partial distinction in subjects and objects. The pivotal feature of Finnish is the DAM that is then replicated in Aanaar Saami by assigning novel syntactic

	Finnish		Aanaar Saami
Total object			
singular:	genitive	→	genitive-accusative
	(nominative	→	nominative)
plural:	nominative	→	nominative
Partial object			
singular:	partitive	→	genitive-accusative
plural:	partitive	→	accusative
Total subject			
singular/plural:	nominative	→	nominative
Partial subject			
singular:	partitive	→	genitive-accusative
plural:	partitive	→	accusative

Figure 1. Mapping of Finnish grammatical cases to the Aanaar Saami ones in the new data

functions to the Aanaar Saami nominative and genitive-accusative cases, thus generating a DAM system in Aanaar Saami. A summary of the mapping of Finnish grammatical cases to the Aanaar Saami ones according to data from new Aanaar Saami is presented in Figure 1.

Following Sasse (1990: 32) and Matras & Sakel (2007), I propose that the reason why a bilingual Aanaar Saami and Finnish speaker attempts to replicate Finnish DAM in Aanaar Saami is communicative: they want to express the same thoughts in both languages. Being bilingual, DAM is already a part of their communicative repertoire, so they turn to creativity to satisfy their communicative needs (Heine & Kuteva 2005: 34–35).

It is often claimed that DAM is one of the most difficult aspects of Finnish grammar for L2 learners of Finnish, and this is corroborated by experimental evidence (e.g. Sikiö 2008; Göken 2012). I am inclined to hypothesize that this works the other way round as well: the DAM system is a deeply ingrained and integral part of the mental grammar of native-level speakers of Finnish. Therefore, if the other language is structurally close enough, providing the material for the replication of a DAM system, this replication is likely to happen. However, more research on the subject is certainly needed.

Nevertheless, there is inter-speaker and even intra-speaker variation in the Aanaar Saami DAM, considering that the same speaker may use both canonical and non-canonical argument marking. This suggests that the replication process is gradual rather than abrupt and, at the present stage, Aanaar Saami DAM appears to be an incipient category (see Heine & Kuteva 2005: 71): its use is not obligatory, and it is not recognized by grammarians or language planners as a grammatical entity but instead labelled as errors or interference (M. Morottaja 2007: 34; Olthuis 2018).

7 Conclusion

In this paper, I have examined the emerging differential argument marking (DAM) in new Aanaar Saami and compared it to Finnish, its apparent model. As it stands, the data from new Aanaar Saami indicates that some younger members of the Aanaar Saami language community, who are at the very least bilingual in Finnish and Aanaar Saami, have begun to use "formulas of equivalence" (Keesing 1991: 327) between grammatical categories when speaking or writing Aanaar Saami. This is done by calquing grammatical structures from Finnish to Aanaar Saami by the process of pattern replication (e.g. Matras & Sakel 2007), so that the functions and meanings of existing Aanaar Saami grammatical forms are adapted to correspond to their Finnish equivalents. This appears to be more prevalent when the syntactic subject exhibits properties of non-prototypical subjects, such as being non-agentive and indefinite, which hints that certain typological tendencies concerning subjecthood may also explain the more extensive propagation of non-canonical subjects. Nevertheless, it is the replication of the Finnish structures that can be identified as the immediate cause.

Since traditional Aanaar Saami does not have DAM based on totality–partiality distinction, a new system is emerging through reorganization: nominative and accusative cases that previously corresponded more or less with the grammatical relations of *subject* and *object* respectively have spread to new domains so that the function of the accusative may also cover partial subjects in intransitive clauses, and at least the nominative plural may signal a total object in transitive clauses. The situation in the singular differs from that in the plural due to case syncretism: Aanaar Saami does not differentiate between the genitive and accusative singular in nouns, which sets a constraint on how the Finnish genitive–partitive distinction in objects as a whole can be replicated.

It is also interesting to note that while the Aanaar Saami community at large has had close contacts with Finnish speakers for around 200 years by now, the language seems to have resisted the type of major syntactic influence of Finnish as investigated in this paper up until thirty years ago: there is no variation in subjects and objects mentioned in earlier descriptions of the language, nor have I been able to find such examples in any earlier text collections or recordings.

Lastly, I will illustrate the beginning and hypothetical endpoint of the grammatical change in argument marking with an example from my data. (36) is a modified version of the actual example sentence and serves to represent the situation in traditional Aanaar Saami without DAM. The examples in (37) show the situation in new Aanaar Saami after the development of DAM: (37a) is structurally identical to (36), but the semantics of the object has changed so that the accusative case is now only used for partial objects, while (37b) is the sentence as found in the data, with the object in the nominative to signal its totality.

(36) *N vaaldij puser-ijd fáárun*
PN take.PST.3SG sweater-PL.ACC along

'N took ((some or all of) the) sweaters with them'

(37) a. *N vaaldij puser-ijd fáárun*
PN take.PST.3SG sweater-PL.ACC along

'N took some sweaters with them'

b. *N vaaldij puser-eh fáárun*
PN take.PST.3SG sweater-PL.NOM along

'N took all of the sweaters with them' [LIT]

Aanaar Saami has a recently stabilized written standard that is taught at educational institutions; there is no spoken standard per se, but in practice much of the older dialects has been replaced by a spoken form influenced by the current written language. When we include both spoken and written Finnish in the picture, we can imagine a complex tug-of-war of interconnected language forms, of which some can foster the emergence of DAM and others hinder it. Therefore, whether the emerging DAM, or any other Finnish grammati-

cal calque for that matter, eventually becomes a normal part of Aanaar Saami grammar remains to be seen.

Acknowledgements

I wish to thank Jussi Ylikoski, Mariann Bernhardt, and two anonymous reviewers for their helpful comments and advice.

Abbreviations

1	1st person
2	2nd person
3	3rd person
ABL	ablative
ADE	adessive
ACC	accusative
ATTR	attributive form
CMPR	comparative
CNG	connegative form
COMP	complementizer
DEM	demonstrative pronoun
DU	dual
ELA	elative
ESS	essive
GEN	genitive
ILL	illative
INE	inessive
INF	infinitive
LOC	locative
NEG	negative verb
NOM	nominative
PASS	passive
PL	plural
PN	personal name
PST	past
PTCP	participle
PTV	partitive

Q question clitic
SG singular
SPRL superlative

References

Aikio, Ante & Arola, Laura & Kunnas, Niina. 2015. Variation in North Saami. In Smakman, Dick & Heinrich, Patrick (eds.), *Globalising sociolinguistics: Challenging and expanding theory*, 243–255. Routledge.

Bartens, Raija. 1972. *Inarinlapin, merilapin ja luulajanlapin kaasussyntaksi*. (Suomalais-Ugrilaisen Seuran toimituksia 148). Suomalais-Ugrilainen Seura.

Beronka, Johan. 1940. *Lappische Kasusstudien 2: Zur Geschichte des Nominativs, des Akkusativs, der allgemeinen und der äusseren Lokalkasus*. Oslo Etnografiske Museums Skrifter.

Blevins, James P. 2003. Passives and impersonals. *Journal of Linguistics* 39(3). 473–520.

Göken, Ina. 2012. *"Tiedän mik- mitä sana tarkoittaa": Suomea toisena kielenä puhuvien objektin sijanvalinta ja sijanvalinnan taustalla oleva päättelyprosessi*. Helsingin yliopiston suomen kielen, suomalais-ugrilaisten ja pohjoismaisten kielten ja kirjallisuuksien laitos. (Master's thesis). (http://urn.fi/URN:NBN:fi-fe2012112310021).

Grinevald, Colette & Bert, Michel. 2011. Speakers and communities. In Austin, Peter K. & Sallabank, Julia (eds.), *The Cambridge handbook of endangered languages*, 45–65. Cambridge University Press. https : / / doi . org / 10 . 1017 / CBO9780511975981

Hakulinen, Auli & Vilkuna, Maria & Korhonen, Riitta & Koivisto, Vesa & Heinonen, Tarja Riitta & Alho, Irja. 2004. *Iso suomen kielioppi*. Suomalaisen Kirjallisuuden Seura. (http://scripta.kotus.fi/visk/etusivu.php). (Accessed 2021-03-23).

Hakulinen, Lauri. 1961. *The structure and development of Finnish language*. Trad. Atkinson, John. (Indiana University Publications, Uralic and Altaic Series 3).

Haspelmath, Martin. 2021. Existential constructions and two types of comparative concepts: Construction-functions and construction-strategies. (Presented at the Descriptive Grammars and Typology 2, Paris). (https://zenodo.org/record/5758945). (Accessed 2021-12-08).

Havas, F. 2008. Unmarked object in the Uralic languages: A diachronic typological approach. *Linguistica Uralica* 43(1). 1–33. https://doi.org/10.3176/lu.2008.1.01

Heine, Bernd & Kuteva, Tania. 2005. *Language contact and grammatical change*. Cambridge University Press. https://doi.org/10.1017/CBO9780511614132

Huumo, Tuomas. 2001. Kvasiresultatiivilauseet: Esimerkki dynaamisesta subjektiivisesta käsitteistyksestä. *Puhe ja kieli* 21(2). 85–101.

— 2003. Incremental existence: The world according to the Finnish existential sentence. *Linguistics* 41(3). 461–493. https://doi.org/10.1515/ling.2003.016

Itkonen, Erkki & Laitinen, Lea. 1992. *Inarinsaamelaisia kielennäytteitä — Aanaarkiela čájttuzeh*. (Suomalais-Ugrilaisen Seuran toimituksia 213). Suomalais-Ugrilainen Seura.

Jaakola, Minna. 2004. *Suomen genetiivi*. (Suomalaisen Kirjallisuuden Seuran Toimituksia 995). (Doctoral dissertation).

Keesing, Roger M. 1991. Substrates, calquing and grammaticalization in Melanesian Pidgin. In Traugott, Elizabeth Closs & Heine, Bernd (eds.), *Approaches to grammaticalization, volume 1: Focus on theoretical and methodological issues*, 315–342. John Benjamins.

Kelomäki, Tapani. 2019. Suomen passiivin ja impersonaalin määritelmistä ja keskinäisistä suhteista. *Virittäjä* 123(1). 110–134. https://doi.org/10.23982/vir.49512

Kiparsky, Paul. 2001. Structural case in Finnish. *Lingua* 111(4). 315–376. https://doi.org/10.1016/S0024-3841(00)00035-8

Kittilä, Seppo & Ylikoski, Jussi. 2018. Some like it transitive: Remarks on verbs of liking and the like in the Saami languages. In Seržant, Ilja A. & Witzlack-Makarevich, Alena (eds.), *Diachrony of differential argument marking*, 455–480. Language Science Press. https://doi.org/10.5281/ZENODO

Koponen, Eino. 2022. Saami: General introduction. In Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena (eds.), *The Oxford guide to the Uralic languages*, 103–112. Oxford University Press. https : / / doi . org / 10 . 1093 / oso / 9780198767664.003

Korhonen, Mikko. 1964. Lapin murteiden keskinäisistä suhteista. In Ahvenainen, Jorma (ed.), *Lapin tutkimusseuran vuosikirja V*, 49–64.

Kroik, David. 2016. *Differential object marking in South Saami*. Department of Language Studies Umeå University. (Licentiate thesis).

Laakso, Johanna. 2010. Contact and the Finno-Ugric languages. In Hickey, Raymond (ed.), *The handbook of language contact*, 598–617. Wiley-Blackwell.

Lehtola, Veli-Pekka. 2012. *Saamelaiset suomalaiset: Kohtaamisia 1896–1953*. (Suomalaisen Kirjallisuuden Seuran Toimituksia 1351). Suomalaisen Kirjallisuuden Seura.

Magga, Ole Henrik. 1978. *Lokative læt-setninger i samisk*. (Dieđot 30(1)). Sámi institut'ta.

Magga, Rita & Pulska, Unna-Maari. 2019. *Davvisámegiela eatnigielat studeanttaid feaillat studeantaiskosiin jagiid 1997–2009*. University of Oulu, Giellagas Institute. (Master's thesis).
(http://urn.fi/URN:NBN:fi:oulu-201906182555). (Accessed 2023-04-19).

Malchukov, Andrej L. & Spencer, Andrew. 2008. Typology of case systems: Parameter of variation. In Malchukov, Andrej L. & Spencer, Andrew (eds.), *The Oxford handbook of case*, 651–667. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199206476.013

Mardale, Alexandru & Karatsareas, Petros. 2020. Differential object marking and language contact: An introduction to this special issue. *Journal of Language Contact* 13(1). 1–16. https://doi.org/10.1163/19552629-bja10001

Matras, Yaron & Sakel, Jeanette. 2007. Investigating the mechanisms of pattern replication in language convergence. *Studies in Language* 31(4). 829–865.

Mettovaara, Jukka & Ylikoski, Jussi. Forthcoming. Structural approach to language revitalization: Revival of Aanaar Saami. In Frick, Maria & Räisänen, Tiina & Ylikoski, Jussi (eds.), *Language contacts and discourses in the far north*. Palgrave Macmillan.

Morottaja, Matti. 1996. Anáraččat: Ávkkástallojuvvon álbmotoassi. *Ođđa Sápmelaš*.

— 2007. *Anarâškielâ ravvuuh*. (Kotimaisten kielten tutkimuskeskuksen julkaisuja 144). Päikkieennâm kielâi tutkâmkuávdáš.

Morottaja, Petter & Olthuis, Marja-Liisa. 2022. *Inarinsaamen taivutusoppi*. Sämitigge.

Morottaja, Petter & Toivonen, Ida. In preparation. Some notes on Inari Saami syntax and morphology. In Tamm, Anne & Vainikka, Anne (eds.), *Uralic syntax*. Cambridge University Press. (Draft version, 2016-06-10).

Müller, Merit. 2017. *Adjektiiviattribuutin vaihtelu Erkki Itkosen "Inarinsaamelaisissa kielinäytteissä"*. University of Tartu, Department of Estonian & General Linguistics. (Bachelor's thesis).

Nelson, Diane & Toivonen, Ida. 2000. Counting and the grammar: Case and numerals in Inari Sami. *The Leeds Working Papers in Linguistics and Phonetics* 8. 179–192. (Nelson, D. & Foulkes, P.).

Olthuis, Marja-Liisa. 2018. Suomâkielâ interferens sämikielâ ceelhâopâlijn ráhtusijn [Finnish interference in Aanaar Saami syntactic structures]. (Presented at the Sámegiela ja sámi girjjálašvuođa symposia, Oulu).

Olthuis, Marja-Liisa & Kivelä, Suvi & Skutnabb-Kangas, Tove. 2013. *Revitalising indigenous languages: How to recreate a lost generation*. (Linguistic Diversity and Language Rights 10). Multilingual Matters.

Pasanen, Annika. 2015. *Kuávsui já peeivičuovâ — Sarastus ja päivänvalo: Inarinsaamen kielen revitalisaatio*. (Uralica Helsingiensia 9). Helsingin yliopisto & Suomalais-Ugrilainen Seura.

— 2020. Uđđâsist šoddâm kielâ "Ett pånyttfött språk": Hur enaresamiskan återhämtade sig. (Presentation presented at the Meänkieliseminarium 13.–14.11.2020). (https://www.youtube.com/watch?v=bJXVbmzstY4). (Accessed 2023-04-19).

Rießler, Michael. 2007. Grammatical borrowing in Kildin Saami. In Matras, Yaron & Sakel, Jeanette (eds.), *Grammatical borrowing in cross-linguistic perspective*, 229–244. Mouton de Gruyter. https://doi.org/10.1515/9783110199192.229

— 2016. *Adjective attribution*. (Studies in Diversity Linguistics 2). Language Science Press.
(http://langsci-press.org/catalog/book/19). (Accessed 2021-03-23).

Rydving, Håkan. 2013. *Words and varieties: Lexical variation in Saami*. (Mémoires de la Société Finno-Ougrienne 269). Société Finno-Ougrienne.

Sakel, Jeanette. 2007. Types of loan: Matter and pattern. In Matras, Yaron & Sakel, Jeanette (eds.), *Grammatical borrowing in cross-linguistic perspective*, 15–29. Mouton de Gruyter. https://doi.org/10.1515/9783110199192

Sammallahti, Pekka. 1998. *The Saami languages: An introduction*. Davvi Girji.

— 2005. *Láidehus sámegiela cealkkaoahpa dutkamii*. Davvi Girji.

Sasse, Hans-Jürgen. 1990. Language decay and contact-induced change: Similarities and differences. In *Arbeitspapier 12*, 30–56. Institut für Sprachwissenschaft, Universität zu Köln.

Seipiharju, Sáárá. 2021. *Maht suomâkielâ interferens vaaigut feeilâi šoddâmân sämikielâst? Ceelhâopâliih feeilah 6–9 luoka uáppei čalluin*. Oulun yliopisto, Giellagas-instituutti. (Bachelor's thesis).

— 2022. *Anarâškielâ uáppei ceelhâopâlij interferensfeeilâi analyys: Maht feeilah šaddeh já maht taid kannat tivvoođ?* Oulun yliopisto, Giellagas-instituutti. (Master's thesis).

Seržant, Ilja A. 2016. The nominative case in Baltic in a typological perspective. In Holvoet, Axel & Nau, Nicole (eds.), *Valency, argument realization and grammatical relations in Baltic, vol. 3*, 137–198. John Benjamins. https://doi.org/10.1075/vargreb.3

Siewierska, Anna. 2013. Passive constructions. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology.
(https://wals.info/chapter/107). (Accessed 2023-04-19).

Sikiö, Sanni. 2008. *Omistuskonstruktiot ja omistusverbit S2-oppijoiden kirjoitelmissa kolmella taitotasolla*. Tampereen yliopisto, kieli-, käännös- ja kirjallisuustieteiden yksikkö, suomen kieli. (Master's thesis).

Tillinger, Gábor. 2014. *Samiska ord för ord: At mäta lexikalt avstånd mellan språk*. (Acta Universitatis Upsaliensis, Studia Uralica Upsaliensia 39). Uppsala universitet.

Toivonen, Ida. 2007. Verbal agreement in Inari Saami. In Toivonen, Ida & Nelson, Diane (eds.), *Saami linguistics*, 227–258. (Current Issues in Linguistic Theory 288). John Benjamins. https://doi.org/10.1075/cilt.288.09toi

Vainikka, Anne & Brattico, Pauli. 2011. The Finnish accusative. In Brattico, Pauli & Vainikka, Anne (eds.), *Biolinguistica Fennica working papers, vol. 2*, 33–58.

Valtonen, Taarna. 2017. Sinirinnan kielellä: Nils-Aslak Valkeapään runouden suhde puhuttuun saamen kieleen. In Valtonen, Taarna & Valkeapää, Leena (eds.), *Minä*

soin – Mun čuojan: Kirjoituksia Nils-Aslak Valkeapään elämäntyöstä, 201–230. Lapland University Press.

Valtonen, Taarna & Ylikoski, Jussi & Aikio, Ante. 2022. Aanaar (Inari) Saami. In Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena (eds.), *The Oxford guide to the Uralic languages*, 178–195. Oxford University Press. https://doi.org/10.1093/oso/9780198767664.003

Ylikoski, Jussi. 2022a. Legenda jo lausuttaessaan: Havaintoja *tehtäessään*-tyyppisistä verbimuodoista. *AFinLA-teema* 14. 175–194. https://doi.org/10.30660/afinla.111244

— 2022b. Lule Saami. In Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena (eds.), *The Oxford guide to the Uralic languages*, 130–146. Oxford University Press. https://doi.org/10.1093/oso/9780198767664.003.0009

Contact information:

Jukka Mettovaara
Giellagas Institute
P. O. Box 1000
FI-90014 University of Oulu
FINLAND
e-mail: jukka(dot)mettovaara(at)oulu(dot)fi

Finnic comparative constructions in an areal context[1]

Miina Norvik
University of Tartu

Eva Saar
University of Tartu

Abstract

This paper discusses comparative constructions in the Finnic languages. The main focus is on the southern Finnic languages with separate attention paid to dialect variation. By including a maximally complete micro-areal dataset, the article reviews already identified patterns and their spread, while also introducing some (later) developments that have received less attention or gone unnoticed. The results are viewed in the context of neighbouring non-cognate contact languages (Latvian, Latgalian, and Russian); some parallels are also drawn with Lithuanian. As appears, due to being at the crossroads of Finnic and Baltic, it is in the southernmost Finnic languages where the genuine separative comparative construction meets various other patterns, including those borrowed from the Baltic languages. In general, the results enable us to shed more light on the outcomes of more ancient as well as more recent contact situations. The linguistic data originate from text collections, language corpora, example sentences in dictionaries and grammar books, and our own field work data.

Keywords: particle comparatives, locational comparatives, morphosyntax, PAT-borrowing, MAT-borrowing, language contacts

1 Introduction

Comparison can be defined "as a mental act by which two objects are assigned a position on a predicative scale" (Stassen 2013; see also Stassen 1985: 24). In the case of different positions, we are dealing with a comparison of inequality, usually conveyed by means of comparative constructions, as in the Standard Estonian examples (1) and (2). Stassen (2013) makes a distinction between Particle, Locational, Exceed, and Conjoined Comparatives, distinguished on

[1] The research has been supported by the Kadri, Nikolai and Gerda Rõuk Legacy Fund.

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

Finnish Journal of Linguistics 36 (2023), 131–176
DOI: 10.61197/fjl.126225

the basis of expressing the standard (for the underlying cognitive schemas, see Heine 1997). European languages mainly contain examples of the Particle and Locational types. The latter includes instances of marking the standard with a locational element such as with the elative case (see example 1), which is generally used to express a separative meaning. The Particle type, in turn, involves a special comparative particle, such as Estonian *kui* (2), English *than* (see translations for 1–2).

(1) Standard Estonian

Koer	*on*	*kassi-st*	*suure-m*
dog	be.3SG	cat-ELA	big-CMPR
comparee		**standard+ standard marker**	**parameter+ parameter marker**[2]

'The dog is bigger than the cat'

(2) Standard Estonian

Koer	*on*	*suure-m*	*kui*	*kass*
dog	be.3SG	big-CMPR	than	cat
comparee		**parameter+ parameter marker**	**standard marker**	**standard**

'The dog is bigger than the cat'

Comparative constructions show several parallels across the languages of Europe. Moreover, the Particle type is listed as a Standard Average European (SAE) trait and could be regarded as a common innovation among Indo-European languages (Haspelmath 2001; Heine & Kuteva 2006). An older strategy was to mark the standard with the ablative case or some other case with a separative meaning (see Haspelmath 1998). Another characteristic of the European languages is the use of a parameter marker (Stassen 2013) such as *-m* in (1–2). In addition to synthetic marking (comparative suffixes), analytic marking (the use of special adverbs, particles) is also found, cf. *jo* in (3) (see more in § 4.1).

[2] In the literature, different terms are used to denote the constitutive elements of a comparative construction. Here we have followed Treis (2018), who also presents an overview of various terms.

(3) Courland Livonian (LELS)

Kaš u'm ***jo*** ***piški***
cat be.3SG PTCL small

'The cat is smaller'

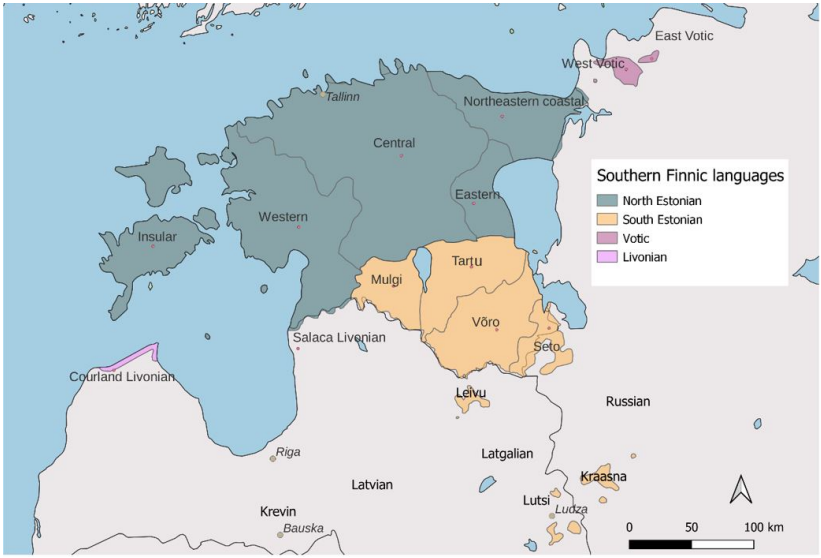
As examples (1–2) illustrate, a language can contain more than one type of comparative construction. This paper analyses the spread of different types and subtypes of comparative constructions with the main focus on instances where two objects expressed with NPs are compared (as in 1–2). Due to the fact that classification of comparative constructions is based on encoding the standard, primary attention is devoted to studying the ways in which the standard is expressed. Only occasionally do we draw some parallels with constructions where two situations (not objects) are compared (see 4). Such examples were only included for background information.

(4) Standard Estonian

Ta on täna ***väsinu-m*** ***kui*** ***eile***
3SG be.3SG today tired-CMPR than yesterday

'Today s/he is (feeling) more tired than yesterday'

The main focus of the current paper lies on the Finnic branch of the Uralic languages, more precisely on the southern Finnic languages (once) spoken to the south of the Gulf of Finland (see Map 1). Traditionally, the languages included in the Southern Finnic branch are Estonian, South Estonian, Livonian, and Votic, while Finnish, Karelian, Ludian, Veps, and Ingrian are grouped into the Northern or Eastern Finnic branch (see e.g., Itkonen 1983: 217–226). In recent times, language historians have offered more elaborate and complex views on the possible diversification and taxonomy of Finnic language varieties (see e.g., Kallio 2014: 162–163; Laakso 2022a: 240–241; Prillop et al. 2020: 24–28). In this paper, we are primarily interested in the spread of patterns in geographically-related areas. Thus, as follows, 'southern' and 'northern' are uncapitalized and used as geographical descriptors. There is a special emphasis on what we refer to here as the *southernmost Finnic varieties*: South Estonian (Mulgi, Tartu, Võro, Seto), South Estonian language island varieties (once) spoken in Latvia (Lutsi, Leivu) and Russia (Kraasna), Livonian (Courland Livonian, Salaca Livonian), Votic (Krevin). With the exception of



Map 1. The southern Finnic languages

the South Estonian Võro and Seto varieties, all of these varieties are critically endangered or extinct.

The current paper has two main objectives:

(i) To provide a detailed study of the comparative constructions in the Finnic languages with a special emphasis on the southernmost Finnic varieties by reviewing already identified patterns but also by looking for additional expressions / matter (MAT)- and pattern (PAT)-borrowings (if any), which may have gone unnoticed. In the case of MAT-borrowings, a phonological form is copied in the recipient language, whereas in the case of PAT-borrowings, certain principles are borrowed, not the form; it is often the case that matter is borrowed together with pattern (Matras & Sakel 2007; Sakel 2007). A more in-depth study is motivated by the rapid growth of the data collected earlier but now made available, and also by the inclusion of relatively recent data. The past decades have seen the emergence of new databases, online dictionaries, grammar books, etc.

(see Table 1 in § 3).

(ii) To offer maximally good areal coverage by viewing the results in a broader areal context, both in the context of the northern Finnic languages (Karelian, Ingrian, Vepsian, and Ludian) as well as non-cognate contact languages (Latvian, Latgalian, Lithuanian, and Russian). The standard varieties of Estonian and Finnish were also included, but only for background information.[3]

In earlier research, Finnic comparative constructions have been studied as part of more general research on Uralic particle comparatives (Raun 1960) and as one of the topics worth discussing from an areal perspective (Koptjevskaja-Tamm & Wälchli 2001), or as the main topic of separate typological research (Stolz 2013). The latter two studies, however, mainly rely on Raun's (1960) work. The grammar books/sketches and other kinds of language/dialect overviews on separate Finnic languages usually comment on comparative constructions, although there has been a tendency to focus on the form of the parameter marker (*-*mpA* > *-*mpi* > -*m*, -*mbi*, -*p*, etc.) rather than on the construction as a whole or its constituent elements or, for instance, the order of those elements.

Considering the aims listed above, we hypothesise that (i) we will be able to identify patterns/constructions that have previously gone unnoticed (at least for some varieties) or show some later developments; (ii) by offering maximally good coverage of the area, we will be able to get a more complete picture of comparative constructions across these varieties.

To collect the data, we turned to original sources: text collections, language corpora, example sentences in dictionaries and grammar books, and our own field work data. Additionally, we included relevant examples mentioned in previous research (if not already present in the dataset). The final dataset contained about 500 linguistic examples, including around 300 examples of constructions with NPs, which are in focus in this paper.

The article is structured as follows. Drawing on previous research, language descriptions, and grammar books, § 2 gives some background on the availability of information on comparative constructions in Finnic and neighbouring languages. § 3 introduces the sources and methods we used to compile

[3] This idea correlates with Johanna Laakso (2022a: 241) who has stated: "Paradoxically enough, although research into Finnic is for practical reasons dominated by Finnish and Estonian, these two languages are genetically heterogeneous, and 'Finnish' and 'Estonian' are not very useful concepts in the historical-comparative study of Finnic."

the dataset. § 4 introduces the main results, which in general show shift from the (more) archaic Separative type towards possibly contact-induced Particle type, especially in southernmost Finnic. In addition to east-west cline we can also observe importance of centre vs. periphery. This is followed by a discussion in § 5 and conclusions in § 6.

2 Previous research on comparative constructions in the Finnic languages and beyond

2.1 General notes

The most comprehensive comparative work on the Uralic languages, including Finnic, is by Raun (1960). He concludes that: "the Finno-Ugric (and Samoyed) languages typically use cases and postpositions for comparison. […] the predominant function of the cases and postpositions used in comparison is separation (disjunctive)", whereas particles are generally a secondary phenomenon (Raun 1960: 232–233). Finnic comparative constructions are also discussed in a study by Stolz (2013) and Koptjevskaja-Tamm & Wälchli (2001), partly drawing on Raun's (1960) work. While Stolz (2013) includes information on various Uralic languages, Koptjevskaja-Tamm & Wälchli (2001) concentrate on the Circum-Baltic languages that include, in addition to non-Uralic languages, the Finnic and Saami languages. The abovementioned studies naturally touch upon different contact situations, but there are also more specific studies on particular areas and their varieties. For instance, Vaba (2011) shows, among other things, that the South Estonian language island varieties Leivu and Lutsi follow the Latvian model and employ a negative particle (translates as 'not than') to express the standard.

For the Finnic languages, comparison tends to be included as a separate section or chapter in grammar books/sketches and other kinds of overviews. The form of the parameter marker (usually a comparative suffix) has tended to gain more attention than the construction as a whole. This holds for South Estonian and Estonian dialects (see e.g., Tanning 1961: 44 for Mulgi, Hagu & Pajusalu 2021: 22 for Seto, or the overview by Viikberg 2020: 153–157, which draws on previous work on Estonian dialects) as well as other Finnic languages (e.g., Laanest 1975: 119; 1986: 124; Virtaranta 1986: 84–85; see the descriptions of Finnic languages also in Bakró-Nagy et al. 2022). The case(s) employed to express the standard are typically also mentioned. For example,

Veps grammars state that the partitive case is used to encode the standard (see Zaiceva 1981: 238–239; Grünthal 2015: 220–223). Valdai Karelian, North Karelian, and Votic sources also mention, in addition to partitive marking, the use of a particle as a standard marker (see Ariste 1948: 58–59; Palmeos 1962: 43; Zaikov 2013: 115; Markus & Rozhanskiy 2017: 610–611, respectively). In Standard Estonian and Estonian dialects, the elative and particle marking are the most common (e.g., Erelt 2017; Viikberg 2020: 156); in South Estonian and occasionally in the Insular dialect of North Estonian, partitive marking is also found (see Raun 1960: 158–159). Still, all this information is scattered among different sources written in different languages at different times. There is a need to bring this information together, while also turning to corpora and adding more recent data (whenever possible). Furthermore, there are also gaps. For instance, the order of the parameter and standard is typically not discussed.

2.2 Historical notes on comparative constructions (standard markers)

The *World Lexicon of Grammaticalization* includes the following sources for comparatives: ABLATIVE, EXCEED, HOW? (W-QUESTION), LOCATIVE, PASS, RECIPIENT, RESEMBLE, UP (see Kuteva et al. 2019: 478). In the Finnic languages, the standard is most commonly encoded by means of a separative case or the particle 'than'. These can be roughly associated with the ABLATIVE, which is a separative case, and HOW? (W-QUESTION), respectively (see more below).

Encoding the standard by means of a **separative case** (e.g., partitive,[4] elative) is claimed to be older than the particle construction, and it has remained prevalent (or the only option) in northern Finnic, i.e., Veps and Ludian (geographically in the northeastern corner of the Finnic-speaking area). As already noted, the separative type was once also characteristic of the Indo-European languages (see § 1). It could be argued that once the partitive lost its separative meaning in the Finnic languages, a newer case with a separative meaning (elative) took over (Prillop et al. 2020: 195–196; Norvik et al. 2021: 64–65).

In general, the **Particle type** is said to arise from bi-clausal sentences such as *Peter is cleverer than Paul is* (Koptjevskaja-Tamm & Wälchli 2001: 683). Finnic *ku-* is usually traced back to an interrogative pronoun, more precisely

[4] In Finnic, the partitive case has developed from an earlier ablative case, which used to identify the 'source' (see Grünthal 2023). Later on, the partitive developed into an object case, although traces of its earlier separative function have been preserved in some constructions in all Finnic languages (see more in Prillop et al. 2020: 192–193).

to an instructive form (*)*kuin* expressing 'how (manner, means)' (FED). It is possible that the spread of *ku-* as a standard marker is an instance of PAT-borrowing reflecting long-lasting contacts with Germanic (for an overview of contacts between Germanic and Finnic languages, see Kallio 2012). Primarily German and Swedish varieties spoken in the Finnic language area can be expected to have played a role. German, for instance, uses the particles *als* and *wie* as standard markers; the latter is, however, termed a regional variety (Duden 2009: 372). The use of *als* as a standard marker is older, while the use of *wie* as a standard marker goes back to the 15th century (DWDS: *wie*). Swedish *än* (6) is related to English *than*. *ku-* is also found in the Saami languages (Raun 1960: 212), which have had contacts with Swedish and Norwegian varieties.

(5) German (DWDS: als)

sie ist nicht ält-er ***als du***
she be.3SG NEG old-CMPR than 2SG

'She is not older than you'

(6) Swedish (SAG 1999: 201)

Jag har sett vackra-re blomm-or ***än de där***
1SG AUX see.PST beautiful-CMPR flower-PL than 3PL there

'I have seen more beautiful flowers than the ones there'

3 Data and methods

Table 1 gives an overview of the language varieties included in the study and the main sources used for collecting the data. Standard Estonian and Standard Finnish are not included in the table since they were considered for background information on the basis of grammar books and previous research (e.g., EKG II: 132–133; Erelt 1986; 2017; VISK). Comparative data on Latvian, Latgalian, Lithuanian, and Russian also mainly originated from grammar books as well as from dialect overviews (e.g., Rudzīte 1964 for Latvian dialects, Cibuļš & Leikuma 2003 for Latgalian).

As Table 1 illustrates, the linguistic data were obtained from various online sources (corpora and dictionaries) as well as published sources. Depending on

Table 1. Data sources on Finnic varieties (southernmost varieties indicated with grey)

Geogr. grouping	Language	Variety	Corpus, online dictionary	Published sources
Southern Finnic	North Estonian	Eastern	EDC, IMS, EMS	
		Northeastern Coastal	EDC, RKS, EMS	
		Western	EDC, EMS	
		Insular	EDC, KES, EMS, EKI WK, HS	Raun 1960
	Votic	West	VKS	Ariste 1948
		East	VKS; EDC	Raun 1960
	South Estonian	Krevin		Winkler 1997
		Mulgi	MS	Tanning 1961
		Võro	EVS, EMS	VVS; Raun 1960
		Seto	SES, SESS, EKI WK	VVS
		Tartu	EDC, EMS	Raun 1960
		Lutsi		EM IX; Raun 1960; Vaba 1977; 2011
		Leivu		EM IX; Raun 1960; Vaba 2011
		Kraasna		EM IX; Kallas 1903; AES 202; Raun 1960
	Livonian	Courland Livonian	LELS; EDC	Sjögren & Wiedemann 1861; LWB; Viitso 2008
		Salaca Livonian		Sjögren & Wiedemann 1861; Winkler & Pajusalu 2018
Northern Finnic	Ingrian	Soikkola, Heva, Lower Luga, Upper Luga*		Ariste 1960; Laanest 1966; Laanest 1997; Nirvi 1971; Raun 1960
	Karelian	North, South, Livvi	KKVS, VepKar	Palmeos 1962; Zaikov 2013
	Veps	North, Central, South	VepKar	Günthal 2015
	Ludian	North, Central, South	VepKar	Ojansuu et al. 1934; Norvik & Saar 2012

* No relevant examples were found.

the variety and source, data collection somewhat differed. As our focus was on the southernmost Finnic varieties, we consulted a greater array of sources to collect examples on these varieties. Thus, for instance, Mulgi, Seto, and Võro examples were extracted from new comprehensive dictionaries (recently published or currently being compiled). For North Estonian, all the relevant examples, which fulfilled the data collection criteria (see below), were collected from the *Estonian Dialect Corpus* (EDC) and online dictionaries on Estonian dialects (if available, e.g., EVS, IMS, KES, RKS) (for this reason they are also presented in separate rows in Table 1). For the northern Finnic languages (Karelian, Veps, Ludian, Ingrian), compilation of the dataset was more random and consisted of noting relevant examples that represented different constructions but not necessarily aiming for including all of the examples featured in a source.

Our approach also differed due to the varying amount of linguistic material available on the selected varieties. For varieties in focus but for which data were scarce (e.g., the extinct Kraasna, Leivu, Lutsi, Salaca Livonian, and Krevin varieties), we opted for maximally complete coverage.

The utilised data collection methods ranged from corpus search and semi-automated search to manual collection of the data (e.g., in the case of an overview on Kraasna published by Kallas in 1903). For Ludian and Veps, we also drew on our own fieldwork data.

It is important to note that the examples in our dataset originate from somewhat different time periods. The South Estonian and North Estonian dialect data mainly represent the language of the second half of the 20th century. The data on the South Estonian language island varieties also partly come from the turn of the 20th century. For instance, all the relevant data on the Kraasna variety goes back to that period. The earliest data on Livonian originate from the mid-19th century. For Salaca Livonian, the mid-19th century data are the only data available. Namely, Salaca Livonian along with Kraasna, Leivu, and Lutsi are all now extinct (for South Estonian language islands, see Balodis & Pajusalu 2021). Courland Livonian data span over 150 years. The Votic, Ingrian, and Karelian data presented in this study were mainly collected in the mid-20th century. Examples of Ludian are from the first decades of the 20th century. Veps data come from the VepKar corpus, which contains texts in different genres from different time periods. Additional Central Ludian and Veps data were collected by the authors of this paper in 2012, 2014, and 2015.

For compiling the final dataset, we only looked for examples where two objects were being compared and thus the standard was overtly expressed with

an NP. The presence of a standard was relevant since distinguishing between different types of comparative constructions is based on encoding the standard (see § 1). Furthermore, as shown on the basis of Estonian, variation in encoding the standard can be expected especially with NPs (see examples 1–2 in § 1), whereas in other cases (e.g., when two situations are compared, cf. example 3) the Particle type prevails (e.g., see Erelt 2017: 432). Still, to be able to give a broader picture, we decided to occasionally include examples where two situations are compared.

The final dataset contained about 500 linguistic examples, including around 300 examples, where two objects (expressed with NPs) are compared. Depending on the variety, there were 10 to 40 examples, which were coded for the following characteristics:

- The standard NP and the form of the standard marker (a case marker, adposition, or particle).
- Based on the marking of the standard, a distinction was made between the main type and subtypes:
  - **Locational type:** (i) Separative type, (ii) Goal/Locative type;
  - **Particle type:** (i) 'than', (ii) negative marker (+ 'than').
- Parameter and the form of the parameter marker (synthetic or analytic).
- Order of the parameter and standard.

Additionally, we noted other occurrences of comparative forms, such as correlative constructions, which provided information on possible developmental paths. Again, such additional examples do not feature in the general counts.

The language examples are presented with different degrees of phonological accuracy in different sources (in most cases, Finno-Ugric transcription is used). For providing examples in this article, we have simplified the transcription of the examples.

4 Results

This section presents the main results by first briefly considering the parameter marker (§ 4.1), and then concentrating on the Locational type (§ 4.2) and Particle type (§ 4.3), both distinguished on the basis of marking the standard.

While §§ 4.2–4.3 include more specific tables, Appendix A gives a general overview of the distribution of these two types across the studied varieties. In order to shed some light on the possible origins and development of standard markers, § 4.4 briefly discusses correlative comparative constructions. § 4.5 provides an overview of the order of parameter and standard in different construction types.

4.1 Marking the parameter

Finnic comparative constructions typically contain a synthetically expressed parameter marker, i.e., the comparative suffix *-*mpA* (> *-*mpi*), which is present in all Finnic varieties, although in different phonetic shapes, e.g., *suur* 'big' → North Estonian *suure-m*, Seto South Estonian *suurõ-*(*m*)*b* ∼ *suurõ-*(*m*)*p*, Standard Finnish *suure-mpi* 'bigger' (for southernmost Finnic see, e.g., Norvik et al. 2021: 46). In Livonian, however, the comparative suffix is present, but it tends to be used with a rather limited number of adjectives, e.g., *madāli* 'low' → *madāl-imi* 'lower' (Viitso 2008: 330). The comparative or the superlative suffix appears instead in some lexicalised forms such as *vaņīmi* 'elder, senior', *vanāva'nbizt* 'grandparents'; a more common way to form comparatives analytically uses the particle *jo* (Laakso 2022b: 384–385), e.g., *pitkā* 'long' → ***jo*** *pitkā* 'longer'. Although analytic constructions for expressing the parameter dominate in Livonian, they are also known in Estonian. The patterns/possibilities occurring in the dataset are the following:

A. The analytic construction does not contain a comparative suffix at all, as in (7);

(7) Courland Livonian (EDC)

jo ***vanā*** *ku* *minā*
PTCL old than 1SG

'older than me'

B. The historical parameter marker is encoded on the adjective, e.g., Estonian *suure-**m*** [big-CMPR] (see examples in § 4.2.1);

C. The historical parameter marker is encoded on the adverb, e.g., Estonian *rohke**m*** 'more', *ena**m*** 'more', *vähe**m*** 'less', which then combines with an adjective in the positive degree (e.g., 8);

(8) Standard Estonian (EKSS: *vähem*)

Sa pole teiste-st ***vähe-m*** ***tark***
2SG NEG other.PL-ELA little-CMPR smart

'you are not less smart than others'

D. The historical parameter marker is marked on the adverb (see above), but it additionally combines with the comparative form of an adjective (e.g., 9). Thus, in such constructions, the comparative is double-marked.

(9) North Estonian (Western; EDC)

seal ol-i üks ***rohke-m vane-m*** *kui ma*
there be-PST.3SG one lot-CMPR old-CMPR than 1SG

'there was one [woman] more older than me'

The patterns C and D are characteristic of Standard Estonian and (mainly North) Estonian dialects. By comparison, in Swedish, unlike in Estonian, the choice between a synthetic or analytic formation in some cases depends on word structure, and double marking does not seem possible, at least not in the standard language (see SAG 1999: 196, 200).

The North Estonian Insular dialect (Hiiu) has occurrences where both the parameter and the standard can be marked with the partitive case (10). In riddles and proverbs this pattern is also known in other parts of Estonia (Raun 1960: 158–159). One example containing a parameter in the partitive case also occurred in Eastern Votic (11).

(10) North Estonian (Insular, Hiiu; Raun 1960: 159)

parama-t te-ɒa *mā olɛ kül̄*
better-PRT 3SG-PRT 1SG be.1SG indeed

'sure (indeed) I am better than he'

(11) Eastern Votic (VKS)

mees ***varmõ-pi-t*** ***tšem*** ***naizikko***
man strong-CMPR-PRT than woman

'a man is stronger than a woman'

4.2 Locational type

The Finnic languages contain two kinds of Locational types, referred to here as the Separative type and the Goal/Locative type. Using Heine's terminology, the respective cognitive schemas at play are the Source Schema, Goal Schema, and Location Schema (see Heine 1997: 111–112).

4.2.1 Separative type (case marking)

The Separative type primarily includes instances where the standard is marked with the elative (12) or partitive (13). Still, while the elative case is commonly used to mark separation/source, the partitive case is associated with the separative function only historically (Lehtinen 2007: 78; Prillop et al. 2020: 192). An original separative function is also attributed to the Russian genitive used in comparative constructions (see Koptjevskaja-Tamm & Wälchli 2001: 683–684).

(12) Võro (EVS)

üts´ kleit´ ***tõsõ-st*** ***uhkõ-mb***
one dress other-ELA gorgeous-CMPR

'One dress [is] more gorgeous than the other'

(13) Western Votic (VKS)

muna on ***viisaa-pi*** ***kannaa***
egg be.3SG smart-CMPR hen.PRT

'The egg is smarter than the hen'

Table 2 provides an overall picture of cases used to encode the standard in the studied varieties. Here and elsewhere, the symbol + is used to mark the presence of a particular method; if it is necessary to identify the presence of a method in a specific variety, an abbreviation in parentheses is used instead (see Ins – insular, OWE – Old Written Estonian). Colour coding is used to illustrate whether a specific method is common/general (dark grey) or restricted/infrequent (light grey).

As can be seen, there is a cline from east to west with partitive marking best preserved in the northern Finnic varieties (geographically located in the

Table 2. Cases used to encode the standard

Case	CLiv	SLiv	StEst	NEst	Mul, Trt	Vro	Set, Kra, Lut, Lei	WVot SIng, LIng, StFin	EVot, HIng, Lud, Kar, Vps
Prt	+	+	(OWE)	(Ins)	+	+	+	+	+
Ela	+		+	+	+	+	+	+	
Gen								+	

east; see example 14) and the least in Livonian, Standard Estonian, and North Estonian varieties (i.e., in southern Finnic languages located geographically in the west). For some varieties, however, it is not always possible to determine, which type prevails; this is the situation for Võro, which has been under stronger influence from Standard Estonian. In general, in languages that show more than one way of marking the standard (e.g., 15), one type can be expected to be primary. For instance, in South Estonian varieties (Seto, Kraasna, Lutsi, Leivu), the partitive construction is prevalent. Western Votic, Soikkola Ingrian, Lower Luga Ingrian, and Standard Finnish show a similar picture, fitting geographically with a continuum from east to west, and at the same time cross-cutting Ingrian and Votic.

(14) North Ludian (Mundjärvi; Ojansuu et al. 1934: 13)

kah-t ***kogo-d*** ***tobd'e-mb***
two-PRT haystack-PRT big-CMPR

'bigger than two haystacks'

(15) Seto (SES)

Vello *olle* ***mui-st*** ***madala-b*** ∼ ***madala-b*** ***mui-d***
brother be.PST.3SG other.PL-ELA low-CMPR ∼ low-CMPR other.PL-PRT

'the brother was shorter than the others'

Table 2 also shows that – of the synthetic constructions (case marking) – elative marking has become the only option in North Estonian, including in Standard Estonian. In the latter, the partitive forms are obsolete – they can be found in

Old Written Estonian (e.g., 16) and to some extent also in 20th century texts. Nowadays, however, the partitive is not mentioned as an option for encoding the standard (see, e.g., Erelt 2017). The North Estonian dialect data contained occurrences mainly from the periphery such as from Hiiumaa Island (17). The South Estonian varieties bordering with North Estonian – Mulgi and Tartu – have been found to contain some remnants of partitive marking primarily in folklore records (e.g., riddles, proverbs, see Raun 1960: 158–159). The retreat of the partitive could be expected, as the standard language is largely based on North Estonian and Mulgi and Tartu are the South Estonian varieties, which have been most affected by the written language.

(16) Old Written Estonian (COWE[1739])

wanne-m min-d
old-CMPR 1SG-PRT

'older than me'

(17) North Estonian (Insular, Hiiu; HS)

sɛɛdus on ***vana-ma-d mei-d***
law be.3SG old-CMPR-PRT 1PL-PRT

'the law is older than us'

Unlike in the rest of the studied varieties, neither type of marking turned out to be preferred in the Livonian varieties (shown in light grey in Table 2). The Salaca Livonian dataset, representing the language use of the mid-19th century, contained no examples of elative marking; there was only one example of the partitive case used with a pronoun (18). Even the Courland Livonian data, which spans over 150 years, contained only a handful of examples with the elative and partitive (mainly from the 19th century). As noted by Sjögren & Wiedemann (1861: 128), the elative was hardly ever heard already in the mid-19th century.

(18) Salaca Livonian (Winkler & Pajusalu 2018)

Ana minne-l ***piški-m sä-da***
give.IMP.2SG 1SG-ALL;ADE small-CMPR DEM-PRT

'Give me less than this'

It is important to note that in the varieties where the use of the partitive for marking the standard has receded, it seems to have been preserved the longest with pronouns. Thus, the only example with the partitive in Salaca Livonian contains a demonstrative pronoun (18) and the examples from the Estonian Insular dialect (e.g., 17) also contained a pronoun. The Estonian dialect data also included examples of the parameter appearing in the partitive case, as in (17).

Although genitive marking of the standard has its own semantic and pragmatic reasons, Finnish still stands out in its use of the genitive for standard marking (19). This phenomenon is restricted to the demonstrative pronouns *se*, *tuo*, *tämä* (VISK § 639). Not surprisingly, we found similar examples from Lower Luga Ingrian spoken in close proximity (see 20).

(19) Standard Finnish (VISK § 639)

Onneksi matka ei ol-lut ***tämä-n pite-mpi***
luckily journey NEG.3SG be-ACT.PST.PTCP DEM-GEN LONG-CMPR

'luckily the journey was no longer than that'

(20) Lower Luga Ingrian (Nirvi 1971: 514)

se ei mää ***se-n hape-mma-ks***
DEM NEG.3SG go.CNG DEM-GEN sour-CMPR-TRA

'this [sauerkraut] doesn't get any sourer than it is'

We are aware that Stolz (2013: 102) did not classify the genitive as a separative case, but since there are examples where the genitive appears in comparative correlative constructions (e.g., 21), we have nevertheless included it in Table 2. The genitive is used in correlative constructions in northern Finnic languages such as Ingrian, Finnish, and Karelian (see also § 4.4). However, at least for Finnish, it has been noted that the more common construction is formed with partitive – partitive, i.e., *mitä – sitä*, instead of with the genitive *sen – mitä* (VISK § 639).

(21) Soikkola Ingrian (Nirvi 1971: 309)

mi-n enemв *palkkaa,* ***se-m paremв***
what-GEN more salary.PRT DEM-GEN better

'the more salary the better'

Table 2 suggests that the Separative type is attested in all of the studied varieties, although it is not necessarily the most common way of expressing comparison of inequality. For instance, Courland Livonian and Salaca Livonian both contained only a few examples, and a more common way to introduce the standard was to use a particle construction (see more in § 4.3).

4.2.2 Goal/Locative type (adpositional marking)

Koptjevskaja-Tamm & Wälchli (2001: 684) present the Goal/Locative type primarily on the basis of the Baltic languages, mainly drawing on prepositions with the original meaning 'behind' and 'above, about' < 'before'; the latter instances are included under "over/for". Here, we have used 'over, about' to capture these meanings. Stolz (2013: 102), in turn, includes Livonian *yl* (*ül*) ~ *i'ļ* 'over' along with the non-standard Estonian *vasta* 'against' under the subtype "other". As already noted by Koptjevskaja-Tamm & Wälchli (2001), it is not always easy to make a distinction between the goal and locative functions, thus, we also have used the term Goal/Locative type. Table 3 includes information on the southernmost Finnic varieties as well as neighbouring non-cognate contact varieties where instances of these types could be found.

Typologically, grammatical markers conveying spatial meanings are regarded as expected sources for the markers of the standard (Kuteva et al. 2019: 268). However, while UP is given separate attention (ibid., 446–448) (it could be roughly associated with 'over, above' and 'on'), BEHIND and AGAINST are not mentioned as separate sources. In general, both the Goal and Locative type seem to be marginal in Europe (Stolz 2013: 32, 319, 322).

The prepositional construction containing 'over, above' could be attested in both Livonian varieties. Still, its occurrence appeared to be more prevalent in the Salaca Livonian dataset – 5 out of 13 examples were instances of the Locational type, and among these, 4 examples were formed with *yl* (e.g., 22) and one with partitive (see § 4.2.1, ex. 18). It is likely that the 'over, above' constructions in Livonian (e.g., 22–23) are instances of PAT-borrowing from Latvian. As regards the collection of the Salaca Livonian data, it has been noted that about 60% of the linguistic data were obtained by asking for translations of Latvian sentences (see Winkler & Pajusalu 2018: 155). Since the respective examples containing a comparative construction are provided with a Latvian counterpart all including *par*, they seem to be instances of translations. Although Sjögren & Wiedemann (1861) have noted down some examples of Courland Livonian *i'ļ*, such pattern does not seem to have spread

Table 3. Goal/Locative type in southernmost Finnic in an areal context

	Lith	Lav	Ltg	CLiv	SLiv	Set, Vro, Lei, Lut	Rus
'on'		*uz*					
'over, above'		*par*	*par*	*i'ļ*	*yl*		
'against'	*prieš*	*pret*				*vasta*	
'behind'	*už*	*aiz*					*za*

in the course of time (indicated in light grey). Latvian *par* was also borrowed into Latgalian to fulfil the same function (Nau 2011: 72). The fact that it was not inherent to Latgalian might also explain why neither Lutsi nor Leivu show traces of *üle* used in comparative constructions (at least we were not able to find any examples in the dataset). Otherwise, Leivu and Lutsi *üle* show functions and meanings concurrent with Latvian *par*, but comparative constructions are not mentioned (see Vaba 2011: 224–225).

(22) Salaca Livonian (Winkler & Pajusalu 2018: 88–89)

Läeli-m ***yl*** ***kaks*** ***birkau***
heavy-CMPR over two ship_pound.GEN

'heavier than two ship-pounds'

(23) Courland Livonian (LWB 308)

ta mõtlõ-b, ku ta vȯl-ks ***i'ļ*** ***munt*** *pōrak, bet*
3SG think-3SG that 3SG be-COND.3SG over other.GEN important but
äb ūo
NEG.3SG be.CNG

's/he thinks that s/he is more important (beautiful) than the others, but s/he is not'

The prepositional constructions containing 'against' occurred only in the southernmost South Estonian varieties: Seto (24), Võro, Lutsi (25), and Leivu (see Table 3). The latter two varieties are now extinct. In Seto, the 'against'-construction is not predominant; though it may be old, it is still common (L1

speaker Maeve Leivo, p. c.). Relevant examples of Võro are collected from the close vicinity of Seto. In Western Võro, such constructions are not known (Mariko Faster, p. c.). In North Estonian (including in Standard Estonian), comparative constructions containing *vastu* (26) can also be found, but they are primarily associated with negative constructions. As explained in EKSS, *vastu* is used to express that someone or something is not comparable with/equal to someone or something. It is important to note that *vasta*/*vastu* can be attested with a wide array of meanings (as is the case of *üle* described above); in addition to its use in comparative constructions, dictionaries (EKSS, SES) also include meanings such as 'towards', 'opposite of', 'in front of', 'before', etc.

(24) Seto (SES)

kuld om ***vasta*** ***hõpõ-t*** *õks viil pallo* ***kalli-p***
gold be.3SG against silver-PRT still more much expensive-CMPR

'gold is still much more expensive than silver'

(25) Lutsi (Raun 1960: 211)

pallo om sjoo-h külä-h ***vasta*** ***minnu*** ***vanõ-mb-it***
many be.3SG DEM-INE village-INE against 1SG.PRT old-CMPR-PL.PRT

'there are many in this village who are older than me'

(26) Standard Estonian (EKSS)

Ei ***rukkileiva*** ***vastu*** *pole teis-t* ***vägeva-ma-t***
NEG rye_bread.GEN against NEG other-PRT mighty-CMPR-PRT

'There is nothing mightier than rye bread'

Again, parallels can be drawn with the Baltic languages. In Lithuanian, the corresponding construction consists of *prieš* 'against' + accusative, which "occasionally denotes an object with which the subject is compared" (see example 27; Ambrazas 1997: 583). The Latvian cognate is *pret*, in which case a similar pattern could be possible, but it seems to be less established; (28) is one of the few examples (Eglė Žilinskaitė-Šinkūnienė, p. c.).

(27) Lithuanian (Ambrazas 1997: 583)

Sūnus ***prieš*** ***tevą*** *negražus*
son.NOM.SG against father.ACC.SG not.handsome.NOM.SG.M

'The son is not handsome compared to (lit. 'against') the father'

(28) a. Latvian (LiLa)

Viņš saprata, ka atrodas nelādzīgā
3SG.M realise.PST.3 that find.PRS.3 unfavourable.LOC.SG
pozīcijā ***pret*** ***stāvētāju***
position.LOC.SG against stander.ACC

b. Lithuanian (LiLa)

Jis suprato, kad ***prieš*** ***stovintįjį*** *yra*
3SG.M realise.PST.3 that against stander.ACC.SG be.3
nepatogioje pozicijoje, atrodė
unfavourable.LOC.SG position.LOC.SG find

'He realised that he was in an unfavourable position compared to (lit. 'against') the stander'

The 'behind' (see 29) and 'on'-subtypes, in turn, do not seem to be present in the Finnic languages. They exist in the Baltic languages, though in Latvian only marginally (see Endzelin 1923: 353–354 for Latvian *aiz* and *uz*). In the neighbouring Slavic languages (e.g., see SRNG for Pskov Russian) 'behind' is most probably a Baltic substrate (Čekmonas 2001: 120; Koptjevskaja-Tamm & Wälchli 2001: 684).

(29) Russian (DRJa 1985: 179)

Mlaže ***za*** ***menja***
young.CMPR behind 1SG.ACC

'younger than me'

In general, it seems that the Goal/Locative constructions are restricted to the southernmost Finnic and Baltic contact area; they are relatively infrequent and seem to be contact-induced. So far, we have not been able to find the

corresponding patterns in the northern Finnic languages, but further research is needed.

4.3 Particle type

Below, § 4.3.1 gives an overview of the particle making use of the pronoun *ku-* and § 4.3.2 takes a closer look at instances where this particle is accompanied by a negative marker or the negative marker is the sole marker of the standard. § 4.3.3 discusses instances, which include the pronoun *mi-* or a corresponding Russian loan. The Particle type is present in all Finnic languages, except in Kraasna (see Appendix A). However, since the marker *ku-* could be attested in the language but the data on Kraasna are the scarcest, it is possible that simply no instances were recorded.

4.3.1 Particle *ku-* 'than'

The majority of the studied varieties make use of the standard marker *ku-*, which originates from an interrogative pronoun (FED; see also § 2). Such comparative constructions have spread in all southern Finnic languages and there are also examples found in Ingrian (e.g., 30), Karelian, and Finnish (31; VISK § 635–637). We were not able to find any examples of *ku-* used in Veps or Ludian as a standard marker. The dataset only contained instances of another pronoun (*mi-*) and a MAT-borrowing from Russian, suggesting that such particle constructions are of later origin (see more in § 4.3.3).

(30) Soikkola Ingrian (Nirvi 1971: 674)

muna ei jo̮ konsā veD ***vīzahe-mp kuiŋ kana***
egg NEG.3SG be.CNG never PTCL smart-CMPR than hen

'An egg is never smarter than a hen'

(31) Standard Finnish (VISK § 635)

Toinen kirja-si on ***parempi kuin ensimmäinen***
second book-POSS.2SG be.3SG better than first

'Your second book is better than the first'

Table 4. Comparative conjunctions using a negative marker

	CLiv	SLiv	Krev	Lei	Lut	Lat	Ltg	Lith	Rus
NEG + 'than'	***äb*** *ku* ∼ ***äb****ku*	***ap*** *ka*		***ei****'ku*, ***ei****kku*, ***e*** *ku*, ***is****ku*	***ei****gu*, ***üs****kui*	***ne****kā*	***na****kai* ∼ ***ni****kai*	***ne****gu* *kaip*, ***ne****kaip*, ***ne****gu*	***ne****želi*
NEG			***ep***	***is***	***ei***	***ne***	***na***	***nei***, ***ne***	

In Standard Estonian and Standard Finnish, the particle *ku-* tends to be interchangeable with the case-marked (partitive, elative) standard when the standard is an NP. However, for a wider range of constructions when comparison is intended (e.g., infinitival constructions), the particle *ku-* is the only option (see for Estonian Erelt 2017: 432, for Finnish VISK § 637). As already noted, it is likely that the Separative type represents an older type and the Particle type emerged later. The stage of transition is thought to have enabled the use of the (original) partitive together with the particle *ku-* in North Estonian (Raun 1960: 159).

4.3.2 (*ku-* +) a negative marker

The use of a particle in combination with a negative marker is a typologically infrequent pattern (see Stassen 1985: 217; Stolz 2013: 73). In the Circum-Baltic Area, the standard marker includes a negative element in the Baltic languages, Polish, and to some extent also Russian (see Koptjevskaja-Tamm & Wälchli 2001: 683). Latvian can be seen as the source for the occurrence of the corresponding markers in Leivu, Lutsi (Vaba 1977: 20, 24), and Livonian. Table 4 gives an overview of the forms occurring in the dataset and in previous literature.

According to Table 4, all Finnic varieties (once) spoken in present-day Latvia make use of a negative marker, either (i) in connection with 'than' (in the Livonian varieties; as in 32), (ii) as the sole standard marker (Krevin; e.g., 33), or (iii) both (Lutsi, Leivu). Still, since Krevin data were scarce, no far-reaching conclusions can be made about the presence/absence of a negative marker used with 'than'.

(32) Courland Livonian (EDC)

Jo kovāl ***äb*** ***ku*** *tämā*
PTCL smart NEG than 3SG

'smarter than him/her'

(33) Krevin (Winkler 1997: 390)

se õikõ käsi väkevä-mp ***ep*** *se kura*
DEM right hand strong-CMPR NEG DEM left

'The right hand is stronger than the left hand'

Table 4 suggests that the Finnic varieties primarily contain examples of PAT-borrowing since the pattern (combining a negative marker with a particle) is taken rather than the matter – the negative marker is of native origin. Furthermore, Lutsi and Leivu examples indicate that both present and past tense markers can be used to introduce the standard: *e*(*i*) refers to present and *is*/*üs* to past. However, a closer look at the examples shows that their use is not dependent on temporal reference but that instead they are used interchangeably (see also Norvik et al. 2021). The Livonian dataset, in turn, only contained instances of the present negative marker used to mark the standard (e.g., 32), although Livonian varieties also make a distinction between present and past negation much as in Lutsi and Leivu.

The use of a sole negative marker seems to represent an older pattern since it is present in now extinct Finnic varieties (Krevin, Leivu, Lutsi), Latgalian *na* is listed as infrequent in Modern Latgalian (Nau 2011: 72), and Latvian *ne* is not mentioned at all in modern grammars (cf. Endzelin 1923: 353–254, where *ne* is listed as one option although *nekā* and *par* are considered the most common).

There are some rare examples of comparative conjunctions including a negative marker also in the northern Finnic languages. Still, these examples are instances of comparison, where two situations are compared, not two NPs (see 34–35; cf. § 1). As argued by Raun (1960: 219), Karelian examples follow the Russian pattern. It is possible that these examples reflect the use of Russian *neželi*, which is termed to be old-fashioned (see Timberlake 2004: 215) and occurs instead in the literary language (36; Tolkovye).

(34) South Karelian (Raun 1960: 221)

nagraja-n šuu-h parembi kaččuo ***ei*** ***kuin*** *itkijän*
laugher-GEN mouth-ILL better look.INF NEG than weeper.GEN

'it is better to look into the mouth of a laughing person than of a weeping person'

(35) North Karelian (KKVS)

poro-lla paremmim peäše-t ***ei*** ***kun*** *hepose-lla*
reindeer-ADE better get_by-2SG NEG than horse-ADE

'one gets by better with an reindeer than with a horse'

(36) Russian (Tolkovye [Stasova, Vospominanija])

Eto byl orator sovsem drugogo tipa, ***neželi*** *Lenin*
DEM be.PST.M orator totally other.GEN type.GEN than Lenin

'It was a totally different kind of orator than Lenin'

4.3.3 Particle 'than' making use of the pronoun *mi-*

The dataset also included examples of the native pronoun *mi-* 'what' used as a standard marker (e.g., 37a). There were also instances of bi-clausal sentences, which serve as further evidence (see 38; cf. § 2.2). These examples can be analysed as instances of PAT-borrowing from Russian, since the interrogative or demonstrative pronoun *čem* 'with what (instrumental)' used in the comparative constructions originates from *čto* 'what' (Koptjevskaja-Tamm & Wälchli 2001: 683). Raun (1960: 222) provides examples from Kildin Saami and Finnic, e.g., Võro, Eastern Votic (Ičäpäivä), Lower Luga Ingrian (Kallivere), Tver Karelian (Bježeck). For its wider spread in the Kola Peninsula and examples in Skolt Saami, see Itkonen (2011: 247–248).

(37) a. Veps (VepKar)

spičk om nügüd' nece pide-mb, ***mi*** *nece d'orš*
match be.3SG nowadays DEM long-CMPR what DEM ruffe

b. Russian

sejčas spički dela-jut dlinn-ee ***čem*** *etot erš*
now match.PL make-3PL long-CMPR than DEM.M ruffe

'a match is nowadays longer than this ruffe (Gymnocephalus cernua)'

(38) Lower Luga Ingrian (Nirvi 1971: 579)

hän tekehüis paremma-ks ***mi-tä*** *on*
3SG pretend.3SG better-TRA what-PRT be.3SG

'he pretends to be better than he is'

In addition to PAT-borrowings, there are also MAT-borrowings. In previous literature, MAT-borrowings (accompanying a PAT-borrowing) have been attested in Mordvin and further to the east (see Raun 1960: 230; Stolz 2013: 102–103). We were also able to find MAT-borrowings in the Finnic languages, e.g., (39) represents the extinct Eastern Votic dialect. Palmeos (1962) lists one Valdai Karelian example (40) and we can contribute an example from our recent fieldwork data (41). Examples (40–41) are, however, instances where two situations are compared. Thus, it seems possible that such structures are especially open to foreign influences, both PAT- and MAT-borrowing (see also 28 above).

(39 = 11) Eastern Votic (VKS)

mees varmõ-pit ***tšem*** *naizikko*
man strong-CMPR.PRT than woman

'a man is stronger than a woman'

(40) South Karelian (Valdai; Palmeos 1962: 43)

enämbi miu-šša razvo-a ***čim*** *šiu-šša*
more 1SG-INE fat-PRT than 2SG-INE

'I have more fat than you (I'm more fat than you)'

(41) Central Ludian (Norvik & Saar 2012)

kala-d ljubi-n enämbä ***čem*** *liha-d*
fish-PRT like-1SG more than meat-PRT

'I like fish more than meat'

4.4 Correlative comparative constructions

Correlative comparative constructions like *the sooner the better* (also referred to as *conditional correlatives*, *propositional correlatives*; see more in Dikken 2005) shed light on the development of parameter markers as well as standard markers. For instance, the Latvian correlative construction *jo – jo* is regarded as the source for the Livonian **parameter marker** *jo*; however, in Standard Latvian, *jo* is always followed by a comparative form (Stolz 2013: 107). In Livonian, however, it tends to combine with a parameter in the positive degree (see 42), including in the case of correlative constructions (43). Use of *jo* + positive degree is claimed to be a characteristic of the Latvian varieties spoken in western Courland (Endzelin 1923: 352; Rudzīte 1964: 221–222). There is also a form with a similar usage in Lithuanian, primarily in Samogitian (spoken in the western part of Lithuania) – *juo*, which is considered an instrumental form of the pronoun *io-* (Endzelin 1923: 353). Thus, this suggests an areal continuum from northern Courland via western Courland to Samogitia. One of the Lutsi examples (44) expressing a superlative degree is composed of *kõige* + *jo* followed by a comparative form.

(42) Courland Livonian

jo *sūr*
PTCL big

'bigger'

(43) Courland Livonian (EDC)

jo *lei'žgili su'g,* ***jo*** *sūr-d mūm-õd*
PTCL close relative PTCL big-PL wedding.present-PL

'the closer the relative, the bigger the wedding-presents'

(44) Lutsi (EM IX: 134)

kõige ***jo*** *nūre-mb* *sõzar*
all.GEN PTCL young-CMPR sister

'the youngest sister (of all)'

The role of correlative constructions in the rise of **standard markers** can be observed in Veps, cf. the use of Russian *čem* and Veps *mi* in correlative constructions (45a–b) and as comparative particles (46a–b) (for *mi* as a PAT-borrowing from Russian, see § 4.2.3). While the Standard Estonian correlative comparative construction (45c) makes use of *mida – seda* (comparable to Veps *mi – se*), unlike Veps and Russian, Estonian constructs the comparative construction in (46c) using *kui* (not *mis*).

(45) a. Veps (VepKar)

Mi *enamb* *rahvas-t,* ***se*** *pahe-mba* *kulu-b*
what more people-PRT DEM worse-CMPR sound-3SG
tapand-se
threshing-DEF

b. Russian

Čem *bol'še* *ljudej* *molotit,* ***tem*** *sil'n-ee* *slyšna*
REL more people.ACC thresh.3SG COREL stong-CMPR hear
molot'ba
threshing

c. Standard Estonian

Mi-da *enam* *rahvas-t,* ***se-da*** *hullemini* *on*
what-PRT more people-PRT DEM-PRT worse be.3SG
rehepeksu *kuul-da*
threshing.PRT hear-INF

'The more people, the louder (worse) the threshing sound'

(46 = 37) a. Veps (VepKar)

spičk om nügüd' nece pide-mb, ***mi*** *nece d'orš*
match be.3SG nowadays DEM long-CMPR what DEM ruffe

b. Russian

sejčas spički delajut dlinn-ee ***čem*** *etot erš*
now match.PL make.3PL long-CMPR than DEM.M ruffe

c. Standard Estonian

tikk on nüüd pike-m ***kui*** *see kiisk*
match be.3SG nowadays long-CMPR than DEM ruffe

'a match is nowadays longer than this ruffe (*Gymnocephalus cernua*)'

The fact that, according to Raun (1960), *mi-* is considered to be marginal and the use of *čem* ∼ *čim* as a standard marker is not mentioned for Finnic languages point to a relatively recent use/emergence of the pronoun *mi* and the Russian loan *čem* ∼ *čim* in comparative constructions.

Correlative comparative constructions may also show the **preservation** of older forms. In (Standard) Estonian, for instance, only the correlative constructions contain (have preserved) the partitive form of pronouns referring to the standard. In Salaca Livonian *too*, *mida – seda* is claimed to have been older and only later replaced by *jo – jo* (Sjögren & Wiedemann 1861: 127). In addition to the partitive, which is found with correlative constructions both in the southern as well as northern Finnic varieties (see, e.g., 45c and 47), in some northern Finnic languages (Ingrian, Finnish, and Karelian), a pronoun can even appear in its genitive form (48).

(47) North Ludian (Virtaranta 1986: 85)

mi-d_*enamban magada-d,* ***ši-da*** *keuhe-mba-kš l'iene-d*
what-PRT_more sleep-2SG, DEM-PRT poor-CMPR-TRA will.be-2SG

'the more you sleep, the poorer you become'

(48) Livvi Karelian (KKVS)

mi-n *bohate-mbi,* ***se-n*** *julgie-mbi*
what-GEN rich-CMPR, DEM-GEN brave-CMPR

'the richer, the braver'

4.5 The order of the standard and parameter

Comparative constructions show strong correlations with basic word order (Stassen 1985: 53–56). For instance, in Stassen's sample, Location Schema is exclusively used with SVO languages and Source Schema (corresponds to Separative type) is preferred with SOV languages. Over time, Finnic languages have shifted to being SVO languages (although, broadly speaking, word order is still quite free / still shows variability, especially in spoken language). However, in most of these languages, the Separative type is common or is even the predominant form (as is the case with the Finnic languages located geographically in the east), which also reflects their history.

The expected order for SOV languages is standard + parameter, as is common in the Uralic languages further in the east (e. g., Stolz 2013). In the Finnic dataset, for the Separative type, the order of the standard and parameter tends to depend on the encoding of the standard marker of the comparative construction. The order standard + parameter was preferred with elative marking (see 49), which was noted as being a newer way of encoding the standard (see § 2.2). For partitive marking, the preferred order was parameter + standard (see 50). The order standard + parameter also occurred in Veps, Ludian, and Karelian though less commonly (e. g., see 14 in § 4.2.1). Interestingly, in Kraasna and Leivu, the parameter appears to precede the standard regardless of the case used for the standard marking.

(49) Standard Estonian

minu-st *targe-m*
1SG-ELA smart-CMPR

'smarter than me'

(50) Seto

targõ-p *minno*
smart-CMPR 1SG.PRT

'smarter than me'

For the Goal/Locative type and Particle type (incl. NEG + PTCL), the order is (primarily) parameter + standard.

5 Discussion

Previous research on comparative constructions has shown that areal continuity is more important than genealogy (e.g., see Heine 1997). This finds support on the micro-areal level too. For instance, the negative marker occurs as part of the standard marker (or as the sole standard marker) in the Finnic varieties (once) spoken in present-day Latvia: Courland Livonian, Salaca Livonian, Lutsi, Leivu, and Krevin. These instances are almost exclusively cases of PAT-borrowing of Latvian *nekā* (Latgalian *nakai*, *nikai*). To compare, we were not able to find any such instances in the Estonian varieties spoken within the borders of present-day Estonia. This seems to suggest that for a negative marker to be employed, long-lasting contact with a language making use of such means is needed. By comparison, although Karelian varieties show traces of a similar development, the corresponding Russian standard marker (*neželi*) has become obsolete, which might be the reason why the respective pattern did not spread in the neighbouring Finnic varieties.

Still, examples of distributions that do not follow political borders were more common. For instance, the Particle type including *ku-* as a standard marker was the most prevalent in the southern Finnic languages but was also attested in Ingrian, Finnish, North Karelian, and Livvi Karelian-speaking areas (but absent in Ludian and Veps), which are all contiguous with one another. Since the use of the Particle type is regarded as a SAE feature (again, as a PAT-borrowing, see § 1), its presence already indicates adherence to SAE. Unlike in the other studied varieties, in Livonian, the Particle type has overridden the Separative type and has ultimately become almost the only option. Thus, it could be concluded that Courland Livonian has come closest to SAE in this respect (Salaca Livonian became extinct already in the 19th century).

The expression of the standard by employing (what were originally) spatial markers as the main means and the existence of several such markers seems to

point to an older or inherent part of the system. For instance, the Separative type, which is associated with the majority of Uralic languages (Stolz 2013: 103) and can be argued to represent the original system (see § 4.2.1), was present in all of the studied Finnic varieties, except in Krevin (probably due to the scarcity of the data). Further evidence of change in the system is the replacement of an older separative marker (partitive case) with a newer one (elative case) to encode the standard. Still, there were differences between the varieties with respect to whether partitive or elative marking was used or if both were possible. We observe a cline from east to west with a more conservative partitive marking better preserved in the east (see Table 2 in § 4.2.1) and elative marking prevailing in the west. It is possible that several factors are at play. Russian may have had a preserving effect, in which case the synthetic marking of the standard has served as a model. At the same time, the respective languages have not retained the order historically associated with the Separative type, i.e., standard + parameter (see § 4.5); it is possible that Russian has provided the model for parameter + standard (e.g., 51). By comparison, no such model has existed in Livonian, for which Latvian is the main contact language. This could be one of the reasons why the Separative type has receded in Livonian.

(51) a. Veps (VepKar)

Proud om kal'he-mb kuuda-d
truth be.3SG precious-CMPR gold-PRT

b. Russian

Pravda dorože zolota
truth precious.CMPR gold.GEN

'The truth is more precious than gold'

Although in general there is a cline from east to west, the centre-periphery division also plays a role. For North Estonian and South Estonian varieties, it can be argued that a location at the periphery may have helped to preserve partitive encoding of the standard. The partitive is best preserved in the southeastern corner (Seto and Võro), though some rare examples can also be found in the western Estonian islands of Saaremaa and Hiiumaa (in the Hiiumaa examples, even the parameter can be in the partitive). One can see that Standard Estonian (mainly based on the Central dialect of North Estonian) has left its imprint on

the bordering South Estonian varieties of Mulgi and Tartu where only some rudiments of partitive marking could be found.

In addition to separation, the languages in the studied area also contained examples of standard markers having developed from markers originally conveying GOAL and/or LOCATION (represented by 'over, above' and 'against' in Finnic). These markers were, however, restricted to the southernmost Finnic and Baltic/Slavic contact area. The 'over, above' type was a clear instance of PAT-borrowing from Latvian into Livonian, with the majority of examples coming from Salaca Livonian. This could be seen as a result of the data collection method – these examples had a Latvian counterpart containing *par*. At the time, the data were collected (mid-19th century), most speakers were bilingual and Salaca Livonian was on the verge of disappearing. By comparison, the Courland Livonian dataset does not show the spread of the 'over, above' type over time.

When compiling the dataset, we also identified several other spatial adpositions that might serve as possible sources for marking the standard in a comparative construction and could be associated withr the Goal/Locative type. These cases involved comparison of the type X is Y by/at/near Z (Seto *man* and *kottal*); X is next to Z (occurs in North Estonian as well as South Estonian); see (52–53). Seto *man* and *kottal* allow us to draw a slight parallel with Hungarian, where a case marker expressing 'by, at, near' (*-nál/-nél*) is the primary method for encoding the standard in a comparative construction (see also Raun 1960: 192). It must be noted that these cases do not contain a parameter in the comparative form; these examples only illustrate the use of spatial sources for comparison in a general sense.

(52) Seto

Serga timä man om poišokanõ
Serga (prop.) 3SG.GEN by be.3SG boy.DIM

'Serga is a boy compared to him'

(53) Seto (EKI WK; SES)

mis tä mu kõrval om, sitt
what 3SG 1SG.GEN beside be.3SG shit

'what is he compared to me, shit'

MAT-borrowings were clearly less common than PAT-borrowings. MAT-borrowings could be attested only in the case of the Particle type standard marker (Valdai Karelian *čim*, Eastern Votic and Central Ludian *čem*) and the parameter marker (Livonian *jo*). While Valdai Karelian represented the language use of the 1960s, Ludian *čem* originated from fieldwork data in 2012. It is important to note that all relevant examples have been collected from bilingual speakers at the last stage of these varieties. By comparison, the entire text collection by Ojansuu et al. (1934) contained no examples of *čem*. Raun (1960), for instance, does not mention corresponding MAT-borrowings for any Finnic languages (only for Mordvin and further east). This seems to suggest that such MAT-borrowings are of relatively recent origin.

The use of *jo* (< MAT-borrowing from Latvian) as the main means for marking the parameter in Courland Livonian makes it stand out among the other Finnic varieties. Namely, Finnic comparative constructions can generally be expected to contain a synthetic parameter marker (< **mpV*), which was also the case for the examples in the dataset. Courland Livonian, by contrast, tends to contain instances of *jo* + positive degree. However, while the Particle type makes Courland Livonian the closest to SAE, not using a synthetic parameter marker but replacing it with an analytic marker could place it even one step closer. As noted, using a special affix for a parameter marker is especially common in the languages spoken in Europe (Stassen 2013; see also § 1). Still, as shown in the examples from certain Germanic languages (see § 4.1), synthetic marking is not always possible and an analytic marker is used instead. For instance, in Swedish, words with certain structures do not permit synthetic parameter marking. At the same time, in North Estonian and especially in its western and insular dialects, the historical parameter marker can be encoded on the parameter (synthetic marking) or on the adverb (analytic marking), the latter then combines with an adjective in the positive or comparative degree, while the choice between synthetic and analytic marking does not depend on the structure of the adjective (e.g., the North Estonian options for saying 'more beautiful': *ilusam* ∼ *rohkem ilus* ∼ *rohkem ilusam*). By comparison, in the eastern Uralic languages it is more common for a parameter marker (either synthetic or analytic) to occur only optionally (see Stolz 2013).

6 Conclusions

This study reviewed already identified patterns of Finnic comparative constructions, while also looking for additional possible ways of expressing comparison and the paths of development for these constructions. The main focus of the study was on the southernmost Finnic varieties. In order to be able to identify the spread of different constructions, linguistic data were included from various sources and time periods. The data were collected from all relevant local varieties; many of which have previously found less attention in comparative work (e.g., North Estonian and South Estonian dialect data). The results were placed in an areal context by including information from the neighbouring non-cognate languages: Latvian, Latgalian, Russian, and to some extent also Lithuanian.

The main types of comparative constructions that were analysed were grouped into two main types, the **Locational type** (i.e., (a) the Separative type containing the standard encoded using the partitive, elative, or genitive case; (b) the Goal/Locative type encoded using the adpositions 'over, above' and 'against') and the **Particle type** (i.e., three possible subtypes: *ku-* 'than', (*ku-* +) a negative marker, *mi-* 'what'). The spread of these construction types shows the importance of areal effects but also points to possible language-internal developments. For instance, for the case marking used with the Separative type (a subtype of the Locational type), it was possible to observe a cline from east to west. Namely, the partitive case, which in the Finnic languages is an old separative case, was better preserved in the east (but also in the peripheral areas of western Estonia). A similar kind of synthetic marking is also used in the neighboring Russian language, which also shows similarities in the order of the parameter and standard. In general, pronouns tended to have the best preservative effect. In North Estonian and the surrounding South Estonian varieties, the elative case (a newer separative case) has taken over. Use of the elative has also extended to the western dialects of Votic and Ingrian. By using a separative case, these varieties, however, also show the need to adhere to the Separative type, which is originally associated with Uralic comparative constructions. Livonian, by preferring the Particle type, has moved furthest from the Separative type, and closest to SAE.

The Goal/Locative type was the other main subtype of locational comparatives discussed in this paper. This type could be presented primarily due to contacts between Finnic, Baltic, and Slavic in the Central Baltic area. The Livonian examples of the 'over, above' type (*yl* (*ül*) ∼ *i'l*), which were clear

instances of PAT-borrowing from Latvian *par*, seem to indicate that replacing native case-marking of the standard with an adpositional construction is a possible outcome of contacts in a multilingual setting. A typological peculiarity we noted was the 'against' construction used in South Estonian (Seto, eastern Võro, Lutsi, Leivu), which correlates with similar (archaic) examples from Latvian and Lithuanian. Still, there were further examples of adpositional constructions, which could be considered possible sources for marking the standard. These examples contained adpositions with the source meaning 'next to', 'near, by, at'. By comparison, Hungarian uses a case with the latter meaning to mark the standard in comparative constructions.

The Particle type was present in all of the studied Finnic varieties, with the native *ku-* (originally an instructive form expressing 'how'; cf. German *wie* 'how') being spread most widely (except in Ludian and Veps), though to different extents. In the Finnic varieties spoken in present-day Latvia, *ku-* could also be accompanied by a negative marker or, less commonly, the negative marker could be the sole marker. Compared to the Separative type, the use of *ku-* in the Finnic languages is a more recent way of encoding the standard; however, given its wide spread, it is still relatively old. This seems be supported by the fact that Veps uses the pronoun *mi-* 'what', which copies the corresponding Russian interrogative marker *čem* 'with what'. Furthermore, in Valdai Karelian (South Karelian) and eastern Votic but also in the recent Ludian data, Russian *čem* even appeared as a MAT-borrowing. Along with Livonian *jo*, which is originally a MAT-borrowing from Latvian, used as an analytical standard marker, *čem* was another example of MAT-borrowing in the area. In general, PAT-borrowings were more common.

To conclude, the development of comparative constructions in Finnic varieties can be observed as an ongoing process where both the language-internal developments and (non-cognate) contact languages have an important role.

Acknowledgements

We are really thankful to the anonymous reviewers for their helpful and insightful comments and suggestions. We would also like to thank Uldis Balodis for proofreading the manuscript.

Abbreviations

1, 2, 3	person
ACC	accusative
ADE	adessive
ALL	allative
ACT–	active
AUX	auxiliary
CL	clitic
CLiv	Courland Livonian
CMPR	comparative
CNG	connegative
COM	comitative
COND	conditional
CONJ	conjunction
DEM	demonstrative
DIM	diminutive
EVot	Eastern Votic
ELA	elative
GEN	genitive
HIng	Heva dialect of Ingrian
ILL	illative
IMP	imperative
INE	inessive
INF	infinitive
INS	insular
IPS	impersonal
Kra	Kraasna
Kre	Krevin
Lav	Latvian
Lei	Leivu
Lith	Lithuanian
LIng	Lower Luga Ingrian
LOC	locative
Ltg	Latgalian
Lud	Ludian
Lut	Lutsi
M	masculine
Mul	Mulgi
NEG	negative
NEst	North Estonian

NKar	North Karelian
NOM	nominative
PL	plural
POSS	possessive
PRS	present
PRT	partitive
PST	past
PTCL	particle
PTCP	participle
Q	question particle
REFL	reflexive
Rus	Russian
Set	Seto
SG	singular
SIng	Soikkola Ingrian
SKar	South Karelian
SLiv	Salaca Livonian
StEst	Standard Estonian
StFin	Standard Finnish
SUP	supine
TERM	terminative
TRA	translative
Trt	Tartu
Vot	Votic
Vps	Veps
Vro	Võro
WVot	Western Votic

Data sources

AES 202 = *Akadeemilise Emakeele Seltsi ülevaated: Häälikuloolisi andmeid ja tekste Kraasna murdest* [Reviews of the Academic Mother Language Society: Soundhistorical data and texts from the Kraasna dialect]. Ed. by Ojansuu, Heikki [1938]. Eesti Keele Instituut. (http://emsuka.eki.ee/view/book/175/0). (Accessed 2022-12-20).

Ariste 1960 = Ariste, Paul. 1960. Isuri keelenäiteid [Text examples of Ingrian]. In Ariste, Paul (ed.), *Soome-ugri keelte küsimusi*, 7–68. (Keele ja Kirjanduse Instituudi uurimused V). Eesti Riiklik Kirjastus.

COWE = *Vana kirjakeele korpus* [Corpus of old written Estonian]. (https://vakk.ut.ee/). (Accessed 2023-01-20).

DRJa = Avanesov, Ruben I. & Gorškov, Aleksandr I. (eds.). 1985. *Dialektografija russkogo jazyka* [Dialectography of the Russian language]. Nauka.

EDC = Lindström, Liina & Todesk, Triin & Pilvik, Maarja-Liisa. 2022. *Eesti murrete korpus* [Estonian dialect corpus]. Tartu Ülikooli eesti ja üldkeeleteaduse instituut. https://doi.org/10.23673/re-365

EKI WK = *Wiedemanni sõnavarakogu* [Wiedemann's archive]. Eesti Keele Instituudi eesti murrete ja soome-ugri keelte arhiiv.

EKSS 2009 = Langemets, Margit & Tiits, Mai & Valdre, Tiia & Veskis, Leidi & Viks, Ülle & Voll, Piret (eds). 2009. *Eesti keele seletav sõnaraamat* [Explanatory dictionary of the Estonian language]. Eesti Keele Sihtasutus. (http://www.eki.ee/dict/ekss/). (Accessed 2023-01-10).

EMS = Haak, Anu & Juhkam, Evi & Lonn, Varje et al. (eds.). 1994–2022. *Eesti murrete sõnaraamat* [Dictionary of Estonian dialects] (*a–puisklema*). Eesti Keele Instituut. (https://www.eki.ee/dict/ems/). (Accessed 2023-01-10).

EM IX = Mets, Mari & Haak, Anu & Iva, Triin & Juhkason, Grethe & Kalmus, Mervi & Norvik, Miina & Pajusalu, Karl & Teras, Pire & Tuisk, Tuuli & Vaba, Lembit (eds.). 2014. *Eesti murded IX: Lõunaeesti keelesaarte tekstid* [Estonian dialects IX: Texts of South Estonian language islands]. Eesti Keele Instituut, Tartu Ülikool.

EVS = *Eesti-võro sõnaraamat / Võro-eesti sõnaraamat* [Estonian-Võro dictionary / Võro-Estonian dictionary]. Võro Instituut. (https://synaq.org/). (Accessed 2023-01-10).

HS = Kokla, Paul. 2015. *Hiiu sõnaraamat* [Hiiu dictionary]. Ed. by Sedrik, Meeli. Eesti Keele Sihtasutus. (http://www.eki.ee/dict/hiiu/). (Accessed 2023-01-06).

IMS = Pall, Valdek. 2016. *Idamurde sõnastik* [Eastern dialect dictionary]. 2., täiendatud ja parandatud trükk. Ed. by Niit, Ellen. Eesti Keele Sihtasutus. (http://www.eki.ee/dict/ida/). (Accessed 2022-12-16).

Kallas 1903 = Kallas, Oskar. 1903. *Kraasna maarahvas* [Kraasna county folk]. Soome Kirjanduse Selts.

KES = Leas, Reene & Könninge, Reti & Murulauk, Silvi & Niit, Ellen. 2016. *Kihnu sõnaraamat* [Kihnu dictionary]. Ed. by Pajusalu, Karl & Viikberg, Jüri. Eesti Keele Sihtasutus. (http://www.eki.ee/dict/kihnu/). (Accessed 2022-12-16).

KKVS = Torikka, Marja & Joki, Leena (eds.). 2009. *Karjalan kielen verkkosanakirja*. (Kotimaisten kielten tutkimuskeskuksen verkkojulkaisuja 18). Kotimaisten Kielten Keskus. (https://kaino.kotus.fi/cgi-bin/kks/karjala.cgi). (Accessed 2022-12-16).

Laanest 1966 = Laanest, Arvo. 1966. *Isuri murdetekste*. [Ingrian dialectal texts]. Valgus.

Laanest 1997 = Laanest, Arvo. 1997. *Isuri keele hevaha murde sõnastik* [Dictionary of the Hevaha dialect of Ingrian]. Eesti Keele Instituut.

LELS = Viitso, Tiit-Rein & Ernštreits, Valts. 2012. *Līvõkīel-ēstikīel-leļkīel sõnārõntõz / Liivi-eesti-läti sõnaraamat / Lībiešu-igauņu-latviešu vārdnīca* [Livonian-Estonian-Latvian dictionary]. Tartu Ülikool, Latviešu valodas aģentūra.

LiLa = *Lithuanian-Latvian-Lithuanian parallel corpus*. (http://lila.korpuss.lv/). (Accessed 2022-12-01).

LWB = Kettunen, Lauri. 1938. *Livisches Wörterbuch mit grammatischen Einleitung*. (Lexica Societatis Fenno-Ugricae V). Suomalais-Ugrilainen Seura.

MS = Ilves, Kristi & Laande, Alli & Pajusalu, Karl & Todesk, Triin & Tomingas, Marili. 2021. *Mulgi sõnaraamat I* [Mulgi dictionary I]. Mulgi Kultuuri Instituut, Eesti Keele Instituut, Tartu Ülikool. (http://www.eki.ee/dict/mulgisuur/). (Accessed 2022-12-01).

Nirvi 1971 = Nirvi, Ruben E. 1971. *Inkeroismurteiden sanakirja*. (Lexica Societas Fenno-Ugricae XVIII). Suomalais-Ugrilainen Seura.

Norvik & Saar 2012 = Norvik, Miina & Saar, Eva. 2012. Unpublished transcriptions of fieldwork recordings.

Ojansuu et al. 1934 = Ojansuu, Heikki & Kujola, Juho & Kalima, Jalo & Kettunen, Lauri (eds.). 1934. *Lyydiläisiä kielennäytteitä*. (Suomalais-Ugrilaisen Seuran Toimituksia LXIX). Suomalais-Ugrilainen Seura.

RKS = Norvik, Piret & Kendra, Heli. 2019. *Kuusalu rannakeele sõnaraamat* [Dictionary of the Coastal dialect of Kuusalu]. Ed. by Oja, Vilja. Emakeele Sihtasutus. (https://www.eki.ee/dict/kuusalu/). (Accessed 2022-11-11).

SES = Saar, Eva & Hagu, Paul & Käsi, Inge & Leivo, Maeve & Mets, Mari & Pook, Hanna & Pajusalu, Karl. In preparation. *Seto-eesti sõnaraamat* [Seto-Estonian dictionary].

SESS = Saar, Eva & Hagu, Paul & Käsi, Inge & Leivo, Maeve & Pook, Hanna & Pajusalu, Karl. 2020. *Seto eripäraste sõnade sõnaraamat* [Seto dictionary of unique words]. Ed. by Saar, Eva & Viikberg, Jüri. Emakeele Sihtasutus. (https://www.eki.ee/dict/setoeri/). (Accessed 2023-07-11).

SRNG = *Slovar' russkix narodnyx govorov 35* [Dictionary of Russian folk dialects 35]. (*Revet'–Rjaščik*). 2001. Rossijskaja akademija nauk, Institut lingvističeskix issledovanij, Nauka.

Tolkovye = *Tolkovye onlajn-slovari russkogo jazyka* [Explanatory online dictionaries of the Russian language]. (https://lexicography.online/explanatory/). (Accessed 2022-02-13).

VepKar = *Open corpus of Veps and Karelian languages*. (http://dictorpus.krc.karelia.ru/en). (Accessed 2022-11-11).

VKS = *Vadja keele sõnaraamat / Vad'd'aa tšeelee sõna-tširja / Словарь водского языка* [Votic dictionary]. 2013. 2., täiendatud ja parandatud trükk. Ed. by Grünberg, Silja. Eesti Keele Instituut, Eesti Keele Sihtasutus.

(https://www.eki.ee/dict/vadja/). (Accessed 2023-09-27).

VVS = Käsi, Inge. 2011. *Vanapärase Võru murde sõnaraamat: Rõuge, Vastseliina, Setu* [The dictionary of old Võru dialect: Rõuge, Vastseliina, Setu]. Ed. by Neetar, Helmi. Eesti Keele Sihtasutus.

References

Ambrazas, Vytautas (ed.). 1997. *Lithuanian grammar*. Baltos lankos.

Ariste, Paul. 1948. *Vadja keele grammatika* [Grammar of Votic]. (Nõukogude soome-ugri teadused 9). Teaduslik Kirjandus.

— 1960. Isuri keelenäiteid [Text examples of Ingrian]. In Ariste, Paul (ed.), *Soome-ugri keelte küsimusi*, 7–68. (Keele ja Kirjanduse Instituudi uurimused V). Eesti Riiklik Kirjastus.

Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena (eds.). 2022. *The Oxford guide to the Uralic languages*. Oxford University Press.

Balodis, Uldis & Pajusalu, Karl. 2021. Introductory survey of the South Estonian language islands. *Eesti ja soome-ugri keeleteaduse ajakiri = Journal of Estonian and Finno-Ugric Linguistics* 12(2). 7–31. https://doi.org/10.12697/jeful.2021.12.2.01

Čekmonas, Valery. 2001. On some Circum-Baltic features of the Pskov-Novgorod (Northwestern Central Russian) dialect. In Dahl, Östen & Koptjevskaja-Tamm, Maria (eds.), *The Circum-Baltic languages: Typology and contact, volume 1: Typology and contact*, 339–359. (Studies in Language Companion Series 55). John Benjamins.

Cibuļš, Juris & Leikuma, Lidija. 2003. *Vasals! Latgaliešu valodas mācība* [Hello! Learning the Latgalian language]. n. i. m. s. (https://www.lu.lv/filol/dialekt/publikacijas/Vasals.pdf). (Accessed 2023-01-30).

Dikken, Marcel. 2005. Comparative correlatives comparatively. *Linguistic Inquiry* 36(4). 497–532.

Duden. 2009. *Die Grammatik: Unentbehrlich für richtiges Deutsch, Band 4*. 8., überarbeitete Auflage. Dudenverlag.

DWDS = *Der deutsche Wortschatz von 1600 bis heute*. 2023. (https://www.dwds.de/). (Accessed 2023-01-24).

EKG II = Erelt, Mati & Kasi, Reet & Metslang, Helle & Rajandi, Henno & Ross, Kristiina & Saari, Henn & Tael, Kaja & Vare, Silvi. 1993. Eesti keele grammatika II: Süntaks: Lisa: kiri [Estonian Grammar II: Syntax: Supplement: letter]. In Erelt, Mati & Erelt, Tiiu & Saari, Henn & Viks, Ülle (eds.), Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut.

Endzelin, Janis. 1923. *Lettische Grammatik*. Carl Winter.

Erelt, Mati. 1986. *Eesti adjektiivisüntaks* [Syntax of Estonian adjectives]. (Eesti NSV Teaduste Akadeemia Emakeele Seltsi toimetised 19). Valgus.

— 2017. Võrdlustarind omadus- ja määrsõnafraasina [Comparative construction used as an adjective and adverb phrase]. In Erelt, Mati & Metslang, Helle (eds.), *Eesti keele süntaks* [Estonian syntax], 430–445. (Eesti keele varamu III). Tartu Ülikooli Kirjastus.

FED = *Suomen etymologinen sanakirja*. 2022. (https://kaino.kotus.fi/suomenetymologinensanakirja). (Accessed 2023-01-06).

Grünthal, Riho. 2015. *Vepsän kielioppi*. (Apuneuvoja suomalais-ugrilaisten kielten opintoja varten XVII). Suomalais-Ugrilainen Seura.

— 2023. Diachronic bottlenecks of the Uralic (ablative-)partitive. *Linguistic Variation* 23(1). 124–156. https://doi.org/10.1075/lv.21003.gru

Hagu, Paul & Pajusalu, Karl. 2021. *Seto keele teejuht* [A guide to the Seto language]. Seto Instituut.

Haspelmath, Martin. 1998. How young is standard average European? *Language Sciences* 20(3). 271–287.

— 2001. The European linguistic area: Standard average European. In Haspelmath, Martin & König, Ekkehard & Oesterreicher, Wulf & Raible, Wolfgang (eds.), *Language typology and language universals: An international handbook, vol. 2*, 1492–1510. (Handbücher zur Sprach- und Kommunikationswissenschaft 20.2). Walter de Gruyter.

Heine, Bernd. 1997. *Cognitive foundations of grammar*. Oxford University Press.

Heine, Bernd & Kuteva, Tania. 2006. *The changing languages of Europe*. Oxford University Press.

Itkonen, Terho. 1983. Välikatsaus suomen kielen juuriin. *Virittäjä* 87(2–3). 190–229, 349–386.

— 2011. *Koltan- ja kuolanlapin sanakirja*. (Lexica Societatis Fenno-Ugricae XV). Suomalais-Ugrilainen Seura.

Kallio, Petri. 2012. The prehistoric Germanic loanword strata in Finnic. In Grünthal, Riho & Kallio, Petri (eds.), *A linguistic map of prehistoric Northern Europe*, 225–238. (Mémoires de la Société Finno-Ougrienne 266).

— 2014. The diversification of Proto-Finnic. In Ahola, Joonas & Frog & Tolley, Clive (eds.), *Fibula, fabula, fact: The viking age in Finland*, 155–170. Suomalaisen Kirjallisuuden Seura.

Koptjevskaja-Tamm, Maria & Wälchli, Bernhard. 2001. The Circum-Baltic languages: An areal-typological approach. In Dahl, Östen & Koptjevskaja-Tamm, Maria (eds.), *The Circum-Baltic languages: Typology and contact, volume 2: Grammar and typology*, 615–750. (Studies in Language Companion Series 55). John Benjamins.

Kuteva, Tania & Heine, Bernd & Hong, Bo & Long, Haiping & Narrog, Heiko & Rhee, Seongha. 2019. *World lexicon of grammaticalization*. 2nd edn. Cambridge University Press.

Laakso, Johanna. 2022a. Finnic. In Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena (eds.), *The Oxford guide to the Uralic languages*, 240–253. Oxford University Press. https://doi.org/10.1093/oso/9780198767664.003

— 2022b. Livonian. In Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena (eds.), *The Oxford guide to the Uralic languages*, 380–391. Oxford University Press. https://doi.org/10.1093/oso/9780198767664.003

Laanest, Arvo. 1975. *Sissejuhatus läänemeresoome keeltesse* [Introduction to Finnic languages]. Ed. by Valmen Hallap. Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut.

— 1986. *Isuri keele ajalooline foneetika ja fonoloogia* [Historical phonetics and phonology of Ingrian]. Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut.

Lehtinen, Tapani. 2007. *Kielen vuosituhannet: Suomen kielen kehitys kantauralista varhaissuomeen* [Xxx.] (Tietolipas 215). Suomalaisen Kirjallisuuden Seura.

Markus, Elena & Rozhanskiy, Fedor. 2017. *Sovremennyj vodskij jazyk*. 2-e izdanije, ispravlennoje i dopolnennoje. Nestor-Istorija.

Matras, Yaron & Sakel, Jeanette. 2007. Investigating the mechanisms of pattern replication in language convergence. *Studies in Language* 31(4). 829–865.

Nau, Nicole. 2011. *A short grammar of Latgalian*. (Languages of the World/Materials 482). Lincom Europa.

Palmeos, Paula. 1962. *Karjala valdai murrak* [The Valdai Karelian dialect]. (Emakeele Seltsi Toimetised 5). Eesti NSV Teaduste Akadeemia.

Prillop, Külli & Pajusalu, Karl & Saar, Eva & Soosaar, Sven-Erik & Viitso, Tiit-Rein. 2020. *Eesti keele ajalugu* [History of the Estonian Language]. (Eesti keele varamu VI). Tartu Ülikooli Kirjastus.

Raun, Alo. 1960. The equivalents of English "than" in Finno-Ugric. In Sebeok, Thomas A. (ed.), *American studies in Uralic linguistics*, 149–247. (Uralic and Altaic Series 1). Indiana University.

Rudzīte, Marta. 1964. *Latviešu dialektoloģija* [Latvian dialectology]. Latvijas valsts izdevniecība.

SAG = Teleman, Ulf & Hellberg, Staffan & Andersson, Erik. 1999. *Svenska Akademiens grammatik 1–4*. Svenska akademien & Norstedts ordbok. (https://svenska.se/grammatik/). (Accessed 2023-07-11).

Sakel, Jeanette. 2007. Types of loan: Matter and pattern. In Matras, Yaron & Sakel, Jeanette (eds.), *Grammatical borrowing in cross-linguistic perspective, empirical approaches to language typology [EALT]*, 15–29. Mouton de Gruyter.

Sjögren, Andreas Johan & Wiedemann, Ferdinand Johann. 1861. *Joh. Andreas Sjögren's gesammelte Schriften, Band II, Theil I: Joh. Andreas Sjögren's Livische Grammatik nebst Sprachproben*.

Stassen, Leon. 1985. *Comparison and Universal Grammar*. Blackwell.

— 2013. Comparative constructions. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology. (http://wals.info/chapter/121). (Accessed 2022-11-11).

Stolz, Thomas. 2013. *Competing comparative constructions in Europe*. Akademie Verlag. https://doi.org/10.1524/9783050064994

Tanning, Salme. 1961. *Mulgi murdetekstid* [Dialectal texts of Mulgi]. (Eesti murded I). Eesti Riiklik Kirjastus.

Timberlake, Alan. 2004. *A reference grammar of Russian*. Cambridge University Press.

Treis, Yvonne. 2018. Comparative constructions: An introduction. *Linguistic Discovery* 16(1). https://doi.org/10.1349/PS1.1537-0852.A.492

Vaba, Lembit. 1977. *Läti laensõnad eesti keeles* [Latvian loanwords in Estonian]. Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut.

— 2011. Kuidas läti-eesti keelekontakt on mõjutanud eesti murdekeele grammatikat ja sõnamoodustust [Impact of Latvian-Estonian language contacts on the grammar and word-formation of Estonian dialect]. *Emakeele Seltsi aastaraamat* 56(1). 204–246. https://doi.org/10.3176/esa56

Viikberg, Jüri. 2020. *Eesti murrete grammatika* [Grammar of Estonian dialects]. (Eesti keele varamu). Tartu Ülikooli Kirjastus.

Viitso, Tiit-Rein. 2008. Liivi keele erijooned läänemeresoome keeleruumis [Special features of Livonian in the context of Finnic languages]. In Pajusalu, Karl & Sutrop, Urmas & Teras, Pire (eds.), *Liivi keel ja läänemeresoome keelemaastikud*, 309–355. Eesti Keele Sihtasutus.

Virtaranta, Pertti. 1986. *Haljärven lyydiläismurteen muoto-oppia*. (Suomalais-Ugrilaisen Seuran toimituksia 190). Suomalais-Ugrilainen Seura.

VISK = Hakulinen, Auli & Vilkuna, Maria & Korhonen, Riitta & Koivisto, Vesa & Heinonen, Tarja Riitta & Alho, Irja. 2004. *Iso suomen kielioppi*. Suomalaisen Kirjallisuuden Seura. (http://scripta.kotus.fi/visk). (Accessed 2023-09-27).

Winkler, Eberhard. 1997. *Krewinisch: Zur erschließung einer ausgestorbenen ostseefinnischen Sprache*. (Veröffentlichungen der Societas Uralo-Altaica 49). Harassowitz.

Winkler, Eberhard & Pajusalu, Karl. 2018. *Salis-Livisch II: Grammatik und Wörterverzeichnis: Mit einem Anhang zu den salis-livischen Sprichwörtern*. Harrassowitz.

Zaiceva, Maria Ivanovna. 1981. *Grammatika vepsskogo jazyka: Fonetika i morfologija* [Grammar of Veps]. Akademia nauk CCCP, Karel'skij filial, Institut Jazyka, Literatury i Istorij.

Zaikov, Pekka. 2013. *Vienankarjalan kielioppi*. Karjalan Sivistysseura.

Appendix A Distribution of comparative constructions in the Finnic varieties

Area	Language / Variety	Separative			Goal/Locative		Particle		
		Prt	Ela	Gen	against	over, above	*ku-; mi-; < čem*	Neg + 'than'	Neg
North Estonian	Northeastern Coastal		+				*ku, kui*		
	Eastern		+				*ku, kui*		
	Insular	+	+				*kui, kuĭe,*		
							kut		
	Western			+			*kui*		
South Estonian	Mulgi	+	+				*ku*		
	Tartu	+	+				*ku*		
	Võro	+	+		*vasta*		*ku*		
	Seto	+	+		*vasta*		*ku, kui*		
Varieties (once) spoken in present-day Latvia	SLivonian	+				*ül*	*ku, kui*	*ap ku, ap ka*	*ep*
	CLivonian	+	+			*i'*	*ku*	*äb ku ~ äbku*	*iś*
	Krevin								
	Leivu	+	+		*vasta*		*ku*	*ei? ku, ei ku, iś ku, nigu*	
	Lutsi	+	+		*vasta*		*kuu*	*eigu, iśkui*	*ei*
Varieties (once) spoken in present-day Russia	Kraasna	+	+						
	Votic	+	+				*ku, kui,*		
	Ingrian	+	+	+			*tšem*		
	Karelian	+		+			*ku, kuin*		
							ku, kui, kun, kuin		
	Väldai (SKar)	+		+			*čim*		
	Ludian	+	+				*čem*		
	Veps	+					*mi*		

Contact information:

Miina Norvik
Institute of Estonian and General Linguistics
Ülikooli 18
50090 Tartu
Estonia
e-mail: miina(dot)norvik(at)ut(dot)ee

Eva Saar
Institute of Estonian and General Linguistics
Ülikooli 18
50090 Tartu
Estonia
e-mail: eva(dot)saar(at)ut(dot)ee

Suomen putoamis- ja kaatumisverbit leksikaalisen typologian näkökulmasta

Johanna Viimaranta
Helsingin yliopisto

Tiivistelmä

Artikkeli käsittelee suomen kielen ylhäältä alaspäin suuntautuvan liikkeen verbejä, joista keskeisimmät ovat *pudota*, *tippua*, *kaatua* ja *romahtaa*. Suomen 34 verbilekseemin käyttömahdollisuuksia tarkastellaan 42 eri kielen aineistoja käsitelleen tutkimushankkeen tulosten pohjalta. Hankkeen tulosten mukaan liikkeeseen ylhäältä alaspäin kuuluu neljä erityyppistä kehystä eli yleistettyä tilannetta. Ensimmäinen kehys on putoaminen korkeammalta tasolta alemmalle, toinen vertikaalisen orientaation menettäminen eli kaatuminen, kolmas jossain kiinni tai johonkin yhteydessä olevan olion irtoaminen ja putoaminen ja neljäs luhistuminen (romahtaminen alas alkuperäisen rakenteen menettäen). Tutkituissa kielissä näitä neljää peruskehystä voidaan kuvata 1–4 eri merkitykseltään laajalla verbillä ja lukuisilla muilla suppeamman semantiikan verbeillä. Suomesta ei löydy verbilekseemiä, jolla voisi kuvata kaikkia näitä neljää peruskehystä, vaan siinä erotetaan toisistaan toisaalta horisontaalinen liike alaspäin ja vertikaalisen orientaation menettäminen (*pudota*/*tippua* vs. *kaatua*). Myös rakenteen menettämiseen luhistumalla liittyvät tilanteet vaativat omat lekseeminsä (mm. *romahtaa*, *luhistua*, *sortua*).

Suomessa, kuten monissa muissakin kielissä putoava tarkoite vaikuttaa verbin valintaan: sateen ja aineiden sekä samankaltaisista pienistä osasista koostuvien tarkoitteiden kuten lehtien tai neulasten putoamisesta käytetään erityisiä lekseemejä. Suomessa voi kuitenkin käyttää samoja verbejä elollisista ja elottomista olioista. Alaspäin suuntautuvat liikkeen lähtöpiste ei ole suomessa keskeinen verbilekseemin valinnan kannalta, mutta sen sijaan putoamistapaa, päätepistettä ja etenkin kontaktin tapaa päätepisteessä voi tarkentaa verbin valinnalla.

Avainsanat: leksikaalinen typologia, liikeverbit, suomi, kielten vertailu, semantiikka, Moskovan semanttinen koulukunta, synonymia

1 Johdanto

Tässä artikkelissa kuvaan suomen kielen ylhäältä alaspäin suuntautuvan liikkeen verbejä leksikaalisen typologian näkökulmasta. Näistä sanoista keskei-

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

Finnish Journal of Linguistics 36 (2023), 177–212
DOI: 10.61197/fjl.113910

simmät ovat *pudota*, *tippua*, *kaatua* ja *romahtaa*, mutta verbejä on myös kymmeniä muita. Tässä artikkelissa ei käsitellä putoamiseen liittyviä kausaatioverbejä kuten *kaataa*, *tiputtaa*, *heittää* eikä muutenkaan paneuduta johdoksiin. Keskiössä ovat verbit, jotka esiintyvät laajimmin erilaisissa ylhäältä alaspäin suuntautuvan liikkeen merkityksissä. Analyysi pohjaa 42 eri kielen[1] aineistoja tutkineeseen projektiin, jonka lopputuotos, *Acta Linguistica Petropolitana* -lehden yli 1 000-sivuinen erikoisnumero ilmestyi vuonna 2020. Hankkeen tutkimuksen kohde oli ”kontrolloimaton painovoiman takia tapahtuva liike ilmassa ilman kontaktia mihinkään pintaan” (Reznikova ym. 2020: 9). Kutsun hanketta tässä artikkelissa yksinkertaisuuden vuoksi *putoamisverbiprojektiksi*. Edellä mainitun kokoelman suomea koskeva venäjänkielinen artikkeli on minun ja Ekaterina Protassovan kirjoittama (Viimaranta & Protassova 2020). Tämä artikkeli tukeutuu samaan aineistoon kuin aiempi venäjänkielinen artikkeli, mutta tässä artikkelissa ei aiemmasta artikkelista poiketen kuvata pelkästään suomea vaan verrataan sitä muihin hankkeen kieliin. Leksikaalistypologinen vertailu muihin kieliin mahdollistaa tarkemman ja objektiivisemman näkökulman suomen verbivalintaan vaikuttaviin seikkoihin.

Putoamisverbiprojektin lähtöoletus oli, että on olemassa universaali putoamisen semanttinen kenttä, jolla on toisiinsa liittyviä kielenkäyttötilanteita eli kehyksiä. Kehysten lisäksi toinen vertailun yksikkö on verbilekseemin käyttömahdollisuus tietyssä kehyksessä. Erityisesti tämä koskee ns. dominanttiverbejä eli merkitykseltään laajoja käsiteltävän semanttisen alueen keskeisiä verbejä. Näiden verbien kuvausta pidetään erityisen tärkeänä, koska eri kielten vertailussa yksittäisten kielten laaja- tai monimerkityksisten lekseemien käyttömahdollisuuksien rajat asettuvat eri kohdille. Toisaalta kuvataan myös, mitkä seikat vaikuttavat muiden samoissa kehyksissä käytettävien merkityksel-

[1] Projektin tutkimuskohteena on ollut 42 kieltä jotka edustavat 12 eri kielikuntaa: indoeurooppalaisia kieliä (armenia, englanti*, espanja*, gudžarati*, hindi-urdu*, jiddiš, norja (bokmål)*, nykykreikka*, pandžabi*, persia*, ranska*, saksa, serbia, sloveeni, šugnan*, tšekki ja venäjä); dravidakieli tamili*; uralilaisia kieliä (ersä, hanti*, komisyrjäänin ižman murre, mokša, pohjoishanti, suomi*, tundranenetsi, unkari ja vuorimari); tšuktšilais-kamtšatkalainen tšuktši*, turkkilaisia kieliä (kazakki* ja turkki*), koilliskaukasialaisia kieliä (adyge*, aguli*, karata* ja rutuli*); seemiläinen tigrinja*, austronesialainen indonesia, austroaasialainen khmer*, sinotiibetiläinen mandariinikiina*, japanilainen kieli japani*, venäläinen viittomakieli* sekä isolaattikielet baski* ja korea*. Osasta kielistä (tähdellä merkityt) on kuvattu koko putoamisen semanttinen alue, osa puolestaan on mukana jossain suppeammasta näkökulmasta tehdyssä usean kielen vertailussa. Myös oppijan ja perintöpuhujan venäjä on mukana, koska kielenoppijoiden virheitä tutkimalla saadaan tietoa niistä leksikon ominaispiirteistä, joita on muuten vaikea huomata. Tämä liittyy siihen, että oman kielen tekemät eronteot tuntuvat helposti luontevilta ja loogisilta, vaikka nekin ovat yhtä arbitraarisia kuin muissa kielissä.

tään suppeampien verbien valintaan. Tämä lähestymistapa ei pyri kuvaamaan minkään lekseemin koko merkitystä tai sen eri merkitysten suhdetta toisiinsa – kiinnostuksen kohteena ovat ainoastaan käytöt tietyissä kehyksissä, ei lekseemien kuvaaminen polysemian näkökulmasta.

Pyrin tässä artikkelissa vastaamaan seuraaviin kysymyksiin: Mitkä seikat vaikuttavat putoamisverbin valintaan suomessa? Miten suomen putoamiseen liittyvät semanttiset eronteot asettuvat 42 kielen vertailussa?

Seuraavassa esittelen ensin lyhyesti tutkimuksen taustalla olleen venäläisen tutkimusperinteen ja siihen asettuvan putoamisverbihankkeen teoreettisine taustoineen. Esittelen lyhyesti myös ne tulokset, jotka on aiemmassa tutkimuksessa saatu suomeen liittyen. Sen jälkeen esitän analyysini suomen sijoittumisesta leksikaalis-typologisessa vertailussa ja pohdin tämän tiedon laajempaa merkitystä.

2 Teoriatausta

2.1 Leksikaalinen typologia

Vaikka kaikilla kielillä voi lähtökohtaisesti kuvata kaikki tilanteet, on kielten välillä suurtakin eroa siinä, mitä seikkoja kielellistetään. Liikeverbien yhteydessä on mm. tunnettua, että slaavilaiset kielet kuvaavat verbiaspektijärjestelmän ja laajan etuliitevalikoiman avulla hyvinkin tarkkaan liikkeen laatua, kun taas monissa muissa kielissä ei ole pakollista erotella liikkeen tapaa niin tarkkaan ja merkitykseltään laaja-alaiset liikeverbit ovat erittäin frekventtejä. Esim. venäjässä ei ole olemassa käyttömahdollisuuksiltaan laajaa liikeverbiä kuten suomen *mennä* tai englannin *go*, vaan eri käyttökonteksteissa on käytettävä kymmeniä eri verbejä.

Leksikaalinen typologia tutkii sanaston typologista variaatiota eri kielissä (Koptjevskaja-Tamm ym. 2016: 434). Se juontaa juurensa 1990-luvulla yleistyneeseen ajatukseen, jonka mukaan myös sanaston rakentumisen tulisi olla typologisen tutkimuksen kohteena (mm. Lehmann 1990; Behrens & Sasse 1997). Erilaiset leksikaalisen typologian mallit lähtevät erilaisista lähtökohdista ja perusyksiköistä, mm. lekseemeistä, merkityksistä, kieliopillisista piirteistä, käsitteistä tai semanttisista alueista. Vertailukelpoisen tiedon saamiseksi eri kielistä leksikaalisessa typologiassa käytetään usein elisitaatiomenetelmiä, mutta myös sekundaariaineistoa tutkimuskirjallisuudesta, sanakirjoja ja elektronisia korpuksia.

Termin ”leksikaalinen typologia” lisäksi puhutaan joskus myös ”semanttisesta typologiasta”. Nämä kaksi eroavat siinä, että siinä missä leksikaalisen typologian kohteena on nimenomaisesti leksikon eli sanaston typologinen luokittelu, on semanttinen typologia terminä laajempi sisältäen muutkin semantiikan vertailut ja muunkinlaiset merkit kuin sanat (Evans 2012: loppuviite). Itse ymmärrän eron niin, että kun kyse on sanojen ominaisuuksista, on tämä nimenomaisesti leksikaalista typologiaa, kun taas käsiteltäessä esim. semantiikan universaaleja voi tarkastelu ulottua laajemman semanttisen typologian puolelle.

Erilaisia leksikaalisen typologian suuntauksia erottaa se, nähdäänkö leksikon kieliopilliset piirteet kuvauksen lähtökohtana vai yhtenä lekseemin ominaisuuksista. Kieliopillisten kategorioiden merkitystä yleisemmässä keskustelussa leksikaalisesta typologiasta lisää se, että semantiikka on mukana paljon käytetyissä kognitiivisen kielitieteen kieltä kokonaisuudessaan kuvaavissa ja myös kielten vertailussa käytettävissä malleissa kuten esim. kognitiivinen kielioppi (Langacker 2008) ja konstruktiokieliopin eri muodot (mm. Croft 2001). Esittelen seuraavaksi Moskovan semanttisen koulukunnan traditioista lähtevän leksikaalisen typologian suuntauksen, jossa kuvauksen lähtökohtana on lekseemin käyttömahdollisuudet.

2.2 Moskovan semanttinen koulukunta ja MLexT-tutkimusryhmä

Moskovan semanttinen koulukunta sai alkunsa 1960-luvulla. Sille on keskeistä ajatus, että kielessä ei ole täysin synonyymisia sanoja tai muita leksikaalisia yksiköitä vaan ainoastaan *näennäissynonyymejä* (ven. *kvazisinonim*[2]), jotka eroavat toisistaan vähintäänkin mahdollisilta käyttökonteksteiltaan. Näennäissynonyymien eroja tutkitaan mm. korvaustesteillä. Niissä etsitään sellaisia kontekseja, joissa sanoja ei voi korvata toisillaan. Näin pyritään määrittelemään ne tekijät, jotka estävät tai mahdollistavat tietyn lekseemin käytön tietyssä kontekstissa. Alkuperäisen Moskovan semanttisen koulukunnan tutkijat olivat

[2] Kyrillisin kirjaimin kirjoitettavien kielten sanat on esitetty tässä artikkelissa translitteroituina. Venäjän kielen translitteroinnissa on käytetty Suomessa käytössä olevaa tieteellistä translitterointikaavaa. Englanninkielisissä lähteissä esiintyvät henkilönimet on kuitenkin esitetty samassa muodossa kuin lähteessä. Samoin jos venäjänkielisessä artikkelissa on ollut myös translitteroitu muoto nimestä, on käytetty tätä samaa muotoa. (Sekä kyrillisin että latinalaisin kirjaimin kirjoitettavaa) serbiaa on kirjoitettu lähteestä poiketen tässä latinalaisin kirjaimin. Kiitän Arja Hamaria ja Tapani Salmista neuvoista koskien tundranenetsin ja komin translitterointia. Muilla kirjaimistoilla ja merkkijärjestelmillä kirjoitettavien kielten sanat on esitetty siinä muodossa, missä ne ovat olleet viitattavassa lähteessä.

leksikologeja ja leksikografeja ja työn ydin olikin nimenomaan venäjän kielen sanakirjatyössä. (Rakhilina & Reznikova 2014.) Moskovan semanttinen koulukunta uskoo, että kielenkuvauksen on oltava yhtenäistä (ven. *integral'nyj*). Tämä tarkoittaa sitä, että sanojen semantiikan kuvauksessa on kuvattava myös niiden kieliopillinen käyttäytyminen eikä esim. sanakirja-artikkelista saa siten jättää pois tarpeellista tietoa valenssista ja kongruenssista.

Moskovan semanttisen koulukunnan työ on ollut aina voimakkaan aineistopohjaista ja pyrkinyt aineistonkeruun ja -käsittelyn kehittämiseen. Se on ollut toimintansa alusta alkaen kiinnostunut konekääntämisestä ja kieliteknologiasta ja myötävaikuttanut sittemmin venäjän kielen suurten korpusten luomiseen. Vaikka neuvostoliittolainen kielitiede kehittyikin osin eristyksissä länsimaisesta, levisivät Moskovan semanttisen koulukunnan ajatukset myös muualle mm. Kanadaan emigroituneen Igor Mel'čukin ansiosta. (Apresjan 2005; Boguslavskij & Iomdin 2009.)

Alkuperäinen Moskovan semanttinen koulukunta ei ollut kiinnostunut kielitypologiasta eikä sen kuvausmallia ollut tarkoitettu kielten vertailuun (Rakhilina & Reznikova 2014: 3). Sen perintöä jatkaa kuitenkin tässä artikkelissa kuvattavan tutkimushankkeen toteuttaja, nykyisessä muodossaan vuonna 2010 työnsä aloittanut Moskovan leksikaalis-typologinen ryhmä (*Moskovskaja leksiko-tipologičeskaja gruppa*, *Moscow Lexical Typology group*, www.lextyp.org, jäljempänä *MLexT*-ryhmä). Leksikaalisen typologian lähtökohtiin kuuluu MLexT-ryhmän mukaan, että sekä leksikko kokonaisuudessaan ja yksittäisten sanojen merkitys muuttuvat paljon nopeammin kuin kielioppi tai äännejärjestelmä, ja tämän takia kieliotokseen otetaan mieluusti mukaan sukukieliä (myös lähisukukieliä) eikä ajatella, että kustakin kieliryhmästä riittäisi yksi edustaja (Reznikova ym. 2020: 19). Kognaattien käyttömahdollisuuksien tarkastelulla lähisukukielissä saadaan tietoa sanaston rakentumisen ja sen muutoksen johdonmukaisista piirteistä.

Sanaston rakentumisessa erityistä huomiota saa MLexT-tutkimusryhmän edustamassa näkemyksessä niiden lekseemien määrä, jotka välttämättä tarvitaan tietyn asian ilmaisemiseen kielessä. Sama kieli voi eritellä hyvinkin tarkkaan yhden toiminnan mutta käyttää toisesta yhtä, merkitykseltään yleistä verbilekseemiä. Kielten vertailun yksikkönä on *kehys* (englanniksi *frame*). Tämä tarkastelumalli tukeutuu Charles Fillmoren kehyssemantiikkaan (esim. Fillmore 1976; Fillmore & Atkins 1992) ja FrameNet-tietokantaan[3], mutta ei ole yhteneväinen sen kanssa. MLexT-tutkimusryhmän kuvausmallissa ei

[3] https://framenet.icsi.berkeley.edu/fndrupal

nimittäin kuvata ensisijaisesti yksittäisiä semanttisia tai syntaktisia piirteitä, vaan kehyksiä ja niissä käytettäviä lekseemejä. Siinä missä FrameNetin kuvauksessa semanttisen verkon yhteydet muodostuvat niin kehysten kuin kehysten elementtien/osasten (mm. aktanttien) välillä yhden kielen sisällä, MLexT-tutkimusryhmän leksikaalisessa typologiassa ajatellaan, että nimenomaan verbilekseemin käyttömahdollisuus useissa kehyksissä muodostaa yhteyden näiden kehysten välille. Myös kehysten nimitykset ovat konkreettisempia ja yleistajuisempia kuin FrameNetin kuvausmallissa eikä niissä käytetä metakieltä, jolla kuvattaisiin kehyksen tai sen aktanttien suhdetta muihin kehyksiin. Tulkinta siitä, voiko lekseemiä käyttää tietyssä kehyksessä, perustuu MLexT-ryhmän leksikaalisessa typologiassa aina aineistossa todennettuihin toistuviin tapauksiin, ei siis esim. semanttisten vektorien tarjoamille todennäköisyyksille sanojen käyttömahdollisuuksista tai muille distributiivisen semantiikan metodeille (vrt. Baroni ym. 2014; Georgakopoulos & Polis 2018).

Kehys tarkoittaa tiettyä kielennettävää tilannetta tai tilaa. Se sisältää puhujan hallussa olevan tietouden maailman toiminnasta ja niistä toimijoista, jotka liittyvät kuhunkin tilanteeseen (esim. lasin kaatumiseen pöydällä). Kehyksen sisältämät tiedot ovat välttämättömiä lekseemien idiomaattiselle käytölle ja tulkinnalle. Niitä ei kuitenkaan ole mahdollista tietojen laajuuden takia kuvata tyhjentävästi, ja toisaalta koska ajatellaan tietouden olevan jaettua, tämä ei ole tarpeenkaan. Riittää siis, että tiedostetaan, että on olemassa kehys ”lasin kaatuminen pöydällä” ja tämän kehyksen kielentäminen on ymmärrettävissä niille eri kielten puhujille, joilla on kokemusta tästä tilanteesta. Toisaalta kehys voi olla laajempikin (säiliön kaatuminen millä tahansa alustalla) tai suppeampikin (muovisen kuohuviinilasin kellahtaminen kumoon), mutta tällöin myös sen sisältämä tieto on erilainen. Kehys on siis yleistetty merkitys, joka on monia kieliä vertailemalla todettu typologisesti merkittäväksi. Se on omanlaisensa vertailukäsite (vrt. Haspelmath 2010). Kehys ei ole kielikohtainen ja se on luotu keinotekoisesti kielten leksikaalista vertailua varten.

MLexT-ryhmä pyrkii selvittämään tietylle semanttiselle alueelle kuuluvat oletettavasti universaalit kehykset ja tarkastelemaan eri kielten lekseemien käyttömahdollisuuksia näissä kehyksissä. Kehys on olemassa, jos edes yhdestä kielestä löytyy lekseemi, jolla voi kuvata vain sitä, muttei muita saman semanttisen alueen kehyksiä. Hyvänä esimerkkinä tästä on suomen lekseemi *kaatua*, jota voi käyttää vertikaalisen orientaation menettämisestä, mutta ei niissä konteksteissa, joissa suomessa käytettäisiin esim. verbiä *pudota*. Toisaalta semanttisen alueen olemassaolon todistaa sellaisen kielen löytyminen, jossa kaikissa alueen kehyksissä voi käyttää yhtä ja samaa lekseemiä (esim. englannin *fall*).

MLexT-ryhmä lähtee tietyn semanttisen alueen vertailusta mahdollisimman monissa eri kielissä. Tällöin kuvataan mahdollisuuksien mukaan kaikki tutkittavien kielten tämän semanttisen alueen lekseemit niiden käyttömahdollisuuksien kautta. Keskeinen aineistonlähde MLexT-ryhmän tutkimusotteessa on elisitaatio, joka tapahtuu laajojen kyselylomakkeiden ja kirvoitusaineiston (mm. videonpätkien) muodossa. Näiden lisäksi käytetään mahdollisuuksien mukaan korpusaineistoja. Korpusaineistojen tärkeys on siinä, että ne sisältävät aitoa kielenkäyttöä aidossa kontekstissa ja siten antavat lisätietoa elisitaation antamalle tiedolle tai riittävän laajojen aineistojen tapauksessa jopa korvaavat sen. Korpusaineistoja voi myös käsitellä automaattisesti, mikä antaa lisämahdollisuuksia leksikaalisen typologian kehitykselle (Koptjevskaja-Tamm ym. 2016).

MLexT-ryhmän leksikaalinen typologia lähtee polysemian sijaan pikemminkin samalla *lekseemillä kuvattavuuden* (*colexification*) ajatuksesta. Tässä mallissa ei niinkään ole tärkeää määritellä, miten lekseemin eri merkitykset ovat syntyneet tai liittyvät toisiinsa, koska typologisen lähestymistavan kiinnostuksen kohteena on kuvata pikemminkin ne eri merkitykset, joissa samat lekseemit esiintyvät eri kielissä ja löytää näitä merkityksiä yhdistävä tekijä. Kun yhteinen tekijä on löydetty, voidaan sen perusteella määritellä kehys, jonka on tarkoitus toimia kuvausta rajaavana yleistyksenä. Kuvausmallissa käytetään havainnollistamiskeinona *semanttisia karttoja* (*semantic map*), joilla kuvataan semanttisia verkkoja eli lekseemien sijoittumista kehyksiin ja kehysten suhteita toisiinsa (vrt. Croft 2001; François 2008). Semanttisten karttojen käyttöä tähän tapaan on kritisoitu siitä, että kun halutaan kuvata kartoilla kaikki mahdolliset koleksifikaatiot, mukaan voi tulla myös hyvinkin marginaalisia yhteyksiä. Kieliotoksen määrän kasvaessa voi lopulta kaikille merkityksille löytää jonkin yhteyden, vaikka tämä olisikin käyttöjen kannalta hyvin harvinaista. (Cysouw 2007.) Vaikka pidänkin tätä kritiikkiä aiheellisena, ovat leksikaaliset kartat mielestäni havainnollinen tapa kuvata eri kielten lekseemien merkitysten distribuutiota.

MLexT-ryhmän nettisivuilla on listattu parikymmentä eri tutkimusprojektia, jotka käsittelevät verbejä tai adjektiiveja eli joko toiminnan tai ominaisuuden kuvailua. Suomen kieli on aiemmin ollut mukana neljässä yksinomaan venäjäksi dokumentoidussa hankkeessa, joiden aiheet ovat liike vedessä eli esim. uiminen, kelluminen ja purjehtiminen (Mustajoki & Protassova 2007), kipuverbit (Nikunlassi 2013), eläinten ääntelyä kuvaavat verbit (Nikunlassi & Protassova 2015) ja lämpötilaa kuvaavat adjektiivit (Nikunlassi & Juvonen 2015).

2.3 MLexT-ryhmän putoamisverbiprojekti

Putoamisverbiprojektin artikkelikokoelman johdantoartikkelissa (Reznikova ym. 2020) todetaan, että kansainvälisesti liikkumista ilmaiseviin verbeihin kohdistunut kiinnostus on kohdistunut pitkälti väyläverbeihin, mutta liikkeen tapaa kuvaavista verbeistä (kuten putoamisverbeistä) on kirjoitettu hyvin vähän, vaikka putoamisesta voi olettaa olevan tarvetta puhua kaikenlaisissa ihmisyhteisöissä maantieteellisistä ja muista oloista riippumatta. Tämän esitetään liittyvän siihen, että vaikka Leonard Talmyn eri kielten vertailuun innoittaneissa liikeverbitutkimuksissa (Talmy 2000) otetaan huomioon väylän lisäksi liikkeen tapa, on väylä helpommin lähestyttävä, koska se on alku- ja loppupisteineen useammin koodattu verbien valenssiin (joko verbin vartaloon tai erillisenä ”satelliittina”). Sama koskee liikeverbeihin liittyviä deiktisiä kysymyksiä kuten tulemista ja menemistä (Fillmore 1975), jotka koodataan nimenomaisesti havainnoijan sijainnilla väylällä eikä liikkeen tavan mukaan (Reznikova ym. 2020).

Verbin valintaan mahdollisesti vaikuttaviksi piirteiksi todettiin putoamisverbiprojektissa putoavan subjektin tarkoitteen ominaisuudet (elollisuus/elottomuus, koko, koostuminen osista, nestemäisyys), putoamisen tapa (nopeus, liikeradan muoto) ja putoamisen alku- ja loppupisteet (tapahtuuko alkupisteessä irtoaminen, mitä oliolle tapahtuu loppupisteessä) (Reznikova ym. 2020).

MLexT-tutkimusryhmän perusajatuksen mukaisesti kielitieteellisillä korpuksilla on keskeinen rooli leksikaalisen semantiikan tutkimuksessa. Niiden avulla määritellään semanttiseen kenttään kuuluvat sanat, analysoidaan sanojen käyttöjen eroja ja hankitaan esimerkkejä. Kaikista kielistä korpusaineistoja ei kuitenkaan ole olemassa. Aineistojen keräämisessä etenkin näistä kielistä oli tämän vuoksi käytössä myös laaja venäjänkielinen kyselylomakkeen runko, joka oli laadittu putoamisverbiprojektin aiemmissa vaiheissa. Kyselylomakkeen tukena oli stimulusaineisto piirrosanimaatioiden muodossa. Piirrosanimaatiot esittävät seuraavat tilanteet: 1) sormus luiskahtaa pois sormesta, 2) sade ropisee maahan, 3) tuulenpuuska vie hatun ihmisen päästä, 4) lehtiä putoilee puusta, 5) pallo lentää havainnoijan jalkoihin, 6) pöydällä oleva pullo kaatuu ja neste leviää pöydälle, 7) poikki sahattu puu kaatuu maahan, 8) kukkaro tippuu käsilaukusta, 9) vettä valuu vesiputouksena kalliojyrkänteeltä, 10) säkissä on reikä ja siitä valuu ulos hiekkaa, 11) mies istuu omenapuun alla ja omena putoaa hänen päähänsä, 12) lentokone putoaa mereen, 13) pyykkinaru, jolla roikkuu pyykkiä, irtoaa toisesta päästään naulasta, 14) vahaa tippuu kynttilästä pöydälle, 15) mieheltä irtoavat vähitellen kaikki hiukset, 16) ihminen putoaa puun

oksalta, 17) mies ampuu ja toinen kaatuu kasvoilleen maahan, 18) kovat sateet huuhtelevat penkereen mukanaan, 19) mies lyö toista miestä nyrkillä ja tämä kaatuu selälleen, 20) pöytä kaatuu ja kaikki sen päällä olevat astiat tippuvat lattialle, 21) linnunpoikanen putoaa pesästä, 22) ihminen tippuu heikkoihin jäihin ja 23) rakennus luhistuu maanjäristyksessä. Nämä animaatiot sisältävät pelkkää liikkuvaa kuvaa (ei siis ääntä).

Putoamisverbiprojektin tulosten mukaan putoamiseen kuuluu neljä erityyppistä kehystä. Ensimmäinen kehys on putoaminen korkeammalta tasolta alemmalle, toinen vertikaalisen orientaation menettäminen eli kaatuminen, kolmas jonkin kiinni tai yhteydessä olevan irtoaminen ja putoaminen ja neljäs luhistuminen alas alkuperäisen rakenteen menettäen. Tutkituissa kielissä näitä neljää peruskehystä voidaan kuvata 1–4 eri dominanttiverbillä. (Reznikova ym. 2020.) Esim. venäjä ja englanti eivät systemaattisesti erota putoamista ja kaatumista toisistaan, mutta mm. suomi erottaa. Kuvausmallissa tämän kehyksen suhteen dominoiviksi kutsutaan sellaisia kieliä, joissa kaikkia neljää kehystä voi kuvata samalla verbilekseemillä ja heikosti dominoiviksi sellaisia kieliä, joissa tarvitaan kaksi lekseemiä. Distributiivisiksi kutsutaan sellaisia kieliä, joissa jokaista neljää kehystä kuvaamaan tarvitaan eri lekseemit ja heikosti distributiivisiksi sellaisia, joissa lekseemejä tarvitaan kolme (mm. suomi).

Putoamisen semanttisen alueen olemassaolon todistavat sellaiset kielet, joissa kaikkia em. neljää kehystä voi kuvata samalla verbilekseemillä. Esimerkiksi englannissa lekseemiä *fall* voidaan käyttää kaikkien näiden kehysten merkitystä ilmaisemaan, vaikkakin kolmannen kehyksen tapauksessa käytetään tyypillisesti (muttei aina) fraasiverbiä *fall off* ja neljännen kehyksen kohdalla *fall apart* (Vinogradova ym. 2020). Samoin norjan (bokmål) *falle*-verbiä voi käyttää kaikissa näissä kehyksissä (Livanova & Mordashova 2020). Myös seuraavissa kielissä on lekseemit, joita voi käyttää kaikissa näissä neljässä merkityksessä: gudžarati (*paṛvuⁿ*), hindi (*girnā*), pandžabi (*ḍigṇā*), tigrinja (*wädäkä*), nykykreikka (*péfto*) ja baski (*erori*). Myös rutulin verbivartalolla *-irxur* voi kuvata kaikkia neljää kehystä. (Reznikova ym. 2020: 24–25.)

On myös monia kieliä, joissa tiettyä verbilekseemiä voi käyttää kolmessa putoamiskehyksessä, mutta neljättä sillä ei voi kuvata, ts. tarvitaan väistämättä kaksi eri lekseemiä. Putoamisverbiprojektin kielissä ei kuitenkaan ole sellaista tapausta, jossa muita kehyksiä voisi kuvata samalla verbillä, mutta korkeammalta putoaminen vaatisi erillistä verbiä. Tällä perusteella on tultu johtopäätökseen, että kaikkein keskeisin putoamistapahtuma on nimenomaisesti korkealta putoaminen ja tämä ydin yhdistää kaikkia muita putoamiskehyksiä. Erityisen yleinen tutkituissa kielissä on sellainen järjestelmä, jossa muita kolmea kehys-

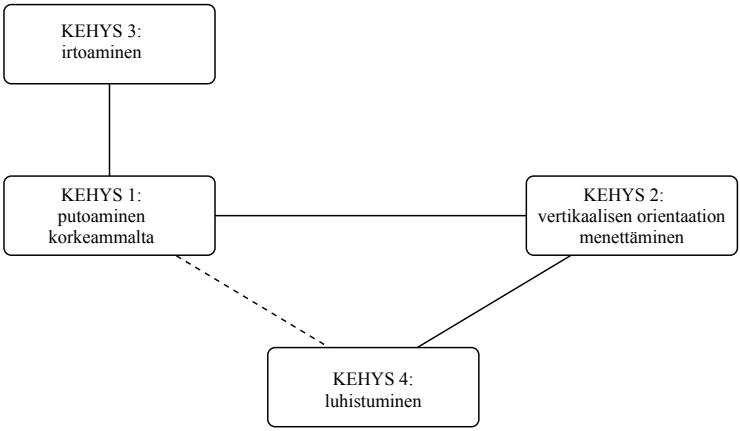
tä voi kuvata samalla verbillä, mutta luhistuminen vaatii oman lekseeminsä. Tällainen laajamerkityksinen verbi kuitenkin ilman rakennuksen yms. luhistumisen merkitystä on mm. ranskan dominanttiverbi *tomber* (Zhukova & Kor Chahine 2020), venäjän *padat'*, korean *tteleci-ta*, turkin *düşmek*, indonesian *jatuh* ja espanjan *caer*(*se*) (Reznikova ym. 2020: 24–25).

Silloin kun neljää peruskehystä kattamaan tarvitaan kolme eri dominanttilekseemiä, on jakauma otoksen kielissä aina samanlainen: ylhäältä putoamisen ja irtoamisen peruskehyksiä voidaan kielellistää samoilla dominanttiverbeillä, mutta luhistumiseen ja kaatumiseen käytetään muita lekseemejä. Putoamisen alkupisteessä tapahtuvaa irtoamista ei ole täten pakko kielellistää ja toisaalta ylhäältä putoaminen, kaatuminen ja luhistuminen nähdään erilaisina toimintoina. Tällaisia kieliä ovat aguli, karata, sloveeni, suomi ja komi. Esimerkiksi karatassa putoamisen ja irtoamisen kehyksiä kuvaamaan käy verbi *t'arała*, kaatumista *karała* ja luhistumista *baχːała*. Muunlaisten kolme dominanttiverbiä vaativien järjestelmien puuttuminen otoksesta tulkitaan niin, että irtoaminen liittyy tiiviimmin korkealta tippumiseen kuin luhistuminen tai kaatuminen. (Reznikova ym. 2020: 27.)

Toinen ääripää ovat kielet, joissa putoamisen neljä peruskehystä vaativat kaikki erilliset lekseeminsä (ns. distributiivinen järjestelmä). Tällaisia ovat putoamisverbiprojektin tietojen mukaan mm. adyge, khmer, unkari ja tšuktši. Esim. adygessa ylhäältä alas putoaminen ilmaistaan vartalolla *-fe-*, vertikaalisen asennon menetys verbillä *wəkʲʷerejə-* ja irtoaminen vartalolla *-zə-*. Luhistumista kuvaamaan käytetään adygessa sen sijaan useita eri lekseemejä. (Reznikova ym. 2020: 26.)

Erikseen voi huomata, että ylipäätään erillistä ylhäältä alaspäin suuntautuvaa liikettä kuvaavaa lekseemiä tai lekseemien joukkoa ei sitäkään välttämättä tarvita – agulissa ei ole mitään varsinaista putoamisverbiä, mutta putoamisen kehyksiä kuvataan muilla dynaamisen liikkeen verbeillä, joiden välinen suhde on täysin kuvattavissa putoamisverbiprojektin parametrien avulla (Reznikova & Merdanova 2020).

Projektin kuvausmallissa saatuja typologisia tietoja hyödynnetään semanttisten karttojen piirtämisessä siten, että putoamisen kentän kuvauksissa neljä peruskehystä ovat eri kielten kuvauksissa samassa järjestyksessä ja niiden semanttiselle typologialle perustuvat suhteet on kuvattu yhtenäisillä viivoilla. Kuviossa 1 nämä viivat kulkevat 1) korkealla putoamisen ja irtoamisen välillä, 2) korkealta putoamisen ja vertikaalisen asennon menettämisen välillä sekä 3) vertikaalisen asennon menettämisen ja luhistumisen (rakenteen menettämisen) välillä. Tämä kuvausmalli perustuu näiden kehysten kuvaamiseen samoilla



Kuvio 1. Putoamisen semanttinen kartta (Reznikova ym. 2020: 29). Käännös venäjän kielestä kirjoittajan.

dominanttiverbeillä eri kielissä. Sen sijaan suhde korkealta putoamisen ja luhistumisen välillä on kuvattu katkoviivalla, koska putoamisverbiprojektin kieliotoksen perusteella nimenomaisesti luhistumisen kehys ei ole kuvattavissa korkealta putoamisen dominanttiverbillä, vaan suhde menee pikemminkin vertikaalisen asennon menettämisen kehyksen kautta.

Tällaista karttakuvausta voi hyödyntää yksittäisten kielten verbien kuvauksessa siten, että kunkin kuvattavan verbin käyttömahdollisuudet lisätään kuvaan omanvärisellään tai omanlaisellaan viivalla, jonka sisällä on kaikki ne käytöt, jotka verbillä voi olla. Näin ollen em. käyttömahdollisuuksia kuvaava viiva englannin *fall*-verbille ympäröisi kaikki neljä peruskehystä.

Putoamisverbien konkreettisten käyttöjen lisäksi putoamisverbiprojekti oli kiinnostunut myös samojen verbien metaforisista käytöistä. Tässä tarkastelussa se viittaa käsitemetaforateorian klassikkoon *Metaphors we live by* (Lakoff & Johnson 1980) ja käyttää sen kuvausmallia, jossa kielenkäytön oletetaan heijastelevan ihmisen käsitteistyksen luomia suhteita konkreettisten ja abstraktien entiteettien välille. Putoamisen tapauksessa keskeinen käsitemetafora on Lakoffin ja Johnsoninkin mainitsema HYVÄ / ENEMMÄN ON YLHÄÄLLÄ, HUONO / VÄHEMMÄN ON ALHAALLA. MLexT-ryhmän kuvausmalli ei ota huomioon käsitemetafora- ja metonymiateorian lukuisia myöhempiä tarkennuksia ja vaihtoehtoisia malleja.

3 Aiempi tutkimus suomen putoamisverbeistä

Suomen kielen vapaan liikkeen verbeistä on hyvin vähän tutkimusta, putoamis- ja kaatumisverbeistä erikseen ei juuri lainkaan. Tippumisen ja putoamisen erot ovat kyllä saaneet osakseen kielenhuollollista huomiota (Palander 2006) ja sanakirjat (mm. *Kielitoimiston sanakirja* vuoden 2018 versioon saakka) eivät ole suositelleet tippua-verbin käyttöä kerrallisesta putoamisesta. Samanaikaisesti aiemman artikkelimme viimeisen version kanssa valmistui meistä riippumatta myös verbien *pudota*, *tippua* ja *tipahtaa* polysemiaa käsittelevä pro gradu -tutkielma (Jeshoi 2019). Saimme iloksemme havaita, että tutkielman tulokset olivat samansuuntaisia omien tulostemme kanssa. Lähtökohtaerona on kuitenkin verbien tarkastelun suunta – Anniina Jeshoin tutkimus lähtee yksittäisten verbien eri merkitysten yhteyksistä, mitä meidän tutkimuksemme ei käsittele lainkaan.

Mielenkiintoni kohteena olevista verbeistä frekventeimmät ovat toki olleet osana laajempaa suomen verbitutkimusta. Ne on luokiteltu verbirektionsa puolesta dynaamisten liikeverbien luokkaan (Pajunen 1999: 55–56). Jos subjektin tarkoite on eloton, näiden liikeverbien kuvaama liike on kontrolloimatonta (Pajunen 1999: 55–56). Niiden liikkeen lähtöpistettä voi kuvata elatiivilla, joten ne ovat lähtemis- tai irrottamisverbejä (Pajunen 1999: 164–165) ja ne kohdistuvat tiettyyn suuntaan, joten ne ovat suuntaisia liikeverbejä (Pajunen 1999: 158). Semanttisesti niitä on luokiteltu mm. kaatumisverbeiksi ja lysähtämis- ja mätkähtämisverbeiksi (Pajunen 2001: 215–217). Liikkeen suuntautuneisuuden ilmaiseminen sijamuodoilla ja adpositioilla (mm. Kittilä 2014) liittyy myös tutkimieni verbien käyttöön, mutta aiempi tutkimus ei ole osoittanut alaspäin suuntautuvan liikkeen verbien eroavan tässä muista verbeistä. Muutenkaan näiden verbien ei ole aiemmin esitetty muodostavan suomessa erillistä ryhmää. Onkin täysin mahdollista, että tällainen ryhmä on olemassa leksikaalisen typologian näkökulmasta, mutta suomen kuvaus yksittäisenä kielenä ei sitä tue.

Suunnan ilmaisemisen lisäksi liikkeen tavasta on tehty tutkimusta fennistiikassa tilagrammien näkökulmasta. Tällöin keskeistä on ollut analysoida erilaisten viittauskehysten ja niiden sisältämien näkökulmien vaikutusta tilagrammien tulkintaan (mm. Huumo 2013; 2019; Ojutkangas 2005; Teeri-Niknammoghadam 2021). Putoamisverbiprojektin kiinnostuksen kohteina olevilla verbeillä kuvataan kuitenkin liikettä, jonka alkupisteessä putoavan tarkoitteen korkein kohta on vertikaalisella akselilla korkeammalla kuin sen loppupisteessä. Viittauskehys on täten aina maan vetovoimaan perustuvana absoluuttinen (vrt. Levinson 2003). Itse käsitteistys ylhäältä alas suuntautuvana

liikkeenä pysyy, vaikka verbin valintaan voi potentiaalisesti vaikuttaa lentoradan muoto ja putoamisen (kaatumisen, luhistumisen, irtoamisen) subjektin tarkoite voi vaihtaa sen aikana monellakin tavalla orientaatiotaan. Putoamisen yksityiskohtia voidaan eksplikoida, mutta sekään ei muuta vertikaalista akselia – ylhäällä oleva pysyy korkeammalla maan pinnasta kuin alempana oleva. Absoluuttinen spatiaalinen viittauskehys yhdistää siis tätä semanttista kenttää riippumatta siitä, onko liikkeen havainnoija itse ylempänä tai alempana, mikä hänen asentonsa on ja mihin putoamisprosessin osaan hän suuntaa mielenkiintonsa.

Aiemmassa artikkelissamme (Viimaranta & Protassova 2020) dokumentoitu suomen kielen tutkimus tehtiin seuraavalla tavalla: Kyselylomakkeen, stimulusvideoiden ja sanakirjojen avulla teimme luettelon verbeistä, jotka liittyvät suomessa putoamisen semanttiseen kenttään. Kyselylomake ja videot olivat siis meidän kahden tutkijan käytössä, joista minä olen suomen äidinkielinen puhuja ja toinen erittäin edistynyt kielenoppija. Emme keränneet aineistoja muilta suomen puhujilta. Kun tutkittavien verbien luettelo oli valmis, niiden käyttömahdollisuuksien analyysi tehtiin korpusaineiston perusteella. Verbeillä *alentua*, *aleta*, *kaatua*, *kellahtaa*, *kellistyä*, *kierähtää*, *kolahtaa*, *kompastua*, *kopsahtaa*, *kupsahtaa*, *kuukahtaa*, *langeta*, *laskea*, *lätsähtää*, *leijailla*, *lentää*, *loiskahtaa*, *luhistua*, *lysähtää*, *lyyhistyä*, *molskahtaa*, *painua*, *plätsähtää*, *pudota*, *roiskahtaa*, *romahtaa*, *sortua*, *tippua*, *tömähtää*, *tuupertua*, *vajota*, *valua*, *varista* ja *vierähtää* tehtiin haut Kielipankin aineistoissa (e-thesis-fi, FinnTreeBank 2, FinnTreeBank 3, KlK-fi, lehdet90off-v2, Suomi24-Korp-2016H2, Suomi24-Korp-2017H2).[4] 34 verbilekseemin lista ei luonnollisesti voi olla tyhjentävä etenkään murre-eroille alttiiden deskriptiivisten verbien osalta. Saatuja hakutuloksia analysoitiin niiden edustamien kehysten ja käyttömahdollisuuksien näkökulmasta, kunnes verbin esiintymissä ei enää tullut esille uudenlaisia käyttöjä.

Taulukossa 1 esitetään aiemman tutkimuksen (Viimaranta & Protassova 2020) mukaisesti ne kehykset ja niihin liittyvät metaforiset merkitykset, joissa tutkittuja suomen putoamisverbejä voi käyttää. Mukaan on otettu nimenomaisesti putoamisverbiprojektin käsittelemät tapaukset, ei verbien muita merkityksiä. Taulukon tiedoista voi huomata, että dominanttiverbeillä *pudota*

[4] Nämä aineistot valikoituivat käytännön syistä. Koska Korp-käyttöliittymä toimi hakuhetkellä (kevät 2019) erittäin hitaasti, oli tulosten saamiseksi käytettävissä olevia korpuksia rajoitettava. Tämä yhdistelmä osoittautui kokeiluissa sellaiseksi, jolla saatiin huomattava määrä erilaisia hakutuloksia ilman käyttöliittymän jumittumista. Lisäksi tämä valikoima oli käytössä siksi, että se sisälsi erityyppisiä tekstejä eri aikakausilta.

Taulukko 1. Suomen putoamis- ja kaatumisverbejä ja niiden käyttöyhteydet putoamisen kehyksissä

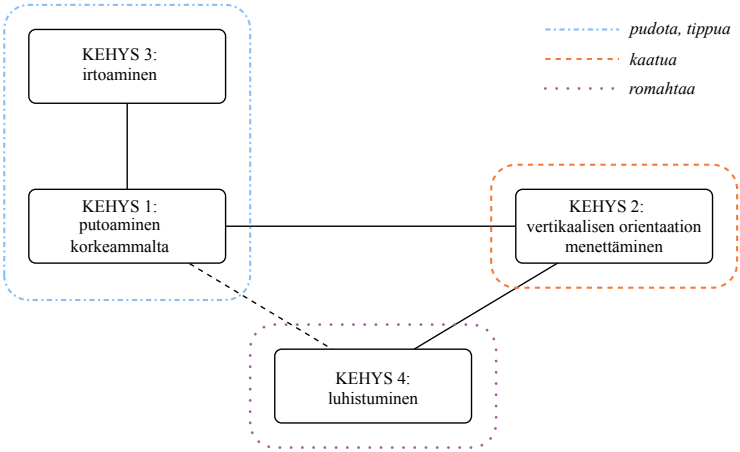
	KEHYKSET				METAFORISET MERKITYKSET						
VERBI	1. putoaminen korkeammalta tasolta	2. vertikaalisen orientaation menettäminen	3. irtoaminen	4. luhistuminen	pinnan aleneminen putoamisena	vähemmän on alhaalla	huonompi taso on alhaalla	toimintakyvyn menettäminen	irtoaminen menettämisenä	kokija tilaan	lopputilalla havainnoija
alentua	×				×	×	×				
aleta	×				×	×	×				
kaatua		×					×	×			
kellahtaa		×						×			
kellistyä		×						×			
kierähtää	×										
kolahtaa	×	×									
kompastua		×						×			
kopsahtaa	×	×									
kupsahtaa		×						×			
kuukahtaa		×						×			
langeta	×	×					×			×	×
laskea					×	×	×				
lätsähtää	×	×			×	×	×				
leijailla	×										
lentää	×										×
loiskahtaa	×	×									
luhistua				×	×		×			×	
lysähtää				×	×		×			×	
lyyhistyä				×	×		×			×	
molskahtaa	×	×									
painua					×						
plätsähtää	×	×									
pudota	×		×			×			×	×	×
roiskahtaa	×	×									
romahtaa				×	×	×	×	×		×	
sortua				×	×			×		×	
tippua	×		×			×	×	×	×	×	×
tömähtää	×	×									
tuupertua		×						×			
vajota	×				×					×	
valua	×					×				×	
varista	×										
vierähtää	×									×	

ja *tippua* on eniten mahdollisia putoamiseen liittyviä käyttöjä. *Tippua*-verbillä on taulukon tietojen mukaan niitä vielä enemmän kuin *pudota*-verbillä, sitä käytetään myös toiminnan loppumisen ja huononemisen kuvainnollisissa merkityksissä, joita *pudota*-verbin aineistoissa ei ollut.

Suomen putoamisverbeistä ehkä erikoisin on *langeta*. Sillä on sekä ylhäältä alas kohdistuvan putoamisen että vertikaalisen asennon menettämisen merkitykset. Myös sen metaforiset merkitykset ovat monipuolisia. *Langeta*-verbin lisäksi sellaisilla verbeillä, joiden merkitys keskittyy putoamisen loppupisteen kuvailuun, on sekä horisontaalisen että vertikaalisen putoamisen merkitys. Voi ajatella, että ne eivät kuvaa itse putoamista ollenkaan, mutta putoaminen on mukana verbin metonyymisessä käytössä. Tällaisia verbejä taulukossa ovat *kolahtaa*, *kopsahtaa*, *lätsähtää*, *loiskahtaa*, *molskahtaa*, *plätsähtää*, *roiskahtaa* ja *tömähtää*.[5]

Putoamisen metaforissa liikettä alaspäin ei tapahdu konkreettisesti, vaan liikkeellä alaspäin käsitteistetään jotain abstraktia asiantilaa. Putoamisverbiprojektissa metaforisten merkitysten tutkimus perustuu tiedoille 20 eri kielestä. Keskeinen tulos on, että verbien lukuisat metaforiset käytöt eri kielissä vahvistavat neljän peruskehyksen olemassaolon – samantyyppiset metaforiset merkitykset esiintyvät aina saman kehyksen sisällä. Kuvausmallissa keskeisessä asemassa ovat konventionaalistuneet metaforat, sillä yksittäisillä okkasionaalisilla käytöillä ei ajatella olevan samanlaista todistusarvoa. (Rakhilina ym. 2020.) Metaforisille käytöille on yhteistä käsitteistys Hyvä / Enemmän on Ylhäällä, Huono / Vähemmän on Alhaalla. Alhaalla oleminen voi edustaa putoajan näkökulman (esim. *hän putosi kärryiltä*) lisäksi myös lopputilan havainnoijan näkökulmaa (esim. *hänelle lankesi vastuu isästä*).

[5] Verbejä on toki lukuisia muitakin, ja niiden ekspressiivisyys ja äännesymboliset ulottuvuudet ovat erillisen tutkimuksen aihe.



Kuvio 2. Putoamisen semanttinen kartta: suomi

4 Analyysi suomen putoamisverbeistä muihin kieliin verrattuna

Tulkintani mukaan suomen dominanttiverbit putoamisen semanttisella kentällä ovat siis *pudota*, *tippua*, *kaatua* ja *romahtaa*. Semanttisena karttana ne voidaan kuvata kuvion 2 muodossa. Siitä näkee, että lähisynonyymeja *pudota* ja *tippua* käytetään korkealta putoamisen ja irtoamisen kehyksissä, kun taas vertikaalisen asennon menettämisen dominanttiverbi on *kaatua* ja luhistumisen (rakenteen menettämisen) dominanttiverbi *romahtaa*. Esittelen seuraavaksi putoamisen kehykset yksi kerrallaan vertaillen suomea muihin kieliin. Tiedot muista kielistä on vertailua varten saatu putoamisverbiprojektin julkaistuista tuloksista. Esittelen myös alustavia huomioita siitä, miten typologisessa tutkimuksessa havaitut putoamiseen liittyvät metaforiset käsitteistykset esiintyvät suomen kielessä.

4.1 Putoaminen korkeammalta tasolta alemmalle

Kontrolloimaton putoaminen korkeammalta tasolta alemmalle on putoamisverbiprojektin tulosten mukaan kaiken putoamisen ydin ja muilla putoamisen tavoilla on yhteys siihen. Suomessa käytetään tässä merkityksessä dominanttiverbejä *pudota* ja *tippua*. Muiden, merkitykseltään suppeampien verbien va-

linnassa vaikuttavat putoavan subjektin tarkoitteen ominaisuudet, putoamisen tapa ja lisäksi putoamisen lähtöpiste ja loppupiste sekä näiden suhde toisiinsa.

4.1.1 Subjektin vaikutus verbin valintaan

Monissa kielissä putoajan elollisuus tai elottomuus vaikuttaa putoamisverbin valintaan. Elollisista subjektin tarkoitteista voi erikseen olla merkitystä sillä, onko kyseessä ihminen vai muu elollinen (eli liikuntakykyinen) olento. Suomessa putoajan elollisuus ei vaikuta dominanttiverbien *pudota* ja *tippua* käyttöön. Molempia verbejä voidaan käyttää sekä elottomien että elollisten putoajien kanssa. Ylhäältä alas putoamisen dominanttiverbit sopivat suomessa kuvaamaan kaikenlaisten ja kaikenkokoisten tarkoitteiden tahdotonta liikettä ylhäältä alas.

Vaikka ihmisyyteen liittyvä tahdonalaisuus ominaisuutena ei välttämättä suoraan liity suomen putoamisverbin valintaan, on suomessakin myös tahdonalaisuutta ja tarkoituksellisuutta ilmaisevia putoamisverbejä, kuten kausatiivinen ja refleksiivinen *pudottautua*.

Ihmisen liikkeestä ylemmältä tasolta alemmalle käytetään verbejä *pudota* ja *tippua* riippumatta siitä, onko alkuasento ollut seisova, istuva vai makaava (esim. 1–3).[6]

(1) Rakennus eteni, mutta eräänä päivänä kyseinen mies putosi telineiltä. [KLK-fi]$_1$

(2) Tytär putosi illalla hevosen selästä. [lehdet90ff-v2]$_4$

(3) Hän tippui sängystä ja iski suunsa suoraan leluun. [KLK-fi]$_6$

Kuitenkin jos putoamisliike sisältää liikkeen olion oman akselin ympäri, niin ihmisistä kuin esineistäkin voidaan käyttää horisontaalisesta asennosta horisontaaliseen asentoon pudotessa verbejä *vierähtää* ja *kierähtää* (4, 5). Suomessa vertikaalista putoamisliikettä ei siis tarvitse ilmaista näissä tapauksissa erikseen.

(4) Ilman aivotoimintaa vierähdin sängystä lattialle kuin sementtisäkki. [KLK-fi]$_{36}$

(5) Itselläni oli sängyn reunalle este ettei vauva voinut kierähtää lattialle. [Suomi24-Korp-2016H2]$_{37}$

[6] Alaindeksit esimerkin jälkeen osoittavat esimerkin numeroa aiemmin julkaistussa artikkelissa (Viimaranta & Protassova 2020). Se sisälsi 135 glossattua, korpuksesta peräisin olevaa muokkaamatonta esimerkkiä.

Subjektin tarkoitteen koko ja muoto voivat vaikuttaa niiden kanssa käytettäviin verbeihin. Esim. norjan verbiä *tumle* voi käyttää vain suurien ja painavien tarkoitteiden kanssa (Livanova & Mordashova 2020: 174). Suomessa dominanttiverbejä *pudota* ja *tippua* voi käyttää kaikenkokoisten putoajien kanssa. Sen sijaan osa suppeamman merkityksen verbeistä, mm. *leijailla* ja *valua*, ovat mahdottomia painavien ja tarkkarajaisten tarkoitteiden kanssa. Suomessa subjektin koon ja muodon vaikutuksen verbin valintaan näkee erityisesti onomatopoeettisen ja deskriptiivisanaston yhteydessä. Esim. pienikokoinen putoaja voi tuskin *tömähtää* maahan.

Myös putoajien määrä vaikuttaa joissain kielissä verbin valintaan. Joissain kielissä yksittäisen putoavan olion ja useiden putoavien olioiden kanssa käytetään systemaattisesti eri verbejä. Näin on tilanne agulissa, jossa yksittäisen liikkuvan tarkoitteen kohdalla käytetään liikettä merkitsevää lekseemiä *-arxas*, joka on mahdoton silloin kuin tarkoitteita on useita (Reznikova ym. 2020: 37). Suomessa putoajien määrä ei vaikuta verbin valintaan, vaikka lauseen tulkintaan se voi toki joissain tapauksissa vaikuttaa, vertaa esim. *puu varisee* ja *lehdet varisevat*, mutta **lehti varisee*. Koska *varisee* merkitsee useiden lehtien toimintaa, tulkitaan yksikössä oleva sana sen kanssa joko metonyymisesti (puu sen osina) tai aineena, esim. *puu*[*hake*] *varisee*.

Siinä missä muiden subjektien tarkoitteiden kanssa käytetään yleisesti samoja dominanttiverbejä, putoamisverbiprojekti on löytänyt kolme laajalti eri kielissä esiintyvää erityistapausta, jotka vaativat säännönmukaisesti muun, merkitykseltään suppeamman verbin käyttöä. Tällaisessa erityisasemassa subjektien tarkoitteiden joukossa ovat sade, neste (tippoja ja vesimassoja lukuun ottamatta) ja ainemaiset tai monista pienistä osasista koostuvat oliot (esim. lehdet ja neulaset). Myös putoavan elottoman esineen koko voi vaikuttaa verbin valintaan. (Reznikova ym. 2020.)

Myös suomessa sateen (kuten myös lumen, rännän, rakeiden) putoamisesta käytetään usein erityistä lekseemiä *sataa*. Kuitenkin myös dominanttiverbejä *pudota* ja *tippua* voi käyttää silloin kun kuvataan nimenomaisesti yksittäisten pisaroiden päätymistä tiettyyn paikkaan eikä sadetta kokonaisuutena (esim. 6 ja 7).

(6) Sadehan mitataan tietyn läpimittaisessa putkessa, johon sade putoaa. [Suomi24-Korp-2016H2][25]

(7) Miten luonnollisesti sade tippuu räystäiltä! [Suomi24-Korp-2017H2][26]

Sen sijaan norona alas valuva vesi ei voi *pudota* tai *tippua*, ainoana poikkeuksena tässä on suuren putoavan vesimassan (esim. vesiputouksen) liike alaspäin, jossa niin suomessa kuin monissa muissakin kielissä voi käyttää ylhäältä alas putoamisen dominanttiverbiä, esim. turkin *düşmek*, pandžabin *ḍigṇā* tai nykykreikan *péfto* (Reznikova ym. 2020: 35). Sen sijaan otannan kielissä vaihtelee se, millaisiin muihin aineisiin tällaiset norona valumisen käytöt ulottuvat. Tyypillistä on, että nesteiden ja hienojakeisten aineiden kuten hiekan valumisesta voi käyttää samoja verbejä (esim. englannin *pour*, baskin *isuri*, serbin *sipati* (*se*)). (Reznikova ym. 2020: 35–36.) Näin on myös suomessa – nesteiden lisäksi mm. hiekka voi *valua*, toisin kuin poikkeukseksi osoittautuneessa venäjässä, jossa nesteiden ja hienojen aineiden kuten hiekan valumisesta käytetään systemaattisesti eri verbejä.[7]

Suomessa lehtien ja neulasten putoamisesta voi käyttää dominanttiverbejä *pudota* ja *tippua*, mutta etenkin korostettaessa putoamisen vähittäistä ja vaiheittaista luonnetta käytetään myös verbejä *varista* ja *karista* (8).

(8) Lehdet varisee puista syksyllä niin kuin ennenkin. [Suomi24-Korp-2016H2]$_{31}$

Ainesanoista nesteistä ja myös jauhemaisista aineista puhuttaessa suomen *valua*-verbiä voi käyttää sekä silloin kun aine valuu vapaana norona, että silloin kun se valuu jotain väylää pitkin (9), pintaa myöten, aineen läpi tai säiliön reunan yli. Ero tässä ei siis automaattisesti vaikuta verbin valintaan.

(9) Vesi valuu ammeen alareunan reikien kautta ulos. [KLK-fi]$_{34}$

Kielet eroavat siinä, miten suurista yksiköistä koostuvista aineista voi käyttää valua-tyyppisiä verbejä. Suomessa vapaassa pudotuksessa esim. kävyt eivät voi *valua*, kun taas jotain pintaa myöten mennessään ne tai isommatkin subjektit voivat. Sen sijaan komissa verbiä *kiśśyny* 'vuotaa, valua' voi käyttää nesteiden ja hiekan valumisen lisäksi myös esim. käpyjen vapaasta pudotuksesta, mutta ei lehtien varisemisesta (Kashkin 2020).

Aineiden liikkeen käsitteistykseen liittyy se niiden ominaispiirre, että itsessään aineella ei ole muotoa, mikä vaikuttaa niiden alaspäin suuntautuvaan liikkeeseen. Jauhemaisten ja nestemäisten aineiden liike voi valumisena muistuttaa huomattavasti toisiaan, mutta tippoja voivat muodostaa vain nesteet, kun

[7] Suomessakin kausaatioverbit erottelevat tarkemmin sitä, mitä kaadetaan ja myös kaatamisen tapaa – esim. jauhoja voi *kaataa*, *sirotella* tai *varistella*, mutta ei ehkä *valuttaa*, nestettä *kaataa*, *valuttaa*, *tiputella* – muttei *sirotella* tai *varistella*.

taas sirottelu onnistuu vain jauhe- tai kidemäisillä aineilla. Silloin kun aineet ovat jossain säiliössä, esim. pakkauksessa, säiliön tai pakkauksen muoto kuten myös se, miten tiuhaan aine on pakattu vaikuttaa siihen, miten liike alaspäin käsitteistetään. Toisaalta esim. mehua sisältävä avaamaton kartonkipakkaus on käsitettävissä tiiliskiven tapaisena esineenä, joka voi sellaisena myös pudota ja kaatua. Avattuun tai rikkoutuneeseen kartonkipakkaukseen on puolestaan muodostettu tai muodostunut väylä, josta aine voi päästä ulos. Väylän muoto ja syntymistapa voi vaikuttaa verbin valintaan. Suomessa aine voi *valua* niin kanonisten väylien kuten kaatoaukkojen kautta, syntyneistä rei'istä kuin esim. reunojen yli tulvienkin.

4.1.2 Putoamisen alku- ja loppupisteen ja putoamisen tavan vaikutus verbin valintaan

Merkittävä verbin valintaan liittyvä seikka on, mitä oliolle tapahtuu sen putoamisen loppupisteessä, esim. meneekö se pinnan läpi kokonaan tai osittain. Suomessa näissä tapauksissa voi käyttää dominanttiverbiä (10, 11) toisin kuin esim. venäjässä, jossa *padat'* ei sovi näihin konteksteihin.

(10) Hän on pudonnut jäihin kahdesti. [KLK-fi]$_7$

(11) Kuulin nimittäin juttua että olisi tippunut katon läpi. [Suomi24-Korp-2016H2]$_{10}$

Toki suomessa on myös muita alaspäin suuntautuvasta liikkeestä käytettäviä verbejä kuten *vajota* ja *upota*. Nämä verbit vaativat jonkin aineen, jonka sisällä liike alaspäin toteutuu.

Myöskään muunlainen lentorata kuin suoraan vertikaalisesti ylhäältä alas menevä ei estä dominanttiverbien käyttöä (12). Kuitenkin näissä muunlaisen lentoradan tapauksissa käytetään usein *lentää*-verbiä (13).

(12) Kun sitten pallo tippuu maahan niin näet että se vierii juuri niin kuin oletit. [Suomi24-Korp-2016H2]$_{13}$

(13) Jopa pallo lensi heidän puolelleen. [Suomi24-Korp-2016H2]$_{38}$

Aineistoni perusteella ei ole selvää, miten lentoradan muoto aivan tarkalleen vaikuttaa suomessa verbin valintaan. Lentoradan muodon lisäksi keskeinen verbin valinnassa lienee myös putoamisen loppupiste.

Ilmavirtojen mukana tapahtuvaa liikettä kuvaava *leijailla*-verbi voi suomessa kuvata putoamista silloin, kun putoamisen tapaa korostetaan ja subjektin tarkoite on lumihiutaleiden lisäksi muut kevyet esineet kuten terälehdet, puiden lehdet tai höyhenet. Varsinaisesti putoamisverbeihin sen voinee laskea vain silloin, kun annetaan joko lähtö- tai tulopiste (14). Alas kohdistuva liike voidaan myös eksplikoida adverbilla *alas* (15).

(14) Maahan leijailevat terälehdet symboloivatkin elämän hetkellisyyttä. [lehdet90ff-v2]$_{40}$

(15) Hetken kuluttua höyhen leijaili alas taivaalta. [Suomi24-Korp-2017H2]$_{42}$

Projektin tietojen mukaan[8] on harvinaista, että on olemassa suomen *leijailla*-verbin tapainen verbi, jolla on muun liikettä tarkoittavan merkityksensä ohessa spesifi putoamismerkitys (kontrolloimaton liike ylhäältä alaspäin ilmavirtojen mukaan), jossa putoamisen loppupisteen voisi spesifioida esim. lisäämällä esimerkin 15 loppuun sanan *maahan*.

4.1.3 Metaforisia käyttöjä

4.1.3.1 VÄHEMMÄN ON ALHAALLA

Metafora VÄHEMMÄN ON ALHAALLA ilmenee numeraalisella asteikolla operoivien määreiden kanssa (lämpötila, paine, hinnat, valuuttakurssit, väkiluvut) ja myös esim. kiinnostuksen, halun tai innostuksen intensiteetin vähenemisen ilmaisuissa. On typologisesti tyypillistä, että näissä voi käyttää putoamisverbejä kuten venäjän *padat'*, englannin *fall*, norjan *falle*, ranskan *tomber* ja *chuter*, hindin *girnā*, pandžabin *ḍignā*, gudžaratin *paṛvuⁿ*, turkin *düşmek*, tigrinjan *wädäḳä*, japanin *ochiru*, tamilin *viẓu* ja korean *tteleci-ta* (Rakhilina ym. 2020: 71). Myös suomessa nämä kaikki voivat *pudota* tai *tippua*, vaikka muu verbi onkin monessa tapauksessa tyypillisempi.

Lisäksi ei-numeerisesta laadun huonontumisesta ja kokonaismäärien vähenemisestä käytetään ylhäältä alas putoamisen verbejä monissa kielissä (Rakhilina ym. 2020), myös suomessa. Lauseessa voi olla esim. sana *määrä* (16) tai

[8] Rakhilina, Ekaterina 2019, henkilökohtainen kommunikaatio.

summa, mutta se ei ole välttämätön sellaisten substantiivien kanssa, jotka sisältävät lähtökohtaisesti kvantifioinnin. Täten esimerkissä (17) ei ole epäselvää, että tarkoitetaan palkkion määrää.

(16) Kansanopistojen määrä putoaa yli kolmanneksella. [e-thesis-fi]$_{133}$

(17) Saatu palkkio tippuu. [lehdet90ff-v2]$_{132}$

Tietyn metaforisen tason alenemisesta käytetään dominanttiputoamisverbien lisäksi myös verbejä *laskea*, *alentua*, *aleta*, joissa metafora liittyy aiemmin saavutetun metaforisen tason alenemiseen (esim. 18, 19, 20). Esimerkeissä voisi näiden verbien lisäksi käyttää dominanttiverbejä *tippua* ja *pudota*.

(18) Ensi vuonna kokonaistuotanto laskee yhden prosentin. [Suomi24-Korp-2017H2]$_{134}$

(19) Opetuksen sekä varsinkin tutkimuksen taso laskee. [lehdet90ff-v2]$_{135}$

(20) Luonnollisen koron taso alentunut euroalueella [lehdet90ff-v2]$_{137}$

Aleneminen putoamisena on projektin tulkinnan mukaan metaforista myös silloin kuin puhutaan konkreettisten pintojen tason laskusta. Tämä johtuu siitä, että tällaisessa tilanteessa aineen määrä jossain säiliössä vähenee, ja sen takia tason pinta liikkuu alaspäin, mutta ei ole mitään oliota, joka konkreettisesti liikkuisi alaspäin. Pinnan tason konkreettista menemistä alaspäin käsitteistetään siten metaforisena putoamisena, ja tämä jo itsessään metaforinen käsitteistys johtaa toiseen metaforiseen käsitteistykseen, jossa aleneminen on vähenemistä.

4.1.3.2 Tilaan päätyvän kokijan metafora

Tietyn fysiologisen, emotionaalisen tai psykologisen tilan alkamisen metaforisessa merkityksessä suomessa käytetään verbejä *pudota*, *tippua*, *vajota*, *langeta*. Näitä verbejä käytettäessä alkavaan tilaan päätyminen käsitteistetään ihmisen tahdosta riippumattona, ihminen on kokijan asemassa (21, 22).

(21) Haluan kuitenkin sanoa, että ei meidän pidä vajota pessimismiin Euroopan unionin kyvyttömyydestä vahvistaa yhteistä kantaa. [FinnTreeBank 3: EuroParl]$_{55}$

(22) Ei Jaanan vielä olisi tarvis epätoivoon langeta, hän sanoi. [lehdet90ff-v2]$_{94}$

Metaforinen putoaminen ei välttämättä ole negatiiviseen tilaan päätymistä, mistä todistaa mm. englannin *fall in love*. Putoamisverbiprojektin tietojen mukaan tämä sanontatapa on levinnyt käännöslainana mm. turkkiin ja japaniin.

4.1.3.3 Lopputilan havainnoijan metaforat

Lopputilan havainnoijan metaforissa putoamisen subjektin tarkoitteena on ääni, valo tai varjo, aika, epämiellyttävät tapahtumat, valintamahdollisuuden puuttuminen ja ihmisten vuorovaikutus (Rakhilina ym. 2020: 80–94). Ihminen on näissä käytöissä havainnoijan roolissa. Suomessa tämä esiintyy valon ja varjon lankeamisen, epämiellyttävien tapahtuminen ja valintamahdollisuuksien puuttumisen tapauksissa. Ihmisten vuorovaikutus tarkoittaisi tapauksia, joissa kommunikaatiotilanne selitetään sillä, että puhekumppani on ”tippunut” (esim. ranskan *tomber sur quelqu'un*), suom. ehkä esim. *törmätä jhnkin*. Jossain kielissä on metaforisia ilmauksia, joissa ääni kuten linnunlaulu voi ”pudota korviin” (gudžarati, hindi-urdu, pandžabi). Typologisen metafora-artikkelin kirjoittajat ovat nähneet tässä yhteyden sen kanssa, miten suomessa sanat voivat putoilla suusta, kun tarkoitetaan katkonaista puhevirtaa (23) (Rakhilina ym. 2020: 81). Myös tiedon saapuminen vähä vähältä on esimerkki lopputilan havainnoijan metaforasta (24).

(23) Sanat putoilevat välillä harvakseltaan. [lehdet90ff-v2]$_{54}$

(24) Tietoa putoilee ajoittain, se vaihtelee päivittäin ja osa on epävirallista. [lehdet90ff-v2]$_{53}$

Valon ja varjon kohdistumisesta voidaan käyttää metaforisesti putoamisverbejä yhtä lailla monissa indoeurooppalaisissa kielissä (mm. englanti, saksa, ranska) kuin esim. suomessa, adygessa ja gudžaratissakin (Rakhilina ym. 2020). Suomessa valo ja varjo voi tyypillisesti *langeta*, joissain tapauksissa myös *pudota*, ei kuitenkaan *tippua* (ellei tarkoiteta konkreettista valonlähdettä kuten lamppua tai valon määrän vähenemistä).

Tämä on yksi käyttöeroista nykysuomen *putoamisen* ja *tippumisen* välillä. Runollisen sävyinen *langeta* sopii tällaisiin konteksteihin hyvin, neutraali *pudota* on myös joskus mahdollinen, mutta *tippua* ei toimi. Esimerkin (25) tapaisissa konteksteissa *tippua* tarkoittaisi varmasti ja *pudota* todennäköisesti valaisimen putoamista.

(25) Nyt valo lankeaa lantakäytävälle ja ruokintapöydälle. [lehdet90ff-v2]$_{90}$

Vuoden- ja vuorokaudenaikojen vaihtumisesta ja sääilmiöiden alkamisesta käytetään monissa kielissä putoamisverbejä. Suomessa esim. syksy ei voi ”pudota” (kuten esim. ruotsissa) tai myrsky ”tippua” (kuten mm. englannissa), mutta *langeta*-verbiä käytetään joskus tällaisissa merkityksissä (*ilta lankeaa*). Lisäksi yleinen metafora, jossa epätoivottujen tapahtumien katsotaan putoavan ihmisen päälle, esiintyy myös suomessa. Tässä merkityksessä käytetään useita eri putoamisverbejä. Myös tapauksissa, joissa ihmisen tai ihmisen suorittaman toiminnan osaksi tulee jokin rooli tai tehtävä, käytetään usein putoamisverbejä. Suomessa tämä verbi on useimmiten *langeta* (26, 27).

(26) Tämän tavoitteen saavuttamisessa keskeinen rooli lankeaa tutkimukselle. [lehdet90ff-v2]$_{91}$

(27) Sielunhoitajan tehtävät lankesivat usein etenkin vanhemmille naisille. [lehdet90ff-v2]$_{92}$

Kaikkinensa suomessa esiintyvät lähes kaikki ne ylhäältä alaspäin suuntautuvan liikkeen metaforiset merkitykset, jotka projekti on löytänyt 20 kielessä. Osassa näistä ei kuitenkaan voi tai ole tyypillistä käyttää dominanttiverbejä *pudota* ja *tippua*.

4.2 Vertikaalisen asennon menettäminen

4.2.1 Vertikaalisen asennon menettämisen konkreettiset käytöt

Kuten todettua, suomi kuuluu niihin melko harvoihin kieliin, jotka tekevät systemaattisesti eron ylhäältä alaspäin putoamisen ja vertikaalisen asennon menettämisen (kaatumisen) välillä. Toisin kuin joissain muissa kielissä, suomessa kaatumisen syy (sisäinen vai ulkoinen) sekä kaatuvan subjektin tarkoitteen ihmisyys tai elollisuus eivät vaikuta kaatumisverbin valintaan.

Ihmisen kaatumisen erityispiirteitä selittää ihmisen etu–taka-akseli. Joissain kielissä (mm. nenetsissä ja koreassa) on erillisiä lekseemejä, joilla kuvataan ihmisen kaatumista tietyn ruumiinosan törmätessä maahan (Reznikova ym. 2020: 41), suomessa tämä ilmaistaan esim. adverbeilla *kyljelleen*, *vatsalleen*, *takaraivolleen* jne.

Joissain kielissä myös muut elolliset olennot kuin ihminen ovat erityisasemassa. Näin on mm. tundranenetsissä, jossa elottomien esineiden kaatumisesta käytetään verbiä *xăwă-* ja elollisten verbiä *mo”na-* (Kaškin ym. 2015: 42). Samoin japanissa verbiä *korobu* voi käyttää vain itsenäisesti liikkuvista elollisista subjekteista. Koreassa puolestaan verbiä *ssuleci-ta* käytetään subjektin

tarkoitteen sisäisistä syistä (esim. heikko terveys) tapahtuvasta kaatumisesta ja *nemeci-ta* silloin kun syy on ulkoinen, esim. hyökkäys. (Reznikova ym. 2020: 41.) Suomessa verbi *kaatua* kävisi kaikkiin näihin tapauksiin.

Elottomista olioista erityisasemassa ovat typologisesti yhtäältä puut, tolpat, aidat ja seinät ja toisaalta autot, junat ja pöydät. Tämä johtuu putoamisverbiprojektin tietojen mukaan siitä, että prototyyppisesti kaatuvia ovat sellaiset esineet, jotka joko ovat kapeita tai niillä on litteä puoli ja täten kaatumisen jälkeen kaikki osat ovat lähellä maan pintaa. Suomessa kaatua voi siis ihminen (28), neljällä jalalla seisova eläin kuten koira (29) ja vertikaalisena hahmotettava esine kuten hylly (30) tai puu (31).

(28) Uskollinen juoksija kaatui maaliviivalle. [KLK-fi]$_{60}$

(29) Minä tiesin sen siitä, että äkkiä koira kaatui ja alkoi tempoilla lumessa. [KLK-fi]$_{61}$

(30) Älä nojaa, hylly voi kaatua. [FinnTreeBank 2]$_{63}$

(31) Luulin ensin, että puu on kaatunut talon päälle. [KLK-fi]$_{65}$

Suomen lekseemi *kaatua* vaikuttaa siis olevan käytöltään hyvin samanlainen kuin monien muiden kielten vertikaalisen aseman menettämistä kuvaavat dominanttilekseemit.

Autojen ja pöytien kaltaisten esineiden tapauksessa vertikaalisen asennon menettäminen johtaa siihen, että osa esineestä (esim. auton päällepäin jäänyt kylki) on edelleen kaukana maasta (32).

(32) Tämä auto tiettävästi kaatui testissä. [FinnTreeBank 3: EuroParl]$_{64}$

Suomessa autot ja pöydätkin voivat kaatua, mutta joissain muissa kielissä on käytettävä toista lekseemiä. Esim. tigrinjassa dominanttiverbiä *wädäḳä* voi käyttää mm. puiden, tolppien ja polkupyörien, mutta ei autojen ja pöytien kaatumisesta (Bulakh 2020: 684–685).

Toinen kaatumisen erityistapaus on putoamisverbiprojektin mukaan pitkänomaiset, seisovat säiliöt (mm. pullot), joita varten kielessä on usein dominantista kaatumisverbistä poikkeava kumoon menemistä merkitsevä verbi, esim. venäjän *oprokinut'sja* 'kaatua, mennä kumoon/nurin'. Suomessa nesteiden päätymisestä säiliön ulkopuolelle käytetään kuitenkin *kaatua*-verbiä (33, 34).

(33) Lähes täysinäinen maitopurkki kaatui pöydälle ja kattilaan jossa oli keittoa. [Suomi24-Korp-2017H2]$_{66}$

(34) Kaiken muun lisäksi kahvi kaatui sänkyyn! [Suomi24-Korp-2017H2]$_{67}$

Jos koko säiliö sisältöineen putoaa, voidaan käyttää putoamisverbiä (35).

(35) Vähän kuin että maito putosi lattialle, oho… [Suomi24-Korp-2017H2]$_{69}$

Kaikkinensa suomen verbillä *kaatua* on vertikaalisen asennon menettämisen kehyksessä erittäin laaja käyttöpiiri. Seuraavassa tarkastelen vielä vertikaalisen asennon menettämisen metaforisia käyttöjä.

4.2.2 Vertikaalisen asennon menettämisen metaforiset käytöt

Vertikaalisen asennon menettämismetafora liittyy monissa kielissä toimintakykyisyyden menettämiseen. Mm. järjestelmät ja suunnitelmat voivat menettää toimintakykyisyytensä tai osoittautua mahdottomiksi (36, 37, 38) ja ihminen voi kaatua eli kuolla (sodassa) (39).[9] Samoin häviäjä *kaatuu* metaforisesti (40), ja ihminen voi *kaatua sänkyyn* kun flunssa vie hänen toimintakykynsä. Tämä metafora rakentuu vertailulle kaatumisen ja toimintakyvyn menettämisen välillä.

(36) Neuvostoliitto kaatui noin 50 vuodessa. [lehdet90ff-v2]$_{84}$

(37) Olkiluodon ydinvoimalahankkeesta tehty valitus kaatui oikeudessa. [lehdet90ff-v2]$_{87}$

(38) Usein avioliitto kompastuu isän, ei tytön itsensä, kieltoon. [lehdet90ff-v2]$_{88}$

(39) Dunckerin lisäksi Hömeforsissa kaatui 211 suomalaista soturia. [lehdet90ff-v2]$_{99}$

(40) Ruotsi kaatui selvästi numeroin 4-1. [lehdet90ff-v2]$_{95}$

Kaatua-verbin lisäksi suomessa käytetään tällaisessa merkityksessä mm. verbejä *kellahtaa*, *kellistyä* (41), *kompastua* (42), *kupsahtaa* ja *kuukahtaa*. Näiden eri verbien käyttöä hyvin samantyyppisissä merkityksissä edistänee niiden ekspressiivisyys.

[9] Vaikka *kaatua*-verbi on vanhastaan tulkittu tällaisessa merkityksessä käännöslainaksi (Hakulinen 1969: 50–51), on myös mahdollista, että tämä metaforiseksi tulkittu kuolemisen merkitys on sen alkuperäinen merkitys, josta konkreettinen kaatuminen on johdettu (Salminen 2012: 353; 2023: 386).

(41) Lohkovoittaja kellistyi puhtaasti. [lehdet90ff-v2]$_{98}$

(42) Suomen jääkiekkojoukkue kompastui vasta loppuottelussa. [KLK-fi]$_{96}$

Niissäkin kielissä, joissa ei lähtökohtaisesti eroteta toisistaan putoamista ylemmältä tasolta alemmalle ja kaatumista, toimintakykyyn yhdistettävän pystyasennon menettäminen selittää vastaavia metaforisia käyttöjä siinä missä tilaan vaipuminen oli samantyyppistä kuin suomessa. Esim. englannissa verbillä *fall in*(*to*), esim. *fall into depression* ilmaistaan masennustilaan vaipumista, mutta verbillä *fall apart* toimintakyvyn menettämistä (Vinogradova ym. 2020).

4.3 Irtoaminen

4.3.1 Konkreettinen irtoaminen

Putoamisverbiprojektin mukaan yhdessä putoamisen peruskehyksistä tapahtuu kaksi erillistä vaihetta, irtoaminen alkuperäisestä paikasta ja putoaminen ylhäältä alas. Yhdellä verbillä kuvatessa nämä vaiheet kuvataan yhdessä. Esimerkkeinä kyselylomakkeessa ja stimulusaineistossa on hatun lentäminen pois päästä, hampaiden ja hiusten irtoaminen, sormuksen luiskahtaminen pois sormesta, kännykän tippuminen käsilaukusta, linnunpoikasen putoaminen pesästä ja pyykkinarun pään irtoaminen naulasta, jossa se on kiinni. Suomessa dominanttiverbien *pudota* ja *tippua* käyttö on kaikissa näissä tapauksissa mahdollista. Tämä onkin projektin tietojen mukaan yleistä, ja monissa kielissä voi jättää irtoamisvaiheen huomiotta ja kuvata tilannetta putoamisverbillä. On kuitenkin myös kieliä (mm. khmer, adyge), joissa putoamisen dominanttiverbejä ei voi käyttää, jos putoamisen alkuvaiheessa on irtoaminen (Reznikova ym. 2020: 43).

Irtoamisen voi toki myös kuvata, ja myös suomessa irtoamista merkitseviä verbejä voi käyttää yksinään kuvaamaan koko tapahtumaa. Typologisesti onkin yleistä, että putoamis- ja irtoamisverbit kilpailevat tällaisissa käytöissä. Verbin valinnassa putoamista merkitsevän dominanttiverbin ja irtoamista merkitsevän verbin välillä voi olla tällöin ratkaisevaa se, miten kiinteästi irtoava asia on kiinni alkupisteessä. On esim. mahdollista, että niissä tapauksissa, joissa putoava esine ei ole kiinteästi kiinnitetty alkupisteeseen (esim. vaatteen putoaminen naulasta, sormuksen sormesta, esim. 43), käytetään yhtä verbiä ja niissä tapauksissa, joissa kiinnitys on (nappi takissa, pyykkinaru sidottuna kiinnikkeeseensä) toista. Esim. venäjässä dominanttiverbiä *padat'* ’pudota, kaatua’

voi käyttää löyhästi kiinnittyneistä tai vain yhteydessä olevista esineistä kuten pyyhkeen tippumisesta naulasta tai sormuksen sormesta, mutta se on mahdoton pyykkinaruista ja kiinnitetyistä napeista, joiden putoamisesta on käytettävä irtoamista merkitsevää verbiä.[10] Suomessa putoamisen dominanttiverbit ovat näissäkin tapauksissa mahdollisia.

Myös hiusten ja hampaiden irtoamisesta ja sormuksen sujahtamisesta pois sormesta voidaan käyttää dominanttiverbejä (43, 44). Putoamisverbiprojektin tietojen mukaan suomessa mahdollinen hiusten ja hampaiden putoamisesta käytettävä *lähteä*-verbi on typologisesti harvinainen.

(43) Se oli pudonnut murhaajan saippuoidusta sormesta. [KLK-fi]$_{21}$

(44) Informoin jos hampaat tippuu suusta. [Suomi24-Korp-2017H2]$_{23}$

Typologisen tiedon mukaan yhtenä verbin valintaan vaikuttavana kriteerinä voi olla se, onko kyse luonnostaan syntyneestä yhteydestä, johon ihminen ei ole vaikuttanut (esim. päässä kasvavat hiukset, lehdet puussa) vai ihmisen toiminnan vaikutuksesta syntyneestä yhteydestä (esim. pyykkinarun pään tai napin kiinnitys). Esim. tigrinjassa irtoavista (putoavista) hiuksista, lehdistä, karvoista, käärmeennahoista jne. voi käyttää erityistä lekseemiä *rägäfä*, mutta se on mahdoton, jos yhteys on syntynyt ihmisen toiminnan seurauksena (Bulakh 2020: 677).

Myös irtoamisen tapa voi vaikuttaa verbin valintaan, ja monissa kielissä kuten suomessa on käytössä esim. luiskahtamiseen viittaavia irtoamisverbejä, joiden kuvaamat tapahtumat voivat sisältää liikkeen alaspäin. Myös liikkeen tavan eksplikoiva verbi, esim. *lentää* saattaa kuvata sellaistakin putoamista, johon sisältyy alussa irtoaminen, esim. *nappi lensi housuista*.

Irtoamiskehyksessä verbin valintaan voivat siis vaikuttaa esineen kiinnittyneisyyden aste alkupisteeseen, irtoamisen tapa ja itse esineen tyyppi (Reznikova ym. 2020: 45). On lisäksi kielikohtaista, mitkä tapaukset käsitteistetään osana irtoamisen ja mitkä ylhäältä alas putoamisen kehystä. Tästä esimerkkinä on linnunpoikasen putoaminen pesästä ja kännykän tippuminen laukusta, jotka suomessa ovat molemmat putoamisen kehyksen osia, mutta joissain muissa kielissä ne käsitteistetään irtoamisena ja verbin valinta on tämän mukainen.

[10] Irrallisen napin putoamisesta esim. kädestä tai pöydältä voi kyllä käyttää *padat'*-verbiä.

4.3.2 Irtoaminen menettämisenä

Kiinnityksen tai yhteyden irtoamisen ja jostain pois putoamisen metaforisissa käytöissä on suomessa vahva kyydistä putoamisen kielikuva, jota edustaa mm. päivärahoilta pois putoaminen (45) ja kuvainnollinen kärryiltä putoaminen joko laajemmin (46) tai ymmärryksen tasolla (47). Myös aseman menettämistä esim. urheilukilpailuissa, uralla jne. voidaan kuvata näillä verbeillä (48). Vastaavissa käytöissä on myös mm. englannin *fall off* (Vinogradova ym. 2020).

(45) Mun toinen vanhempi on kanssa pudonnut päivärahoilta pois heti kättelyssä. [Suomi24 2017H2: 2012]$_{50}$

(46) Yhteiskunnan pyörästä olen pudonnut jo vuosia sitten. [e-thesis-fi]$_{49}$

(47) Siirrytään saksan kieleen, ja minä putoan kyydistä. [lehdet90ff-v2]$_{51}$

(48) Yks joukkue tippuu aina kerrallaan. [lehdet90ff-v2]$_{52}$

4.4 Luhistuminen

4.4.1 Konkreettinen luhistuminen

Luhistumisen kehyksessä on kyse alaspäin menevästä liikkeestä, jonka subjektin tarkoitteena ovat tietyt rajatut esineet: sillat, talot tai vesialtaan (joki, järvi, meri) reunat. Tällaisen subjektin liike alaspäin aiheuttaa muutoksen tarkoitteen muodossa. Sillan luhistumisesta voi useissa kielissä käyttää alaspäin suuntautuvan liikkeen verbiä, mutta talot eivät useinkaan voi ”pudota”. Tämä liittyy siihen, että talon muoto muuttuu luhistumisen seurauksena enemmän kuin sillan. Suomessa luhistumisen kehyksessä käytettäviä verbejä ovat mm. *sortua*, *luhistua*, *lysähtää* ja *romahtaa* (esim. 49, 50). Vesialtaan reunan pettämisen merkityksessä näistä verbeistä tyypillisin on *sortua*.

(49) Talo, joka sortuu, ei ole rakennettu näin. [lehdet90ff-v2]$_{112}$

(50) Ihminen on sitä kuoltuaan, rakennus luhistuttuaan, sademetsä kuivuttuaan autiomaaksi. [lehdet90ff-v2]$_{117}$

4.4.2 Toimintakyvyn menettäminen luhistumisena

Kaatumisen lisäksi toimintakyvyn menettämistä ilmaistaan monissa kielissä luhistumiskehyksen verbeillä. Näissä käytöissä toimintakyvyn menetys on usein totaalisempaa kuin kaatumiskehyksen verbejä käytettäessä.

Myös suomeksi mm. liiketoiminta voi romahtaa, tai esim. yhteiskuntajärjestelmä menettää perustansa (51).

(51) Vanha maailma romahti. [KLK-fi]$_{126}$

Luhistuminen voi merkitä myös nopeaa muutosta laajemmin. Siinä missä esim. valuuttakurssit voivat suomeksi neutraalimmin pudota, erityisen voimakkaasta alenemisesta käytetään *romahtaa*-verbiä.

Erilaisia rakenteen menettämisen verbejä käytetään myös voiman tai vaikutusvallan menettämisen merkityksessä (52).

(52) Yhdysvallat eli Vietnamin sodan ja lysähtäneen hippiaatteen turruttamaa aikaa. [lehdet90ff-v2]$_{127}$

Ihmisen romahtaminen, hajoaminen merkitsee hänen sisäisen rakenteensa muuttumista niin, että normaali toimintakyky ei säily. Tällaisessa merkityksessä käytetään *romahtaa*-verbin lisäksi muita rakenteen menettämistä merkitseviä verbejä kuten *lyyhistyä* (53).

(53) Joinain hetkinä hän oli varma, että pian hän lyyhistyisi ja luovuttaisi. [lehdet90ff-v2]$_{125}$

Epätoivottavasta toiminnasta käytetään ihmisestä myös *sortua*-verbiä (54, 55).

(54) Tekijä kuitenkin tekee tämän maltillisesti sortumatta ylettömiin ylistyssanoihin. [lehdet90ff-v2]$_{123}$

(55) Sepon isä sortui juomaan. [lehdet90ff-v2]$_{124}$

Luhistumiskehyksen metaforiset käytöt muistuttavat täten suomessa kaatumiskehyksen samantyyppisiä käyttöjä (esim. *langeta juomaan* vs. *sortua juomaan*). Ihminen on kuitenkin luhistumiskehyksen käytöissä koko toimintakykynsä menettävä toimija, eikä niinkään väliaikaisesti aktiivisen vertikaalisen asentonsa menettävä aktiivinen toimija. Samantyyppisissä ylhäältä alas putoamisen kehyksen metaforisissa käytöissä (kuten *tippua päivärahalta*) ihminen on puolestaan kokijan roolissa.

Metaforisessa luhistumisessa keskeistä on sama kuin luhistumisen kehyksessä laajemminkin eli alkuperäisen rakenteen menetys alaspäin kohdistuvan liikkeen seurauksena. Se, mitkä alkutilan ominaisuudet mahdollistavat tämän, vaihtelee.

5 Lopuksi

MLexT-tutkimusryhmän lähtökohta leksikaaliseen typologiaan on siis seuraavanlainen: Semanttisen kentän määrittely tehdään sillä perusteella, että tutkituissa kielissä tiedetään olevan lekseemejä, joiden merkitys kattaa kuvattavat tilanteet. Kuvattavat tilanteet kartoitetaan eri kielissä kyselylomakkeiden, videomuotoisen stimulusaineiston ja korpusaineistojen avulla. Esimerkit glossataan ja niiden alkuperä ilmoitetaan. Typologinen puoli sisältää sen, että eri lekseemien käyttömahdollisuudet kehyksissä kuvataan toisiinsa verrattavissa olevien semanttisten karttojen avulla. Näissä semanttisissa kartoissa kuvataan sisäkkäisten laatikoiden avulla yksittäisten lekseemien käyttömahdollisuuksia.

Putoamisen semanttisella kentällä käytettävän verbin valinta voi eri kielissä riippua mm. subjektin tarkoitteen tyypistä, putoavien esineiden määrästä, alkuasennosta, asennosta loppupisteessä tai putoamisen syystä. Eri kielten järjestelmissä neljää määriteltyä putoamisen kehystä kuvataan eri määrällä dominanttiverbejä. Kaikissa kielissä semantiikaltaan eroavia putoamisverbejä on monia, mutta ero on siinä, ovatko semantiikaltaan suppeammat verbit ainoa tapa kuvata tiettyä asiantilaa vai onko mahdollista käyttää myös dominanttiverbiä. Mielenkiintoista ja kielikohtaista on myös se, miltä muilta semanttisilta kentiltä voi lainata verbejä kuvaamaan putoamista.

Tutkitun 42 kielen joukossa suomessa oli putoamisen semanttisella kentällä sekä useita tyypillisiä että epätyypillisiä piirteitä. Tyypillistä suomessa on ensinnäkin se, että erityistapauksia putoamisessa ylhäältä alas ovat sade ja nesteen putoaminen ylipäätään samoin kuin aineiden ja samankaltaisista pienistä osasista koostuvien tarkoitteiden kuten lehtien tai neulasten putoaminen. Samoin suomen putoamis- ja kaatumisverbien metaforiset käytöt ovat nekin hyvin tavallisia otannan kielten joukossa.

Suomi on kieli, jossa putoamisen lähtöpisteellä on vähemmän merkitystä verbin valinnassa, mutta sen sijaan putoamistapaa, päätepistettä ja etenkin kontaktin tapaa päätepisteessä voi kuvata lukuisilla eri verbeillä. Niitä käytettäessä keskeinen ei ole niinkään itse putoaminen prosessina vaan sen lopputulos. Suomen tapauksessa lähtöpisteen kiinnittyneisyyden aste ei sekään vaikuta – samoja verbejä käytetään silloin kun liikkeen alkuun liittyy irtoaminen luonnollisesta kontekstista, putoaminen jostain säiliöstä tai lähteminen tilanteesta, jossa ei ole mitään kiinnittyneisyyttä.

Suomessa, kuten suurimmassa osassa otannan kielistä, millään yhdellä lekseemillä ei voi kuvata kaikkia neljää putoamisen kehystä. Ylhäältä alas putoamisen ja kaatumisen erottaminen toisistaan yhdistää suomen moniin muihin

kieliin mutta samalla erottaa sen useista muista. Toisaalta kuten edellä on tullut ilmi, suomenkin *pudota* ja *tippua* ovat erittäin monimerkityksisiä ja *langeta* on merkitysyhdistelmiltään poikkeuksellinen. Myös nesteiden säiliömetonyyminen kaatuminen esim. pöydälle on kiinnostava tapaus.

Erityisen poikkeava suomi on otoksen 42 kielen joukossa siinä, että MLexT-ryhmä on jo aiempienkin tutkimustensa perusteella todennut sen kieleksi, jossa on harvinaisen paljon eri verbilekseemejä ja myös onomatopoeettisia ja deskriptiivisiä verbejä.

Leksikaaliseen typologian lähtökohdissa on useita erityistä huomiota ansaitsevia seikkoja. Niistä ensimmäinen on semanttisen kentän määrittely. Tässä projektissa putoamiseksi määritellään vapaa liike ylhäältä alas. Kuitenkin lekseemeillä, joita käytetään kuvaamaan kontrolloimatonta liikettä, on tyypillisesti myös muita käyttöjä, kuten valuminen tai vieriminen jotain pitkin, heittäminen jne. Kuvauksen erottaminen pelkästään täysin vapaata liikettä kuvaavaksi (ilman mitään kosketuspintaa tai estettä) on käytännössä mahdotonta. Kuvattavana on tietyt kehykset (vapaan putoamisen lajit), mutta kuvausta laajennetaan myös muihin kehyksiin, joissa käytetään jossain tutkittavista kielissä samoja lekseemejä kuin lähtökohtana pidettävässä kehyksessä. Lisähaasteita tuo myös semantiikaltaan suppeampien verbien mukaanotto kuvaukseen. Halu kuvata kaikki tietyn semanttisen kentän verbit saa väistämättä merkitysten väliset rajat pullistumaan ja rakoilemaan.

Voi lisäksi kysyä, miten tulisi suhtautua sellaisiin tapauksiin, joissa tällainen liike on toki läsnä, mutta verbillä kuvataan ensisijaisesti jotain muuta? Esim. jos maito kaatui pöydälle, on mielenkiinnon kohteena pikemminkin kaatumisen lopputulos kuin itse kaatuminen riippumatta siitä, ilmaistaanko tätä erikseen vai ei.

Myös sen erottaminen, mikä kuuluu itse verbilekseemiin, on typologisesti haastavaa. Tässä artikkelissa käsiteltyjen verbien kanssa käytetään usein määreitä (*satelliitteja* Talmyn terminologiassa, vrt. Talmy 2000) *alas*, *maahan*, *pois* jne., joiden ei kuitenkaan suomessa yleisesti tulkita olevan osa itse verbiä. Esim. englannin tapauksessa verbit *fall*, *fall off* ja *fall down* käsiteltiin osin erillisinä. Venäjässä puolestaan vastaava tieto olisi tyypillisesti verbin etuliitteessä ja etuliite puolestaan vaikuttaa venäjässä verbin rektioon, mikä on peruste käsitellä samavartaloisia eri etuliittein varustettuja verbejä eri verbeinä. Slaavilaisten kielten verbien aspektipareja sen sijaan käsitellään yhtenä verbinä, mikä sekään ei ole kiistaton tulkinta. Eri verbilekseemien semantiikkaa ja käyttömahdollisuuksia typologisesti tarkasteltaessa on tällaisten seikkojen

takia vaikea päästä täysin tyydyttävään ratkaisuun edes siitä, mikä ylipäätään on lekseemi, verbi tai verbilekseemi ja mitä tulisi käyttää vertailukäsitteenä.

Leksikaalisen typologian keinoin voidaan osoittaa, että kaikkien kielten järjestelmät ovat täynnä sattumanvaraisuuksia eivätkä niiden merkityskategoriat ole sen luonnollisempia kuin muissakaan kielissä. Tämän argumentoinnissa auttaa monien kielten aineistojen systemaattinen läpikäynti ja esimerkkien glossaus. Typologisessa tutkimuksessa ei pyritäkään samanlaiseen tarkkuuteen kuin yksittäisten kielten semantiikan kuvauksessa on mahdollista päästä. Semanttisten kategorioiden vertailulla eri kielissä saavutetaan kuitenkin tuloksia, joilla on arvoa myös yksittäisten kielten kuvauksen kannalta.

Putoavan subjektin tarkoitteen vaikutusta verbin valintaan olisi suomen kuvauksen näkökulmasta syytä tarkastella yksityiskohtaisemmin. Myös putoamisverbien metaforiset käytöt olisi mahdollista kuvata huomattavasti tätä artikkelia tarkemmin. On mahdollista, että tässä tutkimuksessa huomioon otettujen typologisesti merkittävien seikkojen lisäksi suomessa on muitakin verbin valintaan vaikuttavia piirteitä. Tarkempaa aineistopohjaista ja kenties myös kokeellisiin menetelmiin perustuvaa tutkimusta tarvitaan täyttämään tämän tutkimuksen harmaita alueita ja aukkoja.

Viitteet

Apresjan, Jurij D. 2005. O moskovskoj semantičeskoj škole. *Voprosy jazykoznanija* 1/2005. 3–30.

Baroni, Marco & Bernardi, Raffaella & Zamparelli, Roberto. 2014. Frege in space: A program for compositional distributional semantics. *Linguistic Issues in Language Technologies* 9/2014. 241–346.

Behrens, Leila & Sasse, Hans-Jürgen. 1997. *Lexical typology: A programmatic sketch*. Universität zu Köln, Institut für Sprachwissenschaft.

Boguslavskij, Igor M. & Iomdin, Leonid L. 2009. Moscow semantic school. Teoksessa Kempgen, Sebastian & Kosta, Peter & Berger, Tilman & Gutschmidt, Karl (toim.), *Die slavischen Sprachen / The Slavic languages*, 813–828. De Gruyter Mouton. https://doi.org/10.1515/9783110214475

Bulakh, M. S. 2020. Verbs of falling in Tigrinya. *Acta Linguistica Petropolitana* 16(1). 677–720. https://doi.org/10.30842/alp2306573716121

Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford University Press.

Cysouw, Michael. 2007. Building semantic maps: The case of person marking. Teoksessa Miestamo, Matti & Wälchli, Bernhard (toim.), *New challenges in typolo-*

gy: Broadening the horizons and redefining the foundations, 225–248. Walter de Gruyter.

Evans, Nicholas. 2012. Semantic typology. Teoksessa Song, Jae Jung (toim.), *The Oxford handbook of linguistic typology*, 504–533. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199281251.001.0001

Fillmore, Charles. 1975. *Santa Cruz lectures in deixis 1971*. Indiana University Linguistics Club.

— 1976. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences* 280(1). 20–32.

Fillmore, Charles & Atkins, B. T. S. 1992. Towards a frame-based lexicon: The semantics of RISK and its neighbors. Teoksessa Lehrer, Adrienne & Kittay, Eva Feder (toim.), *Frames, fields and contrasts: New essays in semantics and lexical organization*. Lawrence Erlbaum.

François, Alexandre. 2008. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. Teoksessa Vanhove, Martine (toim.), *From polysemy to semantic change: Towards a typology of lexical semantic associations*, 163–215. John Benjamins.

Georgakopoulos, Thanasis & Polis, Stéphane. 2018. The semantic map model: State of the art and future avenues for linguistic research. *Language and linguistics compass* 12(2). e12270. https://doi.org/10.1111/lnc3.12270

Hakulinen, Lauri. 1969. *Suomen sanaston käännöslainoja*. Suomalaisen Kirjallisuuden Seura.

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86(3). 663–687.

Huumo, Tuomas. 2013. Kaiken edellä on nainen: Suomen dynaamisten akseligrammien järjestelmä ja sen käyttö paikan ja ajan ilmauksissa. *Sananjalka* 55(1). 7–28.

— 2019. Why Monday is not in front of Tuesday: On the uses of English and Finnish front adpositions in sequence metaphors of time. *Linguistics* 57(3). 607–652.

Jeshoi, Anniina. 2019. *Taivas putoaa niskaan, rahaa ei tipu ja suosio tipahtaa: Verbien* pudota*,* tippua *ja* tipahtaa *polysemiaa*. Helsingin yliopisto. (Pro gradu -tutkielma).

Kashkin, E. V. 2020. Glagoly peremeščenija veščestv v nekotoryh finno-ugorskih jazykah. *Acta Linguistica Petropolitana* 16(1). 463–493.

Kaškin, E. V. & Žornik, D. O. & Zakirova, A. N. & Kožemjakina, A. D. & Plešak, P. S. 2015. K leksičeskoj tipologii glagolov padenija: Dannye ural'skih jazykov. Teoksessa Miščenko, D. F. (toim.), *Dvenadcataja konferencija po tipologii i grammatike dlja molodyh issledovatelej: Tezisy dokladov*, 41–44. Nestor-Istorija.

Kittilä, Seppo. 2014. The (non-)prototypicality of direction: The (allative and illative) case(s) of Finnish. Teoksessa Luraghi, Silvia & Narrog, Heiko (toim.), *Perspectives on semantic roles*, 151–179. John Benjamins.

Koptjevskaja-Tamm, Maria & Rakhilina, Ekaterina & Vanhove, Martine. 2016. The semantics of lexical typology. Teoksessa Riemer, Nick (toim.), *The Routledge handbook of semantics*, 450–470. Routledge.

Lakoff, George & Johnson, Mark. 1980. *Metaphors we live by*. University of Chicago Press.

Langacker, Ronald W. 2008. *Cognitive grammar: A basic introduction*. Oxford University Press.

Lehmann, Christian. 1990. Towards lexical typology. Teoksessa Croft, William & Denning, Keith M. & Kemmer, Suzanne (toim.), *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75th birthday*, 161–185. John Benjamins.

Levinson, Stephen C. 2003. *Space in language and cognition: Explorations in cognitive diversity*. Cambridge University Press.

Livanova, A. N. & Mordashova, D. D. 2020. Glagoly padenija v norvežskom jazyke (bukmol). *Acta Linguistica Petropolitana* 16(1). 152–187. https://doi.org/10.30842/alp2306573716104

Mustajoki, Arto & Protassova, Ekaterina. 2007. Finskie glagoly so značeniem plavanija. Teoksessa Majsak, T. A. & Rakhilina, E. V. (toim.), *Aquamotion: Glagoly dviženija v vode: leksičeskaja tipologija*, 380–405. Indrik.

Nikunlassi, Ahti. 2013. Bolevye predikaty v finskom jazyke. Teoksessa Kibrik, A. E. (toim.), *Lingvističeskij bespredel 2: Sbornik naučnyh trudov k jubileju A. I. Kuznecovoj*, 255–273. Izdatel'stvo Moskovskogo Universiteta.

Nikunlassi, Ahti & Juvonen, Päivi. 2015. Temperature adjectives in Finnish. Teoksessa Koptjevskaja-Tamm, Maria (toim.), *The linguistics of temperature*, 491–536. John Benjamins.

Nikunlassi, Ahti & Protassova, Ekaterina. 2015. Finskie glagoly, opisyvajuščie izdavaemye životnymi zvuki. Teoksessa Rakhilina, Ekaterina & Reznikova, Tatiana & Vyrenkova, Anastasia & Orekhov, Boris & Ryzhova, Daria (toim.), *Glagoly zvukov životnyh: Tipologija metafor*. Jazyki slavjanskoj kul'tury.

Ojutkangas, Krista. 2005. Viittauskehykset ja tarkastelunäkökulma: Miten sijaintia perusakseleilla kuvataan? *Virittäjä* 109(4). 525–551.

Pajunen, Anneli. 1999. *Suomen verbirektiosta: Verbin argumenttirakenteen jäsenten valinnasta*. Turun yliopisto.

— 2001. *Argumenttirakenne: Asiaintilojen luokitus ja verbien käyttäytyminen suomen kielessä*. Suomalaisen Kirjallisuuden Seura.

Rakhilina, E. V. & Reznikova, T. I. & Ryzhova, D. A. 2020. Tipologija metafor padenija. *Acta Linguistica Petropolitana* 16(1). 64–112. https://doi.org/10.30842/alp2306573716102

Rakhilina, Ekaterina V. & Reznikova, Tatiana I. 2014. Doing lexical typology with frames and semantic maps. Teoksessa Higher School of Economics. http://dx.doi.org/10.2139/ssrn.2564011

Reznikova, T. I. & Rakhilina, E. V. & Ryzhova, D. A. 2020. Glagoly padenija v jazykah mira: frejmy, parametry i tipy system. *Acta Linguistica Petropolitana* 16(1). 9–63. https://doi.org/10.30842/alp2306573716101

Salminen, Tapani. 2012. Traces of Proto-Samoyed vowel contrasts in Nenets. Teoksessa Hyytiäinen, Tiina & Jalava, Lotta & Saarikivi, Janne & Sandman, Erika (toim.), *Per Urales ad Orientem: Iter polyphonicum multilingue: Festskrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012*, 339–358. (Suomalais-Ugrilaisen Seuran Toimituksia 264). Suomalais-Ugrilainen Seura.

— 2023. Suomalais-samojedilaisia sanavertailuja. Teoksessa Holopainen, Sampsa & Kim, Jeongdo & Metsäranta, Niklas (toim.), *Elämä ja etymologia: Janne Saarikiven 50-vuotisjuhlakirja*, 377–399. Books on Demand.

Talmy, Leonard. 2000. *Toward a cognitive semantics: Typology and process in concept structuring*. MIT Press.

Teeri-Niknammoghadam, Krista. 2021. Edessä *ja* edellä*: Suomen* ete-*vartaloisten tilagrammien merkitykset ja metaforisuus*. Turun yliopisto. (Väitöskirja).

Viimaranta, Johanna & Protassova, E. U. 2020. Glagoly padenija v finskom jazyke. *Acta Linguistica Petropolitana* 16(1). 389–434. https://doi.org/10.30842/alp2306573716112

Vinogradova, O. I. & Viklova, A. V. & Pospelova, K. V. 2020. Glagoly padenija v anglijskom jazyke: Dannye korpusov i tipologija. *Acta Linguistica Petropolitana* 16(1). 115–151. https://doi.org/10.30842/alp2306573716103

Zhukova, M. V. & Kor Chahine, I. 2020. Francuzskij glagol tomber i ego sinonimy: ‘Sverhu vniz’ i ne tol’ko. *Acta Linguistica Petropolitana* 16(1). 225–247. https://doi.org/10.30842/alp2306573716106

Yhteystiedot:
Johanna Viimaranta
Helsingin yliopisto
FI-00014 Helsingin yliopisto
sähköposti: johanna(dot)viimaranta(at)helsinki(dot)fi

Infinitives in transit: The spreading of shortened infinitive forms from Finland Finnish to Finnish spoken in Sweden

Lasse Vuorsola
Stockholm University

Abstract

The study examines the use of two spoken Finnish features among Sweden Finnish adolescents in a formal education setting. The features in question are the colloquial versions of the A-infinitive lative (e.g., *mä saan pelaa* 'I'm allowed to play') and the MA-infinitive illative (e.g., *mä meen pelaa* 'I'll go and play') forms when used with contracting verbs (e.g., *pelata* 'to play'). These features are most prominent in the spoken Finnish in Helsinki, but the data shows it is also present in the speech of bilingual children in Sweden.

Keywords: Non-finites, minority language, Finnish, spoken language, Finnish in Sweden

1 Introduction

The current squib is a descriptive study of the spoken Finnish of 6th grader pupils attending a bilingual Sweden Finnish independent school (SFS) in the greater Stockholm area. Specifically, I examine to what extent and how a group of Sweden Finnish 6th grader pupils use shortened versions of the A-infinitive lative case e.g., *saan pelata* 'I'm allowed to play' and the MA-infinitive illative case e.g., *menen pelaamaan* 'I'll go and play' in conjunction with so-called contracting verbs in spoken Finnish e.g., *minä pelaan* 'I play'. The shortened versions of the above mentioned examples would be *saan pelaa* and *meen pelaa* respectively.

1.1 Situating the study

Finnish has a centuries-long historical presence in Sweden with an estimated 80,000–450,000 speakers of Finnish today (Lainio 2018: 48). The language

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

Finnish Journal of Linguistics 36 (2023), 213–228
DOI: 10.61197/fjl.124673

is acquired in schools and partly at home, but this varies due to incomplete language transfer among previous generations (Vuorsola 2022: 2). The modern basis for the minority's formation can be traced to the 1960s and 1970s when large groups of Finnish people emigrated from Finland to Sweden to find work. The resulting large groups of Finnish speaking immigrant children led to changes in the Swedish school system, such as the introduction of bilingual classes in the compulsory school. Eventually the support for Finnish education in Sweden vaned and these classes were discontinued and forced assimilation policies emerged (Hyltenstam 1999: 11–13). After this, bilingual formal education in Finnish resumed mostly in so-called bilingual independent schools (Vuorsola 2019: 364). One of the biggest positive developments for the minority was the ratification of the Council of Europe's *Framework Convention for the Protection of National Minorities* (FCNM) and the *European Charter for Regional or Minority Languages* (ECRML) in 2000 and the subsequent classification of Finnish, along with Meänkieli, Yiddish, Romani Chib and Sámi, as a national minority language in Sweden. Despite this status, the Sweden Finnish minority's rights are frequently not met, especially in terms of formal education of Finnish in municipal schools (Lainio 2014; Vuorsola 2019).

1.2 Previous research on Finnish in Sweden

There is a growing body of research on the Sweden Finnish minority and the use of Finnish in Sweden. Studies on Finnish in Sweden often involve translanguaging (Kolu 2017) or code-switching (Muhonen 2014) in formal education settings, error analysis (Valijärvi 2016) and examining of sentence constituents (Nesser 1983) and case endings (Nesser 1986). Juvonen (2000), in turn, has examined the grammaticalisation process of the definite demonstrative pronoun *se* (Eng. a pronoun proximal, near to the interlocutor) into a definitive article. A study closest to the interests of the current squib is Lainio (1993), where he has examined the spoken Finnish in Sweden in the 1980s including the use of the MA-infinitive illative and how different Finnish dialects are present in the speech of Sweden Finnish speakers (Lainio 1993: 36). The current study attempts to add knowledge about the further linguistic development of Finnish in Sweden in an age-group that may be seen as the future language carriers of Finnish in Sweden.

Next, I present the principles for the on-going change affecting the A-infinitive lative and the MA-infinitive illative case of contracting verbs in

spoken Finnish. In Section 3 I discuss the data, followed by the findings in Section 4 as well as the discussion and conclusion in Section 5.

2 Finnish infinitives in standard and spoken forms

The structures under study are 1) the A-infinitive lative and 2) the MA-infinitive illative in spoken Finnish. The A-infinitive lative acts as a complement to other verbs while the MA-infinitive illative denotes movement (Kuparinen 2018: 30). The two structures are called infinitives in the current squib, but they can also be referred to as non-finites (see Ylikoski 2022). Out of all verb types in Finnish only the 1st verb type, i.e., verbs that end in *AA*, *OA*, *eA*, *UA*, *iA* in the A-infinitive form e.g., *sanoa* 'to say' (Mantila 2004: 326–327; Kuparinen 2018: 31) and so-called contracting verbs may occur as shortened versions both in the A-infinitive lative and the MA-infinitive illative structures (Kuparinen 2018: 32). Contracting verbs in Finnish refer to verbs that end in either -*AtA*, -*UtA* or -*OtA* in the A-infinitive form depending on the vowel harmony. Contracting verbs are a productive verb type in Finland Finnish to which recent loaned verbs tend to be accommodated (see Räisänen 1987). It is less common for contracting verbs than verbs belonging to the 1st verb type to occur as shortened infinitives in speech, since the form is mostly used in the Helsinki region and has not yet become an unmarked form of the common spoken variety (Kuparinen 2018: 31).

In Table 1 I exemplify the change with the verb *varata* 'to reserve'. The examples include the verbs *voida* 'can, to be able' and *mennä* 'to go'. For stylistic consistency, I have used the colloquial version *meen* 'I'll go' with the shortened infinitive of the MA-infinitive illative. There is no such colloquial form for the verb *voida*.

For the A-infinitive lative form, the A-infinitive *varata* 'to reserve' is replaced by the 3rd person verb form (*hän*) *varaa* 's/he reserves' and is most likely the product of analogy from how shortened infinitives are constructed with the 1st verb type. Originally the forms in question are most likely the result of Swedish and Finnish speakers being in contact in the southern Tavastian region (Itkonen 1989: 364–365; Kuparinen 2018: 31). Children who are learning to speak are also known to produce the structure (Sorsakivi 1982: 381).

The shortened versions of the infinitives occur in the spoken Finnish among youths in Helsinki (Paunonen 2006: 38) and, as Kuparinen (2018: 38)

Table 1. The standard and shortened versions of A infinitive lative and MA-infinitive illative forms with contracting verbs

	A-infinitive lative	MA-infinitive illative
Standard infinitives	*voin varata* 'I can reserve'	*menen varaamaan* 'I'll reserve', lit. 'I will go and reserve'
Shortened infinitives	*voin varaa* 'I can reserve'	*meen varaa* 'I'll go and reserve'

demonstrates in his recent work, the feature seems to be expanding and becoming more than a youth variant, since more people retain the feature in their speech as adults. Sorsakivi (1982: 390) states already in her article from 1982 that the shortened infinitives are a prestige form in spoken Finnish feature among youths in Helsinki. Next, I will discuss the data and how it was collected.

3 Data and data collection

The data was collected by the author in the spring of 2018 after a completed ethical vetting.[1] The entirety of the data consists of 19 filmed and audio recorded Finnish classes with 6th graders in a bilingual Sweden Finnish school, 15 of which contained uses of the linguistic features that are under study. I implement ethnographic data collection methods such as classroom observations as well as audio and video recordings. In all 12 out of 13 6th graders partook in the study. The composition of the group can be considered as being representative for a class of pupils in a Sweden Finnish bilingual class.

After the recording, the researcher systematically identified the use of A-infinitive lative and MA-infinitive illative constructions of contracting verbs in the data, transcribed the examples and compiled an example list.

The data was gathered for a larger sociolinguistic study and has some limitations. For example, the pupils' personal histories were not systematically collected. To partly amend this, I have included Appendix B which features some details of the pupils' possible migration histories that appear during interactional exchanges in the data. In addition to this, I have added information

[1] Project number 2017/2513-31/5, vetting completed by Regionala etikprövningsnämnden i Stockholm in Stockholm 11 January 2018. The permission included data collection for both sociolinguistic studies and linguistic analysis.

on the pupils' Swedish accents. A Sweden Swedish accent in spoken Swedish might indicate that the student in question has lived in Sweden for an extended period and s/he may have had less contact with on-going linguistic development of spoken Finnish in Finland.

In the next section I analyse how and to which extent shortened infinitives are used by Sweden Finnish adolescents in a formal education setting.

4 Findings

Out of 19 lessons 15 featured one or more either shortened or standard Finnish infinitives and in total the construction appears 62 times in the data. Out of the 62 occurrences 37 appear as the shortened infinitive variant and 25 appear as the standard Finnish variant. All in all, there are 12 participating pupils, 7 of whom use the infinitive forms studied:

- 2 used only the standard forms;
- 2 used only the shortened forms;
- 3 used both.

When looking at the frequency of use, it is indisputable that the shortened infinitives are often present in the spoken Finnish among the 6th graders in the SFS. The verbs used for the infinitive constructions are few and seem to follow the theme of gaming and are mostly loanwords. (See Appendix C for full list of occurrences.)

All of the found occurrences were either the shortened infinitives or standard infinitives. None represented the form of the unmarked MA-infinitive illative *meen pelaan* 'I'll go and play', where the illative is marked with an *-n*. This form is not considered as a shortened infinitive but another spoken Finnish variant of the structure. This variant has spread wider than the shortened infinitive and can be found in the Finnish Ostrobothnia, the Tavastian dialects and even eastern dialects (Mielikäinen 1991: 52–53). This would indicate that the spoken Finnish that is impacting the pupils' use of infinitive forms is specifically from the Helsinki region and not from another part of Finland.

Next, I examine how the pupils use shortened and standard infinitives and whether the construct is used in stylistically consistently in a manner that is expected. I compare this stylistic consistency with the background information presented in Appendix B to try to determine how independently the pupils use the features under study.

4.1 How shortened infinitives are used by Sweden Finnish adolescents

All the following examples are originally in Finnish. The shortened infinitives are in bold typeface in the examples. Examples (1a–b) are from pupil M1 who is discussing a movie.

(1) a. M1: ku emmä haluu **spoilaa** mitä täs tapahtuu
'because I don't want to spoil what happens in this'

b. M1: mut mä en- mä en haluu kir- kirj- **spoilata** mitä siin tapahtuu (.) [teacher's name] mä en haluu **spoilata** mitä siin tapahtuu (.) se on niin hyvä elokuva (.)
'but I don't want to write- write- spoil what happens in it (.) I don't want to spoil what happens in it (.) it's such a good film'

In Example (1a) above M1 uses the shortened infinitive *spoilaa* 'to spoil' and other colloquial forms such as the short 1st person singular pronoun *mä* 'I' which is commonly used in the Helsinki region, Tavastian dialects and southern Ostrobothnia (Mielikäinen 1991: 14; Paunonen 2006: 43) in the assimilated form *emmä* (instead of *en minä* 'not me'), as well as the shortened colloquial forms of words such as *ku* 'because' instead of the full *kun* and *täs* 'in this' instead of *tässä*. In this regard M1 can be said to use the shortened infinitive in a stylistically consistent context. Later in (1b), M1 switches to the standard Finnish infinitive *spoilata* 'to spoil' when giving an answer to his teacher. However, he keeps using similar colloquial Finnish forms such as *mä* 'I' instead of *minä* (Mielikäinen 1991: 14; Paunonen 2006: 43). This could signal that M1 regards the shortened infinitives as a form that is too informal to be used in his spoken Finnish when answering the teacher.

This behaviour is not consistent, as exemplified by M2 in Example (2) below.

(2) M2: tota mun kaveri se halus vaan **pelaa** (.) niin sen vanhemmat teki nii että se sai yhen viikonlopun vain **pelata** sitten se (.) se pelas niim paljo (.) se pelas iha niinku kakstoist tuntii molemmat päivät (.) niin se väsy siihen peliin (.) sitten se teki kaikkee muuta paitsi pelasi (.) ku se väsy sitten niinku siihen ku se sai **pela** (.) se o ku niinku tekee tosi paljo mieli karkkii (.) saa syödä mitem paljoo karkkii vaan halus sit tulee niin pahana (.) paha olo et ei koskaan enää haluu karkkii tosi paljoo

> ‘so my friend he just wanted to play (.) so his parents did so that during one weekend he got to only play and then he (.) he played so much (.) he played like twelve hours on both days (.) so he got tired of the game (.) then he did everything else except played (.) because he got tired like because he got to play (.) it’s like when you are really craving candy (.) you get to eat as much candy as he wanted then you feel sick (.) sick that you never want lots of candy’

As presented in Appendix B, M2 has never lived in Finland but one of his parents is from Helsinki and his extended family still lives in Helsinki. His Swedish accent is local. In this rather lengthy comment M2 alternates between the shortened and the standard infinitive i.e., *pelaa* and *pelata* ‘to play’ respectively. The third example is partly cut off but can be interpreted as the shortened infinitive. In this sequence, the pupil in question is telling the anecdote for the whole class and indirectly to the teacher. M2 uses common colloquialisms which can be found in e.g., the Helsinki region, such as the apocope in the past tense of contracting verbs *halus* ‘wanted’; as standard *halusi* (Mielikäinen 1991: 46) and the demonstrative pronoun *se* ‘it’ instead of *hän* ‘he/she’ (Paunonen 1995: 165–177). In this regard M2’s use of shortened infinitives as a feature among other spoken Finnish forms is consistent. There seems to be no pattern why the shortened or standard infinitive is used, and these can thus far be considered as interchangeable.

The following example exhibits that the pupil M3 is also able to use short infinitives competently.

(3) M3: pitää **eimaa** vähän ylöspäin ja sillon pitää **smoukkaa-** menee heti siihen connecting

‘you have to aim a little up and then you have to smoke- go directly there connecting’

M3 is a bilingual pupil with a local Sweden Swedish accent who frequently opts for using Swedish in discussions with his peers. Example (3) stems from a discussion on online gaming strategy with M1. M3’s language use is marked by colloquialisms. M3 is using two shortened infinitives, *eimaa* ‘to aim’ and *smoukkaa* ‘to smoke’, instead of the A-infinitive standard infinitives but also a spoken Finnish form of *silloin* ‘then’, where the *-i* of the diphthong *-oi* is omitted and becomes *sillon*. This is a widely spread feature of Spoken Finnish (VISK §24). In other words, M3 use the shortened infinitives in a context that features other colloquialisms.

The next example features M4's language use, who is very competent in Finnish and has lived and received formal education in Finland. Along with M2 who has a confirmed link with Helsinki, M4 could potentially be one of the pupils spreading the feature in the group. M4 uses the shortened infinitives repeatedly. In Example (4) below, M4 showcases his mastery of the structure by alternating between the shortened and standard infinitives to convey different roles in conversation. In the example, the pupils are discussing how much they are allowed to play games at home.

(4) 01 M4: no emmä haluu **pelaa** mitään marioo
'I don't want to play any (Super) Mario'

02 M5: minä voisin **pelaa** sellasta
'I could play that'

03 M4: jos mä pelaan ni mä ehin puoltuntii **pelaa** illalla (.) kun mä tuun treeneistä (.) enkä ees sitäkään ku "ei saa **pelata** viikolla"
'If I play, then I have time to play half an hour in the evening (.) when I get home from training (.) maybe not even that because "no playing during the week" ' [imitating a nagging voice]

04 M5: [laughter] samat
'same'

M4 uses the shortened infinitives on lines 01 and 02, but switches to the standard A-infinitive on line 03 when saying *ei saa pelata viikolla* 'no playing during the week' in a nagging voice and presumably takes on the role of his parents. This shift can be seen as an example of *stylisation* which, according to Rampton (2009: 149) "involves reflexive communicative action in which speakers produce specially marked representations of languages, dialects, and styles that lie outside their own habitual repertoire (at least as this is perceived within the situation at hand)". Example (4) indicates that M4 is aware of the difference between the standard Finnish variant and the shortened variant and that the shortened variant is a feature of a youth idiolect.

When examining M4's language use it is apparent that he is competent in using common features of Finnish spoken language. For example, he uses spoken Finnish variant *mä* instead of the full first-person singular pronoun *minä* 'I' (Mielikäinen 1991: 14; Paunonen 2006: 43). The partitive case in the word *tuntii* 'hours' in M4's speech is assimilated, which is common in

Tavastian dialects and in the Helsinki region (Paunonen 2006: 37) and in standard Finnish would be *tuntia*. When combining these common spoken Finnish variants, the rare feature from the Helsinki region *pelaa* 'to play' is non-intrusive.

Example (4) also features M5 who has lived in Finland and his history of not being originally from Sweden is corroborated by his non-local Swedish pronunciation. He also speaks one of the Romance languages at home with at least one parent. As the other pupils, M5 uses shortened infinitives competently but his use is marked with some stylistic variations and differences. On line 02 of Example (4) M5 uses the shortened infinitive *pelaa* 'to play', but instead of being consistently colloquial he uses the full first-person singular personal pronoun *minä* 'I' instead of *mä*. Though this can also be interpreted as an emphasis marking that he would enjoy playing the Mario game and not an inconsistency, M5's speech features similar inconsistencies in other examples.

(5) M5: viikonloppuna, koska yleensä kun on viikonpäivät silloin tulee läksyi, pitää mennä suihkuun, syödä ja niinku ei oo niinku aikaa **pelaa**, niin viikonloppuna kun sä tuut kotiin sä voit mennä heti sohval ja sitten (??) koulust koska huomenna on lauantai niin

'During weekends, because normally during weekdays there's homework, you have to shower, eat and like there's no time to play, so during weekends, when you come you can go straight to the sofa and then (??) from school because tomorrow is Saturday so'

In Example (5) M5 discusses his gaming habits and uses the form *pelaa* 'to play' instead of *pelata* 'to play'. However, this construction is not a shortened infinitive. Instead, M5 uses the short form *pelaa* in an incomplete clause denoting ownership, which in standard Finnish would be (*sinulla*) *ei ole aikaa pelata* 'you have no time to play'. In other words, M5 uses the shortened infinitive differently than the other pupils. When it comes to stylistic consistency in Example (5), M5 uses the colloquial 2nd person pronoun *sä* 'you' which is found in e.g., the spoken Finnish in Helsinki (Mielikäinen 1991: 14; Paunonen 2006: 43), but simultaneously includes the *i* in the second syllable *-oi* diphthong in the word *silloin* 'then', which, as mentioned before, is something that is often omitted in colloquial speech patterns (VISK § 24). A speaker who retains the *-i* in the diphthong may be perceived as foreign or to be using a formal speech pattern (Lehtonen 2015: 116–119). However,

he does omit the *i* from the word *sellasta* 'that' in the previous Example (4), which in standard form would be *sellaista*. It is worth noting that there is variation with regards to retaining the *-i* specifically in the words *sellainen* and *semmoinen* (see Mustanoja 2011).

M5's stylistically uneven production of spoken Finnish suggests that the shortened infinitive feature might not be a completely independent part of M5's language use and I thus interpret it as an example of accommodation through convergence in the above mentioned speech situation (Gallois et al. 2005: 123; Giles 2016) which is a strategy where speakers modify their linguistic output to resemble the output of their co-interlocutor when "they desire recipients' approval" (Gallois et al. 2005: 125). Next, I present examples where shortened infinitives are used beyond their set constructions.

4.2 Infinitives used in other structures

The only time M5 does not follow the pattern of using a previously established shortened infinitive is in the example (6), where the pupils are discussing the length of the recess.

(6) 01 M6: Hei onks meil nytten viidentoista minuutin rasti
'Hey do we have a fifteen-minute recess now'
02 M5: Joo
'Yes'
03 M4: [kahdenkymmenen
'twenty'
04 M5: kahdenkymmenen
'twenty'
'huh'
05 M5: me ehitään (.) mä voin opettaa teille pokeria **pelata**
'we have time (.) I can teach you poker to play'

Here the use of the shortened infinitive deviates from both the colloquial Finnish model and the standard A-infinitive form. The verb 'to play' is a part of a noun phrase that acts as the object to the main predicate 'to teach'. In this context the verb 'to play' in standard Finnish would be nominalized by using the derivative ending *-minen* to *pelaaminen* and, following the inflection rules of Finnish objects, would be in the partitive case, i.e., *mä voin opettaa teille*

pokerin pelaamista ‘I can teach you to play poker’. Admittedly, the example could also be turned into to MA-infinite illative *voin opettaa teitä pelaamaan pokeria* but it would require additional changes.

M2 uses the shortened infinitive similarly in Example (7).

(7) M2: nii samassa lauseessa ja sitte myös j- m- joka kerta kun vanhemmat sanoo että tule alas syömään mä haluun vaa jatkaa **pelaa**
‘in the same sentence and then also j- m- every time when parents tell me to come down to eat I just want to continue playing’

As in the previous example i.e., the sentence requires the nominalization of the verb with the derivative ending in the partitive case i.e., *pelaamista* ‘playing’, in order to work as the object in the sentence. Instead of this nominalization M2 opts to use the shortened infinitive *pelaa* ‘to play’. Since M5 is Swedish-Finnish bilingual but has never lived in Finland, this may very well be the result of interference from Swedish, where the sentence would be *jag vill fortsätta spela* ‘I want to continue playing, lit. (to) play’.

Finally, M1 repeats this pattern for a third time in Example (8).

(8) M1: kyllä mä nyt tuun cs:ää jatkaa **pelaa**
‘I will continue playing CS’ [CS refers to the online game Counter Strike]

Again, the sentence would require the nominalisation of the verb, i.e., *pelaamista* ‘playing’. As with Example (7) the shortened infinitive here could be interference from Swedish, since the translation, *jag kommer att fortsätta spela CS* ‘I will continue to play CS’, as the forms are similar.

Examples (6), (7), and (8), then, indicate that the use of the A-infinitive and its shortened colloquial form may be expanding to encompass the *minen*-derivative ending as well.

5 Discussion and Conclusion

This squib cannot be construed as indicative of the linguistic development of spoken Finnish in Sweden in a broad sense. A larger corpus is needed in order to be able to conduct such a study. The features under study occur infrequently in speech (Sorsakivi 1982: 381) and the amount of research participants is small. As discussed, the shortened infinitives of contracting verbs can originally be traced back to the Southern Tavastian region in Finland and to

the contact between Swedish and Finnish speakers (Itkonen 1989: 364–365; Kuparinen 2018: 31). Considering that the involved languages in the studied group are the same, it is possible that a similar change is occurring currently within the bilingual school. However, the likelier scenario, based on the findings, is that the development of Finnish in Finland and Sweden is linked, at least to some degree.

The use of the infinitives indicates three tendencies among the pupils. The first is that the pupils who use the shortened infinitives do so quite frequently. Secondly, the pupils who use the feature replicate it competently in the way that it is used in the Helsinki region. Some pupils are even using it to mark a shift from a role to another as was the case in Example (4). Thirdly, despite the pupils' ability to use the feature competently, there are differences suggesting that pupils such as M5 could have adopted shortened infinitives from other pupils in the class. Finally, Examples (6–8) exhibit how pupils seem to replicate Swedish structures in Finnish resulting in the use of shortened infinitives instead of the derivate ending *-minen* that is used for nominalisation (for a similar process involving the demonstrative pronouns *tämä*, *tuo*, *se* see Juvonen 1996). This suggests that the use of short infinitives might spread to other grammatical structures.

References

Gallois, Cindy & Ogay, Tania & Giles, Howard. 2005. Communication accommodation theory: A look back and a look ahead. In Gudykunst, William B. (ed.), *Theorizing about intercultural communication*, 121–148. Sage.

Giles, Howard. 2016. The social origins of CAT. In Giles, Howard (ed.), *Communication Accommodation Theory: Negotiating personal relationships and social identities across contexts*, 1–12. Cambridge University Press. https://doi.org/10.1017/CBO9781316226537.001

Hyltenstam, Kenneth. 1999. Inledning: Ideologi, politik och minoritetsspråk. In Hyltenstam, Kenneth (ed.), *Sveriges sju inhemska språk: Ett minoritetsspråkperspektiv*, 11–37. Studentlitteratur.

Itkonen, Terho. 1989. *Nurmijärven murrekirja*. (Suomalaisen Kirjallisuuden Seuran Toimituksia 498). Suomalaisen Kirjallisuuden Seura.

Juvonen, Päivi. 1996. Språkkontakt och språkförändring: Användningen av demonstrativa pronomen i en sverigefinsk kontaktsituation. In Huss, Leena (ed.), *Många vägar till tvåspråkighet: Forskarseminarium vid Göteborgs universitet den 21–22 oktober 1994*, 57–88. Centre for Multiethnic Research, Uppsala University.

— 2000. *Grammaticalizing the definite article: A study of adnominal determiners in a genre of spoken Finnish*. Department of Linguistics, Stockholm University.
Kolu, Jaana. 2017. ”Määhän aina niinku sekotan typ engelska svenska och finska”: Diskurspartiklar som interaktionell och grammatisk resurs i tvåspråkiga ungdomssamtal i Haparanda, Stockholm och Helsingfors. In Sköldberg, Emma & Andréasson, Maia & Adamsson Eryd, Henrietta & Lindahl, Filippa & Lindström, Sven & Prentice, Julia & Sandberg, Malin (eds.), *Svenskans beskrivning 35*, 107–119. Göteborgs universitet.
Kuparinen, Olli. 2018. Infinitiivien variaatio ja muutos Helsingissä. *Virittäjä* 122(1). 29–52. https://doi.org/10.23982/vir.65310
Lainio, Jarmo. 1993. Sweden Finnish. In Extra, Guus & Verhoeven, Ludo (eds.), *Immigrant languages in Europe*, 21–53. Multilingual Matters.
— 2014. The art of societal ambivalence: A retrospective view on Swedish language policies for Finnish in Sweden. In Halonen, Mia & Ihalainen, Pasi & Saarinen, Taina (eds.), *Language policies in Finland and Sweden: Interdisciplinary and multi-sited comparisons*, 116–144. Multilingual Matters. https://doi.org/10.21832/9781783092710-007
— 2018. The five national minorities in Sweden and their languages: The state of the art and ongoing trends. In Forsgård, Erik Nils & Markelin, Lia (eds.), *Perspectives on minorities in the Baltic Sea area*, 45–76. Magma.
Lehtonen, Heini. 2015. Sano vai sanoi – Mitä merkitystä on yhdellä *i*:llä: Jälkitavujen *i*-loppuisen diftongin sosiaalinen indeksisyys monietnistyvässä Helsingissä. In Sorjonen, Marja-Leena & Rouhikoski, Anu & Lehtonen, Heini (eds.), *Helsingissä puhuttavat suomet: Kielen indeksisyys ja sosiaaliset identiteetit*, 444–479. Suomalaisen Kirjallisuuden Seura.
Mantila, Harri. 2004. Murre ja identiteetti. *Virittäjä* 108. 322–346.
Mielikäinen, Aila. 1991. *Murteiden murros: Levikkikarttoja nykypuhekielen piirteistä*. (Jyväskylän yliopiston suomen kielen laitoksen julkaisuja 36). Jyväskylän yliopiston suomen kielen laitos.
Mustanoja, Liisa. 2011. *Idiolekti ja sen muuttuminen: Reaaliaikatutkimus tampereen puhekielestä*. (Doctoral dissertation).
Nesser, Anne. 1983. *Subjekt, objekt och predikativ i sverigefinska barns uppsatser*. (Fuskis/Fidus 5). insk-Ugriska Institutionen, Uppsala universitet.
— 1986. *Se oli kurja käyä koulusa: Bruket av inre och yttre lokalkasus i sverigefinska skolbarns språk*. Acta Universitatis Upsaliensis.
Paunonen, Heikki. 1995. *Suomen kieli Helsingissä: Huomioita Helsingin puhekielen historiallisesta taustasta ja nykyvariaatiosta*. Helsingin yliopiston suomen kielen laitos.
— 2006. Vähemmistökielestä varioivaksi valtakieleksi. In Juusela, Kaisu & Nisula, Katariina (eds.), *Helsinki kieliyhteisönä*, 13–99. Helsingin yliopiston suomen kielen ja kotimaisen kirjallisuuden laitos.

Räisänen, Alpo. 1987. Miten uudet *aa*-, *ää*-loppuiset supistumaverbit tulevat nykysuomeen? *Virittäjä* 91(4). 491–502.
Rampton, Ben. 2009. Interaction ritual and not just artful performance in crossing and stylization. *Language in Society* 38(2). 149–176.
Sorsakivi, Merja. 1982. Infinitiivijärjestelmän muutoksia lasten kielessä. *Virittäjä* 86. 377–391.
Valijärvi, Riitta-Liisa. 2016. Ruotsinsuomalaisten opiskelijoiden kirjallisen tuotoksen morfosyntaksin ja sanaston virheanalyysia. *Fenno-Ugrica Suecana Nova* 15. 175–200.
Vuorsola, Lasse. 2019. Societal support for the educational provisions of Finnish in the Swedish school system in theory and practice. *Language Policy* 18. 363–385. https://doi.org/10.1007/s10993-018-9491-5
— 2022. Peer interaction practices as part of a Sweden Finnish spatial repertoire. *Linguistics and Education* 67. https://doi.org/10.1016/j.linged.2022.101014
Ylikoski, Jussi. 2022. Non-finites. In Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena (eds.), *The Oxford guide to the Uralic languages*, 936–949. Oxford University Press. https://doi.org/10.1093/oso/9780198767664.003

Appendix A Transcription key

(.) pause
- interrupted speech
[..] extralinguistic information or clarification
[??] unintelligible speech

Appendix B Pupils' personal histories and Swedish accents

Pupil	Personal history	Swedish accent
M1	Personal history unknown	Not applicable
M2	Has never lived in Finland but has relatives from and in Turku as well as Helsinki	Sweden Swedish
M3	Personal history unknown	Sweden Swedish
M4	Has lived and received formal education in Finland	Not applicable
M5	Has lived in Finland but there is no data on when and where. Speaks one of the Romance languages as a home language with at least one parent	Sweden Swedish but not native

Appendix C Number of occurrences with translations

Contracting verbs in Finnish	English translation	Occurrences with standard inf.	Occurrences with shortened inf.
betata	to bet	0	2
defusata	to defuse	0	3
depositata	to deposit	0	3
eimata	to aim	0	1
fiksata	to fix	0	1
hypätä	to jump	2	1
joinata	to join	2	0
kämpata	to camp	0	1
levätä	to rest	1	0
mess-upata	to mess up	0	1
naiffata	to knife	0	1
pelata	to play	15	19
plantata	to plant	0	1
smoukata	to smoke	0	1
spoilata	to spoil	3	1
tuhota	to destroy	0	1
vastata	to answer	2	0

Contact information:

Lasse Vuorsola
Stockholm University
Department of Slavic and Baltic Studies, Finnish, Dutch and German
Section for Finnish
SE-106 91 Stockholm
SWEDEN
e-mail: lasse(dot)vuosola(at)finska(dot)su(dot)se

Di Garbo, Francesca & Olsson, Bruno & Wälchli, Bernhard (eds.). 2019. *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*. (Studies in Diversity Linguistics 26). Language Science Press. Pp. 340.

Di Garbo, Francesca & Olsson, Bruno & Wälchli, Bernhard (eds.). 2019. *Grammatical gender and linguistic complexity: Volume II: World-wide comparative studies*. (Studies in Diversity Linguistics 26). Language Science Press. Pp. 389.

Reviewed by Patricia Carvalho Ribeiro & Rahel T. Dires

1 Introduction

Grammatical gender is commonly defined as a grammatical feature, classified as a noun class, which is present in several languages. More precisely, in almost half of the languages in the world (Audring 2016). Such a feature is often linked to biological sex for animate referents. For inanimate referents, however, grammatical gender can be related with other properties of the noun. Another characteristic of grammatical gender is agreement. According to Corbett (2013), "gender is about agreement", which means that evidence of gender should trespass the nouns themselves to agree with other expressions in the phrases it heads. Thus, how does gender agreement occur in different languages? Understanding the nature of grammatical gender across the languages of the word is not a trivial task.

In *Grammatical gender and linguistic complexity, Volume I: General issues and specific studies*, and *Grammatical gender and linguistic complexity, Volume II: World-wide comparative studies*, Francesca Di Garbo, Bruno Olsson, and Bernhard Wälchli present a collection of up-to-date studies investigating the typological aspects of grammatical gender and its evolution in languages of the world. We consider this set of papers to be an important contribution to the field of Typological Linguistics, since it presents traditional and modern tools of analysis to measure gender complexity cross-linguistically, as well as discusses how these complexities evolve and change over time. Hence, these papers go back and forth in time, presenting both diachronic and synchronic studies on the complexity of gender systems, not only in large cross-linguistic samples but also in individual languages not well described in the literature so far.

Finnish Journal of Linguistics 36 (2023), 229–239
DOI: 10.61197/fjl.136275

2 Volume I: General issues and specific studies

The first volume of the publication contains 10 chapters distributed in four parts. Chapter 1, as defined by the editors, is an overview of the books' contents meant to contextualize the volumes in the field of grammatical gender studies, as well as establish the key concepts supporting the questions discussed in the papers.

2.1 Part I: General issues

In the first part of the book, the reader will be introduced to the theoretical foundations of gender complexity in three chapters. In Chapter 2, *Canonical, complex, complicated?*, Jenny Audring applies the canonical typology from Corbett & Fedden (2016) to build up an approach to compare and contrast CANONICITY, COMPLEXITY, and DIFFICULTY. In this study, canonicity is taken as the baseline for assessing complexity, while difficulty is presented as an independent notion. One strength of the chapter is the way in which it provides a nuanced view of the concept of canonical structures. Audring argues that while canonical structures can be useful as a reference point for language analysis, they are often oversimplified and do not accurately reflect the full complexity of linguistic systems. She uses several examples from different languages to demonstrate how apparently canonical structures can vary in different contexts or be subject to various exceptions. Regarding complexity, this work presents a detailed study of the complexity of gender systems. The principles applied in judging complexity are presented here, and a concrete example of it is given on page 23 via the PRINCIPLE OF TRANSPARENCY, which states that markers having different functions are more complex than markers dedicated to only one function. In addition, difficulty is defined as inherently relative and possibly influenced by a range of external factors, thus, it needs to be observed in context. For example, according to the evidence found here, difficulties in acquiring a gender system are related to the frequency with which children are exposed to nouns accompanied by agreeing words. Although the chapter is balanced between theoretical discussion and illustrative examples, the terminology is specific to canonical typology and needs to be considered in light of the descriptions presented by the author, which might negatively affect the impact of the chapter in other contexts of typological studies.

In Chapter 3 *Gender: Esoteric or exoteric?*, Östen Dahl describes how the limitations concerning the data on language ecology[1] hinder the understanding of the relations between grammatical complexity and factors external to the language system. More specifically, Dahl questions to what extent ecological factors influence the growth, maintenance, and loss of gender systems, arguing that it is necessary to go beyond the patterns presented in databases such as WALS to understand in which conditions gender systems emerge and mature. Additionally, Dahl betakes the notions of ESOTERIC NICHE and EXOTERIC NICHE coined by Lupyan & Dale (2010) to differentiate languages with comparatively smaller populations, smaller areas, and fewer linguistic neighbours (esoteric) from those with larger populations, larger areas, and more linguistic neighbours (exoteric). A strong contribution of the paper is that it draws attention to an important shortcoming in the study of the evolution of gender systems. However, the chapter does not propose ways in which the limitations regarding the availability of data on language ecology on the databases could be surpassed.

In Chapter 4 *Why is gender so complex? Some typological considerations*, Johanna Nichols continues the discussion about the relation of grammatical gender and language complexity, testing three hypotheses about the overall complexity of languages with and without gender:

I. Languages with gender are more complex overall than those without gender.

II. Languages with gender are more complex morphologically than genderless languages.

III. Languages with gender have a higher inflectional synthesis of the verb than genderless languages (p. 74).

Nichols' work is innovative in that it measures language complexity not only through the number of elements in the inventory or the values in a system (inventory complexity) but also by the amount of information required to describe a system (descriptive complexity). By doing so, she concludes, in light of a wide cross-linguistic sample, that languages with gender are not more complex than gender-less languages.

[1] Data related to language ecology include information such as community size, degree of contact with other language communities, number of speakers learning or speaking the language non-natively, and so on.

2.2 Part II Africa

In Chapter 5, entitled *Niger-Congo "noun classes" conflate gender with deriflection*, Tom Güldemann and Ines Fiedler use a methodological approach with four analytical concepts to discuss grammatical gender in Niger-Congo: 1. agreement class; 2. gender; 3. nominal form class, and 4. DERIFLECTION – a blend of inflection and derivation which refers to a relevant morphological or phonological phenomenon interacting with gender (p. 95). Proposing an analysis of Niger-Congo languages[2] not based on Bantu systems, this paper creates descriptions that are broader than usual and suitable for the analysis of other language families outside Africa, building an interesting framework to describe and compare gender cross-linguistically.

According to Güldemann & Fiedler's results, nominal form classes[3] and agreement classes conflate in Niger-Congo gender systems. That is the main reason for problems in the analysis of this language family since the traditional Niger-Congo framework fuses two independent linguistic phenomena associated with nouns (gender agreement and deriflection) in the concept of "noun class". The strengths of this chapter lie in its analysis and recognition of nominal prefixes in Akan, which previous authors failed to relate to a nominal system. Additionally, the chapter highlights the specific developments in Bantu philology that have shaped the framework, while also considering the typological treatment of gender. It brings a cross-linguistic approach that offers valuable insights into these systems. Moreover, the chapter's context and insights help shed light on the complexities of gender and deriflection systems in the languages discussed.

In Chapter 6 *Gender in Uduk*, Don Killian focuses on the complexities and features of gender in Uduk – a Koman language spoken on the border of Ethiopia and Sudan. The paper has a special value to the studies of grammatical gender since it discusses some unusual properties of the gender system of Uduk, such as the predominant arbitrariness of gender assignment, where the features of the referents (animacy, sex) most of the time seem to be irrelevant for gender assignment. In addition, Killian's analysis of Uduk demonstrates that a non-canonical gender system can also be relatively simple. By challenging common gender assignment assumptions, this chapter shows its

[2] Namely, three Niger-Congo groups in West Africa: Akan, Guang and Ghana-Togo-Mountain.
[3] "Nominal form classes are established in the present approach by word forms with identical morphological or phonological properties; they represent the counterpart of agreement classes in the realm of morpho(phono)logy." (p. 99)

arbitrariness. It highlights verb agreement intricacies with different genders and unveils distinct gender treatment in a narrative context. Pronominal objects' indexation patterns are also explored. While insightful, further research is needed to clarify some aspects, indicating the evolving nature of linguistic analysis.

2.3 Part III New Guinea

Opening this part dedicated to New Guinea, Chapter 7 *Gender in Walman*, written by Matthew S. Dryer, describes gender and gender-like phenomena in the language of the Torricelli family spoken on the north coast of Papua New Guinea. The chapter studies language samples, revealing how PLURALIA TANTUM and diminutives share gender similarities but differ too. In this sense, pluralia tantum nouns can represent an additional gender form, coexisting with masculine and feminine. The study revealed that the corpus studied contains twice as many pluralia tantum nouns as lexically masculine, while diminutives show less gender-like traits, lacking lexically diminutive nouns.

In Chapter 8 *The gender system of Coastal Marind*, Bruno Olsson provides an overview of the gender system of Coastal Marind, a Papuan language of the Anim family of Southern New Guinea. In this language, nouns are divided into four genders: masculine, feminine and two inanimate genders. The chapter provides a detailed analysis of the different genders and their manifestations in various aspects of the language, including nouns, pronouns, adjectives, and verb agreement. The paper also explores the relationship between gender and plurals, considering the possibility that the Gender IV nouns in Walman could be seen as pluralia tantum. However, the author ultimately rejects this analysis and argues that the Gender IV nouns should be considered a separate gender. Olsson argues pluralia tantum nouns as a precursor to the fourth gender. Thus, an initial 3-gender system might have expanded to 4 genders, potentially spurred by grouping pluralia tantum nouns. The author concludes that the Coastal Marind data implies that Anim languages could represent a distinct case of gender system complexity due to a unique number-related interaction (p. 222). This paper is unique in its in-depth analysis of gender in a specific language, providing valuable insights into the structure and functioning of the gender system in Walman.

Ending this part of the volume about New Guinea, in Chapter 9 *Gender in New Guinea*, Erik Svärd classifies the gender systems of 20 languages in the New Guinea region. Using five criteria established by Di Garbo (2014)

in her analysis of gender systems of African languages,[4] Svärd identified four typologically rare characteristics of the New Guinea languages: 1. size and shape as important criteria of gender assignment, with large/long being masculine and small/short feminine, 2. the co-existence of two separate nominal classification systems, 3. no gender distinctions in pronouns, and 4. verbs as the most common indexing target (p. 225). The comparison between the gender systems of New Guinea and Africa shows that they are very different, with the main difference focusing on the prevalence of non-sex-based gender systems and gender marking on nouns in Africa, while the opposite occurs in New Guinea. Importantly, Svärd defends that pluralia tantum is a significant category of study for gaining a full understanding of gender systems in the New Guinea languages.

2.4 Part IV South Asia

Chapter 10 *Gender typology and gender (in)stability in Hindu Kush Indo-Aryan languages*, by Henrik Liljegren, takes the reader to South Asia. This ambitious contribution aims at the understanding of gender distribution and manifestation across 25 Indo-Aryan languages spoken in the Hindu Kush-Karakorum region – an area composed of mountains located in the Northeastern part of Afghanistan, Northern Pakistan, and the disputed territory of Kashmir. The article presents a micro-typology of gender systems in HKIA, summarizing related language traits (p. 282). The inherited sex-based system endures but fades in some Northwestern languages. In Western languages, an animacy-based system emerges alongside or instead of sex-based structures, impacting their setup. Gender's strength varies, being entrenched in the East but waning in the West. The North shows weaker semantic transparency, while the South emphasizes formal assignment and object agreement. An important finding of this study is the suggestion of non-trivial interactions between neighbouring languages, with languages without gender (or with markedly different assignment systems) possibly influencing the development of gender languages.

[4] The five criteria established by Di Garbo (2014) are: 1. sex-based and non-sex-based gender systems, 2. number of genders, 3. gender assignment, 4. number of gender-indexing targets, and 5. occurrence of gender marking on nouns (p. 230).

3 Volume II: World-wide comparative studies

In the second volume of *Grammatical gender and linguistic complexity*, the reader will be presented to studies that explore several geographic areas, diverse gender systems, and, with special attention, the dynamic nature of gender complexity. Three extensive chapters describe diachronic and typological case studies on gender systems, while a final chapter discusses traditional and modern theoretical and empirical challenges concerning the investigation of grammatical gender. Chapter 1, as presented in volume I, is an overview of the books' contents meant to contextualize the volumes in the field of grammatical gender studies, as well as establish the key concepts supporting the questions discussed in the papers.

In Chapter 2 *The evolving complexity of gender agreement systems*, Francesca Di Garbo & Matti Miestamo defend the necessity of taking diachrony into account to deeply understand the relationship between gender and complexity, i.e., how gender systems vary in complexity and how this variation is distributed cross-linguistically. Matasović (2018) surveys agreement systems, including gender. Being a typology, the approach is similar to Di Garbo and Miestamo's. However, in this chapter from Di Garbo and Miestamo the objective is to use a diachronic perspective to understand how and why gender systems vary in complexity, which contributes particularly to the understanding of the patterns behind the development of gender systems, and their change or loss. The study examines 36 languages in 15 sets, revealing shared traits of borrowed gender agreement patterns which emerge from borrowed nouns and adjectives, are confined to noun phrases, and rely on semantic assignment for conveying properties of nouns and natural gender distinctions (p. 44). Readers of the *Finnish Journal of Linguistics* may be particularly interested in the description of the Swedish gender system development (p. 28–30). The authors mention that many nonstandard varieties of Swedish and other Scandinavian languages maintain a tripartite gender system, which seems to be a previous stage in use before standard varieties with bipartite gender systems, such as Danish and Swedish, started to spread. Another interesting aspect of Swedish is the presence of a pervasive reduction of gender agreement morphology in Karleby Swedish – a variety spoken in the town of Karleby, located in the Finnish region of Ostrobothnia (p. 36). This fact raises the question of a possible influence coming from Finnish, which is a genderless language.

In Chapter 3 *The feminine anaphoric gender gram, incipient gender marking, maturity, and extracting anaphoric gender markers from parallel texts*, Bernhard Wälchli performs a typological investigation of feminine anaphoric gender grams (i.e., grammatical items equivalent to the words she/her in English), in a world-wide sample of 816 languages (with only 187 showing feminine anaphoric gender grams). The paper discusses how simple gender differs from more mature and genealogically stable forms of anaphoric gender. Thus, three simple forms of gender are extracted from the corpus of parallel texts from the New Testament: 1. non-compositional complex noun phrases (e.g.: *that woman*), 2. reduced nominal anaphors (e.g.: *woman*), and 3. general nouns (e.g.: *girl*; *she*). Based on his results, Wälchli concludes that anaphoric gender grams exhibiting suppletion or neutralization must have undergone some kind of grammaticalization process, presupposing earlier stages with simpler gender grams more similar to nouns or developed from markers of other grammatical categories. However, the author explains that not all cases of incipient anaphoric gender markers come from grammaticalization developments since linguistic gender categories can also be a result of language planning. As an example, he mentions that in Swedish the gender-neutral form *hen* has been used to replace the masculine *han* 'he' and the feminine *hon* 'she', especially in generic use (p. 95). One of the strengths of this chapter is Wälchli's use of parallel texts to extract anaphoric gender markers. By comparing the same biblical text translated in different languages, the author can identify patterns in the use of gender markers and provide insights into how they have evolved over time. One of the weaknesses is that the language of the chapter is highly technical, which can prevent the study from being widely applicable. Also, it lacks clear and concise summaries that would help readers understand the key takeaways.

In Chapter 4 *On the distribution and complexity of gender and numeral classifiers*, Kaius Sinnemäki applies the statistical tool of generalized linear mixed models to determine whether there is a trade-off of complexity between gender and numeral classifiers, analysing the presence vs. absence of these variables in a language. The sample contains 360 languages, stratified genealogically and areally. The use of generalized linear mixed models revealed the interesting fact that languages with numeral classifiers are significantly less likely to have gender than those with no numeral classifiers. From the total of 360 surveyed, only 22 languages (6%) attested both gender and numeral classifiers, supporting the idea that languages tend to avoid breaking the principle of economy and prefer not to develop or maintain more than one system

with the same function simultaneously. Conversely, since the use of generalized linear mixed models requires a large and representative dataset, the use of the tool in the study of gender and numeral classifiers in languages with limited data availability or small sample sizes might not be the most effective.

Finally, Chapter 5 *The dynamics of gender complexity* written by Bernhard Wälchli & Francesca Di Garbo ends this volume with a wide-ranging enquiry into the diachrony and complexity of gender systems. The paper presents state-of-the-art research on the topic, re-examining phenomena that are central to studies of gender, such as animacy hierarchy, assignment rules, agreement, and cumulative expression with other inflectional categories. In this work, grammatical gender is seen as a mature phenomenon[5] which functions as a dynamic entity evolving over time. The authors use the formula FROM X TO Y to represent the pathways in which a more mature manifestation of gender is organized in the form of noun classes. In this sense, the main point of the paper is to understand why gender can become quite complex in some languages and remain simple, or develop into being simple once again, in other languages. This dynamic nature of gender systems is the reason for the adoption of what the authors denominate as a dynamic approach to the definition of gender. Thus, considering this dynamic approach, Wälchli and Di Garbo propose the following definition for gender:

> Gender is a grammatical category type with a semantic core of animacy and/or sex reflecting classes of referents, which have a propensity to turn into classes of noun lexemes. It is overtly marked on noun-associated forms. It typically exhibits cumulative exponence with number, case, and/or person. Gender is organized in the form of systems (Volume II: 207).

In sum, one major contribution of this paper is the presentation of an alternative definition for gender in which gender systems are considered as a category centred on the semantics of animacy and/or sex, but with a propensity to turn into noun lexemes, forming what is called lexical gender. The paper describes languages in which it is plausible to assume that gender originates as referent-based items at the top segment of the animacy hierarchy, with a tendency to spread to inanimate referents due to factors such as agentivity, uniqueness or possession. In other words, gender is a category that can change from referent-based to mostly lexical items – which have the possibility of re-

[5] A linguistic feature forged in a lengthy period of historical development, presupposing a non-trivial prehistory (Dahl 2004: 2).

verting to referent-based in some cases. This definition of gender differs from other definitions in the sense that it brings light to the evolution of gender systems over time, putting the semantics of animacy and/or sex at the centre of this process of development. Additionally, the authors defend that gender should not be considered in isolation since it frequently behaves as dependent on other grammatical categories such as number, case, and person. In line with other contributions in this publication, this inquiry finds that number is particularly connected with gender in pluralia tantum and other phenomena related to lexical plurality.

4 Conclusion

This collection of papers composing the two volumes of *Grammatical gender and linguistic complexity* combines sophisticated research methods with linguistic analysis in the investigation of grammatical gender, offering researchers interested in this topic a valuable source of theoretical frameworks, linguistic data, and references for future studies. The collection of papers composing the two volumes presents studies about languages from continents which were not yet well represented in the traditional literature on gender, filling in an important gap in the research about grammatical gender. Grammatical gender is depicted in this publication as a complex system that varies cross-linguistically and can change over time due to several factors. In this sense, the dynamic nature of gender systems emerges as a central issue to be pursued by researchers on this topic.

Reading these papers can be challenging, but certainly enriching. The discussion of complex topics and the description of extensive data are well guided by the systematic division and organization of the contents. The data attached in the appendices of some of the chapters is especially useful for doctoral researchers in typological linguistics investigating gender or other nominal categories. For example, a helpful tool for novice researchers is the appendix to Chapter 5 *The dynamics of gender complexity* (Volume II: 355), which offers a long list of key concepts with brief definitions and page locations. In summary, we strongly recommend these volumes to every linguist interested in cross-linguistic studies, language diversity, language change, and especially in the evolution of gender systems.

References

Audring, Jenny. 2016. Gender. In *Oxford research encyclopedia of linguistics*. Oxford University Press.

Corbett, Greville G. 2013. *The expression of gender*. De Gruyter.

Corbett, Greville G. & Fedden, Sebastian. 2016. Canonical gender. *Journal of Linguistics* 52(3). 495–531.

Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. John Benjamins.

Di Garbo, Francesca. 2014. *Gender and its interaction with number and evaluative morphology: An intra- and intergenealogical typological survey of africa*. Stockholm University. (Doctoral dissertation).

Lupyan, Gary & Dale, Rick. 2010. Language structure is partly determined by social structure. *PLOS one* 5(1). e8559.

Matasović, Ranko. 2018. *An areal typology of agreement systems*. Cambridge University Press.

Contact information:

Patricia Carvalho Ribeiro
Faculty of Arts, Programme in Language Studies
University of Helsinki
e-mail: patricia(dot)ribeiro(at)helsinki(dot)fi

Rahel T. Dires
Faculty of Arts, Programme in Language Studies
University of Helsinki
e-mail: rahel(dot)t(dot)dires(at)helsinki(dot)fi